JSLHR

Research Article

# Developmental Shifts in Detection and Attention for Auditory, Visual, and Audiovisual Speech

Susan Jerger,[a,b] Markus F. Damian,[c] Cassandra Karl,[a,b] and Hervé Abdi[a]

**Purpose:** Successful speech processing depends on our ability to detect and integrate multisensory cues, yet there is minimal research on multisensory speech detection and integration by children. To address this need, we studied the development of speech detection for auditory (A), visual (V), and audiovisual (AV) input.
**Method:** Participants were 115 typically developing children clustered into age groups between 4 and 14 years. Speech detection (quantified by response times [RTs]) was determined for 1 stimulus, /buh/, presented in A, V, and AV modes (articulating vs. static facial conditions). Performance was analyzed not only in terms of traditional mean RTs but also in terms of the faster versus slower RTs (defined by the 1st vs. 3rd quartiles of RT distributions). These time regions were conceptualized respectively as reflecting optimal detection with efficient focused attention versus less optimal detection with inefficient focused attention due to attentional lapses.
**Results:** Mean RTs indicated better detection (a) of multisensory AV speech than A speech only in 4- to 5-year-olds and (b) of A and AV inputs than V input in all age groups. The faster RTs revealed that AV input did not improve detection in any group. The slower RTs indicated that (a) the processing of silent V input was significantly faster for the articulating than static face and (b) AV speech or facial input significantly minimized attentional lapses in all groups except 6- to 7-year-olds (a peaked U-shaped curve). Apparently, the AV benefit observed for mean performance in 4- to 5-year-olds arose from effects of attention.
**Conclusions:** The faster RTs indicated that AV input did not enhance detection in any group, but the slower RTs indicated that AV speech and dynamic V speech (mouthing) significantly minimized attentional lapses and thus did influence performance. Overall, A and AV inputs were detected consistently faster than V input; this result endorsed stimulus-bound auditory processing by these children.

When children engage in face-to-face conversations, they typically detect, discriminate, and identify audiovisual (AV) speech sounds. Detection is the awareness that an AV speech event occurred, discrimination is the awareness that two AV speech sounds differ from each other, and identification is the labeling of the speech sounds. These different levels of speech perception tap different levels of linguistic processing, which are, at least to some extent, hierarchical, and children must detect and discriminate speech sounds before they can identify and label them (e.g., Aslin & Smith, 1988; Jerger, Martin, & Damian, 2002; McClelland & Elman, 1986; Stevenson, Sheffield, Butera, Gifford, & Wallace, 2017). Gogate, Walker-Andrews, and Bahrick's (2001) model of early word acquisition—as it relates to AV speech—is an example of this hierarchical perceptual analysis. The model proposes that, when infants detect the redundancies between speech sounds and their corresponding lip movements/mouth shapes, they can more readily discriminate similar-sounding phonological patterns, such as "pin" and "tin," and thus can recognize/label each pattern and associate it with its concept.

In short, lower level multisensory processes underpin higher level multisensory speech perception and word recognition skills, and altered lower level processes can have cascading effects onto these higher levels of processing. This relation is illustrated by the speech, language, and educational difficulties observed in children with early-onset hearing impairments and by the delayed expressive language skills observed in children with early-onset visual impairments (e.g., Briscoe, Bishop, & Norbury, 2001;

[a]School of Behavioral and Brain Sciences, GR4.1, University of Texas at Dallas, Richardson
[b]Callier Center for Communication Disorders, Richardson, TX
[c]School of Experimental Psychology, University of Bristol, United Kingdom
Correspondence to Susan Jerger: sjerger@utdallas.edu

Editor-in-Chief: Frederick (Erick) Gallun
Editor: Lori J. Leibold
Received September 10, 2017
Revision received January 2, 2018
Accepted July 16, 2018
https://doi.org/10.1044/2018_JSLHR-H-17-0343

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

Eimas & Kavanagh, 1986; Jerger et al., 2006; McConachie & Moore, 1994).

Despite the unquestionable contribution of detection and discrimination abilities to multisensory speech perception and word recognition, these lower levels of multisensory speech processing, particularly detection, are less well studied in children than the higher level speech recognition skills. The extant discrimination literature indicates that visual (V) speech (i.e., the articulatory gestures of talkers) benefits phoneme discrimination in individuals ranging in age from infancy (e.g., Teinonen, Aslin, Alku, & Csibra, 2008) to adulthood (e.g., Files, Tjan, Jiang, & Bernstein, 2015). In children, V speech improves feature contrast discrimination (e.g., vi vs. zi, a place feature contrast; Hnath-Chisolm, Laipply, & Boothroyd, 1998), vowel phoneme monitoring (Fort, Spinelli, Savariaux, & Kandel, 2010), and phoneme discrimination for visually distinct contrasts (e.g., ba vs. ga; LaLonde & Holt, 2015; but see Boothroyd, Eisenberg, & Martinez, 2010, for an exception).

With regard to age, improvements in the benefits from V speech have been observed for syllable/nonword discrimination up to 7 years by Hnath-Chisolm et al. (1998) but up to 10 years by Fort et al. (2010). In distinction to these results, however, Jerger, Damian, McAlpine, and Abdi (2018) recently demonstrated that V speech altered discrimination in all age groups from 4 to 14 years. These researchers administered a same–different syllable discrimination task, with the contrast of the critical syllable pair requiring children to discriminate a syllable with an intact /b/ onset (e.g., /b/i) from the same syllable but with a nonintact (spliced out) /-b/ onset (/-b/i). Results showed that the presence or absence of V speech was critical for perception: The addition of V speech to auditory (A) speech caused children to vote "same" when they listened to the intact–nonintact syllable pairs (e.g., /b/i–/-b/i), a configuration implying that V speech caused the nonintact onsets to be perceived as intact. The degree of this "visual speech fill-in effect" for the nonintact onsets predicted the children's receptive vocabulary skills.

In concert with the speech discrimination literature, the extant multisensory speech detection literature indicates that adults detect AV speech better than A speech (Bernstein, Auer, & Takayanagi, 2004; Grant, 2001; Grant & Seitz, 2000; Kim & Davis, 2003, 2004; LaLonde & Holt, 2016; Tjan, Chao, & Bernstein, 2013; Tye-Murray, Spehar, Myerson, Sommers, & Hale, 2011) and that infants detect equivalent phonetic information in A and V speech and changes in any mode (A, V, or AV speech) for at least some conditions (e.g., Kuhl & Meltzoff, 1982; Lewkowicz, 2000). In children, there is only one study that reported that 6- to 8-year-olds showed an adultlike detection advantage for AV relative to A speech (LaLonde & Holt, 2016).

Although there is a dearth of information about multisensory speech detection by children, there is a tenable child literature on the detection of nonspeech multisensory inputs, such as a noise and a light. This literature used *simple response times* to assess how quickly children can detect a preidentified sensory target and execute a preprogrammed motor response: Faster detection for the multisensory compared with unisensory inputs indicates multisensory facilitation. This literature reports that children aged roughly 7 years and older detect simultaneous A and V nonspeech inputs faster than unisensory inputs (Barutchu, Crewther, & Crewther, 2009; Barutchu et al., 2010, 2011; Brandwein et al., 2011; Gilley, Sharma, Mitchell, & Dorman, 2010). However, the degree of facilitation is smaller and more variable in children than in adults up to about 14–15 years of age.

In short, proficient speech detection is critical for children to have access to the AV cues that underpin speech and language development, yet multisensory speech detection remains understudied in children. To help address this gap in the literature, we studied the development of speech detection as quantified by *simple response times* for unisensory speech (A or V) versus multisensory speech (AV) in children from 4 to 14 years of age. The stimulus in our study consisted of the utterance "buh" presented in A, V, and AV modes. A primary research question was whether children show enhanced detection of AV speech relative to the unisensory inputs.

Such enhanced detection is supported by evoked potential evidence in adults revealing that inputs from the A and V modalities interact at both the early and late stages of sensory processing (e.g., Baart, Stekelenburg, & Vroomen, 2014; Molholm et al., 2002; van Wassenhove, Grant, & Poeppel, 2005). This pattern of evoked potential findings has been interpreted to indicate that multisensory speech perception is a multistaged process with general spatial and temporal AV speech correspondences interacting early in processing and phonetic AV speech features interacting later in processing (Baart et al., 2014; see also Schwartz, Berthommier, & Savariaux, 2004). We should acknowledge that these proposed stages of multisensory speech perception clearly occur before the behavioral response times of individuals, which makes it difficult (as pointed out by Schroger & Widmann, 1998) to specify the stage(s) of processing at which the A and V inputs are interacting. Our experimental design—the children responded to only one preidentified speech syllable "buh" presented in the A, V, or AV modes—clearly minimized the need for phonetic processing to identify the input. That said, as speech input unfolds, it automatically activates corresponding phonological representations according to the match between the evolving input and the representations in memory (e.g., Marslen-Wilson & Zwitserlood, 1989; McClelland & Elman, 1986). Thus, the A and V speech inputs of this research may interact at any or all stages of analysis (see also Davis & Kim, 2004; Reisberg, McLean, & Goldfield, 1987).

Another aspect of our experimental design was that the V input consisted of either the dynamic V speech that produced the auditory "buh" or the talker's static face. We included a static face not only as a control condition but also because different types of previous studies have observed some interesting differences between dynamic versus static faces. First, accuracy on a task monitoring for an A speech syllable in a carrier phrase is significantly better when adults view the talker's dynamic articulating face versus a static face (Davis & Kim, 2004). Second, although a dynamic articulating face and a visual symbol both enhance the detection of A speech in adults, the dynamic articulating face produces a relatively greater degree of multisensory facilitation (Bernstein et al., 2004; but see Tjan et al., 2013). Third, dynamic faces—relative to static faces—enhance the recognition of emotional expressions by adults and of unfamiliar faces by infants (Alves, 2013; Otsuka et al., 2009) possibly because (as proposed by O'Toole, Roark, & Abdi, 2002) motion may enhance the perceptual processing of faces and thus produce richer mental representations. Fourth, a dynamic articulating face generates more extensive cortical activation than a static face on functional magnetic resonance imaging scans (Calvert & Campbell, 2003; Campbell et al., 2001). Overall, the preponderance of this evidence predicts that performance in children may benefit more from the dynamic articulating face than from the static face.

Finally, we should note that dynamic faces are also more ecologically valid because they correspond to everyday social interactions, and this, in turn, may make them more attention provoking. In fact, some investigators propose that V speech may act as a type of alerting mechanism that boosts attention, which helps children detect and process information faster (Campbell, 2006; Wickens, 1974). Thus, we also expect some differential effects of attention on the dynamic versus static faces.

Attention is a key consideration because simple response time tasks as used herein are easy and monotonous—characteristics that are gold standards for assessing sustained attention (e.g., Betts, McKay, Maruff, & Anderson, 2006; Langner & Eickhoff, 2013; Manly et al., 2001). Sustained attention may be defined as "the ability to self-sustain mindful, conscious processing of stimuli whose repetitive, non-arousing qualities would otherwise lead to habituation and distraction." (Robertson, Manly, Andrade, Baddeley, & Yiend, 1997, p. 747). Typically, younger children find it more difficult to sustain attention, and so they may find a simple response task particularly taxing because of their immature frontal cortex, which may limit the use of more automatic strategies (Thillay et al., 2015).

Children continue to improve their capacity to sustain attention up to the preteen/teenage years, with much of the developmental change occurring before 10–11 years old (e.g., Betts et al., 2006; Manly et al., 2001; Thillay et al., 2015). Because of their immature sustained attention, younger children are more likely to experience difficulties in maintaining task goals, and this will increase the number of momentary lapses of attention and produce a larger number of slowed responses. Thus, the number of slowed responses is considered an index of these momentary attentional lapses (Key, Gustafson, Rentmeester, Hornsby, & Bess, 2017; Lewis, Reeve, Kelly, & Johnson, 2017; Venker et al., 2007; Weissman, Roberts, Visscher, & Woldorff, 2006). We predict that these occasional lapses producing slowed responses will create slower mean performance (based on all trials) in younger children than in preteens/teenagers. To the extent that dynamic faces are more richly encoded and more attention provoking than static faces, we predict that performance for the dynamic face will show fewer slowed responses. Below, we describe how we assessed our data on the development of speech detection (as defined by response times) for unisensory versus multisensory inputs with two complementary analyses.

Traditionally, the analysis of simple response times relies on a measure of central tendency—typically the mean (see Laurienti, Burdette, Maldjian, & Wallace, 2006; Miller, 1988). Thus, in the first analysis, we analyzed mean response times in the children divided into chronological age groups. In the second analysis, however, we augmented this traditional approach by an analysis of the faster versus slower response times. The second analysis was motivated by the observation that mean performance does not yield a pure measure of detection because, as noted above, the children's ability to detect sensory input depends on their ability to sustain focused attention (e.g., Barutchu et al., 2009; Betts et al., 2006; Thillay et al., 2015).[1] Researchers studying age-related changes in elderly individuals have also wrestled with the limitations of mean performance (e.g., Rabbitt & Goward, 1994; Rabbitt, Osman, Moore, & Stollery, 2001; Tse, Balota, Yap, Duchek, & McCabe, 2010). Results in this arena that studied faster versus slower response times suggested that elderly participants' fastest times are minimally affected by increasing chronological age and that differences in mean performance with age may disproportionately reflect differences in the number of slowed times (see, e.g., Rabbitt et al., 2001). In our second analysis, we interpreted results based on the rationale that optimal detection and efficient sustained focused attention are located in the faster times and less optimal detection with inefficient sustained focused attention due to attentional lapses is located in the slower times (see Tse et al., 2010, and Zhou & Krott, 2016, for a similar reasoning). Both analyses are introduced by Data Analytic Sections and Research Questions.

## Method

### *Participants*

Participants were 115 native English-speaking children ranging in age from 4;2 to 14;6 years;months (51% boys, 49% girls). The racial distribution was 84% White, 9% Asian, and 7% Black, with 9% reporting Hispanic ethnicity. Hearing sensitivity, visual acuity, auditory word recognition (Ross & Lerman, 1971), vocabulary skills (Dunn & Dunn, 2007), and visual perception (Beery & Beery, 2004) were within normal limits (age based when appropriate) in all participants. Normal hearing sensitivity was defined as bilaterally symmetrical thresholds of ≤ 20 dB HL at all test frequencies between 500 and 4000 Hz (American National Standards Institute, 2010). Normal binocular visual acuity (including

[1]A motor (key-press) component is also involved in the task, but it is assumed to be approximately constant within individuals and is not considered (e.g., Miller & Ulrich, 2003).

children with corrected vision) was defined as eight correct of 10 targets (five each at 20/20 and 20/25 acuity) on the Lea Symbols presented in a light box that provided self-calibrating uniform illumination for testing (e.g., Becker, Hubsch, Graf, & Kaufmann, 2002; Good-Lite Company, http://www.goodlite.com).

Participants were divided into four groups based on age (4- to 5-year-olds: $M$ = 4;11, $SD$ = 0.52, $n$ = 32; 6- to 7-year-olds: $M$ = 7;0, $SD$ = 0.59, $n$ = 25; 8- to 10-year-olds: $M$ = 9;3, $SD$ = 0.89, $n$ = 31; and 11- to 14-year-olds: $M$ = 12;5, $SD$ = 1.17, $n$ = 27). Advances in linguistic skills have been proposed to underlie developmental changes in sensitivity to V speech (e.g., Desjardins, Rogers, & Werker, 1997; Erdener & Burnham, 2013; Jerger, Damian, Spence, Tye-Murray, & Abdi, 2009), and our age groups represented four different linguistic stages:

> Four- to 5-year-olds: immature picture-book readers and immature speakers with articulatory deficiencies for complex sounds such as /sh/
>
> Six- to 7-year-olds: beginning readers whose phonology systems are reorganizing from phonemes as coarticulated indistinct speech sounds to phonemes as separable distinct written sounds and maturing speakers with good articulatory proficiency although with some disfluencies
>
> Eight- to 10-year-olds: maturing readers with a blossoming mastery of phonemes as written and spoken sounds and strong articulatory skills
>
> Eleven- to 14-year-olds: mature readers and speakers

Adults were not included because results in the 11- to 14-year-olds and young adults did not differ statistically. Because auditory response times vary as a function of loudness, we should note that average hearing sensitivity (pure-tone average score at 500, 1000, and 2000 Hz) was similar across the groups, ranging from 5.41 dB HL in 4- to 5-year-olds to 2.24 dB HL in 11- to 14-year-olds.

## *Materials and Instrumentation: Stimuli and Response Times*

### Recording

The stimulus "buh" was recorded—as part of a set of QuickTime (Apple Inc., 2001) movie files for associated projects—by an 11-year-old male actor with clearly intelligible speech without pubertal characteristics (f0 of 203 Hz). His full facial image and upper chest were recorded, and he started and ended each utterance with a neutral face/closed mouth. The color video signal was digitized at 30 frames per second with a 24-bit resolution at a 720 × 480 pixel size. The auditory signal was digitized at a 48-kHz sampling rate with a 16-bit amplitude resolution. The video track was routed to a high-resolution computer monitor, and the auditory track was routed through a speech audiometer to a loudspeaker atop the monitor (see Jerger, Damian, Tye-Murray, & Abdi, 2014, for further details). For this project, the stimulus started with the frame containing the auditory onset, and the talker's lips in this beginning frame remained closed but were no longer in a neutral position.

### Stimulus

The stimulus "buh" was presented in three modes: AV, A, and V. For the AV presentation, children saw and heard the talker; for the A presentation, the computer screen was blank; and for the V presentation, the loudspeaker was muted. Testing in these three modes was carried out in two separate conditions: one with a dynamic face articulating the utterance and one with an artificially static face (i.e., the child heard the same auditory track, but the video track was edited, with Adobe Premiere Pro [Adobe Systems Inc., 2003], to contain only the talker's still face and upper chest of the first frame). Hence, the two facial conditions consisted of presenting these two sets of items: (a) AV dynamic face, V dynamic face, and A (no face) or (b) AV static face, V static face, and A (no face). The A stimuli are the same in both facial conditions, thus allowing us to estimate test–retest reliability.

We formed one list of 39 test items (13 in each mode) for each facial (dynamic and static) condition (each list was presented forward and backward to yield two variations). The items of each list were randomized with the constraint that /buh/ was presented once in each mode for each triplet of items (e.g., two-triplet sequence = A/ AV/ V/ V/ A/ AV). This design ensured that any changes in performance due to personal factors (e.g., fatigue, practice) would be equally distributed over all modes.

### Response Times

To obtain response times, the computer triggered a counter/timer (resolution less than 1 ms) at the initiation of a stimulus. The stimulus continued until pressure on a response (telegraph) key stopped the counter/timer. The response board contained two keys separated by a distance of approximately 12 cm. A green square beside each key designated the start position for the child's hand. The key corresponding to the response (right vs. left) was counterbalanced across participants, and a small temporary box covered the unused key.

## *Procedure*

Testing was carried out within a double-walled sound-treated booth. The data of this study were gathered in one session of a multiple-day experimental protocol (e.g., Jerger et al., 2014; Jerger, Damian, Parra, & Abdi, 2017; Jerger, Damian, Tye-Murray, & Abdi, 2016, 2017). The presentation order of the facial conditions was counterbalanced across participants in each age group. One facial condition (either dynamic or static) was administered, followed by about 30 min of other testing and then by the administration of the other facial condition. For the formal testing, a tester sat at a computer workstation and initiated each trial, in an arrhythmic manner, when the child appeared ready by pressing a touch pad (out of the child's sight). A co-tester sat alongside each child to help keep the child "on task" at least overtly at the start of each trial—defined as sitting attentively and looking at the monitor with his or her hand on the start position. The children sat at a distance of 71 cm directly in front of a height-adjustable table containing the computer monitor and loudspeaker. The children's view of the talker's face subtended a visual angle of 7.17° vertically (eyebrow to chin) and 10.71° horizontally (eye level). The children heard the A input at an intensity of approximately 70 dB SPL.

The children were told that they would sometimes hear, sometimes see, and sometimes hear and see a boy. When the boy was talking, he would always be saying "buh." When they saw the boy, they were told that they would see a movie of the boy (dynamic face) for one facial condition and a photo of the boy (static face) for the other facial condition. Before each condition, the children were shown the stimulus for each mode (A, V, and AV). They were told to push the key as fast as possible to the onset of any of these targets with a whole-hand response (the tester illustrated and the child imitated). The children were told to always start with their hand on the green square and, as soon as they hit the key, to be sure to put their hand back on the square and get ready for the next target. Before the administration of each facial condition, practice trials were administered until response times had stabilized across a two-triplet sequence. Flawed trials (i.e., on rare occasions, the equipment malfunctioned or the child moved out of position to do something after the trial started) were deleted online and readministered at the end of the list.

### *Analysis of Mean Response Times*

#### Data Analysis

We compared mean performance in each mode for each facial condition. Mean values are preferred because median values can provide biased estimates for response time distributions with different skewness and/or different or small sample sizes (Miller, 1988; Whelan, 2008). The mean values are reported in the text/graphs because they clearly show how performance differed between the age groups and the modes, but for all inferential statistical analyses, the individual values were log transformed to normalize the distribution (Heathcote, Popiel, & Mewhort, 1991; Whelan, 2008). The Bonferroni correction controlled the familywise alpha (Abdi, Edelman, Valentin, & Dowling, 2009).

To determine whether AV speech produced faster detection for each facial condition, we evaluated the difference between response times in the AV mode minus the fastest unisensory mode as per the fixed favored dimension model for multidimensional stimuli (e.g., Biederman & Checkosky, 1970; Mordkoff & Yantis, 1993; Stevenson et al., 2014). Both the dynamic and static faces were viewed as multidimensional AV stimuli because individuals can accurately match unfamiliar voices to both dynamic and static unfamiliar faces well above chance; this pattern of results indicates that voices share source-identity information with both types of faces (Krauss, Freyberg, & Morsella, 2002; Mavica & Barenholtz, 2013; H. Smith, Dunn, Baguley, & Stacey, 2016a, 2016b; but see Lachs & Pisoni, 2004). Accurate voice–face matching would be particularly prominent in our children because they were familiar with the talker's face and voice from the other tasks they performed in our multiple-day experimental protocol. We predicted that the A response times would comprise the fastest unisensory mode because our pilot data in children and an extensive literature in adults indicate that response times are faster for the A than V mode (e.g., Diederich & Colonius, 2004; Harrar et al., 2014; Vickers, 2007; Woodworth & Schlosberg, 1954). Our research questions were as follows: (a) "Do children respond faster to A than V input as indicated in the adult literature?", (b) "Do children respond faster to AV input than to the fastest unisensory input?", (c) "Do children's response times differ in the facial conditions?", and (d) "Are children's response times reliable?"

## Results

### *Mean Response Times*

Figure 1 compares response times in the A, V, and AV modes for the static and dynamic faces in the four age groups and in the entire group. Statistical analyses (summarized in Table 1) were performed with a mixed-design analysis of variance (ANOVA) with one between-participant factor (age group: 4–5, 6–7, 8–10, and 11–14 years) and two within-participant factors (mode: V, A, and AV; facial condition: static vs. dynamic). Results revealed a significant age group effect, which occurred because response times (collapsed across modes and facial conditions) were slower in the younger than older children: Mean response times were 814 ms in 4- to 5-year-olds but 508 ms in 11- to 14-year-olds. A significant mode effect was also observed, which occurred because response times (collapsed across age groups and facial conditions) were significantly faster for the A and AV modes (592 and 577 ms, respectively) than for the V mode (752 ms). A straightforward interpretation of this latter result was complicated, however, by a significant mode × facial condition interaction, which occurred because mean response times (collapsed across age

**Figure 1.** Mean response times in the auditory (Aud), visual (Vis), and audiovisual (AV) modes for the static and dynamic faces in the four age groups and in all participants. The error bars are ± 1 *SEM*.

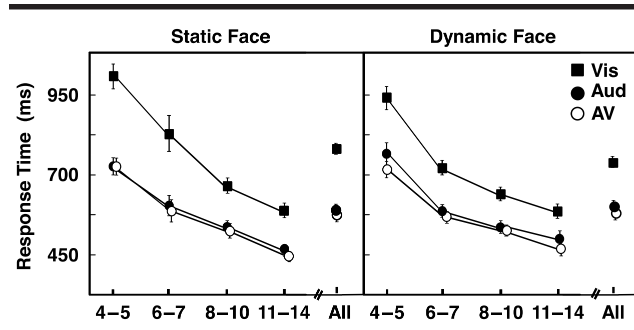


**Table 1.** Results of a mixed-design analysis of variance.

| Factors | *MSE* | *F* | *p* | Partial $\eta^2$ |
|---|---|---|---|---|
| Age group | 0.040 | 34.80 | < .0001 | .485 |
| Mode | 0.002 | 524.76 | < .0001 | .825 |
| Facial condition | 0.005 | 1.11 | *ns* | .011 |
| Mode × age group | 0.002 | 1.95 | *ns* | .051 |
| Facial condition × age group | 0.005 | 0.89 | *ns* | .023 |
| Mode × facial condition | 0.001 | 15.11 | < .0001 | .121 |
| Mode × facial condition × age group | 0.001 | 2.02 | *ns* | .050 |

*Note.* The analysis of variance contained one between-participant factor (age group: 4–5, 6–7, 8–10, and 11–14 years) and two within-participant factors (mode: visual, auditory, and audiovisual; facial condition: static vs. dynamic). The dependent variable was the log-transformed response times. The degrees of freedom were 3,111 for age group and facial condition × age group; 1,111 for facial condition; 2,222 for mode and facial condition × mode; and 6,222 for mode × age group and facial condition × mode × age group. Initially, we conducted this analysis with gender as a factor, but gender did not influence the results. Thus, gender was eliminated. *ns* = not significant.

groups; see "All" in Figure 1) were faster for the dynamic than static face for V input (728–776 ms) but not for A and AV inputs (587–597 ms for A input and 575–578 ms for AV input).

Below, as we turn to analyzing whether the unisensory inputs differed, the above results inform us about the V versus A modes. The significant mode effect indicated that the A response times were faster than the V response times. The significant mode × facial condition interaction indicated that this difference between the V and A response times was greater for the static face (189 ms) than the dynamic face (131 ms). There was no significant interaction involving the age groups; thus (as shown in 1), these significant differences characterized all groups. Below, we addressed whether the AV and A modes differed in any of the age groups or facial conditions.

**AV vs. A Modes**

To probe whether responses to AV input were faster than responses to A input (the fastest unisensory input), we carried out planned orthogonal contrasts for each facial condition in each age group (Abdi & Williams, 2010). Results indicated that the dynamic face (i.e., dynamic AV speech) was associated with faster responses only in 4- to 5-year-olds, $F_{\text{contrast}}(1, 110) = 9.73$, $MSE = 0.001$, $p = .002$, partial $\eta^2 = .042$. No other significant contrast was observed.

**Reliability**

To assess test–retest performance for the A response times, we reformatted the data to represent the first versus second tests (the two facial conditions were counterbalanced such that each occurred as the first test half of the time). The response times were statistically evaluated with a mixed-design ANOVA with one between-participant factor (age group: 4–5, 6–7, 8–10, and 11–14 years) and one within-participant factor (test: first vs. second). Results indicated that there was no significant effect of test nor any Test × Group interaction. A follow-up simple regression analysis (Abdi et al., 2009) in the entire group indicated that the children's A response times for the first and second tests were significantly correlated, $r = .840$, $F(1, 114) = 270.12$, $p < .0001$. The slope of the regression line was 0.768, which indicates that there was a 0.768-unit change in the second-session responses for each 1-unit change in the first-session responses. The variance (mean square) residual, or the degree of variability of the individual data about the regression line, was 0.004. The mean auditory response times for the first and second test sessions in the entire group were 599 ms ($SD$ = 190 ms) and 585 ms ($SD$ = 153 ms), and the individual difference scores for the first test minus the second test averaged 15 ms, with a 95% confidence interval ranging from −5 to 35 ms.

**Summary**

The children's mean response times became significantly faster as age increased—a result that agrees with previous findings (e.g., Goodenough, 1935; Jerger, Martin, & Pirozzolo, 1988). The children also responded faster to the A input than the V input—a pattern consistent with the literature noted above. This A-faster-than-V pattern of results was observed in 97%–98% of the children for the two facial conditions. With regard to whether the children responded faster to AV than A input, the addition of V speech was associated with faster responses but only in 4- to 5-year-olds. The AV-faster-than-A pattern of results in the dynamic facial condition was observed in 78% of the 4- to 5-year-olds. A silent V speech (i.e., mouthing) effect was also observed in that responses in the V mode were faster for the dynamic facial condition than the static facial condition. This mean pattern of results was observed in 67% of the children. The evaluation of test–retest performance established highly reliable results.

## Analysis of Faster vs. Slower Response Times

### *Data Analysis*

Mean performance in the above analyses may reflect a shift of the entire response time distribution or a shift of only the slow tail or the skewness of the distribution (e.g., see Balota & Yap, 2011; Rabbitt et al., 2001). We explored possible differences in the faster versus slower times with response time distributions computed by Vincentile analysis—a nonparametric technique that preserves the component distributions' shapes and does not make assumptions about the underlying distribution (see Jiang, Rouder, & Speckman, 2004; Ratcliff, 1979). Vincentile analysis is especially recommended because it provides stable estimates even with a small number of response times per participant/condition.

To obtain the Vincentile distributions, each child's response times—for each mode/facial condition—were rank-ordered and then initially divided into sequential bins of 10% (deciles). A cumulative distribution function (CDF) was obtained for each age group by averaging each of the bins across the participants in that group for each facial condition/mode. In Appendix A, the CDFs for the A, AV, and V modes in the static (see Panel A) and dynamic (see Panel B) facial conditions for all age groups are portrayed. In adults, CDFs such as these are explored with ex-Gaussian analyses of the response distributions, but we did not have a sufficient number of trials to conduct this type of analysis (Heathcote et al., 1991). Thus, we computed another set of Vincentile distributions by dividing each child's rank-ordered response times—for each mode/facial condition—into sequential bins of 25% (quartiles). Statistically, we investigated whether our effects of interest appeared in the faster and/or slower response times by analyzing the 25th and 75th (i.e., first and third) quartiles of the Vincentile CDFs. Again, our assumptions for interpreting the results are that optimal detection and efficient focused attention are located in the faster times (first quartile) and less optimal detection with inefficient focused attention due to attentional lapses is located in the slower times (third quartile). We were interested in whether the pattern of mean results reported above was observed at both quartiles (results influenced by both detection and attention) or at only one of the quartiles (results influenced by only detection or attention). To assess this, we carried out contrast analyses (Abdi & Williams, 2010) on the log-transformed response times at the first/faster and third/slower quartiles for each facial condition in each age group with a Bonferroni correction to control the familywise alpha. Our focused research questions were as follows: (a) "Do the A versus V inputs differ in the age groups at both quartiles or only the first/faster or third/slower quartile?", (b) "Do the AV versus fastest unisensory input differ in any age group at one or both quartiles?", and (c) "Does the facial condition affect these results?"

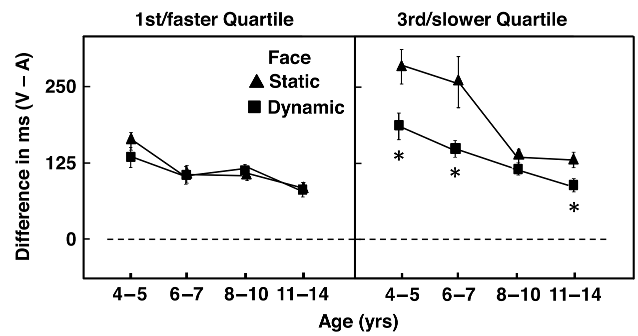
## Results

### *Faster vs. Slower Response Times*

#### V vs. A Modes

Figure 2 shows the mean difference scores (V response times − A response times) in the age groups at each quartile for the static and dynamic facial conditions. Appendix B presents the $F_{\text{contrast}}$ results for the V versus A modes. The large positive difference scores in Figure 2, along with the statistical results, documented that the V response times were significantly slower than the A response times in all age groups at both quartiles for both facial conditions. Relative to the V input, the A input was detected faster and with significantly fewer attentional lapses (see also CDFs in Appendix A). Faster A-than-V responses were observed in about 97% of children for both facial conditions at both quartiles.

**Figure 2.** The mean difference scores (visual [V] response times − auditory [A] response times) in the age groups for the static and dynamic faces at the first/faster and third/slower quartiles of the cumulative distribution functions. The error bars are ± 1 *SEM*. Every data point showed a significant difference for the V versus A modes. An asterisk indicates the data points showing a significant difference for the static versus dynamic silent faces.

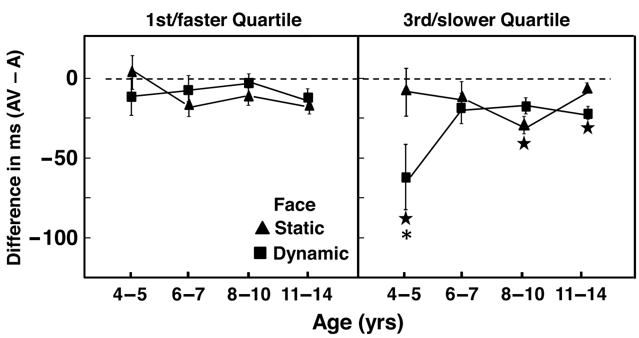



As indicated by the asterisks in Figure 2 and as documented by the $F_{\text{contrast}}$ results for the dynamic versus static faces in Appendix C, dynamic V speech—relative to a static face—decreased the mean difference scores significantly at the third quartile/slower responses but not at the first quartile/faster responses, with the exception of results in 8- to 10-year-olds, which did not differ for the facial conditions at either quartile. These results indicate that dynamic V speech captured attention and reduced attentional lapses more than the static face, with about 75% of children, not including the 8- to 10-year-olds, showing this pattern of results. Reasons for the different patterns of results in 8- to 10-year-olds are unclear, and indeed, about 60% of these children showed the typical pattern of results for the dynamic versus static facial conditions.

#### AV vs. A Modes

Figure 3 shows the mean difference scores (AV response times − A response times) in the age groups at each quartile for the static and dynamic facial conditions. Appendix D presents the $F_{\text{contrast}}$ results for the AV versus A modes. Statistical findings in Appendix D and the differences scores in Figure 3 for the first/faster quartile showed that multisensory AV input did not improve detection in any age group. With regard to the third/slower quartile, AV dynamic speech captured and benefited attention in 4- to 5-year-olds and 11- to 14-year-olds, and static facial input benefited attention

**Figure 3.** The mean difference scores (audiovisual [AV] response times – auditory [A] response times) in the age groups for the static and dynamic faces at the first/faster and third/slower quartiles of the cumulative distribution functions. The error bars are ± 1 *SEM*. Stars indicate the data points showing a significant difference for the AV versus A modes; an asterisk indicates the data point showing a significant difference for the static versus dynamic faces.




in 8- to 10-year-olds. This pattern of results was observed in about 75% of children in each of these age groups. Finally, as indicated by the asterisk in Figure 3 and as documented by the $F_{contrast}$ results for the dynamic versus static faces in Appendix E, differences between the facial conditions achieved statistical significance only in 4- to 5-year-olds at the third/slower quartile, with 55% of these children showing a greater difference score for the dynamic face.

Because we know little about the influence of attention on AV multisensory speech perception by children, we reassessed the results in Figure 3 at the third/slower quartile with a mixed-design ANOVA with one between-participant factor (age group: 4–5, 6–7, 8–10, and 11–14 years) and two within-participant factors (mode: A vs. AV; facial condition: static vs. dynamic).[2] As always, the individual values were log transformed to normalize the distribution (Heathcote et al., 1991; Whelan, 2008), and the Bonferroni correction controlled the familywise alpha (Abdi et al., 2009). However, two considerations influenced how we carried out the current Bonferroni correction. First, a standard omnibus ANOVA is a nonspecific, global test that seeks any differences within or between factors (even ones that are not of interest) and suffers from low statistical power relative to procedures that decompose the systematic variance into meaningful contrasts (Rosenthal, Rosnow, & Rubin, 2000). Second, false negatives can be a more fundamental problem than false positives in an area with little evidence because they may retard further meaningful growth of knowledge (Fiedler, Kutzner, & Krueger, 2012). Thus, as recommended when some $F$ values in an omnibus ANOVA are more important than others a priori, we allocated the individual alphas per family of tests unequally for the Bonferroni correction (Abdi & Williams, 2007). We tested the critical mode × facial condition × Age Group interaction with an $\alpha = .04$ and shared the remaining .01 between the other $F$ tests, which were evaluated with an $\alpha = .0017$.

Statistical findings are summarized in Table 2. Results revealed a significant age group effect, which occurred because response times (collapsed across modes and facial conditions) were slower in younger than older children, as noted previously. A significant mode effect was also observed, which occurred because response times (collapsed across age groups and facial conditions) were significantly faster for the AV mode than the A mode (591 and 615 ms, respectively). A straightforward interpretation of this latter result was complicated, however, by a significant mode × facial condition × age group interaction, which indicated that the relationship between the AV and A response times differed for the facial conditions but in inconsistent ways across the age groups. Critically, this interaction points out that the relationship between the AV and A response times varied across the age groups. To probe this pattern of interaction, we conducted $t$ tests on the difference between the AV versus A response times in each age group for each facial condition. Results are summarized in Table 3. Results mirrored the previously obtained $F_{contrast}$ findings. The significant differences between the AV and A response times indicated that AV dynamic speech benefited attention in 4- to 5-year-olds and 11- to 14-year-olds, and static facial input benefited attention in 8- to 10-year-olds. In short, facial input (either AV dynamic speech or a static face) significantly influenced attention in all age groups, except the 6- to 7-year-olds.

## Discussion

Everyday tasks depend on our ability to detect and integrate information from multiple sensory modalities. Despite the acknowledged importance of this lower level of processing for speech, however, we know little about children's multisensory speech detection abilities. The purpose of this research was to study the development of speech detection for A, V, and AV inputs in children from 4 to 14 years of age. Our experimental design featured two novel approaches. First, our V input consisted of both static and dynamic faces, which allowed us to determine whether effects on performance reflected a facial effect or an articulating face-specific effect (influenced only by the dynamic face). Second, we assessed development not only in terms of the traditional mean response times but also in terms of the faster versus slower response times. We should acknowledge that some of the slower response times in these children may have been reflecting motivational factors rather than attentional lapses (see Reinvang, 1998). This research, however, minimized this possibility by having a co-tester who tried to keep the children engaged in the task. We should also note that there were only 13 trials per condition (78 trials in total) due to the limited testing time available with young children. Importantly, however, we selected a technique (Vincentizing) that is especially suitable for analyzing data with only a few observations per condition (i.e., it has been shown that Vincentizing

[2]We thank one of the reviewers for recommending this analysis.

**Table 2.** Results of a mixed-design analysis of variance.

| Factors | *MSE* | *F* | *p* | Partial $\eta^2$ |
|---|---|---|---|---|
| Age group | 0.028 | 31.14 | < .0001 | .462 |
| Mode | 0.001 | 29.51 | < .0001 | .210 |
| Facial condition | 0.004 | 0.47 | *ns* | .005 |
| Mode × age group | 0.001 | 0.59 | *ns* | .012 |
| Facial condition × age group | 0.004 | 0.68 | *ns* | .018 |
| Mode × facial condition | 0.001 | 3.85 | *ns* | .034 |
| Mode × facial condition × age group | 0.001 | 3.47 | .018 | .086 |

*Note.* The analysis of variance contained one between-participant factor (age group: 4–5, 6–7, 8–10, and 11–14 years) and two within-participant factors (mode: auditory vs. audiovisual; facial condition: static vs. dynamic). The dependent variable was the log-transformed response times at the third/slower quartile. The degrees of freedoms were 3,111 for age group, mode × age group, facial condition × age group, and mode × facial condition × age group and 1,111 for mode, facial condition, and mode × facial condition. *ns* = not significant.

provides stable estimates even with only 10–20 trials per participant/condition; see Jiang et al., 2004; Ratcliff, 1979).

We discuss the results below in terms of the unisensory inputs (V vs. A) and the multisensory input versus the fastest unisensory input (AV vs. A). A focus is to understand how the results for the first/faster and third/slower quartiles contributed to the interpretation of mean performance in children. These two time regions were respectively conceptualized as reflecting optimal detection with efficient focused attention versus less optimal detection with inefficient focused attention due to attentional lapses.

## *V vs. A Inputs*

Mean performance in the age groups indicated significantly faster A than V response times and significantly faster V responses for the silent dynamic face (i.e., mouthing) than the static face. The A-faster-than-V outcome agrees with long-term previous findings in adults (Diederich & Colonius, 2004; Harrar et al., 2014; Vickers, 2007; Woodworth & Schlosberg, 1954; see Brandwein et al., 2011, and Gilley et al., 2010, for exceptions). Analysis of the faster versus slower response times indicated that (a) A input relative to V input not only facilitated the children's ability to detect the input but also reduced their attentional lapses, whereas (b) silent dynamic V speech (mouthing) relative to a static face only reduced attentional lapses. This latter finding supports the proposal that a dynamic face may be more richly encoded and thus more attention provoking than a static face (Calvert & Campbell, 2003; Campbell et al., 2001; O'Toole et al., 2002). Overall, the pattern of results implies that the changes in mean performance could be reflecting the effects of detection and/or attention.

**Table 3.** Results of paired *t* tests in each age group for each facial condition.

| Facial condition | *t* | *p* | Partial $\eta^2$ |
|---|---|---|---|
| | 4–5 years | | |
| Static face | 0.38 | *ns* | .001 |
| Dynamic face | 3.19 | .003 | .246 |
| | 6–7 years | | |
| Static face | 1.42 | *ns* | .053 |
| Dynamic face | 1.50 | *ns* | .091 |
| | 8–10 years | | |
| Static face | 4.69 | < .0001 | .421 |
| Dynamic face | 2.44 | *ns* | .154 |
| | 11–14 years | | |
| Static face | 0.63 | *ns* | .015 |
| Dynamic face | 3.28 | .003 | .294 |

*Note.* The dependent variable was the log-transformed response times at the third/slower quartile for the audiovisual versus auditory modes. *ns* = not significant.

The significantly faster speed of processing for A than V input strongly supports stimulus-bound auditory processing and an automatic capture of attention by A input in these children (e.g., Napolitano & Sloutsky, 2004; Sloutsky & Napolitano, 2003). These results are reminiscent of the auditory distraction literature in adults (e.g., Macken, Phelps, & Jones, 2009; Watkins, Dalton, Lavie, & Rees, 2007), which emphasizes the capacity of A input to capture attention despite adults' attempts to "not listen." Such findings have impactful implications for speech and language development in children. As an example—if we view the speech input more narrowly as A only and the V input more broadly as environmental objects—pretend that a parent looks and points to an object while saying "lamp" to his or her preschoolers. The V input in this example is permanent, but the A input is fleeting. If the children fail to see the "lamp" at first glance, they can easily see it by taking another look. If, however, the children fail to hear the word at first listen, they cannot easily hear it by taking another listen. Thus, the automatic capture of attention by A input in young children may critically nurture speech and language development because it helps children perceive words that are "written on the wind."

The unequal detection of the A and V dimensions of speech in this research may reflect, at least to some degree, the conscious behaviors demanded by our experimental protocol. That said—to the extent that these results generalize to AV speech perception with its more unconscious detection of the A and V dimensions—these results may inform the interpretation of studies that manipulated the onsets of the A and V cues and found that individuals are more likely to synthesize these cues when the V speech starts before the A speech than vice versa. For example, in adults, AV interactions occur even when the V speech leads the A speech by 170–180 ms (Munhall, Gribble, Sacco, & Ward, 1996; van Wassenhove, Grant, & Poeppel, 2007). In contrast, when A speech leads the V speech, AV interactions occur only up to an asynchrony of 30 ms (e.g., van Wassenhove et al., 2007). This pattern of AV interactions for asynchronous speech appears to be adultlike by 7 years of age, although children do not show the same degree of AV interactivity as adults (Hillock-Dunn, Grantham, & Wallace, 2016). A greater tolerance for V-speech–leading asynchronies seems to have ecological validity because V cues frequently start before A cues in everyday speech (e.g., Bell-Berti & Harris, 1981). That said, the current research suggests that the greater tolerance of V-speech–leading asynchronies may also be reflecting people's slowness in detecting V speech relative to A speech.

### *AV vs. A Inputs*

Mean performance showed that response times were faster for dynamic AV input than A input but only in 4- to 5-year-olds. Analysis of the faster and slower times, however, indicated that AV dynamic speech did not influence detection (i.e., responses at the first/faster quartile) in any group. These results disagree with one previous study of speech detection by children, which reported adultlike benefits from AV speech in 6- to 8-year-olds on a task requiring detection of speech in noise (Lalonde & Holt, 2016). Our results also show a different developmental course from the one characterizing the detection advantage for nonspeech multisensory A and V inputs. The nonspeech child literature was introduced because there are few multisensory speech detection studies in children. We should note, however, that this nonspeech A and V literature cannot be directly related to the AV speech findings because speech dimensions/cues are processed in an interdependent (conjoined) manner (Garner, 1974; Green & Kuhl, 1989; Jerger, Martin, Pearson, & Dinh, 1995; Jerger et al., 1993; Tomiak, Mullennix, & Sawusch, 1987), whereas arbitrarily paired inputs such as a noise and a light are typically processed in an independent (separable) manner (e.g., Garner, 1974; Marks, 2004). Thus, our different results are difficult to interpret due to the pronounced task differences along with different perceptual processing structures that preclude an unambiguous comparison of speech versus nonspeech research.

With regard to the third quartile/slower response times, AV dynamic speech captured attention and thus significantly minimized slowed responses relative to A speech in 4- to 5-year-olds and 11- to 14-year-olds. This AV effect seems reminiscent of the U-shaped curve we observed previously in which AV phonologically related speech distractors primed picture naming in 4- to 5-year-olds and 10- to 14-year-olds but not in children of in-between ages (Jerger et al., 2009). The current results, however, additionally revealed that AV static facial input significantly minimizes attentional lapses and thus slowed responses in 8- to 10-year-olds as well. In short, V speech or facial input relative to A speech significantly impacted results in all age groups except in 6- to 7-year-olds (a peaked U-shaped curve).

Previously, Jerger et al. (2009) related their U-shaped results to dynamic systems theory (e.g., L. Smith & Thelen, 2003), which proposes that (a) multiple factors typically underlie developmental change and (b) a lack of any effect in children may be reflecting a period of transition (not a lack of effect) during which immature knowledge and processing subsystems are reorganized and restructured into more mature, elaborated, and robust forms. During these developmental transitions, processing systems are less robust, and children cannot easily use their cognitive resources; consequently, during these transitional stages, children's performance can be unstable and affected by methodological approaches and task demands (Evans, 2002).

We propose that the developmental shifts in AV performance for the slowed times reflect different stages of reorganization and transition. With regard to 4- to 5-year-olds and 11- to 14-year-olds, we should note that alike performance in these groups may not be reflecting alike underlying mechanisms. Whereas performance in 11- to 14-year-olds is mature and reflects dynamic AV speech capturing attention and minimizing attentional lapses, performance in 4- to 5-year-olds is immature and may be reflecting a dynamic AV speech effect and/or other factors. For example, 3-year-olds and thus perhaps 4- to 5-year-olds attend preferentially to dynamic over static faces (Libertus, Landa, & Haworth, 2017), and younger children with less mature articulatory proficiency observe V speech more, perhaps to cement their knowledge of the acoustic consequences of articulatory gestures (Desjardins et al., 1997; Dodd, McIntosh, Erdener, & Burnham, 2008).

Performance in 6- to 7-year-olds did not show any influence of either type of face, but performance in 8- to 10-year-olds revealed the minimization of attentional lapses by AV static facial input—an effect that may reflect the simultaneous or correlated onsets interacting to produce a more emphatic onset-alerting signal. As noted previously, voices share source-identity information with both the dynamic and static faces (Krauss et al., 2002; Mavica & Barenholtz, 2013; H. Smith et al., 2016a, 2016b). We propose that the different results in 6- to 7-year-olds and 8- to 10-year-olds occurred because the relevant knowledge and processing subsystems, particularly phonology, were reorganizing between roughly 6 and 9 years of age into more mature resources for a wider range of activities (see Jerger et al., 2009, for a discussion and references). Phonological processes are particularly relevant because, although this task minimized phonological processing demands, speech input automatically activates corresponding phonological representations as it unfolds, as noted previously (e.g., Marslen-Wilson & Zwitserlood, 1989; McClelland & Elman, 1986). Thus, the A and V inputs of this research may interact at multiple stages of analysis, which can also be influenced by cognitive resources such as attention (e.g., Davis & Kim, 2004; Reisberg et al., 1987). Finally, we should acknowledge that both this research and the Jerger et al. (2009) research studied response times. The measurement of processing speed can be a more sensitive measure of task proficiency. That said, all methods of identifying and quantifying multisensory interactions have advantages and disadvantages (Stevenson et al., 2014).

## Conclusions

These results emphasized the pronounced ability of both AV speech and silent dynamic V speech (mouthing) to minimize attentional lapses and thus influence detection. Such findings demonstrate the usefulness of V speech even in situations that do not involve impoverished A input. Another primary result was that response times were always faster to A and AV inputs than V input. Our overall results strongly endorsed stimulus-bound auditory processing by these children. Such findings are good news for children who must listen to learn.

## Acknowledgments

This research was supported by the National Institute on Deafness and Other Communication Disorders Grant DC-000421 to the University of Texas at Dallas. We thank the children and parents who participated and the researchers who assisted, namely, Aisha Aguilera, Carissa Dees, Nina Dinh, Nadia Dunkerton, Alycia Elkins, Brittany Hernandez, Demi Krieger, Rachel Parra McAlpine, Michelle McNeal, Jeffrey Okonye, 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 Drs. Brent Spehar and Nancy Tye-Murray of the Washington University School of Medicine (computer programming and stimuli recording/editing).

## References

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

**Abdi, H., & Williams, L.** (2007). Bonferroni and Sidak corrections for multiple comparisons. In N. Salkind (Ed.), *Encyclopedia of measurement and statistics* (pp. 103–107). Thousand Oaks, CA: Sage.

**Abdi, H., & Williams, L.** (2010). Contrast analysis. In N. Salkind (Ed.), *Encyclopedia of research design* (pp. 243–251). Thousand Oaks, CA: Sage.

**Adobe Systems Inc.** (2003). Adobe Premiere Pro [Computer software]. Retrieved from https://www.adobe.com/products/premiere.html

**Alves, N.** (2013). Recognition of static and dynamic facial expressions: A study review. *Estudos de Psicologia, 18,* 125–130.

**American National Standards Institute.** (2010). *Specifications for audiometers* (ANSI/ASA S3.6-2010 [R2010]). New York, NY: Author.

**Apple Inc.** (2001). QuickTime File Format Specification (Classic version) [Computer software]. Retrieved from https://developer.apple.com

**Aslin, R., & Smith, L.** (1988). Perceptual development. *Annual Review of Psychology, 39,* 435–473.

**Baart, M., Stekelenburg, J., & Vroomen, J.** (2014). Electrophysiological evidence for speech-specific audiovisual integration. *Neuropsychologia, 53,* 115–121.

**Balota, D., & Yap, M.** (2011). Moving beyond the mean in studies of mental chronometry: The power of response time distributional analyses. *Current Directions in Psychological Science, 20,* 160–166.

**Barutchu, A., Crewther, D., & Crewther, S.** (2009). The race that precedes coactivation: Development of multisensory facilitation in children. *Developmental Science, 12,* 464–473.

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

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

**Becker, R., Hubsch, S., Graf, M., & Kaufmann, H.** (2002). Examination of young children with Lea symbols. *British Journal of Ophthalmology, 86,* 513–516.

**Beery, K., & Beery, N.** (2004). *The Beery–Buktenica Developmental Test of Visual–Motor Integration with Supplemental Developmental Tests of Visual Perception and Motor Coordination–Fifth Edition (Beery VMI 5)*. Minneapolis, MN: NCS Pearson.

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

**Bernstein, L., Auer, E., Jr., & Takayanagi, S.** (2004). Auditory speech detection in noise enhanced by lipreading. *Speech Communication, 44,* 5–18.

**Betts, J., McKay, J., Maruff, P., & Anderson, V.** (2006). The development of sustained attention in children: The effect of age and task load. *Child Neuropsychology, 12,* 205–221.

**Biederman, I., & Checkosky, S.** (1970). Processing redundant information. *Journal of Experimental Psychology, 83,* 486–490.

**Boothroyd, A., Eisenberg, L., & Martinez, A.** (2010). An on-line imitative test of speech-pattern contrast perception (OlimSpac): Developmental effects in normally hearing children. *Journal of Speech, Language, and Hearing Research, 53,* 531–542.

**Brandwein, A., Foxe, J., Russo, N., Altschuler, T., Gomes, H., & Molholm, S.** (2011). The development of audiovisual multisensory integration across childhood and early adolescence: A high-density electrical mapping study. *Cerebral Cortex, 21,* 1042–1055.

**Briscoe, J., Bishop, D., & Norbury, C.** (2001). Phonological processing, language, and literacy: A comparison of children with mild-to-moderate sensorineural hearing loss and those with specific language impairment. *Journal of Child Psychology and Psychiatry and Allied Disciplines, 42,* 329–340.

**Calvert, G., & Campbell, R.** (2003). Reading speech from still and moving faces: The neural substrates of visible speech. *Journal of Cognitive Neuroscience, 15,* 57–70.

**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, the Netherlands: Elsevier.

**Campbell, R., MacSweeney, M., Surguladze, S., Calvert, G., McGuire, P., Suckling, J., . . . David, A.** (2001). Cortical substrates for the perception of face actions: An fMRI study of the specificity of activation for seen speech and for meaningless lower-face acts (gurning). *Cognitive Brain Research, 12,* 233–243.

**Davis, C., & Kim, J.** (2004). Audio-visual interactions with intact clearly audible speech. *The Quarterly Journal of Experimental Psychology: Human Experimental Psychology, 57A*(6), 1103–1121.

**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.

**Diederich, A., & Colonius, H.** (2004). Bimodal and trimodal multisensory enhancement: Effects of stimulus onset and intensity on reaction time. *Perception & Psychophysics, 66,* 1388–1404.

**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). *Peabody Picture Vocabulary Test–Fourth Edition (PPVT-4)*. Minneapolis, MN: Pearson.

**Eimas, P., & Kavanagh, J.** (1986). Otitis media, hearing loss, and child development: A NICHD conference summary. *Public Health Reports, 101,* 289–293.

**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.

**Evans, J.** (2002). Variability in comprehension strategy use in children with SLI: A dynamical systems account. *International Journal of Language & Communication Disorders, 37,* 95–116.

**Fiedler, K., Kutzner, F., & Krueger, J.** (2012). The long way from α-error control to validity proper: Problems with short-sighted false-positive debate. *Perspectives on Psychological Science, 7,* 661–669.

**Files, B., Tjan, B., Jiang, J., & Bernstein, L.** (2015). Visual speech discrimination and identification of natural and synthetic consonant stimuli. *Frontiers in Psychology, 6,* 878. https://doi.org/10.3389/fpsyg.2015.00878

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

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

**Gilley, P., Sharma, A., Mitchell, T., & Dorman, M.** (2010). The influence of a sensitive period for auditory–visual integration in children with cochlear implants. *Restorative Neurology and Neuroscience, 28,* 207–218.

**Gogate, L., Walker-Andrews, A., & Bahrick, L.** (2001). The intersensory origins of word comprehension: An ecological-dynamic systems view. *Developmental Science, 4,* 1–37.

**Goodenough, F.** (1935). The development of the reactive process from early childhood to maturity. *Journal of Experimental Psychology, 18,* 431–450.

**Grant, K.** (2001). The effect of speechreading on masked detection thresholds for filtered speech. *The Journal of the Acoustical Society of America, 109,* 2272–2275.

**Grant, K., & Seitz, P.** (2000). The use of visible speech cues for improving auditory detection of spoken sentences. *The Journal of the Acoustical Society of America, 108,* 1197–1208.

**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.

**Harrar, V., Tammam, J., Perez-Bellido, A., Pitt, A., Stein, J., & Spence, C.** (2014). Multisensory integration and attention in developmental dyslexia. *Current Biology, 24,* 531–535.

**Heathcote, A., Popiel, S., & Mewhort, D.** (1991). Analysis of response time distributions: An example using the Stroop task. *Psychological Bulletin, 109,* 340–347.

**Hillock-Dunn, A., Grantham, D., & Wallace, M.** (2016). The temporal binding window for audiovisual speech: Children are like little adults. *Neuropsychologia, 88,* 74–82.

**Hnath-Chisolm, T., Laipply, E., & Boothroyd, A.** (1998). Age-related changes on a children's test of sensory-level speech perception capacity. *Journal of Speech, Language, and Hearing Research, 41,* 94–106.

**Jerger, S., Damian, M., McAlpine, R., & Abdi, H.** (2018). Visual speech fills in both discrimination and identification of non-intact auditory speech in children. *Journal of Child Language, 45,* 392–414.

**Jerger, S., Damian, M., Parra, R., & Abdi, H.** (2017). Visual speech alters the discrimination and identification of non-intact auditory speech in children with hearing loss. *International Journal of Pediatric Otorhinologyngology, 94,* 127–137.

**Jerger, S., Damian, M., 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., Damian, M., Tye-Murray, N., & Abdi, H.** (2014). Children use visual speech to compensate for non-intact auditory speech. *Journal of Experimental Child Psychology, 126,* 295–312.

**Jerger, S., Damian, M., Tye-Murray, N., & Abdi, H.** (2016). Phonological priming in children with hearing loss: Effect of speech mode, fidelity, and lexical status. *Ear and Hearing, 37,* 623–633.

**Jerger, S., Damian, M., Tye-Murray, N., & Abdi, H.** (2017). Children perceive speech onsets by ear and eye. *Journal of Child Language, 44,* 185–215.

**Jerger, S., Damian, M., Tye-Murray, N., Dougherty, M., Mehta, J., & Spence, M.** (2006). Effects of childhood hearing loss on organization of semantic memory: Typicality and relatedness. *Ear and Hearing, 27,* 686–702.

**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., Martin, R., Pearson, D., & Dinh, T.** (1995). Childhood hearing impairment: Auditory and linguistic interactions during multidimensional speech processing. *Journal of Speech & Hearing Research, 38,* 930–948.

**Jerger, S., Martin, R., & Pirozzolo, F.** (1988). A developmental study of the auditory Stroop effect. *Journal of Brain and Language, 35,* 86–104.

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

**Jiang, Y., Rouder, J., & Speckman, P.** (2004). A note on the sampling properties of the Vincentizing (quantile averaging) procedure. *Journal of Mathematical Psychology, 48,* 186–195.

**Key, A., Gustafson, S., Rentmeester, L., Hornsby, B., & Bess, F.** (2017). Speech-processing fatigue in children: Auditory event-related potential and behavioral measures. *Journal of Speech, Language, and Hearing Research, 60,* 2090–2104.

**Kim, J., & Davis, C.** (2003). Hearing foreign voices: Does knowing what is said affect visual-masked-speech detection. *Perception, 32,* 111–120.

**Kim, J., & Davis, C.** (2004). Investigating the audio-visual speech detection advantage. *Speech Communication, 44,* 19–30.

**Krauss, R., Freyberg, R., & Morsella, E.** (2002). Inferring speakers' physical attributes from their voices. *Journal of Experimental Social Psychology, 38,* 618–625.

**Kuhl, P., & Meltzoff, A.** (1982). The bimodal perception of speech in infancy. *Science, 218,* 1138–1141.
**Lachs, L., & Pisoni, D.** (2004). Crossmodal source identification in speech perception. *Ecological Psychology, 16,* 159–187.
**Lalonde, K., & Holt, R.** (2015). Preschoolers benefit from visually salient speech cues. *Journal of Speech, Language, and Hearing Research, 58,* 135–150. https://doi.org/10.1044/2014_JSLHR-H-13-0343
**Lalonde, K., & Holt, R.** (2016). Audiovisual speech perception development at varying levels of perceptual processing. *The Journal of the Acoustical Society of America, 139,* 1713–1723.
**Langner, R., & Eickhoff, S.** (2013). Sustaining attention to simple tasks: A meta-analytic review of the neural mechanisms of vigilant attention. *Psychological Bulletin, 139,* 870–900.
**Laurienti, P., Burdette, J., Maldjian, J., & Wallace, M.** (2006). Enhanced multisensory integration in older adults. *Neurobiology of Aging, 27,* 1155–1163.
**Lewis, F., Reeve, R., Kelly, S., & Johnson, K.** (2017). Evidence of substantial development of inhibitory control and sustained attention between 6 and 8 years of age on an unpredictable go/no-go task. *Journal of Experimental Child Psychology, 157,* 66–80.
**Lewkowicz, D.** (2000). Infants' perception of the audible, visible, and bimodal attributes of multimodal syllables. *Child Development, 71,* 1241–1257.
**Libertus, K., Landa, R., & Haworth, J.** (2017). Development of attention to faces during the first 3 years: Influences of stimulus type. *Frontiers in Psychology, 8,* 1976. https://doi.org/10.3389/fpsyg.2017.01976
**Macken, W., Phelps, F., & Jones, D.** (2009). What causes auditory distraction? *Psychonomic Bulletin & Review, 16,* 139–144.
**Manly, T., Anderson, V., Nimmo-Smith, I., Turner, A., Watson, P., & Robertson, I.** (2001). The differential assessment of children's attention: *The Test of Everyday Attention for Children* (*TEA-Ch*), normative sample and ADHD performance. *Journal of Psychology & Psychiatry, 42,* 1065–1081.
**Marks, L.** (2004). Cross-modal interactions in speeded classification. In G. Calvert, C. Spence, & B. Stein (Eds.), *The handbook of multisensory processes* (pp. 85–105). Cambridge, MA: MIT Press.
**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.
**Mavica, L., & Barenholtz, E.** (2013). Matching voice and face identity from static images. *Journal of Experimental Psychology: Human Perception and Performance, 39,* 307–312.
**McClelland, J., & Elman, J.** (1986). The TRACE model of speech perception. *Cognitive Psychology, 18,* 1–86.
**McConachie, H., & Moore, V.** (1994). Early expressive language of severely visually impaired children. *Developmental Medicine & Child Neurology, 36,* 230–240.
**Miller, J.** (1988). A warning about median reaction time. *Journal of Experimental Psychology: Human Perception & Performance, 14,* 539–543.
**Miller, J., & Ulrich, R.** (2003). Simple reaction time and statistical facilitation: A parallel grains model. *Cognitive Psychology, 46,* 101–151.
**Molholm, S., Ritter, W., Murray, M., Javitt, D., Schroeder, C., & Foxe, J.** (2002). Multisensory auditory–visual interactions during early sensory processing in humans: A high-density electrical mapping study. *Cognitive Brain Research, 14,* 115–128.
**Mordkoff, T., & Yantis, S.** (1993). Dividing attention between color and shape: Evidence of coactivation. *Perception & Psychophysics, 53,* 357–366.
**Munhall, K., Gribble, P., Sacco, L., & Ward, M.** (1996). Temporal constraints on the McGurk effect. *Perception & Psychophysics, 58,* 351–362.
**Napolitano, A., & Sloutsky, V.** (2004). Is a picture worth a thousand words? The flexible nature of modality dominance in young children. *Child Development, 75,* 1850–1870.
**O'Toole, A., Roark, D., & Abdi, H.** (2002). Recognizing moving faces: A psychological and neural synthesis. *Trends in Cognitive Sciences, 6,* 261–266.
**Otsuka, Y., Konishi, Y., Kanazawa, S., Yamaguchi, M., Abdi, H., & O'Toole, A.** (2009). Recognition of moving and static faces by young infants. *Child Development, 80,* 1259–1271.
**Rabbitt, P., & Goward, L.** (1994). Age, information processing speed, and intelligence. *The Quarterly Journal of Experimental Psychology: Human Experimental Psychology, 47*(3), 741–760.
**Rabbitt, P., Osman, P., Moore, B., & Stollery, B.** (2001). There are stable individual differences in performance variability, both from moment to moment and from day to day. *The Quarterly Journal of Experimental Psychology: Human Experimental Psychology, 54*(A), 981–1003.
**Ratcliff, R.** (1979). Group reaction time distributions and an analysis of distribution statistics. *Psychological Bulletin, 86,* 446–461.
**Reinvang, I.** (1998). Validation of reaction time in continuous performance tasks as an index of attention by electrophysiological measures. *Journal of Clinical and Experimental Neuropsychology, 20,* 885–897.
**Reisberg, D., McLean, J., & Goldfield, A.** (1987). Easy to hear but hard to understand: A lip-reading advantage with intact auditory stimuli. In B. Dodd & R. Campbell (Eds.), *Hearing by eye: The psychology of lip-reading* (pp. 97–111). Hillsdale, NJ: Erlbaum.
**Robertson, I., Manly, T., Andrade, J., Baddeley, B., & Yiend, J.** (1997). 'Oops!': Performance correlates of everyday attentional failures in traumatic brain injured & normal subjects. *Neuropsychologia, 35,* 747–758.
**Rosenthal, R., Rosnow, R., & Rubin, D.** (2000). *Contrasts and effect sizes in behavioral research. A correlational approach.* New York, NY: Cambridge University Press.
**Ross, M., & Lerman, J.** (1971). *Word Intelligibility by Picture Identification Test (WIPI)*. Pittsburgh, PA: Stanwix House.
**Schroger, E., & Widmann, A.** (1998). Speeded responses to audiovisual signal changes result from bimodal integration. *Psychophysiology, 35,* 755–759.
**Schwartz, J., Berthommier, F., & Savariaux, C.** (2004). Seeing to hear better: Evidence for early audio-visual interactions in speech identification. *Cognition, 93,* B69–B78.
**Sloutsky, V., & Napolitano, A.** (2003). Is a picture worth a thousand words? Preference for auditory modality in young children. *Child Development, 74,* 822–833.
**Smith, H., Dunn, A., Baguley, T., & Stacey, P.** (2016a). Concordant cues in faces and voices: Testing the back-up signal hypothesis. *Evolutionary Psychology,* 1–10, https://doi.org/10.1177/1474704916630317
**Smith, H., Dunn, A., Baguley, T., & Stacey, P.** (2016b). Matching novel face and voice identity using static and dynamic facial images. *Attention, Perception & Psychophysics, 78,* 868–879.
**Smith, L., & Thelen, E.** (2003). Development as a dynamic system. *Trends in Cognitive Sciences, 7,* 343–348.
**Stevenson, R., Ghose, D., Fister, J., Sarko, D., Altieri, N., Nidiffer, A., . . . Wallace, M.** (2014). Identifying and quantifying multisensory integration: A tutorial review. *Brain Topography, 27,* 707–730.
**Stevenson, R., Sheffield, S., Butera, I., Gifford, R., & Wallace, M.** (2017). Multisensory integration in cochlear implant recipients. *Ear and Hearing, 38,* 521–538.

**Teinonen, T., Aslin, R., Alku, P., & Csibra, G.** (2008). Visual speech contributes to phonetic learning in 6-month-old infants. *Cognition, 108,* 850–855.

**Thillay, A., Roux, S., Gissot, V., Carteau-Martin, I., Knight, R., Bonnet-Brilhault, F., & Bidet-Caulet, A.** (2015). Sustained attention and prediction: Distinct brain maturation trajectories during adolescence. *Frontiers in Human Neuroscience, 9,* 519. https://doi.org/10.3389/fnhum.2015.00519

**Tjan, B., Chao, E., & Bernstein, L.** (2013). A visual or tactile signal makes auditory speech detection more efficient by reducing uncertainty. *European Journal of Neuroscience, 39,* 1323–1331.

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

**Tse, C., Balota, D., Yap, M., Duchek, J., & McCabe, D.** (2010). Effects of healthy aging and early stage dementia of the Alzheimer's type on components of response time distributions in three attention tasks. *Neuropsychology, 24,* 300–315.

**Tye-Murray, N., Spehar, B., Myerson, J., Sommers, M., & Hale, S.** (2011). Cross-modal enhancement of speech detection in young and older adults: Does signal content matter? *Ear and Hearing, 32,* 650–655.

**van Wassenhove, V., Grant, K., & Poeppel, D.** (2005). Visual speech speeds up the neural processing of auditory speech. *Proceedings of the National Academy of Sciences of the United States of America, 102,* 1181–1186.

**van Wassenhove, V., Grant, K., & Poeppel, D.** (2007). Temporal window of integration in auditory–visual speech perception. *Neuropsychologia, 45,* 598–607.

**Venker, C., Goodwin, J., Roe, D., Kaemingk, K., Mulvaney, S., & Quan, S.** (2007). Normative psychomotor vigilance task performance in children ages 6 to 11—*The Tucson Children's Assessment of Sleep Apnea* (*TuCASA*). *Sleep & Breathing, 11,* 217–224.

**Vickers, J.** (2007). *Perception, cognition, and decision training: The quiet eye in action* (pp. 47–64). Champaign, IL: Human Kinetics.

**Watkins, S., Dalton, P., Lavie, N., & Rees, G.** (2007). Brain mechanisms mediating auditory attentional capture in humans. *Cerebral Cortex, 17,* 1694–1700.

**Weissman, D., Roberts, K., Visscher, K., & Woldorff, M.** (2006). The neural bases of momentary lapses in attention. *Nature Neuroscience, 9,* 971–978.

**Whelan, R.** (2008). Effective analysis of reaction time data. *The Psychological Record, 58,* 475–482.

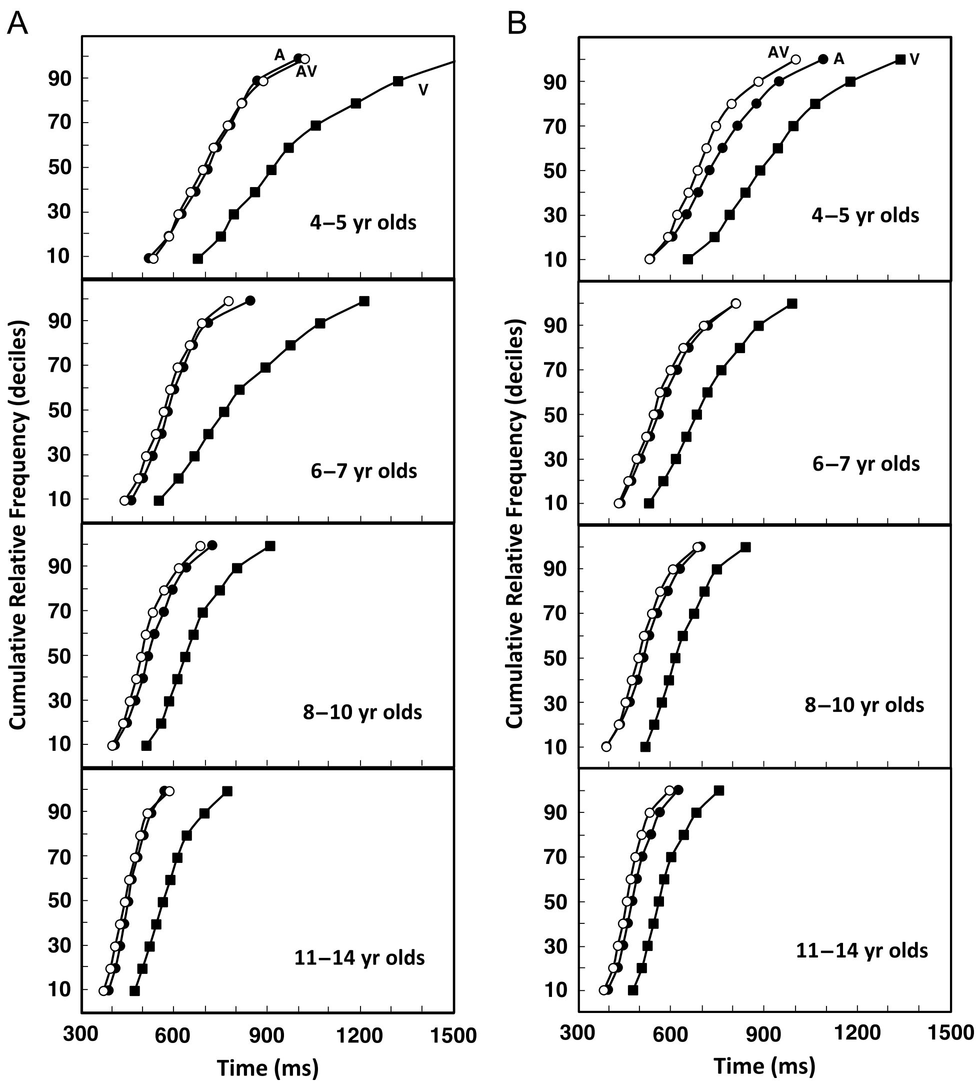
**Wickens, C.** (1974). Temporal limits of human information processing: A developmental study. *Psychological Bulletin, 81,* 739–755.

**Woodworth, R., & Schlosberg, H.** (1954). *Experimental psychology*. New York, NY: Henry Holt and Company.

**Zhou, B., & Krott, A.** (2016). Bilingualism enhances attentional control in non-verbal conflict tasks—Evidence from ex-Gaussian analyses. *Bilingualism: Language and Cognition, 21,* 162–180. https://doi.org/10.1017/S1366728916000869

## Appendix A

The Cumulative Distribution Functions for the Auditory, Audiovisual, and Visual Modes in the Static (Panel A) and Dynamic (Panel B) Facial Conditions for All Age Groups



## Appendix B

$F_{contrast}$ Analyses to Determine Whether the Visual (V) vs. Auditory (A) Response Times Differ at Each Quartile for Each Facial Condition in the Age Groups

| Quartile and facial condition | Mode | | | | |
|---|---|---|---|---|---|
| | V | A | $F_{contrast}$ | $p$ | Partial $\eta^2$ |
| 4–5 years | | | | | |
| First (fast) quartile | | | | | |
| Static face | 726 | 561 | 354.23 | < .0001 | .761 |
| Dynamic face | 711 | 580 | 228.82 | < .0001 | .673 |
| Third (slow) quartile | | | | | |
| Static face | 1046 | 765 | 507.09 | < .0001 | .820 |
| Dynamic face | 983 | 800 | 233.88 | < .0001 | .678 |
| 6–7 years | | | | | |
| First (fast) quartile | | | | | |
| Static face | 594 | 485 | 228.82 | < .0001 | .673 |
| Dynamic face | 562 | 460 | 208.76 | < .0001 | .653 |
| Third (slow) quartile | | | | | |
| Static face | 871 | 617 | 537.67 | < .0001 | .829 |
| Dynamic face | 753 | 608 | 254.91 | < .0001 | .697 |
| 8–10 years | | | | | |
| First (fast) quartile | | | | | |
| Static face | 541 | 434 | 276.72 | < .0001 | .714 |
| Dynamic face | 537 | 421 | 341.76 | < .0001 | .755 |
| Third (slow) quartile | | | | | |
| Static face | 687 | 554 | 244.20 | < .0001 | .688 |
| Dynamic face | 664 | 548 | 204.08 | < .0001 | .648 |
| 11–14 years | | | | | |
| First (fast) quartile | | | | | |
| Static face | 489 | 401 | 218.69 | < .0001 | .663 |
| Dynamic face | 496 | 415 | 199.22 | < .0001 | .642 |
| Third (slow) quartile | | | | | |
| Static face | 602 | 474 | 305.35 | < .0001 | .733 |
| Dynamic face | 596 | 502 | 180.72 | < .0001 | .619 |

*Note.* Results were based on a mixed-design analysis of variance with one between-participant factor (age group: 4–5, 6–7, 8–10, and 11–14 years) and three within-participant factors (mode: V vs. A; facial condition: static vs. dynamic; quartile: first vs. third). Although mean response times are presented to ease understanding, the dependent variable for analyses was the log-transformed response times. For all values of $F_{contrast}$, mean square error = 0.0005 and $df$ = 1,111.

## Appendix C

$F_{contrast}$ Analyses to Determine Whether the Visual–Auditory Difference Scores for the Static (Stat) vs. Dynamic (Dynam) Facial Conditions Differ at Each Quartile in the Age Groups

| Quartile | Facial condition | | $F_{contrast}$ | $p$ | Partial $\eta^2$ |
|---|---|---|---|---|---|
| | Stat | Dynam | | | |
| | | 4–5 years | | | |
| First (fast) quartile | 165 | 131 | 6.04 | *ns* | .052 |
| Third (slow) quartile | 281 | 183 | 26.39 | < .0001 | .192 |
| | | 6–7 years | | | |
| First (fast) quartile | 109 | 102 | 0.10 | *ns* | .001 |
| Third (slow) quartile | 254 | 145 | 25.22 | < .0001 | .185 |
| | | 8–10 years | | | |
| First (fast) quartile | 107 | 116 | 1.36 | *ns* | .012 |
| Third (slow) quartile | 133 | 116 | 0.88 | *ns* | .008 |
| | | 11–14 years | | | |
| First (fast) quartile | 88 | 81 | 0.29 | *ns* | .003 |
| Third (slow) quartile | 128 | 94 | 199.22 | .006 | .066 |

*Note.* Results were based on a mixed-design analysis of variance with one between-participant factor (age group: 4–5, 6–7, 8–10, and 11–14 years) and two within-participant factors (facial condition: static vs. dynamic; quartile: first vs. third). Although mean difference scores (visual–auditory) are presented to ease understanding, the dependent variable for analyses was always the log-transformed difference scores. For all values of $F_{contrast}$, mean square error = 0.0010 and *df* = 1,111. *ns* = not significant.

## Appendix D

$F_{contrast}$ Analyses to Determine Whether the Audiovisual (AV) vs. Auditory (A) Response Times Differ at Each Quartile for Each Facial Condition in Age Groups

| Quartile and facial condition | Mode | | $F_{contrast}$ | $p$ | Partial $\eta^2$ |
|---|---|---|---|---|---|
| | AV | A | | | |
| | | 4–5 years | | | |
| First (fast) quartile | | | | | |
| Static face | 566 | 561 | 0.20 | *ns* | .002 |
| Dynamic face | 568 | 580 | 1.01 | *ns* | .009 |
| Third (slow) quartile | | | | | |
| Static face | 758 | 765 | 0.40 | *ns* | .004 |
| Dynamic face | 737 | 800 | 26.46 | < .0001 | .192 |
| | | 6–7 years | | | |
| First (fast) quartile | | | | | |
| Static face | 468 | 485 | 7.20 | *ns* | .061 |
| Dynamic face | 452 | 460 | 1.41 | *ns* | .012 |
| Third (slow) quartile | | | | | |
| Static face | 605 | 617 | 3.63 | *ns* | .032 |
| Dynamic face | 589 | 608 | 4.24 | *ns* | .037 |
| | | 8–10 years | | | |
| First (fast) quartile | | | | | |
| Static face | 423 | 434 | 3.63 | *ns* | .032 |
| Dynamic face | 418 | 421 | 2.83 | *ns* | .025 |
| Third (slow) quartile | | | | | |
| Static face | 525 | 554 | 15.55 | .0001 | .123 |
| Dynamic face | 531 | 548 | 4.24 | *ns* | .095 |
| | | 11–14 years | | | |
| First (fast) quartile | | | | | |
| Static face | 386 | 401 | 6.66 | *ns* | .057 |
| Dynamic face | 402 | 415 | 4.24 | *ns* | .037 |
| Third (slow) quartile | | | | | |
| Static face | 466 | 474 | 0.40 | *ns* | .004 |
| Dynamic face | 480 | 502 | 10.51 | .002 | .086 |

*Note.* Results were based on a mixed-design analysis of variance with one between-participant factor (age group: 4–5, 6–7, 8–10, and 11–14 years) and three within-participant factors (mode: AV vs. A; facial condition: static vs. dynamic; quartile: first vs. third). Although mean response times are presented to ease understanding, the dependent variable for analyses was the log-transformed response times. For all values of $F_{contrast}$, mean square error = 0.0005 and *df* = 1,111. *ns* = not significant.

## Appendix E

$F_{contrast}$ Analyses to Determine Whether the Audiovisual–Auditory Difference Scores for the Static (Stat) vs. Dynamic (Dynam) Facial Conditions Differ at Each Quartile in the Age Groups

| | Facial condition | | | | |
|---|---|---|---|---|---|
| Quartile | Stat | Dynam | $F_{contrast}$ | *p* | Partial $\eta^2$ |
| | | 4–5 years | | | |
| First (fast) quartile | 5 | –12 | 0.92 | *ns* | .008 |
| Third (slow) quartile | –7 | –63 | 9.76 | .002 | .081 |
| | | 6–7 years | | | |
| First (fast) quartile | –17 | –8 | 0.91 | *ns* | .008 |
| Third (slow) quartile | –12 | –19 | 0.01 | *ns* | .000 |
| | | 8–10 years | | | |
| First (fast) quartile | –11 | –3 | 0.91 | *ns* | .008 |
| Third (slow) quartile | –28 | –17 | 1.72 | *ns* | .015 |
| | | 11–14 years | | | |
| First (fast) quartile | –15 | –13 | 0.40 | *ns* | .004 |
| Third (slow) quartile | –8 | –22 | 2.83 | *ns* | .025 |

*Note.* Results were based on a mixed-design analysis of variance with one between-participant factor (age group: 4–5, 6–7, 8–10, and 11–14 years) and two within-participant factors (facial condition: static vs. dynamic; quartile: first vs. third). Although mean difference scores (AV–A) are presented to ease understanding, the dependent variable for analyses was always the log-transformed difference scores. For all values of $F_{contrast}$, mean square error = 0.0010 and $df$ = 1,111. *ns* = not significant.