



HAL
open science

Gaze Awareness in Agent-Based Early-Childhood Learning Application

Deepak Akkil, Prasenjit Dey, Deepshika Salian, Nitendra Rajput

► **To cite this version:**

Deepak Akkil, Prasenjit Dey, Deepshika Salian, Nitendra Rajput. Gaze Awareness in Agent-Based Early-Childhood Learning Application. 16th IFIP Conference on Human-Computer Interaction (INTERACT), Sep 2017, Bombay, India. pp.447-466, 10.1007/978-3-319-67684-5_28 . hal-01678416

HAL Id: hal-01678416

<https://inria.hal.science/hal-01678416>

Submitted on 9 Jan 2018

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

Gaze Awareness in Agent-Based Early-Childhood Learning Application

Deepak Akkil¹, Prasenjit Dey², Deepshika Salian³, Nitendra Rajput^{4*}

¹ University of Tampere, Tampere, Finland

{deepak.akkil}@uta.fi

² IBM Research, Bangalore, India

prasenjit.dey@in.ibm.com

³ Kidzee Kindergarten Domlur, Bangalore, India

kidzeedomlur6@gmail.com

⁴ Infoedge India Limited¹

nitendra@acm.org

Abstract. Use of technological devices for early childhood learning is increasing. Now, kindergarten and primary school children use interactive applications on mobile phones and tablet computers to support and complement classroom learning. With increase in cognitive technologies, there is further potential to make such applications more engaging by understanding the user context. In this paper, we present the Little Bear, a gaze aware pedagogical agent, that tailors its verbal and non-verbal behavior based on the visual attention of the child and employs means to reorient the attention of the child, when distracted from the learning activity. We used the Little Bear agent in a learning application to enable teaching the vocabulary of everyday fruits and vegetables. Our user-study (n=12) with preschoolers shows that children interacted longer and showed improved short-term retention of the vocabulary using the gaze aware agent compared to a baseline touch-based application. Our results demonstrate the potential of gaze aware application design for early childhood learning.

Keywords. Gaze, Touch, Pedagogical Agent, Early-childhood learning, Vocabulary building, Games, Mobile devices, Engagement, Comparative study.

1 Introduction

The use of developmentally and pedagogically appropriate technology to promote early childhood learning is on the rise. Recently, a variety of technological tools, such as digital whiteboards [29], touch sensitive tabletops [10], Kinect sensors for whole body gestural interaction [17], and tangible objects [6] are beginning to be applied in the early childhood pedagogical environment. In addition, mobile phones and tablet computers are among the most popular technological devices applied in early childhood learning environment. Growing popularity and access to technology enable learners to

* This work was done when Nitendra Rajput was working for IBM Research.

become familiar in interacting with these devices at a very young age and provide scope for designing innovative, ubiquitous, and constructive learning experiences [34].

Currently, touch-based interaction is the most common technique in learning applications on mobile devices. Direct-touch interactive applications enhance learning compared to non-interactive video viewing [1] and is also known to be both easier and preferred by children than indirect interfaces such as a mouse [26]. However, there are also several challenges in designing touch-based interactive applications for children. Plowman and Mcpake [31] note that if children do not understand what they need to do, the interactivity offered by such devices may be counter-productive to learning. Touch interactions require the children to reach to the screen and tap the screen or do specific touch gestures, to perform an action. This requires careful design of the stimuli and prompts to make it intuitive [16]. Further, reaching to touch the screen maybe difficult depending on how the screen is positioned in front of the child [35]. Also, for early learners, the problem of accidental touches, and need for fine motor skills to perform complex actions has been recognised as challenges for using touch interactivity.

Another challenge in designing optimal learning experiences for young children is that the children have very limited attention span and get easily distracted (e.g. a noise from the hallway or colorful object in the tablet screen of a peer)) [12]. Luna [27] notes that younger the child is, the more easily distractible they are. It is hence important in educational applications designed for early learners to be aware of children's attention and employ ways to reorient the attention when the child is distracted to facilitate continuous learning.

Recent advancements in computer vision software and hardware technology, have made gaze tracking cheaper, more accurate and ergonomic to use. The technology is increasingly seen as a viable and potentially beneficial interaction technique in mobile devices [22]. Gaze tracking has been previously used with children. For example, as a methodology in early developmental research [11], as a tool to diagnose and understand different psychiatric disorders in young children [39], and as an assistive technology for children with physical [5,18] or learning [37] disabilities. However, the potential of gaze-based interaction remains unexplored for early childhood learning, among the general children population.

Gaze provides implicit information about the attention, intention, and rich cues of the interest of a child. Unlike touch-based interaction which needs an explicit action, a gaze-based application could adapt itself based on visual attention of the child, integrating learning with their curious visual exploration [20] of young minds, providing a rich and embodied experience. Gaze aware applications could also keep track of the visual attention of the child and employ means to reorient the attention, when the child is distracted from the learning activity.

In this paper, we present the Little Bear, a gaze-aware pedagogical agent designed for early childhood learners. The Little Bear tailors its verbal and non-verbal behavior in response to the visual attention of the child. The agent also exhibits emotional responses of sadness and boredom when the user is distracted from the learning activity, to help reorient the attention of the child to the learning activity. We developed an educational application using the Little Bear as the central character, to teach the vocab-

ulary of everyday fruits and vegetables. Further, to evaluate the feasibility and usefulness of the gaze-aware pedagogical agent, we conducted two user studies with children from the 4-5 years age group. The purpose of the first study was to understand how children interacted with the application using the Little Bear and compare it with a baseline application that makes use of conventional touch-based interaction. The purpose of the second study was to understand the feasibility of integrating the gaze-aware agent in a touch-based application.

1: Comparing gaze-aware application with the baseline touch-based application: Different fruits and vegetables appeared onscreen and the learner was free to explore them. When the user activated an object, by either looking at it (*LittleBear*) or touching the item (*baseline*), the agent spoke interesting details about the item. Additionally, in the *LittleBear* application, the agent exhibited verbal and non-verbal behavior to help reorient the attention of the child, when the child looked away from the tablet screen. In this study, we wanted to understand how learners interacted with the application based on the conditions and observe differences in engagement, if any.

2: Feasibility of combining gaze and touch input: The agent asked a question regarding an onscreen fruit or vegetable (e.g. *could you show me the fruit banana*) and the learner had to touch the corresponding fruit. The agent proactively encouraged the learner if they gazed at the correct item before touching it or provided clues to guide the learner to the correct fruit if they gazed at the wrong item without touching it. With this observational study, we wanted to know if the two modalities can be combined in the current state of the technologies for children.

We begin by reviewing the relevant related work. We then describe the agent-based learning application we developed and the two user studies we conducted with it. Next, we report the results, followed by discussion of our key findings and observations.

2 Background Literature

2.1 Touch-based interaction for early learners

Aladé et al. [1] studied the pedagogical value of touchscreen technology for preschoolers and found that such interactive technology can facilitate and enhance learning compared to conventional mediums. Previous research has also investigated the challenges of using a touchscreen by young children. Hourcade [19] notes that children under the age of seven have underdeveloped fine motor skills and not yet well established hand preferences, and these affect how they use touch to interact. Vatavu et al. [38] studied touchscreen interaction for children between 3 to 6 years and found that children are significantly slower than adults and made more mistakes in performing simple touch interaction tasks. They partly attributed this to the under-developed finger dexterity and visuospatial processing abilities. Nacher [28] investigated the feasibility of multi-touch gesture for pre-kindergarten (2-3 years) children and found that some gestures such as double tap, long tap and two finger rotations are especially difficult to

perform. McKnight and Fitton [13] studied young children's abilities to use touchscreen and found that children are more prone to accidental or unintended touches.

Complementary input modalities such as gaze tracking are expected to be available in mobile devices in the future. Our work investigates the feasibility and usefulness of applying gaze as an input modality in an agent-based learning application.

2.2 Animated pedagogical agent-based learning application

Animated pedagogical agents are virtual characters designed to teach or guide users in an educational environment. While the pedagogical value of such agents are still debated [14], a well-established benefit of having a life-like tutoring agent is the engagement and motivational benefit it provides. Studies have shown that even the mere presence of a lifelike agent, even if not interactive, is capable of fostering interest and attention. Black et al. [4] studied how children converse with virtual agents and humans. They found that children tend to smile more, sit up and stay oriented towards the dyad when speaking to an agent rather than with humans.

Atkinson et al. [2] showed that the visual presence of the agent when combined with spoken instruction is more effective at promoting learning than an agent incapable of speaking. Krämer and Bente [23] note that improvement in non-verbal communication is required for agents to be more successful. While pedagogical agents were widely expected to transfer the benefits of human-human tutoring, the current generation of agents do not exhibit sophisticated non-verbal communication nor do they exert a social influence. Pedagogical agents that tailor their behaviour based on the emotional and cognitive state of the learner and show capacities for non-verbal communication, may have more pedagogical value [23].

Our study extends the previous knowledge on animated pedagogical agent. We designed an animated pedagogical agent capable of oral communication that is also aware of the visual attention of the learner. The agent exhibits emotional states, depending on whether the user is paying attention to the learning activity and, also uses the gaze information of the user to establish joint attention (agent looks where the child is looking), then guides the user based on the visual interest.

2.3 Gaze-aware virtual agents

Gaze awareness has been studied in virtual agents previously. Lahiri et al. [25] studied how gaze data of adolescents with autism, recorded while interacting with a virtual agent could be used to provide personalised feedback regarding appropriateness of gaze behaviour. Similarly, Ramloll et al. [32] developed a gaze contingent environment for fostering social attention in autistic children, using an avatar that rewards socially appropriate gaze behaviour of the user. Eichner et al. [9] investigated gaze awareness in virtual agents capable of giving oral presentations and found that agents that tailor the presentation based on interest and disinterest of the user provide a more natural interaction experience. Bee et al. [3] notes that agents that use the gaze information of the user to exhibit socially acceptable mutual gaze and gaze aversion behaviour are rated more positively. D'Mello et al. [7] studied gaze aware tutoring agents for young adults

and found that using audio prompts such as “*please pay attention*” was an effective way in dynamically orienting the user’s attention.

Previous research on gaze-aware virtual agents has shown they provide therapeutic benefits for children with autism, and can make the interaction more natural and effective for normal users. Our work complements the previous work in this by investigating the value of gaze-aware agents, exhibiting emotional response, for early learners.

3 Comparing Gaze-Aware Application with Baseline Touch-based Application

3.1 Design of the gaze-aware learning application

We designed a learning application with the “Little Bear”, a bear character, as the pedagogical agent. The application was designed to teach children the names of different everyday fruits and vegetables. The application was set in a garden-like 3D environment, where the agent would take the child for a walk. Different fruits and vegetables would appear on screen at pre-defined locations during the walk. The environment was chosen to provide a realistic context to the learning, similar to a human companion taking the child to a garden and teaching about the different items (see Figure 1). Once a fruit or vegetable appeared on screen, it could be activated by glancing at the fruit, making the agent speak an interesting detail about the fruit.

For gaze tracking, we used an off-the-shelf gaze tracker. Since accuracy of tracking maybe affected by the frequent movement of children, the gaze reactive areas were made slightly larger than the size of the fruit (see Figure 1). When the fruits were visible, background image was dimmed, so that the fruits are the most salient objects in the vicinity to reduce the probability of the child looking something in the background activating the fruit. Once the fruit was activated, a short animation was played where the fruits were scaled to 25% larger than their size then back to the original size and the agent also spoke an interesting detail about the fruit. The agent communicated orally using pre-recorded voice generated using the IBM Watson Text-to-Speech service. The parameters for speed of speech, pitch and, pauses between words were carefully chosen to make the speech feel natural, fitting to the character and easy to understand. The agent knew four different facts about each fruit (colour, size, where it grows etc.) and told one fact at a time in order, once the item was activated. Every fact included the name of the fruit (e.g. “*apple is red in colour*”). Once a fruit was activated, the agent speech lasted 3-5 seconds and during this window the application did not respond to any other gaze activations of fruits until the agent finished speaking.

In addition to speaking about the fruit or vegetable based on the current visual attention of the child, the agent also used the gaze information to reorient the attention of the child back to the learning activity when distracted. When the child is distracted



Fig. 1. Learning application using the Little Bear. The red semi-transparent boxes indicate the gaze interactive area. The application used in the baseline condition also had the same visual appearance and layout

and does not look at the tablet screen, the character becomes sad (see Figure 2a) and uses speech to attract attention by saying “*I become sad when you do not look at me*”.

Further, when the agent was not speaking, the head of the agent was oriented towards the direction the user is looking at (obtained from gaze tracker), giving an implicit feedback of gaze tracking and helping establish joint attention (see Figure 2 (b)-(e)). However, when the agent was speaking, the head of the character was oriented directly ahead, abiding by the established social conventions of eye contact during face to face conversation. To further improve the realism of the agent, the agent also exhibited realistic lip movement and blink behavior to complement the speech.

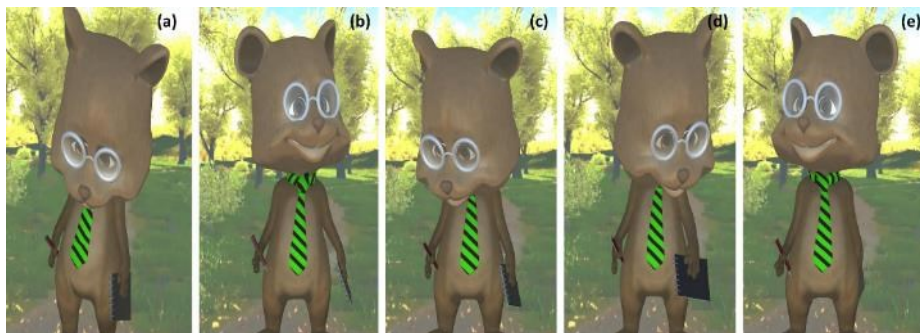


Fig. 2. Agent non-verbal behaviour in response to visual attention of the child. (a) A frame from the ‘sad’ animation when the child does not look at the screen. (b)-(e) head orientation of the bear changes based on the visual attention of the child (agent looks where the child is looking).

Agent response	Trigger in <i>LittleBear</i> condition	Trigger in <i>baseline</i> condition
Agent shows signs of sadness and utters “ <i>I become sad when you do not look at me</i> ”	User is distracted and looks away from the screen for 1.25sec.	Not possible
The agent acts happy and energetic.	User attends to the screen	Always
The agent orients its head in real-time to the direction the user is looking at.	User looks around different parts of the screen without activating any onscreen object	Not possible
The agent utters “ <i>I love the attention you are giving me, but did you notice the fruits we just found?</i> ”	User looks at the agent for a predefined duration (3.5sec) without looking at other fruits and vegetables on the screen.	User taps the agent
The agent speaks a detail about the item each time it is activated. E.g. name of the fruit, its colour, size etc.	User looks at the fruit or vegetable for 500ms.	User taps a fruit

Table 1. Description of various behavior of the agent-guided interaction for various modalities in the application.

3.2 Design of the baseline touch-based application

The baseline touch-based application used the identical learning context and visual appearance of the agent as in the gaze-aware application (See Figure 1) and the difference was only in the interaction mechanism. In the baseline condition, touching the on-screen fruit or vegetable activated it and the agent was not aware of attention of the child and always appeared happy with the head oriented directly ahead. We did not employ methods to predict to the attention using the frequency of touch activity (e.g. assuming that the child is not attending to the screen, if no touch activity is detected for a specific duration). Our application design was based on the affordances offered by modality, i.e., gaze has a strong association with attention, and looking away strongly indicates distraction, while a gap in touch interaction need not always mean lack of attention. Table 1 lists the verbal and non-verbal behaviour of the agent depending on the visual attention of the child.

3.3 User Study

The purpose of the study was to understand and compare the engagement and learning benefits provided by gaze awareness in agent-based learning application compared to the baseline touch-based application. We initially conducted a pilot study with 2 participants (1 male and 1 female) of the age group 4-5 years. The values for the dwell-duration for the character (3.5sec), duration to activate the fruits (500ms), the duration of time for which the user looked away from the screen before the character responded

visually (1.25sec) and verbally (6sec) with sad emotion were chosen based on the pilot evaluation. The chosen value for dwell duration for activating a fruit was short (500ms). The median fixation duration for children (4-6yrs) in free viewing task is 280ms [15] and it is also known that children have less ability to perform longer fixations at a target than adults [40]. We noticed that a longer dwell was considerably difficult for our pilot participants. Choosing a relatively short dwell time allowed our application to be implicitly reactive to the interest/visual attention of the child than requiring an explicit gaze action.

Participants.

We worked with Kidzee Kindergarten in Bangalore, India for the user study. We invited 12 children (5 females, 7 males) enrolled at the kindergarten to take part in the study. The participants were between the age group 4-5 years old and all participants had normal vision. English was not the first language of any of the participants and the participants were nominated by the teachers based on the English language proficiency. All the participants were familiar with mobile devices. The kindergarten itself has the policy of using learning applications on mobile devices, allowing every child to play with the devices once a week. Informed consent for the study was received from the teachers. No personally identifiable data of the child was collected or maintained.

Design.

We chose a within-subject design. The two experimental conditions were labelled as follows: *LittleBear*: participants used the gaze-aware learning application with the Little Bear agent and used gaze input only and *Baseline*: participants interacted using touch input only (see Table 1).

We constructed two lists of 20 fruits and vegetables each for the test. Each list was constructed such that both had approximately the same number of everyday and uncommon fruits, making it safe to assume that participants were equally likely to know the names of the same number of fruits in both the lists before the test. The two lists were used in the application for the two different conditions. The order of the condition and the list for the condition were counter-balanced.

There were three dependent measures: (i) the duration of time participants interacted with the application, (ii) the number of times on average the participants activated each fruit, and (iii) the short-term retention of the vocabulary soon after the interaction as measured using a paper-based evaluation at the end of the session. For the paper-based evaluation, we showed the participant a paper with pictures of four fruits each and asked them to point at a specific fruit (e.g. *can you show me the fruit banana?*). The evaluation consisted of 10 such questions for each condition and was conducted soon after the interaction for the condition.

To test for differences between the *LittleBear* and *Baseline* conditions we used a non-parametric pair-wise (Monte Carlo) randomisation test [8]. In the randomisation test, the null hypothesis is that pair-wise differences are equally likely to be positive or negative. Repeated resampling ($n=10,000$) with a random assignment of signs for the

difference between the conditions gives us a sampling distribution of the mean difference. The observed mean difference is compared to the sampling distribution to estimate how likely the observed difference is by chance.

Apparatus.

The learning application was developed using Unity 5.3 on a Microsoft surface pro 4 touchscreen tablet. We used a Tobii EyeX gaze tracker to record the gaze information. The tracker was mounted at the bottom of the tablet. See Figure 3(a) for the experimental set up.



Fig. 3. (a) Experimental setup. (b) A frame from the interactive introduction of the Little Bear gaze-aware agent. The green dot indicates the current gaze point of the user.

Procedure.

The test was conducted in the Kidzee kindergarten premises. The participants were seated facing the tablet, which was positioned on a table in normal viewing position, approximately 40cm away (similar to situations when using tables with special holders for mobile devices). In the baseline condition, the participants were free to lean forward or re-position themselves as they felt comfortable. The gaze tracker was first calibrated. We used the built-in Tobii EyeX “pop the dots” calibration process. It used seven calibration points. The visual stimuli used for the calibration were coloured dots, which burst when the user looked at one for a specific duration. We used the built-in calibration procedure because we felt the playful element in the calibration procedure was well-suited for children.

Before the *LittleBear* condition, a short interactive introduction was shown to the participant. In the introduction material, the character spoke “*I will tell you a secret. I have eyes and I know where you are looking*”. We used the metaphor that the agent has eyes to convey the gaze awareness feature of the application. In the interactive introduction, the child could see the smoothed gaze point of where they were looking and the character would also look at the same area (shown in figure 3(b)). The bear also showed the emotions of sadness and boredom when the participant did not look at the screen. The moderator encouraged the participants to look away from the screen and observe the response of the character (as shown in Table 1). After the interactive character introduction, the learning application was started.

In both the conditions, the character introduced the environment as a garden with a lot of fruits and vegetables and encouraged the participant to walk with him to find interesting items. The character then turned around and started walking in the garden, where we used background animation and music to give a perception of a virtual tour around the garden. At different points during the walk, four different fruits/vegetables appeared at the sides of the screen (see Figure 1). The character encouraged the participant to explore the fruits by saying, for example, “*we just found more fruits, do you know what these are?*”.

In total, there were five phases for each condition and each phase had four items appear onscreen (altogether 20 fruits/vegetables). In the first phase, the moderator encouraged the participant to activate the different onscreen fruits (by touching them or dwelling at them by using gaze) and this phase was treated as practice and excluded from all the analysis. For each condition, the participant was encouraged to interact with the different onscreen fruits and vegetables, listen to what the agent said and repeat after the character. The application did not require the participants to activate any fruits to go to the next phase. When the participant wanted to go to the next location they had to press the “NEXT” button (see Figure 1) on the screen, which made the character walk to a new location and four new items then appeared on screen, starting the next phase.

After each session, we used a five level smiley-meter [41] to record participant response along with qualitative feedback. We used the smiley-meter with two questions: “*Did you like the game you played?*” and “*Did you like the bear character?*” and the participants had to respond by selecting an appropriate smiley face from the smiley-meter shown on paper. In addition, we also logged the interaction events for further analysis.

Results.

Average time spent per phase.

Figure 4(a) shows the time spent per phase interacting with the application for the two conditions. This study was about free exploration and the application did not have any restriction on how many times user had to activate each fruit. The extent of time spent on each phase was hence completely controlled by the child in both conditions. Therefore, we can assume that the time spent in each of the phases in different conditions reflects the overall engagement. The median value indicates that our participants spent almost twice as much time interacting with the application in the *LittleBear* condition than baseline. The difference was found to be statistically significant using the pair-wise randomisation test ($p=0.02$).

Average activations per fruit.

Figure 4(b) shows the average number of activations per fruit for the two conditions. The median value indicates that participants activated each fruit almost twice as often in the *LittleBear* condition as in the baseline condition and hence indulged in more learning activity. The difference was found to be statistically significant using the pair-wise randomisation test ($p=0.001$).

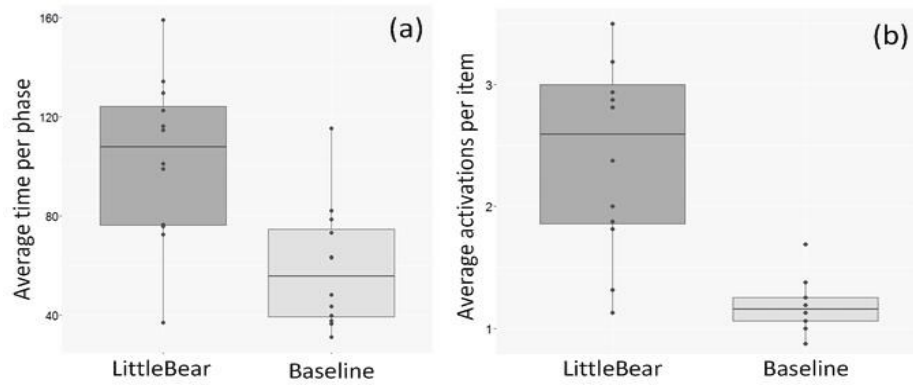


Fig. 4. (a) Average time (in seconds) spent per phase for the two conditions. (b) Average number of activations per fruits for the two conditions

Number of items retained after the interaction.

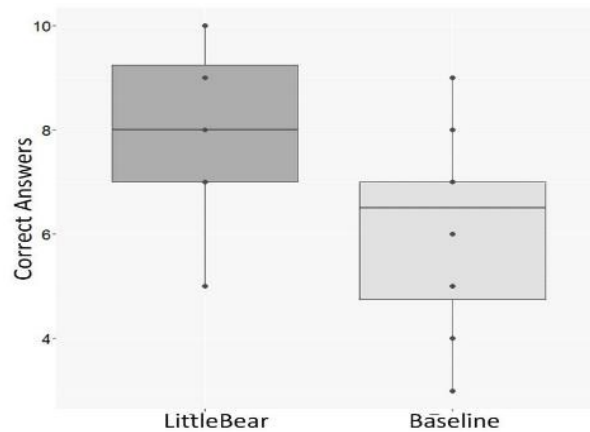


Fig. 5. Boxplot showing the number of fruits the participants recalled correctly in the paper-based evaluation following the interaction. There were in total 10 questions and each question had four different alternatives.

Figure 5 shows the boxplot for the number of correct answers in the paper-based evaluation following each condition. This evaluation was used as a measure of the *short-term retention* of the vocabulary, following the interaction. The median values indicate that our participants retained about 23% more names of fruits and vegetables after using the gaze-based learning application when compared to touch-based interaction. The difference was found to be statistically significant using the pair-wise randomisation test ($p=0.03$)

Smiley-Meter Data

We did not notice any difference between the Smiley-Meter data for the questions “*Did you like the game?*” and “*Did you like the little bear character?*” for the two conditions. Both the conditions were rated very highly (4 or 5 out of the 5 levels) for both the questions. Randomisation test showed that the differences were not statistically significant ($p > 0.05$).

Other Observations: Children interacting with gaze-aware Agent

Using the agent’s emotion to regain the child’s visual attention was a useful technique. All the participants reported that they did not want to make the character sad or tried their best to keep the character happy. This proved helpful in two ways: first, to attract the participant’s attention when they were distracted or looking away from the screen and second, the participants could successfully tell when the agent could not see them (i.e. gaze tracker could not track their eyes) as the bear appeared sad (e.g. when the child moved out of the view of the tracker), and hence re-positioned themselves, making the agent happy again.

Often, our participants would be distracted by external factors, e.g. by the objects or noise in the environment. However, this made the agent appear sad and was a motivator for them to return their attention to the application. This was also clear in two separate scenarios when two of the participants were casually interacting with the application after the test, in the presence of other children observing them. The users specifically wanted the observers to leave the room as their presence and conversations were distracting the user, which in turn was making the character appear sad.

Further, our participants sometimes made significant movements of their head and upper body while interacting with the application. This, at times, meant that they moved out of the view of the tracker, which again made the character appear sad. In the beginning phase of the study this sometimes caused confusion. For example, one of the participants commented “*why cannot he see me, I am looking at him?*” and another participant leaned very close to the display and commented “*he still cannot see me, my bear has bad eyes*”. These situations required moderator assistance to help the participants re-position themselves such that they were in the view of the gaze tracker. However, the required moderator assistance reduced in the later phases of the study and the participants could work independently. In contrast, the children worked almost always independently in the baseline condition.

Also, the notion of an agent “with eyes and emotion” was received positively and sparked playful interactions where the children covered their faces, deliberately moved away so that the bear could not see them anymore, or tried to cover the eyes of the bear on the screen with their hands. When asked why they were “making the character sad”, in all cases, children reported that they were simply playing “hide and seek” with the character. Overall, gaze-based interaction was well received. During the test, children did not show any signs of confusion or not able to control the application with the eyes due to Midas Touch. They could correctly associate the fruit that they looked and the

spoken description of the agent. Figure 6 shows a representative gaze distribution pattern for the participant P11, interacting with the Little Bear.

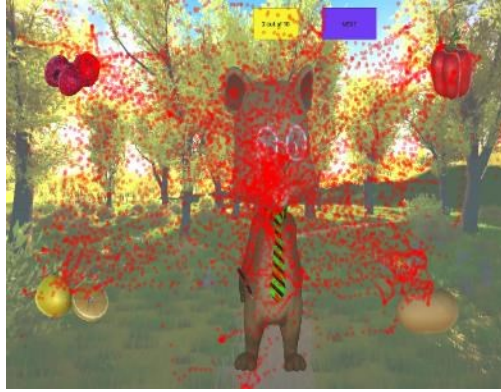


Fig. 6. An example gaze distribution pattern for the LittleBear condition

4 Feasibility of Combining Gaze and Touch for Children

Next, we wanted to understand the feasibility of combining implicit gaze-input and explicit touch input for children in their early learning years. There are numerous previous works that have investigated the combination of interaction techniques for adults [36, 44]. While touch is a familiar, and the primary, input modality in mobile devices, gaze provides a wealth of information from a pedagogy point of view. Gaze could be used as a cue to evaluate common ground in conversational interfaces [8], be useful to predict intentions to provide proactive guidance (like in our Study-2), as a reliable way to evaluate comprehension [11] etc. Hence, gaze and touch is a valuable combination of input modalities for early learners. We extended the gaze-aware agent application that we used in the previous comparative evaluation. The main difference was that instead of allowing the users to freely explore the fruits and vegetables, the agent asked the user a question (e.g. *can you show me the fruit banana?*) and the user had to touch the correct fruit, out of the four on-screen alternatives. The agent implicitly used the gaze information to understand if the user was confused or unsure about the answer. For example, if the user glanced at the wrong fruit for 500ms, the agent would proactively tell the user how the fruit they were looking at is different in terms of size and colour from the fruit they need to find (e.g. *“you are looking at apple, apple is red in colour, while banana is yellow in colour”*) or if the user looked at the correct fruit while being hesitant to touch it, the agent would encourage the user to touch it (e.g. *“Yes, that is the one, go ahead and touch it!”*). When the user touched the wrong fruit, the agent gave the same response, comparing the fruit that was touched with the fruit they needed to find. If the user touched the correct fruit, the agent thanked the user for helping him find the fruit and danced for a few seconds, before moving to the next location where the user had to help find a different fruit.

We felt that with such gaze-based proactive intervention could potentially reduce the cognitive load associated with selecting a choice, and even possibly enhance learning [36]. The focus of the study was to observe how children would interact with an application that combines implicit gaze-input with explicit touch input and understand the real-world feasibility of combining gaze and touch input, in the current state of the technology, in an early childhood learning application.

4.1 User Study

We used the same participants and same experimental set up as in the previous study. The participants completed 4 onscreen questions and each question had four onscreen alternatives (same as Figure 1). We used the fruits/vegetables which the participants had incorrectly answered in the paper-based evaluation of Study-1. In a few cases, the participants had less than 4 incorrect answers in paper-based evaluation after Study-1. In those cases, we used a predefined set of fruits which were not part of the evaluation to be used in the Study-2. The task for the participants was to find the specific fruit asked for by the agent among other onscreen fruits and touch it.

Results and Observations.

As we had anticipated, almost all participants had one or more questions about which they were noticeably unsure or had selected the wrong answer at first. However, gaze-based guidance was rarely invoked due to issues with gaze tracking. Figure 7 shows the gaze data loss for each participant. The gaze data loss was calculated based on the number of valid samples expected, based on the approximate sampling rate of the tracker and the number of actual valid samples returned by the tracker.

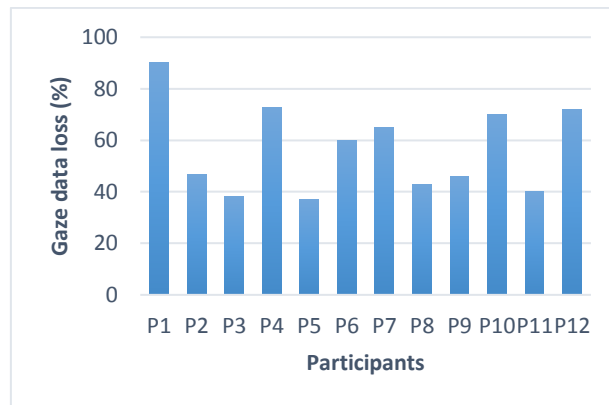


Fig. 7. Percentage of time gaze data could not be tracked for the participants.

There were two separate issues that caused problems in gaze tracking:

1. Short arms of the participants meant that, as soon as a question was asked by the agent, they leaned forward to touch the screen and the distance from the screen was insufficient for the gaze tracker to work. Optimal working distance of Tobii EyeX tracker is 40cm-90cm.
2. In our study, the tracker was positioned at the bottom part of the tablet. This meant that normally when the children used their hands to touch the screen, the hands would obstruct the gaze tracking sensor causing tracking issues.

There were also differences between how participants interacted with the application. E.g. P1 was always close to the tablet making gaze tracking difficult almost all the time, while P3 and P4 moved closer when the interaction required touch activation. For all participants, there were problems in gaze tracking just before and while they were touching the screen for interaction. In the discussion section, we reflect on the technological improvements required to make the combination of gaze and touch work smoothly for early learners.

5 Discussion

We investigated the value of gaze awareness in an agent-based learning application for children. Before the study, we did not know if the gaze input is feasible; how children would react to the notion of an "agent with eyes"; nor how it compares against baseline touch-based interaction. Our results from study-1, suggest that an application that uses the attention information of the children to tailor the agent's verbal and non-verbal behavior may lead to better engagement, help develop a closer connection with the agent and provide learning benefits than a baseline touch-based application, that is unaware of the attention of the user.

All the participants in our experiment (study-1) completed both the conditions successfully. While we did encounter challenges with continuously gaze tracking the user when the children moved a lot, we did not face any issues with gaze tracker calibration and accuracy of tracking. Our study shows the feasibility of gaze-based interaction in early childhood learning applications. Our participants also reacted very positively to the concept of a gaze-aware pedagogical agent or "bear with eyes and emotions". Our results from Study-1 suggest that such gaze-aware applications are not only engaging but could also help motivate the children to attend to the learning material, and prove beneficial in learning as indicated by our results on improved short-term retention. In the *LittleBear* condition, our participants developed a closer emotional connection with the agent and all the participants reported that they did not want to make the agent sad. The agent, who became sad when not looked at, may have motivated the participants to attend more to the on-screen items.

There are two different factors that could have jointly contributed to the improved overall engagement and performance in the paper-based evaluation in the *LittleBear* application. First is the input modality that was used to activate the fruits. The Little

Bear used the visual attention of the user to tailor its response and spoke about the fruits and vegetables that were of current interest to the user. Eichner et al. [9] showed that an agent that tailor its verbal response based on the attention of the user improves the overall engagement in adult users. In some situations, interacting with the application may have been easier with gaze than using explicit touch actions; for example, when the user wanted to activate the same fruits multiple times. While interacting with the Little Bear, children often exhibited a gaze behavior where they kept looking at the same fruit even after the agent had finished speaking a detail for that fruit. A cue that can be considered as if the child "wants to know more" about that fruit. This meant that the agent continued explaining the next detail about the same fruit, as opposed to the baseline condition which required another explicit touch action for the next activation. The fact that our participants spent more time using the application in the LittleBear condition could have influenced how well they performed in the paper-based evaluation.

Second, Little bear also exhibited feelings of sadness when the child looked away, as a mean to reorient the attention of the user back to the learning activity. D'Mello et al. [7] have shown that using simple audio prompts to regain the attention of a distracted user to a learning application could produce significant learning gains, even at deeper levels of comprehension. This emotional response may have motivated our participants to focus on the learning activity and this additional engagement possibility may also be one of the reasons the users interacted longer with the Little bear compared to the baseline condition. Overall, our results show the strong potential of gaze-based educational applications for early learners.

Our participants rated both the conditions very highly in the smiley-meter based questions. There could be two plausible explanations for this. First, it could be because our participants did not have any strong preference between the two conditions and enjoyed using both equally well. Touch-based interaction used in the baseline condition was a familiar interaction technique for our participants. While our participants were noticeably more engaged interacting with the Little Bear using gaze, they were also noticeably more confident and independent in the baseline condition. Second, it could have been a problem with our methodology of getting feedback from the participants. Previous research had also reported similar results using a smiley-meter with rural Indian children of age group 6-7 years. Kam et al. [21] note that their participants chose smiling faces, simply because they were more aesthetically appealing than frowns. We keep this as an important area to address in future studies.

Further, we also investigated the feasibility of combining gaze and touch-based interaction for early learners. Both gaze and touch modalities worked well when used as the sole interaction techniques, however, we faced unexpected technical challenges relating to gaze tracking robustness when using the combination of gaze and touch to interact. The two main reasons for the tracking issue was the sub-optimal working distance of gaze trackers, and hands obstructing the tracker view while interacting by using touch. In the current state of the gaze tracking technology, it seems difficult to combine gaze-input with touch interaction in mobile devices for early learners. Technological improvements are needed before the combination would work seamlessly. In the next

section, we discuss the technological improvements required for gaze tracking to be useful for field applications for children.

5.1 Technological improvement required in gaze trackers

When using touch-based interaction, children position the mobile device very close to them, to perform easy touch actions. This distance is often less than the optimal working distance of commercial gaze trackers. Unlike adult users, children are even more restless and move a lot while interacting with technological devices. It would hence be beneficial for gaze trackers to have a larger range of operating distance and a larger field of view to efficiently track children.

Further, almost all gaze trackers are mounted below the screen, possibly to avoid the occlusion of the eyes by the eyelid and get a clearer image of the pupil. However, when mounted below the screen, frequent touch-based operations would cause the hands to obstruct the view of the tracker. While this is also an issue with older users [30], it is aggravated in the case of young users, as they do not show any specific preference of hands when touching the screen [19], and as we noticed in our study, sometimes even use both hands together to point at and touch an item. A potential solution to this would be to mount the camera above the screen or use multiple gaze tracking cameras to improve the accuracy and robustness of tracking.

Further, collaborative learning is common and encouraged in early childhood pedagogical environment (e.g. two children sharing a tablet). In our study, two of our participants wanted the few curious onlookers to leave the room because they were distracting the participant, which in turn was making the agent sad. While this shows that our approach of using agent emotion to reorient the child's attention may not be conducive to collaborative learning, this is also an inherent problem with gaze trackers. Currently, commercial gaze trackers do not support tracking multiple users simultaneously. It would hence be beneficial for gaze tracking devices, that are meant to be applied in early childhood learning environments, to support the inherent pedagogical practices of collaborative learning and playing.

5.2 Limitations and future work

We have identified five limitations in our study. First, the study was conducted by a trained external moderator and the reported results are based on short-term evaluation. There is growing understanding in the Child-Computer Interaction research community for the need of long-term evaluation of technologies in naturalistic settings. The presence of the moderator, who is unfamiliar to the children may have influenced how long they interacted with the systems. Novelty of the gaze-based interaction may have influenced the engagement. Future work will investigate the long term engagement and pedagogical value of gaze-aware learning applications, for example, by training the teachers to help in the data collection [33].

Second, the task used in our study-1 involved only simple selection/activation of fruits (with either gaze in the *LittleBear* condition or touch in the *baseline* condition). Using gaze as the sole input modality may only be suitable for such simple interaction

tasks. Other input modalities such as touch/tilting the device affords more complex interaction possibilities. The purpose of the study was to demonstrate the engagement possibility and learning benefit offered by gaze awareness in an agent-based application for children and further research is required to understand how this gaze awareness can be best combined with other expressive interaction.

Third, we placed the device in a viewing position. In study-1, baseline condition, the children were free to lean forward or re-position themselves as they felt comfortable. However, in a naturalistic setting children may position the device differently (e.g. on the lap or flat on the table). The positioning of the device may have an influence on how children interact with the device using touch [35]. Further research is required to understand the effect of device positioning on our results.

Fourth, the small sample size (N=12) is a limitation of the study. While many of the reported results were statistically significant, a larger number of participants could have been beneficial to understand any difference based on gender of the participant, or familiarity of mobile devices.

Finally, all our participants were between the age group of 4-5 years old and from an urban community. Age and culture may have an influence on how people respond to novel technologies, pedagogical agents, or the acceptable social gaze behaviour [24] with the agent. Future work should also investigate if our results could be applicable to children of different age groups and cultural representation.

6 Conclusion

We investigated how children of the age group 4-5 years interact and engage with a gaze-aware pedagogical agent that tailors both its verbal and non-verbal behaviour to the visual attention of the user. Our results suggest that gaze is a promising input modality for early learners. Gaze awareness in the agent-based application not only improved the children's engagement with the application, but also improved their short-term retention of the vocabulary learnt, and helped develop a closer connection with the agent. We further investigated the feasibility of combining implicit gaze and explicit touch input in a learning application. While the combination of modality has strong potential in applications for children, our results suggests that several technical improvements are required for this combination to work seamlessly for children.

Acknowledgement

We thank Nathalie Henry Riche for her valuable inputs while shepherding this paper. We also thank the staff members at Kidzee Kindergarten, Domlur, Bangalore for their support to conduct the study.

References

1. Aladé F, Lauricella AR, Beaudoin-Ryan L, Wartella E. Measuring with Murray: Touchscreen technology and preschoolers' STEM learning. *Computers in Human Behavior*. 62:433-41 (2016).
2. Atkinson RK. Optimizing learning from examples using animated pedagogical agents. *Journal of Educational Psychology*. 94(2):416 (2002).
3. Bee N, Wagner J, André E, Charles F, Pizzi D, Cavazza M. Interacting with a gaze-aware virtual character. In *Proceedings of the 2010 workshop on Eye gaze in intelligent human machine interaction*, pp. 71-77. ACM (2010)
4. Black M, Chang J, Chang J, Narayanan S. Comparison of child-human and child-computer interactions based on manual annotations. In *Proceedings of the 2nd Workshop on Child, Computer and Interaction*, p. 2. ACM. (2009)
5. Borgestig M, Sandqvist J, Parsons R, Falkmer T, Hemmingsson H. Eye gaze performance for children with severe physical impairments using gaze-based assistive technology: A longitudinal study. *Assistive technology*. 28(2):93-102 (2016).
6. Chipman G, Druin A, Beer D, Fails JA, Guha ML, Simms S. A case study of tangible flags: a collaborative technology to enhance field trips. In *Proceedings of conference on Interaction design and children*, pp. 1-8. ACM (2006).
7. D'Mello S, Olney A, Williams C, Hays P. Gaze tutor: A gaze-reactive intelligent tutoring system. *International Journal of human-computer studies*. 70(5):377-98 (2012).
8. Dugard P. Randomization tests: A new gold standard? In *Journal of Contextual Behavioral Science* 3(1) (2014).
9. Eichner T, Prendinger H, André E, Ishizuka M. Attentive presentation agents. In *International Workshop on Intelligent Virtual Agents*, pp. 283-295. Springer (2007).
10. Evans MA, Drechsel E, Woods E, Cui G. Multi-touch tabletop computing for early childhood mathematics: 3D interaction with tangible user interfaces. In *Proceedings of the 9th International Conference of the Learning Sciences*, pp. 274-275 (2010).
11. Feng G. Eye Tracking: A brief guide for developmental researchers. *Journal of Cognition and Development*. 12(1):1-1 (2011).
12. Fisher AV, Godwin KE, Seltman H. Visual environment, attention allocation, and learning in young children when too much of a good thing may be bad. *Psychological science*. 1;25(7):1362-70 (2014).
13. McKnight L, Fitton D. Touch-screen technology for children: giving the right instructions and getting the right responses. In *Proceedings of the 9th international conference on interaction design and children*, pp. 238-241. ACM (2010).
14. Heidig S, Clarebout G. Do pedagogical agents make a difference to student motivation and learning?. *Educational Research Review*. 6(1):27-54 (2011).
15. Helo A, Pannasch S, Sirri L, Rämä P. The maturation of eye movement behavior: Scene viewing characteristics in children and adults. *Vision research*. 103:83-91 (2014).
16. Hiniker A, Sobel K, Hong SR, Suh H, Kim D, Kientz JA. Touchscreen prompts for preschoolers: designing developmentally appropriate techniques for teaching young children to perform gestures. In *Proceedings of International Conference on Interaction Design and Children*, pp. 109-118. ACM (2015).

17. Homer BD, Kinzer CK, Plass JL, Letourneau SM, Hoffman D, Bromley M, Hayward EO, Turkey S, Kornak Y. Moved to learn: The effects of interactivity in a Kinect-based literacy game for beginning readers. *Computers & Education*.74:37-49 (2014).
18. Hornof AJ, Cavender A. EyeDraw: enabling children with severe motor impairments to draw with their eyes. In *Proceedings of the SIGCHI conference on Human factors in computing systems*. pp. 161-170. ACM (2005).
19. Hourcade JP. Interaction design and children. *Foundations and Trends in Human-Computer Interaction*.1(4):277-392 (2008).
20. Gottlieb J, Oudeyer PY, Lopes M, Baranes A. Information-seeking, curiosity, and attention: computational and neural mechanisms. *Trends in cognitive sciences*. 17(11):585-93 (2013).
21. Kam M, Rudraraju V, Tewari A, Canny J. Mobile gaming with children in rural India: Contextual factors in the use of game design patterns. In *Proceedings of 3rd Digital Games Research Association International Conference* (2007).
22. Kangas J, Akkil D, Rantala J, Isokoski P, Majaranta P, Raisamo R. Gaze gestures and haptic feedback in mobile devices. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* pp. 435-438. ACM (2014).
23. Krämer NC, Bente G. Personalizing e-learning. The social effects of pedagogical agents. *Educational Psychology Review*. 22(1):71-87 (2010).
24. LaFrance M, Mayo C. Cultural aspects of nonverbal communication. *International Journal of Intercultural Relations*. 2(1):71-89 (1978).
25. Lahiri U, Warren Z, Sarkar N. Design of a gaze-sensitive virtual social interactive system for children with autism. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*.4:443-52 (2014).
26. Lu C, Frye D. Mastering the machine: A comparison of the mouse and touch screen for children's use of computers. In *International Conference on Computer Assisted Learning*, pp. 417-427. Springer (1992).
27. Luna B. Developmental changes in cognitive control through adolescence. *Advances in child development and behavior*. 37:233-78(2009).
28. Nacher V, Jaen J, Navarro E, Catala A, González P. Multi-touch gestures for pre-kindergarten children. *International Journal of Human-Computer Studies*. 73:37-51 (2015).
29. Ovaska S, Hietala P, Kangassalo M. Electronic whiteboard in kindergarten: Opportunities and requirements. In *Proceedings of the 2003 conference on Interaction design and children*, pp. 15-22. ACM (2003).
30. Pfeuffer K, Alexander J, Chong MK, Gellersen H. Gaze-touch: combining gaze with multi-touch for interaction on the same surface. In *Proceedings of the 27th annual ACM symposium on User interface software and technology*, pp. 509-518, ACM (2014).
31. Plowman L, McPake J. Seven myths about young children and technology. *Childhood Education*. 89(1):27-33 (2013).
32. Ramloll R, Trepagnier C, Sebrechts M, Finkelmeyer A. A gaze contingent environment for fostering social attention in autistic children. In *Proceedings of the 2004 symposium on Eye tracking research & applications*, pp. 19-26. ACM (2004).
33. Robertson J, Macvean A, Howland K. Robust evaluation for a maturing field: the train the teacher method. *International Journal of Child-Computer Interaction*. 1(2):50-60 (2013).
34. Rogers Y, Price S, Randell C, Fraser DS, Weal M, Fitzpatrick G. Ubi-learning integrates indoor and outdoor experiences. *Communications of the ACM*.48(1):55-9 (2005).
35. Romeo G, Edwards S, McNamara S, Walker I, Zигuras C. Touching the screen: Issues related to the use of touchscreen technology in early childhood education. *British Journal of Educational Technology*. 34(3):329-39 (2003).

36. Schroeder EL, Kirkorian HL. When Seeing Is Better than Doing: Preschoolers' Transfer of STEM Skills Using Touchscreen Games. *Frontiers in Psychology* (2016).
37. Sibert JL, Gokturk M, Lavine RA. The reading assistant: eye gaze triggered auditory prompting for reading remediation. In *Proceedings of the 13th annual ACM symposium on User interface software and technology*, pp. 101-107. ACM (2000).
38. Vatavu RD, Cramariuc G, Schipor DM. Touch interaction for children aged 3 to 6 years: Experimental findings and relationship to motor skills. *International Journal of Human-Computer Studies*. 74:54-76 (2015).
39. Vidal M, Turner J, Bulling A, Gellersen H. Wearable eye tracking for mental health monitoring. *Computer Communications*.35(11):1306-11 (2011).
40. Ygge J, Aring E, Han Y, Bolzani R, Hellström A. Fixation stability in normal children. *Annals of the New York Academy of Sciences*. 1039(1):480-3 (2005).
41. Zaman B, Abeele VV, De Grooff D. Measuring product liking in preschool children: An evaluation of the Smileyometer and This or That methods. *International Journal of Child-Computer Interaction*. 1(2):61-70 (2013).