

Chapter 4

TALKING TO DIGITAL FISH

Designing effective conversational interfaces for educational software

Courtney Darves and Sharon Oviatt

A Locanian, having plucked all the feathers off from a nightingale and seeing what a little body it had, "Surely," quoth he, "thou art all voice and nothing else."

--Plutarch, Laconic Apothegms

Abstract Conversational interfaces that incorporate animated characters potentially are well suited for educational software, since they can engage children as active learners and support question asking skills. In the present research, a simulation study was conducted in which twenty-four 7-to-10-year-old children used speech and pen input to converse directly with animated fish as they learned about marine biology. The animated fish responded with TTS voices crafted to sound either extroverted or introverted in accordance with the speech signal literature. During these interactions, children became highly engaged, asking an average of 152 questions during a 45-minute session. Self-report measures further confirmed that children liked “talking to the animals” and that the TTS and interface were easy to understand and use. The auditory embodiment of animated characters as TTS output also had a significant selective impact on children's engagement in asking science questions. Specifically, children asked +16% more science questions when conversing with animated characters embodying an extrovert TTS voice that resembled the speech of a master teacher (e.g., higher volume and pitch, wider pitch range), rather than an introvert TTS voice, although no differential impact was found on social questions. These findings reveal that conversational interfaces can be designed that effectively stimulate children during learning activities, thereby supporting the goals of next-generation educational software.

Keywords: Children's speech, animated software characters, conversational interfaces, text-to-speech, audio interface design, science education, educational software.

1. Introduction

Researchers, educators, and industry alike all have explored the effective integration of computers in education (Cassell and Ryokai, 2000; Haughland, 2000; Kulik, Bangert and Williams, 1983). Computer-based instruction can lead to higher test scores and better attitudes toward computers, and can reduce the amount of time needed to learn a subject matter (Haughland, 2000; Kulik et al., 1983). It clearly also can play a major role in making education more accessible and better tailored for all children. For example, audio-visual multimedia systems have the potential to provide better support for children with different learning styles, in comparison with previous text-based information delivery (Druin, 1996; Moreno, Mayer, Spires and Lester, 2001).

Although computer-based instruction historically has relied on rote-learning strategies, newer interfaces now are being designed that give children a more active role in the learning experience (Haughland, 2000). Recently, researchers have emphasized the importance of children having more active control and creative input during computer-based learning. One trend has been to make students designers of the technology they use (Cassell and Ryokai, 2000; Druin, 1996; Druin, 1999). For example, in one application children added the graphics, animation, and sound to their own original story (Cassell and Ryokai, 2000). In another application, students learned about plant physiology by designing roots, stems, and leaves appropriate for different environments (Moreno et al., 2001).

Another emerging trend in educational software is the incorporation of animated characters, which can provide an interface design vehicle for engaging children and managing the overall tutorial exchange (Lester, Converse, Stone, Kahler and Barlow, 1997; Oviatt and Adams, 2000). When animated characters are embedded within a conversational interface, they quite naturally can become the central focus of the content exchange as an interlocutor, rather than playing a subsidiary and sometimes distracting “help-agent” role. As an example, in the course of learning about science, a child could converse directly with an animated parasite or sea creature to extract information about it. The immediacy of such an interaction could be designed to facilitate children’s engagement as “active learners” such that they seek answers to questions that they care about as they construct an understanding of science (Richmond, Gorham and Mccroskey, 1987). Consistent with a constructivist view of educational theory, one goal of the present research was to investigate how animated character technology can be designed to bring out the best in student’s question asking skills.

1.1 Evaluating Animated Characters in Educational Software

While past research has confirmed animated characters’ ability to engage and motivate users (Cassell, Sullivan, Prevost and Churchill, 2000; Dehn and Van Mulken, 2000; Lester et al., 1997; Moreno et al., 2001), it rarely has shown any task-relevant performance enhancement as a function of a character’s presence or specific design

(Dehn and Van Mulken, 2000). Likewise, most research on animated character design has focused on rendering them with high-fidelity graphics and animation, and on the impact of visual embodiment, but has ignored the question of whether auditory embodiment also can provide powerful cues that influence user behavior. However, in one study involving web-based book reviews, the TTS voice used for animated characters influenced users' self-reported book preferences and purchasing behavior (Nass and Lee, 2001). In another web-based study, children rated TTS sentences with focal F0 doubling as more fun than control sentences, whereas they rated sentences with focal duration lengthening as more boring than control sentences (Gustavson and House, 2001). In another web-based study, the presence of animated agents that actively monitored users' behavior as they worked was found to decrease users' performance and increase their anxiety level (Rickenberg and Reeves, 2000). Unfortunately, there are few compelling demonstrations that animated characters significantly improve users' learning-oriented behavior in any way during a tutorial exchange. One exception to this is a recent study which demonstrated that when students interacted with animated characters that provided spoken versus text-based feedback, they showed increased interest, learning, and performance on transfer tests (Moreno et al., 2001). However, the specifics of spoken interface design and different TTS voice types in particular were not assessed.

Since conversational interfaces are social in nature (Nass, Isbister and Lee, 2000; Nass and Lee, 2001; Nass, Steuer and Tauber, 1994), in the present research the voice characteristics of a "master" teacher were used as a design metaphor for integrating animated characters into an educational software application. The education literature indicates that students respond with increased attention and on-task behavior to dynamic and energetic speech (Bettencourt, Gillett, Gall and Hull, 1983; Sallinen-Kuparinen, 1992), or to an extroverted speech style that is higher in volume and pitch and more expanded in pitch range (Nass and Lee, 2001; Nass and Lee, 2000; Scherer, 1979). As a result, we might expect that animated characters that respond in an extroverted voice would be more effective in stimulating children's learning-oriented behavior, including their level of spontaneous question asking about educational content.

1.2 Goals of the Study

In the present study, children conversed directly with animated fish using the Immersive Science Education for Elementary kids (*ISEE*) interface as they learned about marine biology. This research was designed to:

- Explore whether conversational interaction with animated characters can be engaging for children, as measured by time spent interacting with the software, quantity of spontaneous question asking, and children's self reports

- Determine whether the TTS voice characteristics used for animated characters influence children’s learning-oriented behavior (e.g., question asking), and what the implications are for designing educational software
- Assess the overall usability of the *I SEE!* conversational interface prototype

With respect to the second goal, children’s queries were compared when they interacted with animated characters embodying different TTS voice profiles. In a comparison of introvert versus extrovert voices, it was predicted that an extrovert voice that shares features in common with master teachers’ speech would be more effective in stimulating children to ask task-appropriate questions during learning activities. In particular, it was predicted that children would ask more biology questions when conversing with an extrovert TTS voice (compared with an introvert voice), although no differential impact would occur for general social-interaction questions. The long-term goal of this research is to design effective conversational interfaces, in particular ones that have a desirable behavioral impact on users for the application being designed.

2. Methods

2.1 Participants, Task, and Procedure

Twenty-four elementary-school children participated in this study as paid volunteers. A median split divided the participants into two age groups, younger children (mean age 8;2, range 7;7. to 8;8.), and older ones (mean age 9;7, range 9;4 to 10;1), with each age group gender balanced. All participants were native English speakers without known behavioral or linguistic impairments. Participants also represented different personality types as assessed by parent and teacher ratings, with 13 rated by parents and/or teachers as extroverts versus 11 rated as moderately introverted. Participation was conducted at an elementary school field site.

Children interacted with an educational software program called Immersive Science Education for Elementary kids (*I SEE!*), which is an application designed to teach children about marine biology, simple data tabulation, and graphing. The interface permitted children to use speech, pen, or multimodal (speech and pen) input while conversing with animated software characters as they learned about marine biology. Figure 1 illustrates the *I SEE!* Interface. An animated “Spin the Dolphin” character, shown in the lower right corner of Figure 4-1, also was co-present on the screen and available as a conversational partner. The child could control the marine animals by asking Spin the Dolphin to start, stop, or replay the videos. When the movie stopped, the marine animal was embellished with animated eyes that gazed at the child and blinked naturally. At this point, the animal became available as a “conversational partner” for answering questions about itself using text-to-speech (TTS) output. For example, an animated manatee could identify its species, diet, habitat, unique behavior, endangered species status, and so forth. Essentially, the animated eyes that gazed at the

child and blinked provided attentional cues that marked the transition from a passive movie-viewing experience to active availability of the animal as the child's conversational partner.

During data collection, children queried the marine animals to collect information and build a graph representing information about them (e.g., "Can this animal change colors rapidly?"). Children also were encouraged to ask any questions they wished and to have fun learning new things about the animals. The marine animals were responsive but did not direct the conversation. Therefore, children's spontaneous conversations with the animals primarily were self-initiated, reflecting their own curiosity and interests about the marine creatures. When each child had finished asking a marine animal questions, he or she could request that Spin the Dolphin start the next movie with a new animal.

Before starting a session, each child received instructions and practice with a science teacher on how to use the *I SEE!* interface on a small hand-held computer, shown in Figure 4-2. During the ten- to fifteen-minute hands-on orientation, children interacted with and graphed information about three practice animals. Following the practice session, the experimenter left the room, and the child used the *I SEE!* application to view and interact with 24 different marine animals (e.g., octopus, shown in Figure 4-1) that were sequenced in three task sets of eight animals apiece. Each task set had a different target question (e.g., "Is this animal common or endangered?") and presented a new set of animals. Children could spend as much time as they liked interacting with each individual animal, and whenever they were ready could ask to see the next one.

During data collection, no teacher or adult was present to influence what children asked, or how long they interacted with the animals in *I SEE!* After the child finished interacting with all 24 animals, the experimenter returned and conducted a post-experimental interview related to the *I SEE!* system and its animated characters. A video record was made of all human-computer interaction with the interface during each session and interview, including all of children's spoken and pen-based input and a close-up view of their face and hands.

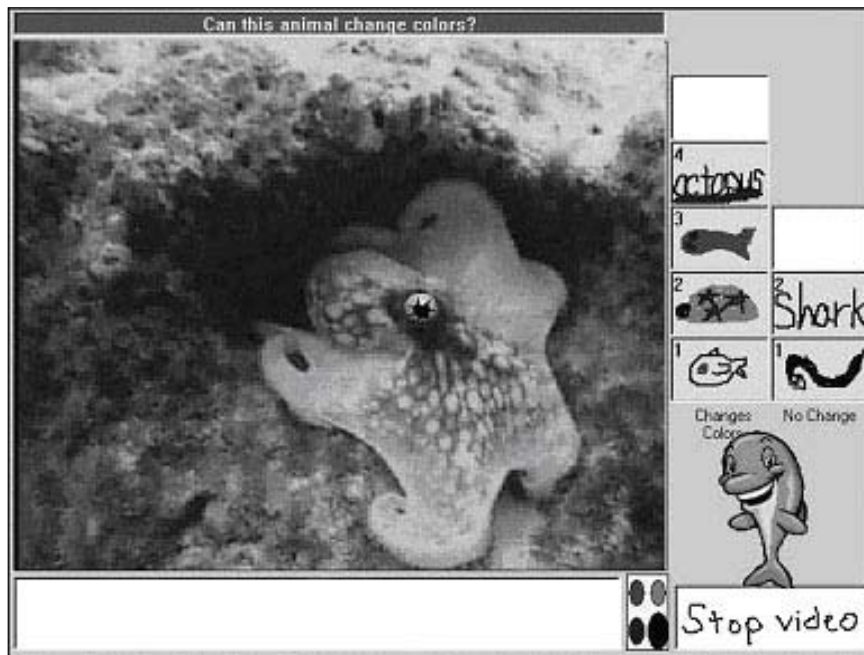


Figure 4.1. I SEE! Interface



Figure 4.2. Eight-year old boy at school as he asks an animated marine character questions about itself.

2.2 Simulation Environment

The *I SEE!* interface is a simulated conversational system that was designed to support proactive research on conversational interaction and interface design. As such, children's input was received by an informed assistant who interpreted their queries and provided system responses. System responses to high-frequency child queries were pre-loaded into a database, which is a feature that supported rapid simulated responding for the majority of children's questions. An automatic error generator produced general failure-to-understand messages in response to 4-5% of child queries in order to enhance the simulation's credibility. During testing, children believed that they were interacting with a fully functional system. The simulation environment ran on a PC, and it received input from a Fujitsu Stylistic™ 2300 that was used by the children. Details of the simulation infrastructure, its performance, and its use in research with children have been described elsewhere (Oviatt and Adams, 2000).

2.3 Text to Speech Manipulation

Text-to-speech voices from Lernout and Hauspie's TTS 3000 were used to convey the animated characters' spoken output. TTS voices were tailored for intelligibility of pronunciation. They included both male and female American English prototype voices, which were further tailored to represent opposite ends of the introvert-extrovert personality spectrum as indicated by the speech signal literature (Scherer, 1979; Smith, Brown, Strong and Rencher, 1995; Tusing and Dillard, 2000). Introvert and extrovert voices were selected because they are relatively well understood, highly marked paralinguistically, and have been used in previous research on the design of animated characters (Nass and Lee, 2001). In addition, the extrovert TTS voice was selected to model the speech of an engaging teacher (e.g., higher volume and pitch, wider pitch range), as described in the education literature and in the introduction of this chapter (Bettencourt et al., 1983; Sallinen-Kuparinen, 1992). As such, comparison of children's behavior in the two TTS conditions permitted an assessment of whether an interface with a TTS voice modeled after that of an expert teacher (Extrovert), as opposed to its acoustic opposite (Introvert), could be used to facilitate children's question-asking behavior. In total, four TTS voices were used in this study: (1) Male Extrovert (ME), (2) Male Introvert (MI), (3) Female Extrovert (FE), and (4) Female Introvert (FI).

The introvert and extrovert TTS voices were designed to represent the upper and lower bounds of speech signal features (e.g., amplitude, duration) in order to determine whether users' behavior could be influenced by a TTS target voice when an optimal degree of contrast is present. The TTS voices also were designed to maintain an identifiable social presence. To achieve this, features of the extrovert and introvert TTS voices were manipulated together as they tend to co-vary in real speakers' voices. That is, individual parameters were not manipulated in isolation in order to avoid producing artificial voices with no recognizable social personality or realism. Table 4-1 summarizes these differences in global speech signal features between the introvert and extrovert TTS voices.

It is important to note that due to pre-loading of system responses, lexical content was controlled in the different TTS voice conditions. In addition, the TTS voice conditions were counterbalanced across task sets, which controlled for the visual appearance of different animated characters presented during the study. Therefore, the only experimentally manipulated variable was the acoustic-prosodic characteristics of the TTS output.

Table 4.1. Characteristics of the four TTS voice conditions

TTS Voice	Mean Amplitude	Mean Pitch Range	Utterance Rate	Dialogue Response Latency
Type	(dB)	(Hz)	(syl/sec)	(sec)
FE	60	186	5.2	1.65
ME	58	106	5.2	1.65
FI	45	71	3.3	3.36
MI	44	58	3.3	3.36

2.4 Research Design and Analyses

The research design for the larger data collection effort, within which this study was situated, was a completely crossed factorial. The main within-subject factor was (1) Type of TTS Voice (Introvert, Extrovert). This factor remained constant for the first 16 animals, but switched for the remaining 8 (from I to E, or E to I). To test the generality of any TTS effects, I and E voices were tested using both male and female voice prototypes, which resulted in four voices total (ME, FE, MI, FI). Other between-subject comparisons included (2) Child Gender (Male, Female) and (3) Child Age (Young, Old), which was categorized using a median split to divide children into a younger (average 8 yrs., 2 mos.) and older (average 9 yrs, 7 mos.) group.

With respect to the main comparison involving TTS voice type, the marine animals were assigned one of the four TTS voices during practice and task sets 1 and 2. However, the introvert-extrovert dimension of the TTS voice then was switched for task set 3 (e.g., MI switched to ME; FE switched to FI). The TTS voices were distributed equally across subjects, with 6 children assigned each of the 4 voices for the initial task set. Participants were assigned semi-randomly to ensure equal numbers of male and female and older and younger children in each of the 4 TTS voice conditions. Figure 4-3 illustrates the main manipulation involving the TTS voice types and their switch for the marine characters before task 3 during each session.

In the present evaluation, time to complete activity, number of questions asked, and self-report comments were used to assess children's engagement with the interface. In addition, the total number of different types of questions that children asked (e.g., biology, social, interface help) was evaluated to compare the impact of E versus I TTS

voices on children's active initiation of learning-oriented behavior. Further details of the dependent measures used in this study are outlined in section 2.5. In accord with the hypotheses and directional predictions outlined in section 1.2, *a priori* paired t-tests were used to compare children's specific question asking behavior during the E versus I TTS voice conditions. Independent t tests were used to confirm that children's differential question asking behavior in the E versus I TTS voice conditions did not vary as a function of TTS voice gender, children's gender, or children's age.

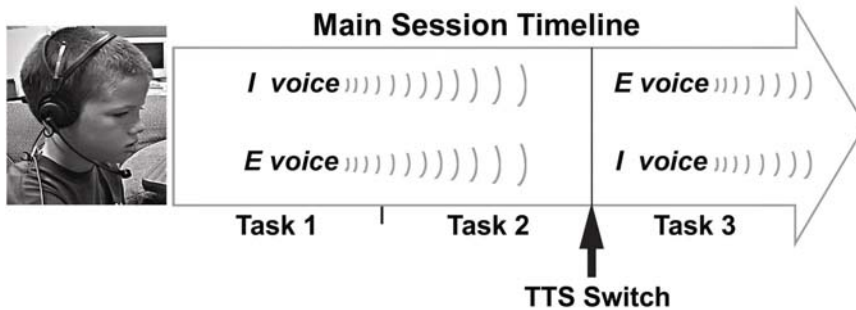


Figure 4.3. TTS voice for animated characters before and after switch task

2.5 Data Coding and Dependent Measures

Human-computer interaction was videotaped and conversational interaction transcribed. Children's conversations with the animated characters were coded for the following dependent measures: (1) time to complete activity, (2) number and type of child questions, and (3) children's self-report comments about the interface and its ease of use.

Time to Complete Activity

For all subjects, total time spent engaged with the *I SEE!* interface after practice was measured to the nearest second.

Number and Type of Self-Initiated Queries

The number and type of children's spontaneous queries to the animated characters and Spin the Dolphin were counted and coded into separate genre types. Table 4-2 provides descriptions and sample questions representing the main genres. The four genres were used to classify the questions into the following categories: (1) Biology, (2) Social, (3) Interface Help, and (4) Other questions. Questions coded in the Biology genre focused on factual educational information about the marine animal, including its diet, habitat, predators, and so forth. In contrast, the Social genre encompassed questions that were social-interactive in nature, including questions about family life, friends, personal preferences, as well as ritualized social greetings. The Interface Help genre included questions on how to use the *I SEE!* Interface. The Help genre served as an indicator of ease and naturalness of the interface. The Other genre included

miscellaneous questions not classifiable into the other main genres, for example “What’s behind you?” In addition, the number of child requests for an animated character to repeat an utterance was counted separately to assess TTS intelligibility. Children’s commands (e.g., to start the movies introducing new animals), responses to system initiations, and simple acknowledgments were relatively infrequent, and were separated from the other main categories of interest.

Table 4.2. Description of query genres

Genre	Description and Examples
Biology	Questions about biology. <ul style="list-style-type: none"> - <i>What kind of marine animal are you?</i> - <i>How do you defend yourself?</i>
Social	Questions about social and personal issues. <ul style="list-style-type: none"> - <i>What’s your name?</i> - <i>What’s your favorite color?</i>
Help	Questions about how to use the <i>I SEE!</i> interface. <ul style="list-style-type: none"> - <i>How do I stop the movie?</i> - <i>How do I change the ink color?</i>

Interview Self-Reports

At the end of each child’s session, the experimenter returned to interview the child. Children were told that their responses would be used to help improve future versions of the computer. The following questions were used to summarize the children’s appraisal of the interface and its ease of use:

- (1) What did you think of this computer?
- (2) Would you like to have a computer like this? (If so, what would you do with it?)
- (3) Was the computer easy or hard to use? (If hard, what was hard?)
- (4) What did you think of the ocean animals?
- (5) What were the animals like—a stranger, friend, teacher, parent, or what?
- (6) What about this computer did you like?

Children’s responses to these questions were easily separated into categories for qualitative description. The percentage of children who gave positive appraisals of the system was summarized based on responses to the first question, and the percentage of children who appraised the marine animals positively was assessed using the fourth question. Assessments of children’s desire to own such a computer, its ease of use, and beliefs about what the animals were like were evaluated from the second, third, and

fifth questions, respectively. Finally, children's qualitative comments about the animated characters and system were summarized based on responses to the fourth and sixth questions.

Inter-coder Reliability

In total, 17% of child queries were second-scored by an independent coder for genre classifications, and these judgments between coders matched over 99% of the time.

3. RESULTS

Our study of twenty-four children's interactions with the *I SEE!* interface yielded approximately 36 hours of videotape data. A total of 3,643 child queries were coded for genre classification, of which 3,340 were directed to the animated marine animals, and another 303 to Spin the Dolphin. Figure 4-4 shows an example of a transcribed question-answer exchange between a child and a marine animal, which illustrates very typical content for these dialogues.

3.1 Engagement in Interface and Ease of Use

Even though children were alone in the classroom with no teacher present, they spontaneously asked an average of 152 queries of the animated marine animals while engaged with the interface. The total questions asked per child ranged from 62 to 309. Figure 4-5 displays the total number of questions asked by each subject. During these interactions, children spent an average of 45.9 minutes engaged in conversation with the animated characters.

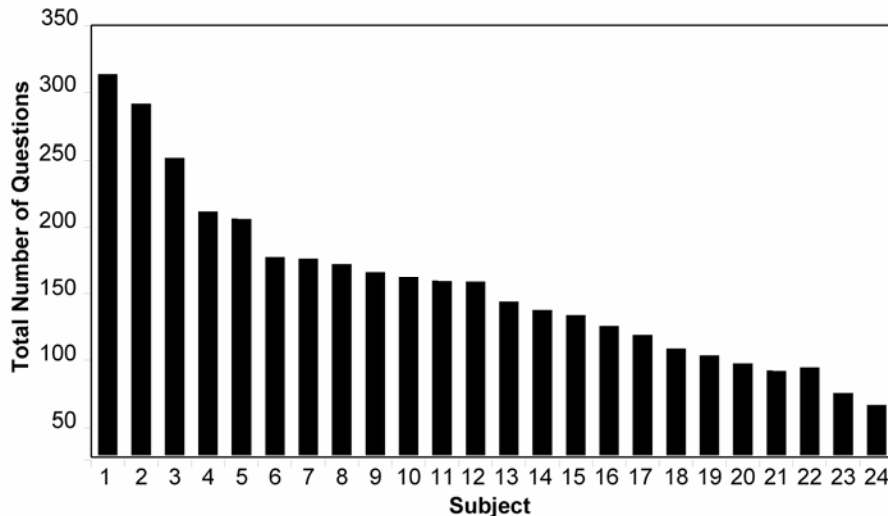


Figure 4.5. Individual differences in total number of questions asked

Child: Are you an octopus?

Octopus: I'm an octopus.

Child: What is your name?

Octopus: My name is Ollie the Octopus. What is your name?

Child: Jake.

Octopus: Hello, Jake.

Child: Do you have any brothers and sisters?

Octopus: I have many brothers and sisters.

Child: Do you know any of them?

Octopus: They have gone away.

Child: Do you have teeth?

Octopus: I don't have teeth. I have a beak instead.

Child: What do you eat?

Octopus: I eat crustaceans and sometimes smaller mollusks. I hunt them at night.

Child: How warm is it in the ocean?

Octopus: About 73 degrees.

Child: Do you move a lot?

Octopus: I can crawl with my arms. I can also swim using my arms or by shooting out water from my siphon.

Child: What kind of octopus are you?

Octopus: I'm an Atlantic octopus.

Child: Can I see another movie now?

Figure 4.4. Typical conversational question-answer exchange between a child and octopus animated marine character.

In spite of the fact that children were introduced to Spin the Dolphin as a character who could provide them with help using the computer, less than 0.25% of all children's

queries to either the animated marine characters or to Spin involved requests for help with the interface, including help constructing graphs. In addition, children rarely (less than 1% of the time) requested repetition of TTS feedback from the animated characters or Spin, which confirmed that the TTS was adequately intelligible for the present application.

Based on self-reports collected during post-experimental interviews, 100% of the 24 children gave a positive assessment of the interface, with 79% reporting that it was “easy to use,” and 96% reporting that they wanted one to own. Typical qualitative comments included that the computer was “cool,” “fun,” and something they would “like to play with all day.” Children’s most common spontaneous comments were that they liked “talking to the animals” (50%), “being able to write and speak to the computer” (29%), and “being able to get answers to questions and learn things” (21%). In terms of the animated characters, 96% of children assessed them positively, with 83% describing them as being like “friends” or “teachers” (i.e., rather than parents, strangers, or other).

3.2 Distribution of Question Types

As shown in Table 4-3, the majority of children’s queries to the animated marine characters (75%) focused on marine biology factual information. The remaining questions (24%) were social in nature, with only a small percentage on miscellaneous topics. As shown in Figure 4-6, all but one child asked more factual questions than social questions

Table 4.3. Distribution of total queries to marine animals by topic

Genre	Occurrences	Percent of Corpus
Biology	2493	74.6
Social	794	23.8
Other	53	1.6
Interface Help	0	0

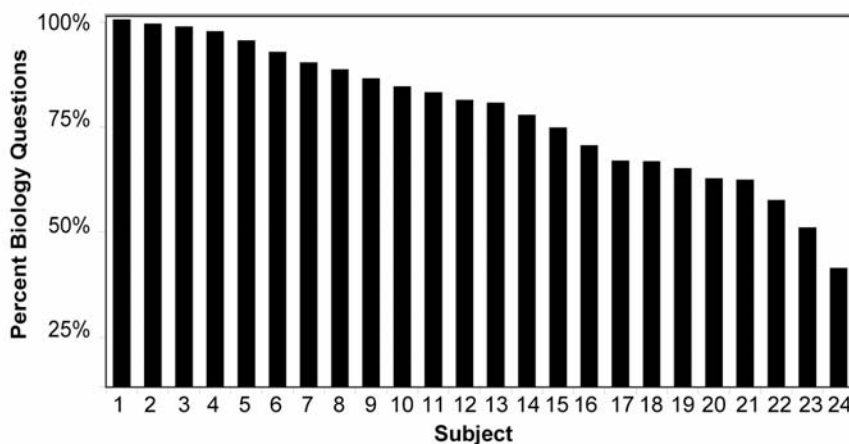


Figure 4.6. Percentage biology questions asked by each subject of their total social and biology queries

3.3 Impact of TTS Voice Type on Child Queries

Children asked more questions overall when interacting with animated marine characters embodying the extrovert TTS voice, compared with the introvert voice (mean 141 vs. 126 questions, respectively). Figure 4-7 illustrates children's differential level of question asking when interacting with the introvert and extrovert voices, broken down into the two main genre types of biology versus social questions. *A priori* paired t-tests confirmed that children asked a greater number of biology questions when conversing with the extrovert voice, rather than the introvert one (mean 108 and 93 biology queries, respectively), paired $t=2.08$ ($df=23$), $p < .025$, one-tailed. This represented a 16% overall increase in children's educationally-relevant question asking when interacting with the extrovert TTS voice. Further, the majority of children, or 17 of 24, responded in this manner. Table 4-4 shows individual differences in the relative increase in total biology questions asked when children interacted with characters speaking in extrovert versus introvert TTS voices. In contrast, no significant difference was found in the level of children's social queries when interacting with these two voice types, $t < 1$, N.S.

Comparison of the difference in biology questions asked as a function of the TTS voice condition revealed no significant difference between the younger and older children, $t = 0.165$ ($df = 22$), $p = 0.871$ (separate variances), N.S. There also was no difference between male and female children, $t = 0.465$ ($df = 22$), $p = 0.647$, N.S. Finally, these results also generalized across testing with the male and female TTS voice prototypes, for which no significant differences were observed, $t = 1.377$ ($df = 22$), $p = 0.18$, two-tailed, N.S.. That is, all of these analyses confirmed that the extrovert TTS voice stimulated significantly and selectively more biology queries. A summary of these results on the impact of TTS voice type on children's question asking is shown in Tables 4.5 and 4.6.

Table 4.4. Individual differences in percentage increase in biology questions asked when interacting with extrovert versus introvert TTS voices

Subject	Percent Change
S1	100%
S2	68%
S3	64%
S4	58%
S5	52%
S6	37%
S7	34%
S8	33%
S9	31%
S10	25%
S11	24%
S12	17%
S13	13%
S14	10%
S15	9%
S16	6%
S17	3%
S18	-1%
S19	-13%
S20	-14%
S21	-17%
S22	-24%
S23	-39%
S24	-53%

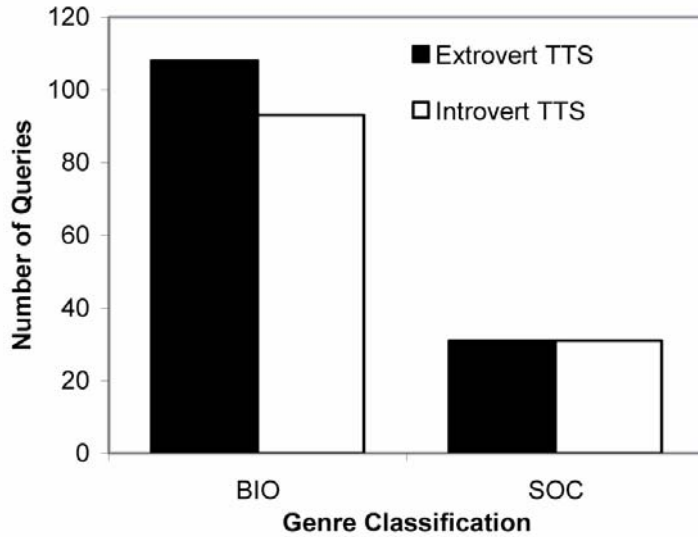


Figure 4.7. Number of biology (BIO) and social (SOC) queries asked by children when interacting with characters using extrovert versus introvert TTS voices.

Table 4.5. Impact of extrovert (E) versus introvert (I) TTS voice type on number of educationally-relevant child queries.

Dependent Measure	P value	Magnitude and Direction of Effect
Total queries	<.09	+12% more questions in E condition
Biology queries	<.025*	+16% more questions in E condition
Social queries	0.968	N.S.

Table 4.6. Generality of effect (increased biology-content questions) across child gender, age, and TTS voice gender.

Comparison Groups	P value	Magnitude and Direction of Effect
Male TTS vs. Female TTS	0.183	N.S.
Male children vs. Female children	0.647	N.S.
Younger children vs. Older children	0.871	N.S.

4. DISCUSSION

The primary aim of this chapter has been to explore aspects of conversational interfaces and animated character design within the context of an empirical evaluation of an educational interface prototype. We explored both the auditory embodiment of animated characters, which will be especially important for future mobile interfaces, as well as how to design animated characters for educational software in a way that engages children and facilitates their learning-oriented behavior. Within the framework of proposed evaluation taxonomies for animated character design (Ibister and Doyle, 2004), the present work represents an *Application Domain* research focus.

4.1 Acoustic characteristics of animated character design

Auditory embodiment alone, independent of an animated character's visual appearance or lexical output, can be highly influential in stimulating users' behavior in task-appropriate ways. In the present conversational interface, children's question asking was substantially affected by the acoustic-prosodic features of the TTS output they heard. When interacting with the extrovert voiced characters, which in many ways represented the rhetorical style of a master teacher (Bettencourt et al., 1983; Sallinen-Kuparinen, 1992), children were stimulated to ask 16% more marine biology questions. In contrast, children's general social questions were not differentially affected by the same introvert and extrovert voices. In other words, using an extrovert TTS voice that was louder, faster, higher in pitch, and wider in pitch range had a *selective impact* on children's educationally relevant question-asking behavior. The extrovert voice essentially was more successful in motivating and managing a tutorial exchange.

This finding underscores the important role of TTS design in the success of future conversational interfaces. Matching an appropriate TTS voice to an application domain can be a tool for influencing user behavior, and may be particularly useful for mobile audio-only interfaces. As computer interfaces evolve toward multimodal and speech-based communication, it will become increasingly important to understand the impact of acoustic-prosodic TTS parameters on users' learning and behavior.

4.2 Conversational interfaces as educational interfaces

One goal of this research was to investigate the quantity and quality of children's question asking when using a relatively unstructured conversational interface. We found that when left alone, children spontaneously asked an average of 152 questions of the digital fish, and in some cases over 300 questions. The majority of children's questions, or 75%, focused on marine biology, and this pattern was consistent for the majority of the children. The large volume of questions focusing on marine biology suggests that conversational interfaces can be successfully designed to promote children's mastery of science through active question-asking (Richmond et al., 1987). Beyond this, even children's questions that focused on social interaction and bonding with the marine animals may have indirectly stimulated learning by engaging them in marine science content.

Children's most common positive comment about the computer was that they liked "talking to the animals," which may in part reflect the "immediacy characteristics" of this interface (Richmond et al., 1987), as well as the self-reinforcing nature of conversation itself. The majority of children also reported that the animated characters were like "friends" or "teachers", with 96% of the children assessing the characters positively. Children's engagement with the characters was corroborated by the social quality of their conversations. For example, they gave the fish compliments ("You're pretty"), showed empathy toward them ("I'm sorry you're endangered"), and displayed emotional attachment ("I'll miss you, Spin!"). Finally, past work with the *I SEE!*

interface has revealed that children predominantly use personal pronouns when addressing these animated characters (Oviatt and Adams, 2000).

The ease and naturalness of conversational interfaces make them good candidates for educational software, especially among young children who may be unable to read or spell consistently. After only brief exposure, young children using *I SEE!* were able to converse with the fish, extract large amounts of information about marine biology, and construct graphs tabulating this new information. In post-experimental interviews, 79% of children reported that the system was “easy to use,” and children rarely requested help using the interface. Future uses of conversational interfaces may include handheld computers that serve as “interactive tour guides” at aquariums, zoos, or museums. Rather than passively reading exhibit signs or listening to an audio-taped tour, visitors could use the conversational interface to obtain answers to specific questions quickly and easily.

4.3 Conclusion

In the present research, the interpersonal and educational literature provided a point of departure for designing effective TTS voices for a tutorial exchange. TTS voices modeled after expert teachers' voices were the most effective at stimulating learning-oriented interactions. Specifically, these voices increased the number of science questions asked by children, although they had no differential impact on the number of general social questions asked. Furthermore, both behavioral and self-report measures indicated that the young children in this study enjoyed the conversational interface, and especially “talking to the animals.” When working alone in a classroom, children asked the digital fish over 150 questions during a 45-minute session. They also found the interface easy to use and rarely requested help or repetition of the TTS output.

These results indicate that conversational interfaces can facilitate learning through direct question and answer exchanges that are easy, natural, and highly engaging for users. Furthermore, the specific characteristics of animated character voices can influence users' behavior in task-relevant ways. In different application contexts with other user groups, undoubtedly different social metaphors and associated TTS profiles will be needed to achieve the most desirable impact. Future research should continue to pursue understudied aspects of the design of animated characters, including their auditory embodiment. The long-term goal of this research is the design of effective conversational interfaces, in particular ones that have a task-appropriate behavioral impact on users for the application being designed.

Acknowledgements

This research was supported in part by Grants IRI-9530666 and IIS-0117868 from the National Science Foundation, Special Extension for Creativity (SEC) Grant IIS-9530666 from NSF, and a gift from the Intel Research Council to the second author. Thanks to Matt Wesson for implementing simulation, transcription, and data analysis tools and for assisting during testing. Thanks also to Rachel Coulston for reliability

scoring of the genre coding. Finally, we are grateful to the students who participated in this research.

References

- Bettencourt E., Gillett M., Gall M. and Hull R. (1983) Effects of teacher enthusiasm training on student on-task behavior and achievement, *American Educational Research Journal*, 20(4), pp. 435-450.
- Cassell J. and Ryokai K. (2000) Story Spaces: Interfaces for Children's voices. *Proc. of ACM SIGCHI Conference of Human Factors in Computing Systems: CHI 2000*, pp. 243-244.
- Cassell J., Sullivan J., Prevost S. and Churchill E. (eds.) (2000) *Embodied conversational agents*. MIT Press: Cambridge, MA.
- Dehn D.M. and Van Mulken S. (2000) The impact of animated interface agents: A review of empirical research, *International Journal of Human-Computer Studies*, 52, pp. 1-22.
- Druin A.e. (ed.) (1996) *Designing Multimedia Environments for Children*. John Wiley & Sons: USA.
- Druin A.e. (ed.) (1999) *The Design of Children's Technology*. Morgan Kaufmann: San Francisco, CA.
- Everyday Classroom Tools Project, Harvard Graduate School of Education <http://www.harvard.edu/ECT/Inquiry/inquiry1.html>.
- Gustavson K. and House D. (2001) Fun or Boring? A web-based evaluation of expressive synthesis for children. *Proc. of Eurospeech 2001*, Aalborg, Denmark, pp. 565-568.
- Haughland S.W. (2000) *Computers and young children*, Report available from the Educational Resources Information Center, ERIC Document Reproduction Service No. ED 438 926.
- Ibister K. and Doyle J. (2004) Meaningful evaluations of interdisciplinary efforts: A proposed taxonomy for ECA research, In: Pelachaud C. and Ruttkay Z., (eds.) *Evaluating Embodied Conversational Agents*, Kluwer Academic Publishers.
- Kulik J.A., Bangert R.L. and Williams G.W. (1983) Effects of computer-based education on secondary school students, *Journal of Educational Psychology*, 75(1), pp. 19-26.
- Lester J.C., Converse S.A., Stone B.A., Kahler S. and Barlow T. (1997) Animated pedagogical agents and problem-solving effectiveness: A large-scale empirical evaluation. *Proc. of Eighth World Conference on Artificial Intelligence in Education*, Kobe, Japan, pp. 23-30.
- Moreno R., Mayer R., Spires H. and Lester J. (2001) The case for social agency in computer-based teaching: Do students learn more deeply when they interact with animated pedagogical agents?, *Cognition and Instruction*, 19(2), pp. 177-213.
- Nass C., Ibister K. and Lee E. (2000) Truth is beauty: Researching embodied conversational agents, In: Cassell J., Sullivan J., Prevost S. and Churchill E., (eds.) *Embodied Conversational Agents*, MIT Press, Cambridge, MA, pp. 374-402.
- Nass C. and Lee K. (2001) Does computer-synthesized speech manifest personality? Experimental tests of recognition, similarity-attraction, and consistency-attraction, *Journal of Experimental Psychology: Applied*, 7(3), pp. 171-181.
- Nass C. and Lee K.L. (2000) Does computer-generated speech manifest personality? An experimental test of similarity-attraction. *Proc. of Conference on Human Factors in Computing Systems: CHI 2000*, ACM Press, New York, NY., pp. 329-336.

- Nass C., Steuer J. and Tauber E. (1994) Computers are social actors. *Proc. of Conference on Human Factors in Computing Systems: CHI 1994*, ACM Press, Boston, MA, pp. 72-78.
- Oviatt S.L. and Adams B. (2000) Designing and evaluating conversational interfaces with animated characters, In: Cassell J., Sullivan J., Prevost S. and Churchill E., (eds.) *Embodied Conversational Agents*, MIT Press, Cambridge, MA, pp. 319-343.
- Richmond V., Gorham J. and McCroskey J. (1987) The relationship between selected immediacy behaviors and cognitive learning, *Communication Yearbook*, 10, pp. 574-590.
- Rickenberg R. and Reeves B. (2000) The effects of animated characters on anxiety, task performance, and evaluations of user interfaces. *Proc. of Conference on Human Factors in Computer Systems: CHI 2000*, ACM Press: The Hague, Amsterdam, pp. 49-56.
- Sallinen-Kuparinen A. (1992) Teacher communicator style, *Communication Education*, 41(2), pp. 153-166.
- Scherer K.R. (1979) Personality markers in speech, In: Scherer K.R. and Giles H., (eds.) *Social Markers in Speech*, Cambridge Univ. Press, Cambridge, UK., pp. 147-209.
- Smith B.L., Brown B.L., Strong W.J. and Rencher A.C. (1995) Effects of speech rate on personality perception, *Language and Speech*, 18, pp. 145-152.
- Tusing K.J. and Dillard J.P. (2000) The sounds of dominance: Vocal precursors of perceived dominance during interpersonal influence, *Human Communication Research*, 26, pp. 148-171.