# Effects of Looking Behavior on Listening and Understanding in a Simulated Classroom

**Dawna E. Lewis, PhD**
Boys Town National Research Hospital, Omaha, NE

**Shannon Wannagot, BA**
Boys Town National Research Hospital, Omaha, NE
University of Connecticut, Storrs, CT

**Audiovisual cues can improve speech perception in adverse acoustical environments when compared to auditory cues alone. In classrooms, where acoustics often are less than ideal, the availability of visual cues has the potential to benefit children during learning activities. The current study evaluated the effects of looking behavior on speech understanding of children (8-11 years) and adults during comprehension and sentence repetition tasks in a simulated classroom environment. For the comprehension task, results revealed an effect of looking behavior (looking required versus looking not required) for older children and adults only. Within the looking-behavior conditions, age effects also were evident. There was no effect of looking behavior for the sentence-repetition task (looking versus no looking) but an age effect also was found. The current findings suggest that looking behavior may impact speech understanding differently depending on the task and the age of the listener. In classrooms, these potential differences should be taken into account when designing learning tasks.**

## Introduction

Spoken speech is a complex signal containing numerous auditory cues that can be used to aid in understanding (Pisoni & Remez, 2008). In natural environments visual cues also play a role in that understanding. Research has shown that the availability of audiovisual cues improves speech-perception performance in conditions where the auditory signal is degraded for both adults and children over that seen with auditory or visual input alone (Erber, 1969; Massaro & Cohen, 1995; Ross, Saint-Amour, Leavitt, Javitt, & Foxe, 2007; Sumby & Pollack, 1954; Wightman, Kistler, & Brungart, 2006). The addition of this information also has been shown to positively impact speech understanding when degraded acoustics are not an issue (Arnold & Hill, 1976; McGurk & MacDonald, 1976).

Development plays a role in children's ability to understand speech that is presented auditory-only in noise and reverberation and in the ability to integrate auditory and visual input. Separately and together, noise and reverberation have a greater effect on children than on adults (Crandell & Smaldino, 2000; Neuman & Hochberg, 1983; Neuman, Wroblewski, Hajicek, & Rubenstein, 2010; Wroblewski, Lewis, Valente, & Stelmachowicz, 2012; Yang & Bradley, 2009), with some skills improving into adolescence (Johnson, 2000). Although infants have been shown to use visual information in the perception of speech (Desjardins & Werker, 2004; Kuhl & Meltzoff, 1982; 1984), this early skill is rudimentary and continues to develop through childhood and into adolescence (Desjardins, Rogers, & Werker, 1997; Massaro, 1984; Massaro, Thompson, Barron & Laren, 1986; McGurk & McDonald, 1976; Ross, Molholm, Blanco, Gomez-Ramirez, Saint-Amour & Fox, 2011; Wightman, et al., 2006; see also Soto-Faraco, Calabresi, Navarra, Werker, & Lewkowicz, 2012, for a review).

In classrooms, elementary-age children often encounter unfavorable listening conditions, where poor signal-to-noise ratios (SNR) and extended reverberation times (RT) may affect hearing, understanding and overall academic success (Dockrell & Shield, 2006; Jamieson, Kranjc, Yu, & Hodgetts, 2004; Klatte, Hellbruck, Seidel, & Leistner, 2010; Klatte, Lachman, & Meis, 2010). ANSI S12.60-2002 provides standards for SNR and RT in schools (American National Standards Institute [ANSI], 2002). However, numerous studies have shown that typical classrooms often exceed the recommended SNR and RT (see Picard & Bradley, 2001, for a review; Nelson, Smaldino, Erler, & Garstecki, 2008). Under such conditions, children may be less able to process the compromised auditory signal for understanding. This may be especially true for younger children whose abilities to understand speech in noise and reverberation are still developing. As a result, the benefit of combined auditory and visual input has the potential to improve speech perception in this population. However, multiple factors can influence that potential benefit.

In classroom environments many listening tasks may require attention to multiple talkers and not all talkers will be easily visualized. Variability among talkers has the potential to negatively impact speech perception in adults and children (Mullennix, Pisoni, & Martin, 1989; Ryalls & Pisoni, 1997), and recent research has suggested that, at least for adults, this may occur for both auditory-only and auditory-visual presentations of speech (Heald & Nusbaum, 2014). To benefit from combined auditory and visual input under such conditions, the listener must be able to locate the appropriate talker while that person is speaking. It also is possible that some actual and potential talkers could distract the listener from the speech signal of interest. In those instances, both auditory and visual attention could serve as distracters (Ricketts & Galster, 2008; Valente, Plevinsky, Franco, Heinrichs-Graham, & Lewis, 2012). In these instances, the effort required to locate and understand multiple talkers may expend cognitive resources that otherwise would be used for comprehension, with potentially negative consequences.

In view of the potential for both positive and negative effects of looking behavior when attempting to understand speech, it can be helpful to examine that behavior during speech perception tasks. Potential effects of looking behavior during complex listening tasks were examined as part of recent studies in our laboratory investigating speech comprehension and sentence repetition in a simulated classroom environment (Valente et al., 2012). As part of those studies, participants were asked to attend to audiovisual recordings of a teacher and four students who were reading lines from a play which were presented from loudspeakers and LCD monitors located around the listener. Participants were told that they could look as much or as little as they needed to understand the play. At the end of the play, they answered factual questions about its content. Looking behavior during the comprehension task was monitored via a head-worn gyroscope that recorded head movements in the horizontal plane. The results of the gyroscopic recordings were analyzed in two ways. The first measurement (proportion of events visualized, POEV) represented the proportion of time a listener looked directly at each talker when he/she was speaking. The second measurement (overall looking behavior) represented general looking across all talkers during the task and provided an indication of participants' attempts to look at talkers. The same participants repeated auditory-only sentences that contained three keywords each and were presented from the same locations used for the comprehension task.

As part of the first experiment reported in Valente et al. (2012), adults and children (8-12 years) with normal hearing performed the task in an acoustic environment with an SNR of 10 dB and an RT of 0.6 seconds. This acoustic environment is comparable to that found in many classrooms (Bradley & Sato, 2008). Although results revealed ceiling or near-ceiling performance for both children and adults on the sentence repetition task, children exhibited lower scores than adults on the comprehension task. Thus, under typical classroom acoustic conditions, children demonstrated a high level of sentence-recognition ability that was similar to that of adults. However, under those same conditions their performance on a more complex listening task was poorer.

Analysis of looking behavior during the comprehension task revealed that both adults and children looked directly at the talkers as they were speaking less than 50% of the time. This finding was not surprising given that the task required participants to follow multiple talkers who changed often. As a result, even if participants attempted to look at all talkers, they may not have been able to visualize them as they were speaking. However, children did localize the talkers significantly more often than adults.

For the measure of overall looking behavior, children exhibited greater overall looking than adults and participants who exhibited higher looking behavior showed poorer performance on the task. These finding suggested that, although children were more likely to attempt to look at the talkers during the comprehension task, their understanding of the material may not have benefited from looking.

During listening tasks in which it may be difficult to visualize talkers as they speak, attempts to access visual information may not always be beneficial. In such activities attempts to locate talkers may use cognitive resources that are needed for comprehension. When interpreting the relationship between looking behavior and comprehension in Valente et al. (2012), it is important to remember that participants were instructed to look as much or as little as they felt necessary during the comprehension task. If individuals' strategies for listening and looking were chosen to optimize understanding, it is possible that some participants made better choices for that particular task than others. Adult-child differences in both performance and looking behavior could indicate that children were less adept at choosing the appropriate looking strategies during this particular task.

Just as there is development in the ability to benefit from audiovisual input, there also may be development in the ability to decide how and when to use that information appropriately for a given task. As a first step in examining this possibility, the current study evaluated the effects of required looking behavior on speech understanding of children and adults with normal hearing (NH) during the comprehension and sentence-repetition tasks used in Valente et al. (2012) and described above. During the comprehension task, participants were instructed to look at each talker as he/she spoke. Results were compared to those of listeners with NH from Valente et al. (2012) and a subsequent study (Lewis, Valente & Spalding, 2014) using the same task during which participants also had not been required to look at the talkers. In those studies, looking behavior had not been examined during the sentence-repetition tasks. Thus, for the current study participants were instructed to locate the loudspeaker for half of the sentence presentations and to look straight ahead for the other half allowing a within-subject examination of looking behavior to be completed for the sentence-repetition task.

## Method

### Participants

Forty children (8-11 years) and 10 adults (19-29 years) with audiometric thresholds within the normal range of hearing (NH; ≤20 dB HL for octave frequencies from 250 to 8000 Hz) participated. No child scored greater than 1 SD below the mean for receptive vocabulary as measured by the Peabody Picture Vocabulary Test (PPVT-IV, Dunn & Dunn, 2007). All children were typically developing by report and were native speakers of English.

Results from 39 children (8-11 years old) and 20 adults with NH from previous studies in our lab (Valente et al., 2012; Lewis et al., 2014) were included in the analysis of the comprehension task to compare performance between listeners who were required to look (current study) and listeners who were instructed that they could look as much or as little as they needed to understand the talkers (previous studies). The children from the previous studies had the same demographic characteristics the currently evaluated group. Table 1 lists the number of participants in each age group.

**Table 1.** Number of participants in each age group from current and previous studies.

| Age Group | 8 years | 9 years | 10 years | 11 years | Adult |
|---|---|---|---|---|---|
| N (previous studies) | 10 | 9 | 9 | 11 | 20 |
| N (current study) | 10 | 10 | 10 | 10 | 10 |

This study was approved by the Institutional Review Board (IRB) for Boys Town National Research Hospital and assent/consent was obtained for all children. Children were paid $15 per hour for their participation and received a book at the completion of the study.

**Simulated acoustic environment**

The simulated acoustic environment was the same as reported in Valente et al. (2012). Readers are referred to that paper for a detailed description of the creation and validation of the environment and it will be described briefly here. The simulated environment was comprised of a physical room and a virtually-modeled room. The physical room was acoustically treated and loudspeakers and LCD monitors were placed on small tables around a participant's location (Figure 1). In the physical room ambient acoustics at the participant's location (RT = 0.35 sec and background noise = 35 dBA) were below those selected for the simulated environment. Real-time simulation techniques (Braasch, Peters & Valente, 2008) were used to create a virtually-modeled room (Virtual Microphone Control; ViMiC) with the same dimensions as the physical room and with virtual microphones and sound sources positioned at the same locations as the loudspeakers and monitors in the physical room. The simulated room (which combined the physical and virtual rooms) included direct sound and both first-order and late reflections. As a result, room acoustics could be set to represent those found in a classroom. Background noise within the simulated room had a spectrum similar to that of heating, ventilation, and air-conditioning (HVAC) systems.

**Figure 1.** Simulated classroom set-up. Used with permission from Lewis, D., Valente, D.L., & Spalding, J. (2015). Effect of minimal/mild hearing loss on children's speech understanding in a simulated classroom. *Ear & Hearing, 36*, 136-144.

For both tasks, the level of speech at the participant's location was 60 dBA and RT was set to 0.6 s. For the comprehension task, noise was adjusted for a 10 dB SNR. Due to the fact that sentence repetition at a 10 dB SNR resulted in ceiling or near ceiling performance in the Valente et al. (2012) study, a SNR of 0 dB was chosen for the sentence repetition task in the current study to reduce potential ceiling effects that could obscure differences across looking conditions.

**Procedures**

Comprehension. The video recordings from Valente et al. (2012) were used. In those recordings, a teacher and four students read lines from an unfamiliar 10-minute elementary-age-appropriate Reader's Theater play (Shepard 2010). For the task, each talker was located at one of the five loudspeaker/ monitor arrangements around the participant. Each talker acted as a different character and there was no overlapping speech across talkers. Participants were instructed to look at each of the talkers whenever they spoke. At the end of the presentation, participants were asked 18 factual questions (e.g., Where did the troll go each day while Leif worked? What did Leif think was going to happen when the troll was chasing him and Master Maid?). Listeners responded orally and their answers were written down by the experimenter for later percent-correct scoring.

Looking Behavior. To monitor looking behavior during the comprehension task, each child wore a custom-designed gyroscope (Analog Devices, EVAL-ADXRS610) attached to a headband. The gyroscope tracked head movement in the horizontal plane. Output from the gyroscope was converted into a digital signal for analysis (Teabox; Electrotap LLC).

Participants' looking behavior was analyzed in two ways using the procedures from Valente et al. (2012). *Proportion of events visualized* (POEV) signified the proportion of time a participant looked directly at each talker (+/- $15^{o}$) when he/she was speaking. For this measure, the location of each of the simulated talkers relative to the participant location was compared to the gyroscopic data that were obtained for that participant. *Overall looking behavior* was measured using the head-angle recordings of the participant. The standard deviation (SD) of the head tracking measurement represented the degree of head movement relative to 0o azimuth. Following the convention developed for Valente et al. (2012), overall looking behavior was classified into three categories: low (SD < $20^{o}$), medium ($20^{o}$ < SD < $45^{o}$) and high (SD > $45^{o}$).

Sentence repetition. Participants heard and repeated 50 auditory-only sentences with three key words (Bamford-Kowal-Bench sentences [BKB]; Bench, Kowal, & Bamford, 1979) spoken by a female talker from each of the five loudspeakers. The sentences were digitally recorded in a sound booth using a condenser microphone (AKG Acoustics C535 EB) with a flat frequency response (±2 dB from 0.2 to 20 kHz). Sentences were presented one at a time. The screen of each monitor was lighted when a sentence was being presented from that location. Although the lighted screens did not provide content information, spatial visual cues have been shown to improve speech intelligibility in adverse listening environments (Best, Ozmeral, & Shinn-Cunningham, 2007). Thus, this option offered a visual cue to assist participants in determining talker location. Two looking conditions were utilized during the speech-perception task. In this within-subject task these two conditions were different from those described for the comprehension task. In the *looking* condition, participants were instructed to localize and visualize the lighted screen before repeating the sentence. In the *no-looking* condition, participants were instructed to look straight ahead at the front screen (i.e., they were instructed not to look around). Each participant completed the sentence repetition task in both conditions, listening to half of the sentences in one condition and half in the other. The starting condition for the task was alternated across listeners in each age group. Responses were scored for correct repetition of each of the three target words in each sentence.

## Results

### Comprehension Task

Individual results for performance on the comprehension task and overall looking behavior are shown in Figure 2 for the previous studies (left panel) and the current study (right panel). When looking was not required, the range of scores for 8-10 year olds was more widely distributed than that of the 11-year olds and adults. However, the older children and adults for whom looking was required showed a distribution of scores that was more similar to that of the younger children in either looking-behavior condition.

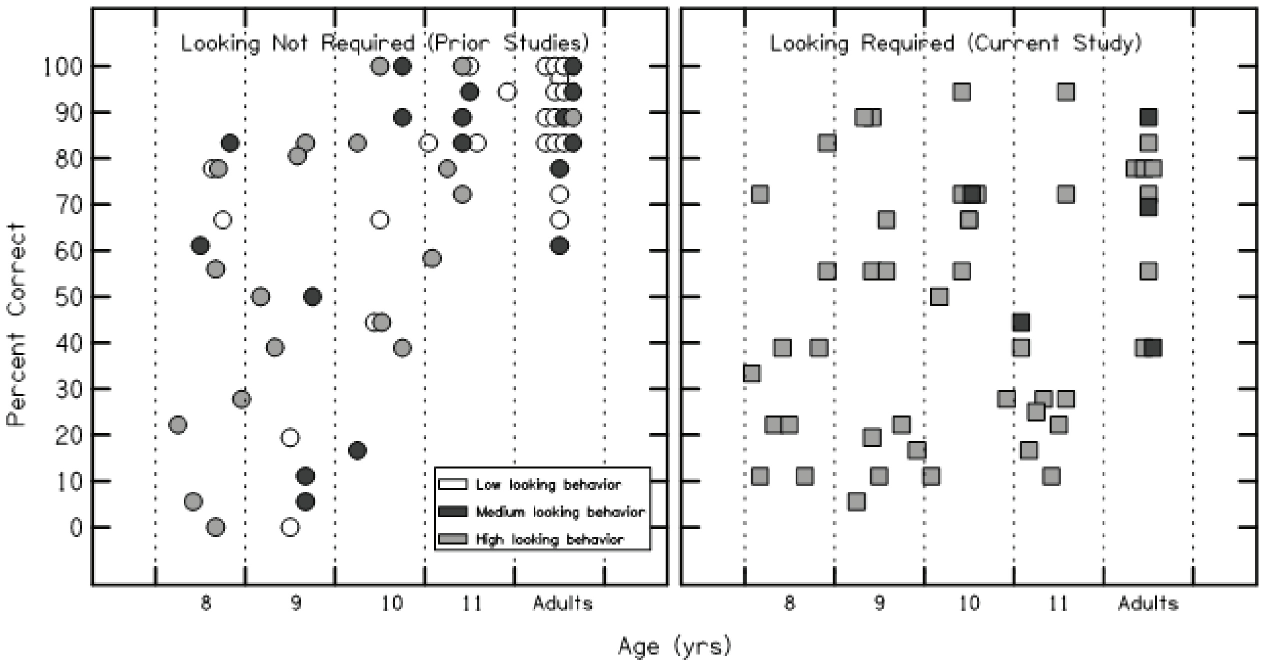


**Figure 2.** Individual comprehension scores and overall looking behavior for the classroom listening task when participants were not required (left) and were required (right) to look at the talkers. Looking behavior is represented as low (white), medium (black) or high (gray). For the children (8-11 yrs), age within each year is indicated by the horizontal position of the symbols. For the adults, symbols are randomly jittered on the horizontal axis for visibility only.

Means and standard deviations for comprehension scores across looking condition and age group are shown in Table 2. A two-way analysis of variance (ANOVA) was conducted with age group and looking condition as between-subject factors. Results revealed significant main effects on comprehension of looking condition ($F_{(1,99)}$ = 10.409; $p$ =.002, $\eta_p^2$ =0 .095) and age group ($F_{(4,99)}$ = 9.197; $p$<.0001, $\eta_p^2$ = 0.271) as well as a looking condition by age group interaction ($F_{(4,99)}$ = 3.503; $p$ =.010, $\eta_p^2$ = 0.124). Overall, performance was higher when looking was not required. Minimal mean differences (Fisher's LSD) revealed that across looking conditions adults and 11-year olds for whom looking was not required performed better than those for whom looking was required. Across age groups, adults and 11-year olds performed better than 8-, 9-, and 10-year olds and 10-year olds performed better than 8- and 9-year olds when looking was not required. When looking was required, adults performed better than 8-, 9-, and 11-year olds and 10-year olds performed better than 11-year olds.

**Table 2.** Means and standard deviations (in parentheses) for comprehension scores across looking condition and age group.

| | **8 years** | **9 years** | **10 years** | **11 years** | **Adult** |
|---|---|---|---|---|---|
| **Looking Not Required** | 47.8 (31.2) | 37.7 (31.1) | 64.8 (30.1) | 85.1 (12.6) | 87.6 (11.3) |
| **Looking Required** | 38.9 (24.7) | 43.1 (31.9) | 58.9 (24.2) | 38.1 (26.2) | 68.1 (17.7) |

In Figure 2 overall looking behavior for individual participants also is categorized: low (white), medium (black) or high (gray). As expected, in the current study most participants (all except three adults, one 10 year-old, and one 11 year-old) demonstrated high looking behavior (SD >45°). Those five participants displayed medium looking behavior (20°<SD<45°). These findings indicate that the majority of participants in the current study followed the directions to attempt to look at the talkers as they were speaking. In the previous studies when participants were allowed to look as much or as little as they chose, overall looking behavior was variable. The majority of adults exhibited low looking behavior, with only one showing high overall looking. In contrast, the majority of the youngest children exhibited high looking behavior, with only two showing low looking. By 11 years of age there were almost equal numbers of children showing high and low looking behavior.

Despite the pattern of high overall looking behaviors in the required-looking condition, the mean proportion of time participants were looking at talkers while they were speaking (POEV) was still less than 0.60 for all ages. POEV as function of age group and looking condition is shown in Figure 3 and means and standard deviations are shown in Table 3.

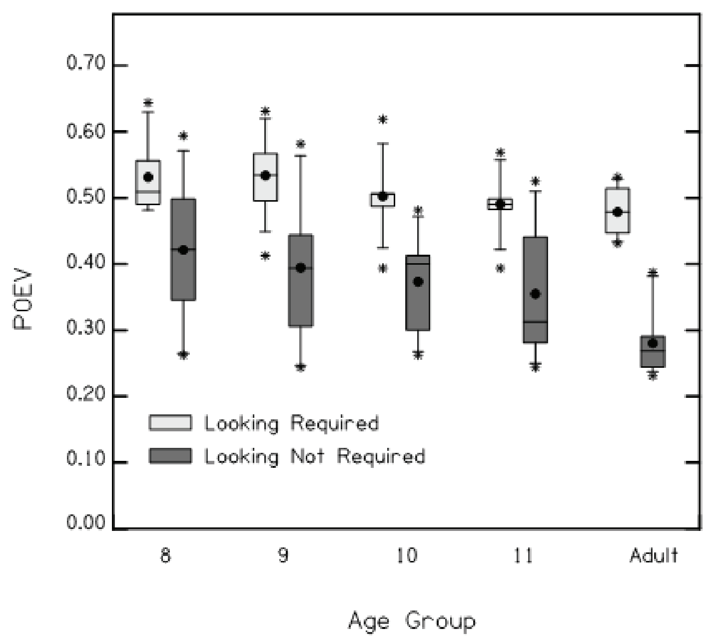


**Figure 3.** Proportion of events visualized (POEV) during the comprehension task for listeners who were required (light gray) and who were not required to look (dark gray) at the talkers. Boxes represent the interquartile range and whiskers represent the 5th and 95th percentiles. For each box, lines represent the median and filled circles represent the mean scores. Asterisks represent scores that fell outside the 5th-95th percentiles.

A two-way ANOVA revealed significant main effects of looking condition ($F(1,99) = 94.436$; $p < .0001$, $\eta p2 = 0.488$) and age group ($F(4,99) = 6.028$; $p < .0001$, $\eta p2 =0 .0263$) but no looking condition by age group interaction ($F(4,99) = 1.209$; $p =.312$, $\eta p2 =0 .047$). Participants who were required to look exhibited higher POEV than those who were not required to look. In addition, POEV was lower for adults than for any of the groups of children. No other age-group differences were significant.

**Table 3.** Means and standard deviations (in parentheses) for Proportion of events visualized (POEV) across looking condition and age group.

| | 8 years | 9 years | 10 years | 11 years | Adult |
|---|---|---|---|---|---|
| **Looking Not Required** | 0.421 (0.024) | 0.394 (0.025) | 0.374 (0.025) | 0.355 (0.023) | 0.280 (0.017) |
| **Looking Required** | 0.531 (0.024) | 0.534 (0.024) | 0.502 (0.024) | 0.491 (0.024) | 0.479 (0.024) |

**Sentence Repetition Task**

Recall that a within-subject design was used in the current study to examine the relation between sentence repetition and looking behavior. Percent-correct sentence repetition as a function of age group and looking condition is shown in Figure 4 with means and SDs shown in Table 4.

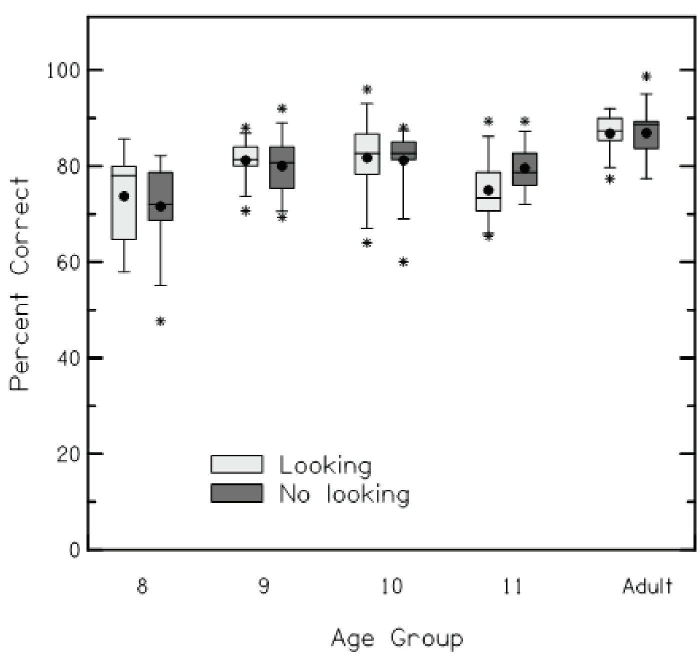


**Figure 4.** Sentence-repetition scores for the looking (light gray) and no looking (dark gray) conditions. Boxes represent the interquartile range and whiskers represent the $5^{th}$ and $95^{th}$ percentiles. For each box, lines represent the median and filled circles represent the mean scores. Asterisks represent scores that fell outside the $5^{th}$-$95^{th}$ percentiles.

A mixed-model ANOVA was conducted with looking condition as a within-subject factor and age group as a between-subject factor. Results revealed a significant main effect of age group ($F(4,44) = 5.815$; $p =.001$, $\eta p2 = 0.329$). There was no significant effect of looking condition ($F(4,44) =.028$; $p = 0.867$, $\eta p2 = 0.001$) and no looking condition by age group interaction ($F(4,44) =1.246$; $p = 0.305$, $\eta p2 = 0.102$). Adults performed better than 8- and 11-year olds. No other effects of age group were noted. (Note: One 11-year-old was not included in statistical analyses due to inattention during the no-looking portion of this task).

**Table 4.** Means and standard deviations (in parentheses) for sentence repetition scores across looking condition and age group.

| | 8 years | 9 years | 10 years | 11 years | Adult |
|---|---|---|---|---|---|
| **No Looking** | 71.6 (10.5) | 80.0 (6.8) | 81.2 (7.9) | 79.6 (5.8) | 86.9 (6.5) |
| **Looking** | 73.7 (10.75) | 81.2 (4.9) | 81.73 (8.0) | 75.0 (7.7) | 86.8 (4.5) |

## Discussion

The current study examined the effect of looking behavior on children's speech understanding in a simulated classroom environment. During the comprehension task, participants in the required-looking condition located and looked at talkers as they were speaking less than 60% of the time. This occurred despite the fact that overall looking behavior was high for all but five participants, indicating that the majority were following instructions and making an attempt to locate the talkers. Similar to the previous studies (Valente et al., 2012; Lewis et al., 2014), it is likely that the discrepancy between overall looking behavior and POEV was impacted, at least in part, by the rapid transitions among talkers and their locations relative to the participant. In addition, it is important to note that a head-mounted gyroscope was used to monitor head movements. This method would not pick up changes in looking behavior that did not require head movements (i.e., those for which eye movements alone would be sufficient to allow visual attention to the talker or where there was a combination of head and eye movement). As a result, POEV may under-estimate visual attention to the talker to some degree. Other methods that could provide information regarding visual attention based on eye-movements (e.g., eye-tracking) were not possible in the current experimental paradigm due to the range of locations for the five talkers.

When looking was required for the comprehension task, adults and 11-year olds demonstrated significantly poorer performance than those that were not required to look. Based on our previous studies (Valente et al., 2012; Lewis et al., 2014), we hypothesized that the cognitive resources required to perform the listening task may impact children's ability to process the speech signal under adverse conditions. This hypothesis also is supported by research with adults and children showing that factors that increase listening effort (defined as attention and cognitive effort required for speech understanding) can negatively impact speech understanding (Fraser, Gagne, Alepins & Dubois, 2010; Gosselin & Gagne, 2011; Hicks & Tharpe, 2002; Larsby, Hallgren, Lyxell & Arlinger, 2005; Pichora-Fuller, Schneider & Daneman, 1995).

For the comprehension task, the cognitive load/listening effort could be impacted by multiple factors: the task itself (processing new information over a 10-minute period of time, linguistic complexity of the material), the acoustic environment (realistic noise and reverberation), age of the participants, and looking behavior (required versus not-required looking). The factor that differed between our current and previous tasks was that of looking behavior. As noted previously, one could assume that the previous participants chose looking strategies that they believed would optimize their understanding during the task, with some participants making better choices regarding looking than others. Examination of Figure 2 suggests that there was little difference in the range of scores for the younger children (8-10 years) when comparing required and not-required looking. In these age groups, few participants demonstrated low overall looking behavior when allowed to choose. However, 11-year olds and adults who were required to look exhibited a wider range of scores and poorer mean performance compared to those who were not required to look. In addition, very few of the participants in those age groups exhibited high overall looking behavior when given the choice. These findings suggest the possibility that high looking behavior was not the optimal looking option for the older children and adults and this requirement may have consumed cognitive resources that had been used for comprehension in the previous studies. Many of the younger children may still be learning how to maximize looking and listening behaviors and may be less likely to choose a strategy that will optimize their performance.

While differences in performance across looking conditions were significant, it should be noted that there was only one story available for the comprehension task. This limitation precludes examination of within-subject performance under the two conditions, which impacts interpretation of the results. Further studies, which are in preparation, will allow examination of comprehension across a variety of tasks and acoustic environments.

There were no significant within-subject effects of looking condition for the sentence-repetition task. In addition, while there were age effects, there was no consistent developmental trend. These findings suggest that during simple speech perception tasks, attempts to locate the talker may not negatively impact cognitive resources necessary to complete the task. It should be recalled that the sentence-repetition task was presented auditory-only. Although there was a visual cue to help listeners locate the signal source, they did not receive any additional visual cues that would have been available from seeing the talker's face. Thus, it is possible that with an audiovisual presentation there may have been an improvement in scores in the required looking condition. However, the relatively high scores in both the looking and non-looking conditions suggest that any potential improvement under the current acoustic conditions would have been small.

The sentence recognition task, as used in this study, was chosen to because word- and sentence-recognition tasks are commonly used in studies that examine the effects of room acoustics on speech perception (e.g., Crandell & Smaldino, 2000; Neuman et al., 2010; Yang & Bradley, 2009). Although these high-context, short sentences provide information about children's speech understanding, the cognitive effort required to fill in any missing acoustic-phonetic information in individual sentences would be expected to be lower than that of tasks children regularly encounter in classrooms. The comprehension task represented a classroom task in which there are multiple talkers who change rapidly (e.g., a classroom discussion; group interaction). In order to answer comprehension questions about the story, the listener had to be able to attend to all of the talkers over the entirety of the story, place that information in memory and process it for later recall. Such a task would utilize greater cognitive resources than simply repeating individual sentences. Under conditions that have the potential of further impacting the cognitive load of the speech perception task (e.g. poor acoustics, additional simultaneous tasks), the resources that are available for storage and recall may be negatively impacted. Further examination of this issue in greater detail is in progress.

Although, as previously discussed, the addition of visual cues can improve speech understanding in many situations, attempts to look for and at talkers may not always be beneficial for listeners. Although having someone else determine when looking is required may not typically occur in natural environments (decisions are made by the individual), the findings reported here suggest looking behavior may impact comprehension differently depending on the task and the age of the listener. During complex listening tasks with frequent changes in talkers, attempts to locate talkers may not always aid in comprehension, while in less complex listening tasks, looking strategies may have less of an impact on understanding. During a complex listening task, younger children may employ different looking strategies than older children and adults.

Differences in both looking strategies and potential benefits/limitations of such strategies should be taken into account when considering how learning tasks will be presented in classrooms. For example, in a task with multiple talkers, the environment could be arranged so that all talkers are easily visualized with limited looking effort (e.g., placing all children who are reading at the front of the room; arranging desks in a u-shape so that each of the children can more easily see all of the other children). For some tasks, working in small groups may provide better audiovisual access than would be available in large groups. For children with hearing loss, audibility of the speech signal may be poorer than for their peers with normal hearing. For these children, poor acoustics may place an even greater load on cognitive resources and the children may rely more heavily on combined audiovisual input for understanding. Future research examining how audiovisual cues and looking behavior impact this population relative to those with normal hearing can provide importance guidance for educational personnel.

**Acknowledgements**

This work was supported by grants R03 DC009675, T32 DC000013, and P30 DC004662 from the National Instituted of Deafness and Communication Disorders and grant P20 GM109023 from the National Institute of General Medical Sciences, both of the National Institutes of Health. The content of this article is the responsibility and opinions of the authors and does not represent the views of the NIH. The first author serves on the Pediatric Advisory Board of Phonak, but there are no conflicts with the work presented here. No other conflicts of interest are noted by the authors.

## References

American National Standards Institute (2010). *Acoustical Performance Criteria, Design Requirements and Guidelines for Schools, Part 1: Permanent Schools*. ANSI 12.60-2010/Part 1. New York: ANSI.

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

Bench, J., Kowal, A., & Bamford, J. (1979). The BKB (Bamford-Kowal-Bench) sentence lists for partially-hearing children. *British Journal of Audiology*, 13, 108–112.

Best, V., Ozmeral, E. & Shinn-Cunningham, B. (2007). Visually-guided attention enhances target identification in a complex auditory scene. *Journal of the Association for Res in Otolaryngology*. 8, 294-304

Braasch, J., Peters, N. & Valente, D.L. (2008). A loudspeaker-based projection technique for spatial music applications using virtual microphone control. *Computer Music Journal*, 32, 55-71.

Bradley, J. & Sato, H. (2008). The intelligibility of speech in elementary school classrooms. *Journal of the Acoustical Society of America*, 123, 2078-2086.

Crandell, C. & Smaldino, J. (2000). Classroom acoustics for children with normal hearing and with hearing impairment. *Language Speech and Hearing Services in Schools*, 31, 362-370.

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 Psychology*, 45, 187-203.

Dockrell, J. & Shield, B. (2006). Acoustical barriers in classrooms: the impact of noise on performance in the classroom. *British Educational Research Journal*, 32, 509-525.

Dunn, L.M. & Dunn, D.M. (2007). *The Peabody Picture Vocabulary Test*, Fourth Edition, Bloomington, MN: NCS Pearson, Inc.

Erber, N. (1969). Interaction of audition and vision in the recognition of oral speech stimuli. *Journal of Speech and Hearing Research*, 12, 423-425.

Fraser, S., Gagne, J.-P., Alepins, M. & Dubois, P. (2010). Evaluating the effort expended to understand speech in noise using a dual-task paradigm: The effects of providing visual cues. *Journal of Speech, Language, Hearing Research*, 53, 18-33.

Gosselin, P. & Gagne, J.-P. (2010). Use of a dual-task paradigm to measure listening effort. *Canadian Journal of Speech-Language Pathology and Audiology*, 34, 43-51.

Heald, S. & Nusbaum, H. (2014). Talker variability in audio-visual speech perception. *Frontiers in Psychology*, 5, 1-9. Doi: 10.3389/fpsyg.2014.00698.

Hicks, C. & Tharpe, A.M. (2002). Listening effort and fatigue in school-age children with and without hearing loss. *Journal of Speech, Language, Hearing Research*, 45, 573-584.

Jamieson, D., Kranjc, G., Yu, K. & Hodgetts, W. (2004). Speech intelligibility of young school-aged children in the presence of real-life classroom noise. *Journal of the American Academy of Audiology*, 15, 508-517.

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

Klatte, M., Hellbrück, J., Seidel, J. & Leistner, P. (2010). Effects of classroom acoustics on performance and well-being in elementary school children: a field study. *Environment & Behavior*, 42, 659-692.

Klatte, M., Lachmann, T. & Meis, M. (2010). Effects of noise and reverberation on speech perception and listening comprehension of children and adults in a classroom-like setting. *Noise Health: Special Issue on Noise, Memory and Learning 12*, 270-282.

Kuhl, P. & Meltzoff, A. (1982). The bimodal perception of speech in infancy. *Science, 218*, 1138-1141.

Kuhl, P. & Meltzoff, A. (1984). The intermodal representation of speech in infants. *Infant Behavior & Development, 7*, 361-381.

Larsby, B., Hallgren, M., Lyxell, B. & Arlinger, S. (2005). Cognitive performance and perceived effort in speech processing tasks: Effects of different noise backgrounds in normal-hearing and hearing-impaired subjects. *International Journal of Audiology, 44*, 131-143.

Lewis, D., Valente, D.L., & Spalding, J. (2015). Effect of minimal/mild hearing loss on children's speech understanding in a simulated classroom. *Ear & Hearing, 36*, 136-144.

Massaro, D. (1984). Children's perception of visual and auditory speech. *Child Development, 55*, 1777-1788.

Massaro, D. & Cohen, M. (1995). Perceiving talking faces. *Current Directions in Psychological. Science, 4*, 104-109.

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.

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

Mullenix, J., Pisoni, D. & Martin, C. (1989). Some effects of talker variability on spoken word recognition. *Journal of the Acoustical Society of America, 85*, 365-378.

Nelson, E., Smaldino, J., Erler, S., & Garstecki, D. (2008). Background noise levels and reverberation times in old and new elementary school classrooms. *Journal of Educational Audiology, 14,* 16-22.

Neuman, A. & Hochberg, I. (1983). Children's perception of speech in reverberation. *Journal of the Acoustical Society of America, 74*, 215-219.

Neuman, A., Wroblewski, M., Hajicek, J. & Rubenstein, A. (2010). Combined effects of noise and reverberation on speech recognition performance of normal-hearing children and adults. *Ear & Hearing, 31*, 336-344.

Picard, M. & Bradley, J.S. (2001). Revisiting speech interference in classrooms. *Audiology 40*, 221-244.

Pichora-Fuller, M.K., Schneider, B. & Daneman, M. (1995). How young and old adults listen to and remember speech in noise. *Journal of the Acoustical Society of America, 97*, 593-608.

Pisoni, D. & Remez, R. (Eds.). (2008). *The Handbook of Speech Perception*. Malden, MA: Blackwell Publishing.

Ricketts, T. & Galster, J. (2008). Head angle and elevation in classroom environments: Implications for amplification. *Journal of Speech Language Hearing Research, 51*, 516-525.

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

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

Ryalls, B. & Pisoni, D. (1997). The effect of talker variability on word recognition in preschool children. *Developmental Psychology, 33*, 441-452.

Shepard, A. (2010). "Aaron Shepard's RT Page: Scripts and tips for reader's theater," http://www.aaronshep.com/rt/index.html#RTE (Last viewed December 18, 2014).

Soto-Faraco, S., Calabresi, M., Navarra, J., Werker, J. & Lewkowicz, D. (2012). The development of audiovisual speech perception. In A. Bremmer, D. Lewkowicz, and C. Spence (Eds.). *Multisensory Development* (pp. 207-228). Oxford, UK: Oxford University Press.

Sumby, W. & Pollack, I. (1954). Visual contribution to speech intelligibility in noise. *Journal of the Acoustical Society of America, 26*(2), 212-215.

Valente, D.L., Plevinsky, H.M., Franco, J.M., Heinrichs-Graham, E.C. & Lewis, D.E. (2012), Experimental investigation of the effects of the acoustical conditions in a simulated classroom on speech recognition and learning in children *Journal of the Acoustical Society of America, 131*, 232-246.

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.

Wroblewski, M., Lewis, D., Valente, D. & Stelmachowicz, P. (2012). Effects of reverberation on speech recognition in stationary and modulated noise by school-aged children and young adults. *Ear & Hearing, 33*, 731-744.

Yang, W. & Bradley, J. (2009). Effects of room acoustics on the intelligibility of speech in classrooms for young children. *Journal of the Acoustical Society of America, 125*, 922-933.