Journal of Experimental Child Psychology 102 (2009) 40–59

Contents lists available at ScienceDirect

Journal of Experimental Child Psychology

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

Developmental shifts in children's sensitivity to visual speech: A new multimodal picture–word task

Susan Jerger [a,b,c,*], Markus F. Damian [d], Melanie J. Spence [a,b], Nancy Tye-Murray [a,c], Herve Abdi [a]

[a] School of Behavioral and Brain Sciences, University of Texas at Dallas, Richardson, TX 75083, USA
[b] Callier Center for Communication Disorders, University of Texas at Dallas, Dallas, TX 75235, USA
[c] Central Institute for the Deaf and Washington University School of Medicine, St. Louis, MO 63110, USA
[d] Department of Experimental Psychology, University of Bristol, Bristol BS8 1TU, UK

ARTICLE INFO

Article history:
Received 23 September 2007
Revised 11 August 2008
Available online 1 October 2008

Keywords:
Picture–word task
Audiovisual speech perception
U-shaped developmental function
Phonological processing
Picture word interference
Picture naming
Multimodal speech processing
Dynamic systems theory

ABSTRACT

This research developed a multimodal picture–word task for assessing the influence of visual speech on phonological processing by 100 children between 4 and 14 years of age. We assessed how manipulation of seemingly to-be-ignored auditory (A) and audiovisual (AV) phonological distractors affected picture naming without participants consciously trying to respond to the manipulation. Results varied in complex ways as a function of age and type and modality of distractors. Results for congruent AV distractors yielded an inverted U-shaped function with a significant influence of visual speech in 4-year-olds and 10- to 14-year-olds but not in 5- to 9-year-olds. In concert with dynamic systems theory, we proposed that the temporary loss of sensitivity to visual speech was reflecting reorganization of relevant knowledge and processing subsystems, particularly phonology. We speculated that reorganization may be associated with (a) formal literacy instruction and (b) developmental changes in multimodal processing and auditory perceptual, linguistic, and cognitive skills.

Introduction

Speech communication by adults is naturally a multimodal event with auditory and visual speech integrated mandatorily. This basic property of mature speech perception is illustrated dramatically by

* Corresponding author. Address: School of Behavioral and Brain Sciences, University of Texas at Dallas, Richardson, TX 75083, USA. Fax: +1 972 883 2491.
E-mail address: sjerger@utdallas.edu (S. Jerger).

0022-0965/$ - see front matter © 2008 Elsevier Inc. All rights reserved.
doi:10.1016/j.jecp.2008.08.002

McGurk effects (McGurk & MacDonald, 1976). In the McGurk task, individuals hear a syllable whose onset has one place of articulation while seeing a talker simultaneously mouthing a syllable whose onset has a different place of articulation (e.g., auditory /ba/ and visual /ga/). Adults typically experience the illusion of perceiving /da/ or /ða/, a blend of the auditory and visual inputs. The McGurk illusion is consistent with the idea that auditory and visual speech interact prior to the classification of phonetic features such as place of articulation (Green, 1998). Integrating auditory and visual speech without conscious effort clearly has adaptive value. Seeing a talker's face facilitates listening in noisy soundscapes and in clear environments containing unfamiliar or complex content (Arnold & Hill, 2001; MacLeod & Summerfield, 1987; Massaro, 1998).

Multimodal speech perception in infants and children

In contrast to performance in adults, multimodal speech perception in children is not well understood. The literature suggests that, at least in some respects, infants are more inclined than children to perceive speech multimodally. For example, infants demonstrate multimodal integration (Burnham & Dodd, 2004; Rosenblum, Schmuckler, & Johnson, 1997). When infants are habituated to a McGurk-like stimulus (auditory /ba/ and visual /ga/) and then presented with either an auditory /ba/, /da/ or /ða/, they respond as if the auditory /ba/ is unfamiliar and the integrated percepts of /da/ or /ða/ are familiar. Infants also detect equivalent phonetic information in auditory and visual speech (Kuhl & Meltzoff, 1982; Patterson & Werker, 1999, 2003). When infants hear a vowel while watching side-by-side images of two talkers, one mouthing the heard vowel and one mouthing a different vowel, they look significantly longer at the talker whose articulatory movements match the heard speech. Such findings suggest that the correspondences between auditory and visual speech are recognized without extensive perceptual–motor experience. A complication to these findings has developed recently with the observation of inconsistent results in infants dependent on testing and stimulus conditions (Desjardins & Werker, 2004). Nonetheless, the evidence as a whole continues to suggest that visual speech may play an important role in learning the phonological structure of spoken language (Dodd, 1979, 1987; Locke, 1993; Mills, 1987; Weikum et al., 2007).

In contrast to the infant and adult literatures, the child literature emphasizes that visual speech has less influence on speech perception by children. In their initial research, McGurk and MacDonald (1976) noted that significantly fewer children than adults show an influence of visual speech on perception. In response to one type of McGurk stimulus (auditory /ba/–visual /ga/), the percentage of individuals who reported hearing /ba/ (auditory capture) was 40% to 60% of children but only 10% of adults. This pattern of results has been replicated and extended to other tasks (Desjardins, Rogers, & Werker, 1997; Dupont, Aubin, & Menard, 2005; Hockley & Polka, 1994; Massaro, 1984; Massaro, Thompson, Barron, & Laren, 1986; Sekiyama & Burnham, 2004; Wightman, Kistler, & Brungart, 2006). Overall results are consistent with the idea that performance is dominated by auditory input in children and visual input in adults, agreeing with the general observation of a bias toward the auditory modality in young children (Sloutsky & Napolitano, 2003).

Children's visual speech perception improves with increasing age, but the time course of developmental change is not well understood. A few studies have observed benefit from visual speech by the preteen/teenage years (Conrad, 1977; Dodd, 1977, 1980; Hockley & Polka, 1994), with one report citing an earlier age of 8 years (Sekiyama & Burnham, 2004). Developmental improvement has been attributed to experience in producing speech, changes in the emphasis and perceptual weight given to visual speech cues, and age-related advances in speechreading skills and/or linguistic skills, perhaps as a consequence of educational training (Desjardins et al., 1997; Green, 1998; Massaro et al., 1986; Sekiyama & Burnham, 2004).

Age-related effects versus task demand effects

The nature and extent of audiovisual speech perception appear to differ in children versus infants and adults. Some investigators have cautioned, however, that the observed performance differences, particularly between infants and children, might not be reflecting age-related change in multimodal speech processing. Instead, the differences may be experimentally induced effects from varying pro-

cedures, stimuli, and task demands (Bjorklund, 2005; Desjardins et al., 1997; Fernald, Swingley, & Pinto, 2001; Green, 1998). A point is that infants' knowledge has been assessed indirectly via procedures such as looking time, whereas children's knowledge has been accessed directly via a variety of offline tasks requiring voluntary conscious retrieval of knowledge and formulation of responses during a poststimulus interval. The concepts of indirect and direct tasks are demarcated herein on the basis of task instructions as recommended by Merikle and Reingold (1991). Indirect measures do not direct participants' attention to the experimental manipulation of interest, whereas direct measures unambiguously instruct participants to respond to the experimental manipulation. The extent to which age-related differences in multimodal speech processing are reflecting development change versus varying task demands remains an important unresolved issue.

The purpose of this research was to assess the influence of visual speech on phonological processing by children with an indirect approach, namely, the multimodal picture–word task. The task was adapted from the children's cross-modal picture–word task of Jerger, Martin, and Damian (2002) and is appropriate for a broad range of ages. Our experimental tasks qualify as indirect measures because we assess how manipulation of seemingly to-be-ignored distractors affects performance without the participants being informed of, or consciously trying to respond to, the manipulation. The value of an indirect approach for studying visual speech has been demonstrated previously by research showing an indirect effect of visual speech on performance in adults who had difficulty in directly identifying the visual speech stimuli (Jordan & Bevan, 1997). Facial expressions also seem to indirectly influence judgments of vocal speech expressions (happy–fearful) in individuals with severe impairments in directly processing facial expressions (de Gelder, Pourtois, Vroomen, & Bachoud-Levi, 2000). These results provide specific evidence that performance on direct and indirect tasks may differ. We propose that more precisely detailed visual speech representations are required for direct tasks requiring conscious access and retrieval of information relative to indirect tasks. Below we briefly describe the original cross-modal test and our new adaptation.

Children's multimodal picture–word task

In the children's cross-modal picture–word task (Jerger et al., 2002), children are asked to name a picture while attempting to ignore a nominally irrelevant auditory distractor. The connection between the picture–distractor pairs is varied systematically to reflect either a congruent, conflicting, or neutral relationship between the picture–distractor items. The dependent measure is the speed of picture naming, and the goal is to determine whether congruent or conflicting relationships speed up or slow down naming, respectively, relative to neutral (or baseline) relationships. Relative to the pictures, the entire set of distractors represents phonologically onset-related, semantically related, and unrelated items. More specifically, the phonologically related distractors are composed of onsets that are congruent, conflicting in place of articulation, or conflicting in voicing (e.g., the picture *pizza* coupled with *peach, teacher,* and *beast,* respectively). The semantic distractors are composed of categorically related and unrelated pairs (e.g., the picture *pizza* coupled with *hot dog* and *horse,* respectively), whereas the unrelated distractors are composed of vowel nucleus onsets (e.g., the picture *pizza* coupled with *eagle*).

The onset of the distractors is varied to be before or after the onset of the picture, referred to as the stimulus onset asynchrony (SOA). Whether the distractor influences picture naming depends on the SOA and the type of distractor. With regard to SOA, the effect of an onset-related phonological distractor is typically greater when the distractor lags the onset of the picture (Damian & Martin, 1999; Schriefers, Meyer, & Levelt, 1990). With regard to the type of distractor, phonologically related distractors speed up naming when the onsets are congruent but slow down naming when the onsets are conflicting in place or voicing relative to unrelated distractors. When congruent or conflicting auditory distractors speed up or slow down naming, performance is assumed to reflect crosstalk between speech production and perception (Levelt et al., 1991).

Fig. 1 illustrates this crosstalk for a lagging distractor in terms of the stages of processing characterizing production (top line) and perception (bottom line). In the figure, a speaker is naming the picture *pizza* while hearing the phonologically congruent distractor *peach.* The stages of processing for producing and perceiving speech proceed in opposite directions. Whether discrete, interactive, or cascaded, most models of picture naming assume the following stages: (a) conceptual processing and

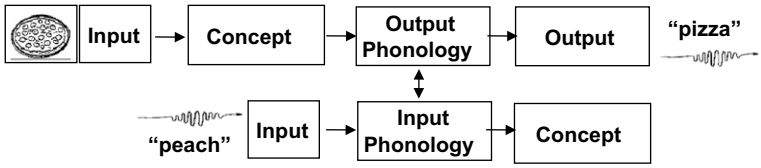

Fig. 1. Simplified stages of processing for speech production (top line) and perception (bottom line) for a speaker naming the picture *pizza* while hearing the phonologically congruent distractor *peach*.

activation of a set of meaning-related lexical items, (b) output phonological processing of the selected item, and (c) articulatory motor programming and output. In terms of perceiving speech, processing of the distractor is assumed to consist of the following stages: (a) input auditory/phonetic processing, (b) input phonological processing with activation of a set of phonologically related items, and (c) lexical–semantic and conceptual processing of the selected item. The output and input phonological processes are typically assumed to be separable interacting systems (Martin, Lesch, & Bartha, 1999).

The crosstalk between speech production and perception is assumed to occur when the picture naming process is occupied with output phonology and the distractor perceptual process is occupied with input phonology. Congruent distractors are assumed to speed up picture naming by activating input phonological representations whose activation spreads to output phonological representations, thereby allowing speech segments to be selected more rapidly during the naming process. Conflicting distractors are assumed to slow down naming by activating conflicting output phonological representations that compete with the picture's output phonology for control of the response. A novel contribution of the current research was the presentation of distractors both auditorily and audiovisually.

Our new picture–word task should provide an estimate of multimodal speech processing that is less sensitive to developmental differences in task demands such as the conscious access and retrieval of information required by direct procedures (Bertelson & de Gelder, 2004). That said, performance on both indirect and direct multimodal speech tasks remains susceptible to developmental changes in a variety of cognitive–linguistic skills. Below we detail our primary and some secondary research questions and predict how developmental changes in relevant cognitive–linguistic factors may affect selected components of our task.

Research questions and predicted results

Our primary research question concerned whether and how visual speech may enhance phonological processing by children over the age range of 4 to 14 years relative to auditory speech. In agreement with Campbell (1988), we view visual speech as an extra phonetic resource, perhaps adding another type of phonetic feature, that should enhance facilitatory and interference effects relative to auditory speech only. Possible age-related influences in children's sensitivity to visual speech may be predicted in terms of interactive developmental changes in (a) input/output coding processes, (b) phonological representational knowledge, and (c) general information processing.

Input/Output coding processes

Evidence indicates that younger children with less mature perceptual skills process auditory speech cues less efficiently. Relative to adults, they require a greater amount of, and a higher fidelity of, input for auditory word recognition (Cameron & Dillon, 2007; Elliott, Hammer, & Evan, 1987). These results suggest that younger children may need to rely on visual speech to supplement their less efficient processing of auditory speech cues. Visual speech might enhance phonological effects on performance by providing additional phonetic information along with speech envelope information that aids extraction of the auditory cues. Facial expressions might also supplement less efficient auditory speech processing by providing easier nonverbal information that promotes understanding of the intent of what was heard (Doherty-Sneddon & Kent, 1996).

The above evidence about input coding processes leads to the prediction that visual speech will enhance phonological processing by younger children. Some proposals about output coding processes bolster this prediction, suggesting that younger children with less mature articulatory proficiency observe visual speech disproportionately in order to cement their knowledge of the relation between articulatory gestures and their acoustic consequences (Dodd, 1987; Gordon, 1990). With regard to predicting performance across a broad age range, the evidence suggests that there may be developmental shifts in the processing weights assigned to the auditory and visual speech modalities. This in turn may cause apparent developmental shifts in children's sensitivity to visual speech (Brainerd, 2004). The time course of developmental effects is difficult to predict from the literature.

Phonological representational knowledge

A broad literature suggests that younger children have less detailed phonological representations and less efficient mapping of acoustic information onto the representations (see Snowling & Hulme, 1994). This evidence predicts that the additional phonetic information provided by visual speech will enhance phonological effects on performance in younger children. With regard to performance across a broader age range, some developmental changes in phonological representational knowledge seem to occur at around 5 or 6 years of age. First, data suggest that phonological processes become sufficiently proficient at around this age to begin using "inner" speech for learning, remembering, and problem solving (Conrad, 1971). Second, and perhaps more important, the initiation of literacy instruction at around this age triggers dramatic changes in phonological skills (Bentin, Hammer, & Cahan, 1991; de Gelder & Morais, 1995; Morais, Bertelson, Cary, & Alegria, 1986; Morrison, Smith, & Dow-Ehrensberger, 1995). Some authorities propose that as children's experience transmutes from phonemes as coarticulated nondistinct speech elements to phonemes as separable distinct written elements, phonological knowledge and awareness of phonemes become more highly detailed and specified (for discussions, see Anthony & Francis, 2005; Bryant, 1995). The time frame required for systematizing the knowledge gained during literacy learning for a language such as English with complicated print–sound mappings is estimated as 3 years (Anthony & Francis, 2005).

Overall, phonological knowledge appears to reorganize into a more elaborated, systematized, and robust resource for supporting a wider range of activities, such as reading and using inner speech to think and reason, from roughly 6 to 9 years of age. To the extent that the phonological knowledge supporting visual speech processing is not as readily accessed and/or retrieved during this process of restructuring, results predict that we may observe a developmental shift in children's sensitivity to visual speech during this time period. An intimate link between visual speech skills and the phonological knowledge gained by becoming literate is supported by findings that older individuals with reading disorders exhibit significantly less influence of visual speech on performance and unusually poor visual speechreading skills (de Gelder & Vroomen, 1998a; Ramirez & Mann, 2005).

Information processing

Information processing skills have been addressed in terms of general attentional resources, multimodal stimuli, and face processing. First, with regard to general resources, to the extent that phonological representational knowledge undergoes restructuring as discussed above, this reorganization may demand a disproportionate share of a child's limited processing capacity. To the extent that overloading available information processing resources creates an obstacle to processing visual speech, results predict that we may see less influence of visual speech on performance in the age range from 6 to 9 years.

A number of general processing mechanisms may also be enhanced by external cues, and younger children with immature processing skills may benefit disproportionately from such cues. For example, some experts propose that visual speech acts as a type of "alerting" or "motivational" mechanism (Arnold & Hill, 2001; Campbell, 2006). This viewpoint suggests that visual speech may boost attention, orienting, arousal, and/or motor preparedness, which would aid detection, discrimination, and rapid information processing (Wickens, 1974). This evidence predicts that younger children may benefit from visual speech due to processing-enhancing mechanisms that boost less mature skills. Available evidence does not allow prediction of the time course of the developmental effects.

With regard to multimodal stimuli, theorists propose that development consists of transitioning from processing of multimodal inputs more holistically to true multimodal integration of differentiated sensory modalities (for a review, see Lickliter & Bahrick, 2004). This predicts that we may observe developmental shifts in children's sensitivity to visual speech because of transitions in the processing of auditory and visual speech inputs from a supramodal manner to a modality-specific manner. The time course of developmental effects is difficult to predict.

Finally, with regard to face processing, evidence suggests that the talker's face is encoded during speechreading (Campbell & De Haan, 1998). An association between speechreading and face processing is supported by the observation that patients with severe face processing deficits due to prosopagnosia may show a loss of visual speechreading ability (de Gelder & Vroomen, 1998b). Children have some difficulties in processing faces and the full range of facial expressions up to around the preteen–teenage years (Campbell, Walker, & Baron-Cohen, 1995; Carey, Diamond, & Woods, 1980; Durand, Gallay, Seigneuric, Robichon, & Baudouin, 2007; Mondloch, Geldart, Maurer, & Le-Grand, 2003). Face-to-face communication may also hinder, rather than help, performance on some types of tasks in children (Doherty-Sneddon, Bonner, & Bruce, 2001; Doherty-Sneddon et al., 2000). This latter finding may be related to the more general phenomenon of gaze aversion, in which individuals reduce environmental stimulation in an attempt to reduce cognitive load and enhance processing (Glenberg, Schroeder, & Robertson, 1998). This predicts that we may observe a developmental shift in the influence of visual speech on phonological processing due to a transition in the processing of the facial context of visual speech around the preteen–teenage years. In short, multiple complex interactive factors may produce developmental shifts in children's sensitivity to visual speech.

In addition to our primary research question, secondary research questions addressed whether phonologically related distractors consistently speed up phonological processing when they are congruent and slow down processing when they are conflicting relative to the baseline distractors as well as whether the magnitude of phonological effects on performance declines systematically with age. Scant evidence in children with cross-modal picture–word tasks indicates that phonologically related auditory distractors consistently facilitate picture naming when they are congruent and disrupt naming when they are conflicting relative to a baseline condition (Brooks & MacWhinney, 2000; Jerger, Lai, & Marchman, 2002; Jerger et al., 2002). Effects on performance are more pronounced in younger children than in older children. We expect the experimental manipulations of our multimodal approach to produce comparable effects on phonological processing, thereby allowing us to address our primary question in a sensitive manner. Results will contribute new evidence about how phonological processing is influenced by visual speech over a broad range of ages on the same task and whether results on an indirect task mirror results across studies in the literature on direct tasks.

Method

Participants

Participants were 100 children (50 girls and 50 boys) ranging in age from 4 years 3 months to 14 years 0 months. The racial distribution was 85% Whites, 5% Asians, 3% Blacks, 3% Indians, and 4% multiracial, with 12% of Hispanic ethnicity. The children were formed into five groups of 20 each according to age: 4-year-olds, 5-year-olds, 6- and 7-year-olds, 8- and 9-year-olds, and 10- to 14-year-olds. The rationale for grouping by variable age intervals was that speech development in children is a nonlinear process in which developmental growth is more active and changing during earlier years than during later years (American Speech Language Hearing Association., 2008). Thus, as age progresses, one can group children by larger age intervals while maintaining reasonably homogeneous speech skills. The criteria for participation were (a) no diagnosed or suspected disabilities and (b) English as the native language. All children passed standardized or laboratory measures establishing the normalcy of hearing sensitivity, visual acuity (including corrected to normal), visual perception, spoken word recognition, vocabulary skills, articulatory proficiency, phoneme discrimination, and oral–motor function. The average Hollingshead (1975) social strata score was 1.5, which is consistent with a major business and professional socioeconomic status.

With regard to pronunciation of the names of the pictures, all participants pronounced the onsets accurately. The offsets of the picture names were also pronounced correctly except by 19 children, of whom 53% were 4-year-olds, 26% were 5-year-olds, 16% were 6- or 7-year-olds, and 5% were 8- or 9-year-olds. These children mispronounced either the /th/ in *teeth,* the /mp/ in *pumpkin,* the /r/ in *deer,* or the /z/ in *pizza* during speeded naming. With regard to identification of the auditory distractors, all children showed near ceiling performance on an auditory-only task. With regard to visual speechreading skills, scores on the Children's Audiovisual Enhancement Test (Tye-Murray & Geers, 2001) improved noticeably with age. Visual-only performance scored in terms of words averaged approximately 4% in the 4- and 5-year-olds, 15% in the 6- to 9-year-olds, and 23% in the 10- to 14-year-olds. Visual-only performance for word onsets scored in terms of visemes, or the smallest distinguishable units of speech defined by lip movements (Fisher, 1968), averaged approximately 35% in the 4- and 5-year-olds, 58% in the 6- to 9-year-olds, and 73% in the 10- to 14-year-olds.

Materials and instrumentation for picture–word task

Stimulus preparation

All stimuli were recorded by an 11-year-old boy. He wore a solid navy shirt and lip gloss, and he looked directly into the camera. The rationale for a child talker was to increase attention and interest for the child participants. Our informal experience with children and formal evidence in infants (Bahrick, Netto, & Hernandez-Reif, 1998) suggest a strong preference for child faces over adult faces. The recording setting was the Audiovisual Stimulus Preparation Laboratory of the University of Texas at Dallas with recording equipment, soundproofing, and supplemental lighting and reflectors. The talker started and ended each utterance with a neutral face/closed mouth position. The full facial image and upper chest of the talker were recorded. Full facial image stimuli yield more accurate speechreading performance (Greenberg & Bode, 1968), supporting the idea that facial movements other than the mouth area may contribute to speechreading (Munhall & Vatikiotis-Bateson, 1998).

The audiovisual recordings were digitized via a Macintosh G4 computer with Apple Fire Wire, Final Cut Pro, and Quicktime software. Color video was digitized at 30 frames/s with 24-bit resolution at a 720×480-pixel size. Auditory input was digitized at a 22-kHz sampling rate with a 16-bit amplitude resolution. The pool of utterances was edited to an average root mean square (RMS) level of -14 dB. The average fundamental frequency was 202 Hz.

Stimulus onset asynchrony

The colored pictures were scanned into a computer and edited to achieve objects of a similar size and complexity on a white background. The size of the pictures was edited to be the width of the face at eye level. Each picture was pasted onto the upper chest of the talker in exactly the same time frame for both auditory and audiovisual items. The pictures were pasted twice to form stimulus onset asynchronies (SOAs) of -165 and +165 ms (the onset of the distractor was 165 ms or 5 frames before and after the picture, respectively). To be consistent with current practice, we defined a distractor's onset on the basis of its auditory onset. Technically, a picture can be pasted onto an audiovisual stimulus only at the beginning of a frame (every 33 ms). To illustrate our pasting strategy, we use an imaginary SOA of 0 ms (simultaneous picture–distractor onsets). The goal was that the onset of a picture should be in the frame nearest the auditory onset. Thus, if the auditory onset was in the first half of a frame, we pasted the picture at the beginning of that frame; if the auditory onset was in the last half of a frame, we pasted the picture in the beginning of the nearer following frame. This strategy yielded an average SOA with a maximum variability of approximately 16 ms.

In the literature, leading and lagging SOAs are reported both in combination and in isolation. With regard to using SOAs in combination, researchers have chosen this approach when they were interested in tracking the time course of phonological or semantic activation. These results have yielded interesting differences between the different types of distractors. More specifically, findings have shown that a lagging SOA of roughly 100–200 ms tends to maximize any phonological effects on performance due to interactions between input and output phonology (Damian & Martin, 1999; Schriefers et al., 1990) (see Fig. 1). When phonological distractors are presented at a leading SOA of approximately 100 to 200 ms, on the other hand, phonological effects on performance are typically

very small. In this latter case, less interaction is attributed to less temporal overlap between the two types of phonology, with activation of the input phonological representations decaying prior to the output phonological encoding of the picture. In contrast to these findings, results for semantic distractors have yielded the opposite pattern of interaction. Semantic effects on performance are typically negligible at lagging SOAs and prominent at leading SOAs. With regard to a focus on only one SOA, researchers have chosen this approach when they wished to investigate differing effects on performance produced by differing types of phonological or semantic distractors. In this case, they typically focus on the SOA maximizing any effect, that is, lagging for phonological distractors and leading for semantic distractors. Research questions about the time course of activation versus the effects of differing types of distractors are usually reported in separate articles, although all data may be gathered simultaneously, particularly in children, due to the difficulties and expense of recruiting and testing the participants. It is also the case that an inconsistent relationship between the picture–distractor pairs is viewed as boosting listeners' attempts to disregard the distractors. In this article, aimed at explicating the developmental course of children's sensitivity to visual speech, we focus only on the phonological distractors at the lagging SOA.

Pictures and distractors

Development of specific test items and conditions comprising the children's cross-modal picture–word task has been detailed previously (Jerger & Martin et al., 2002). The pictured objects of this study are the same pictures as those used previously; however, the distractors differ. Table A1 (see Appendix A) details the individual picture and distractor word items, and Table A2 summarizes linguistic statistics for the phonology pictures and distractors. In brief, the test materials are of high familiarity, high concreteness, high imagery, high phonotactic probabilities, low word frequency, and early age of acquisition (Carroll & White, 1973; Coltheart, 1981; Cortese & Fugett, 2004; Dale & Fenson, 1996; Gilhooly & Logie, 1980; Morrison, Chappell, & Ellis, 1997; Nusbaum, Pisoni, & Davis, 1984; Snodgrass & Vanderwart, 1980; Vitevitch & Luce, 2004). In brief, the onsets of the pictures always began with /b/, /p/, /t/, or /d/ coupled with the vowels /i/ or / ∧ / Previous research has established that speechreading performance for these onsets is equivalent for /i/ and / ∧ / vowel contexts (Owens & Blazek, 1985). Our rationales for selecting the onsets were twofold. First, the onsets represent developmentally early phonetic achievements and reduced articulatory demands (Dodd, Holm, Hua, & Crosbie, 2003; Smit, Hand, Freilinger, Bernthal, & Bird, 1990). To the extent that phonological development is a dynamic process, with knowledge improving from unstable, minimally specified, and harder-to-access/retrieve representations to stable, robustly detailed, and easier-to-access/retrieve representations, it seems important for an initial study to assess early acquired phonemes that children are more likely to have mastered (for similar reasoning about semantic knowledge, see McGregor, Friedman, Reilly, & Newman, 2002).

Second, the onsets represent variations in place of articulation (/b/–/d/ vs. /p/–/t/) and voicing (/b/–/p/ vs. /d/–/t/), two phonetic features that are traditionally thought to be differentially dependent on auditory versus visual speech. Previous findings, based on lip or lower face visual images, indicate that place of articulation is easier to discriminate visually, whereas voicing is easier to discriminate auditorily (Miller & Nicely, 1955; Owens & Blazek, 1985). Each picture was administered in the presence of the four types of distractors described previously: congruent, one feature conflicting in place of articulation, one feature conflicting in voicing, and vowel onset baseline distractors.

Experimental instrumentation

To administer picture–word items, the video track of the Quicktime movie file was routed to a high-resolution computer monitor and the auditory track was routed through a speech audiometer to a loudspeaker. For audiovisual trials, each trial contained 1000 ms of the talker's still neutral face and upper chest, followed by presentation of one colored picture on the chest and an audiovisual utterance of one distractor word, followed by 1000 ms of the still neutral face and the colored picture. For auditory-only trials, each trial contained 1000 ms of the still neutral face and upper chest, followed by a continuation of the still neutral face, presentation of one colored picture on the chest, and an auditory-only utterance of one distractor word, followed by 1000 ms continuation of the still face and the colored picture. Each picture was pasted in exactly the same time frame for both audi-

tory and audiovisual items. Thus, the only difference between the auditory and audiovisual conditions was that the auditory items have a neutral face and the audiovisual items have a dynamic face.

The computer monitor and the loudspeaker were mounted on an adjustable-height table directly in front of the child at a distance of approximately 90 cm. To name each picture, the child spoke into a unidirectional microphone mounted on an adjustable stand. To obtain naming latency, the computer triggered a counter/timer with better than 1 ms resolution at the initiation of a movie file. The timer was stopped by the onset of the child's vocal response into the microphone, which was fed through a stereo mixing console amplifier and 1-dB step attenuator to a voice-operated relay (VOR). A pulse from the VOR stopped the timing board via a data module board. We verified that the VOR was not triggered by the auditory distractors. The counter timer values were corrected for the amount of silence in each movie file before the onset of the picture. Naming times were digitally recorded for offline analysis in all children with flawed pronunciations.

Procedure

Participants were tested in two separate sessions approximately 12 days apart: one for auditory testing and one for audiovisual testing. The modality of the first and second sessions was counterbalanced across participants. The first session always began with a practice task. A tester showed each picture on a 5 × 5-inch card, asking the child to name the picture and teaching the child the target names of any pictures named incorrectly. Next, the tester flashed some picture cards quickly and modeled speeded naming. The child was asked to copy the tester for another few pictures. Speeded naming practice trials went back and forth between tester and child until the child was naming pictures fluently, particularly without saying "a" before names. The second session always began with a mini-practice task.

The experimental trials consisted of two practice items followed by presentation of all the pictures with each type of speech distractor in a random order within one unblocked condition (for a discussion, see Starreveld, 2000). No individual picture or word distractor was allowed to recur without at least two intervening trials. The child sat at a child-sized table in a double-walled sound-treated booth. The tester sat at a computer workstation, and a cotester sat alongside the child, keeping him or her on task. Each trial was initiated by the tester's pushing the space bar (out of the participant's sight). The child was instructed to name each picture and disregard the speech distractor. The child was told that "Andy" (a pseudonym) was wearing a picture on his chest and wanted to know what it was. The child was to say the name as quickly as possible to say it correctly. The microphone was placed approximately 12 inches from the child's mouth without blocking his or her view of the monitor. If necessary, the child's speaking level, the position of the microphone or child, and/or the setting on the 1-dB step attenuator between the microphone and VOR were adjusted to ensure that the VOR was triggering reliably. The intensity level of the distractors was approximately 70 dB sound pressure level (SPL), as measured at the imagined center of the participant's head with a sound level meter.

Measures

The dependent measures were picture naming times in the presence of both the auditory and audiovisual distractors. The picture–distractor pairs represented congruent, conflicting in place of articulation, conflicting in voicing, and neutral (or baseline) relationships. With regard to the characteristics of these data, 5.52% of all trials were excluded or missing for the following reasons. Naming responses that were more than 3 standard deviations from an item's conditional mean were discarded. This procedure excluded 1.68% of trials. Naming responses that were flawed, on the other hand, were deleted online and readministered after intervening items. The percentage of overall trials judged to be flawed (e.g., lapses of attention, squirming out of position, triggering the microphone in a flawed manner) was 17.45%, ranging from 24.48% in the younger children to 7.35% in the older children. The percentage of missing trials remaining at the end because the readministered trial was also flawed was 6.35% in the younger children and less than 1% in the older children, averaging 3.84% of overall trials.

Results

Analysis plan

Naming times were analyzed with factorial mixed-design analyses of variance, regression analyses, and *t* tests (Abdi, Edelman, Valentin, & Dowling, in press). The overall set of variables was composed of a between-participants factor (five age groups) and within-participants factors representing the modality of the distractor (auditory or audiovisual) and the type of condition (congruent, conflicting in place of articulation, conflicting in voicing, or baseline). The problem of multiple comparisons was controlled with the false discovery rate (FDR) procedure (Benjamini & Hochberg, 1995; Benjamini, Krieger, & Yekutieli, 2006). The FDR approach controls the expected proportion of false-positive findings among rejected hypotheses. A value of the approach is its demonstrated applicability to repeated-measures designs. For the experimental conditions, we quantified the degree of facilitation and interference from congruent and conflicting onsets, respectively, with adjusted naming times. Adjusted times were derived by subtracting each participant's vowel baseline naming times from his or her congruent and conflicting times, as was done in our previous studies (Jerger & Lai et al., 2002; Jerger & Martin et al., 2002). This approach controls for developmental differences in detecting and responding to stimuli and allows each picture to serve as its own control without affecting the differences among the types of distractors.

Baseline condition

Fig. 2 shows average naming times in the age groups for the vowel onset distractors presented in the auditory versus audiovisual modalities. Naming times for the /i/ and / ∧ / onsets were statistically equivalent; results are collapsed across vowels. Omnibus statistical analysis of the data included one between-participants factor (age groups) and one within-participants factor (modality: auditory vs. audiovisual). Results indicated that age significantly affected overall naming times, $F(4, 95) = 23.109$, $p < .0001$. No other significant effects or interactions were observed.

To obtain a more precise understanding of the effects of age, we carried out a multiple regression analysis. Results indicated a significant decrease in naming times with increasing age, $F(1, 196) = 142.182$, $p_{\text{one-way}} < .0001$. A linear trend accounted for approximately 99% of the age-related decline in naming times for each modality. The slopes of the auditory and audiovisual functions (–10 vs. –9 ms/month, respectively) did not differ statistically, indicating that naming times improved in a comparable way with age for each mode. The intercepts of the auditory and audiovisual functions (2524 vs. 2471 ms, respectively) also did not differ. In the presence of homogeneous slopes, equivalent intercepts indicate equivalent absolute naming times; that is, the auditory developmental function was not shifted relative to the audiovisual function. Naming times collapsed across modality de-

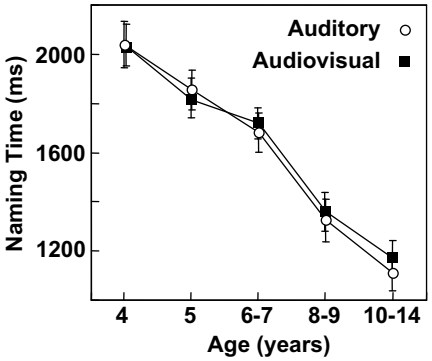

Fig. 2. Average absolute naming latencies in five age groups for vowel onset baseline distractors presented in the auditory versus audiovisual modalities.

creased from approximately 2035 ms in the 4-year-olds to 1140 ms in the 10- to 14-year-olds. An age-related improvement in absolute picture naming times agrees with previous findings (Brooks & MacWhinney, 2000; Jerger & Lai et al., 2002; Jerger & Martin et al., 2002; Jescheniak, Hahne, Hoffmann, & Wagner, 2006; Melnick, Conture, & Ohde, 2003).

Experimental conditions

Initially, we conducted an omnibus analysis of the experimental conditions with one between-participants factor (age groups) and two within-participants factors (modality: auditory or audiovisual; condition: congruent, conflicting in place, or conflicting in voicing). Results indicated that adjusted naming times were significantly influenced by age, $F(4, 95) = 3.795$, $p = .007$, and condition, $F(2, 190) = 201.684$, $p < .001$. The effect of the condition on performance varied in complex ways, however, as a function of age and the modality of the distractor, with a significant Condition × Age Group interaction, $F(8, 190) = 6.242$, $p < .001$, Condition × Modality interaction, $F(2, 190) = 3.260$, $p = .041$, and Condition × Age Group × Modality three-way interaction, $F(8, 190) = 2.463$, $p = .015$. No other significant effects were observed. These complex interactions were probed by analyzing each condition separately.

Congruent condition

Fig. 3 shows the degree of facilitation as quantified by adjusted naming times for auditory and audiovisual congruent distractors in the age groups. The zero baseline of the ordinate represents naming times for the vowel onset baseline distractors (Fig. 2). Results of multiple regression analysis did not indicate a significant general effect of age on adjusted naming times. However, the individual developmental curves characterizing the auditory and audiovisual functions differed significantly from each other, $F(1, 196) = 3.952$, $p_{\text{one-way}} = .024$. Whereas a quadratic trend accounted for the largest proportion (74%) of age-related variability for audiovisual distractors, a linear trend accounted for the largest proportion (89%) of the variability for auditory distractors. For audiovisual distractors, both the linear and quadratic trends were significant, $F(1, 95) = 2.79$, $p_{\text{one-way}} = .049$, and $F(1, 95) = 8.60$, $p_{\text{one-way}} = .002$, respectively. For auditory distractors, the linear trend approached significance, $F(1, 95) = 2.28$, $p_{\text{one-way}} = .066$. Previous results on the children's cross-modal picture–word task with auditory distractors have consistently shown a greater degree of facilitation for younger children than for older children (Brooks & MacWhinney, 2000; Jerger &Lai et al., 2002; Jerger & Martin et al., 2002; Jescheniak et al., 2006).

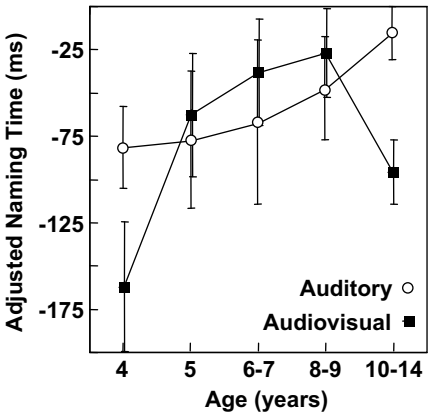

Fig. 3. Congruent distractors: Degree of facilitation for auditory versus audiovisual modalities as quantified by adjusted naming latencies in five age groups. The zero baseline of the ordinate represents naming times for vowel onset baseline distractors (Fig. 2). A larger negative value indicates more facilitation.

Multiple *t* tests with the FDR method controlling for multiplicity indicated that the degree of facilitation was significantly greater for audiovisual distractors than for auditory distractors in the 4-year-olds and 10- to 14-year-olds. All other groups showed equivalent degrees of facilitation for both types of distractors. Multiple *t* tests with the FDR method assessing whether the adjusted naming times differed significantly from zero indicated significant facilitation in the 4-, 5-, and 10-year-olds for audiovisual distractors and in the 4- and 5-year-olds for auditory distractors. FDR results in the 6- and 7-year-olds and 8- and 9-year-olds for auditory distractors approached significance.

Conflicting in voicing condition

Fig. 4 shows the degree of interference as quantified by adjusted naming times in the age groups for auditory and audiovisual distractors conflicting in voicing. Again, the zero baseline of the ordinate represents naming times for the vowel onset baseline distractors. Results of multiple regression indicated a significant decrease in interference with increasing age, $F(1, 196) = 24.049$, $p_{\text{one-way}} < .0001$. The degree of age-related change differed significantly, however, for the auditory and audiovisual functions, with the developmental trajectory significantly steeper for the audiovisual modality, $F(1, 196) = 3.580$, $p_{\text{one-way}} = .030$. A linear trend accounted for 90% of the between-groups variability for audiovisual distractors but only 55% of the variability for auditory distractors. The trends were significant for both functions: audiovisual, $F(1, 95) = 24.46$, $p_{\text{one-way}} < .0001$, and auditory, $F(1, 95) = 5.76$, $p_{\text{one-way}} = .009$. The curvilinear trends did not achieve significance. The slopes of the functions declined by 2 ms/month for the audiovisual mode but only by 1 ms/month for the auditory mode. The nonparallel slopes for the functions rendered testing differences between the intercepts irrelevant.

Multiple *t* tests with the FDR method controlling for multiplicity indicated significantly greater interference from audiovisual distractors than from auditory distractors in the 4-year-olds. All other groups showed equivalent degrees of interference for both types of distractors. Multiple *t* tests with the FDR method assessing whether all adjusted naming times differed significantly from zero indicated a significant degree of interference in all groups for both auditory and audiovisual distractors. These findings agree with previous findings for auditory-only conflicting in voicing distractors (Jerger & Lai et al., 2002; Jerger & Martin et al., 2002).

Conflicting in place condition

Fig. 5 shows the degree of interference as quantified by adjusted naming times in the age groups for auditory and audiovisual distractors conflicting in place. Results of multiple regression analysis indi-

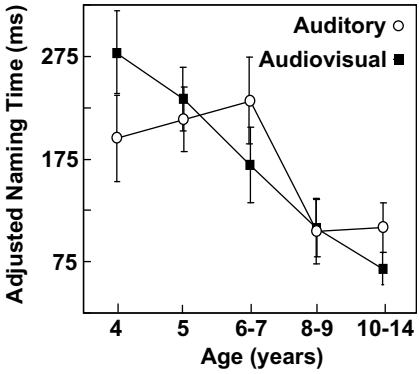

Fig. 4. Conflicting in voicing distractors: Degree of interference for auditory versus audiovisual modalities as quantified by adjusted naming latencies in five age groups. The zero baseline of the ordinate represents naming times for vowel onset baseline distractors (Fig. 2). A larger positive value indicates more interference.

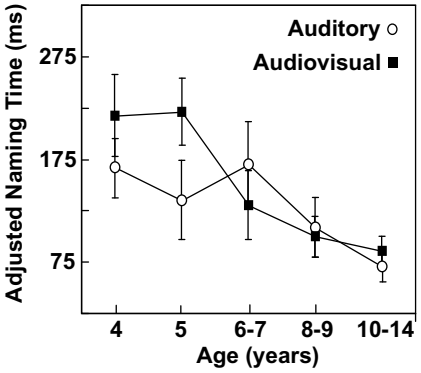

Fig. 5. Conflicting in place distractors: Degree of interference for auditory versus audiovisual modalities as quantified by adjusted naming latencies in five age groups. The zero baseline of the ordinate represents naming times for vowel onset baseline distractors (Fig. 2). A larger positive value indicates more interference.

cated a significant decrease in interference with age, $F(1, 197) = 15.579$, $p_{\text{one-way}} < .0001$. A linear trend accounted for 85% of the between-groups variability for audiovisual distractors and 66% of the variability for auditory distractors. The trends were significant for both functions: audiovisual, $F(1, 95) = 13.95$, $p_{\text{one-way}} < .0001$, and auditory, $F(1, 95) = 4.18$, $p_{\text{one-way}} = .022$. The curvilinear trends did not achieve significance. The developmental functions for the audiovisual and auditory modalities were characterized by statistically equivalent slopes (−1 ms/month) and intercepts (254 ms). Homogeneous slopes and intercepts signify comparable adjusted naming times and a uniform degree of age-related decline for the two modalities. That said, the notable trend suggesting greater interference from the audiovisual distractors in the younger children, particularly the 5-year-olds, clearly seems worth mentioning. Finally, results of multiple *t* tests with the FDR method indicated significant interference in all groups for both auditory and audiovisual distractors. These findings are consistent with previous findings for auditory-only conflicting in place distractors (Jerger & Lai et al., 2002; Jerger & Martin et al., 2002).

Discussion

This research modified the children's cross-modal picture–word task (Jerger & Martin et al., 2002) into a multimodal procedure for assessing indirectly the influence of visual speech on phonological processing. Results varied as a function of age and the type and modality of the distractors in complex ways. For distractors conflicting in place of articulation, the groups showed statistically equivalent interference for the auditory and audiovisual distractors. There was a notable trend suggesting greater interference from the audiovisual distractors in the younger children, particularly the 5-year-olds, but we lacked the statistical power to detect the effect when correcting for multiple comparisons. The degree of interference decreased significantly with increasing age in an equivalent manner across modalities.

For distractors conflicting in voicing, the degree of interference also decreased significantly with increasing age, but the auditory and audiovisual functions exhibited significantly different developmental trajectories. The degree of age-related change was greater for the audiovisual function because the 4-year-olds showed significantly greater interference from audiovisual distractors than from auditory distractors, whereas all other groups showed equivalent degrees of interference from both types of distractors. Although results in the 4-year-olds appear to be inconsistent with the literature indicating that voicing is difficult to discriminate visually on the lips (Tye-Murray, 1998), they are consistent with more recent data suggesting that some visemes, such as /p/ versus /b/, may be more readily

discriminated visually when individuals view full facial images as used herein (Bernstein, Iverson, & Auer, 1997, as cited in Bernstein, Demorest, & Tucker, 2000).[1]

For the congruent distractors, the auditory and audiovisual functions exhibited significantly different developmental trajectories. The audiovisual function showed a unique, significant quadratic trend due to audiovisual distractors producing significantly greater facilitation than auditory distractors in the 4-year-olds and 10- to 14-year-olds but not in the other age groups. The degree of facilitation also varied considerably. Congruent distractors produced significant facilitation in the 4-, 5-, and 10-year-olds for audiovisual distractors and in the 4- and 5-year-olds for auditory distractors, with results in the 6- and 7-year-olds and 8- and 9-year-olds being of borderline significance.

Our results showing a pronounced influence of visual speech on performance for congruent and conflicting distractors in the 4-year-olds are difficult to relate to the literature because previous studies have pooled results with older ages. To the extent that the previous amalgamated data are reflecting the performance of 4-year-olds accurately, our results on an indirect task disagree with the results obtained on direct testing measures (Desjardins et al., 1997; Dupont et al., 2005; Massaro, 1984; Massaro et al., 1986). Further research on indirect versus direct testing approaches in 4-year-olds is warranted. Our results showing a lack of influence of visual speech on performance in the 5- to 9-year-olds agree with previous data on a variety of direct testing measures (Desjardins et al., 1997; Dupont et al., 2005; Hockley & Polka, 1994; Massaro, 1984; Massaro et al., 1986; McGurk & MacDonald, 1976; Sekiyama & Burnham, 2004; Wightman et al., 2006). Children within this age range are less influenced by visual speech on both indirect and direct tasks. The data support the conclusion that the negative findings in the 5- to 9-year-olds represent an age-related effect rather than differences in the task demands of indirect and direct procedures.

Some previous investigators have attributed the reduced influence of visual speech on performance in children to their poorer speechreading abilities (Massaro et al., 1986; Wightman et al., 2006). A relation between the influence of visual speech on performance and speechreading skills seems undeniably reasonable. That said, such a relationship cannot explain the developmental shifts noted in our research. As detailed in the Method section, visual-only speechreading scores were comparable in the 4- and 5-year-olds (4% words, 35% visemes) and were poorer in these children than in the 6- to 9-year-olds (15% words, 58% visemes). Thus, speechreading skills cannot account for our results indicating a greater influence of visual speech on performance in the 4-year-olds than in the 5- to 9-year-olds.

Our findings in the 10- to 14-year-olds indicating that congruent audiovisual distractors produced significantly more facilitation agree with results in the literature (Conrad, 1977; Dodd, 1977, 1980; Hockley & Polka, 1994). A novel finding of this research was the observation that conflicting audiovisual distractors did not produce significantly more interference in the 10- to 14-year-olds. A significant influence of visual speech on the degree of facilitation but not on the degree of interference may be associated with the more advanced cognitive abilities of preteen and teenage children and adults, particularly in terms of inhibiting conflicting information and resisting interference (Bjorklund, 2005; Jerger, Pearson, & Spence, 1999). Finally, we should note that previous investigators have compared children's performance with adults' performance rather than with 10- to 14-year-olds' performance. We also tested a group of college students (18–38 years of age). The data were not included because the pattern of results in the 10- to 14-year-olds was adult-like, mirroring findings in the college students.

[1] To probe Bernstein and colleagues' (2000) suggestion, the speech readability of the onsets of the distractors (/b/, /d/, /p/, and /t/) in terms of place of articulation (labial vs. not labial) and voicing (voiced vs. not voiced) was assessed in a pilot study with 20 normal adults (13 women and 7 men) ranging in age from 19 to 32 years ($M = 22.6$ years). The students watched visual-only presentations of the distractors intermixed with filler items and classified each onset in terms of lips/not lips or voicing/not voicing. The order of classifying on the basis of place of articulation or voicing was counterbalanced across students. Immediately before testing each type of classification, the students classified a practice list with feedback. Classification of the distractors' onsets averaged 90.5% correct (range = 70–100) for place of articulation and 65.0% correct (range = 45–80) for voicing. Classification was significantly above chance for both labial place of articulation, $t(19) = 22.84$, $p < .0001$, and voicing, $t(19) = 7.55$, $p < .0001$. The amount of visual speech information observed for distractors conflicting in voicing seems to have been sufficient to significantly influence performance on a classification task in adults and on our indirect task in 4-year-olds. More research is needed in this area.

The intriguing finding of the current study was the inverted U-shaped developmental function observed for congruent audiovisual distractors. Performance showed an influence of visual speech at the youngest and oldest ages but not at the intermediate ages. As noted earlier, speechreading skills were consistently improving with age. Why then was performance in the intermediate-aged children less influenced by visual speech, as seen in Fig. 3? U-shaped functions have been carefully scrutinized by dynamic systems theorists (Smith & Thelen, 2003), who propose that the plateau of the U-shaped trajectory is reflecting a period of transition rather than an actual loss of visual speech on performance. The idea is that the components of early skills are softly assembled behaviors (i.e., malleable configurations) that reorganize over time into more mature, stable, and flexible forms (Gershkoff-Stowe & Thelen, 2004). The dynamic systems model also assumes that *multiple* interactive factors, rather than one single factor, typically form the basis of developmental change. From this viewpoint, the temporary decline in the influence of visual speech on performance is viewed as reflecting a reorganization of relevant knowledge and processing subsystems in response to internal and environmental forces. Using knowledge mechanisms that are in a period of significant dynamic growth may require more resources and overload the processing system, resulting in a temporary decrease in processing efficiency.

With regard to the developmental changes in face processing, our results do not seem consistent with a lack of influence of visual speech on performance due to difficulties in processing faces and the full range of facial expressions until the preteen–teenage years (Campbell et al., 1995; Carey et al., 1980; Mondloch et al., 2003). The age range supporting the establishment of adult-like face processing skills does correspond closely to our age range showing the reestablishment of an influence of visual speech on performance, namely, 10- to 14-year-olds. That said, an effect on performance due to immaturities in face processing seems contradicted by the data in the current 4-year-olds, who showed a pronounced influence of visual speech on performance.

With regard to the influence of developmental changes in input/output processing skills, our results are consistent with the transitions in the processing weights of the auditory versus visual speech modalities that were proposed in the Introduction. An important idea of dynamic systems theory is that the ends of the U-shaped trajectory, which seem to reflect identical outcomes in the 4-year-olds and 10- to 14-year-olds, might not be reflecting identical underlying mechanisms. With regard to information processing attentional resources, our results are consistent with the proposal that visual speech may act as an external cue that disproportionately benefits information processing in preschool children with less mature skills, creating an indirect influence of visual speech on performance. With regard to multimodal processing, the U-shaped results are consistent with the proposal of developmental shifts due to transitions in the processing of multimodal inputs from an undifferentiated holistic manner to a modality-specific manner.

Our results are also consistent with the proposal that phonological representational knowledge reorganizes during the kindergarten–early elementary school years. The age at which literacy instruction begins and speech coding becomes sufficient to begin using inner speech for learning, remembering, and problem solving, around 5 or 6 years, is uncannily similar to the age at which an effect of visual speech on phonological processing seems to disappear in the current study. The age range of our results showing a lack of visual speech on performance is uncannily similar to the estimate that it requires a period of 3 years to systematize the knowledge gained during literacy learning for a language such as English (Anthony & Francis, 2005). To the extent that temporary periods of reorganization and dynamic growth may be characterized by less robust processing systems and decreases in processing efficiency, the influence of visual speech may vary as a function of the processing demands of different tasks. Higher demand tasks that stress processing may reveal developmental shifts more readily than do lower demand tasks that do not create the same degree of stress. Future research should explore the effects of visual speech on performance with tasks that manipulate information processing requirements.

In sum, a complex array of factors may influence the processing of multimodal stimuli. The U-shaped developmental function for congruent audiovisual distractors might be reflecting any or all of the above considerations with the possible exception of immaturities in face processing. Multimodal speech processing clearly seems to involve diverse component processes that require a multidisciplinary perspective.

Acknowledgments

This work was supported by the National Institute on Deafness and Other Communication Disorders (Grant DC-00421 to the University of Texas at Dallas). We thank Alice O'Toole for her generous advice and assistance in recording our audiovisual stimuli and interpreting data. We appreciate the thoughtful comments of Virgina Marchman on an earlier version of the manuscript. We thank the children and parents who participated, and we thank the students who assisted, namely Shaumika Ball, Karen Banzon, Katie Battenfield, Sarah Joyce Bessonette, K. Meaghan Dougherty, Irma Garza, Stephanie Hirsch, Kelley Leach, Anne Pham, Lori Pressley, Anastasia Villescas (all with data collection, analysis, and/or presentation), and Derek Hammons (with computer programming).

Appendix A. *Table. A1*

Pictures and auditory/auditory–visual distractors of Children's Multimodal Picture–Word Test

Pictured objects		Distractors			
Phonology items:					
Bees	Pizza	Beach	Demon	Peacock	Teacher
Bus	Pumpkin	Beast	Detective	Peach	Tee-shirt
Deer	Teeth	Buckle	Dumbo	Potato	Tomato
Duck	Tongue	Butterfly	Dumptruck	Puddle	Tugboat
			Eagle	Onion	
Semantic filler items:					
Boot	Pickle	Bear	Flag	Puppet	
Dog	Pizza	Bed	Glove	Shirt	
Doll	Tiger	Cat	Horse	Slipper	
Pants		Cheese	Hotdog	Worm	
		Dress	Lemon		

Note. Phonology pictures were administered in the presence of three types of distractors (congruent, one feature conflicting in place, and one feature conflicting in voicing onsets, e.g., *bus–buckle, bus–dumptruck, bus–potato*) plus the baseline distractor *onion* for / ʌ / vowel-nucleus pictures or *eagle* for /i/ vowel-nucleus pictures. Filler item pictures were presented in the presence of two types of distractors (semantically related and unrelated, e.g., *boot–slipper, boot–flag*).

Table. A2

Linguistic statistics for picture and distracter word items of phonology condition for Children's Multimodal Picture–Word Test

Source	Scale	Pictures (n = 8)	Distracters (n = 18)
Concreteness	*Very concrete = 7 or 700*		
Overall average	7 point or adjusted 7 point	6.27 (5)	5.80 (12)
Gilhooly and Logie (1980)	End point = 7	6.69 (2)	6.30 (6)
Coltheart (1981)	End point = 700	617.20 (5)	569.08 (12)
Imagery	*High imageability = 7 or 700*		
Overall average	7 point or adjusted 7 point	6.43 (7)	5.93 (14)
Morrison et al. (1997)	End point = 7	6.38 (5)	6.21 (7)
Coltheart (1981)	End point = 700	622.20 (5)	586.85 (13)
Cortese and Fugett (2004)	End point = 7	6.68 (6)	6.33 (3)
Word familiarity	*Very familiar* = 5, 7, or 700		
Overall average	7 point or adjusted 7 point	5.41 (8)	5.53 (15)

Table A2 (*continued*)

Source	Scale	Pictures (n = 8)	Distracters (n = 18)
Morrison et al. (1997)	End point = 5	2.57 (5)	3.03 (7)
Snodgrass and Vanderwart (1980)	End point = 5	3.05 (5)	2.98 (7)
Coltheart (1981)	End point = 700	543.20 (5)	517.62 (13)
Nusbaum et al. (1984)	End point = 7	6.93 (6)	6.97 (13)
Age of acquisition	*13+ years* = 7 or 9		
Overall average	7 point or adjusted 7 point	2.41 (6)	2.85 (10)
Morrison et al. (1997)	End point = 7	2.22 (5)	2.69 (7)
Carroll and White (1973)	End point = 9	3.19 (4)	3.40 (4)
Gilhooly and Logie (1980)	End point = 7	2.23 (2)	2.92 (6)
Word frequency			
Toddler data			
Dale and Fenson (1996)	Proportion of children understanding/producing words at 30 months	88.79 (7)	78.93 (4)
Adult data			
Kucera & Francis[a]	Printed occurrences per million	29.57 (7)	18.64 (14)
Word recognition			
Jerger et al. (2007)[b]	Percentage of children recognizing words from six alternative picture choices		
	Preschool	–[c]	91.90 (18)
	Elementary	–[c]	98.89 (18)
Phonotactic probability			
Vitevitch and Luce (2004)	Positional segment frequency		
	Sum	.1731 (8)	.2159 (18)
	Onset	.0580 (8)	.0524 (18)
	Position-specific biphone frequency		
	Sum	.0065 (8)	.0115 (18)
	Onset	.0023 (8)	.0025 (18)

Note. Each of the overall averages was obtained by averaging data across resources for each item and then averaging across mean item values for each subset. Numbers of items contributing to averages across resources are presented in parentheses. No average could be determined for word frequency.

[a] As cited in Coltheart (1981).

[b] Jerger, S., Bessonette, S. J., Davies, K. L., & Battenfield, K. (2007). Unpublished data, University of Texas at Dallas.

[c] Picture readability in young children was established previously (Jerger, Martin, & Damian, 2002).

References

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

American Speech Language Hearing Association. (2008). *Typical Speech and Language Development*. Retrieved March 28, 2008, Available from http://www.asha.org/public/speech/development/default.htm.

Anthony, J., & Francis, D. (2005). Development of phonological awareness. *Current Directions in Psychological Science, 14*, 255–259.

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

Bahrick, L., Netto, D., & Hernandez-Reif, M. (1998). Intermodal perception of adult and child faces and voices by infants. *Child Development, 69*, 1263–1275.

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, 57*, 289–300.

Benjamini, Y., Krieger, A., & Yekutieli, D. (2006). Adaptive linear step-up procedures that control the false discovery rate. *Biometrika, 93*, 491–507.
Bentin, S., Hammer, R., & Cahan, S. (1991). The effects of aging and first grade schooling on the development of phonological awareness. *Psychological Science, 2*, 271–274.
Bernstein, L., Demorest, M., & Tucker, P. (2000). Speech perception without hearing. *Perception & Psychophysics, 62*, 233–252.
Bertelson, P., & de Gelder, B. (2004). The psychology of multimodal perception. In C. Spence & J. Driver (Eds.), *Crossmodal space and crossmodal attention* (pp. 141–177). Oxford, UK: Oxford University Press.
Bjorklund, D. (2005). *Children's thinking* (4th ed.). *Cognitive development and individual differences*. Belmont, CA: Wadsworth/Thomson Learning.
Brainerd, C. (2004). Dropping the other U: An alternative approach to U-shaped developmental functions. *Journal of Cognition and Development, 5*, 81–88.
Brooks, P., & MacWhinney, B. (2000). Phonological priming in children's picture naming. *Journal of Child Language, 27*, 335–366.
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: Lawrence Erlbaum/Taylor & Francis.
Burnham, D., & Dodd, B. (2004). Auditory–visual speech integration by prelinguistic infants: Perception of an emergent consonant in the McGurk effect. *Developmental Psychobiology, 44*, 209–220.
Cameron, S., & Dillon, H. (2007). Development of the Listening in Spatialized Noise–Sentences Test (LISN-S). *Ear and Hearing, 28*, 196–211.
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.
Campbell, R., & De Haan, E. (1998). Repetition priming for face speech images: Speech-reading primes face identification. *British Journal of Psychology, 89*, 309–323.
Campbell, R., Walker, J., & Baron-Cohen, S. (1995). The development of differential use of inner and outer face features in familiar face identification. *Journal of Experimental Child Psychology, 59*, 196–210.
Carey, S., Diamond, R., & Woods, B. (1980). Development of face recognition: A maturational component? *Developmental Psychology, 16*, 257–269.
Carroll, J. B., & White, M. N. (1973). Age-of-acquisition norms for 220 picturable nouns. *Journal of Verbal Learning and Verbal Behavior, 12*, 563–576.
Coltheart, M. (1981). The MRC psycholinguistic database. Retrieved August 9, 2006, Available from http://www.psy.uwa.edu.au/mrcdatabase/uwa_mrc.htm.
Conrad, R. (1971). The chronology of the development of covert speech in children. *Developmental Psychology, 5*, 398–405.
Conrad, R. (1977). Lipreading by deaf and hearing children. *British Journal of Educational Psychology, 47*, 60–65.
Cortese, M., & Fugett, A. (2004). Imageability ratings for 3000 monosyllabic words. *Behavioral Research Methods, Instruments, and Computers, 36*, 384–387.
Dale, P., & Fenson, L. (1996). Lexical development norms for young children. *Behavioral Research Methods, Instruments, and Computers, 28*, 125–127.
Damian, M., & Martin, R. (1999). Semantic and phonological codes interact in single word production. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 25*, 345–361.
de Gelder, B., & Morais, J. (1995). Speech and reading: One side to two coins. In B. de Gelder & J. Morais (Eds.), *Speech and reading: A comparative approach* (pp. 1–13). Hove, UK: Lawrence Erlbaum/Taylor & Francis.
de Gelder, B., Pourtois, G., Vroomen, J., & Bachoud-Levi, A. (2000). Covert processing of faces in prosopagnosia is restricted to facial expressions: Evidence from cross-modal bias. *Brain and Cognition, 44*, 425–444.
de Gelder, B., & Vroomen, J. (1998a). Impaired speech perception in poor readers: Evidence from hearing and speech reading. *Brain and Language, 64*, 269–281.
de Gelder, B., & Vroomen, J. (1998b). Impairment of speech-reading in prosopagnosia. *Speech Communication, 26*, 89–96.
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.
Desjardins, R., & Werker, J. (2004). Is the integration of heard and seen speech mandatory for infants? *Developmental Psychobiology, 45*, 187–203.
Dodd, B. (1977). The role of vision in the perception of speech. *Perception, 6*, 31–40.
Dodd, B. (1979). Lip reading in infants: Attention to speech presented in- and out-of-synchrony. *Cognitive Psychology, 11*, 478–484.
Dodd, B. (1980). Interaction of auditory and visual information in speech perception. *British Journal of Psychology, 71*, 541–549.
Dodd, B. (1987). The acquisition of lipreading skills by normally hearing children. In B. Dodd & R. Campbell (Eds.), *Hearing by eye: The psychology of lipreading* (pp. 163–175). London: Lawrence Erlbaum.
Dodd, B., Holm, A., Hua, Z., & Crosbie, S. (2003). Phonological development: A normative study of British English-speaking children. *Clinical Linguistics and Phonetics, 17*, 617–643.
Doherty-Sneddon, G., Bonner, L., & Bruce, V. (2001). Cognitive demands of face monitoring: Evidence for visuospatial overload. *Memory and Cognition, 29*, 909–919.
Doherty-Sneddon, G., & Kent, G. (1996). Visual signals and the communication abilities of children. *Journal of child psychology and psychiatry, 37*, 949–959.
Doherty-Sneddon, G., McAuley, S., Bruce, V., Langton, S., Blokland, A., & Anderson, A. (2000). Visual signals and children's communication: Negative effects on task outcome. *British Journal of Developmental Psychology, 18*, 595–608.
Dupont, S., Aubin, J., & Menard, L. (2005). A study of the McGurk effect in 4- and 5-year-old French Canadian children. *ZAS Papers in Linguistics, 40*, 1–17.
Durand, K., Gallay, M., Seigneuric, A., Robichon, F., & Baudouin, J. (2007). The development of facial emotion recognition: The role of configural information. *Journal of Experimental Child Psychology, 97*, 14–27.
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.

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.
Fisher, C. (1968). Confusions among visually perceived consonants. *Journal of Speech and Hearing Research, 11*, 796–804.
Gershkoff-Stowe, L., & Thelen, E. (2004). U-shaped changes in behavior: A dynamic systems perspective. *Journal of Cognition and Development, 5*, 11–36.
Gilhooly, K. J., & Logie, R. H. (1980). Age-of-acquisition, imagery, concreteness, familiarity, and ambiguity measures for 1,944 words. *Behavior Research Methods and Instrumentation, 12*, 395–427.
Glenberg, A., Schroeder, J., & Robertson, D. (1998). Averting the gaze disengages the environment and facilitates remembering. *Memory and Cognition, 26*, 651–658.
Gordon, P. (1990). Perceptual–motor processing in speech. In R. Proctor & T. Reeve (Eds.), *Stimulus–response compatibility*. North-Holland: Elsevier Science.
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, Vol. 2. Advances in the psychology of speechreading and auditory–visual speech (pp. 3-25). Hove, UK: Taylor & Francis.
Greenberg, H., & Bode, D. (1968). Visual discrimination of consonants. *Journal of Speech and Hearing Research, 11*, 869–874.
Hockley, N., & Polka, L. (1994). A developmental study of audiovisual speech perception using the McGurk paradigm. *Journal of the Acoustical Society of America, 96*, 3309.
Hollingshead, A. (1975). *Four factor index of social status*. New Haven, CT: Yale University, Department of Sociology.
Jerger, S., Lai, L., & Marchman, V. (2002). Picture naming by children with hearing loss: II. Effect of phonologically-related auditory distractors. *Journal of the American Academy of Audiology, 13*, 478–492.
Jerger, S., Martin, R., & Damian, M. (2002). Semantic and phonological influences on picture naming by children and teenagers. *Journal of Memory and Language, 47*, 229–249.
Jerger, S., Pearson, D., & Spence, M. (1999). Developmental course of auditory processing interactions: Garner interference and Simon interference. *Journal of Experimental Child Psychology, 74*, 44–67.
Jescheniak, J., Hahne, A., Hoffmann, S., & Wagner, V. (2006). Phonological activation of category coordinates during speech planning is observable in children but not in adults: Evidence for cascaded processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 32*, 373–386.
Jordan, T., & Bevan, K. (1997). Seeing and hearing rotated faces: Influences of facial orientation on visual and audiovisual speech recognition. *Journal of Experimental Psychology: Human Perception and Performance, 23*, 388–403.
Kuhl, P., & Meltzoff, A. (1982). The bimodal perception of speech in infancy. *Science, 218*, 1138–1141.
Levelt, W., Schriefers, H., Vorberg, D., Meyer, A., Pechmann, T., & Havinga, J. (1991). The time course of lexical access in speech production: A study of picture naming. *Psychological Review, 98*, 122–142.
Lickliter, R., & Bahrick, L. (2004). Perceptual development and the origins of multisensory responsiveness. In C. Calvert, C. Spence, & B. Stein (Eds.), *The handbook of multisensory processes* (pp. 643–654). Cambridge, MA: MIT Press.
Locke, J. (1993). *The child's path to spoken language*. Cambridge, MA: Harvard University Press.
MacLeod, A., & Summerfield, Q. (1987). Quantifying the contribution of vision to speech perception in noise. *British Journal of Audiology, 21*, 131–141.
Martin, R., Lesch, M., & Bartha, M. (1999). Independence of input and output phonology in word processing and short-term memory. *Journal of Memory and Language, 41*, 3–29.
Massaro, D. (1984). Children's perception of visual and auditory speech. *Child Development, 55*, 1777–1788.
Massaro, D. (1998). *Perceiving talking faces: From speech perception to a behavioral principle*. Cambridge, MA: MIT Press.
Massaro, D., Thompson, L., Barron, B., & Laren, E. (1986). Developmental changes in visual and auditory contributions to speech perception. *Journal of Experimental Child Psychology, 41*, 93–113.
McGregor, K., Friedman, R., Reilly, R., & Newman, R. (2002). Semantic representation and naming in young children. *Journal of Speech, Language, and Hearing Research, 45*, 332–346.
McGurk, H., & MacDonald, M. (1976). Hearing lips and seeing voices. *Nature, 264*, 746–748.
Melnick, K., Conture, E., & Ohde, R. (2003). Phonological priming in picture naming of young children who stutter. *Journal of Speech, Language, and Hearing Research, 46*, 1428–1443.
Merikle, P., & Reingold, E. (1991). Comparing direct (explicit) and indirect (implicit) measures to study unconscious memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 17*, 224–233.
Miller, G., & Nicely, P. (1955). An analysis of perceptual confusions among some English consonants. *Journal of the Acoustical Society of America, 27*, 338–352.
Mills, A. (1987). The development of phonology in the blind child. In B. Dodd & R. Campbell (Eds.), *Hearing eye: The Psychology of lipreading* (pp. 145–161). London: Lawrence Erlbaum.
Mondloch, C., Geldart, S., Maurer, D., & LeGrand, R. (2003). Developmental changes in face processing skills. *Journal of Experimental Child Psychology, 86*, 67–84.
Morais, J., Bertelson, P., Cary, L., & Alegria, J. (1986). Literacy training and speech segmentation. *Cognition, 24*, 45–64.
Morrison, C., Chappell, T., & Ellis, A. (1997). Age of acquisition norms for a large set of object names and their relation to adult estimates and other variables. *Quarterly Journal of Experimental Psychology A, 50*, 528–559.
Morrison, F., Smith, L., & Dow-Ehrensberger, M. (1995). Education and cognitive development: A natural experiment. *Developmental Psychology, 31*, 789–799.
Munhall, K., & Vatikiotis-Bateson, E. (1998). The moving face during speech communication. In R. Campbell, B. Dodd, & D. Burnham (Eds.), *Hearing by eye II: Advances in the psychology of speechreading and auditory–visual speech* (pp. 123–139). Hove, UK: Psychology Press.
Nusbaum, H. C., Pisoni, D. B., & Davis, C. K. (1984). Sizing up the Hoosier Mental Lexicon: Measuring the familiarity of 20,000 words (Research on Speech Perception Progress Report No. 10). Bloomington: Indiana University, Department of Psychology, Speech Research Laboratory.
Owens, E., & Blazek, B. (1985). Visemes observed by hearing-impaired and normal-hearing adult viewers. *Journal of Speech and Hearing Research, 28*, 381–393.

Patterson, M., & Werker, J. (1999). Matching phonetic information in lips and voice is robust in 4. *5-month-old infants. Infant Behavior and Development, 22*, 237–247.
Patterson, M., & Werker, J. (2003). Two-month-old infants match phonetic information in lips and voice. *Developmental Science, 6*, 191–196.
Ramirez, J., & Mann, V. (2005). Using auditory–visual speech to probe the basis of noise-impaired consonant–vowel perception in dyslexia and auditory neuropathy. *Journal of the Acoustical Society of America, 118*, 1122–1133.
Rosenblum, L., Schmuckler, M., & Johnson, J. (1997). The McGurk effect in infants. *Perception and Psychophysics, 59*, 347–357.
Schriefers, H., Meyer, A., & Levelt, W. (1990). Exploring the time course of lexical access in language production: Picture–word interference studies. *Journal of Memory and Language, 29*, 86–102.
Sekiyama, K., & Burnham, D. (2004). Issues in the development of auditory–visual speech perception: Adults, infants, and children. In S. H. Kim & D. H. Youn (Eds.). *Proceedings of International Conference on Spoken Language Processing 2004* (Vol. 2, pp. 1137–1140). Jeju Island, South Korea: International Conference on Spoken Language Processing.
Sloutsky, V., & Napolitano, A. (2003). Is a picture worth a thousand words? Preference for auditory modality in young children. *Child Development, 74*, 822–833.
Smit, A., Hand, L., Freilinger, J., Bernthal, J., & Bird, A. (1990). The Iowa Articulation Norms Project and its Nebraska replication. *Journal of Speech and Hearing Disorders, 55*, 779–798.
Smith, L., & Thelen, E. (2003). Development as a dynamic system. *Trends in Cognitive Sciences, 7*, 343–348.
Snodgrass, J. G., & Vanderwart, M. (1980). A standardized set of 260 pictures: Norms for name agreement, image agreement, familiarity, and visual complexity. *Journal of Experimental Psychology: Human Learning and Memory, 6*, 174–215.
Snowling, M., & Hulme, C. (1994). The development of phonological skills. *Philosophical Transactions of the Royal Society of London B, 346*, 21–27.
Starreveld, P. (2000). On the interpretation of onsets of auditory context effects in word production. *Journal of Memory and Language, 42*, 497–525.
Tye-Murray, N. (1998). *Foundations of aural rehabilitation*. San Diego: Singular Publishing Group.
Tye-Murray, N., & Geers, A. (2001). *Children's Audio–Visual Enhancement Test*. St. Louis, MO: Central Institute for the Deaf.
Vitevitch, M. S., & Luce, P. A. (2004). A web-based interface to calculate phonotactic probability for words and nonwords in English. *Behavior Research Methods, Instruments, and Computers, 36*, 481–487.
Weikum, W., Vouloumanos, A., Navarra, J., Soto-Faraco, S., Sebastian-Galles, N., & Werker, J. (2007). Visual language discrimination in infancy. *Science, 316*, 1159.
Wickens, C. (1974). Temporal limits of human information processing: A developmental study. *Psychological Bulletin, 81*, 739–755.
Wightman, F., Kistler, D., & Brungart, D. (2006). Informational masking of speech in children: Auditory–visual integration. *Journal of the Acoustical Society of America, 119*, 3940–3949.