Journal of Experimental Child Psychology 126 (2014) 295–312

Contents lists available at ScienceDirect

Journal of Experimental Child Psychology

journal homepage: www.elsevier.com/locate/jecp

Children use visual speech to compensate for non-intact auditory speech

Susan Jerger [a,*], Markus F. Damian [b], Nancy Tye-Murray [c], Hervé Abdi [a]

[a] School of Behavioral and Brain Sciences, University of Texas at Dallas, Richardson, TX 75080, USA
[b] School of Experimental Psychology, University of Bristol, Bristol BS8 1TU, UK
[c] Department of Otolaryngology, Washington University School of Medicine, St. Louis, MO 63110, USA

ARTICLE INFO

Article history:
Received 5 September 2013
Revised 14 May 2014

Keywords:
McGurk effect
Speechreading
Children
Audiovisual speech perception
Lipreading
Development

ABSTRACT

We investigated whether visual speech fills in non-intact auditory speech (excised consonant onsets) in typically developing children from 4 to 14 years of age. Stimuli with the excised auditory onsets were presented in the audiovisual (AV) and auditory-only (AO) modes. A visual speech fill-in effect occurs when listeners experience hearing the same non-intact auditory stimulus (e.g., /-b/ag) as different depending on the presence/absence of visual speech such as hearing /bag/ in the AV mode but hearing /ag/ in the AO mode. We quantified the visual speech fill-in effect by the difference in the number of correct consonant onset responses between the modes. We found that easy visual speech cues /b/ provided greater filling in than difficult cues /g/. Only older children benefited from difficult visual speech cues, whereas all children benefited from easy visual speech cues, although 4- and 5-year-olds did not benefit as much as older children. To explore task demands, we compared results on our new task with those on the McGurk task. The influence of visual speech was uniquely associated with age and vocabulary abilities for the visual speech fill-in effect but was uniquely associated with speechreading skills for the McGurk effect. This dissociation implies that visual speech—as processed by children—is a complicated and multifaceted phenomenon underpinned by heterogeneous abilities. These results emphasize that children perceive a speaker's utterance rather than the auditory stimulus per se. In children, as in adults, there is more to speech perception than meets the ear.

* Corresponding author. Fax: +1 972 883 2491.
E-mail address: sjerger@utdallas.edu (S. Jerger).

http://dx.doi.org/10.1016/j.jecp.2014.05.003
0022-0965/© 2014 Elsevier Inc. All rights reserved.

Introduction

When adults engage in casual conversations in noisy environments, they typically understand each other without effort. Such skilled understanding in degraded soundscapes seems related to the inherently multimodal nature of speech perception, as dramatically illustrated by the classic McGurk effect (McGurk & MacDonald, 1976). In this task, an audiovisual speech stimulus with mismatched auditory and visual onsets (e.g., hearing /ba/ while seeing /ga/) is presented to participants. Adults typically perceive a mixture of the auditory and visual inputs (i.e., /da/ or /ða/). Our ability to integrate auditory and visual speech helps us to understand speech in noisy environments as well as unfamiliar content in clear environments (Arnold & Hill, 2001; MacLeod & Summerfield, 1987). Because visual speech is so useful to communication when the message is complex/degraded or the environment is noisy (e.g., classrooms), it is paramount to investigate the development of multimodal speech perception during the preschool and elementary school years.

Development of multimodal speech perception

Extant studies with multiple types of tasks report that, compared with adults, children from around 5 years of age to the preteen/teenage years show reduced sensitivity to visual speech (e.g., Desjardins, Rogers, & Werker, 1997; Erdener & Burnham, 2013; Jerger, Damian, Spence, Tye-Murray, & Abdi, 2009; McGurk & MacDonald, 1976; Ross et al., 2011; Sekiyama & Burnham, 2008; Tremblay et al., 2007; see Fort, Spinelli, Savariaux, & Kandel, 2010, for an exception regarding vowel monitoring). For example, McGurk and MacDonald (1976) noted that children were influenced by visual speech only roughly half as often as adults. Poorer sensitivity to visual speech in children has been attributed to developmental differences in articulatory proficiency or speechreading skills (Desjardins et al., 1997; Erdener & Burnham, 2013), linguistic experiences and perceptual tuning into language-specific phonemes (Erdener & Burnham, 2013; Sekiyama & Burnham, 2008), or differences in the perceptual weighting of visual speech cues (Green, 1998).

Supplementing these more specific theories, Jerger et al. (2009), who observed a U-shaped developmental function with a significant influence of congruent visual speech on phonological priming in 4- and 12-year-olds but not in 5- to 9-year-olds, adopted a dynamic systems viewpoint (Smith & Thelen, 2003). Jerger and colleagues hypothesized that children's poorer sensitivity to visual speech from 5 to 9 years of age was reflecting a period of dynamic growth as relevant perceptual, linguistic, and cognitive skills were reorganizing in response to external and internal factors. Externally, reorganization may be due to literacy instruction at around 5 or 6 years of age, during which time knowledge transmutes from phonemes as coarticulated nondistinct parts of speech into phonemes as separable distinct written and heard elements (Bryant, 1995; Burnham, 2003; Horlyck, Reid, & Burnham, 2012). Internally, reorganization may be due to phonological processes becoming sufficiently proficient (at around the same ages) to support the use of inner speech for learning, remembering, and problem solving (Conrad, 1971). The actuality of reorganization is confirmed by evoked potential studies indicating—during this age period—developmental restructuring of the lexical phonological system (Bonte & Blomert, 2004).

A dynamic systems viewpoint (Smith & Thelen, 2003) stresses two points motivating this research. First, periods of poorer sensitivity to visual speech may reflect a transition period—in contrast to a loss of ability—during which time relevant perceptual, linguistic, and cognitive resources are harder to access. Second, dynamic periods of reorganization and growth are characterized by less robust processing systems and decreases in processing efficiencies. So according to a dynamic system viewpoint, the influence of visual speech may vary as a function of task/stimulus demands for these softly reassembled resources. This suggests that tasks with less complex stimuli and/or lower task demands might not tax a child's limited processing resources as readily. This would make the harder to access resources more accessible; therefore, the child's performance might reveal greater sensitivity to visual speech. Below we introduce our new task to set up reviewing the related literature.

The new visual speech fill-in effect

Our approach assesses performance for words and nonwords with intact visual speech coupled to non-intact auditory speech (excised consonant onsets). The strategy was to insert visual speech into the gap created by the excised auditory onset to study the possibility of a visual speech fill-in effect—operationally defined by the difference in performance between the audiovisual (AV) and auditory-only (AO) modes. As an example, stimuli for the word *bag* would be as follows: (1) for AV, intact visual (/b/ag) coupled to non-intact auditory (/–b/ag); (2) for AO, static face coupled to the same non-intact auditory (/–b/ag). The visual speech fill-in effect occurs when listeners experience hearing the same auditory stimulus as different depending on the presence/absence of visual speech such as hearing /bag/ in the AV mode but hearing /ag/ in the AO mode. The AO mode also controls for the influence on performance of remaining coarticulatory cues in the stimulus. Below we review related studies that investigated the perception of AV or AO speech containing an excised segment—replaced with noise, however, instead of visual speech.

Previous studies

Studies with adults indicate that listeners report hearing AO speech with an excised segment replaced with noise as intact, a phenomenon that has been called illusory filling in, illusory continuity, auditory induction, and perceptual or phonemic restoration (Bashford & Warren, 1987; Samuel, 1981; Shahin, Bishop, & Miller, 2009; Warren, 1970). Of particular interest to the current approach—inserting visual speech into the auditory gap—three of the studies questioned whether the addition of visual speech enhances this type of illusory phenomenon. The studies investigated adults' ability to discriminate between AV speech with an excised auditory phoneme replaced with noise versus AV speech with an intact auditory phoneme with superimposed noise. Two studies (Shahin, Kerlin, Bhat, & Miller, 2012; Shahin & Miller, 2009) varied the auditory gap duration to determine whether AV congruent speech lengthened the gap duration producing illusory continuity relative to auditory plus incongruent or static visual speech. Results showed that participants perceived the stimulus as continuous—rather than interrupted—over longer gaps when they saw and heard congruent speech compared with static or incongruent speech. These investigators reasoned that congruent visual speech or visual context may have made the noise sound more speech-like or may have helped to restore the speech envelope during the interruption, thereby enhancing the illusion of continuity. The other study (Trout & Poser, 1990) investigated whether, relative to AO speech, AV speech enhanced the ability to discriminate the noise-replacing versus noise-superimposed types of stimuli when the place of articulation was easy to see (e.g., bilabial) versus difficult to see (e.g., alveolar, velar). Although results were difficult to interpret unequivocally, visual speech did not appear to influence the results. In short, the previous adult studies suggest opposing conclusions about whether visual speech influences this type of illusory phenomenon.

In contrast to the above adult studies, studies in children have focused on AO speech. Results have shown that, compared with adults, recognition of AO speech with excised phonemes replaced by noise is unusually disrupted in 5-year-olds (Newman, 2004; Walley, 1988). For Walley (1988), this indicated that children require more intact AO speech to identify even familiar words, although Walley also cautioned that the noise may have influenced performance. Newman (2004) extended her research paradigm and included a comparison of AO speech with excised phonemes replaced by noise versus silence. Children showed an adult-like advantage when noise filled the gap, and so Newman concluded that—even though children have remarkable difficulty in understanding non-intact AO speech—children show adult-like perceptual restoration. Overall, these results make it difficult to draw strong conclusions about this type of illusory phenomenon in children.

In overview, the literature does not provide clear-cut predictions about performance by children on the new visual speech fill-in task. The child literature on the development of AV speech perception, however, clearly indicates that the influence of visual speech is not as consistent and robust in children as in adults. To account for this difference, Jerger et al. (2009) theorized that the influence of visual speech on performance in children might be modulated by the information processing requirements of tasks. The current research reports two projects that systematically manipulated aspects of information processing. Study 1 assessed the effects of age, salience of visual speech cues, and lexical status on the visual speech fill-in effect. Study 2 examined the effects of task demands and child factors on visually influenced performance by comparing results on the visual speech fill-in and McGurk tasks. Below we address some relevant issues for selecting test stimuli.

Study 1

Stimuli construction issues

Speech in noise

To reduce the ceiling effect for AO speech, previous AV studies have studied AO speech in noise. The ability to identify AO speech in noise, however, does not reach adult-like performance until around 11 to 14 years of age for consonants and 10 years of age for vowels (Johnson, 2000). Thus, we reasoned that noise might have diminished the effect of visual speech on previous tasks because children had difficulty in inhibiting task-irrelevant input and resisting interference (Bjorklund & Harnishfeger, 1995). Therefore, we chose to reduce the ceiling effect for AO performance with non-intact speech.

Salience of speech

Our task focuses on onsets to evaluate whether children's performance can benefit from visual speech. Relative to the other parts of an utterance, onsets are easier to speechread, more reliable with less articulatory variability, and more stressed (Gow, Melvold, & Manuel, 1996). Phonemes are also more easily processed in the onset position; onsets reveal better accuracy on phonological awareness tasks in children (Treiman & Zukowski, 1996) and on nonword tasks in adults (Gupta, Lipinski, Abbs, & Lin, 2005). Finally, onsets are important because speech perception proceeds incrementally even in infants (Fernald, Swingley, & Pinto, 2001). These effects are compatible with theories proposing that speech input activates phonological and lexical–semantic information as it unfolds, according to the match between the evolving input and representations in memory (e.g., Marslen-Wilson & Zwitserlood, 1989).

Synchrony between visual and auditory speech onsets

Our task presents a visual consonant + vowel onset coupled to an auditory silent gap + vowel onset (our methodological criterion created a silent gap of approximately 50 ms for words and 65 ms for nonwords; see Method). A question is whether this change in the normal synchrony relation between the visual and auditory speech onsets affects AV integration. The literature suggests that it does not. Listeners normally have access to visual cues before auditory cues (Bell-Berti & Harris, 1981). Adults synthesize visual and auditory cues (without any detection of asynchrony or any effect on intelligibility) when visual speech leads auditory speech by as much as 200 ms (Grant, van Wassenhove, & Poeppel, 2004). Visual speech can lead auditory speech by as much as 180 ms without altering the McGurk effect (Munhall, Gribble, Sacco, & Ward, 1996). This literature suggests that cross-modal synchrony is probably not the basis of AV integration (Munhall & Tohkura, 1998).

Integrality between speech cues

This issue has been studied with the Garner (1974) task in which participants selectively attend to features of stimuli (e.g., consonants vs. vowels). Participants are asked to classify the stimuli on the basis of one of the features (e.g., /b/ vs. /g/), while the other feature either remains constant (control condition; /ba/ vs. /ga/) or varies irrelevantly (/ba/, /bi/ vs. /ga/, /gi/). Results have shown that irrelevant variation in vowels interferes with classifying consonants and vice versa (see, e.g., Tomiak, Mullennix, & Sawusch, 1987). Such a pattern of interaction indicates that AO speech cues are perceived in a mutually interdependent manner. A question is whether this tight coupling between AO speech cues generalizes to AV speech cues. Such a generalization is supported by a study with the Garner task (Green & Kuhl, 1989) showing that speech cues (e.g., voice onset time, place of articulation) are perceived in an interdependent manner not only when both of the cues are specified by AO speech but also when the voice onset time is specified by AO speech and the place of articulation is specified by a combination of AV speech (i.e., McGurk stimuli). These results imply that the auditory and visual speech cues of this research should be processed in an interdependent manner. Even though most participants listening to our stimuli in the AO mode report hearing a vowel onset (see Method and Results), the vowel still contains some lawful variation from being produced in a consonant–vowel–consonant context rather than in isolation. To the extent that the speech perceptual system uses lawful variation (Tomiak et al., 1987), our visual speech onset cues may be more easily grafted onto the remaining compatible visual and auditory vowel–consonant cues to yield a unified AV percept. These results imply that the auditory and visual speech cues of this study should be processed in an interdependent manner. From this viewpoint, children's performance may be more sensitive to visual speech on our task than, for example, on the McGurk effect with its non-compatible auditory and visual content. This possibility is addressed directly in Study 2. Below we predict results based on the effects of age, salience of the visual speech cues, and lexical status.

Predicted results

Age

The evidence reviewed above on the development of AV speech perception predicts that children from around 5 years of age to the preteen/teenage years will show reduced benefit from visual speech. To the extent that our new task is more sensitive to the influence of visual speech, other evidence predicts other possible age-related differences. Overall, however, the predictions are inconsistent.

Processing skills. Younger children process AO speech cues less efficiently. As an example, gating studies document that younger children require more AO input to recognize words than teenagers (Elliott, Hammer, & Evan, 1987). Younger children also have less detailed and harder to access phonological representations (Snowling & Hulme, 1994). All of this suggests that younger children may rely more on visual speech to supplement their less efficient processing of AO speech. Visual speech may also enhance performance because it (a) facilitates the detection of AO phonemes (Fort, Spinelli, Savariaux, & Kandel, 2012), (b) provides extra phonetic information that facilitates the extraction of AO cues and reduces uncertainty (Campbell, 1988; Dodd, 1977), and (c) acts as a type of alerting mechanism that benefits younger children's immature attentional skills (Campbell, 2006). Finally, younger children with less mature articulatory proficiency tend to observe visual speech more—perhaps in order to cement their knowledge of the acoustic consequences of articulatory gestures (Desjardins et al., 1997; Dodd, McIntosh, Erdener, & Burnham, 2008). These results suggest that the processing weights assigned to the auditory and visual modalities may shift with age, and younger children may show a greater influence of visual speech than older children.

Speechreading skills. If any visual speech fill-in effect depends on visual speechreading, then older participants—being more proficient—should show a greater visual speech fill-in effect than younger children.

Visual speech cues/phonotactic probabilities

The bilabial /b/ is more accurately speechread and is more common as an onset in English than the velar /g/ (Storkel & Hoover, 2010; Tye-Murray, 2009; Vitevitch & Luce, 2004; see Appendix Table A1 for phonotactic probabilities). These differences predict that the non-intact /b/ onset will be more readily restored than the non-intact /g/ onset.

Lexical status

To determine whether the availability of a lexical representation influences the visual speech fill-in effect, we compared performance for words versus nonwords (e.g., bag vs. baz). As noted previously, theories propose that speech automatically activates its lexical representation and that activation of this knowledge (a) reduces the demand for processing resources and (b) facilitates the detection of phonemes (Bouton, Cole, & Serniclaes, 2012; Fort et al., 2010; Newman & Twieg, 2001; Rubin, Turvey, & Van Gelder, 1976). Lexical status also affects visually influenced performance on the McGurk task, with the McGurk effect being more prevalent for words than for nonwords (Barutchu, Crewther, Kiely, Murphy, & Crewther, 2008) and for stimuli in which the visual stimulus forms a word and the auditory stimulus forms a nonword (Brancazio, 2004). To the extent that these results generalize to our task, we may see a greater visual speech fill-in effect for words than for nonwords. An exception to this prediction is raised, however, by the finding that lexical–semantic access requires more attentional resources in 4- and 5-year-olds than in older children (Jerger et al., 2013). This outcome predicts that the word stimuli may disproportionately drain the younger children's processing resources and reduce sensitivity to visual speech for words.

In addition to predicting results from the literature, we can predict results from theories of speech perception. A particularly relevant model for predicting the effects of lexical status is the TRACE theory of auditory speech perception with its interactive activation architecture consisting of acoustic feature, phoneme, and lexical levels (McClelland & Elman, 1986). In this model, information flows forward (from feature level to lexical level) and backward (from lexical level to feature level). Thus, the model proposes that the activation level of a phoneme is determined by activation from both the feature and lexical levels. For AV speech, Campbell (1988) proposed that visual speech adds visual features that feed forward to activate their associated phonemes. Thus, for our AV stimuli, the visual speech features corresponding to the onset would be activated and would feed forward to activate the phoneme. With regard to the words versus nonwords, the response to nonwords in the TRACE model would be based on activation only at the feature and phoneme levels in the purest sense. In this case, performance would reflect the feature and phoneme pattern of activation. If, however, the nonwords partially activate a similar lexical item that in turn supports phonological processing (see Gathercole, Willis, Emslie, & Baddeley, 1991), then performance might be driven by the lexical level as well. Our definition of the visual speech fill-in effect as the difference between performance for the AV mode and that for the AO mode should control for any lexical influences on performance that are not unique to visual speech.

Another relevant theory is the hierarchical model of speech segmentation (Mattys, White, & Melhorn, 2005), which proposes that in optimal situations listeners assign the greatest weight to lexical–semantic content. If the lexical–semantic content is compromised, listeners switch and assign the greatest weight to phonetic–phonological content. If both the lexical–semantic content and phonetic–phonological content are compromised, listeners switch and assign the greatest weight to acoustic–temporal (prosodic) content. It is also the case that monosyllabic words such as ours may activate their lexical representations without requiring phonological decomposition, whereas nonwords require phonological decomposition (Mattys, 2014). If this model generalizes to our task, children in both the AO and AV modes should assign the greatest weight to lexical–semantic content for words and to phonetic–phonological content for nonwords. To the extent that a greater weight on phonetics–phonology for nonwords increases children's attention to visual speech cues, we predict a greater visual speech fill-in effect for nonwords than for words.

Method

Participants

The children were 92 native English speakers ranging in age (years;months) from 4;2 to 14;5 (53% boys and 47% girls). The racial distribution was 87% White, 7% Asian, and 4% Black, with 9% of participants reporting Hispanic ethnicity. The children were divided into four age groups: 4- and 5-year-olds (*M* = 4;11), 6- and 7-year-olds (*M* = 6;11), 8- and 9-year-olds (*M* = 8;10), and 10- to 14-year-olds (*M* = 11;7). Each group consisted of 22 children except the 10- to 14-year-old group, which contained 26 children. Visual perception, articulatory proficiency, and hearing and vision sensitivity were within normal limits for chronological age. Other demographic characteristics are detailed in Study 2.

Materials and instrumentation

Stimuli

The stimuli were monosyllabic words and nonwords beginning with the consonant /b/ or /g/ coupled to the vowel /i/, /æ/, /ʌ/, or /o/ (see Appendix Table A1). The stimuli were recorded at the Audiovisual Recording Lab at Washington University School of Medicine. The talker was an 11-year-old trained boy actor with clearly intelligible speech without pubertal characteristics (f_0 of 203 Hz). His full facial image and upper chest were recorded. He started and ended each utterance with a neutral face/closed mouth. The color video signal was digitized at 30 frames/s with 24-bit resolution at a 720 × 480-pixel size. The auditory signal was digitized at a 48-kHz sampling rate with 16-bit amplitude resolution. The utterances were adjusted to equivalent A-weighted root mean square sound levels.

Editing the auditory onsets

To edit the auditory track, we located the /b/ or /g/ onset visually and auditorily with Adobe Premiere Pro and Soundbooth (Adobe Systems, San Jose, CA, USA) and loudspeakers. We applied a perceptual criterion to operationally define a non-intact onset. We excised the waveform in 1-ms steps from the identified auditory onset to the point in the adjacent vowel for which at least four of five trained listeners (AO mode) heard the vowel as the onset. Splice points were always at zero axis crossings. Using this perceptual criterion, we excised on average 52 ms (/b/) and 50 ms (/g/) from the word onsets and 63 ms (/b/) and 72 ms (/g/) from the nonword onsets. Performance by young untrained adults for words ($n = 10$) and for nonwords ($n = 10$) did not differ from the results presented here for the 10- to 14-year-old group. The visual track of the words and nonwords was also edited to form AV (dynamic face) versus AO (static face) modes of presentation.

AV versus AO modes

All stimuli were presented as Quicktime movie files. The AV mode consisted of the talker's still neutral face and upper chest, followed by an AV utterance of a word or nonword, followed by the talker's still neutral face and upper chest. The AO mode consisted of the same auditory track, but the visual track contained the talker's still neutral face and upper chest for the entire trial. The video track was routed to a high-resolution computer monitor, and the auditory track was routed through a speech audiometer to a loudspeaker.

Final set of items

The AV and AO modes of the word (or nonword) test items with intact and non-intact /b/ and /g/ auditory onsets were randomly intermixed and formed into lists, which also contained 14 filler items presented in the AV and AO modes. The filler items consisted of words (or nonwords) with intact not /b/ or /g/ consonant or vowel /i/, /æ/, /ʌ/, or /o/ onsets. Illustrative filler items are the word/nonword pairs of eagle/eeble, apple/apper, cheese/cheeg, and table/tavel. Thus, listeners heard trials randomly alternating between intact and non-intact auditory onsets, AV and AO modes, and test and filler items. Each test item (intact and non-intact) was presented twice in each mode. These 64 test trials were intermixed with 48 filler trials, yielding 57% test trials. The set of 112 trials was divided into four lists (presented forward or backward for eight variations). The items comprising a list varied randomly under the constraints that (a) no onset could repeat, (b) the intact and non-intact pairs [e.g., bag and (–b)ag] could not occur without at least two intervening items, (c) a non-intact onset must be followed by an intact onset, (d) the mode must alternate after three repetitions, and (e) all types of onsets (intact /b/ and /g/, non-intact /b/ and /g/, vowels, and not /b/ or /g/) needed to be dispersed uniformly throughout the lists. The number of intervening items between the intact and non-intact pairs averaged 12 items. The intensity level of the stimuli was approximately 70 dB SPL (sound pressure level). The responses of the participants were digitally recorded.

Procedure

The tester sat at a computer workstation and initiated each trial by pressing a touch pad (out of children's sight). Children, with a co-tester alongside, sat at a distance of 71 cm directly in front of an adjustable height table containing the computer monitor and loudspeaker. Their view of the talker's face subtended a visual angle of 7.17° vertically (eyebrow–chin) and 10.71° horizontally (eye level). Children completed the word/nonword repetition tasks along with other procedures in three sessions scheduled approximately 10 days apart. In the first session, children completed three of the word (or nonword) lists in separated listening conditions; in the second session, children completed the fourth word (or nonword) list and the first nonword (or word) list in separated conditions; and in the third session, children completed the remaining three nonword (or word) lists in separated conditions. The order of presentation of the words versus nonwords was counterbalanced across participants in each age group. The analyses below were collapsed across the counterbalancing conditions.

Children were instructed to repeat exactly what the talker said. We told younger children that the task was a copycat game. For the words, participants were told that they might hear words or nonwords. For the nonwords, they were told that none of the utterances would be words. Due to the multiple procedures of our protocol, children had heard the words and nonwords previously as distracters for the multimodal picture–word task (Jerger et al., 2009).

Children's utterances were transcribed independently by the tester and co-tester. For the utterances with non-intact onsets, the transcribers disagreed on 1.83% of word responses and on 2.68% of nonword responses. For responses that were in disagreement, another trained listener independently transcribed the recorded utterances. Her transcription, which always agreed with one of the other transcribers, was recorded as the response. The criteria for scoring responses to the non-intact onsets were as follows:

1. Correct vowel onsets ["ean" for "(–b)ean"] were scored as an auditory-based response for both modes.
2. Correct consonant onsets ["bag" for "(–b)ag"] were scored as a visual-based response for the AV mode and as a coarticulatory/lexical-based response for the AO mode. Visemes (visually indistinguishable phonemes) of a consonant were counted as correct for the AV mode. Viseme alternatives represented less than 1% of correct responses for both words and nonwords.
3. Incorrect vowel or consonant onsets ["dear" for "(–g)ear"] were scored as errors.

Determination of word knowledge

Parents identified each word that their children knew. For the remaining words, the word was considered as known if children could identify the word from a set of six alternatives and tell us about it. The number of unknown words identified by this approach averaged from 0.91 in the 4- and 5-year-olds to 0.00 in the 10- to 14-year-olds. All unidentified words were taught to children. The results below did not differ for taught versus previously known words.

Results

Accuracy for words and nonwords with intact or non-intact onsets

The accuracy of repeating the intact words and nonwords in the two modes for all groups was near ceiling (>96%) for the onsets and offsets (i.e., the remainder of the utterances after the onsets). The accuracy of repeating the offsets of the non-intact stimuli in the AO versus AV modes averaged 98.57% versus 98.78% (words) and 96.40% versus 94.77% (nonwords), respectively. Below we analyze the onset responses for the non-intact stimuli. The overall proportion of onset errors for the AO versus AV modes averaged 5.16% versus 2.38% (non-intact words) and 7.47% versus 2.79% (non-intact nonwords), respectively. To ensure that the responses in the children who made errors contributed equally to the group averages, the number of correct vowel and consonant onsets was normalized such that the sum of the correct responses always equaled eight. For example, if a girl had five correct vowel responses, two correct consonant responses, and one error, her normalized data were 5.71 vowel responses and 2.29 consonant responses. Our initial analysis addressed whether performance for the AO baselines differed as a function of age, lexical status, or onset.

Stimuli with non-intact onsets: AO baselines

Fig. 1 displays the proportion of correct consonant onset responses for the /b/ versus /g/ onsets of the words and nonwords in the AO mode. Results were analyzed with an analysis of variance (ANOVA) with one between-participants factor (Group: 4- and 5-year-olds vs. 6- and 7-year-olds vs. 8- and 9-year-olds vs. 10- to 14-year-olds) and two within-participants factors (Stimulus: words vs.

nonwords; Onset: /b/ vs. /g/). The overall collapsed results showed a significant age-related change with more correct consonant onset responses in the 4- and 5-year-olds than in the 10- to 14-year-olds (50% vs. 28%), $F(3,88) = 5.16$, $MSE = 12.532$, $p = .002$, partial $\eta^2 = .150$. The overall proportion of correct consonant onset responses was also significantly greater for the words than for the nonwords (51% vs. 23%) and for the /b/ onsets than for the /g/ onsets (39% vs. 34%): stimulus, $F(1,88) = 127.66$, $MSE = 9.593$, $p < .0001$, partial $\eta^2 = .592$; onset, $F(1,88) = 7.41$, $MSE = 0.758$, $p = .008$, partial $\eta^2 = .078$. Finally, the /b/ onsets showed significantly more correct consonant onset responses than the /g/ onsets for the words (58% vs. 44%) but not for the nonwords (21% vs. 24%), with a significant Stimulus × Onset interaction, $F(1,88) = 32.11$, $MSE = 1.480$, $p < .0001$, partial $\eta^2 = .267$. No other significant effects or interactions were observed.

In short, performance for the AO baselines showed more coarticulatory or lexically based responses for the words than for the nonwords and for the younger children than for the older children. To probe whether this outcome was reflecting remaining coarticulatory cues or a lexical effect, we evaluated performance for the words in children receiving the opposing counterbalancing conditions (words first vs. nonwords first). Results supported a change in the weighting of the lexical–semantic information. When only children in the nonwords first condition were considered, the proportion of consonant onset responses for the word baselines dropped to 34% (instead of 51%). If performance had been reflecting remaining coarticulatory evidence in the waveform, results for the same auditory waveform should not have changed across the counterbalancing conditions. Finally, it is possible that results for the words may have differed depending on whether the non-intact word did or did not form a new vowel onset word [(–g)ear vs. (–g)uts]. To address this possibility, we carried out an item analysis. Performance did not differ as a function of whether the non-intact words did or did not form a new vowel onset word. Overall results in Fig. 1 indicate that the AO baselines for both the words and nonwords are sufficiently below ceiling to allow us to evaluate the difference in the proportion of correct responses for the AV versus AO modes (visual speech fill-in effect).

Visual speech fill-in effect

Fig. 2 shows the difference (AV–AO) in the proportion of correct consonant onset responses for the non-intact words and nonwords as a function of the onset. Results were analyzed with an ANOVA consisting of one between-participants factor (Group: 4- and 5-year-olds vs. 6- and 7-year-olds vs. 8- and 9-year-olds vs. 10- to 14-year-olds) and two within-participants factors (Stimulus: words vs. nonwords; Onset: /b/ vs. /g/). The findings indicated a significant age-related increase in the visual speech fill-in effect, with an overall magnitude of only 6% in the 4- and 5-year-olds but 25% in the 10- to 14-year-olds, $F(3,88) = 16.13$, $MSE = 2.795$, $p < .0001$, partial $\eta^2 = .355$. The visual speech fill-in effect was

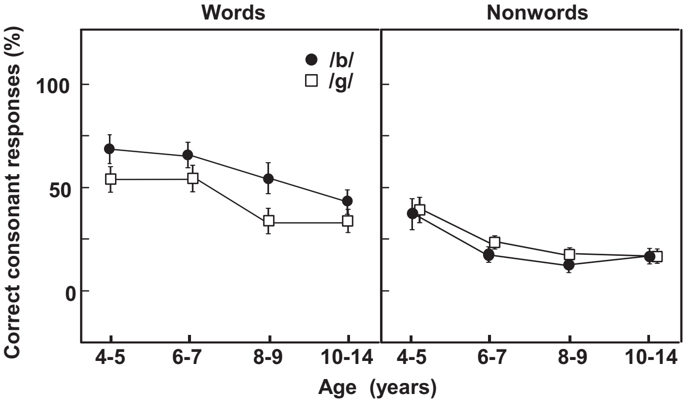

Fig. 1. Comparison of AO baselines for words and nonwords in the four age groups. A consonant onset response indicates that remaining coarticulatory information in the vowel or lexical effects formed the basis of the response.

also significantly larger for nonwords than for words and for the /b/ onsets than for the /g/ onsets, with an overall magnitude of 25% for nonwords but only 12% for words and 35% for /b/ onsets but only 2% for /g/ onsets: stimulus, $F(1,88) = 41.52$, $MSE = 2.435$, $p < .0001$, partial $\eta^2 = .321$; onset, $F(1,88) = 190.81$, $MSE = 3.189$, $p < .0001$, partial $\eta^2 = .684$.

With regard to the interactions, the /b/ onsets showed a significantly larger visual speech fill-in effect for nonwords than for words (47% vs. 22%), whereas the /g/ onsets did not (2% for both types of stimuli), producing a significant Onset × Stimulus interaction, $F(1,88) = 42.77$, $MSE = 2.235$, $p < .0001$, partial $\eta^2 = .327$. Finally, the visual speech fill-in effect showed a significantly smaller difference between the /b/ and /g/ onsets in the 4- and 5-year-olds (21%) than in the older children (approximately 36%), producing a significant Onset × Age Group interaction, $F(3,88) = 3.20$, $MSE = 3.189$, $p = .027$, partial $\eta^2 = .098$. No other significant effects or interactions were observed.

To determine whether each age group showed a significant visual speech fill-in effect, we conducted multiple t tests assessing whether each difference score differed from zero. The multiple comparison problem was controlled with the false discovery rate (FDR) procedure (Benjamini & Hochberg, 1995). Results for the nonwords showed a significant visual speech fill-in effect in all age groups for the /b/ onsets and in the 8- and 9-year-olds for the /g/ onsets. Results for the words showed a significant visual speech fill-in effect for the 6- and 7-year-olds, 8- and 9-year-olds, and 10- to 14-year-olds for the /b/ onsets. FDR results for the /g/ onsets approached significance in the 10- to 14-year-olds for both the words and nonwords.

To address whether the smaller visual speech fill-in effect for the words relative to the nonwords—with /b/ onsets—was associated with differences in the AO baselines (Fig. 1), we evaluated performance only in children in the nonwords first condition who had lower AO baseline performance (34%). Results for the /b/ onsets continued to show a significantly smaller visual speech fill-in for the words than for the nonwords, $F(1,42) = 13.65$, $MSE = 3.533$, $p = .0006$, partial $\eta^2 = .245$. Thus, the smaller visual speech fill-in effect for the words with /b/ onsets was not reflecting any limitation on performance associated with the slightly higher AO baseline seen in Fig. 1. In short, Study 1 indicated that age, lexical status, and speechreadability/phonotactic probability influenced the visual speech fill-in effect. In the next study, we investigated the effects of task demands and child factors on visually influenced performance by comparing results on the visual speech fill-in and McGurk tasks.

Study 2

As discussed earlier, a dynamic systems viewpoint (Smith & Thelen, 2003) suggests that the influence of visual speech in children may vary as a function of task demands. To probe the effect of task

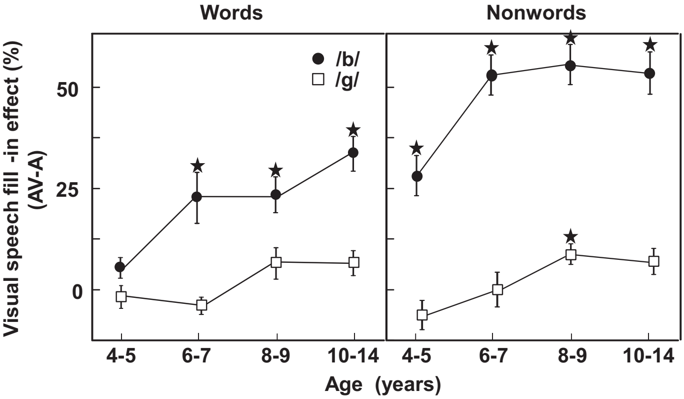

Fig. 2. Comparison of visual speech fill-in effect (difference in number of correct consonant onset responses for AV minus AO modes) for words and nonwords in the four age groups. A star indicates results that significantly differed from zero when controlling for multiple comparisons.

demands and to identify the child factors that influence performance, we compared the influence of visual speech on performance for the nonwords of our new task (naturally compatible intact visual and non-intact auditory /b/ onsets) with that of the McGurk task (conflicting auditory and visual /b/ and /g/ onsets: auditory /bʌ/ and visual /gʌ/) (McGurk & MacDonald, 1976). Below we predict results from our theories and the literature.

Predicted results

Relative weighting of auditory versus visual speech

The literature shows a shift in the relative weights of the auditory and visual modes as the quality of the input shifts. For example, children with normal hearing, listening to McGurk stimuli with lower fidelity (spectrally reduced) auditory speech, respond more on the basis of the intact visual input (Huyse, Berthommier, & Leybaert, 2013), but when the visual input is also artificially degraded the children shift and respond more on the basis of the lower fidelity auditory input. Children with normal hearing or mild to moderate hearing loss and good auditory word recognition—listening to conflicting auditory and visual inputs such as auditory /meat/ coupled to visual /street/—respond on the basis of the auditory input (Seewald, Ross, Giolas, & Yonovitz, 1985). In contrast, children whose hearing loss is more severe—and whose perceived auditory input is more degraded—respond more on the basis of the visual input. Finally, Japanese adults and children barely show a McGurk effect for intact auditory input but show a significant McGurk effect when the auditory input is degraded by noise or an unfamiliar (e.g., foreign) accent (Sekiyama & Burnham, 2008; Sekiyama & Tohkura, 1991). These results suggest that children exploit the relative quality of stimulus attributes to modulate the relative weighting of auditory and visual speech. We hypothesize that children's relative weighting of the auditory versus visual speech inputs will depend on the quality of the auditory input. With intact auditory input, such as McGurk stimuli, speech perception in children will be more auditory bound; with non-intact auditory onsets coupled to intact visual onsets, however, children's performance will depend more on visual speech. If so, performance for our task will be more sensitive to the influence of visual speech than will the McGurk stimuli.

Compatible versus incompatible AV input

A compatible AV utterance whose onsets are within the time window producing a perception of synchrony (Grant et al., 2004; Munhall et al., 1996) is more likely to be treated as a single multisensory event (Vatakis & Spence, 2007). For example, Vatakis and Spence (2007) manipulated the temporal onsets of auditory and visual inputs that were matching or not. Listeners were significantly less sensitive to temporal differences between the matched onsets. We view the silent period characterizing our non-intact auditory onsets as more harmonious with the concurrent visual speech onsets than a McGurk stimulus that has conflicting intact onsets. Finally, another seemingly relevant consideration is that AO and AV consonant–vowel stimuli are processed in an interdependent manner on the Garner (1974) task by adults (Tomiak et al., 1987) and children (Jerger et al., 1993). The latter study assessed performance for other types of speech cues with the Garner task in individuals from 3 to 79 years of age and found interdependent processing at all ages. To the extent that these results generalize to our task, results on the Garner task further suggest that our auditory and visual onsets should be processed interdependently. Overall, these data suggest that listeners will more likely show a greater influence of visual speech on our task than on the McGurk task.

Child characteristics underpinning task performance

Speechreadability

Study 1 indicates that children have difficulty in speechreading the /g/ onset, a finding that agrees with the literature (Tye-Murray, 2009). This suggests that listeners will likely show a greater influence of visual speech on our task, particularly for the /b/ visual onset, than on the McGurk task with its /g/

visual onset. To the extent that performance is reflecting speechreading, older children with better speechreading skills should show greater sensitivity to visual speech.

Language

In relation to the TRACE model discussed previously, older children may benefit more from visual speech because they have more robust detailed phonological and lexical representations (Snowling & Hulme, 1994) that are more easily activated by sensory input. Furthermore, older children—having stronger language skills—may be able to use the activation pattern across the feature, phoneme, and word levels better than younger children. Thus, children with more mature language skills may show greater sensitivity to the influence of visual speech.

Method

Participants

Participants were the 92 children of Study 1. Table 1 summarizes their average ages, vocabulary skills, and visual speechreading skills. Vocabulary skills were within normal limits for all groups. To quantify visual speechreading ability in the following regression analyses, we used the results scored by word onsets.

Stimuli and procedure

Receptive vocabulary skills were estimated with the Peabody Picture Vocabulary Test–Fourth Edition (Dunn & Dunn, 2007); children heard a word and pointed to the picture—out of four alternatives—illustrating that word's meaning. Speechreading skills were estimated with the Children's Audio–Visual Enhancement Test (Tye-Murray & Geers, 2001); children repeated words presented in the AO and visual-only (VO) modes. The stimuli for the visual speech fill-in task were the nonwords with /b/ onsets. The stimuli for the McGurk task (McGurk & MacDonald, 1976) were /bʌ/ and /gʌ/ utterances recorded at the same Audiovisual Recording Lab by the same talker described above. The auditory track of /bʌ/ was combined with the visual track of /gʌ/ with Adobe Premiere Pro and Soundbooth. The auditory and visual utterances were aligned during the release of the consonant. The final McGurk stimulus (auditory /bʌ/–visual /gʌ/) was presented as a Quicktime movie with the talker's still neutral face and upper chest, followed by the AV utterance of the stimulus, followed by the talker's still neutral face and upper chest. We appended the McGurk stimulus to the end of each of the nonword lists described above. The influence of visual speech was quantified (a) by the difference in the number of correct consonant onset responses for the AV versus AO modes for the visual speech fill-in task ($n = 8$) and (b) by the absolute number of visually influenced responses for the McGurk task ($n = 4$). These derived measures were tallied as reflecting the influence of visual speech if results showed (a) a visual speech fill-in effect of at least 25% (difference score $\geqslant 2$) or (b) at least 25% visually based responses for the McGurk trials ($n \geqslant 1$).

Table 1
Average ages, vocabulary skills, and speechreading abilities (and standard deviations) in the four age groups of children ($N = 92$).

Measure	Age group (years)			
	4–5	6–7	8–9	10–14
Age (years;months)	4;11 (0.73)	6;11 (0.67)	8;10 (0.68)	11;70 (1.38)
Vocabulary (standard score)	120.48 (10.29)	116.75 (13.02)	121.82 (12.11)	121.35 (11.72)
Speechreading: visual only[a] (% correct)				
Scored by words	4.86 (5.35)	9.97 (8.09)	17.72 (13.72)	24.59 (12.43)
Scored by word onsets[b]	46.06 (21.30)	53.73 (20.53)	66.14 (14.43)	72.86 (13.07)

[a] AO results were at ceiling.
[b] Onsets were scored with visemes counted as correct (e.g., p̲at for b̲at).

Results

Comparison of visual speech fill-in versus McGurk effects

Visual speech influenced performance in 82% of children for the visual speech fill-in task and 52% of children for the McGurk task. The visually influenced McGurk responses in children consisted of /dʌ/, /ðʌ/, and /gʌ/. Of the children showing an influence of visual speech on only one task, 10% showed only a McGurk effect but 39% showed only a visual speech fill-in effect. We conducted a multiple regression analysis for each task with predictor variables of age, vocabulary skill, and visual speechreading skill and a criterion variable of the quantified effect of visual speech. The intercorrelations among this set were .526 (age and visual speechreading), .105 (age and vocabulary), and −.058 (vocabulary and visual speechreading).

Table 2 summarizes the regression results, with the multiple correlation coefficients and omnibus *F* statistics for all of the variables considered simultaneously followed by the part (also called semipartial) correlation coefficients and the partial *F* statistics evaluating the variation in performance uniquely accounted for (after removing the influence of the other variables) by each individual variable (Abdi, Edelman, Valentin, & Dowling, 2009). The set of variables significantly predicted the influence of visual speech on performance, with children's ages, vocabulary skills, and visual speechreading skills together predicting approximately 17% or 18% of the variance in performance for each task. Part correlations (expressing the unique influence of each variable) indicated that the visual speech fill-in effect varied significantly as a function of children's ages and vocabulary skills, each uniquely accounting for approximately 7% or 8% of the variance in performance. In contrast, the influence of visual speech on McGurk performance varied significantly only as a function of children's speechreading skills, which uniquely accounted for approximately 13% of the variance in performance.

In short, the influence of visual speech on the visual speech fill-in and McGurk tasks appears to be underpinned by dissociable independent variables. A possible caveat is that results for the visual speech fill-in and McGurk tasks may have been influenced by the different baseline levels of performance. Thus, we analyzed the visual speech fill-in effect for the nonwords with /g/ onsets (for which visual speech also exerted a lesser influence on performance; see Study 1). The set of variables significantly predicted the visual speech fill-in effect and accounted for 9% of the variance in performance, $R = .346$, $F(3,88) = 3.89$, $p = .012$. Part correlations indicated that the visual speech fill-in effect varied significantly as a unique function of children's age, $F(1,88) = 5.69$, $p = .019$, with neither their vocabulary skills ($p = .12$) nor their speechreading skills ($p = .639$) achieving significance. These data support the conclusion that age and vocabulary skills underlie the visual speech fill-in effect, but speechreading skills underlie the McGurk effect.

Discussion

In this research, Study 1 assessed the effects of age, salience of visual speech cues, and lexical status on the new visual speech fill-in effect, and Study 2 assessed the effects of task/stimulus demands and

Table 2
Multiple correlation coefficients and omnibus *F*s for all of the variables considered simultaneously followed by the part correlation coefficients and the partial *F* statistics evaluating the variation in performance uniquely accounted for (after removing the influence of the other variables) by age, vocabulary skills, or visual speechreading skills for onsets.

Variable	Visual speech fill-in effect /b/ onsets			McGurk effect		
	Multiple *R*	Omnibus *F*	*p*	Multiple *R*	Omnibus *F*	*p*
All	.421*	6.16	.001	.407*	5.71	.001
	Part *r*	Partial *F*	*p*	Part *r*	Partial *F*	*p*
Age	.288*	8.64	.004	.000	0.05	*ns*
Vocabulary skills	.261*	7.13	.009	.045	0.17	*ns*
Visual speechreading (onsets)	.032	0.03	*ns*	.355*	13.03	.001

Note. The influence of visual speech was quantified by (a) the difference in the number of correct consonant onset responses for the AV–AO modes for the visual speech fill-in effect and (b) the number of visually influenced responses for the McGurk effect. *ns*, not significant; *df* = 3, 88 for omnibus *F* and *df* = 1, 88 for partial *F*. The asterisk (*) indicates statistically significant result.

child factors by comparing results on the visual speech fill-in and McGurk tasks. Results of Study 1 showed—contrary to previous findings—that children from 4 to 14 years of age significantly benefited from visual speech. However, this benefit critically depended on task/stimulus properties and individual factors. With regard to age, results for nonwords showed a significant visual speech fill-in effect in all age groups (4–14 years) for easy visual speech cues (/b/) and in the 8- and 9-year-olds for difficult visual speech cues (/g/). Results for the words showed a significant visual speech fill-in effect in all age groups except the 4- and 5-year-olds for the easy visual speech cues. Results for both the nonwords and words approached significance for the difficult visual speech cues (/g/) in the 10- to 14-year-olds when controlling for multiple comparisons. These results disagree with predictions that younger children with less mature linguistic and processing skills will rely on visual speech to a greater extent than older children. The results do, however, support the prediction that word stimuli disproportionately drain processing resources in 4- and 5-year-olds and reduce sensitivity to visual speech because lexical–semantic access requires more attentional resources in these younger children than in older children (Jerger et al., 2013).

Our finding of an age-related increase in children's sensitivity to visual speech agrees with the literature in general; however, our results document an influence of visual speech at much younger ages than previously observed (e.g., Desjardins et al., 1997; Erdener & Burnham, 2013; Ross et al., 2011; Sekiyama & Burnham, 2008; Tremblay et al., 2007). The current developmental functions are not, however, consistent with the U-shaped developmental functions observed on the multimodal picture–word (phonological priming) task (Jerger et al., 2009), which showed a significant influence of congruent visual speech in 4- and 12-year-olds but not in 5- to 9-year-olds. Clearly, sensitivity to visual speech varies dramatically as a function of task/stimulus demands.

As implied in the developmental patterns above, other critical task/stimulus properties were the salience of the visual speech cues and semantic content. The visual speech fill-in effect was significantly larger for easy speech cues than for difficult speech cues and for nonwords than for words. The latter results imply that children's sensitivity to visual speech may vary depending on their relative weighting and decomposition of the phonetic–phonological content in agreement with the proposals of the hierarchical model of speech segmentation (Mattys et al., 2005). Our results also suggest that compatible AV nonwords are more optimal stimuli than words when research goals are concerned with assessing children's sensitivity to visual speech. That said, words are clearly the building blocks of language and remain essential stimuli for studying the contributions of visual speech to children's communicative abilities.

Results of Study 2 showed that the McGurk task significantly underestimates children's sensitivity to visual speech compared with the visual speech fill-in task. Visually influenced performance was revealed in approximately 80% of children on the visual speech fill-in task but in only approximately 50% of children on the McGurk task. In addition, 39% of children showed only a visual speech fill-in effect. The relatively greater sensitivity to visual speech of our task may be reflecting previous findings indicating that a compatible AV utterance is more likely to be viewed as a single multisensory event and more likely to be synthesized and, thus, processed in an interdependent manner (Tomiak et al., 1987; Vatakis & Spence, 2007). The current outcome is also consistent with the research reporting that children's relative weighting of auditory and visual inputs depends on the quality of the input—with speech perception for intact auditory input such as McGurk stimuli more auditory bound (e.g., Huyse et al., 2013; Seewald et al., 1985; Sekiyama & Burnham, 2008). Importantly, Study 2 also documented that sensitivity to visual speech varies in the same children depending on the task/stimulus demands.

Finally, and perhaps most important, is the finding that age and vocabulary uniquely determine visually influenced performance for the visual speech fill-in task, whereas speechreading skills uniquely determine visually influenced performance on the McGurk task. Findings for the McGurk task agree with previous investigators who proposed that greater sensitivity to visual speech reflects greater speechreading skills (Erdener & Burnham, 2013; see Jerger et al., 2009, for opposing evidence regarding phonological priming). Findings for the visual speech fill-in effect may agree with previous investigators who proposed that greater sensitivity to visual speech is due to a heightened focus on the phonemic distinctions of one's native language (e.g., Erdener & Burnham, 2013). To the extent that phonological skills are critical to how well children learn vocabulary (e.g., Gathercole & Baddeley, 1989), our vocabulary association may be compatible with Erdener and Burnham's (2013) native phonology association, with both effects representing skills that are contingent on more basic phonological mechanisms.

The association between language skills and the visual speech fill-in effect may also explain why younger children benefit less from visual speech. Again, the TRACE theory of AO speech perception proposes that input activates acoustic feature, phoneme, and lexical levels (McClelland & Elman, 1986). For AV speech, Campbell (1988) proposed that visual speech activates visual features and their associated phonemes, and so older children may benefit more from visual speech because they have more robust detailed phonological and lexical representations (Snowling & Hulme, 1994) that are more easily activated by sensory input. Furthermore, the older children—having stronger language skills—may be able to use the activation pattern across the feature, phoneme, and word levels better than the younger children.

Finally, Study 2 documents that benefit from visual speech is not a unitary phenomenon. The McGurk effect was affected by one child factor (speechreading) but not by the other two (age and vocabulary), whereas the visual speech fill-in effect was affected by two child factors (age and vocabulary) but not by the other one (speechreading). This dissociation implies that benefit from visual speech is a complicated multifaceted phenomenon underpinned by heterogeneous abilities.

In conclusion, these results show that—under some conditions—preschoolers and elementary school-aged children benefit from visual speech during multimodal speech perception. Importantly, our new task extends the range of measures that can be used to assess visual speech processing by children and provides results critical to integrating visual speech into our theories of speech perceptual development. We conceptualize the new visual speech fill-in effect as tapping into a perceptual process that may enhance the accuracy of speech perception in everyday listening conditions. These results emphasize that children experience hearing a speaker's utterance rather than the auditory stimulus per se. In children, as in adults, there is more to speech perception than meets the ear.

Acknowledgments

This research was supported by the National Institute on Deafness and Other Communication Disorders (DC-00421). Sincere appreciation goes to the speech science colleagues who advised us to adopt a perceptual criterion for editing the non-intact stimuli. We thank Brent Spar for recording the audiovisual stimuli. We thank the children and parents who participated and the research staff who assisted, namely Aisha Aguilera, Carissa Dees, Nina Dinh, Nadia Dunkerton, Alycia Elkins, Brittany Hernandez, Cassandra Karl, Demi Krieger, Michelle McNeal, Jeffrey Okonye, Rachel Parra, and Kimberly Periman of the University of Texas at Dallas (data collection, analysis, and presentation) as well as Derek Hammons and Scott Hawkins of the University of Texas at Dallas and Brent Spehar of Central Institute for the Deaf (CID) – Washington University (computer programming).

Appendix Table A1

Word and nonword items that were constructed to have as comparable phonotactic probabilities as possible

Word	Nonword	Segment frequency		Biphone frequency	
		Word	Nonword	Word	Nonword
(A) Adult values (Vitevitch & Luce, 2004)					
bag	baz	.1486	.1507	.0087	.0066
bean	beece	.1791	.1619	.0053	.0040
bone	bohs	.1996	.1793	.0045	.0047
bus	buhl	.1692	.1641	.0073	.0080

(*continued on next page*)

Appendix Table A1 (*continued*)

Word	Nonword	Segment frequency		Biphone frequency	
		Word	Nonword	Word	Nonword
	Average	*.1741*	*.1640*	*.0064*	*.0058*
gap	gak	.1425	.1589	.0081	.0086
gear	Geen	.1361	.1538	.0001	.0031
gold	Gothd	.1181	.1187	.0015	.0016
guts	Guks	.1813	.1687	.0048	.0084
	Average	*.1445*	*.1500*	*.0036*	*.0054*
Overall	Average	.1593	.1570	.0050	.0056
(B) Child values (Storkel & Hoover, 2010)					
bag	baz	.1776	.1827	.0108	.0089
bean	beece	.2244	.1813	.0082	.0063
bone	bohs	.2391	.1961	.0080	.0073
bus	buv	.2003	.1601	.0110	.0089
	Average	.2103	.1801	.0095	.0079
gap	gak	.1481	.1763	.0063	.0088
gear	geen	.1642	.1816	.0008	.0044
gold	gothd	.1580	.1568	.0031	.0030
guts	guks	.2170	.2094	.0082	.0120
	Average	.1718	.1810	.0046	.0054
Overall	Average	.1911	.1805	.0071	.0074

Note. The positional segment frequency is the sum of the likelihoods of occurrence of each phoneme in its position within the utterance; the biphone frequency is the sum of the likelihoods of co-occurrence of each two adjacent phonemes.

References

Abdi, H., Edelman, B., Valentin, D., & Dowling, W. (2009). *Experimental design and analysis for psychology*. New York: Oxford University Press.

Arnold, P., & Hill, F. (2001). Bisensory augmentation: A speechreading advantage when speech is clearly audible and intact. *British Journal of Psychology, 92*, 339–355.

Barutchu, A., Crewther, S., Kiely, P., Murphy, M., & Crewther, D. (2008). When /b/ill with /g/ill becomes /d/ill: Evidence for a lexical effect in audiovisual speech perception. *European Journal of Cognitive Psychology, 20*, 1–11.

Bashford, J., & Warren, R. (1987). Multiple phonemic restorations follow the rules for auditory induction. *Perception & Psychophysics, 42*, 114–121.

Bell-Berti, F., & Harris, K. (1981). A temporal model of speech production. *Phonetica, 38*, 9–20.

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society B: Methodological, 57*, 289–300.

Bjorklund, D., & Harnishfeger, K. (1995). The evolution of inhibition mechanisms and their role in human cognition and behavior. In F. Dempster & C. Brainerd (Eds.), *Interference and inhibition in cognition* (pp. 141–173). San Diego: Academic Press.

Bonte, M., & Blomert, L. (2004). Developmental changes in ERP correlates of spoken word recognition during early school years: A phonological priming study. *Clinical Neurophysiology, 115*, 409–423.

Bouton, S., Cole, P., & Serniclaes, W. (2012). The influence of lexical knowledge on phoneme discrimination in deaf children with cochlear implants. *Speech Communication, 54*, 189–198.

Brancazio, L. (2004). Lexical influences in audiovisual speech perception. *Journal of Experimental Psychology: Human Perception and Performance, 30*, 445–463.

Bryant, P. (1995). Phonological and grammatical skills in learning to read. In B. de Gelder & J. Morais (Eds.), *Speech and reading: A comparative approach* (pp. 249–266). Hove, UK: Psychology Press.

Burnham, D. (2003). Language specific speech perception and the onset of reading. *Reading and Writing, 16*, 573–609.

Campbell, R. (1988). Tracing lip movements: Making speech visible. *Visible Language, 22*, 32–57.

Campbell, R. (2006). Audio-visual speech processing. In K. Brown, A. Anderson, L. Bauer, M. Berns, G. Hirst, & J. Miller (Eds.), *The encyclopedia of language and linguistics* (pp. 562–569). Amsterdam: Elsevier.

Conrad, R. (1971). The chronology of the development of covert speech in children. *Developmental Psychology, 5*, 398–405.

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

Dodd, B. (1977). The role of vision in the perception of speech. *Perception, 6*, 31–40.

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

Dunn, L., & Dunn, D. (2007). *The peabody picture vocabulary test-IV* (4th ed.). Minneapolis, MN: NCS Pearson.

Elliott, L., Hammer, M., & Evan, K. (1987). Perception of gated, highly familiar spoken monosyllabic nouns by children, teenagers, and older adults. *Perception and Psychophysics, 42*, 150–157.

Erdener, D., & Burnham, D. (2013). The relationship between auditory-visual speech perception and language-specific speech perception at the onset of reading instruction in English-speaking children. *Journal of Experimental Child Psychology, 114*, 120–138.

Fernald, A., Swingley, D., & Pinto, J. (2001). When half a word is enough: Infants can recognize spoken words using partial phonetic information. *Child Development, 72*, 1003–1015.

Fort, M., Spinelli, E., Savariaux, C., & Kandel, S. (2010). The word superiority effect in audiovisual speech perception. *Speech Communication, 52*, 525–532.

Fort, M., Spinelli, E., Savariaux, C., & Kandel, S. (2012). Audiovisual vowel monitoring and the word superiority effect in children. *International Journal of Behavioral Development, 36*, 457–467.

Garner, W. (1974). *The processing of information and structure*. Potomac, MD: Lawrence Erlbaum.

Gathercole, S., & Baddeley, A. (1989). Evaluation of the role of phonological STM in the development of vocabulary in children: A longitudinal study. *Journal of Memory and Language, 28*, 200–213.

Gathercole, S., Willis, C., Emslie, H., & Baddeley, A. (1991). The influence of number of syllables and wordlikeness on children's repetition of nonwords. *Applied Psycholinguistics, 12*, 349–367.

***Gow, D., Melvold, J., & Manuel, S. (1996). How word onsets drive lexical access and segmentation: Evidence from acoustics, phonology, and processing. In Proceedings of the Fourth International Conference on Spoken Language Processing (Vol. 1, pp. 66–69). Philadelphia: IEEE.

Grant, K., van Wassenhove, V., & Poeppel, D. (2004). Detection of auditory (cross-spectral) and auditory-visual (cross-modal) synchrony. *Speech Communication, 44*, 43–53.

Green, K. (1998). The use of auditory and visual information during phonetic processing: Implications for theories of speech perception. In R. Campbell, B. Dodd, & D. Burnham (Eds.), *Hearing by eye II: Advances in the psychology of speechreading and auditory-visual speech*. Hove, UK: Taylor & Francis.

Green, K., & Kuhl, P. (1989). The role of visual information in the processing of place and manner features in speech perception. *Perception & Psychophysics, 45*, 34–42.

Gupta, P., Lipinski, J., Abbs, B., & Lin, P. H. (2005). Serial position effects in nonword repetition. *Journal of Memory and Language, 53*, 141–162.

Horlyck, S., Reid, A., & Burnham, D. (2012). The relationship between learning to read and language-specific speech perception: Maturation versus experience. *Scientific Studies of Reading, 16*, 218–239.

Huyse, A., Berthommier, F., & Leybaert, J. (2013). Degradation of labial information modifies audiovisual speech perception in cochlear-implanted children. *Ear and Hearing, 34*, 110–121.

Jerger, S., Damian, M., Mills, C., Bartlett, J., Tye-Murray, N., & Abdi, H. (2013). Effect of perceptual load on semantic access by speech in children. *Journal of Speech, Language, and Hearing Research, 56*, 388–403.

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

Jerger, S., Pirozzolo, F., Jerger, J., Elizondo, R., Desai, S., Wright, E., et al (1993). Developmental trends in the interaction between auditory and linguistic processing. *Perception & Psychophysics, 54*, 310–320.

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

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

Marslen-Wilson, W., & Zwitserlood, P. (1989). Accessing spoken words: The importance of word onsets. *Journal of Experimental Psychology: Human Perception and Performance, 15*, 576–585.

Mattys, S. (2014). Speech perception. In D. Reisberg (Ed.), *The Oxford handbook of cognitive psychology* (pp. 391–412). Oxford, UK: Oxford University Press.

Mattys, S., White, L., & Melhorn, J. (2005). Integration of multiple speech segmentation cues: A hierarchical framework. *Journal of Experimental Psychology: General, 134*, 477–500.

McClelland, J., & Elman, J. (1986). The TRACE model of speech perception. *Cognitive Psychology, 18*, 1–86.

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

Munhall, K., Gribble, P., Sacco, L., & Ward, M. (1996). Temporal constraints on the McGurk effect. *Perception & Psychophysics, 58*, 351–362.

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

Newman, R. (2004). Perceptual restoration in children versus adults. *Applied Psycholinguistics, 25*, 481–493.

Newman, S., & Twieg, D. (2001). Differences in auditory processing of words and pseudowords: An fMRI study. *Human Brain Mapping, 14*, 39–47.

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

Rubin, P., Turvey, M. T., & Van Gelder, P. (1976). Initial phonemes are detected faster in spoken words than in spoken nonwords. *Perception & Psychophysics, 19*, 394–398.

Samuel, A. G. (1981). Phonemic restoration: Insights from a new methodology. *Journal of Experimental Psychology: General, 110*, 474–494.

Seewald, R. C., Ross, M., Giolas, T. G., & Yonovitz, A. (1985). Primary modality for speech perception in children with normal and impaired hearing. *Journal of Speech and Hearing Research, 28*, 36–46.

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

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

Shahin, A., Bishop, C., & Miller, L. (2009). Neural mechanisms for illusory filling-in of degraded speech. *NeuroImage, 44*, 1133–1143.

Shahin, A., Kerlin, J., Bhat, J., & Miller, L. (2012). Neural restoration of degraded audiovisual speech. *NeuroImage, 60*, 530–538.

Shahin, A., & Miller, L. (2009). Multisensory integration enhances phonemic restoration. *Journal of the Acoustical Society of America, 125*, 1744–1750.

Smith, L., & Thelen, E. (2003). Development as a dynamic system. *Trends in Cognitive Sciences, 7*, 343–348.

Snowling, M., & Hulme, C. (1994). The development of phonological skills. *Philosophical Transactions of the Royal Society of London B: Biological Sciences, 346*, 21–27.

Storkel, H., & Hoover, J. (2010). An on-line calculator to compute phonotactic probability and neighborhood density based on child corpora of spoken American English. *Behavior Research Methods, 42*, 497–506.

Tomiak, G., Mullennix, J., & Sawusch, J. (1987). Integral processing of phonemes: Evidence for a phonetic mode of perception. *Journal of the Acoustical Society of America, 81*, 755–764.

Treiman, R., & Zukowski, A. (1996). Children's sensitivity to syllables, onsets, rimes, and phonemes. *Journal of Experimental Child Psychology, 62*, 432–455.

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

Trout, J. D., & Poser, W. J. (1990). Auditory and visual influences on phonemic restoration. *Language and Speech, 33*, 121–135.

Tye-Murray, N. (2009). *Foundations of aural rehabilitation: Children, adults, and their family members* (3rd ed.). San Diego: Singular Publishing.

Tye-Murray, N., & Geers, A. (2001). *Children's audio-visual enhancement test*. St. Louis, MO: Central Institute for the Deaf.

Vatakis, A., & Spence, C. (2007). Crossmodal binding: Evaluating the "unity assumption" using audiovisual speech stimuli. *Perception & Psychophysics, 69*, 744–756.

Vitevitch, M., & Luce, P. (2004). A web-based interface to calculate phonotactic probability for words and nonwords in English. *Behavior Research Methods, Instruments, & Computers, 36*, 481–487.

Walley, A. C. (1988). Spoken word recognition by young children and adults. *Cognitive Development, 3*, 137–165.

Warren, R. M. (1970). Perceptual restoration of missing speech sounds. *Science, 167*, 392–393.