CrossMark

BRIEF REPORT

Brief Report: Just-in-Time Visual Supports to Children with Autism via the Apple Watch:® A Pilot Feasibility Study

Amanda O'Brien[1] · Ralf W. Schlosser[1,2,3] · Howard C. Shane[1,4] · Jennifer Abramson[1] ·
Anna A. Allen[4] · Suzanne Flynn[5] · Christina Yu · Katherine Dimery[1]

Abstracts Using augmented input might be an effective means for supplementing spoken language for children with autism who have difficulties following spoken directives. This study aimed to (a) explore whether JIT-delivered scene cues (photos, video clips) via the Apple Watch® enable children with autism to carry out directives they were unable to implement with speech alone, and (b) test the feasibility of the Apple Watch® (with a focus on display size). Results indicated that the hierarchical JIT supports enabled five children with autism to carry out the majority of directives. Hence, the relatively small display size of the Apple Watch does not seem to hinder children with autism to glean critical information from visual supports.

✉ Ralf W. Schlosser
  R.Schlosser@northeastern.edu; R.Schlosser@neu.edu

[1] Department of Otolaryngology and Communication Enhancement, Boston Children's Hospital, 9 Hope Ave, Waltham, MA 02453, USA

[2] Department of Communication Sciences and Disorders, Department of Applied Psychology, Northeastern University, 360 Huntington Ave, Boston, MA 02115, USA

[3] Centre for Augmentative and Alternative Communication, University of Pretoria, Private Bag X20, Hatfield, Pretoria 0028, South Africa

[4] MGH Institute of Health Professions, 36 1st Ave., Boston, MA 02129, USA

[5] Department of Linguistics, Massachusetts Institute of Technology, 77 Massachusetts Ave., Bldg. 32-D808, Cambridge, MA 02139, USA

Introduction

For children with autism, receptive language difficulties have been understudied relative to expressive language issues (Sevcik 2006) despite documented difficulties with understanding language concepts (Mechling and Hunnicutt 2011). More often than not, children with autism are presented with spoken input (Hall et al. 1995) even though deficits in comprehending spoken language are well-documented (Von Tetzchner et al. 2004). As a result, some researchers have advocated for reducing the complexity of the auditory environment by using other forms of input (Hodgdon 1995).

Augmented input refers to strategies to supplement "the input provided to AAC users during communication interaction or during instruction in AAC use" (Wood et al. 1998, p. 261). For example, a child may be provided with oral instructions for a recipe in cooking class during which the instructions are embellished with line drawings to aid in comprehension (Wood et al. 1998). More recently, scene cues have been proposed as beneficial modalities for augmented input. Scene cues are images that portray relevant concepts and their relationships in context through pictorial forms (e.g., line drawings), photos, or full-motion video clip (Shane 2006). Scene cues may be static or dynamic. *Static scene cues* are images that portray relevant concepts and their relationships in context through pictorial form (e.g., line drawings) or photos (Shane 2006). *Dynamic scene cues* are images that portray relevant concepts and their relationships in context through full motion video clips (Shane 2006). In a recent study, nine children with

autism were presented with prepositional directives to place figurines on the table top in a particular arrangement in three input conditions: (a) spoken input, (b) static scene cues plus spoken input, and (c) dynamic scene cues plus spoken input. The children followed instructions more effectively when presented with scene cues (static or dynamic) relative to spoken input alone (Schlosser et al. 2013). This study showed the potential of scene cues as an augmented input modality over spoken-only cues by directly comparing each input condition in a within-subjects design. Scene cues, however, have the potential to be provided on an as needed or just-in-time (JIT) basis rather than as a matter of fact. That is, should communication partners recognize that a child does not understand spoken input, only then will the partner supply the scene cue.

The JIT construct is gaining traction as a method for providing augmentative and alternative communication and visual supports to children with developmental disabilities (Schlosser et al. 2016), in part fueled by the mobile technology revolution (Shane et al. 2012). JIT supports have the potential to (a) lower working memory demands, (b) provide a context via situated cognition, and (c) capitalize on teachable moments (Schlosser et al. 2016). The Apple Watch[®1] (https://www.apple.com/watch/), a wearable technology that vibrates on the wrist when a new text message arrives, has great potential to deliver JIT visual supports such as scene cues in an unobtrusive and discreet manner. Using the Apple Watch[®] to receive scene cues requires the child to have several operational and related skills, but perhaps the ability to view images on its relatively small display is the most pivotal skill. If children were unable to recognize the images, there would be no reason to examine other operational and related skills such as the ability to tolerate wearing the watch on the wrist. As a result, in this study we did not ask the children to wear the watch and receive scene cues via text message; rather, the instructor held the watch in front of the child to show scene cues. The purpose of this feasibility study was two-fold: (a) to explore whether scene cues delivered in a JIT manner enable children with autism to carry out directives; and (b) to test the feasibility of the Apple Watch[®] as a means to present JIT visual supports.

Methods

Participants, Setting, and Experimenter

In order to be selected, participants had to meet the following criteria: (a) an unequivocal primary diagnosis of autism spectrum disorder (based on medical or school records); (b) chronological ages of 6–17 years; (c) hearing and vision within normal limits (as determined by medical

records or parental reports); (d) demonstrated strong interest in visuals including the use of media (based on parent report); and (e) ability to perform screening tasks (see Procedures below) without edible reinforcement (only social reinforcement such as praise will be given). Five children met the above inclusion criteria. Their characteristics are summarized in Table 1.

The study was carried out in a 12×10 square feet clinical room at a pediatric Autism clinic in the Northeast of the U.S. The child sat at a table adjacent to the experimenter with figurines placed on the table top. A licensed speech-language pathologist in the Autism Language Program served as the experimenter and a graduate student intern completing her Master's degree in speech-language pathology served as the independent observer.

Dependent Measure

A response was considered correct if the child carried out the directive with the appropriate figurines/objects on the table top within 10 s of the spoken directive, the static scene cue, or the dynamic scene cue. The number of directives implemented correctly served as the dependent measure.

An independent observer coded the dependent variable (correct, incorrect) in 20 % of the sessions. Inter-observer agreement (IOA) was calculated by dividing the total number of agreements by the number of agreements plus disagreements multiplied by 100. Analysis revealed 100 % agreement between the instructor and the independent observer.

Materials

Materials included an iPad,[®1] the Apple Watch[®] Sport[2] (Model A 1554, 42 mm size, 1.65″ Ion-X glass retina display, 312×390 pixels resolution, composite black), objects and photographs for the screening task (i.e., ball, bottle, boy, Cookie Monster, duck, lamp), five spoken directives and their corresponding scene cues for the screening task (have the boy kick the ball; have Cookie Monster jump; have the duck drink from the bottle; put the duck on the lamp; and put the boy behind the lamp), 10 spoken directives involving prepositional phrases and their corresponding static and dynamic scene cues for the experimental task ("block in cup," "dog on block," "girl on block," "girl in car," "dog on car," "girl up ladder," "dog up ladder," "block down slide," "dog down slide," and "girl push car"), and objects and figurines for the

Table 1 Participant characteristics

Participant	Diagnosis	Chronological age	Gender	Description of speech
1	Autism; ADHD	8;3	Female	Uses 1–3 word scripted phrases to request and protest
				Speech is highly echoic at the simple sentence level
2	Autism; Anxiety disorder	13;9	Male	Uses 1–3 word scripted phrases to request and protest
				Uses single words to label nouns
3	Autism; ADHD	8;0	Male	Uses simple sentences with some grammatical errors to request, protest, and comment
				Speech contains familiar scripts at the simple sentence level
4	Autism	8;4	Female	Uses 1–3 word scripted phrases to request and protest
				Uses single words to label nouns
5	Autism	12;2	Female	Uses single words to request, protest, and label
				Speech is highly echoic at the single word level

experimental task (e.g., block, cup, dog, girl, car, ladder, swing set).

Procedures

Screening Tasks

Two screening tasks were carried out to rule in or out potential participants. The first task involved the matching of six photographs to their corresponding objects. Specifically, children were presented with one full-screen size photograph at a time on the iPad® and asked to match it to the corresponding object from an array of six objects displayed on the table top within 10 s of the instruction "match _____ (name of object)." Children were provided with intermittent non-specific reinforcement (e.g., "keep up the good work") to sustain participation. In order to be counted as a correct response the child had to point to or pick up the corresponding object within the allotted time. In order to be included in the study, children needed to achieve at least 50 % (i.e., three matches) accuracy.

The second task asked potential participants to carry out five directives with figurines and objects on the table top when presented with three input conditions in a sequential JIT manner: (a) spoken cues only; (b) static scene cues on the iPad® plus spoken cues; and (c) dynamic scene cues on the iPad® plus spoken cues. Specifically, each directive was presented first with speech alone. If the child was unable to carry out the directive within 10 s, the child was presented with a static scene cue of the same directive along with speech. If the child still did not carry out the directive accurately, the child was presented with a dynamic scene cue along with speech. As before, children were provided with intermittent non-specific reinforcement only. Children who were able to follow all of the five directives when presented with speech alone, were
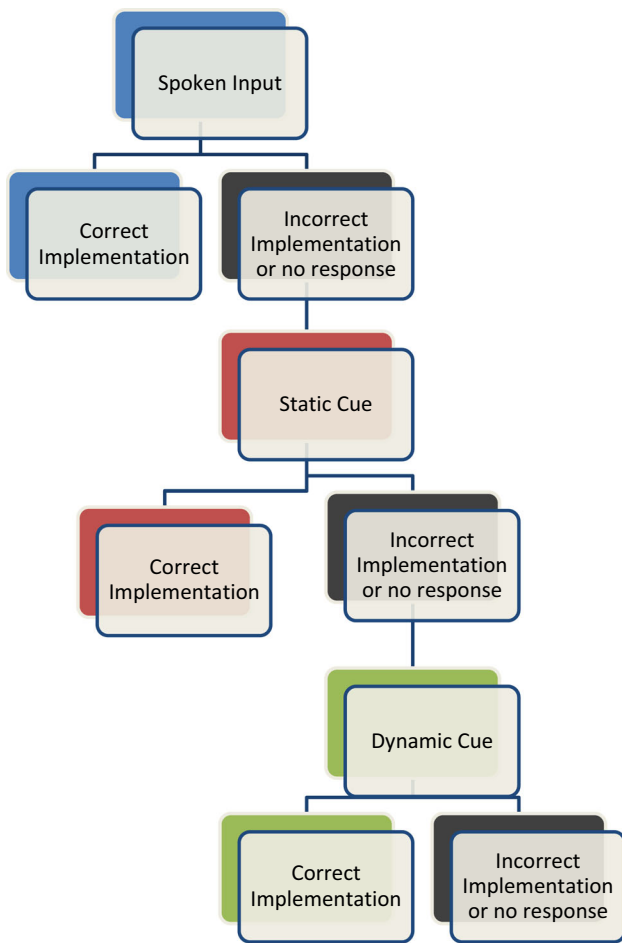
excluded from the study. Children who required visual supports and were able to implement at least 3 out of 5 directives (i.e., 80 %) when presented with static or dynamic scene cues, qualified for participation in the study.

Experimental Task

As with the second screening task, participants were presented with directives and provided with scene cues in a JIT manner as outlined below (and illustrated in Fig. 1), except that this time the scene cues were provided on the Apple Watch® instead of the iPad® and the 10 directives were different from those in the screening task. As before, each directive was presented initially in spoken form, and the child had 10 s to carry out the directives with the figurines and objects provided on the table top. If necessary, the spoken directives were repeated twice for a total of three times. If the child failed to implement the directive accurately or did not respond after the third presentation, the experimenter showed a static scene cue for the same directive on the Apple Watch, holding it approximately 1 foot away from the child at eye level. If necessary, the static scene cue was presented two more times for a total of three times. Again, the child had 10 s to carry out the directive and, if unsuccessful, the experimenter presented and activated the dynamic scene cue on the Apple Watch.® As before, dynamic scene cues were repeated for a total of three times as necessary. Throughout, the children were provided with non-specific intermittent feedback to sustain motivation.
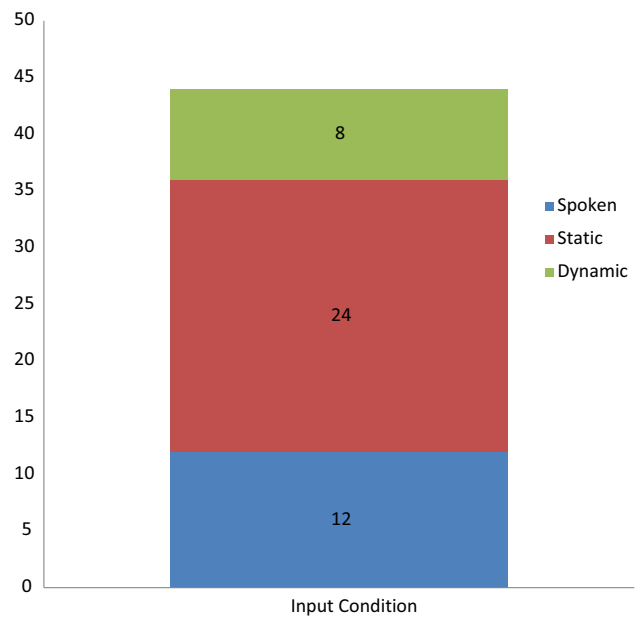
Results

Due to the small n, we refrained from statistical analyses. At a group level, the study involved a total of 50 directives across the five participants. In absolute terms, the children
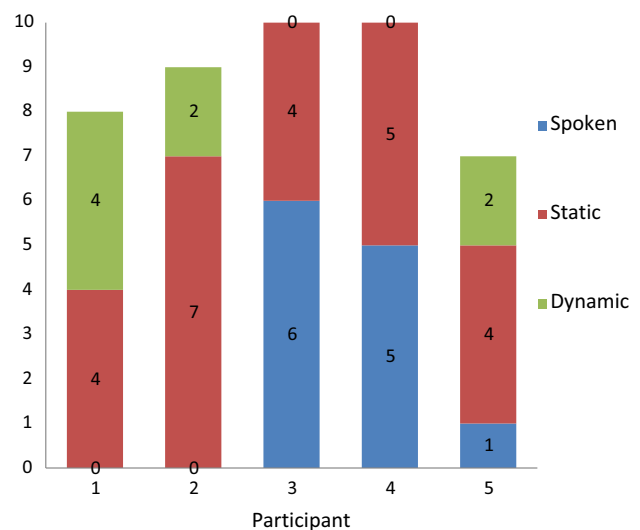
Fig. 1 Hierarchical organization of just-in-time input conditions

Fig. 2 Total number of correct responses to directives across participants based on the three hierarchical input conditions (spoken, static, dynamic)

At an individual level, there was some variability in the extent to which children were able to follow spoken cues, ranging from 0 to 6 (0 to 60 %), with three participants (#1, 2, and 5) at the lower end with 0–1 (0–10 %) and two participants (# 3 and 4) at the upper end with 5–6 (50–60 %) (see Fig. 3). Interestingly, the two participants at the upper end of following spoken directives were able to carry out all of the remaining directives with static scene cues, negating the need for dynamic scene cues. The children at the lower end of being able to follow spoken

successfully implemented 12 (24 %) of the directives with spoken cues only, 24 (48 %) of directives with static scene cues, and 8 (16 %) of the directives with dynamic scene cues. Six (12 %) of the directives were not implemented correctly or resulted in no response (see Fig. 2). Given that the nature of these data are hierarchical (e.g., static scene cues only come into play when the child does not comprehend the spoken only cues), the data can also be reported another way. Since 12 (24 %) of the directives were completed successfully with spoken cues, it was not necessary to supply scene cues for these directives. For the remaining 38 (76 %) of the directives, however, it was warranted to present JIT support via scene cues. Of these remaining directives (the new 100 %), 24 (63.16 %) were successfully implemented when presented with static scene cues plus speech and 14 (36.84 %) directives were carried out incorrectly. For the 14 remaining directives (100 %), 8 (57.14 %) were carried out correctly when presented with dynamic scene cues. Six of the 50 directives (12 %) were not carried out correctly even after dynamic scene cues were provided.

Fig. 3 Individual participant data of correctly followed directives per hierarchical input condition (spoken, static, dynamic)

directives, however, benefitted from both static and dynamic scene cues and, by the end of the study, approximated a near perfect score with 8–9 correct (80–90 %).

Discussion

This study aimed to explore whether JIT-delivered scene cues enable children with autism to carry out directives that they were unable to carry out when provided with spoken input alone. The data provide preliminary support for the processing advantages of scene cues when provided in a JIT manner following the realization that the children failed to respond to spoken only cues. This extends the findings from a previous study in which the non-JIT provision of scene cues resulted in superior direction following compared to spoken cues only (Schlosser et al. 2013). The strong performance with static scene cues provides preliminary data-based support for their placement within the hierarchy of JIT supports (i.e., before dynamic scene cues) for directives that involve prepositional phrases (see Fig. 1). In other words, it appears logical to provide static scene cues before dynamic scene cues, analogous to a least-to-most prompting hierarchy. For directives involving prepositional phases, static scene cues show the placement of a figurine in its *final* position relative to the object rather than *being suggestive of a movement* to get to this position. Hence, static scene cues may be particularly effective in representing directives involving prepositional phrases. It remains to be seen whether that would be the case for directives involving actions ("make the dinosaur hop") where static scene cues do imply movement.

These pilot data suggest that even children who have some ability to follow directives in the spoken modality can still benefit from static scene cues. The two performance profiles gleaned from the analysis of individual variation give rise to the hypothesis that children's ability to follow spoken directives is correlated with the degree to which they can take advantage of static scene cues as well as their need to receive dynamic scene cues. That is, children with good ability to follow spoken cues seemed able to fully capitalize on static scene cues without needing dynamic scene cues as the next level of JIT support. On the other hand, children with extremely limited spoken comprehension skills seemed to benefit from static scene cues to some degree, but also required dynamic scene cues for some directives.

A related purpose of this study was to examine the feasibility of the Apple Watch® as a means to deliver JIT visual supports. Using the proposed taxonomy of classifying JIT supports by (a) intended purpose, (b) modalities, (c) source, and (d) delivery method (Schlosser et al. 2016), the scene cues in this study served as prompts, in the visual modality, were mentor-generated, and delivered face-to-face via the Apple Watch.® Holding up the scene cues on the Apple Watch,® rather than having the child wear the watch and send the scene cues via text, permitted the removal of any potential sensory issues and a focus solely on the viewing of the small display size. Based on the successful implementation of directives when presented with static and dynamic scene cues, the children managed to retrieve pertinent information from the scene cues despite the relatively small display of the Apple Watch.® While the sample size was too small to be able to extrapolate to the larger population of children with ASD, all enrolled children seemed capable of using the small display size.

Now that it is clear that the children in this sample were able to act on the small display size, future research should attend to additional operational competencies needed in order to harness the full potential of the Apple Watch® for delivering visual supports in a JIT manner and to do so unobtrusively and discreetly. In addition to tolerating the watch on the wrist and being able to process vibro-tactile cues (i.e., no hypersensitivity to touch or tactile defensiveness), a user of the watch needs to raise one's arm to view a static scene cue, and (if applicable) to touch the image in order to activate a dynamic scene cue.

Although replication with a larger number of participants is needed, this study offers preliminary evidence that scene cues can be successfully provided in a JIT manner via the Apple Watch,® when spoken directives are not understood. The study also shows that children with autism can extract important visual information despite the small screen size of the Apple Watch,® suggesting it is a potentially viable technology for the delivery of JIT visual supports.

Author Contributions AO conceived of the study, participated in its design and coordination, provided technical solutions in the use of the Apple Watch, collected data, trained independent observers, entered the data, analyzed the data, interpreted the data, helped draft the manuscript, and helped revise the manuscript. RS conceived of the study, conducted the literature review, took a leadership role relative to research design and measurement, analyzed and interpreted the data, drafted the manuscript, and spearheaded the revision of the manuscript. HS conceived of the study, provided input into technical aspects of data collection, helped situate the study within the JIT taxonomy, interpreted the data, helped draft the manuscript, and assisted in revising the manuscript. JA conceived of the study, provided technical solutions for Apple Watch, interpreted the data, helped draft the manuscript. AA and SF conceived of the study, assisted in interpreting the data, and helped draft the manuscript. CY conceived of the study, collected data, assisted in interpreting the data, and helped draft the manuscript. All authors read and approved the final manuscript.

Compliance with Ethical Standards

Conflict of interest The authors report no conflicts of interests.

Human and Animal Rights This research was approved by the appropriate Institutional Review board (IRB).

Informed Consent Informed consent was obtained from the legal guardians/parents of each participating child.

References

Hall, L. J., McClannahan, L. E., & Krantz, P. J. (1995). Promoting independence in integrated classrooms by teaching aides to use activity schedules and decreased prompts. *Education and Training in Mental Retardation and Developmental Disabilities, 30*, 208–217.

Hodgdon, L. (1995). *Visual strategies for improving communication*. Troy, MI: Quirk-Roberts.

Mechling, L. C., & Hunnicutt, J. R. (2011). Computer-based video self-modeling to teach receptive understanding of prepositions by students with intellectual disabilities. *Education and Training in Autism and Developmental Disabilities, 46*, 369–385.

Schlosser, R. W., Laubscher, E., Sorce, J., Koul, R., Flynn, S., Hotz, L., et al. (2013). Implementing directives that involve prepositions with children with autism: A comparison of spoken cues

with two types of augmented input. *Augmentative and Alternative Communications, 29*, 132–145.

Schlosser, R. W., Shane, H. C., Allen, A. A., Abramson, J., Laubscher, E., & Dimery, K. (2016). Just-in-time supports in augmentative and alternative communication. *Journal of Developmental and Physical Disabilities, 28*, 177–193.

Sevcik, R. A. (2006). Comprehension: An overlooked component in augmented language development. *Disability and Rehabilitation, 28*, 159–167.

Shane, H. C. (2006). Using visual scene displays to improve communication and communication instruction in persons with Autism Spectrum Disorders. *Perspectives in Augmentative and Alternative Communication, 15*, 7–13.

Shane, H. C., Laubscher, E., Schlosser, R. W., Flynn, S., Sorce, J. F., & Abramson, J. (2012). Applying technology to visually support language and communication in individuals with ASD. *Journal of Autism and Developmental Disorders, 42*, 1228–1235.

Von Tetzchner, S. R., Øvreeide, K. D., Jørgensen, K. K., Ormhaug, B. M., Oxholm, B. B., & Warme, R. R. (2004). Acquisition of graphic communication by a young girl without comprehension of spoken language. *Disability and Rehabilitation: An International, Multidisciplinary Journal, 26*, 1335–1346.

Wood, L. A., Lasker, J., Siegel-Causey, E., Beukelman, D. R., & Ball, L. (1998). Input framework for augmentative and alternative communication. *Augmentative and Alternative Communication, 14*, 261–267.