



Published in final edited form as:

Ear Hear. 2018 ; 39(4): 783–794. doi:10.1097/AUD.0000000000000534.

Looking Behavior and Audiovisual Speech Understanding in Children with Normal Hearing and Children with Mild Bilateral or Unilateral Hearing Loss

Dawna E. Lewis¹, Nicholas A. Smith¹, Jody L. Spalding¹, and Daniel L. Valente^{1,2}

¹Boys Town National Research Hospital, Omaha, NE

Abstract

Objectives—Visual information from talkers facilitates speech intelligibility for listeners when audibility is challenged by environmental noise and hearing loss. Less is known about how listeners actively process and attend to visual information from different talkers in complex multi-talker environments. This study tracked looking behavior in children with normal hearing (NH), mild bilateral hearing loss (MBHL), and unilateral hearing loss (UHL) in a complex multi-talker environment to examine the extent to which children look at talkers and whether looking patterns relate to performance on a speech-understanding task. It was hypothesized that performance would decrease as perceptual complexity increased and that children with hearing loss would perform more poorly than their peers with NH. Children with MBHL or UHL were expected to demonstrate greater attention to individual talkers during multi-talker exchanges, indicating that they were more likely to attempt to use visual information from talkers to assist in speech understanding in adverse acoustics. It also was of interest to examine whether MBHL, versus UHL, would differentially affect performance and looking behavior.

Design—Eighteen children with NH, 8 children with MBHL and 10 children with UHL participated (8–12 years). They followed audiovisual instructions for placing objects on a mat under three conditions: a single talker providing instructions via a video monitor, four possible talkers alternately providing instructions on separate monitors in front of the listener, and the same four talkers providing both target and non-target information. Multi-talker background noise was presented at a 5 dB signal-to-noise ratio during testing. An eye tracker monitored looking behavior while children performed the experimental task.

Results—Behavioral task performance was higher for children with NH than for either group of children with hearing loss. There were no differences in performance between children with UHL and children with MBHL. Eye-tracker analysis revealed that children with NH looked more at the screens overall than did children with MBHL or UHL, though individual differences were greater in the groups with hearing loss. Listeners in all groups spent a small proportion of time looking at relevant screens as talkers spoke. Although looking was distributed across all screens, there was a

Address correspondence to Dawna E. Lewis, Boys Town National Research Hospital, 555 N. 30th Street, Omaha, NE 68131, USA. 402-498-6351 (fax). 402-498-6607 (phone). dawna.lewis@boystown.org.

²Current Affiliation: JMP Division, SAS Institute Inc., Cary, NC

Conflicts of Interest: The first author (DL) is a member of the Phonak Pediatric Advisory Board but that membership has no conflicts with the content of this manuscript.

bias toward the right side of the display. There was no relationship between overall looking behavior and performance on the task.

Conclusions—The current study examined the processing of audiovisual speech in the context of a naturalistic task. Results demonstrated that children distributed their looking to a variety of sources during the task, but that children with NH were more likely to look at screens than were those with MBHL/UHL. However, all groups looked at the relevant talkers as they were speaking only a small proportion of the time. Despite variability in looking behavior, listeners were able to follow the audiovisual instructions and children with NH demonstrated better performance than children with MBHL/UHL. These results suggest that performance on some challenging multi-talker audiovisual tasks is not dependent on visual fixation to relevant talkers for children with NH or with MBHL/UHL.

INTRODUCTION

In the United States, children with mild bilateral hearing loss (MBHL) or unilateral hearing loss (UHL) make up over 5% of the school-age population (Niskar et al. 1998; NIDCD 2006). These children often demonstrate poorer speech perception in noise than peers with normal hearing (NH; Bess & Tharpe 1986; Crandell 1993; Johnson et al. 1997). In complex environments such as classrooms, where poor acoustics may negatively affect access to auditory information, children with MBHL or UHL may be more reliant on visual information to aid in speech understanding than are peers with NH. It is important, therefore, to examine looking behaviors of children with NH and children with MBHL or UHL in these complex multisensory and multi-talker environments.

Previous work has demonstrated that hearing loss affects listeners' behavior in complex environments. Ricketts and Galster (2008) reported that children with hearing loss were more likely to turn toward brief utterances by secondary talkers in classrooms than were children with NH, and suggested that such attempts could have negative consequences if they draw attention away from the primary talker. Sandgren and colleagues (Sandgren et al. 2013; 2014) reported that children with bilateral mild-moderate hearing loss were more likely to look at the face of their partner with NH during a dyadic referential communication task than vice versa. The same may be true for children with MBHL or UHL, for whom the auditory signal is degraded in a more chronic and persistent way than in peers with NH. This study tracked listeners' speech understanding and looking behavior in a complex multi-talker environment, to test the degree to which children look at talkers and whether looking patterns relate to performance on a speech comprehension task.

In complex multi-talker environments, the listener first must perceptually segregate individual talkers from each other (and from other background sounds) and then selectively attend to the talker of interest (Cherry 1953; Bregman 1990). Children are poorer than adults at segregating sound sources and selectively attending to target speech, particularly when the non-target signal also is speech (Hall et al. 2002; Leibold & Buss 2013). As listening tasks in classrooms become more complex, the cognitive demands increase (Picard & Bradley 2001; Klatte, Hellbrück et al. 2010; Klatte, Lachmann et al. 2010). The negative effects of degraded environmental acoustics (e.g., background noise) on a listener's ability to process

speech input also may increase (Klatte, Lachmann et al. 2010; Shield & Dockrell 2008). If the listening effort required to process the acoustic-phonetic aspects of the speech signal is increased in multi-talker listening environments, children may have fewer resources available for comprehension of the message.

The presence of visual information plays an important role in the processing of speech in multi-talker environments. In addition to the auditory information available in these environments, listeners/observers actively sample and selectively attend to visual information by moving their own eyes, head and body to orient to sources of information (Findlay & Gilchrist 2003; Hayhoe & Ballard 2005). In the presence of background noise or hearing loss, visual information provides a valuable supplementary source of input that can help listeners compensate for a degraded auditory signal.

When noise reduces the audibility of speech, the ability to see the talker increases intelligibility for listeners (Sumbly & Pollack 1954; Grant & Seitz 2000; Arnold & Hill 2001; Helfer & Freyman 2005; Ross et al. 2007). Much research has focused on how individuals use visual information from the talker's mouth (lips, tongue, and teeth) to derive a perceptual benefit beyond the acoustical signal (e.g., Summerfield 1992; Grant et al. 1998; Bernstein et al. 2004; Chandrasekaran, et al. 2009; Scarborough et al. 2009; Files et al. 2015). However, it is clear that visual information beyond the talker's mouth contributes to speech intelligibility. For example, visible movements of the talker's head and face enhance speech recognition even when the mouth movements are not visible (Munhall, Jones et al. 2004; Thomas & Jordan 2004; Davis & Kim 2006; Jordan & Thomas 2011).

Visual information beyond the mouth contains other paralinguistic information that is relevant to communication, such as indicating active listening, providing feedback, or signaling turn taking (see Wagner et al. 2014 for a review). Visual emotion recognition also is associated with movement of the eyes and whole face (Guo 2012; Wegrzyn et al. 2017). In addition, iconic gestures from the talker's hands can provide information that supplements that of the auditory signal during narratives or conversations, particularly regarding relative position and size (Holler et al. 2009; Drijvers & Özyürek 2017).

Being able to see talkers can improve localization (Shelton & Searle 1980), and visual cues often are more salient than auditory cues in determining the location of an audiovisual source in space (Bertelson & Radeau 1981). This is particularly important in multi-talker environments and research in adults has suggested that the availability of a visual cue to the location of a sound can enhance identification of auditory stimuli for adults with NH (Best et al. 2007) and hearing loss (Best et al. 2009) when that visual stimulus can be easily located. Altogether, the above research demonstrates the importance of visual information, beyond lipreading, to the processing of audiovisual speech.

Although children benefit from the availability of visual cues to aid in understanding speech, they may not use those cues to the same extent as adults. For example, the ability to benefit from visual information in speech perception under adverse listening conditions shows a protracted development through childhood (Ross et al. 2011; Maidment et al. 2015), and interacts with other developing aspects of language processing at multiple levels (Sekiyama

& Burnham 2008; Fort et al. 2012; Erdener & Burnham 2013; Jerger, et al. 2014; Lalonde & Holt 2016). Children also may be less accurate at judging visual cues that are not tied to the phonology of the speech signal (e.g., facial cues from the talker indicating uncertainty; Krahmer & Swerts 2005).

Despite research showing benefits of combined auditory and visual input for speech understanding, the addition of visual information has the potential to compete with other task-related processing under conditions of high cognitive demands, negatively affecting adults' speech understanding (Alsius et al. 2005; Mishra et al. 2013). In addition, both adults and children may avert their gaze from talkers during some complex tasks as a means of emphasizing other processing (Ehrlichman 1981; Doherty-Sneddon et al. 2001).

Recent studies have used eye tracking to examine how observers distribute their visual attention to regions on the faces of single talkers, in adults (Vatikiotis-Bateson et al. 1998; Lansing & McConkie 1999; Buchan et al. 2007; 2008;), children (Irwin et al. 2011; Irwin & Brancazio 2014;), and infants (Lewkowicz & Hansen-Tift 2012; Smith et al. 2013). However, single-talker experiments are not representative of the complex social environments in which perceivers must dynamically navigate and select among multiple talkers. In single-talker paradigms there are fewer alternatives for visual attention, which could promote an assumption that observers will attend to talkers' faces when available. However, this assumption has not received rigorous testing, and may not be true for complex environments. Foulsham et al. (2010) found that observers' gaze did track who was talking in a multi-talker conversation, but that allocation was driven by factors such as social status. Other work has shown that the presence of sound increases attentional synchrony in multi-talker displays (Foulsham & Sanderson 2013).

In this study, children's looking behavior was tested in the context of a speech perception testing procedure previously described in Lewis et al. (2014), which compared children's ability to follow verbal instructions for placing objects on a mat under three audiovisual contexts. The single-talker condition (ST) represented a simple one-on-one interaction, with one talker providing instructions via a video monitor. In the multi-talker (MT) condition, task-relevant information came alternately from four possible talkers on separate monitors in front of the listener. The multi-talker with comments (MTC) condition differed from the MT condition in that the four talkers produced both target and non-target information, with occasional interruptions. Overall, these three conditions varied the perceptual complexity of the social environment by manipulating the distribution of information across talkers as well as the relevance of individual talkers' speech to the experimental task. Lewis et al. (2014) completed testing in the presence of noise and reverberation. The current study used background noise only.

The experimental task required children to attend to and process spoken instructions. Doing so could be affected by the availability of both auditory and visual input from target and non-target talkers. Children also needed to find the correct object for placement among a group of objects and place that object in its correct position on the mat. This task, therefore, required the simultaneous coordination and interaction of auditory, visual, linguistic, and cognitive processes.

The assessment of performance accuracy served as a measure of speech understanding. Based on our previous work (Lewis et al. 2014), it was hypothesized that speech understanding would decrease as perceptual complexity increased both for children with NH and children with MBHL or UHL. It also was hypothesized that children with hearing loss would perform more poorly than their peers with NH. For children with MBHL, reduced audibility in both ears could have negative consequences on speech understanding in the experimental task. While it was possible that children with UHL would demonstrate better performance than those with MBHL because they have NH in one ear, poor speech recognition in noise has been shown for children with UHL even when speech is presented from the front or toward the ear with NH (Bess et al. 1986). Thus, it also was possible that the two groups of children with hearing loss would demonstrate similar performance on this task.

Eye tracking provided measures of children's overall looking behavior and its relationship to task performance. The current study analyzed general looking behaviors (i.e., whether participants looked at the screens with talkers) rather than whether participants' looking focused on talkers mouths (i.e., lipreading). Comparing children with NH to children with MBHL or UHL provided information about how hearing loss may influence looking strategies in these children and how those strategies, in turn, could support speech understanding. By tracking where children looked during the task, we examined how their looking behaviors related to their performance on the task. It was hypothesized that children with MBHL or UHL and children with NH would differ, with the children with hearing loss demonstrating greater attention to individual talkers during multi-talker exchanges. Such results could indicate that the children with hearing loss were more likely to attempt to use visual information from those talkers to aid in speech understanding during the complex instruction-following task. Attempts to focus on individual talkers could interact with accuracy on the instruction-following task in a variety of ways. It was possible that children who looked at talkers more often and for a longer amount of time used the combined auditory and visual information to enhance performance when compared to those who looked less. Conversely, it was possible that such looking strategies could interfere with processing of auditory information, resulting in poorer scores for those children, particularly in the most complex multi-talker condition. Although children with UHL have normal hearing in one ear, benefits of binaural processing that can improve speech understanding may be limited or absent, depending on the degree of hearing loss in the poorer hearing ear (Hartvig Jensen et al., 1989; Humes et al. 1980; Newton 1983). Some of the acoustic cues that aid in localization will be less salient for children with UHL than for children with NH or MBHL (for whom the signal is equally audible in both ears), resulting in reduced accuracy and speed when attempting to locate talkers using the auditory signal. Such potential differences could interact with looking behavior across groups of children.

MATERIALS AND METHODS

Participants

Eighteen 8- to 12-year-old children with NH and 18 children with MBHL or UHL participated (see Table 1 for hearing-status group definitions). Per parent report, children had

normal or corrected-to-normal vision and were native speakers of English. All children scored within 1.25 SD of the mean on the Wechsler Abbreviated Scale of Intelligence (Wechsler, 1999). No child scored >2 SD below the mean on the Peabody Picture Vocabulary Test (PPVT IV; Dunn & Dunn 2007) and only 1 child (with MBHL) scored >1 SD below the mean (See Table 2).

Eight of the children with MBHL or UHL had congenital hearing loss, three had acquired hearing loss, and the etiology for seven was unknown. In order to recruit sufficient numbers of children with MBHL or UHL, the number of children at each year of age across the range from 8–12 years was not the same. However, the NH and combined MBHL/UHL groups were age-matched, resulting in the same age distributions within this range for the groups. Children with UHL and MBHL may not use hearing aids or may be inconsistent in their hearing-aid use (Davis et al. 2002; Fitzpatrick et al. 2010; American Academy of Audiology 2013; Walker et al. 2013; Fitzpatrick et al. 2014). To assess performance effects of hearing loss without the benefit of amplification, children with hearing loss did not use amplification during any of the listening tasks.

The Institutional Review Board of Boys Town National Research Hospital (BTNRH) approved this study. Written assent and consent to participate were obtained for each child and they were compensated monetarily for participation. The simulated environment, test materials, procedures and scoring methods are described in detail in Lewis et al. (2014) and will be described briefly here.

Simulated Environment

An audiovisual (AV) instruction-following task was created using individual video recordings of either a single child talker (female) or four different child talkers (2 male, 2 female) who were 8–12 years old and spoke general American English. These children provided instructions for placing objects on a mat that was situated in front of the listener. Video recordings were made in a meeting room at BTNRH using four video cameras (JVC Evario Advanced Video Coding High Definition hard disk cameras, GZ-MG630, Wayne, NJ). For the recordings, the four cameras were placed at a position representing that of the seated listener and faced each talker. Talkers wore omnidirectional lavalier microphones (Shure, ULX1-J1 pack, WL-93 microphone, Niles, IL) to record their speech. To ensure a natural flow to the interaction during recordings, all talkers were present during the video recordings and were free to look at each other during the instructional task. They were not required to look at the cameras when giving instructions so, although all talkers were clearly visible, participants did not always have a straight-ahead view of the talker's face. These recording strategies provided a more realistic visual representation of talkers for our experimental task than one in which all talkers would be required to look directly at the camera when they spoke. However, these more realistic interactions do not preclude use of information from talkers' lips and mouths to aid in speech understanding, as supported by studies demonstrating that audiovisual speech perception is robust to major occlusions of the face (Jordan & Thomas, 2011) and to a wide range of viewing orientations (Jordan & Thomas, 2001).

Test Materials and Instructions

Ten instructions were given, each requiring the child to place 1 of 16 objects that varied on several dimensions (e.g., size, color, shape; see Appendix A) in a specified location/orientation relative to the edges of a rectangular mat or to other items. Sentences used for instructions in the current study were age-appropriate and their structure covered a range of complexity (e.g., Put the red airplane in the center of the mat with the front pointing to you). The instructions followed the same format in all conditions (See Appendix B for an example of the instructions). Recordings were created for three experimental conditions: single talker (ST); multiple talkers (MT); and multiple talkers with competing comments (MTC). For the first two experimental conditions (ST, MT), each child talker read his/her instructions and any talkers in the condition followed that instruction for placing objects on mats before the next instruction. The participant could not see the mats of the talkers and did not know how well they were following instructions. For the third condition (MTC), a scripted dialog was created so that the child talkers made other comments in addition to the instructions for the task. The audio recordings were equalized for sound pressure level across all talkers.

Eye Tracking

Children's looking behavior was recorded using a faceLAB 4 video-based eye-tracking system (Seeing Machines, Canberra, Australia). The system hardware consisted of an infrared stereo camera. The faceLAB system tracked the participant's gaze by first tracking the position of the head based on individualized models of the participant's head geometry. During the initial calibration procedure, these models were generated using the full head model option in faceLAB in which facial reference points were identified in a set of three snapshots with the participant's head faced forward (0°) and 90° to the left and right. After generating the participant's head model, their gaze was calibrated by asking them to fixate on the left, then the right, camera of the eye tracker. The faceLAB system permitted the tracking of eye gaze to multiple objects or surfaces represented in a world model. In this world model, five planar regions were defined corresponding to each of the four talker screens, along with the mat on which the participant performed the task.

The eye tracker data consists of a number of measures describing the position and orientation of the participant's head and eyes, obtained at a sampling rate of 60 Hz. In the current application, the gaze object intersection is a categorical data value that reflects the real world object (talker screen or mat) that intersects the participant's gaze at a particular point in time. Eye tracking studies are prone to missing and noisy data due to occasional lapses in the visibility of the participant's eyes by the eye tracker cameras (Wass et al. 2013). The faceLAB analysis software provides gaze quality measures for each eye tracker sample that specify whether the gaze estimates provided are based on usable data from one or both eyes, or are merely inferred on the basis of the more robust measure of head position and orientation. The eye tracking data analysis below makes use of both gaze object intersection and gaze quality measures.

A Mac Pro computer running Max/MSP 5 software (Cycling 74, Walnut, CA) controlled stimulus presentation. Eye-tracker data were streamed to this computer in UDP format via an Ethernet connection. The Max software simultaneously recorded the faceLAB and

QuickTime frame numbers to permit a temporal alignment of stimulus and eye-tracking data.

Scoring

Listening tasks were scored based on 3–4 properties of each instruction: item, position, step (1–10), and orientation (where applicable). For each task, every correct property was worth one point. The total number of correct properties was divided by the possible total points for a percent correct score. For example, if the fourth instruction was “Put the green car in the bottom right corner of the mat with the front pointing to the large bone”, it was worth four points (item=green car; position=bottom right corner; step=4 (i.e., fourth instruction); orientation= pointing to large bone). If one object was placed on the mat incorrectly but another object was correctly placed in relation to it at a later step, the second item’s placement was scored as correct. For example, if the child had placed the large bone incorrectly on the mat following an earlier instruction but placed the car correctly in relation to the bone, the car would be scored correct for orientation. In addition, if the wrong object was placed on the mat or a correct object was positioned incorrectly during one step and corrected at a later step, the score for that item would be adjusted (e.g., after the change, the position or item might now be correct but the step would be incorrect).

A video camera above the child recorded his/her response. A researcher scored the tasks as they were administered then reviewed the video recording to re-check scores and modify them as needed.

Procedure

The simulated AV instruction-following task, with the video recordings described above, was implemented in a sound booth at BTNRH with the following baseline acoustic characteristics: background noise level = 37 dBA; reverberation time (T30mid) = 0.06 sec; early decay time = 0.05 sec. Speech was presented at 65 dB SPL at the listener’s location. Multi-talker noise (20 talkers; Auditec, Inc., St. Louis, MO) was mixed with the recordings at a 5 dB signal-to-noise ratio. This noise represented a potential noise level in an occupied classroom (Dockrell & Shield 2006; Finitzo 1988; Jamieson et al. 2004). For the experimental task, the child sat at a table facing four LCD monitors (Screens 1, 4: Dell 2007 WFPb, Plano, TX; Screens 2, 3: Samsung Syncmaster 225BW, Ridgefield Park, NJ) with loudspeakers (M-Audio Studiophile AV 40, Cumberland, RI) mounted on top (Figure 1). Across screens, talkers’ heads subtended angles of roughly $7.7^\circ \times 17.7^\circ$.

Children were familiarized with the task prior to testing. They followed sets of 10 AV instructions for placing objects on a mat. They could look at talkers as much or as little as they chose during the task. There were three conditions and their presentation order was the same for all participants. In the first condition (ST), a single talker presented instructions on screen 2. That talker was used for this condition only. This condition served as the baseline for the child’s ability to follow the instructions. The second condition (MT) utilized four different talkers on screens 1–4 across from the child. The talkers took turns providing instructions for the task and all were visible at all times. The order in which the talkers spoke was varied to prevent the child from anticipating the next talker. The third condition (MTC)

used the same four child talkers as in the second. However, while one talker was providing an instruction, the other talkers intermittently interrupted or talked to each other. This condition represented a task during which listeners must separate target speech from distracting speech. The order of conditions followed that of Lewis et al. (2014) and was selected due to potential effects on children's willingness to complete the task if some of them considered the initial condition too difficult. Each condition lasted less than 5 minutes. All children were able to complete each of the conditions.

RESULTS

The data consisted of behavioral performance scores (percent correct) for each participant in each of three conditions (ST, MT, MTC), together with eye-tracking measures of where participants were looking as they performed in each of these conditions. Results begin with an analysis of behavioral performance, followed by an analysis of listeners' looking behavior.

Behavioral Task Performance

Percent correct scores were examined statistically within a linear mixed-effects model framework using the LME4 (Bates et al. 2015) package version (1.1–12) in R (The R Foundation for Statistical Computing; Vienna, Austria). This model is akin to a regression analysis that describes the relation between multiple predictor variables and a criterion variable (behavioral performance), but which is appropriate for repeated-measure designs. In this model, the predictor variables, hearing status and experimental condition, were treated as fixed effects, while participant was treated as a random effect.

The fixed effects factors were zero-sum coded as dummy variables. For hearing status, one variable coded whether the listeners were in the NH group (+0.5) or were part of the combined (MBHL/UHL) hearing loss subgroups (−0.5). The second variable distinguished between the children based on hearing loss type: MBHL (+0.5), UHL (−0.5), NH (0). The same approach was used for experimental condition, with one variable contrasting the ST condition (+0.5) with the two multi-talker conditions (MT and MTC; −0.5), and a second variable contrasting all conditions: MT (+0.5), MTC (−0.5), ST (0). Participant was treated as a random effect in the model. This allowed for appropriate accounting of variance related to the repeated-measures design of this study. In this analysis, models were generated using random effects structures that included random intercepts.

Task performance is shown in Figure 2. Children in the NH group performed significantly better ($M = 87.8\%$, $SD = 11.7$), than those in the combined hearing loss groups, ($M = 80.3\%$, $SD = 13.4$), $t(34.74) = 2.05$, $p = .048$. No significant effect of hearing loss type was found, $t(34.74) = 0.004$, $p = .99$. A corollary approach to evaluating the influence of a factor is to compare models that include and do not include the factor of interest. This form of multi-model comparison found that a model that included the hearing loss type contrast did not significantly outperform a model that only included the NH vs combined MBHL/UHL contrast, $\chi^2(3) = 4.68$, $p = .20$, based on values of AIC (Akaike's Information Criterion). Comparisons between AIC values are akin to comparisons between adjusted R^2 values in traditional regression analyses.

A significant effect of experimental condition was found, with performance in the ST condition ($M = 91.0\%$, $SD = 10.0$) higher than the combined multi-talker conditions ($M = 80.6\%$, $SD = 13.1$), $t(66) = 6.78$, $p = .0001$. A significant effect of multi-talker condition was also found, $t(66) = 3.41$, $p = .001$, with better performance in the MT condition ($M = 83.5\%$, $SD = 12.5$) than MTC condition ($M = 77.6\%$, $SD = 13.1$). No significant interactions between hearing status groups and condition were found for either the ST vs MT/MTC, $t(66) = -1.12$, $p = .26$, or MT vs MTC contrasts, $t(66) = -1.78$, $p = .08$.

Multi-model comparisons based on values of AIC found that a model that included ST vs MT/MTC contrast outperformed a baseline model containing only the random effects, $X^2(1) = 30.81$, $p < .0001$, and that the addition of the MT vs MTC contrast significantly further improved the model fit, $X^2(1) = 9.81$, $p = .002$. A significant improvement was found when hearing status was added, $X^2(3) = 8.86$, $p = .03$, but no further improvement was found with the addition of hearing loss type, $X^2(3) = 4.68$, $p = .20$.

Looking Behavior

The raw eye-tracker data consist of series of gaze object intersection values (i.e., whether the participant was looking at Screen 1, 2, 3, 4) at particular points in time, along with associated measures of the gaze quality, recorded at a rate of 60 samples per second. An attempt was made to reference eye gaze to the mat. However, the angle of the child's head when looking down made examining this region of interest difficult. Consequently, analyses are limited to the four screens. The number of samples was proportional to the total stimulus duration for each condition (ST: 243.6 sec = 14,077 samples; MT: 149.8 sec = 8,986 samples; MTC: 190.8 sec = 11,450 samples). These data, therefore, permitted the calculation of looking time measures for each region of interest (four different screens showing four different talkers).

The analysis adopted a multi-tiered approach. The first analysis assessed participant looking to screens with individual talkers during their utterances, permitting an examination of attention to the relevant talkers at particular points in time. The second analysis examined the overall distribution of looking to all of the screens over the course of the entire interaction in a condition as a whole.

In both of these analyses, a subset of samples meeting a gaze quality criterion of 2 (on a 0–3 categorical scale specific to the faceLAB software) for at least one eye was extracted from the total eye-tracker dataset for an individual participant in each condition. Decreases in gaze quality reflected losses of visibility of the face by the eye tracker due to the participant moving their head out of the field view of cameras, looking away from the cameras (as was sometimes the case when participants looked down to manipulate objects on the mat), or obstructing their face with their hands. Overall, gaze quality for this experiment met criterion for 68.1 % of all samples, with similar levels for the NH ($M = 66\%$), MBHL ($M = 70.1\%$) and UHL ($M = 70.5\%$) groups. However, due to some calibration problems with four participants (2 with NH, 2 with MBHL/UHL), it was not possible to reference their eye gaze data with the various screen locations. These four participants were not included in the eye-tracker analyses.

Looking at relevant screens—The first step in the analysis examined the proportion of time that children spent looking at the relevant screen, meaning the one displaying an active talker who was providing task relevant information. In the ST condition, the relevant screen was always screen 2, but in the MT and MTC conditions, the relevant screen changed over the course of interactions among talkers.

As can be seen in Figure 3, the overall proportion of time looking to the relevant screen was low, with children from all groups looking at the relevant screen an average of 19.1% of the time during instructions. No effect was found for the NH vs MBHL/UHL contrast, $t(33.46) = 0.315, p = .75$, or the MBHL vs. UHL contrast, $t(33.46) = -0.274, p = .79$. No significant effects of experimental condition were found between the ST and multi-talker conditions, $t(57.45) = -.37, p = .71$, or between the two multi-talker conditions, $t(57.78) = 0.16, p = .88$. As above, a multi-model analysis was completed. The addition of factors related to condition, hearing status or hearing loss type did not lead to significant improvements in the fit of the model beyond the baseline containing only random effects.

The previous analyses focused on aggregate measures of overall looking. To examine the looking patterns over time, a measure of gaze switching was derived from the data. In this analysis, a switch consisted of a transition from looking at one screen to looking at a different screen. A switch rate (*spm*, switches per minute) was calculated by dividing the total number of switches in a block by the block duration (Figure 4).

Significantly more switching was found in the MTC condition ($M = 24.3, SD = 12.2$ spm) than the MT condition ($M = 21.6, SD = 11.7$), $t(58.1) = -2.97, p = .004$. This effect of multi-talker condition interacted significantly with hearing loss type, $t(58.04) = 2.01, p = .049$, with children in the MBHL group showing greater increases in switch rate across multi-talker conditions (MT = 19.7 spm vs. MTC = 26.6 spm) than children in the UHL group (MT = 24.6 spm vs. MTC = 25.3 spm). No significant differences in switch rate were found between the two hearing status groups, $t(30.42) = -.0365, p = .72$, or between the two hearing loss types, $t(30.3) = 0.23, p = .82$. Switch-rate differences between the multi-talker conditions and the single-talker condition also were not significant, $t(58.2) = -1.68, p = .10$.

Looking at any screen—Figure 5 illustrates the proportion of total time in each condition during which children were looking at any of the four screens. As can be seen in this figure, the majority of the children with NH (left panel) spent >75% of the time looking at the screens. However, there was a split in the looking behavior of the children with hearing loss. Approximately half of the children with hearing loss exhibited a pattern that was similar to that of the children with NH but the proportion of looking time was lower for the remainder of the children. Statistical analysis revealed that, overall, children with NH looked at the screens significantly more ($M = 84.9\%, SD = 22.1$) than did children with MBHL/UHL ($M = 63.0\%, SD = 32.0$), $t(29.03) = 2.12, p = .04$. No significant effect of hearing loss type was found, $t(29.03) = -0.30, p = .77$, and the addition of hearing loss type did not significantly improve the model fit, $X^2(3) = 2.60, p = .46$. No significant effects for experimental condition, $t(57.01) = -0.85, p = 0.40$, or interactions between experimental condition and hearing status group were found. A multi-model analysis was carried out. The addition of any factors related to experimental condition, hearing status, or hearing loss type

did not lead to significant improvements in the fit of the model beyond the baseline containing only random effects.

Since children spent the majority of time looking at screens, but not necessarily the relevant screen, the breakdown of looking to each of the four screens was examined. Figure 6 shows the proportion of time looking to each of the four screens, relative to the time looking to any screen. Participants in all groups distributed their looking broadly, spending some time looking at all four screens even in the ST condition in which the talker was only displayed on Screen 2. In the MT and MTC conditions, talkers were visible on all four screens, and produced relevant speech roughly equal proportions of the time. Nevertheless, listeners showed an overall right-side bias, looking at Screens 3 and 4 65% of the time. Screen 1 on the far left was attended to only 9% of the time overall.

Relation between looking and performance—The final step in the analysis was to examine how looking behavior related to performance in the task. A new model was created with looking to any screen and hearing status (NH vs. MBHL/UHL) as predictors. Neither the main effect of looking, $t(29.26) = -1.23, p = .29$, nor the interaction of looking \times hearing status, $t(29.26) = 0.655, p = .52$, were significant. This analysis was repeated using looking to the *relevant* screen and hearing status (NH vs. MBHL/UHL) as predictors. Once again, neither the main effect of looking, $t(88.13) = 0.687, p = .49$, nor the interaction of looking \times hearing status, $t(88.13) = -1.488, p = .14$, were significant.

DISCUSSION

The purpose of the current study was to examine looking behavior and performance during an instruction-following task in children with NH and children with MBHL or UHL. In particular, we examined how children performed in a complex multi-talker environment, where they looked, how looking patterns differed as a function of hearing status, and whether looking behavior was related to performance on the task.

Behavioral Task Performance

Results supported the initial hypothesis that performance would decrease as perceptual complexity increased for both children with NH and children with MBHL/UHL. These findings are in agreement with those from Lewis et al. (2014) that examined performance for children with NH on the same task in noise and reverberation. In the current study, children with hearing loss also performed more poorly overall than their peers with NH, supporting our hypothesis. The degree of hearing loss for the children in the current study falls within a restricted threshold range for children with MBHL and, although of potentially greater degree, affects only one ear of children with UHL. Even these small differences in access to the auditory signal when compared to children with NH contributed to hearing-status group differences. When comparing the subgroups of children with hearing loss, however, results showed no significant differences in performance as a result of whether the loss was bilateral or unilateral. A similar absence of bilateral versus unilateral hearing loss effects was found by Lewis et al. (2015) when examining performance for these same children on a comprehension task in a simulated classroom environment. Thus, the presence of hearing loss, but not whether that loss was unilateral or bilateral resulted in reduced performance

relative to children with NH. The absence of any effects of hearing-loss subgroup on performance in this study could be related to interactions between the complex nature of the task and the hearing loss. Children with MBHL and UHL may perform similarly in their ability to complete the task but the underlying factors affecting performance may differ. For example, for children with UHL the degree of hearing loss could affect access to binaural cues that assist listeners in complex environments. In this study, participants were not recruited to provide a representative range of degree of hearing loss in the group with UHL. It also is possible that unilateral/bilateral differences would have played a greater role if children with greater degrees of bilateral hearing loss were included. Future research should address these potential factors with larger subject groups representing a wide range of hearing thresholds. However, it is noteworthy that even a mild degree of bilateral hearing loss or hearing loss in only one ear affected performance in these children, who were not using amplification during testing. Recall that some children may have worn amplification in daily life.

Looking Behavior

Looking behavior during the instruction-following task was examined using eye-tracking procedures. It was hypothesized that children with MBHL/UHL would focus on individual talkers more than would their peers with NH. High percentages of time looking to relevant screens would support that hypothesis. However, the results demonstrated that the overall proportion of time spent looking at talkers as they were providing instructions was low for the majority of participants with both NH and MBHL/UHL and there were no hearing-status group differences. These results do not support the hypothesis. Thus, regardless of condition or the ease with which talkers might be identified, children did not consistently look at individual talkers as they spoke. The absence of a difference between children with hearing loss and children with NH was unexpected. It had been expected that children with MBHL/UHL would attempt to look at talkers as a means of augmenting an auditory signal that was degraded by both noise and hearing loss. These results indicate that children with NH and those with hearing loss were not focusing on talkers for the duration of their spoken instructions in this task.

Measures of gaze switching provided information about the variation in participants' looking over time during the task. Results revealed that children with MBHL showed a greater increase in switch rate from the MT to MTC condition than children with UHL. These findings suggest that, when compared to UHL, hearing loss in both ears may result in more attempts to locate talkers as the perceptual complexity of a multi-talker task increases.

The analysis examining children's looking at relevant talkers showed that participants spent a small proportion of the time looking at those talkers as they spoke. It also was useful to know the proportion of time that listeners looked at any talker out of the total time talkers were speaking during the listening conditions. For this measure, there was a significant difference between children with NH and children with MBHL/UHL. All but two children with NH spent a high proportion of their time looking at screens. Given the low proportion of time children spent looking at relevant talkers as they spoke, these results could suggest that they were scanning the talkers throughout the task.

An examination of the distribution of looking across the four screens helped to determine whether children were scanning across all screens equally or were showing any looking biases. Although looking was distributed across all screens, when comparing the right versus left of center, there was a bias toward the right (65%) and very little time was spent looking toward screen 1 (9%), which was farthest to the left. These findings suggest that children were looking at the various screens, but not equally. It is important to note, however, that looking behavior cannot be tied directly to locating or following talkers since the proportion of looking times was similar in the ST condition, when the only talker was on screen 2, to the multi-talker conditions when there were talkers on all screens. Our current eye-tracking procedures do not allow us to examine these looking strategies more specifically, which should be addressed in future studies.

An important point should be made about what eye tracking does and does not say about visual processing. Eye-tracking systems attempt to estimate which parts of the visual scene project onto the observers' fovea – an area of the retina specialized for high spatial frequency processing of visual details. However, information in the visual periphery projects onto non-foveal regions, which are processed through channels more specialized at dynamic visual information at lower spatial frequencies (Shapley, 1990; Merigan & Maunsell, 1993). Our finding that listeners were not foveating on the relevant talkers does not mean that visual information from the talkers was not perceived or processed. Visual information also could have been processed through peripheral channels, which are very important for audiovisual speech perception. For example, perceptual studies using spatially filtered audiovisual speech stimuli have shown that lower spatial frequencies are the main contributors to audiovisual speech perception (Munhall, Kroos et al. 2004), although there are individual differences in sensitivity to high spatial frequency information in speech (Alsius, et al. 2016). Furthermore, studies of audiovisual speech integration have shown that the McGurk effect (McGurk & MacDonald 1976) persists even when the visual stimulus is up to 40° in the periphery (Paré, et al. 2003).

For the children with MBHL and UHL, the results (as displayed in Figure 5) were divided between those who spent the majority of time looking at the screens and those who looked at screens a low proportion of the time. Low proportions of looking suggest that those children were focusing their visual attention elsewhere. One possibility is that children who were not looking at talkers were instead focusing on the mat during the task. Research has shown that visual attention on to-be-processed items may reduce the amount of information that must be held in working memory during natural tasks (Droll & Hayhoe 2007). A child who focused on the screens during most of the instructions would have to hold the talkers' instructions in memory as they looked to the mat to select and place objects. In contrast, a child who focused on the mat may be able to begin to follow the auditory instruction as it was being presented. The latter scenario would support children's ability to switch attention to different auditory sources without attending to those sources visually. Focusing on the mat could have been a strategy used by some children to complete the task while reducing the need to hold all of the spoken information in working memory before following a particular instruction. Alternatively, some children may not have looked at either the mat or the talkers, gazing away as they processed the auditory information from the instructions (Ehrlichman, 1981; Doherty-Sneddon et al., 2001).

Droll and Hayhoe (2007) suggest that individual strategies in visual attention may be based on the cost or benefit that individuals associate with those strategies. It is possible that the current results represent different strategies adopted for addressing both visual and auditory information in a complex listening task. Some children may rely primarily on information about instructions held in working memory while others use the visual information supplied by the objects and the mat to supplement that information. By attending to objects on the mat, children may forgo visual enhancement effects associated with attending to the talker in order to reduce the working memory demands of the task by attending to the task space. To further understand looking strategies in a perceptually complex task, future research should examine exactly where children are looking during the task when they are not looking at talkers. Such an assessment was not possible using the eye-tracking system in the current study.

Interestingly, there was no relationship between overall looking behavior during this task and performance. Children who tracked the relevant talker did not tend to perform differently on the task than children who did not. This result suggests that the relationship between attention to visual and auditory stimuli in this instruction-following task was not simple. Although children with MBHL/UHL performed more poorly than children with NH, they did not spend more time looking at relevant talkers. The pattern of overall looking, however, suggests differences in the strategies among children, with greater differences in the group of children with MBHL/UHL than in the group of children with NH. Almost all of the children with NH chose to look at the talker space as they listened to the instructions but almost half of the children with hearing loss looked away from talkers during that time. Looking at talkers and looking elsewhere in the task space may both confer performance benefits, but because these options are mutually exclusive, different individuals may weigh these options differently, settling on a trade-off that maximizes their own performance in the task. Doherty-Sneddon et al. (2001) concluded that “if the visual signals do not have an “added value” for the task at hand, the cognitive resources required to process them may detract from processing of task-relevant information” (pp. 917–918). For some of the children with hearing loss, visual information in the talker space may have detracted from performance or, at least, may not have provided useful benefit.

How children come to adopt particular strategies and trade-offs may be affected by multiple complex factors. It is possible, for example, that increasing the cognitive load of the instruction-following task, either by additional levels of complexity or further degrading acoustics, would have resulted in different looking strategies than those seen here and may have produced different outcomes when comparing children with MBHL and UHL. This study also did not include measures of children’s working memory or auditory and visual attention skills to examine how those skills play specific roles in performance on complex listening tasks. Such measures should be included in follow-up studies.

The current study examined audiovisual speech understanding in the context of a naturalistic task that placed a collection of perceptual and task demands on the participant. Examining looking behavior and performance in this kind of task provides information about how children balance these demands. Many paradigms involving audiovisual speech perception present a single, centrally fixated, forward looking talker in the context of a simple response

task. In this kind of scenario, one might too easily presume that listeners give focused and undivided attention to the stimulus talker. However, these paradigms may have limited generalizability to real-life social situations. The current study challenges the assumption that audiovisual speech perception in the real world operates in a similar way. Our task more closely approximates a real-world multi-talker situation, in which focus on the relevant talker is less obvious and obligatory. One major finding of our study is that when the obligatory-attentional constraint is removed from the task, participants' looking is quite variable. Despite this variability, children in the current study were able to follow the audiovisual instructions, and children with NH demonstrated better performance than children with MBHL/UHL. These results suggest that performance on some challenging multi-talker audiovisual tasks is not dependent on visual fixation to relevant talkers for children with NH or with MBHL/UHL.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This work was supported by NIH grants R03 DC009675, R03 DC009884, T32 DC000013, P20 GM109023, and P30 DC004662. The content of this manuscript is the responsibility and opinions of the authors and does not necessarily represent the views of the National Institutes of Health. The first author (DL) is a member of the Phonak Pediatric Advisory Board but that membership has no conflicts with the content of this manuscript. Portions of the findings have been presented at professional meetings. The authors thank Bob McMurray, Lori Leibold, and Ryan McCreery for discussions regarding this manuscript and Christine Hammans for proofreading assistance.

Source of Funding:

This research was funded by the NIH/NIDCD/NIGMS.

References

- Alsius A, Navarra J, Campbell R, et al. Audiovisual integration of speech falters under high attention demands. *Current Biology*. 2005; 15:839–843. [PubMed: 15886102]
- Alsius A, Wayne RV, Paré M, et al. High visual resolution matters in audiovisual speech perception, but only for some. *Atten Percep Psychophys*. 2016; 78:1472–1487.
- American Academy of Audiology. Clinical Practice Guidelines. *Pediatric Amplification*. 2013. Downloaded September 24, 2015 from http://audiology-web.s3.amazonaws.com/migrated/PediatricAmplificationGuidelines.pdf_539975b3e7e9f1.74471798.pdf
- Arnold P, Hill F. Bisensory augmentation: A speechreading advantage when speech is clearly audible and intact. *Br J Psychol*. 2001; 92:339–353.
- Bates D, Mächler M, Bolker, et al. Fitting linear mixed-effects models using lme4. *J Statist Software*. 2015; 67:1–48.
- Bess F, Tharpe A. Case history data on unilaterally hearing-impaired children. *Ear Hear*. 1986; 7:14–19. [PubMed: 3949096]
- Bess F, Tharpe AM, Gibler AM. Auditory performance of children with unilateral sensorineural hearing loss. *Ear Hear*. 1986; 7:20–26. [PubMed: 3949097]
- Best V, Marrone N, Mason C, Kidd G, Shinn-Cunningham B. Effects of sensorineural hearing loss on visually guided attention in a multitalker environment. *J Assoc Res Otolaryngol*. 2009; 10:142–149. [PubMed: 19009321]
- Best V, Ozmeral E, Shinn-Cunningham B. Visually-guided attention enhances target identification in a complex auditory scene. *J Assoc Res Otolaryngol*. 2007; 8:294–304. [PubMed: 17453308]

- Bernstein L, Auer E, Takayanagi S. Auditory speech detection in noise enhanced by lipreading. *Speech Comm.* 2004; 44:5–18.
- Bertelson P, Radeau M. Cross-modal bias and perceptual fusion with auditory-visual spatial discordance. *Percept Psychophys.* 1981; 29:578–584.
- Bregman, A. *Auditory Scene Analysis: The Perceptual Organization of Sound.* Cambridge, MA: The MIT Press; 1990.
- Buchan JN, Paré M, Munhall KG. Spatial statistics of gaze fixations during dynamic face processing. *Soc Neurosci.* 2007; 2(1):1–13. [PubMed: 18633803]
- Buchan JN, Paré M, Munhall KG. The effect of varying talker identity and listening conditions on gaze behavior during audiovisual speech perception. *Brain Res.* 2008; 1242:162. [PubMed: 18621032]
- Chandrasekaran C, Trubanova A, Stillitano S, et al. The natural statistics of audiovisual speech. *PLoS Comput Biol.* 2009; 5:e1000436.doi: 10.1371/journal.pcbi.1000436 [PubMed: 19609344]
- Cherry C. Some experiments on the recognition of speech, with one and with two ears. *J Acoust Soc Amer.* 1953; 25(5):975–979.
- Crandell C. Speech recognition in noise by children with minimal degrees of sensorineural hearing loss. *Ear Hear.* 1993; 14:210–216. [PubMed: 8344478]
- Davis C, Kim J. Audio-visual speech perception off the top of the head. *Cognition.* 2006; 100:B21–B31. [PubMed: 16289070]
- Davis, A., Reeve, K., Hind, S., Bamford, J. Children with mild and unilateral hearing impairment. In: Seewald, R., Gravel, J., editors. *A Sound Foundation Through Early Amplification 2001: Proceedings of the Second International Conference.* Stäfa, Switzerland: Phonak Communications AG; 2002. p. 179-186.
- Dockrell J, Shield BM. Acoustical barriers in classrooms: The impact of noise on performance in the classroom. *Brit Educ Res J.* 2006; 32:509–525.
- Doherty-Sneddon G, Bonner L, Bruce V. Cognitive demands of face monitoring: Evidence for visuospatial overload. *Mem Cognit.* 2001; 29:909–919.
- Drijvers L, Özyürek A. Visual context enhanced: The joint contribution of iconic gestures and visible speech to degraded speech comprehension. *J Speech Lang Hear Res.* 2017; 60:212–222. [PubMed: 27960196]
- Droll J, Hayhoe M. Trade-offs between gaze and working memory use. *J Exper Psychol: Hum Percept Perform.* 2007; 33:1362–1365.
- Dunn, LM., Dunn, DM. *The Peabody Picture Vocabulary Test. Fourth.* Bloomington, MN: NCS Pearson, Inc; 2007.
- Ehrlichman H. From gaze aversion to eye-movement suppression: An investigation of the cognitive interference explanation of gaze patterns during conversation. *Brit J Soc Psychol.* 1981; 20:233–241.
- Erdener D, Burnham D. The relationship between auditory-visual speech perception and language-specific speech perception at the onset of reading instruction in English-speaking children. *J Exper Child Psychol.* 2013; 116:120–138. [PubMed: 23773915]
- Files B, Tjan B, Jiang J, et al. Visual speech discrimination and identification of natural and synthetic consonant stimuli. *Front Psychol.* 2015; 6 <https://doi.org/10.3389/fpsyg.2015.00878>.
- Findlay, JM., Gilchrist, ID. *Active vision: The psychology of looking and seeing.* Oxford University Press; Oxford: 2003.
- Finitzo, T. Classroom acoustics. In: Roeser, R., Downs, M., editors. *Auditory disorders in school children.* 2nd. New York, NY: Thieme Medical; 1988. p. 221-233.
- Fitzpatrick E, Durieux-Smith A, Whittingham J. Clinical practice for children with mild bilateral and unilateral hearing loss. *Ear Hear.* 2010; 31:392–400. [PubMed: 20054278]
- Fitzpatrick E, Whittingham J, Durieux-Smith A. Mild bilateral and unilateral hearing loss in childhood: a 20-year view of hearing characteristics, and audiological practices before and after newborn hearing screening. *Ear Hear.* 2014; 35:10–18. [PubMed: 24300117]
- Fort M, Spinelli E, Savariaux C, et al. Audiovisual vowel monitoring and the word superiority effect in children. *Int J Behav Develop.* 2012; 36:457–467.

- Foulsham T, Cheng J, Tracy J, Henrich J, Kingstone A. Gaze allocation in a dynamic situation: Effects of social status and speaking. *Cognition*. 2010; 117:319–331. [PubMed: 20965502]
- Foulsham T, Sanderson L. Look who's talking? Sound changes gaze behavior in a dynamic social scene. *Vis Cognition*. 2013; 21:922–944.
- Grant KW, Seitz PF. The use of visible speech cues for improving auditory detection of spoken sentences. *J Acoust Soc Am*. 2000; 108:1197–1208. [PubMed: 11008820]
- Grant K, Walden B, Seitz P. Auditory-visual speech recognition by hearing-impaired subjects: Consonant recognition, sentence recognition, and audio-visual integration. *J Acoust Soc Am*. 1998; 103:2677–2690. [PubMed: 9604361]
- Guo K. Holistic gaze strategy to categorize facial expression of varying intensities. *PLoS ONE*. 2012; 7:e42585. doi: 10.1371/journal.pone.0042585 [PubMed: 22880043]
- Hall J, Grose J, Buss E, et al. Spondee recognition in a two-talker masker and a speech-shaped noise masker in adults and children. *Ear Hear*. 2002; 23:159–165. [PubMed: 11951851]
- Hartvig Jensen J, Johansen PA, Borre S. Unilateral sensorineural hearing loss in children and auditory performance with respect to right/left ear differences. *Brit J Audiol*. 1989; 23:207–213. [PubMed: 2790305]
- Hayhoe M, Ballard D. Eye movements in natural behavior. *Trends in Cognitive Sciences*. 2005; 9(4): 188–194. <http://doi.org/10.1016/j.tics.2005.02.009>. [PubMed: 15808501]
- Helfer K, Freyman R. The role of visual speech cues in reducing energetic and informational masking. *J Acoust Soc Amer*. 2005; 117:842–849. [PubMed: 15759704]
- Holler J, Shovelton H, Beattie G. Do iconic hand gestures really contribute to the communication of semantic information in a face-to-face context? *J Nonverbal Behav*. 2009; 33:73–78.
- Humes L, Allen S, Bess F. Horizontal sound localization skills of unilaterally hearing-impaired children. *Audiology*. 1980; 19:508–518. [PubMed: 7425955]
- Irwin JR, Brancazio L. Seeing to hear? Patterns of gaze to speaking faces in children with autism spectrum disorders. *Front Psychol*. 2014; 5(397):1–10. [PubMed: 24474945]
- Irwin JR, Tornatore LA, Brancazio L, et al. Can Children With Autism Spectrum Disorders “Hear” a Speaking Face? *Child Develop*. 2011; 82:1397–1403. [PubMed: 21790542]
- Jerger S, Damian MF, Tye-Murray N, et al. Children use visual speech to compensate for non-intact auditory speech. *J Exper Child Psychol*. 2014; 126:295–312. [PubMed: 24974346]
- Jamieson DG, Kranjc G, Yu K, et al. Speech intelligibility of young school-aged children in the presence of real-life classroom noise. *J Acoust Soc Amer*. 2004; 15:508–517.
- Johnson C, Stein R, Broadway A, et al. “Minimal” high-frequency hearing loss and school-age children: Speech recognition in a classroom. *Lang Speech Hear Serv Schools*. 1997; 28:77–85.
- Jordan TR, Thomas SM. Effects of horizontal viewing angle on visual and audiovisual speech recognition. *J Exper Psychol: Hum Percep Perform*. 2001; 27:1386–1403.
- Jordan TR, Thomas SM. When half a face is as good as a whole: Effects of simple substantial occlusion on visual and audiovisual speech perception. *Atten Percep Psychophys*. 2011; 73:2270–2285.
- Klatte M, Hellbrück J, Seidel J, et al. Effects of classroom acoustics on performance and well-being in elementary school children: A field study. *Environ Beh*. 2010; 42:659–692.
- Klatte M, Lachmann T, Meis M. Effects of noise and reverberation on speech perception and listening comprehension of children and adults in a classroom-like setting. *Noise Health*. 2010; 12(49):270–282. [PubMed: 20871182]
- Krahmer E, Swerts M. How children and adults produce and perceive uncertainty in audiovisual speech. *Lang Speech*. 2005; 48:29–53. [PubMed: 16161471]
- Lalonde K, Holt RF. Audiovisual speech perception development at varying levels of perceptual processing. *J Acoust Soc Amer*. 2016; 139:1713–1723. [PubMed: 27106318]
- Lansing CR, McConkie GW. Attention to facial regions in segmental and prosodic visual speech perception tasks. *J Speech Lang Hear Res*. 1999; 42:526. [PubMed: 10391620]
- Leibold L, Buss E. Children's identification of consonants in a speech-shaped noise or a two-talker masker. *J Speech Lang Hear Res*. 2013; 56:1144–1155. [PubMed: 23785181]

- Lewis D, Manninen C, Valente DL, et al. Children's understanding of instructions presented in noise and reverberation. *Amer J Audiol*. 2014; 23:236–336.
- Lewis D, Valente DL, Spalding J. Effect of minimal/mild hearing loss on children's speech understanding in a simulated classroom. *Ear Hear*. 2015; 36:136–144. [PubMed: 25170780]
- Lewkowicz DJ, Hansen-Tift AM. Infants deploy selective attention to the mouth of a talking face when learning speech. *Proceeding of the National Academy of Sciences*. 2012; 109:1431–1436. <http://doi.org/10.1073/pnas.1114783109>.
- Maidment DW, Kang HJ, Stewart HJ, et al. Audiovisual integration in children listening to spectrally degraded speech. *J Speech Lang Hear Res*. 2015; 58:61–68. [PubMed: 25203539]
- McGurk H, MacDonald JW. Hearing lips and seeing voices. *Nature*. 1976; 264:746–748. [PubMed: 1012311]
- Merigan WH, Maunsell JHR. How parallel are the primate visual pathways. *Ann Rev Neurosci*. 1993; 16:369–402. [PubMed: 8460898]
- Mishra S, Lunner T, Stenfelt S, et al. Visual information can hinder working memory processing of speech. *J Speech Lang Hear Res*. 2013; 56:1120–1132. [PubMed: 23785180]
- Munhall K, Jones J, Callan D, et al. Visual prosody and speech intelligibility. *Psychol Sci*. 2004; 15:123–137.
- Munhall KG, Kroos C, Jozan G, et al. Spatial frequency requirements for audiovisual speech perception. *Percep Psychophys*. 2004; 66:574–583.
- Newton V. Sound localization in children with a severe unilateral hearing loss. *Audiology*. 1983; 22:189–198. [PubMed: 6847534]
- NIDCD. Statistical Report: Prevalence of hearing loss in U.S. children, 2005. 2006. Dec. Retrieved September 24, 2015, from <http://www.nidcd.nih.gov/funding/programs/hb/outcomes/Pages/report.aspx>
- Niskar A, Kieszak S, Holmes A, et al. Prevalence of hearing loss among children 6 to 19 years of age. *J Amer Med Assoc*. 1998; 279:1071–1075.
- Paré M, Richler RC, ten Hove M, et al. Gaze behavior in audiovisual speech perception: The influence of ocular fixations on the McGurk effect. *Percep Psychophys*. 2003; 65:553–567.
- Picard M, Bradley JS. Revisiting speech interference in classrooms. *Audiology*. 2001; 40(5):221–244. [PubMed: 11688542]
- Ricketts T, Galster J. Head angle and elevation in classroom environments: implications for amplification. *J Speech Lang Hear Res*. 2008; 51:516–525. [PubMed: 18367693]
- Ross L, Molholm S, Blanco D, et al. The development of multisensory speech perception continues into the late childhood years. *Eur J Neurosci*. 2011; 33:2329–2337. [PubMed: 21615556]
- Ross L, Saint-Amour D, Leavitt V, et al. Do you see what I am saying? Exploring visual enhancement of speech comprehension in noisy environments. *Cerebral Cortex*. 2007; 17:1147–1153. [PubMed: 16785256]
- Sandgren O, Andersson R, van de Weijer J, et al. Impact of cognitive and linguistic ability on gaze behavior in children with hearing impairment. *Front Psychol*. 2013; 4:1–8. DOI: 10.3389/fpsyg.2013.00856 [PubMed: 23382719]
- Sandgren O, Andersson R, van de Weijer J, et al. Coordination of gaze and speech in communication between children with hearing impairment and normal-hearing peers. *J Speech Lang Hear Res*. 2014; 57:942–951. [PubMed: 24167237]
- Scarborough R, Keating P, Mattys S, et al. Optical phonetics and visual perception of lexical and phrasal stress in English. *Lang Speech*. 2009; 51:135–175.
- Sekiyama K, Burnham D. Impact of language on development of auditory-visual speech perception. *Develop Sci*. 2008; 11:306–320.
- Shapley R. Visual sensitivity and parallel retinocortical channels. *Annual Rev Psychol*. 1990; 41:635–658. [PubMed: 2407178]
- Shelton B, Searle C. The influence of vision on the absolute identification of sound-source position. *Percep Psychophys*. 1980; 28:589–596.

- Shield BM, Dockrell JE. The effects of environmental and classroom noise on the academic attainments of primary school children. *J Acoust Soc Amer*. 2008; 123(1):133–144. [PubMed: 18177145]
- Smith NA, Gibilisco CR, Meisinger RE, et al. Asymmetry in infants' selective attention to facial features during visual processing of infant-directed speech. *Front Psychol*. 2013; 4:1–8. <http://doi.org/10.3389/fpsyg.2013.00601/abstract>. [PubMed: 23382719]
- Sumby W, Pollack I. Visual contribution to speech intelligibility in noise. *J Acoust Soc Amer*. 1954; 26(2):212–215.
- Summerfield Q. Lipreading and audio-visual speech perception. *Philos Trans: Biolog Sci*. 1992; 335:71–78.
- Thomas MR, Jordan TR. Contributions of oral and extraoral facial movement to visual and audiovisual speech perception. *J Exper Psychol: Hum Percep Perform*. 2004; 30:873–888.
- Vatikiotis-Bateson E, Eigsti IM, Yano S, et al. Eye movement of perceivers during audiovisual speech perception. *Percept Psychophys*. 1998; 60:926–940.
- Wagner P, Malisz Z, Kopp S. Gesture and speech in interaction: An overview. *Speech Comm*. 2014; 57:209–232.
- Walker EA, Spratford M, Moeller MP, et al. Predictors of hearing aid use time in children with mild-to-severe hearing loss. *Lang Speech Hear Serv Schools*. 2013; 44:73–88.
- Wass SV, Smith TJ, Johnson MH. Parsing eye-tracking data of variable quality to provide accurate fixation duration estimates in infants and adults. *Behav Res Meth*. 2013; 45(1):229–250. <http://doi.org/10.3758/s13428-012-0245-6>.
- Wechsler, D. Wechsler Abbreviated Scale of Intelligence. San Antonio, TX: Pearson; 1999.
- Wegrzyn M, Vogt M, Kireclioglu B, et al. Mapping the emotional face. How individual face parts contribute to successful emotion recognition. *PLoS ONE*. 2017; 12:e0177239. <https://doi.org/10.1371/journal.pone.0177239>. [PubMed: 28493921]



Figure 1. Experimental set-up. Screens and loudspeakers are numbered 1–4 from left to right. Parental permission was obtained for use of images.

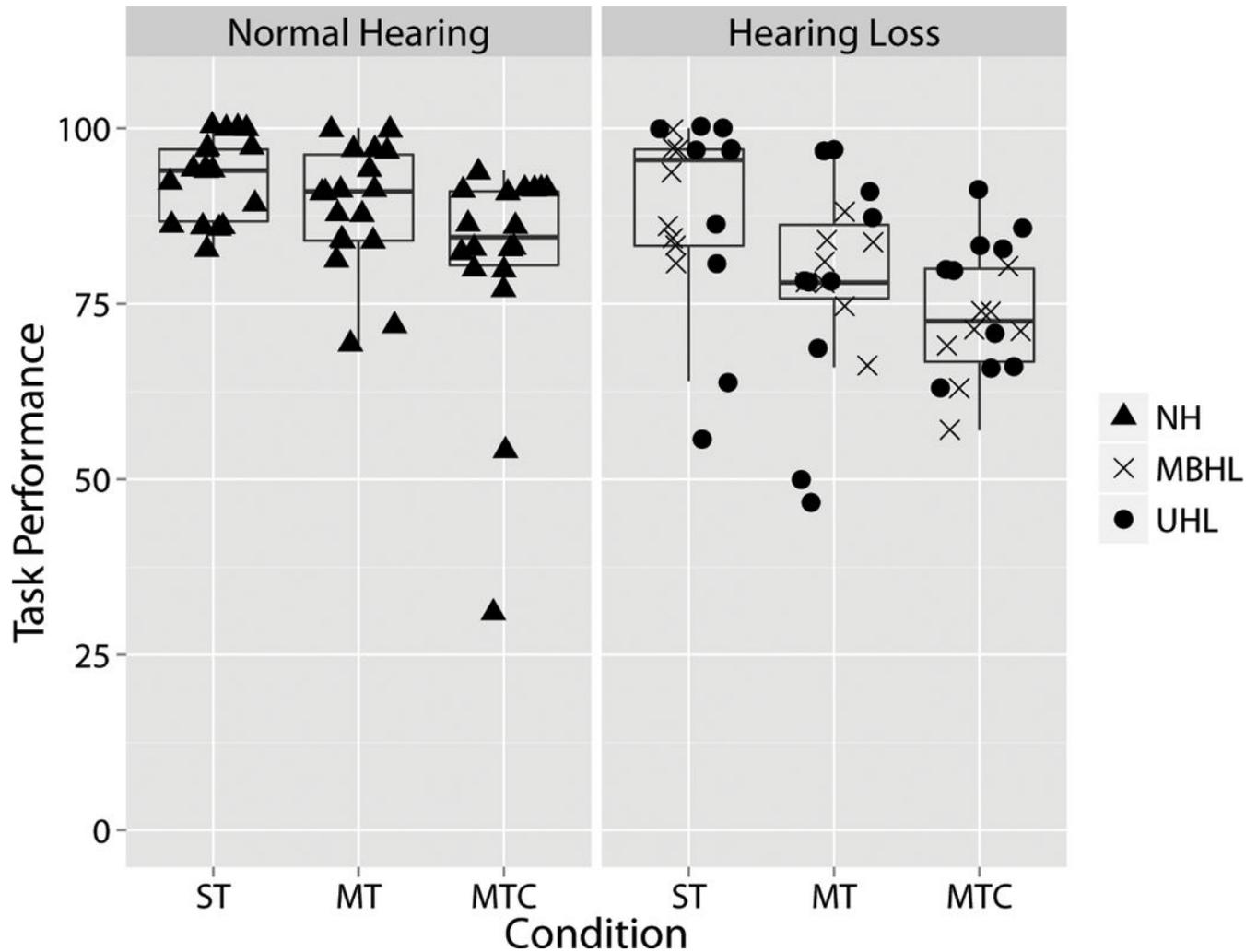


Figure 2.

Percent-correct performance across conditions (ST, MT, MTC) for the children with NH (left panel) and the children with MBHL/UHL (right panel). In this and subsequent figures, boxes represent the 25th to 75th percentiles, whiskers represent the 5th to 95th percentiles and lines within boxes represent medians. Individual data are shown for children with NH (filled triangles), MBHL (Xs) and UHL (filled circles). ST = single talker; MT = multi-talker; MTC = multi-talker with comments; NH = normal hearing; MBHL = mild bilateral hearing loss; UHL = unilateral hearing loss.

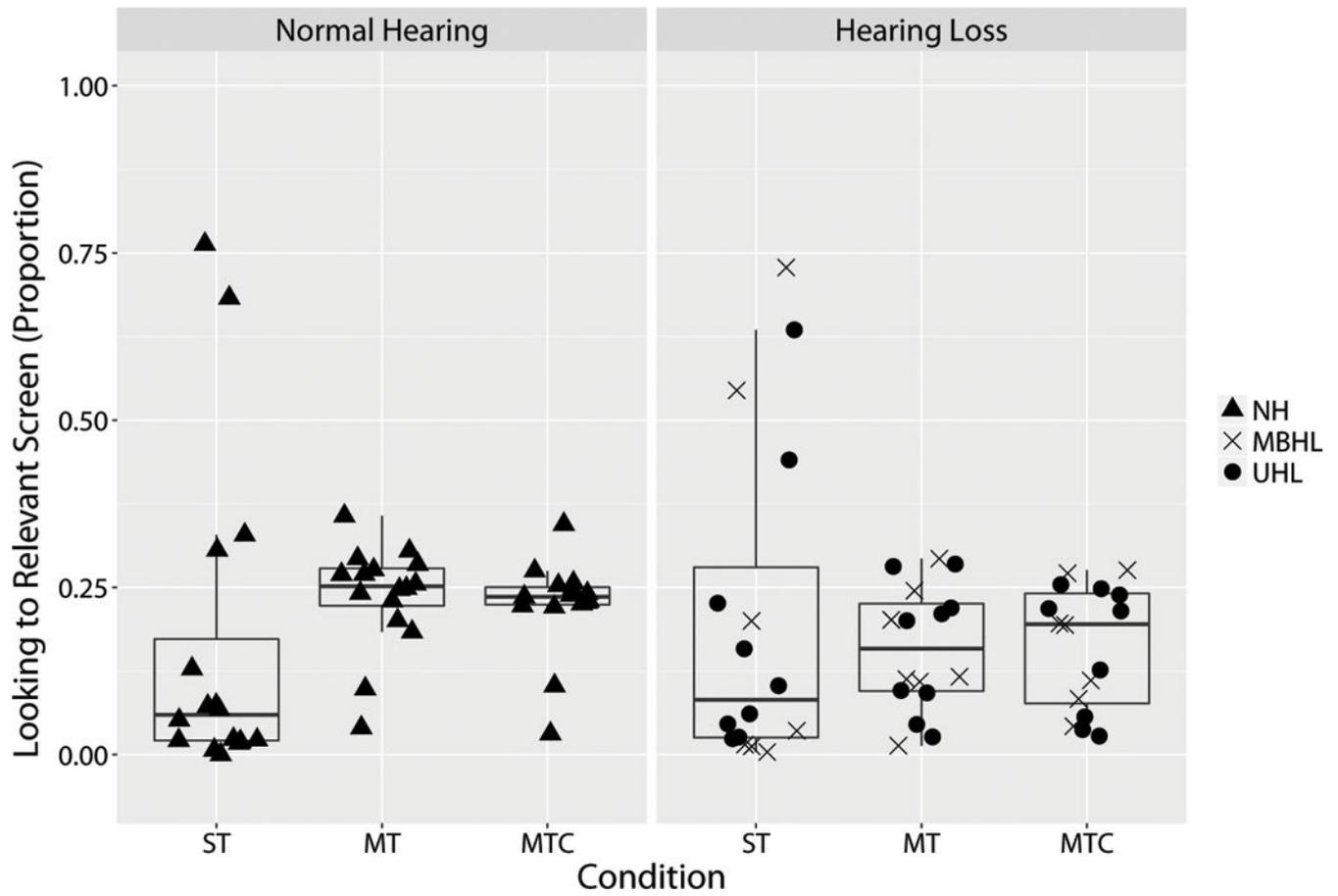


Figure 3.

Proportion of time spent looking at relevant screens as a function of condition (ST, MT, MTC) for children with NH (left panel) and children with MBHL/UHL (right panel). ST = single talker; MT = multi-talker; MTC = multi-talker with comments; NH = normal hearing; MBHL = mild bilateral hearing loss; UHL = unilateral hearing loss.

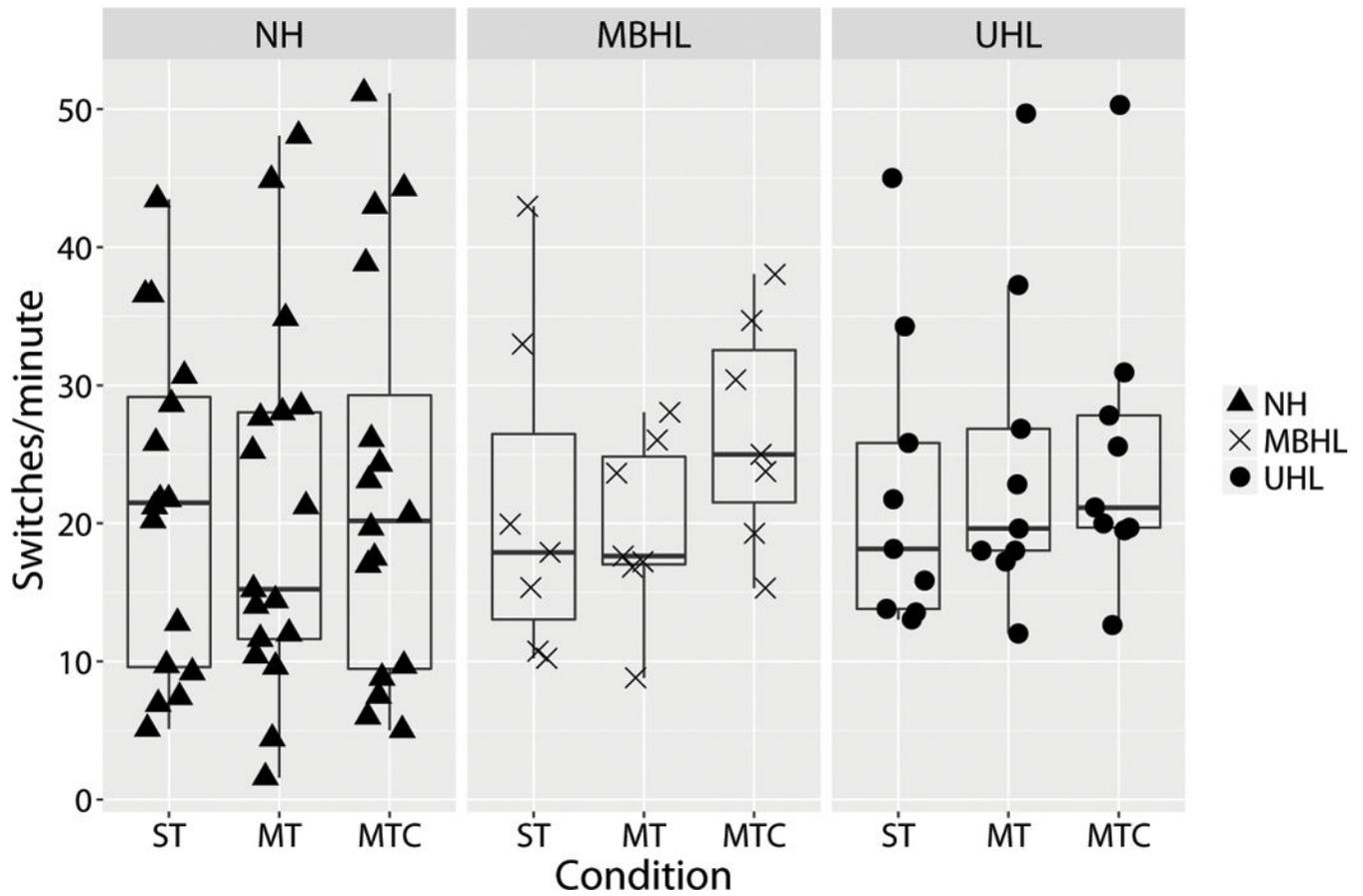


Figure 4. Switch rate (switches/minute) as a function of condition (ST, MT, MTC) for children with NH (left panel), children with MBHL (center panel) and children with UHL (right panel). ST = single talker; MT = multi-talker; MTC = multi-talker with comments; NH = normal hearing; MBHL = mild bilateral hearing loss; UHL = unilateral hearing loss.

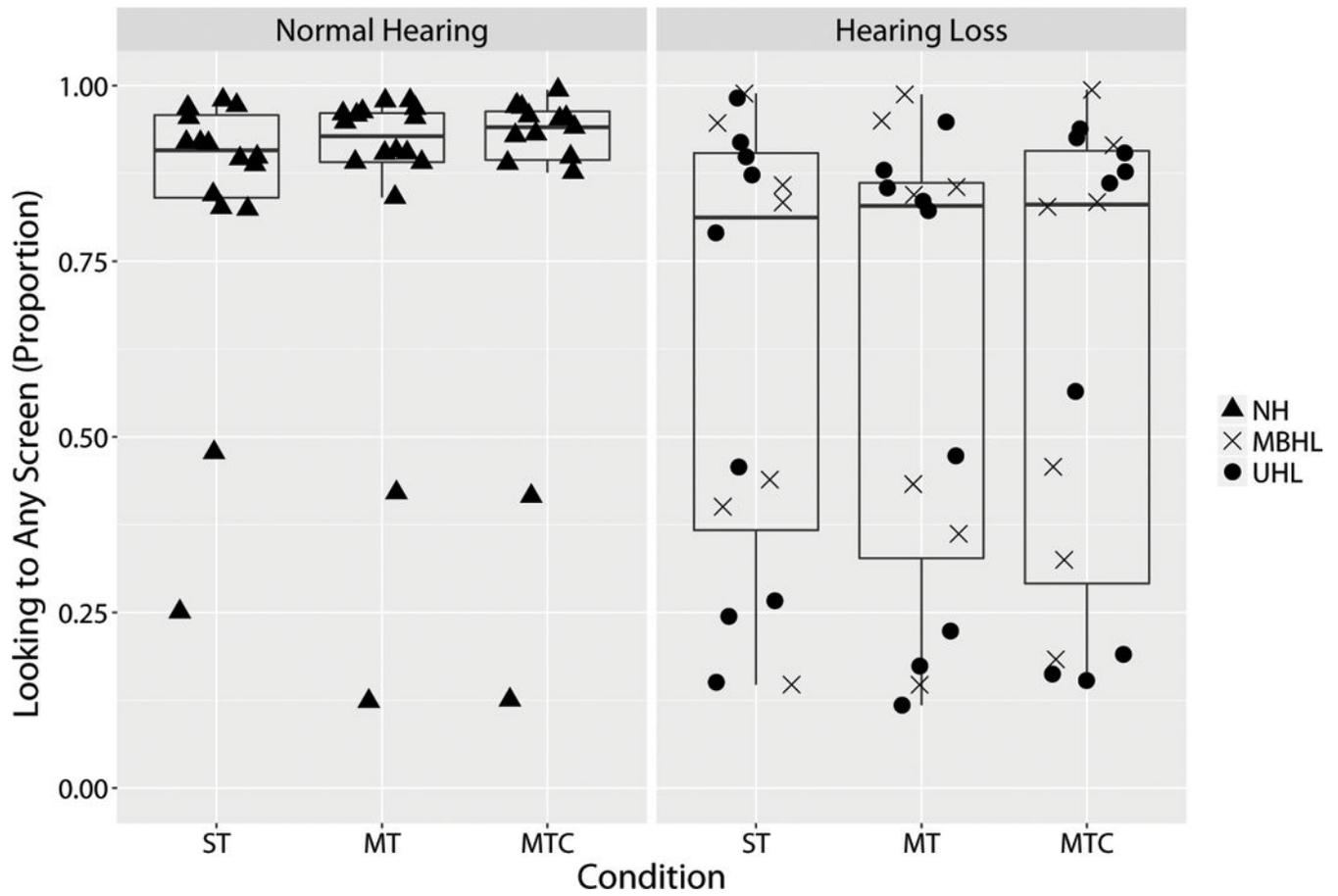


Figure 5. Proportion of time spent looking at any screen as a function of condition (ST, MT, MTC) for children with NH (left panel) and children with MBHL/UHL (right panel). ST = single talker; MT = multi-talker; MTC = multi-talker with comments; NH = normal hearing; MBHL = mild bilateral hearing loss; UHL = unilateral hearing loss.

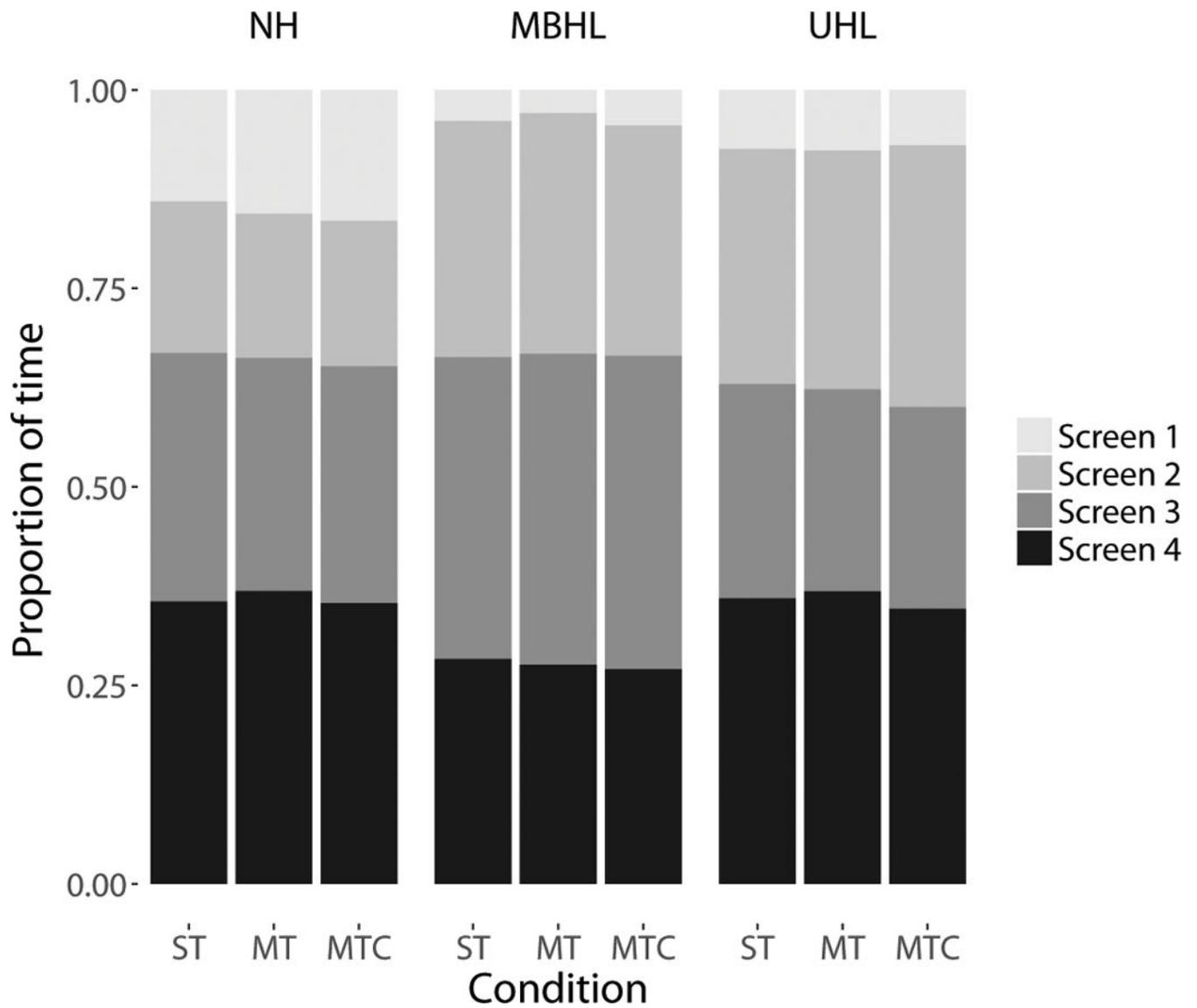


Figure 6.

Proportion of time spent looking at each of the four screens (varying shades of gray) out of the total time looking at any screen for children with NH (left panel), children with MBHL (center panel) and children with UHL (right panel). ST = single talker; MT = multi-talker; MTC = multi-talker with comments; NH = normal hearing; MBHL = mild bilateral hearing loss; UHL = unilateral hearing loss.

Table 1

Definition of hearing-status categories.

	NH	MBHL	UHL
N	18	8	10
Hearing-Status Group Definition	Thresholds ≤ 15 dB HL for octave frequencies 250–8000 Hz.	Three-frequency PTA ≤ 20 and ≤ 45 dB HL OR thresholds >25 dB HL at one or more frequencies above 2 kHz in both ears.	Three-frequency pure-tone average (PTA) >20 dB HL in the poorer ear and ≤ 15 dB HL in the better ear OR thresholds >25 dB HL at one or more frequencies above 2 kHz in the poorer ear.
PTA		For 7 children, mean better ear PTA = 34.2 dB HL (SD = 8.3 dB) and poorer ear PTA = 36.7 dB HL (SD = 7.1 dB). One child had a <u>high-frequency hearing loss</u> with better ear high-frequency average thresholds for 6–8 kHz (left ear) = 45 dB HL and poorer ear threshold at 8 kHz (right ear) = 50 dB HL.	For 8 children, poorer ear PTA = 55 dB HL (SD = 11.7 dB). Two participants had <u>high-frequency hearing loss</u> . For one, the high-frequency average thresholds for 3–8 kHz = 82.5 dB HL (right ear), and for the other average thresholds for 4–8 kHz = 88.3 dB HL (right ear).
Mean Age in yrs (Range)	10.61 (8.0–12.42)	10.74 (8.99–12.33)	10.60 (8.0–12.42)
Gender	8 male, 10 female	4 male, 4 female	5 male, 5 female

Table 2

Standard scores for the Wechsler Abbreviated Scale of Intelligence (WASI) and Peabody Picture Vocabulary Test (PPVT).

Test	Group	Mean	Std. Deviation
WASI	NH	104.7	8.0
	MBHL	97.6	7.2
	UHL	98.8	10.4
PPVT	NH	108.9	16.1
	MBHL	100.8	12.4
	UHL	108.11	15.3

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript