Published in final edited form as:
J Speech Lang Hear Res. 2011 December ; 54(6): 1644–1657. doi:10.1044/1092-4388(2011/10-0098).

Preliminary investigation of visual attention to human figures in photographs: Potential considerations for the design of aided AAC visual scene displays

Krista M. Wilkinson, Ph.D and **Janice Light, Ph.D**
Communication Sciences & Disorders, The Pennsylvania State University

Abstract

Purpose—Many individuals with complex communication needs may benefit from visual aided augmentative and alternative communication systems. In Visual Scene Displays (VSDs), language concepts are embedded into a photograph of a naturalistic event. Humans play a central role in communication development, and might be important elements in VSDs. However, many VSDs omit human figures. In this study, we sought to describe the distribution of visual attention to humans in naturalistic scenes as compared to other elements.

Method—Nineteen college students observed 8 photographs in which a human figure appeared near one or more items that might be expected to compete for visual attention (such as a Christmas tree, or a table loaded with food). Eye-tracking technology allowed precise recording of participants' gaze. The fixation duration over a 7-second viewing period and latency to view elements in the photograph were measured.

Results—Participants fixated on the human figures more rapidly and for longer than expected based on their size, regardless of the other elements in the scene.

Conclusions—Human figures attract attention in a photograph even when presented alongside other attractive distracters. Results suggest that humans may be a powerful means to attract visual attention to key elements in VSDs.

A substantial body of clinical practice in speech-language pathology involves the use of visual supports such as picture schedules, communication books or boards, or high technology speech-generating devices that offer voice output to improve communication. Together, these techniques are called aided augmentative and alternative communication (AAC; see Beukelman & Mirenda, 2005; or www.isaac-online.org). The effectiveness of aided AAC interventions in facilitating receptive and expressive communication outcomes in persons with communication disabilities has been demonstrated in numerous empirical studies (e.g., see Bondy & Frost, 2001; Bopp, Brown, & Mirenda, 2004; Harris & Reichle, 2004; Johnston & Reichle, 1993; Romski & Sevcik, 1996; see Mirenda, 2009, or Wilkinson & Hennig, 2007, for reviews).

The majority of aided AAC interventions rely on a visual channel to foster self-expression by individuals with complex communication needs and/or facilitate their comprehension of language input (e.g. Romski & Sevcik, 1996). It has been long recognized that effective oral/aural language interventions must take into consideration our knowledge about basic principles by which children process the auditory signal (e.g., ASHA, 2007). It seems equally likely that interventions using a visual modality can benefit from knowledge of basic

Contact Information: Krista M. Wilkinson, Communication Sciences & Disorders, 308 Ford Hall, The Pennsylvania State University, University Park, PA, 16802, voice: 814-863-2206, fax: 814-863-8759, kmw22@psu.edu.

principles of visual processing. Wilkinson and Jagaroo (2004) have proposed that aligning aided AAC displays with these known principles of human visual processing should serve to reduce the perceptual and processing demands imposed by aided AAC displays and thereby facilitate functional communicative outcomes. Yet we know very little about how visual information presented on aided AAC displays is attended, perceived, or processed by users.

One common format for aided AAC display design is the Visual Scene Display (VSD). VSDs are displays in which the language concepts (symbols) are embedded directly into a photograph or scanned image of a naturalistic event. Consider Figure 1, which illustrates a simple scene in which a young child is playing with a telephone with her mother. This VSD represents a game of playing telephone, into which language concepts are embedded by the programmer and accessed by the users of the VSD. For instance, the phone might be programmed to speak the word “telephone” and a sound effect of a ringing phone when selected, while selection of the area over the mother’s face would produce the spoken word “mommy” and selection of the child might produce the spoken word “hi”, etc. Empirical support is emerging for the effectiveness of such VSDs for communication with beginning communicators. Drager and her colleagues (Drager, Light, Curran-Speltz, Fallon, & Jeffries, 2003) found that typically-developing toddlers performed more accurately with VSDs than with traditional grid displays, and Light and Drager (2008) have reported that infants with complex communication needs used such VSDs to participate within social interactions before they were a year old.

In articulating the rationale for using VSDs, Drager and her colleagues (Drager et al., 2003; Light et al., 2004) noted that early language learning occurs within the context of a rich, event-based and experiential context. Children learn about the word “dog” and its referent not from hearing the word in isolation, but from hearing it in a variety of experiential contexts, which are unified by the presence of the referent and its label. Thus children learn about and hear the label for dogs as they see dogs in the park, pat them at their relatives’ houses, get kisses from them, and so forth. These predictable routines, described by Nelson (1986) as event schema, may be critical in facilitating language development because they provide contextual support for acquisition of new concepts/words. As Drager, Light, and colleagues have argued (2003, 2004), it seems reasonable to consider whether event-based representations – that is, VSDs featuring events like a child receiving a kiss from a dog, while mom looks on – might also be effective in aiding symbolic development in beginning aided AAC language learners.

Key elements of the event schemas that support the language development of young children are the individuals depicted within them (in Figure 1, the child and mother). People – mothers, fathers, smiling relatives – are central to basic social interactions involving infants and toddlers. Early social transactional routines (such as mutual smiling games) form the basis for development of communicative intentions (Bruner, 1983). One key function of prelinguistic intentional communication is social interaction, that is, communication produced purely to maintain an interaction between the child and the partner (Wetherby & Prizant, 1993).

In addition to the important symbolic and social role played by humans, animate figures, particularly humans, are also a key attractor of visual attention. Very young infants are drawn to examine human faces, especially the eyes (Hopkins, 1980; Buswell, 1935), not just when these stimuli appear in isolation but even when they are presented within complex arrays containing multiple objects (Gliga, Elsabbagh, Andravizou, & Johnson, 2009). Animate figures also appear to be key to visual processing of natural scenes by older children and adults. Fletcher-Watson and colleagues (Fletcher-Watson, Findlay, Leekam, & Benson, 2008) used eye-tracking technology to record point of gaze while participants

viewed split-screen presentation of two photographs, in which one of the photographs contained a human figure and the other did not. Scenes containing humans attracted participants' visual attention more rapidly and for longer than scenes that did not. Using similar technology, Smilek, Birmingham, Cameron, Bischof, and Kingstone (2006) presented viewers with single photographs that included a human, and found that viewers spent more time examining the human than other elements of the photograph, with most attention focused on the face and head region. In addition to these studies of nondisabled individuals, a recent body of research has begun to describe patterns of visual attention to humans in individuals with disabilities, including autism and Williams syndrome (e.g., Fletcher-Watson, Leekam, Benson, Frank, & Findlay, 2009; Klin, Jones, Schultz, Volkmar, & Cohen, 2002; Riby & Hancock, 2008, 2009a, 2009b; see Ames & Fletcher-Watson, 2010, for an extended review of the research related to autism). Specific implications relevant to our particular work and goals are considered in the discussion section.

While emerging evidence supports the use of VSD displays with beginning communicators, we know very little, empirically, about what kinds of elements contribute to displays that attract attention, are easy to process visually, readily understood, and used effectively for functional communication. In clinical practice, humans or social routines have often been overlooked in aided AAC design. For instance, the content of AAC interventions for beginning communicators typically has focused on snack and other simple needs and wants routines, as reflected in the focus of many approaches at initiation of intervention (e.g., Bondy & Frost, 2001) as well as the absence until recently of direct research on social closeness or joint attention functions (e.g., Light, Parsons, & Drager, 2002; see Wilkinson & Reichle, 2009). In these types of routines, the focus is on preferred objects, rather than people; as a result AAC displays for those routines have primarily included representations of inanimate objects as vocabulary items (e.g., cookie, juice) but have not typically incorporated either human elements or social communication functions. Furthermore, in the recently introduced VSDs being made available from some of the assistive technology manufacturers, the VSDs often represent backdrops of places (the kitchen, the living room, school) but contain few humans or social activities. Given the critical role that human figures play in early communication development, and the possible attraction of human figures in visual processing, it seems necessary to consider how human figures are attended to in potential VSDs.

In this study, our overarching goal was to describe the naturally occurring distribution of attention within scenes in which a human was present but not prominent. Smilek and colleagues (2006) argued that a careful delineation of naturally-occurring behavior allows measurement of ecologically-relevant attentional patterns, a suggestion echoed by Ames and Fletcher-Watson, who noted that "such methods have much greater ecological validity than most attentional paradigms and can tell us about how the attention… may be distributed in the real-world" (2010, p. 61). We therefore chose to track spontaneous patterns of visual attention during viewing in which no explicit task instructions were given. This approach is widely used in studies with infants through adults (Fletcher-Watson et al., 2008; Gliga et al., 2009; Smilek et al., 2006) as well as with individuals with disabilities (Fletcher-Watson et al., 2009; Riby & Hancock, 2008, 2009a, 2009b). The viewing patterns were recorded through eye-tracking technology similar to that reported in these other studies.

Our specific aim was to describe the extent to which human figures capture visual attention when they are not centrally prominent, including when they are small and/or offset from the center of the photograph, presented alongside stimuli that are vibrantly colored, near items that are complex and visually interesting, or with items that might be prominent in some other way. The reason for this particular focus is that events captured in VSDs often involve multiple elements, some of which are relevant and some of which are not. Opening of gifts

at a holiday celebration, for instance, would likely involve not just the child, the family, and the gift, but likely include at least some of the trappings of the holiday celebration (a Christmas tree, a menorah, birthday decorations) as well as non-relevant items like pictures on the wall behind the participants in the event. It is important to consider the extent to which viewers are able to see/fixate on the central figure(s) in the presence of these competing elements.

Because the study was an exploratory examination of how well small/offset human figures in photographs attracted and maintained visual attention when presented alongside more prominent items, and under conditions of free viewing, all but one of our analyses reflect descriptive rather than inferential analyses. We evaluated (1) how many of our 19 participants fixated at all on a given element over the 7-second viewing period, as a measure of how likely an element was to attract attention; (2) the mean total time spent by our 19 participants on each element, as a measure of how well the element maintained attention ("attention holding" elements, in the words of Gliga et al., 2009), and (3) the mean latency to first fixation on each element by our 19 participants, as a measure of the speed with which elements attracted attention ("attention grabbing" elements; Gliga et al., 2009). The results set the stage for experimental manipulations in future research.

Method

Participants

Nineteen participants between the ages of 18 and 22 years enrolled in the study. Participants were recruited through flyers on a large university campus. All participants were college students with normal vision or vision corrected to within normal limits according to self report. Four participants were male and 15 were female. The research was approved through the Human Participants Institutional Review Board of the Pennsylvania State University.

Higginbotham (1995) pointed out that gathering information on processes in individuals who do not have complicating sensory, social, emotional, intellectual, or other disabilities is an important step in many types of AAC research. In the case of this particular research, it was essential to initiate study with adult participants with no disabilities because so little is yet known about allocation of visual attention to humans in scenes even in the cognitive sciences. What research is available (Fletcher-Watson et al., 2008; Smilek et al., 2006) was not designed to ask questions of specific relevance to constructing displays such as those used in AAC. Thus, it was necessary to begin by mapping out basic patterns of visual attention in nondisabled participants as a precursor to considering the potential impact of development (how and when do these patterns become established?) and disability (in what ways do various disabilities alter these expected patterns, if at all?). Without this baseline, we would be unable to determine whether viewing patterns vary across ages, whether they differ in individuals with disabilities, or whether differences in viewing patterns across disability categories reflect etiology-specific influences on visual processing. By this same logic, we considered it critical to describe visual attention in a laboratory setting prior to extending the work to naturalistic settings, in order to map out the fundamental influence of visual dimensions under controlled conditions prior to extending the work to naturalistic communication contexts. Clearly, however, further research with specific populations that vary in age and disability status and within more naturalistic contexts will be necessary to delineate the similarities and differences within and across populations and in different settings; we consider both of these issues further in the discussion.

Materials

To address our research questions, we examined the viewing patterns of 19 participants for each of 8 photographs. The use of eight different photographs allowed us to examine the extent to which human figures attract attention, through systematic replications across photographs that included a diverse set of potential competitors for attention. This approach is similar to that of Smilek et al (2006), in which eye gaze patterns were examined in detail across two photographs that varied the context in which humans appeared.

Eight photographs were selected from personal photograph albums of various members of the research group and from a picture dictionary available for this type of research (Lang, Bradley, & Cuthbert, 2008). The 8 photographs reported here were intermixed with 31 other photographs, for a total of 39 items viewed during each data collection session. The other photographs addressed research questions different from those reported here; for instance, one subset addressed a question of attention to a human figure as compared to an equally prominent familiar animal. Although reported separately, the data were collected as a group in part to ensure variety among the stimuli (e.g., so that participants weren't viewing the same type of stimuli over and over) and in part for efficiency of data collection.

The primary criterion for selecting our photographs was that the main human figure(s) did not occupy a substantial proportion of the photographic space; they occupied between 1.3% an 18.3% of the space. An independent viewer verified that the human figure(s) of interest was the subject of each photograph. This viewer was asked to look at each photograph and write down what she thought the subject of the photograph was; in all cases this viewer identified the same central human subject identified by the experimenters. For this paper, we refer to this element as the "human figure" (in some cases, there were other people in the background or periphery; these people will be referred to as "bystanders"). In most photographs the human figures were far smaller than many other elements. A second criterion was that the photographs contain at least one other element that, based on subjective judgment, might compete for visual attention. Figure 2 presents all 8 photographs. As Figure 2 illustrates, in each of the photographs the human figures are either small, offset, and/or presented alongside other prominent elements.

Data acquisition and eye tracking equipment

We used eye-tracking technology similar to that reported in a number of studies of visual attention reviewed in the introduction; such technology has also been used in recent studies of language processing in aphasia, including semantic priming (Odekar, Hallowell, Kruse, Moates, & Lee, 2009), and studies of methods for assessment of comprehension (Heuer & Hallowell, 2007). Data were collected through the synchronization of two computers, a 20-inch iMac© that was responsible for displaying the visual stimuli to the participant and an adapted PC that was specially equipped with the ISCAN® ETL-300 Binocular Free-head Eye-tracking laboratory system for data acquisition. The experimenter advanced the presentation slide show on the iMac© and triggered the onset and offset of data acquisition on the data acquisition PC so that the onset/offset of stimulus presentation and data acquisition were synchronized.

Visual stimuli were presented to participants on the 20-inch iMac© computer monitor within a Powerpoint© presentation set on "slideshow" display, so that the photograph took up the entire display. The photographs were each displayed for approximately 7 seconds with a 3-second inter-trial interval between them. The order of presentation of the photographs was randomized prior to onset of study and presented in this fixed order to all participants. In the inter-trial interval, a white screen was presented that had a red dot in the upper center of the display. Participants were asked to look at the red dot during the 3-

second interval. This maximized the likelihood that the point of fixation was the same at the outset of display. Participants rested their chins on a chinrest approximately 2 feet away from the screen.

During each photograph presentation, the exact point of gaze was recorded on the PC computer set up with the ISCAN® ETL-300 Binocular Free-head Eye tracking system. This system uses a remote infrared camera to detect light reflected off of participant's pupils and corneas to determine point of visual gaze; this camera was mounted just below the iMac stimulus presentation monitor and had no direct contact with the participant. The dual points of reflected light entered into the ISCAN® Raw Eye Movement Data Acquisition (DQW) software as a series of X-Y coordinates that provided a running record of exactly where gaze was directed at all times, in chronological order. Fixation was defined as a gaze duration that exceeded 40 milliseconds. When that fixation occurred within the set of X-Y coordinates that made up a coded element (region of interest), the time, length, and location of that fixation was calculated (a fixation to an X-Y coordinate not defined into a coded element was recorded the same way, as a fixation to "no element"). We retained this fairly short fixation threshold in order to maximize sensitivity to views to even the smallest coded elements, given that very small human figures were the specific focus of interest in this research. The exploratory nature of this analysis builds the framework for later experimental manipulations that might involve a longer threshold.

Calibration between the camera and the software program was conducted at the outset of each session, to ensure accurate data acquisition. During calibration, participants were instructed to look at a series of 5 dots on the screen, one in the center and four placed along the corners of the screen. Calibration generally took between 15 and 30 seconds, and no more than 1–2 minutes to achieve. Although the remote camera accommodates small head movements during data acquisition after calibration, participants used the chinrest during the session, thus limiting head movements. After calibration, participants were told "You are going to see a series of pictures; just look at them as you would normally." After the session, participants were asked whether they felt self-conscious about the setup; none reported discomfort.

Measures for Data Analysis

To prepare the photographs for analysis, areas of interest (elements) were identified and mapped onto each stimulus photograph through a drawing program provided on the ISCAN® PRZ software. Identification of the areas of interest was conducted by the research team prior to data collection/analysis. This element coding involved consensus coding conducted in two steps. First, the two research assistants worked together to create the file that contained all identified elements. The research assistants used the following decision-making guides about what to enclose within each area of interest, or element: (a) *composition*, such that objects were coded in their entirety rather than being separated into their parts (for instance, the entire Christmas tree was considered a single element), (b) *category membership*, such that elements of shared category membership were placed together (for instance, the dog and the cat in *child with Xmas tree* were considered a single element, being animals, while another element consisted of all of the pictures on the wall), and (c) *representation as a distracter of interest*, as described in the Materials section. In this last category, anything that represented a distracter of interest to the research could be separated even if it overrode composition or category membership; thus, for instance, the luminant patch of sunlight was coded separately from the rest of the floor because we were interested in the extent to which such a luminant spot might capture attention. After the research assistants had created the areas of interest, the first author reviewed the files. Thus, the final product reflected the consensus among three viewers, given the identified guidelines.

During element coding, the "drawing" function in the eyetracking software (similar to the Powerpoint™ drawing tool) was used to enclose each defined area of interest. In all photographs, there were some interstices between the defined elements that were not enclosed, in order to ensure that the defined areas did not overlap. The mean amount of each image enclosed within a coded element was 85% (median = 86%). To illustrate how the uncoded space was distributed, Figure 3 presents the enclosed areas for two images, *family at statue* and *man with fountain*, which had 83.3% and 84.2% of the space enclosed in coded elements, respectively.

Dependent measures

The software matched the running record of X-Y coordinates of the eye movements supplied from the infrared eye-tracking camera onto the locations enclosed within the areas of interest, as described above. The ISCAN PRZ software provided summary data concerning all fixations that occurred. Specifically the software provided a listing of element fixations to each element/area of interest, in chronological order of occurrence and with a time stamp. The software also calculated the percentage of the viewing time within each area of interest and the number of participants who fixated on each element. The space occupied by each element was also calculated by the software.

We were interested in the distribution of visual attention to elements in the photographs under spontaneous viewing conditions. If, as expected attention, was unequally divided among elements, some elements would be expected to receive fixations from only some participants, or for only brief periods of time. Thus, it was necessary to evaluate not only direct measures of fixation (how long, how rapid fixations were) but also to measure the number/percentage of participants who noticed each element within the 7-second viewing period. We therefore used the following measures: (1) the number of participants (n = 19) who fixated at all on a given element over the 7-second viewing period, as a measure of how likely an element was to attract attention; (2) the mean total time spent by the 19 participants on each element, as a measure of how well the element maintained attention ("attention holding" elements, in the words of Gliga et al., 2009), and (3) the mean latency to first fixation on each element by the 19 participants, as a measure of the speed with which elements attracted attention ("attention grabbing" elements; Gliga et al., 2009).

Our second measure, the total percentage of time each participant spent viewing each element across the 7-second period, was evaluated in two different ways. First, the simple total percent of time was calculated as a gross measure of attention. However, in addition to this an adjustment was also made because of the natural differences in the sizes of our elements. As Fletcher-Watson et al (2008) pointed out, an element that occupies 10% of the space but which is viewed for 20% of the time is receiving more attention than would be expected, based on its size. Similarly, an element that also is viewed for 20% of the time but which occupies 40% of the space is receiving less attention than would be expected, based on its size. We therefore also used Fletcher-Watson et al.'s (2008) method for adjusting for the relative sizes of elements. Specifically, we calculated the ratio of the time spent on an element compared to the space occupied by that element in the photograph. Ratios over 1 reflect greater looking time than would be expected based on the size of the element; ratios under 1 reflect less looking time than would be expected. This ratio was calculated for each participant's viewing patterns.

Results

Results for all analyses are presented, separately for each photograph, in Table 1.

Number of participants fixating on each element (attraction of attention)

The human figure of each photograph attracted fixations from a majority of participants (over 70%) in all 8 photographs, and from all 19 participants in 5 of the 8 photographs. The photograph with the lowest proportion of participants fixating on the human figure was the photograph of *family at statue* (14 of 19, or 74% of participants); in this photograph the human figures were also the smallest of all of the photographs, occupying only 1.3% of the space and positioned under the centrally located statue, which occupied 9% of the space.

In all photographs, some of the other elements attracted fixation from a majority of participants while others did not. The elements that did not attract attention from a majority of participants fell into the category of background elements, for instance, the stop sign, van, flag, and flowers (*family at statue*), the pictures, cabinets, and ceiling (*child with Xmas tree*), the mountain and grass (*children in garden*), sky and trash cans (*family at China tower*), trees, sand, and sky (*women with pillars*), ground and trash cans (*man at fountain*), cars and pavement (*children at fountain*), and trees, cars, baby stroller, and grocery bag (*women at table*). The elements that did attract attention from a majority of participants were primarily ones that were prominent and/or centrally located elements (the China tower, the pillars, the Christmas tree, the fountains). The other types of element that attracted fixations from a majority of viewers were the animate figures that were not the subjects of the photograph (the animals in the *child with Xmas tree,* and the bystanders in the photographs that contained them).

This pattern indicates that while the visual attention of most participants was directed to the human figure(s), it was not directed exclusively so; other elements did indeed attract fixations. The pattern also indicates, however, that as expected, attention was not attracted equally by all elements in the photographs. Rather, elements that made up the background (sky, ceiling, etc) tended to be observed within the 7 second viewing time by only a minority of the 19 participants.

Time spent on each element (scrutiny/maintenance of attention)

The second measure evaluated how well each element maintained attention during viewing. As can be seen in Table 1, the total percent time spent fixating does not sum to 100%. This is because fixations represent only those periods in which the viewer's gaze dwelled beyond the defined fixation threshold. The remainder of the times would include whenever the viewer's eye spent time moving between fixations/dwells (saccades), times when the participant blinked, periods in which eyetracker was re-establishing the crosshair calibration (generally, after blinks), or fixations within the uncoded interstices. Particularly for subjects who blinked frequently or for those who scanned the photograph rapidly rather than fixating on fewer elements, the absolute proportion of time spent fixating would be expected to be lower.

Descriptive analysis—In terms of the overall percent of time participants spent on each element, the human figure attracted substantial attention in all of the 8 photographs, receiving the greatest or second greatest total attention (or being tied for either position) in all 8 of the photographs. In all the photographs in which the human attracted the second greatest attention, the element that surpassed the human figure was the large, prominent, centrally located element (the statue in the *family at statue*, the Christmas tree in *child with Xmas tree,* the fountain in *man with fountain,* and the tower in the *family at China tower)*.

As described earlier, we also calculated a ratio of the time spent on each element relative to the space it occupied in the photograph to determine whether the time spent on the human figure was greater than expected based on size. Ratios over 1 reflect greater time spent than

would be predicted simply on the basis of the element size while ratios under 1 reflect less time spent than would be predicted based on size. Descriptively, the human figures attracted more attention than expected based on their size in 7 of the 8 photographs.

In all photographs but *women with pillars*, at least one other element attracted greater attention than would be expected based on size. These elements included the statue figures and the medallion of George Washington's head (*family at statue*), the dog and cat and the Christmas tree (*child with Xmas tree*), the fence and colorful tree (*children in garden*), the water fountain (*man at fountain, children at fountain*), and the items on the table in *women at table*. In all photographs, there were elements that were not examined for longer than would be expected based on their size. Some of these elements caught the attention of more than 50% of participants but did not maintain that attention. For instance, the pillars in *women with pillars* were received fixations from all 19 participants, but this attention was maintained for less time than would be predicted based on the size of the pillars (ratio = .37). More commonly, elements neither attracted nor maintained our participants' attention; thus, the sand, and sky of *women with pillars* were examined only by a small number of participants and for less time than expected. In most instances, these were background items like trees, sky, and grass. Similarly, in all photographs, some coded elements caught the attention of only a minority of participants, but were scrutinized by those participants for longer than would be predicted based on size. For instance, the stop sign in *family at statue* received fixations from 7 of the 19 participants, but was viewed by those participants for over three times longer than would be expected based on size. Likewise, only nine participants fixated on the water bottle being thrown by the man in *man at fountain*, but the ratio of time spent viewing it to its size was 6.67.

Inferential analysis: Attention to human figures—The ratio of time spent to size was greater than 1 for the human figure in 7 of the 8 photographs. One-tailed t-tests evaluated whether the calculated ratio was significantly different from 1, for each picture. The p-value was initially set to .05, then divided by 8 (the number of photographs) to adjust for multiple comparisons; thus, the final p-value criterion for judging whether a ratio was different from 1 was p = .006. The ratio was determined to be of statistical significance at this adjusted p value in five of the eight photographs; *family at statue,* $t(1,18) = 2.99$, observed p = .004;, *women with pillars*, $t(1,18) = 4.14$, observed p < .000; *child with Xmas tree,* $t(1,18) = 3.38$, observed p = .002; *children in garden,* $t(1,18) = 4.25$, observed p < .000; and *family at China tower,* $t(1,18) = 3.46$, observed p = .001. Although the ratios for *children at fountain* and *man at fountain* were both above 1, the differences were not of statistical significance; the ratio was below 1 but not significantly so for *women at table*. It is interesting to note, as a potential avenue for future research, that the human figures in these latter three photographs were actually those that occupied the most space (12%, 9%, and 16%, respectively).

Latency to first view

In addition to the time spent viewing each element, we also evaluated the amount of time that elapsed between the onset of the visual stimulus and the first fixation (i.e., latency to first view). Because attention was not evenly distributed across elements, some elements never received fixations. Rather than omit those elements from analysis, the latency for any cell with no fixation was set at the maximal length of each observation period.

Despite the fact that they occupied so little space, the human figures ranked as the element receiving the most rapid fixation in five of the eight photographs, as the second element in two photographs, and as the third element in one photograph. The mean latency to fixate on the human figure(s) was 1.9 seconds (range = 1.32 – 3.12). In the three photographs in which the human element did not receive the shortest latency to first fixation, the elements

that received earlier fixations were substantially larger and were situated in the location of the initial fixation point (the upper center of the photograph); the Christmas tree (*child with Xmas tree*; 14.8% of the space), the China tower (*family at China tower*; 23.8% of the space), and the statue figures and base (*family at statue*; 9 and 9.3% of the space, respectively).

DISCUSSION

We explored visual attention to human figures in a series of photographs in which human figures were quite small, offset from the center, and/or presented beside larger, prominent, and colorful objects. Even under these conditions, the human figures were among the earliest elements to attract attention, maintain the greatest attention across the viewing period, and to do so from the majority of participants. The speed and overall maintenance of attention capture is particularly compelling in light of the fact that the human elements occupied only between 1.3% and 18% of the space. These results suggest that attention was drawn to the human figures irrespective of whether they were centrally or peripherally located or posed near other items that might compete for visual attention. It appears that human figures attracted and maintained substantial visual attention irrespective of what else appeared in the photographed scene.

While human figures attracted attention, participants also distributed their attention to other elements in the scene. This attention was not distributed evenly across all the possible other elements that could have been examined, however. Rather, when another element competed for the first rank in any of our dependent measures, that element was generally a large, prominent, centrally located item (such as the tower, or the statue). Elements that served as "background" were less likely to attract or maintain attention.

Implications of findings for VSD construction

Our study suggests that humans may be key elements that capture and maintain visual attention, even in scenes with many other potential competitors. This finding is consistent with those of a recent study of infant eye gaze to faces when presented alongside a number of other elements (Gliga et al., 2009). We noted in the introduction that humans are critical in early social and communicative development. Babies from early on prefer to look at human faces over other stimuli, and early social transactional routines between infants and caregivers are established in the first year of life that form the foundation for later social, communicative, and language development (Bruner, 1983; Hopkins, 1980; Wetherby & Prizant, 1993). Our study underscored the powerful draw of human figures on visual attention across different and distinct contexts, even in adults and even in the presence of attractive and interesting distracters.

Although the human figures captured initial visual attention and sustained that attention over time, it is also critically important to note that they were not the only elements to garner visual attention. Within a span of 7 seconds, participants fixated on and examined other elements in the photographs. This finding indicates that inclusion of humans did not capture attention so exclusively that viewers failed to notice anything else. This finding is important because in order to benefit from VSDs such as the simple scene in Figure 1 for meaningful communication, the user needs to attend not only to the people in them but also to the social event being depicted (the phone). Importantly, however, background elements, by themselves, did not compete with the human figures or with these other types of elements for the visual attention of our non-disabled adult participants.

Limitations and future research

This analysis was just the first step in systematically delineating visual attention as it relates to the construction of optimal VSDs for aided AAC. Future research is required to investigate other variables that may impact visual attention and, ultimately, use of VSDs.

Participant group—Our participants were nondisabled college students. It is critical to examine the developmental trajectory of the findings we report. We felt it was necessary to begin with this population to determine whether indeed the expectations from visual cognitive neuroscience might apply to individuals with fully-developed language, cognitive, visual, and attentional processing. Having found that they did, it is now critical to examine how younger children view these same stimuli. Indeed, VSDs have been recommended and studied with very young children at the earliest communication levels (Light & Drager, 2008). We must therefore evaluate the effects of humans in scenes on visual attention in these populations.

It is also critical to examine these patterns in individuals with the kinds of developmental disabilities that often lead to the implementation of AAC. Eye gaze analyses have suggested that individuals with autism may show some alterations of gaze patterns when viewing stimuli that include humans. For instance, when human faces are presented in isolation, with little else to look at, older high functioning individuals with autism show fixations on the face, including the eyes and mouth, though reduced relative to nondisabled peers (Pelphrey, Sasson, Reznick, Paul, Goldman, & Piven, 2002). Klin et al (2002) examined viewing patterns of individuals with high functioning autism while they watched short black and white videos, and observed reduced fixations to social stimuli and increased viewing of objects relative to the number produced by participants without autism. Riby and Hancock (2008a, 2009) demonstrated similar reductions relative to individuals without autism during observation of both videos of humans as well as still photographs, by participants with autism as young as 6 years of age.

Despite these reductions, it is important to note that viewing of humans in individuals with autism is not wholly absent, and may be potentially mediated by the type of stimulus. van der Geest and colleagues (2002) reported on a study that tracked the eye gaze of children with autism and of nondisabled peers as they examined drawings (not photographs) of scenes that included a cartoon of a human figure. Both groups of children looked at the cartoon figure more quickly and for longer durations that the surrounding inanimate elements in the scenes. Furthermore, they found that the participants with autism looked at the cartoon figure of the human as often and for similar durations as the nondisabled children. Riby and Hancock (2009) examined view patterns of children between the ages of 6 and 18, as they viewed both humans as well as realistic cartoons (from Tintin) in either video or still presentation. While nondisabled individuals viewing humans in either format tended to focus primarily on the face (40% of view time) and less on body or background (between 20–40%), those with autism showed a reverse pattern, focusing on the background around 50% of the view time, the body approximately 20–30% of view time, with less attention to the face (approximately 20% of view time). However, when the presentation involved the Tintin cartoons, participants with autism showed greater face-oriented viewing, with both the face and the background viewed between 30% and 40% of the view time. Fletcher-Watson et al (2009) showed that while individuals with autism preferentially look at a photograph containing a human over one in which there is no human, when presented side-by-side, they show reductions relative to persons without autism in attention to the human face and do not follow the direction of gaze of the person in the photograph.

These data indicate that while viewing patterns of those with autism may be different from those of nondisabled individuals, there is still attention to humans, including to faces.

Having mapped out viewing patterns to humans in our study with nondisabled participants, it is now important to determine the extent to which attention to human figures occurs in individuals with autism, cerebral palsy, Down syndrome, and other disabilities when the human figures are less prominent. This would be a necessary first step before drawing any conclusions about the optimal role of humans in VSDs for children with complex communication needs.

Stimulus set—Our research did not control systematically the dimensions of the elements within photographic stimuli. Our findings offer some guidance and hypotheses for future studies that exert greater experimental control. For instance, the placement of items (central versus peripheral, foreground versus background) and relative sizes may be of interest to manipulate, to determine whether they exert influence on visual attention. In addition, altering the instructions to change the task from a spontaneous viewing to a directed search ("find the….") would allow analysis of how attention patterns change. This will be important to study as a user of a VSD will likely be searching for a desired target (the content of their intended message), and thus we must determine the extent to which the task itself exerts influence over gaze patterns.

Content—The critical role of the social event depicted in the photographed scenes was not examined in this research. VSDs are not intended simply to be backdrops of locations or to be portraits of people. Rather, to be consistent with Nelson's (1986) concept of event-based learning, VSDs must include the event being engaged in by the people. Our stimuli were chosen to evaluate the attention to small/nonprominent human figures, and thus were not necessarily optimal VSDs in this regard. Further research is needed to determine to what extent events depicted within VSDs attract visual attention, and how readily those events are interpreted.

Extension to functional communication—Finally, the extent to which this laboratory-based research represents viewing patterns that would occur under conditions of actual communication using VSDs is clearly a significant question. It was critical to initiate analysis in the laboratory in order to evaluate viewing patterns under controlled conditions. Without this framework, there would be little means to disentangle the relative contributions of the visual attention from the influences of the communication environment. It may now be possible to evaluate related questions in more naturalistic environments, with the goal of examining any potential alterations that such settings and communication tasks impose.

Summary

Our study adds to the growing evidence that humans attract and maintain attention in naturalistic scenes. This study described gaze patterns when the human figures were small and placed near objects that might be expected to compete for visual attention. Though the human figures did not dominate attention to the exclusion of these other elements, they certainly seemed to be key aspects to ensuring attention to the content of the images. Perhaps makers of VSDs might seek to exploit this attraction to maximize interest by the user and, in turn, facilitate communication. Through research that will allow a better understanding of factors that affect these variables, we may be able to improve design of AAC systems and enhance functional outcomes.

Acknowledgments

The research would not have been possible without the contributions of Kelly McStravock, Kaitlyn Fratantoni, and Jessie Miller, who helped with all aspects of data collection, coding, and analysis. Parts of this work were presented as a symposium at the 2009 conference of the American Speech-Language-Hearing Association. This research was supported in part through two grants: (1) the Communication Enhancement Rehabilitation Engineering Research

Center (AAC RERC), a virtual research center that is funded by the National Institute on Disability and Rehabilitation Research (NIDRR) under grant H133E030018, and (2) grant #P01 HD25995 from the National Institute of Child Health and Human Development (NICHD). The opinions contained in this publication are those of the grantees and do not necessarily reflect those of the granting agencies.

References

American Speech-Language-Hearing Association. Scope of Practice in Speech-Language Pathology. Scope of Practice. 2007. Available from www.asha.org/policy

Ames C, Fletcher-Watson S. A review of methods in the study of attention in autism. Developmental Review. 2010; 30:52–73.

Beukelman, DR.; Mirenda, P. Augmentative and Alternative Communication: Supporting children and adults with complex communication needs. 2. Baltimore, MD: Paul H. Brookes; 2005.

Bondy A, Frost L. The picture exchange communication system. Behavior Modification. 2001; 25:725–745. [PubMed: 11573337]

Bopp KD, Brown KE, Mirenda P. Speech-language pathologists' roles in the delivery of positive behavior support for individuals with developmental disabilities. American Journal of Speech-Language Pathology. 2004; 13:5–19. [PubMed: 15101810]

Bruner, J. Child's talk: Learning to use language. New York: Norton; 1983.

Buswell, GT. How people look at pictures: A study of the psychology of perception in art. Chicago: The University of Chicago Press; 1935.

Drager K, Light J, Curran-Speltz J, Fallon K, Jeffries L. The performance of typically developing 2 ½-year-olds on dynamic display AAC technologies with different system layouts and language organizations. Journal of Speech, Language and Hearing Research. 2003; 46:298–312.

Fletcher-Watson S, Findlay JM, Leekam SR, Benson V. Rapid detection of person information in a naturalistic scene. Perception. 2008:571–583. [PubMed: 18546664]

Fletcher-Watson S, Leekam SR, Benson V, Frank MC, Findlay JM. Eye-movements reveal attention to social information in autism spectrum disorder. Neuropsychologia. 2009; 47:248–257. [PubMed: 18706434]

Gliga T, Elsabbagh M, Andravizou A, Johnson M. Faces attract infants' attention in complex displays. Infancy. 2009; 14:550–562.

Harris MD, Reichle J. The impact of aided language stimulation on symbol comprehension and production in children with moderate cognitive disabilities. American Journal of Speech Language Pathology. 2004; 13:155–167. [PubMed: 15198634]

Heuer S, Hallowell B. An evaluation of multiple-choice test images for *comprehension* assessment in aphasia. Aphasiology. Sep; 2007 21(9):883–900.

Higginbotham J. Use of nondisabled subjects in AAC research: Confessions of a research infidel. Augmentative and Alternative Communication. 1995; 11:2–5.

Hopkins KA. Why do babies find faces attractive? Australian Journal of Early Childhood. 1980; 5:25–28.

Jagaroo, V.; Wilkinson, KM.; Light, J. Current views on scene perception and its relation to aided AAC displays. (in preparation)

Johnston SS, Reichle J. Designing and implementing interventions to decrease problem behaviors. Language, Speech, and Hearing Services in Schools. 1993; 24:225–235.

Kirchner H, Bacon N, Thorpe SJ. "In which of two scenes is the animal?" Ultra-rapid visual processing demonstrated with saccadic eye movements. Perception. 2003; 32(Supplement):170.

Klin A, Jones W, Schultz R, Volkmar F, Cohen D. Visual fixation patterns during viewing of naturalistic social situations as predictors of social competence in individuals with autism. Archives of General Psychiatry. 2002; 59:809–816. [PubMed: 12215080]

Lang, PJ.; Bradley, MM.; Cuthbert, BN. Technical Report A-8. University of Florida; Gainesville, FL: 2008. International Affective Picture System (IAPS): Affective ratings of pictures and instruction manual.

Light, J.; Drager, K. Evidence-based AAC interventions to build language and communication skills with infants, toddlers, and preschoolers. Seminar presented at the ISAAC conference; Montreal, Canada. 2008 Aug.

Light, J.; Parsons, AR.; Drager, K. “There’s more to life than cookies”: Developing interventions for social closeness with beginning communicators who use AAC. In: Reichle, J.; Beukelman, DR.; Light, J., editors. Exemplary Practices for Beginning Communicators. Baltimore: Paul H. Brookes; 2002. p. 187-218.

Mirenda, P.; Iacono, T., editors. Autism spectrum disorders and AAC. Baltimore: Paul H. Brookes;

Nelson, K. Event knowledge: Structure and function in development. Hillsdale, N.J: L. Erlbaum Associates; 1986.

Odekar A, Hallowell B, Kruse H, Moates D, Lee C. Validity of eye movement methods and indices for capturing semantic (associative) priming effects. Journal of Speech, Language, and Hearing Research. 2009; 52:31–48.

Pelphrey KA, Sasson NJ, Reznick JS, Paul G, Goldman BD, Piven J. Visual scanning of faces in autism. Journal of Autism and Developmental Disorders. 2002; 32:49–261.

Riby D, Hancock PJB. Viewing it differently: Social scene perception in Williams syndrome and Autism. Neuropsychologia. 2008; 46:2855–2860. [PubMed: 18561959]

Riby D, Hancock PJB. Looking at movies and cartoons: Eye tracking evidence from Williams Syndrome and autism. Journal of Intellectual Disability Research. 2009a; 53:169–181. [PubMed: 19192099]

Riby D, Hancock PJB. Do faces capture the attention of individuals with Williams syndrome or autism? Evidence from tracking eye movements. Journal of Autism and Developmental Disorders. 2009b; 39:421–431. [PubMed: 18787936]

Romski, MA.; Sevcik, RA. Breaking the speech barrier: Language development through augmented means. Baltimore: Brookes; 1996.

Smilek D, Birmingham E, Cameron D, Bischof W, Kingstone A. Cognitive ethology and exploring attention in real-world scenes. Brain Research. 2006; 1080:101–119. [PubMed: 16480691]

van der Geest JN, Kemner C, Camfferman G, Verbaten MN, van Engeland H. Looking at images with human figures; comparison between autistic and normal children. Journal of Autism and Developmental Disorders. 2002; 32:69–75. [PubMed: 12058845]

Wetherby, A.; Prizant, B. The Communication and Symbolic Behavior Scales. Baltimore: Paul H. Brookes; 1993.

Wilkinson KM, Carlin M, Jagaroo V. Preschoolers’ speed of locating a target symbol under different color conditions. Augmentative and Alternative Communication. 2006; 22:123–133. [PubMed: 17114170]

Wilkinson KM, Carlin M, Thistle J. The role of color cues in facilitating accurate and rapid location of aided symbols by children with and without Down syndrome. American Journal of Speech-Language-Pathology. 2008; 17:179–193. [PubMed: 18448605]

Wilkinson KM, Hennig S. The state of research and practice in augmentative and alternative communication for children with developmental/intellectual disabilities. Mental Retardation and Developmental Disabilities Research Reviews. 2007; 13:58–69. [PubMed: 17326111]

Wilkinson, KM.; Carlin, M.; McIlvane, WJ. Visual processing in individuals with ID: Implications for the construction of visual supports. Poster presented at the annual Gatlinburg Conference on Research and Theory in Mental Retardation/Developmental Disabilities; New Orleans, LA. March; 2009.

Wilkinson KM, Jagaroo V. Contributions of visual cognitive neuroscience to AAC display design. Augmentative and Alternative Communication. 2004; 20:123–136.

Wilkinson, KM.; Light, J.; Drager, KD. Principles of visual processing and the design of visual scene displays for beginning communicators. 2009. Manuscript under review

Wilkinson, KM.; Light, J.; McStravock, K.; Fratantoni, K.; Miller, J.; Drager, K. Improving visual scene display design through consideration of visual processes. Miniseminar submitted to the annual conference of the American Speech-Language-Hearing Association; New Orleans, LA. November; 2009.

Wilkinson, KM.; Reichle, J. The role of aided AAC in replacing unconventional communicative acts with more conventional ones. In: Mirenda, P.; Iacono, T.; Light, J., editors. Autism spectrum disorders and AAC. Vol. chapter 13. Baltimore: Paul H. Brookes; 2009.

Figure 1.
Simple VSD of child using phone

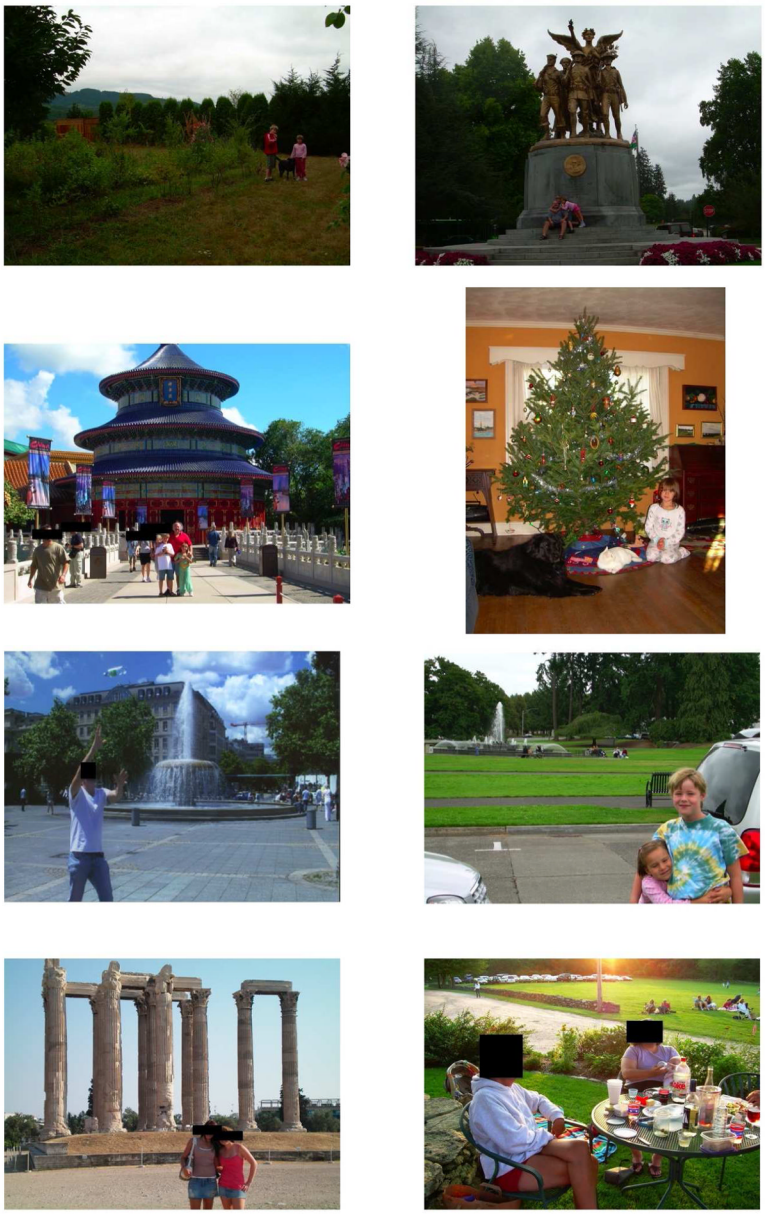

Figure 2.
The eight photographs used for the current study

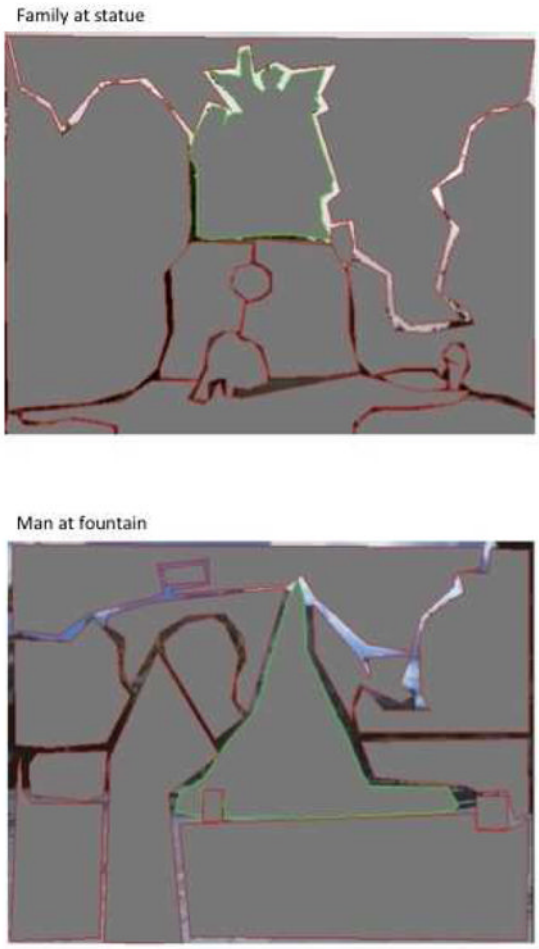

Figure 3.
An example of the enclosed elements for one photograph

Table 1

Participants' eye gaze patterns (number noting, time spent fixating, latency to first fixation, and ranks for the human figure element on these measures) across elements in the 8 photographs

	# NOTING		TIME SPENT ON ELEMENT		Latency to first fixation
Element name (% of space occupied); in order of size	**Total # Ss noting**	**% of Ss (rank, + if tied)**	**Total % time spent on element (rank, + if tied)**	**ratio: time/space**	**In seconds (rank)**
Children in Garden					
sky (22.3)	15	79%	2	0.09	5.70
garden (20.9)	16	84%	4	0.19	3.93
grass (19.2)	4	21%	0	0.00	6.91
dark bg trees (13.7)	13	68%	2	0.15	5.88
midline trees (6.8)	18	95%	5	0.74	3.90
children & dog (2.5)	**19**	**100% (1)**	**12 (1)**	**4.80***	**1.75 (1)**
colorful tree (1.4)	14	74%	2	1.43	5.01
fence (1.1)	14	74%	3	2.73	5.62
mountain (.9)	6	32%	1	1.11	6.75
Family at China Tower					
tower (23.8)	19	100%	16	0.67	2.08
sky (22.8)	8	42%	1	0.04	4.98
bushes, fence (20.3)	13	68%	2	0.10	5.28
bystanders (6.1)	13	68%	5	0.82	4.58
posters (5.9)	12	63%	2	0.34	5.56
small family (2.7)	**17**	**89% (2)**	**8 (2)**	**2.96***	**2.88 (2)**
trash cans (1.5)	2	11%	0	0.00	6.63
sign (.7)	11	58%	3	4.29	5.18
Women with Pillars					
pillars (50.7)	19	100% (1+)	19 (1+)	0.37	2.01
sand (11.1)	2	11%	0	0.00	6.42
sky (11.2)	1	5%	0	0.00	6.63
base (8.9)	10	53%	2	0.22	6.30
two women (6.7)	**19**	**100% (1+)**	**19 (1+)**	**2.84***	**1.38 (1)**
trees (2.8)	4	21%	1	0.36	6.30

J Speech Lang Hear Res. Author manuscript; available in PMC 2012 December 01.

	# NOTING		TIME SPENT ON ELEMENT		Latency to first fixation
Element name (% of space occupied); in order of size	**Total # Ss noting**	**% of Ss (rank, + if tied)**	**Total % time spent on element (rank, + if tied)**	**ratio: time/space**	**In seconds (rank)**
Man at Fountain					
ground (24.8)	3	16%	0	0.00	6.82
trees (17)	18	95%	4	0.24	4.55
sky (11.2)	19	100% (1+)	4	0.36	4.77
water fountain (10.4)	19	100% (1+)	12	1.15	4.04
man (9)	**19**	**100% (1+)**	**11 (2)**	**1.22 (ns)**	**1.35 (1)**
buildings (5.5)	19	100%	5	0.91	3.89
bystanders (5.2)	10	53%	1	0.19	5.85
trash cans (.8)	0	0%	0	0.00	7.00
water bottle (.3)	9	47%	2	6.67	5.79
Children at Fountain					
trees (28)	16	84%	5	0.18	5.09
grassy space (18.8)	18	95%	4	0.21	4.38
cement (14.9)	4	21%	0	0.00	7.00
children (12)	**19**	**100% (1+)**	**13 (1)**	**1.08(ns)**	**1.32 (1)**
front cars (9)	1	5%	0	0.00	6.93
fountain (2.3)	19	100% (1+)	9	3.91	3.72
objects distance (1.5)	13	68%	3	2.00	5.24
bystanders (1)	15	79%	1	1.00	5.86
bench (.8)	10	53%	1	1.25	6.09
Child at Xmas Tree					
tree (14.8)	19	100%	20	1.35	1.21
floor (9)	10	53%	1	0.11	6.71
ceiling (6.7)	8	42%	1	0.15	6.63
curtains (6.4)	10	53%	1	0.16	5.98
dog and cat (4)	17	89% (2+)	9	2.25	3.34
cabinets (3.5)	7	37%	1	0.29	6.05
pictures (3.4)	8	42%	2	0.59	5.86
child (1.8)	**17**	**89% (2+)**	**10 (2)**	**5.56** *	**2.13 (2)**
sunlight (.6)	6	32%	1	1.67	6.44

J Speech Lang Hear Res. Author manuscript; available in PMC 2012 December 01.

	# NOTING		TIME SPENT ON ELEMENT		Latency to first fixation
Element name (% of space occupied); in order of size	**Total # Ss noting**	**% of Ss (rank, + if tied)**	**Total % time spent on element (rank, + if tied)**	**ratio: time/space**	**In seconds (rank)**
Family at Statue					
trees (36.3)	11	58%	2	0.06	5.91
sky (23)	16	84%	3	0.13	4.76
base (9.3)	19	100%	9 (2+)	0.97	2.87
statue people (9)	19	100%	17	1.89	1.91
flowers (3.3)	2	11%	0	0.00	7.00
small family (1.3)	**14**	**74% (4)**	**9 (2+)**	**6.92***	**3.12 (3)**
medallion (.5)	11	58%	1	2.00	6.05
van (.4)	6	32%	0	0.00	6.90
flag (.3)	1	5%	0	0.00	7.00
stop sign (.3)	7	37%	1	3.33	6.81
Women at Table					
women (16.8)	**19**	**100% (1)**	**12 (1+)**	**0.71 (ns)**	**1.57 (1)**
items on table (11)	17	89%	12 (1+)	1.09	3.28
grass (12.6)	17	89%	4	0.32	5.30
bushes (11.6)	16	84%	4	0.34	3.55
back trees (6.9)	4	21%	1	0.14	5.94
driveway (5.1)	10	53%	1	0.20	6.63
cars (3.4)	7	37%	1	0.29	6.87
stone wall (1.8)	10	53%	1	0.56	6.02
bystanders (1.7)	13	68%	2	1.18	5.64
carriage (1.6)	3	16%	0	0.00	7.00
grocerybag (.9)	2	11%	0	0.00	7.00

Note: The first two columns report the raw number and percentage of participants who fixated on each element (and the rank of the human in attracting these fixations, among all elements). The second two columns report the overall time spent looking at each element (and the rank of the human in maintaining this scrutiny) and the ratio of time spent to size occupied (along with the results of the inferential statistical analysis for this measure). The final column reports the mean latency to first fixation for each element (and the rank of the human in this attention capture). Numbers in parentheses next to each element represent the space occupied in the photograph by that element; numbers in parentheses next to each dependent measure represents the rank of the human element relative to all other elements in the photograph, on that measure;

* means that the adjusted ratio of time spent looking at the human element relative to its size is significantly greater than 1, at p = .006; + reflects that this element tied with another for the rank indicated.