Speech Cueing on the Web by ‘The Little Dude’: Multimedia Instruction for Young Children

Bruce L. Mann, Memorial University, Canada
Henry Schulz, Memorial University, Canada
Jianping Cui, Bow Valley College, Canada

ABSTRACT
There is a real need for studies on learning from multimedia with school-age children, even pre-school children. In this research, temporal speech cueing was proposed to help young children as they listened to a speaking pedagogical agent direct their attention to details in on-screen text and graphics. An experiment was conducted with 4th and 5th graders (n = 133) who read on-screen text, and listened to cues presented by a pedagogical agent. Results showed that children in the speech cueing group out-performed those in the on-screen text group in immediate and delayed post-tests. Agent movement had no effect. Implications are discussed for helping young children to learn from the on-screen text presented in contemporary educational multimedia.

Keywords: Agent, Attention, Auditory, Cue, Modality, Multimedia, Sound, Structured Sound Function (SSF)

INTRODUCTION
Research with young children on learning from multimedia is scarce, which is surprising given the cartoon-like quality of the materials. We believe that young children would be as accustomed, and perhaps even more accustomed than undergraduate students to learn from multimedia and cartoon-like characters (i.e., agents), and would therefore be less likely to experience novelty effects from speech cues presented by a pedagogical agent.

DOl: 10.4018/ijopcd.2012070103
they spend much of their day in school, have not yet been streamlined by their abilities nor would they have to devote most of their class time to intensive curricular study (Case, 1985). Most beginning readers rely on context to read (Goldsmith-Phillips, 1989), showing a heavy reliance on contextual facilitation of word perception and are less adept at contextual facilitation of comprehension than children in the higher grades, in accordance with Stanovitch’s (1980) interactive-compensatory hypothesis.

LITERATURE REVIEW

Persistent Problem

Some students experience a persistent problem when trying to learn difficult or unfamiliar content from on-screen text. Although multimedia is pleasing to them, their enjoyment is either uncorrelated or negatively correlated with learning (Clark & Feldon, 2005). There are at least two reasons why young children will ignore or forget to read instructions and feedback in educational multimedia.

Reason 1: Inadequate Attentional Control

One reason is their inability to adequately control their attention during multimedia learning. Unlike most entertainment multimedia, educational multimedia requires active listening and reading instructions and feedback presented in the program or website. Reading instructions and feedback requires mental articulation of that instruction, feedback, hint, program direction or map coordinate. A student reading text within a multimedia environment must be able to mentally articulate their own version of the meaning in the text, or compensate. To engage the appropriate attentive state, a student must self-initiate the appropriate system of information processing (Borich & Tombari, 1995). The student must consciously focus their attention “to bind the separate features of a stimulus—such as the colour, shape, words, into a unitary object” (Matlin, 2002, p. 57).

Reason 2: Under-Developed Phonological Loop

A second reason that accounts for the problem in trying to learn difficult or unfamiliar content from on-screen text, is their insufficient mental articulation of spoken and screen text due to an under-developed phonological loop. The phonological loop is critical for learning difficult or unfamiliar content from on-screen text. The loop plays a crucial role in learning syntax as well as in the acquisition of the phonological form of lexical items. “The loop system mediates the acquisition of syntactic knowledge, as well as the learning of individual words… not to remember familiar words, but to help learn new words (Baddeley, Gathercole, & Papagno, 1998, p. 158, p. 166). Figure 1 illustrates the structure and process of learning difficult or unfamiliar content from on-screen text, according to the attentional control theory of multimedia learning.

Adults and adolescents can hear a sound presented in multimedia and encode the gist directly into their phonological store, encoding the details indirectly through their articulatory loop. Their analysis of difficult or unfamiliar on-screen text is fed into the phonological store by means of sub-vocal speech using an articulatory system, like an inner voice to an inner ear (Baddeley, Gathercole, & Papagno, 1998). Young children are not fully capable of mentally articulating difficult or unfamiliar content presented in text. Their auditory memory consists of a phonological store without a phonological loop (Gathercole, Pickering, Ambridge, & Wearing, 2004). Unarticulated material in young children is analogous to extraneous cognitive load reported in adults (Kalyuga, Chandler, & Sweller, 1999). They must rely on the context. Reading context decreases as a function of reading development and ability (Goldsmith-Phillips, 1989; Swantes, 1991).

Proposed Solution

In this research temporal convergent sound cues were proposed to solve the problem by
providing a reading context for the visual events in multimedia. Sound can evoke responses in visual areas of the brain in very young children (Goswami, 2004). Memory for material presented in sound is more durable and resistant to interference from other modalities than visually presented material and more resilient to forgetting than visual traces (Broadbent, Vines, & Broadbent, 1978). Reports like these about the durability of sound and its resistance to interference and forgetting provide support for including sound in multimedia (Mann, 2008). Yet sound alone is insufficient to learn from multimedia, hence the need for purposeful advice on how to enhance learning from technology with sound. A recent review of purposeful advice reported eight different guidelines for designing educational multimedia, each one differing from the others in its scope or depth of advice, and carrying a different assumption about how people learn from multimedia (Mann, 2008). The assumption in this research was that learning from multimedia for young children begins when they hear a speech cue that alerts, cautions, warns or directs their attention to a future event, or reminds them of a past event displayed, as a visual event in the computer program or Internet site.

In this research the pedagogical agent was an appealing computer-generated cartoon character that moved around the screen, gesturing and cueing students in speech or bubble text. The pedagogical agent spoke only the gist, not explained or narrated, leaving the details in text and illustrations. The pedagogical agent was an appealing computer-generated cartoon character that moved around the screen, gesturing and speaking to students in speech or bubble text. The agent’s audio script took the specific purpose of directing attention to the
Details in the text, graphic or animation. The agent’s audio script took the specific purpose of directing attention to the details in the text, graphic or animation. In selecting cues for the children, a convergent goal was used with continuous constancy (Mann, 2008). Hints and reminders frequently cued students to take action with the information presented in the visual event, usually requesting that the student write something on paper or type some text online. In this study, our expectations were that:

1. Children reading with a pedagogical agent that speaks in temporal speech cues would outperform those with a non-speaking agent reading temporal cues in on-screen text.
2. Children reading with a moving pedagogical agent would outperform those reading with a stationary agent.

**METHOD**

**Participants**

Participants were 133 enthusiastic 4th and 5th grade students (aged 9-12 years) enrolled at an urban public school in Canada that had a consent form that had been signed by a parent or guardian, from an initial group of 162. “Children at this level are eager to cooperate in experiments and are able to persist at them for long periods of time” (Case, 1985, p. 182). This school was selected for this study for its history of below-average literacy (Maguire, 2001). Below-average literacy was a precondition for this study, in accordance with the individual differences principle and to prevent the possibility of expertise reversal effect. The individual differences principle describes a learning outcome with university students wherein design effects are strongest for low prior knowledge learners and hardly present for high prior knowledge students (Mayer, 2001). Expertise reversal effect describes a learning outcome in which an instructional technique that is effective with below-average students loses some of its effectiveness and even becomes ineffective with average and above average students (Paas, Renkl, & Sweller, 2003). None of the participants had been introduced to singular and plural ownership or the proper usage of apostrophe prior to or during the experiment, and although they had been exposed to educational multimedia in the school setting, none of these children had seen a pedagogical agent.

**Research Materials**

Two of the school’s best-equipped computer labs were made available for the experiment. One lab was set-up for students in the MT and NMT conditions to watch a pedagogical agent called “Genie”, and read the on-screen instruction and feedback involving the proper usage of apostrophe to show singular and plural ownership. The other computer lab was configured for students in the MS and NMS conditions to watch agent “Genie” and to listen to equivalent instruction and feedback through headphones. In both labs, and prior to administration of the treatment, the investigator used a big screen television connected to a computer to demonstrate how students should navigate the program.

An educational website was developed and the files uploaded to an independent server, to prevent accidental prior access to the content by the students. The content of the educational website was a guided tutorial on two rules regarding possessive expression with apostrophes, namely, to express singular and plural possession. The language on the website was child-directed and designed to assist their learning, as suggested by Tomasello and Brooks (1999). A literacy expert examined the content and judged it to be appropriate and interesting for 4th and 5th grade students. Following a quality review, the website was revised to reduce clutter and reduce extraneous cognitive load, consistent with the coherence principle (Moreno & Mayer, 2000) and cognitive load theory (Pollock, Chandler, & Sweller, 2002).

A pedagogical agent was programmed so as to be integrated in physical proximity to the on-screen text, in keeping with the spatial contiguity principle (Moreno & Mayer, 1999). Textual statements were presented in all four
treatments, as was “Genie the Magician”. In every condition, the Genie flew in from the top left hand corner of the screen to an initial position on the top left of the screen near the text. Scripting of the pedagogical agent’s temporal speech cues were developed with the primary aim of directing attention to the details in the on-screen text, graphic or animation on the screen, as suggested in the structured sound function (SSF) model (Mann, 2006).

In the speech cues with agent movement (MS) condition the Genie flew in, immediately gestured to the text on the screen with its outstretched arm, and spoke speech cues in a synthesized voice about a future event (direction, instruction hint) or a past event (feedback or reminder). Periodically, Genie would scratch his head, do a magic trick, and fly to a different position on the screen.

In the on-screen text cues with agent movement (MT) condition (Figure 2), the Genie flew in and gestured to the text on the screen with its outstretched arm. The prompts appeared as text bubble in a text box, with identical information to the speech condition. The Genie’s movements in the MT condition were identical to those in the MS condition.

In the speech cues with no agent movement (NMS) condition, the Genie flew in and there-after remained motionless. The Genie in the NMS condition spoke the same temporal speech prompts as the Genie in the MS condition. But there was no gesturing to the text, no changing position on the screen, and no magic tricks.

In the on-screen text cues with no agent movement (NMT) condition, Genie flew in and remained motionless. Genie spoke as silent text bubble appeared above its head. But there was no gesturing to the text, no changing position on the screen, and no magic tricks.

The students were asked to discriminate the correct format of the possessive expression from the incorrect ones. Further, students were asked to transfer the knowledge of the correct usage of apostrophe to make up sentences according to different graphics and pictures. The structure and content of the program were held constant across the treatment conditions as the animation and modality were varied.

Research Design

We included a delayed post-test in this study to determine whether the original knowledge in the children’s long-term memories had been restructured, or overwritten over time. In this way the delayed post-test served to test the main purpose of instruction, namely as a knowledge-builder in long-term memory (Mann, 2008; Sweller, 2004). The study used a 3-factor repeated measures (pre-test, immediate post-test, delayed post-test) design with three independent between-subjects factors: agent modality (speech cues vs. text cues), agent animation (movement or no movement), and student grade level (grade 4 or grade 5).

Research Instruments

Consistent with the Stanovitch hypothesis, that poor and beginning readers gain more from context than more accomplished readers (Goldsmith-Phillips, 1989; Nickerson, 1991; Yeu & Goetz, 1994), we implemented a single 12-item test as the pretest, immediate post-test, and delayed post-test on correct apostrophe usage. The test was administered from the Web. Before implementation of the test, 4 students of the target grade level reviewed the items, and the materials were subsequently revised. The test was graded out of total of 30 possible points and checked by the investigators. There were 10 multiple choice questions each worth 2 points that asked the participant to choose which sentence was correct, and 2 short-answer questions each worth 5 points that asked the participant to type a sentence showing plural ownership based on a picture. The four experimental conditions in the study were deemed equivalent. The probability that this read-then-type requirement might disadvantage speech-cued participants, as suggested (Segers, Verhoeven, & Hulstein-Hendrikse, 2008) was low because the speech-cued participants had their cognitive load shared between auditory and visual memory systems, and were required to
process only the details portion only by reading the text, and the gist portion by listening to speech-cues.

**Research Procedure**

From the initial group of 162 students, 133 4th and 5th grade students returned signed consent forms and took the pre-test of apostrophe usage and interpretation of rules. Administration of the pre-test took about 20 minutes. For each experimental session, the teachers brought the participating students into the labs. Their names were coded and then anonymously and randomly assigned to one of the treatment groups, either: speech cues with agent movement (M, S), speech cues with no agent movement (NM, S), on-screen text cues with agent movement (M, T), or on-screen text cues with no agent movement (NM, T). Four weeks later, these students took the treatment and immediate post-test. Administration of treatment and immediate post-test took about 50 minutes. Six weeks later, the students took the delayed post-test, which lasted about 20 minutes. The experiment was conducted in three scheduled sessions over a 10-week period.

**RESULTS AND DISCUSSION**

**Results**

A preliminary analysis of variance (ANOVA) on the pre-test means of the two grade levels crossed with the four treatment groups indicated that they did not differ significantly, $F(7,129) = 1.19 \quad (p = .312)$. This suggested that the eight groups were similar with low prior knowledge of apostrophe usage, and that that knowledge was uniformly low. Low prior knowledge was a necessary pre-condition in this study, as explained earlier. Statistical imputations were conducted by multiple regression with a stochastic component on cases that showed a student had taken the treatment but were absent for a test to retain as much of the data as possible, in accordance with recommended procedures (Little & Rubin, 1990; Smits, Mellenbergh & Vorst, 2002). There were 11 missing cases on the pre-test and 4 on the delayed post-test, evenly distributed across the treatment groups (from 2 to 4 imputations per group on the pre-test, and from 0 to 2 on the delayed post-test). Missing retention scores were then replaced with the value predicted from multiple regression of the retention test on the pre- and post-tests based on the original complete data. To this replacement...
was added a random normal component with mean = 0 and standard deviation = standard error of regression. Smits, Mellenbergh, and Vorst (2002) found these procedures to be effective for missing data, particularly where listwise deletion would be inappropriate. Listwise deletion methods usually discard important data available on other variables, and argued to be less appropriate than stochastic conditional imputation (Little & Rubin, 1987; 1990; Wothke, 2000). The procedures used permitted analysis of a complete data set that reflected the expected characteristics of the scores (Little & Rubin, 1987; Peugh & Enders, 2004; Raaijmakers, 1999; Raymond & Roberts, 1987; Roth, 1994). This resulted in complete data for 133 students (see Table 1 for numbers in each group).

The design then, consisted of a within-subjects factor (the three repeated measurements: pre-, post-, and delayed test) crossed with the three between-subjects factors (agent animation, modality, and grade level). Treatment effects of animation and modality would be evidenced by interaction with the repeated measures. There was only one statistically significant interaction for the within-subjects analysis: the interaction of the repeated measures with modality, $F(2, 250) = 5.02 \ (p = .007)$. While highly significant, it should be noted that this interaction effect accounted for approximately 4% of the variance attributed to the within-subjects part of the analysis, and should therefore be described as a small to medium effect. There was a significant within-subjects main effect due to the repeated measures, $F(2, 250) = 22.29 \ (p = .000)$, and a significant between-subjects effect due to modality $F(1, 125) = 21.54 \ (p = .000)$. While these main effects were highly significant, they were not pursued directly as the interaction noted above was significant, and it included both of these factors.

The results showed that the pattern of change from pre- to post- and delayed post-test was different for two modality conditions, Speech and Text. Since there were no more complex interactions involving modality, the eight groups were collapsed into the two modality groups. The means of the two groups are depicted in Figure 3. The pattern for the Speech group appears distinctly different from that for the Text group: the Speech group means changed from 10.95 on the pretest to 15.73 on the posttest and 13.65 on the delayed post-test, whereas the corresponding Text group means changed from 8.80 to 10.60 and 9.56.

Simple main effects were tested for the three testing times within the Speech group and within the Text group to determine which of these differences were significant (using procedures outlined for repeated-measures analyses by Keselman & Keselman, 1993). There was a highly significant change over time for the Speech group accounting for over 27% of the variance associated with this group, $F(2, 130) = 24.51 \ (p = .000)$. Further post hoc tests using the Bonferroni critical value of .017 for the three pair wise tests indicated that all three pairs of means differed significantly. The pretest and post-test means ($p = .000$) had a medium to large effect size of .76, the pretest and delayed post-test means ($p = .001$) had a medium effect size of .43, and the post-test and delayed post-test means ($p = .001$) had a small effect size of .31. Means for the speech group were highest on the post-test, followed by the delayed post-test, and then the pretest (Figure 2). Thus, significant learning occurred for this group, and this learning maintained to the delayed post-test but with some loss.

The simple main effect results were also significant for the Text group but not as strong, accounting for only approximately 5% of the variance associated with this group on the three tests, $F(2, 132) = 3.27 \ (p = .041)$. Post hoc tests using the critical value of .017 could not detect where these differences occurred, pretest and post-test means ($p = .030$), pretest and delayed post-test means ($p = .268$), post-test and delayed post-test means ($p = .116$). The means for the Text group were more similar to one another than the means in the Speech group (Figure 2). It is not as clear whether there actually was learning with the Text group, although some change did occur from pre- to post- and delayed post-testing.
Discussion

This research has provided needed data about how young children learn from temporal speech cueing and pedagogical agents. The findings with these 4\textsuperscript{th} and 5\textsuperscript{th} graders were consistent with results in previous research with adults on graphics and narration (Chandler & Sweller, 1992; Mayer, 1997; Moreno & Mayer, 2000). Continued research is needed on listening to auditory events in multimedia, and how school age and even pre-school children might focus attention with a speaking pedagogical agent. Physiological studies with infants have shown that even the youngest child can focus their attention (Cowan, Elliott, Saults, Morey, Mattox, Hismjatullina, & Conway, 2005). O’Leary and Rhodes (1984) reported that when babies...

Table 1. Means on apostrophe test for the eight groups over the three testing times

<table>
<thead>
<tr>
<th>Agent</th>
<th>Modality</th>
<th>Grade</th>
<th>n</th>
<th>M</th>
<th>s</th>
<th>M</th>
<th>s</th>
<th>M</th>
<th>s</th>
</tr>
</thead>
<tbody>
<tr>
<td>an.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Move.</td>
<td>Speech</td>
<td>4</td>
<td>20</td>
<td>11.94</td>
<td>5.27</td>
<td>15.80</td>
<td>5.39</td>
<td>13.29</td>
<td>5.11</td>
</tr>
<tr>
<td></td>
<td></td>
<td>5</td>
<td>10</td>
<td>11.92</td>
<td>6.71</td>
<td>16.90</td>
<td>8.74</td>
<td>18.00</td>
<td>6.99</td>
</tr>
<tr>
<td>Text</td>
<td></td>
<td>4</td>
<td>15</td>
<td>9.72</td>
<td>6.91</td>
<td>11.13</td>
<td>7.61</td>
<td>12.28</td>
<td>6.58</td>
</tr>
<tr>
<td></td>
<td></td>
<td>5</td>
<td>25</td>
<td>10.60</td>
<td>6.78</td>
<td>13.44</td>
<td>8.41</td>
<td>14.57</td>
<td>7.19</td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>38</td>
<td>15</td>
<td>10.33</td>
<td>4.84</td>
<td>13.05</td>
<td>5.54</td>
<td>10.89</td>
<td>5.25</td>
</tr>
<tr>
<td></td>
<td></td>
<td>63</td>
<td>10.44</td>
<td>5.64</td>
<td>13.21</td>
<td>6.76</td>
<td>12.35</td>
<td>6.31</td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>Speech</td>
<td>4</td>
<td>21</td>
<td>9.51</td>
<td>3.56</td>
<td>14.10</td>
<td>6.02</td>
<td>11.92</td>
<td>5.74</td>
</tr>
<tr>
<td></td>
<td></td>
<td>5</td>
<td>15</td>
<td>11.02</td>
<td>8.38</td>
<td>17.13</td>
<td>7.41</td>
<td>13.67</td>
<td>8.41</td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>36</td>
<td>10.14</td>
<td>5.99</td>
<td>15.36</td>
<td>6.70</td>
<td>12.65</td>
<td>6.92</td>
<td></td>
</tr>
<tr>
<td>Text</td>
<td></td>
<td>4</td>
<td>18</td>
<td>8.79</td>
<td>4.43</td>
<td>10.22</td>
<td>3.69</td>
<td>9.11</td>
<td>5.38</td>
</tr>
<tr>
<td></td>
<td></td>
<td>5</td>
<td>16</td>
<td>8.25</td>
<td>4.31</td>
<td>11.19</td>
<td>8.02</td>
<td>9.00</td>
<td>5.74</td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>34</td>
<td>8.54</td>
<td>4.32</td>
<td>10.68</td>
<td>6.04</td>
<td>9.06</td>
<td>5.47</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>4</td>
<td>39</td>
<td>9.18</td>
<td>3.95</td>
<td>12.31</td>
<td>5.38</td>
<td>10.62</td>
<td>5.68</td>
</tr>
<tr>
<td></td>
<td></td>
<td>5</td>
<td>31</td>
<td>9.59</td>
<td>6.64</td>
<td>14.06</td>
<td>8.18</td>
<td>11.26</td>
<td>7.42</td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>70</td>
<td>9.36</td>
<td>5.27</td>
<td>13.09</td>
<td>6.77</td>
<td>10.90</td>
<td>6.47</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>Speech</td>
<td>4</td>
<td>41</td>
<td>10.69</td>
<td>4.59</td>
<td>14.93</td>
<td>5.71</td>
<td>12.59</td>
<td>5.42</td>
</tr>
<tr>
<td></td>
<td></td>
<td>5</td>
<td>25</td>
<td>11.38</td>
<td>7.62</td>
<td>17.04</td>
<td>7.79</td>
<td>15.40</td>
<td>8.02</td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>66</td>
<td>10.95</td>
<td>5.87</td>
<td>15.73</td>
<td>6.60</td>
<td>13.65</td>
<td>6.61</td>
<td></td>
</tr>
<tr>
<td>Text</td>
<td></td>
<td>4</td>
<td>36</td>
<td>8.66</td>
<td>4.01</td>
<td>10.11</td>
<td>3.76</td>
<td>8.67</td>
<td>4.71</td>
</tr>
<tr>
<td></td>
<td></td>
<td>5</td>
<td>31</td>
<td>8.96</td>
<td>5.67</td>
<td>11.16</td>
<td>7.69</td>
<td>10.59</td>
<td>6.28</td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>67</td>
<td>8.80</td>
<td>4.81</td>
<td>10.60</td>
<td>5.89</td>
<td>9.56</td>
<td>5.54</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>4</td>
<td>77</td>
<td>9.74</td>
<td>4.42</td>
<td>12.68</td>
<td>5.43</td>
<td>10.75</td>
<td>5.44</td>
</tr>
<tr>
<td></td>
<td></td>
<td>5</td>
<td>56</td>
<td>10.04</td>
<td>6.66</td>
<td>13.79</td>
<td>8.21</td>
<td>12.74</td>
<td>7.44</td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>133</td>
<td>9.87</td>
<td>5.45</td>
<td>13.14</td>
<td>6.74</td>
<td>11.59</td>
<td>6.41</td>
<td></td>
</tr>
</tbody>
</table>

Copyright © 2012, IGI Global. Copying or distributing in print or electronic forms without written permission of IGI Global is prohibited.
Table 2. Apostrophe analysis of variance: repeated measures

<table>
<thead>
<tr>
<th>Source</th>
<th>df</th>
<th>F</th>
<th>η²</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Between subjects</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Agent animation</td>
<td>1</td>
<td>1.74</td>
<td>.014</td>
<td>.190</td>
</tr>
<tr>
<td>Modality</td>
<td>1</td>
<td>21.54</td>
<td>.147</td>
<td>.000</td>
</tr>
<tr>
<td>Grade</td>
<td>1</td>
<td>3.20</td>
<td>.025</td>
<td>.076</td>
</tr>
<tr>
<td>Agent animation * Modality</td>
<td>1</td>
<td>0.47</td>
<td>.004</td>
<td>.496</td>
</tr>
<tr>
<td>Agent animation * Grade</td>
<td>1</td>
<td>0.28</td>
<td>.002</td>
<td>.597</td>
</tr>
<tr>
<td>Modality * Grade</td>
<td>1</td>
<td>0.26</td>
<td>.002</td>
<td>.609</td>
</tr>
<tr>
<td>Agent animation * Modality * Grade</td>
<td>1</td>
<td>0.39</td>
<td>.003</td>
<td>.532</td>
</tr>
<tr>
<td><strong>Error</strong></td>
<td>125</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Within subjects</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pre-Post-Delayed</td>
<td>2</td>
<td>22.29</td>
<td>.151</td>
<td>.000</td>
</tr>
<tr>
<td>Pre-Post-Del * Agent animation</td>
<td>2</td>
<td>1.46</td>
<td>.012</td>
<td>.233</td>
</tr>
<tr>
<td>Pre-Post-Del * Modality</td>
<td>2</td>
<td>5.02</td>
<td>.039</td>
<td>.007</td>
</tr>
<tr>
<td>Pre-Post-Del * Grade</td>
<td>2</td>
<td>2.10</td>
<td>.017</td>
<td>.125</td>
</tr>
<tr>
<td>Pre-Post-Del * Agent animation * Modality</td>
<td>2</td>
<td>0.08</td>
<td>.001</td>
<td>.922</td>
</tr>
<tr>
<td>Pre-Post-Del * Agent animation * Grade</td>
<td>2</td>
<td>2.69</td>
<td>.021</td>
<td>.070</td>
</tr>
<tr>
<td>Pre-Post-Del * Modality * Grade</td>
<td>2</td>
<td>0.09</td>
<td>.001</td>
<td>.913</td>
</tr>
<tr>
<td>Pre-Post-Del * Agent animation * Modality * Grade</td>
<td>2</td>
<td>0.13</td>
<td>.001</td>
<td>.878</td>
</tr>
<tr>
<td><strong>Error (within subjects)</strong></td>
<td>250</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Figure 3. Means on apostrophe test for the Speech and Text groups

![Graph showing means on apostrophe test for Speech and Text groups]
listened to an audio recording of one woman from a speaker located halfway between two videos of different women speaking simultaneously, the babies preferred to watch the face that belonged with the voice they were hearing. The babies shifted their attention until they associated the auditory and visual events, consistent with event-related potential (ERP) evidence, which is a measure of the electrical activity produced by the brain in response to a sensory stimulus (event) in the environment (Desjardins & Werker, 2004; Patterson & Werker, 2003; Vouloumanos & Werker, 2004), and with Baddeley’s (1992, 2002) revised model of working memory in children.

Our expectations that these children would learn from a moving agent was not supported. Despite the apparent fascination with the agent or “little Dude” however, as he was called by some of the children. This finding was also consistent with adult learning. The speculation was that agent motion and gesturing would capture attention and offer feedback without breaking the student’s train of thought (Craig, Gholson, & Driscoll, 2002; Johnson, Rickel, & Lester, 2000; Veronikas & Shaughnessy, 2005), yet the moving agent didn’t get any better results from that of the agent that did not move. Deeper investigation is needed to determine how children best learn from the presence of a moving and gesturing pedagogical agent.

Limitations

Limitations of the research must also be noted. One limitation was that the experimental method restricted our investigations to a tightly controlled, narrowly defined curricular focus, namely apostrophe usage by 4th and 5th graders. All the teachers of these students were completely excluded from the experiments, quite unlike their daily multimedia learning experiences. The results indicate therefore, that whereas the nature of the presentations used and results from the experiments were valid, they may not be completely generalizable to school-age students in normal lab settings. A second limitation of the research was that motivational strategies were left unplanned because the focus of this study was solely informational. Only simple learner controls on the interface were used instead of more sophisticated methods evident in children’s educational software, such as learner-control with advisement. A third limitation was that although they were randomly assigned to treatment groups, the children in both experiments were nevertheless, a population of convenience wherein whole (intact) classes of students were used.

Future Directions

Spoken directions, instruction and feedback should therefore, continue to be an integral part of multimedia designed for young children, and continue to be critical elements of a pedagogical agent’s persona. Eventually, this kind of research may enable researchers and developers to create effective pedagogical agents to reach the widest possible range of learners from diverse backgrounds.

REFERENCES


Bruce Mann, PhD, LLM, is a Professor of Education at Memorial University in St. John’s, Canada. His research interests include: instructional design with an avatar, non-player character or pedagogical agent; and technology law.

Henry Schulz, PhD, is an Associate Professor of Education at Memorial University in St. John’s, Canada. His research interests include: quantitative research methods, validation and generalizability.

Jianping Cui, MEd, is a Teaching and Learning Consultant at Bow Valley College in Calgary, Canada. Prior to this position Ms. Cui held Instructional Designer positions at Bow Valley College, College of the North Atlantic, and the Faculty of Medicine at Memorial University.