Towards Adapting Fantasy, Curiosity and Challenge in Multimodal Dialogue Systems for Preschoolers

Theofanis Kannetis and Alexandros Potamianos
Dept. of Elec. & Comp. Engineering, Technical Univ. of Crete, Chania 73100, Greece
thkannetis@telecom.tuc.gr, potam@telecom.tuc.gr

ABSTRACT

We investigate how fantasy, curiosity and challenge contribute to the user experience in multimodal dialogue computer games for preschool children. For this purpose, an on-line multimodal platform has been designed, implemented and used as a starting point to develop web-based speech-enabled applications for children. Five task oriented games suitable for preschoolers have been implemented with varying levels of fantasy and curiosity elements, as well as, variable difficulty levels. Nine preschool children, ages 4-6, were asked to play these games in three sessions; in each session only one of the fantasy, curiosity or challenge factor was evaluated. Both objective and subjective criteria were used to evaluate the factors and applications. Results show that fantasy and curiosity are correlated with children's entertainment, while the level of difficulty seems to depend on each child's individual preferences and capabilities. In addition, high speech usage and high curiosity levels in the application correlate well with task completion, showing that preschoolers become more engaged when multimodal interfaces are speech enabled and contain curiosity elements.

Categories and Subject Descriptors

H.5.2 [**User Interfaces**]: [Evaluation/methodology; Voice I/O; Natural language; Graphical user interfaces (GUI)]

General Terms

Experimentation, Human Factors, Measurement, Performance

1. INTRODUCTION

In the past few years, multimodal systems are becoming increasingly part of our everyday life, e.g., mobile communication devices. Multimodal systems combine multiple input and output modalities, such as, keyboard, pen, speech, touch/multi-touch, in order to increase the naturalness, robustness and efficiency of human-computer interaction. One interesting and relevant field of research in this area is multimodal dialogue systems for children. Although children are early adopters of new technologies and interfaces, designing multimodal systems for children is challenging both from the core technology development and the human factors standpoint. Core technology challenges include getting speech recognition technology to work for children users. Interface and human factor challenges have to do with the interaction patterns of children (mix of exploration and exploitation) and the variable capability in using a specific modality (e.g., language, mouse). Overall, variability is one of the greatest challenges when designing multimodal interfaces for children, one size does not fit all.

Recent studies have shown that acoustic and linguistic variability is higher for children than for adults, and this seriously affects speech recognition performance [11]. Also, as shown in [12], children are very different compared to adults at the acoustic, linguistic and interaction levels. In studying integration patterns (voice and gestures) in children, Xiao and colleagues [28] have shown that modality usage was similar between children and adults, although children tend to use both input modes simultaneously rather than sequentially.

Although higher variability and different interaction patterns create additional challenges, there has been notable efforts in the literature for designing, implementing and testing prototype multimodal systems for children. Early speech-enabled prototypes specifically aimed at children included word games for preschoolers [21], aids for reading [15] and pronunciation tutoring [20]. Recently a number of systems with advanced spoken dialogue interfaces, multimodal interaction capabilities and/or embodied conversational characters have been implemented [12, 9, 5, 8]. However, almost all of these systems have focused in the age group 6-15.

A significant advantage when working with the 6-15 age group is that experimental conditions can be more easily controlled, subjects are collaborating and the subjective evaluation results are easier to interpret. For the 4-6 age group speech technology can be very relevant, especially since children are not very adept at using traditional human-computer interfaces, i.e., keyboard and mouse. In this work, we have designed and implemented an on-line (web-based) multimodal platform, in order to be able to quickly prototype, deploy and evaluate multimodal dialogue systems for preschoolers. Using this platform we have designed five such games that use speech and mouse as input modalities.

At ages 4-6, learning and playing are intertwined activi-

ICMI-MLMI'09, November 2–4, 2009, Cambridge, MA, USA.

ties. Thus the main goal of a successful game for preschoolers is to provide fun, excitement and engagement. Several theoretical studies have attempted to identify what is "fun" in a game. According to Malone [14] the essential characteristics of a good computer game can be organized into three categories: fantasy, curiosity and challenge. Alternatively, Lazzaro [10] identified 4 relevant categories (hard fun, easy fun, altered states and the people factor) based on Malone's factors and facial expressions/data obtained from actual games. Another well known study is the theory of flow [7], i.e., strong involvement in a task occurs when the skill of an individual meets the challenges of the task. Finally, in the field of entertainment capture, Yannakakis [22] showed that the player-opponent interaction is a major factor in entertainment.

Based on these prior works, our goal in this paper is to identify how fantasy, curiosity and challenge affect the entertainment value of multimodal dialogue computer games for preschoolers. In previous work with children (ages 8-10) playing on Playware Game Platform, Yannakakis et al [23] has shown that fantasy is correlated with entertainment but curiosity and difficulty depends on each child's preferences. However, it is unclear if these results hold for younger children interacting using a multimodal dialogue interface. Our second goal, is to investigate how these factors can be adapted to increase the entertainment value of the game, i.e., which are good indicators (or features) of the "right" fantasy, curiosity and challenge level in a game. For this purpose both interaction patterns and acoustic features of the speech input will be studied.

The remainder of this paper is organized as follows. First the architecture of the multimodal platform is described in Section 2. Then the functionality and user interface of the five multimodal games are presented in Section 3.1. The development of fantasy and curiosity triggers, as well as, the different levels of game difficulty are outlined in Section 3.2. In Section 4, objective and subjective evaluation results are presented for nine subjects. The implications for designing adaptive multimodal games for preschoolers is given in Section 5 and the paper concludes with Section 6.

2. ON-LINE MULTIMODAL PLATFORM

The main advantage of building a web-based platform for multimodal game-development is that it can be used for (remote) data collection and analysis of educational software and games. The collected data can be used to train language and acoustic models for automatic speech recognition (ASR) and for analysis of user interaction patterns to improve or adapt the user interface. Educational software and games are also used extensively by linguists and psychologists, e.g., to diagnose and solve language development problems. In this work, we use the platform to study children-computer interaction for preschool children and to investigate how we can adapt Malone's quality characteristics in order to improve the user experience.

2.1 System architecture

The system follows a modular architecture, the full functionality of the system being the result of the collaboration between the modules. The architecture of the system is shown in Fig. 1.

Since this is a web-based platform it is (by nature) distributed. The *Application Manager* is responsible for the

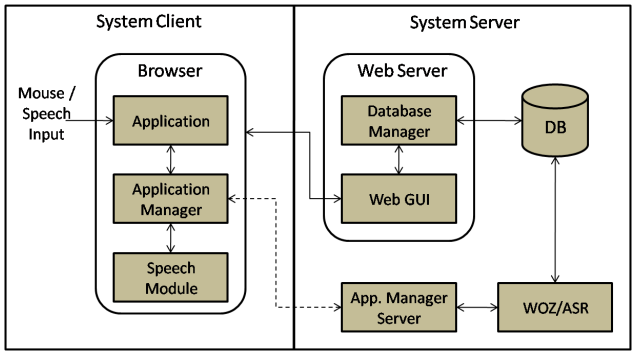


Figure 1: The modular architecture of the platform.

synchronization and cooperation of the modules. It consists of two parts that follow the client/server architecture, i.e., the client/browser side and the server side. The two parts communicate through a two way socket connection. The *Speech Module* is responsible for capturing and streaming the audio, as well as, performing the voice activity detection (VAD) to determine if the user is speaking. When voice activity is detected the Speech Module initiates the audio capture. At the same time, the streaming of audio data to the client part of Application Manager begins. Finally, the multimodal *Application* module may contain any interactive application implemented by the system designer. In our case, we have integrated five preschool games into a single application, as explained in Section 3. All games were implemented in Flash [1] in order to provide an easy and platform-independent way to manipulate sounds, animations and graphics.

On the server side of the system, the *Application Manager* (server part) is responsible for executing the speech requests that are being received from the client side of the platform. It receives and then streams the audio data to the *ASR/WoZ* module for automatic recognition or manual transcription (by the wizard). The recognition results are sent back to the client part of the Application Manager. The module also receives and stores the necessary log files for further processing. On the server side of the system, we have also implemented the *Web Interface Module* and the *Database Communication Module*. The web interface was designed using Java Server Pages (JSP) technology. Using the web interface users can register and login to the platform. Functionality such as profile management and preferences configuration (e.g., microphone configuration) is also provided. The Database Communication Module is responsible for all the necessary database queries. Both the Web

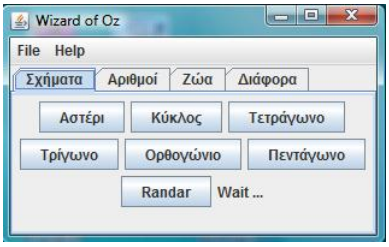
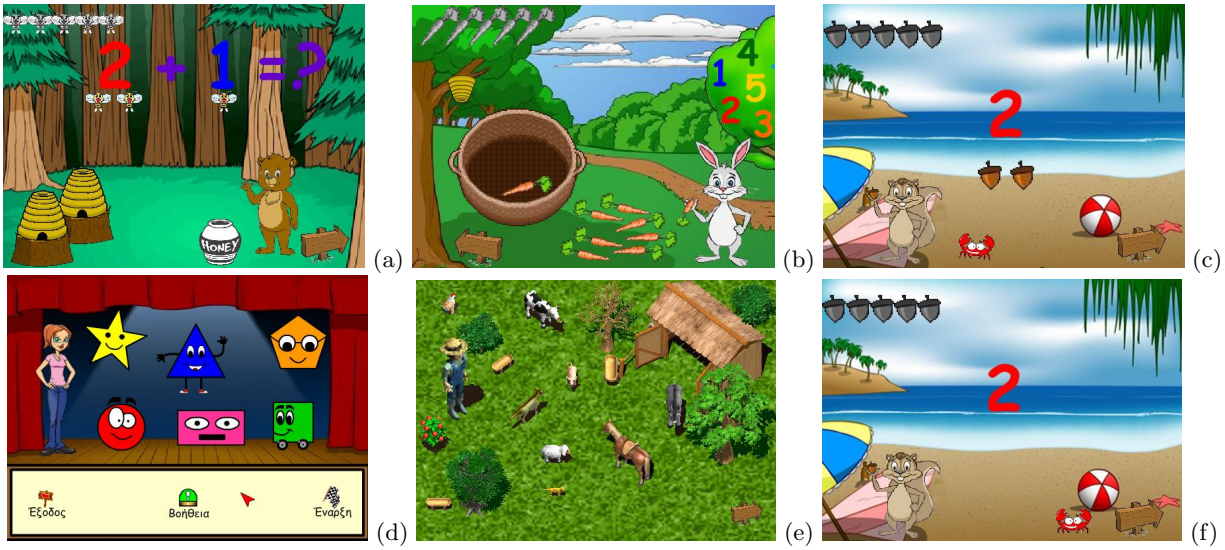


Figure 2: The WoZ module of the platform. Tabs represent different games.



(a) (b) (c) (d) (e) (f)

Figure 3: Example screen-shots of the five tasks: (a) addition, (b) quantity comparison, (c),(f) number recognition, (d) shape recognition, and (e) animal recognition.

Interface and Database Communication modules are parts of the Apache Tomcat Web Server.

The *ASR/WOZ* module has been implemented on the server side of the system. In this study, the ASR module has been replaced by a Wizard of Oz (WoZ) module which is operated by a human transcriber. The WoZ module is actually a graphic user interface (GUI) that plays the audio stream received by the Application Manager and allows the wizard to supply the appropriate transcription via a GUI interface (see Fig. 2). Both the audio and transcription files are stored in the database.

3. APPLICATION & INTERFACE DESIGN

One of the main principles that we must keep in mind when design interfaces for children is usability (see the design guidelines in [3]). Although technology is an enabler, usability is the prerequisite for learning and entertainment. Note, however, that for children user requirements vary significantly with age. According to Piaget [17] the cognitive development of a child can be divided into a series of stages with different characteristics. Attention span varies with age, with older children capable of longer periods of attention than younger children [19]. As a result a different mix of sounds, good animation and graphics has to be used for each child to keep him/her engaged. Moreover, most children in the 4-6 age group are at a preliterate level, so the use of text as an output modality must be avoided. Text output is thus substituted or complemented with sounds, graphics and animation.

Recent work in the field of multimodal dialogue systems shows that children enjoy to interact with virtual characters and that the user experience is enhanced if the animated characters possess a specific "personality" and/or social role [12, 8]. For this reason, we have implemented an animated character for each of the task-oriented games; the animated character was (in most cases) assigned the social role of a friend/helper.

Most preschool children cannot use keyboard and mouse efficiently. They can click on specific targets on the screen using the mouse, provided such targets are relatively large. For these children (especially the 3 and 4 year-olds) speech and touch are the most natural interaction modalities. Although speech is a good choice for the specific target group, the age-dependent acoustic and linguistic variability in children's speech [12, 11] makes automatic speech recognition for children more difficult. In recent years, various algorithms have been proposed to address this problem, e.g., by adapting the acoustic space of children's speech to that of adults [18]. In order to avoid speech recognition errors, we have use a WoZ setup instead, in our study. Henceforth, we assume that there are no speech recognition errors (with the exception of the occasional wizard error).

3.1 Game functionality

We have built a single application consisting of five tasks based on popular preschool activities. The tasks selected were (the target age group for each task is shown in parenthesis): animal recognition (ages 3-4), shape recognition (ages 4-5), quantity comparison (ages 3-4), number recognition (ages 5-6) and addition (ages 5-6). Example screenshots of the implemented tasks are shown in Fig. 3. For each game an embodied agent guides the child through the task. Both mouse and speech are enabled as input modalities. Animation, sounds, graphics, prerecorded prompts and synthesized text-to-speech prompts (where necessary) were used as output. The list of tasks is described next:

- The *animal recognition* task is taking place in a farm. First the voice of an animal is heard and then the farmer asks the child to select the appropriate animal in order to guide it into the farm. There are (up to) nine different farm animals in the game (see Fig. 3(e)).
- The *number recognition* task takes place at the beach where an animated character (squirrel) asks the child to identify which number (1-9) is shown on the screen (see Fig. 3(c),(f)).
- The *shape recognition* game takes place in a theater. Each time one of the shapes (star, circle, square,

rectangle, triangle, pentagon) appears on stage and the child must identify it. The animated character (teacher) provides help and guides the child through the task (see Fig. 3(d)).

- For the *quantity comparison* task, an animated character (rabbit) puts some items inside and some outside of a basket. The rabbit asks then the child to determine whether the items inside the basket are more (or less) than those outside (see Fig. 3(b)).
- For the *addition* task, the child must help an animated character (bear) to collect some honey from the beehives. A simple addition task appears on the screen (two numbers summing up to 2-9). For each correct answer, a bee fills a honey jar with honey (see Fig. 3(a)).

3.2 Fantasy, curiosity and challenge

In this section, we describe how variable levels of fantasy, curiosity and challenge have been implemented for each of the five tasks. According to Malone [14], curiosity is the motivator to learn independently of any goal seeking or fantasy-fulfillment. Specifically "games can evoke players curiosity by providing environments that have an optimal level of informational complexity". That means that those environments "should be neither too complicated nor too simple and should be novel and surprising but not completely incomprehensible". Finally in order for a computer game to be challenging according to Malone a goal must be provided whose attainment is uncertain.

The challenge element is crucial in a game. If a game is too easy, the outcome is likely to be certain, making the game predictable and boring. If it's too hard players are quickly demotivated. This is well-understood by the gaming industry and thus most computer games are playable at different difficulty levels. We have implemented three different levels of difficulty for each of the five tasks. For example, for number recognition the system asks for numbers from one to five at difficulty level 0, from five to nine at level 1, and from one to nine but without the helping items underneath each number at level 2 (compare Fig. 3(c) and (f)). The implementation of the three difficulty levels is shown in Table 1 for each task. Henceforth, the terms challenge and difficulty are used interchangeably.

Fantasy often makes computer games more interesting. Almost every game requires the player to take on a new role (fantasy identity), a process that is apparently very fulfilling. In our work, we use the intrinsic type of fantasy as defined by Malone [14], i.e., the use of a skill is required to achieve some fantasy goals. We have implemented intrinsic fantasy by taking the existing task oriented games and adding to them a fantasy goal, namely, helping an alien that crashed to earth return to his planet. In order to implement different fantasy levels, short animations were also added to each task (triggered fantasy elements). For example, for the numbers recognition task (see Fig. 3(c) and 3(f)), the crab starts walking around making noises when the child clicks on the crab or says "crab". Thus, in our implementation the three different fantasy levels are: without story, with story but without fantasy triggers, and with story and fantasy triggers.

Curiosity is the less obvious factor. Malone identifies two main features of curiosity: sensory curiosity, or the attraction to the environment (sounds, movement, images, etc) and cognitive curiosity or a desire to bring better "form" to one's knowledge structures. Some of the ways to achieve this according to Malone are: rewards, information representation system and surprising feedback. We have implemented several of these elements in our application. A bar representing (progress with) correct answers has been added at the top of the screen for each task. Furthermore we have implemented the incentive of a reward. When a child wins a game task an object passes to his possession. According to Malone, the "easy" way to engage users' curiosity and have surprising feedback is by using randomness. For example, the animated characters now randomly appear in each task depending on the curiosity level. Also the system proposes random tasks to the children based on the curiosity level and children's age. Finally some graphics that appear on stage (e.g., answer bar items) are now selected randomly. Table 1 summarizes how varying degrees of fantasy, curiosity and difficulty have been implemented in our application.

3.3 Game flow diagram

All five tasks follow the same flow diagram shown in Fig. 4. After a small introduction, children can choose to proceed to the main task or leave the game. If a child chooses to play a specific game, the system generates a question (based on difficulty level) and the animated agent asks the child to answer it. At that stage, children can either provide an answer or trigger some fantasy elements in the game.

If the child gives the correct answer then the system generates a new question. When a wrong answer is given or the child does not provide any input (time-out), the agent repeats the same question. After three wrong answers, the agent provides the correct answer and the system generates the next question. Each game concludes after the child gives five correct answers. The child can leave the task or trigger some fantasy elements at any time. We have also implemented a time-out; when a child delays an answer or takes no action, the agent repeats the question. Note that based on the curiosity value, the game selects the agent with whom the child will interact and displays (or not) the answer bar.

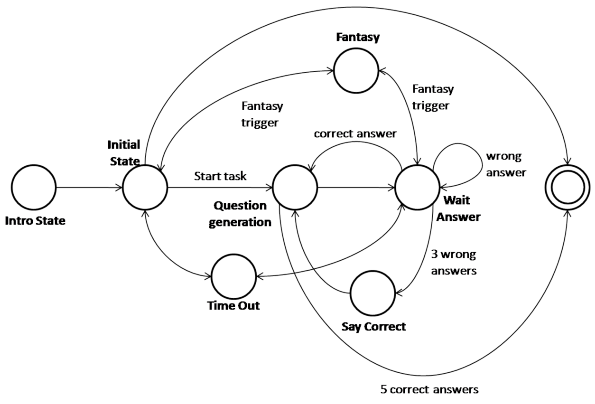


Figure 4: Game flow diagram.

4. METHODOLOGY AND RESULTS

The evaluation of the system took place in a noisy preschool environment using a WoZ experimental setup. Nine native Greek speakers, four to six years old, participated in the study by playing different versions of the ap-

Value	Fantasy	Curiosity	Difficulty				
			Farm	**More/Less**	**Numbers**	**Addition**	**Shapes**
0	No story or fantasy triggers	No correct answers bar and no randomness	Select from 5 different animals	Item difference is 6-8	Numbers from 1-5 with item help	Add up to 2-5 with item help	Star, circle, square
1	Story but no fantasy triggers	Correct answers bar but no randomness	Select from 7 different animals	Item difference is 3-5	Numbers from 5-9 with item help	Add up to 5-9 with item help	Star, circle, square, triangle
2	Both story and fantasy triggers	Both correct answers bar and randomness	Select from 9 different animals	Item difference is 1-2	Numbers from 1-9 without item help	Add up to 2-9 without item help	Star, circle, square, triangle, rectangle and pentagon

Table 1: The three levels of fantasy, curiosity and difficulty as implemented in our application. Implementation of difficulty is task dependent.

plication (at different values of fantasy, curiosity and difficulty). Five of them were boys and four of them girls. All subjects believed that they were interacting with an automated system, i.e., they had no knowledge of the existence of a wizard.

In order to familiarize themselves with the system, each child played the tasks appropriate for his/her age once, using both mouse and voice. After finishing the demo session, each child was asked to play 3 different versions (sessions) of the game and choose the one that he/she enjoyed most[1]. At each session only the value of one factor (fantasy, curiosity or difficulty) was modified, while the values of the two other factors remained constant at level 1. The order that each factor and factor level was presented to the child was randomized. Thus, at each session the child played three versions of the application (one for each of the three levels of the relevant factor). Each user played at least once all the tasks that are suitable for his/her age as discussed in Section 3.1. Overall, each child played 3 sessions, corresponding to 9 different application setups (for each application run each child played 3-5 different tasks). Note that an adult was present during the data collection (sitting next to the child) to help and guide the child through the application, as needed. After the completion of each session the children were asked to evaluate the system by participating in a subjective assessment.

4.1 Evaluation methodology

The evaluation of multimodal dialogue systems is a complex task and different metrics (objective and/or subjective) are typically used to evaluate different aspects of such systems [2]. Since we are mainly interested in investigating how fantasy, curiosity and challenge affect children satisfaction, it is important to measure the correlation between these three factors and objective/subjective interaction metrics. The following objective criteria are reported: average response time, task completion, wrong answers and input modality usage. Response time is defined as the time that elapses from the end of a system prompt until the child completes his/her answer (stops talking or clicks the mouse on a valid target). We separated the response time to inactivity time (end of a system prompt until first voice or mouse activity detection) and interaction time (response minus inactivity time)[16]. In addition to these objective metrics, we also report the most enjoyable system setup that each child selected for each session. At the end of each session, children participated also in an exit interview; a summary of these subjective opinions is also presented in Section 4.3.

[1]To avoid overloading/tiring the child each session was played at different visits to the preschool. In each visit, a single factor was evaluated, e.g., fantasy.

4.2 Objective evaluation results

In Table 2, a summary of the objective evaluation metrics is shown as a function of age and gender. Specifically, for each age and gender, average response times[2] (sec), inactivity and interaction time, percent number of turns of speech and mouse input, and task completion are shown.

	Age			Gender	
	4	**5**	**6**	**M**	**F**
Avg. Resp. Time	4.78	3.78	3.64	3.78	4.38
Avg. Inact. Time	0.99	1.12	1.04	0.89	1.29
Avg. Inter. Time	3.79	2.66	2.60	2.89	3.09
Mouse usage(%)	16.14	18.90	23.55	20.88	19.79
Speech usage(%)	83.85	81.09	76.44	79.11	80.20
Task compl.(%)	89.65	97.14	97.37	90.32	97.62

Table 2: Objective metrics per age and gender.

Four year-olds have higher average response time (by 1 sec) than five and six year-olds (no significant difference in response time between ages five and six). There is no significant difference in inactivity time for all three ages, thus the 1 sec difference is attributed to interaction time. Five and six year-old children have significantly better task completion statistics (around 97%), while 4 year-olds are close to 90%. This is mainly due to the fact that younger children, when facing a difficulty in a task, they often chose to play another task. Older children are usually more persistent and insist until completing the task at hand. The average response time for girls is slightly higher than that of boys (both inactivity and interaction times are higher). However, task completion percentage for girls is significantly

[2]Note that the average response time is computed using only the games that are suitable for each age group (as discussed in Section 3.1).

higher than that of boys (97.62% and 90.32% respectively). Again this difference can be attributed to persistence. In terms of modality usage, we observe a drop of speech usage with increasing age. At the age of four the mouse input usage is close to 16%. At the age of five 19% and at the age of six 23%. This is partly due to the familiarity that older children have with the mouse input device, as explained next. However, speech remains the main input modality for all age groups. Modality usage is similar for boys and girls.

Factor pair	Corr. Coef.	p-value
Resp. Time/Age	-0.2202	0.0712
Mouse Skill/Age	0.6272	0
Resp. Time/Mouse skill	-0.2540	0.0366
Resp. Time/Wrong Ans.	0.4143	0.0005
Inact. Time/Gender	0.2886	0

Table 3: Correlations for time, mouse skill, age.

In Table 3, the correlation between various objective factors and age is shown along with their p-values computed using the one-way ANOVA test. As expected, there is a negative correlation between response time and age, i.e., as the children grow their response time improves. Also there is positive correlation between mouse skill[3] and age. Thus as the children grow, the mouse skill improves and they use mouse input more. Also note the negative correlation between average response time and mouse skill, i.e., as the mouse skill improves the average response time falls[4]. As expected, response time is positively correlated with wrong answers, children become increasingly cautious when they make mistakes. Finally, gender and inactivity time are positively correlated (girls have on average higher inactivity time than boys).

Factor pair	Corr. Coef.	p-value
Speech usage/Age	-0.2579	0.0338
Speech usage/Resp. Time	0.1996	0.1026
Speech usage/Inter. Time	0.1816	0.1287
Speech usage/Wrong answers	0.2726	0.0245
Speech usage/Task compl.	0.4061	0.0006

Table 4: Correlations for speech usage.

In Table 4, the correlation between various objective metrics and speech usage is shown. As shown also in Table 2, there is negative correlation between speech usage and age, due to the limited mouse skills of youngsters. There is also moderate correlation between speech usage and response time. This is expected because for the majority of tasks in our applications, speech is "slower" than mouse as the input modality. Note the positive correlation between speech usage and wrong answers. Interestingly there is also positive correlation between speech usage and task completion. We conclude that when interacting by speech, children are both more spontaneous (thus the higher rate of errors) and more motivated/engaged (thus the higher task completion). This is an interesting case, where although mouse input is the more efficient modality in terms of speed, speech is more efficient in terms of task completion!

[3]Mouse skill was evaluated (at three levels) by the person performing the exit interview, based on the iteration of the child with the system.

[4]We have verified that this trend is not due to the negative correlation between response time and age, in fact, it is especially pronounced for the 4 year-old age group.

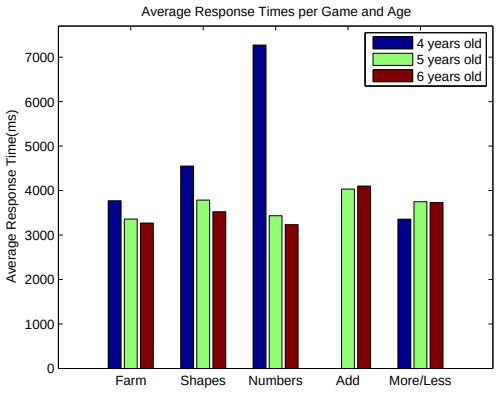


Figure 5: Avg. response times per task and age.

In Fig. 5, the average response time per task and age are shown. As expected, average response times for four year-olds children are higher than five and six year-olds. Also five year-olds and six year-olds have similar response times in most tasks. The trend is consistent across tasks, with the exception of the comparison task ("More/Less"). The very high response time for the "Numbers" task is due the difficulty that four year-old have performing this task.

4.3 Subjective evaluation results

Next we evaluate how fantasy, curiosity and difficulty/challenge affect the user experience. As shown in Fig. 6(a), most children preferred the application with higher levels of fantasy and curiosity. Specifically, six out of the nine children picked the version of the game with story and fantasy triggers (fantasy level 2). Also six out of the nine children chose the game version with randomly created characters, random task proposals and answer bar (curiosity level 2). In Fig. 6(b), the selected "best" system configuration is shown. Systems with high values of fantasy, curiosity and difficulty were the most popular among the children.

In order to compute the correlation between the three factors and entertainment, we have labeled each system version as "entertaining" or "not entertaining" based on the child's preferences, i.e., for each session/factor one system setup

Factor	Corr. Coef.	p-value
Fantasy/Entert.	0.2778	0.0120
Curiosity/Entert.	0.2778	0.0120
Difficulty/Entert.	0.1667	0.1370

Table 5: Correlation between entertainment and the three factors.

(the one picked by the child) is labeled "entertaining" and the other two "not entertaining". Table 5 shows the correlation coefficients (and their corresponding p-values) between the level of each factor (fantasy, curiosity, challenge) and entertainment (binary variable defined above). Both fantasy and curiosity are positively correlated with child's entertainment.

Per the challenge factor, it seems that the preferred level of difficulty is very much child-dependent. Two children chose

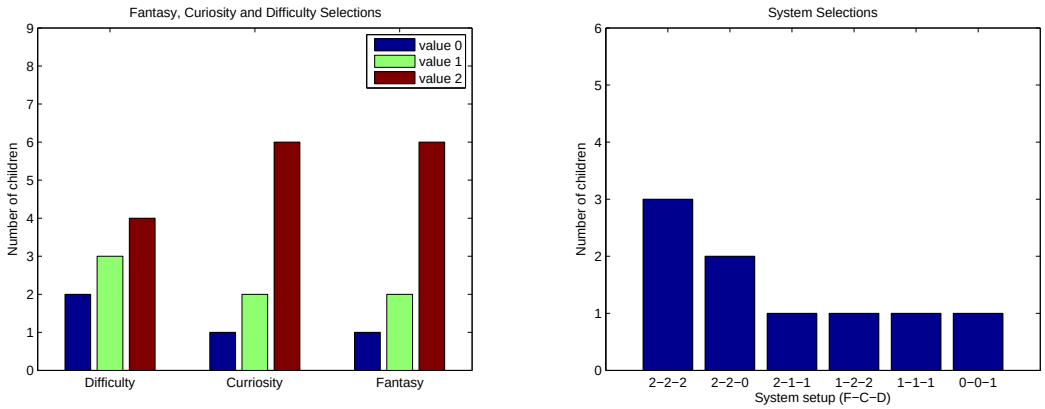


Figure 6: (a) Histogram of subjective optimal levels of fantasy, curiosity and difficulty. (b) Histogram of system picked best overall (by the user), e.g., 2-1-0 corresponds to fantasy level 2, curiosity 1, difficulty 0.

easy difficulty, three medium and four children selected the games with high difficulty level. As a result the correlation between difficulty and entertainment is modest[5].

Factor pair	Corr. Coef.	p-value
Fantasy/Speech usage	0.2236	0.0668
Curiosity/Inter. Time	0.1910	0.1186
Curiosity/Task Compl.	0.2850	0.0185
Difficulty/Wrong Ans.	0.1909	0.1190
Difficulty/Inact. Time	0.1985	0.1046

Table 6: Correlation between the three factors and objective metrics.

In Table 6, the correlation between the three Malone factors and various objective metrics is shown. The results show correlation between fantasy and speech usage, i.e., higher levels of fantasy motivate higher usage of the speech input modality. There is also correlation between curiosity and interaction time, as well as, between curiosity and task completion. This indicates that high levels of curiosity is another motivation for children to complete the selected task (unlike speech usage, however, curiosity elements also increase cognitive load and/or reduce spontaneity). Finally, as expected there is positive correlation between difficulty and wrong answers, as well as, between difficulty and inactivity time due to increased cognitive load for more difficult tasks.

The results from the exit interview are shown in Table 7. Note that for the last three questions only the most popular answers are shown. Most children enjoyed interacting with the system using speech, liked the graphics, sounds and animation of the games, and enjoyed the underlying story. Finally, most children would like to interact again with the application in the future.

5. TOWARDS FACTOR ADAPTATION

In order to create games that adapt to the children's preferred level of fantasy, curiosity and difficulty we have investigated the correlation between various objective metrics

[5]It would be interesting to define a user-dependent challenge metric based on the capabilities of the child for a specific task; perhaps this metric would be better correlated with entertainment value.

Question	YES	NO
Did you like that you could speak to game characters?	96.3%	3.7%
Did you like the game graphics, sound and animations?	88.8%	11.2%
Did the characters listen to you when you talk to them?	92.6%	7.4%
Did you understand what they said to you?	92.2%	7.8%
Did you like the story?	88.8%	11.2%
Would you like to see a new story in the future?	96.3%	3.7%
Does the headset annoy you?	11.2%	88.8%
Would you like to play the games again in the future?	85.2%	14.8%
Who was your favorite character?	Rabbit, Farmer	
What was your favorite game?	Farm, Addition	
What did you not like about the games?	Bees, Numbers recogn. task	

Table 7: Exit interview results.

and these factors. A preliminary analysis of interaction patterns, such as, average interaction time, hesitations, inactivity times, correct/wrong answers have been analyzed. In addition, information from other sources such as voice, video and physiological measurements could be used as features. In [24, 25, 26], physiological measurements such as children's heart rate (HR), blood volume pulse (BVP) and skin conductance (SC) signals, are used as features to predict engagement. Similarly in [6], physiological measurements are used in order to adapt difficulty for a Tetris game.

More recently there has been interest in emotion recognition and modeling of children's mood in spoken dialogue and gaming applications [27]. Emotions are an important part of the gaming experience. Identifying negative emotions can help identify hot-spots in the interaction. Audio, linguistic, pragmatic and visual information can be combined to obtain a good prediction of the child's emotional state [4].

Preliminary analysis of user interaction patterns and features extracted from the user's speech input (e.g., pitch and energy statistics) have shown moderate correlation with optimal levels of fantasy, curiosity and difficulty, e.g., corre-

lation values between pitch statistics (average session pitch minus average speaker pitch) and fantasy was 0.1621. More research work (and more data) are needed to identify good predictors of user preferences, as well as, to combine them in a multimodal framework (interaction, audio, video, physiological features) to maximize the engagement and enjoyability of child-computer interaction.

6. CONCLUSIONS

In this work, we designed and implemented an on-line multimodal platform in order to examine child-computer interaction for ages 4-6. Five preschool games were implemented at various levels of fantasy, curiosity and difficulty. The main conclusions from the evaluation were that: (i) fantasy and curiosity are positively correlated with children's entertainment, while the level of difficulty seems to depend on each child's individual preferences, (ii) when interacting by speech, children are both more spontaneous and more motivated, leading to higher task completion, (iii) high fantasy levels correlate with higher speech usage, and (iv) high levels of curiosity motivates children and leads to higher task completion. Preliminary experiments also showed that interaction patterns and acoustic features are indicators of optimal levels of fantasy, curiosity and difficulty. Nevertheless more experiments with more subjects and different system setups are needed in order to better understand how to design adaptive multimodal dialogue systems for preschool children that maximize engagement and enjoyability.

Acknowledgments

The authors wish to thank Prof. Georgios N. Yannakakis (IT University of Copenhagen) and Dr. Angeliki Mouzaki (Univ. of Crete) for valuable discussions, Manolis Perakakis for comments on evaluation issues, Spyros Meliopoulos for implementing the "Farm" task, and Giorgos Evgeniadis for helping out in the evaluation. We also wish to thank the preschool teachers for their help during the data collection.

7. REFERENCES

[1] Macromedia Flash support page. Available at: http://www.adobe.com/products/.

[2] Beringerb, N., Kartal, U., Louka, K., Schiel, F., and Turk, U., "Promice: A procedure for multimodal interactive system evaluation," in *Proc. of the LREC Workshop on Multimodal Resources and Multimodal Systems Evaluation*, Las Palmas, Spain, 2002.

[3] Bruckman, A. and Bandlow, A., "Human-computer interaction for kids," In *the Human-Computer interaction Handbook: Fundamentals, Evolving Technologies and Emerging Applications*, J. A. Jacko and A. Sears, Eds. Human Factors And Ergonomics. L. Erlbaum Associates, Hillsdale, NJ, 428-440, 2003.

[4] Busso, C. *et al*, "Analysis of emotion recognition using facial expressions, speech and multimodal information," in *Proc. of ICMI*, Oct. 2004.

[5] Cassell, J. and Ryokai, K., "Making Space for Voice: Technologies to Support Children's Fantasy and Storytelling," *Personal Technologies*, vol. 5, 2001.

[6] Chanel, G., Rebetez, C., Betrancourt, M., and Pun, T., "Boredom, Engagement and Anxiety as Indicators for Adaptation to Difficulty in Games," in *Proc. of the 12th Intl. Conf. on Entertainment and Media in the Ubiquitous Era*, Tampere, Finland, Oct. 2008.

[7] Csikszentmihalyi, M., "Flow: The Psychology of Optimal Experience," New York: Harper & Row, 1990.

[8] Gustafson, J., Bell, L., Boye, J., Lindstrom, A., and Wiren, M., "The NICE Fairy-tale Game System," in *Proc. of SIGdial 04*, Boston, MA, Apr. 2004.

[9] Hagen, A., Pellom, B., and Cole, R., "Children's speech recognition with application to interactive book and tutors," in *Proc. ASRU Workshop,* 2003.

[10] Lazzaro, N., "Why We Play Games: Four Keys to More Emotion Without Story," *Technical Report*, XEO Design Inc., 2004, available at: http://www.xeodesign.com.

[11] Lee, S., Potamianos, A., and Narayanan, S., "Acoustics of children's speech: Developmental changes of temporal and spectral parameters," *Journal of the Acoustical Society of America*, vol. 105, pp. 1455-1468, Mar. 1999.

[12] Narayanan, S. and Potamianos, A., "Creating conversational interfaces for children," *IEEE Trans. on Speech and Audio Processing*, vol. 10, pp. 65-78, Feb. 2002

[13] Narayanan S., Potamianos A., and Wang H., "Multimodal systems for children: building a prototype," in *Proc. EUROSPEECH*, Budapest, Hungary, Sept. 1999.

[14] Malone, T. W, "What make things fun to learn? A study of intrinsically motivating computer games," In *Proc. of the 3rd ACM SIGSMALL Symposium and the First SIGPC Symposium on Small Systems*, Palo Alto, California, Sept. 1980.

[15] Mostow, J., Hauptmann, A. G., and Roth, S. F., "Demonstration of a reading coach that listens," in *Proc. of the ACM Symposium on User Interface Software and Technology*, pp. 77-78, 1995.

[16] Perakakis, M. and Potamianos, A., "A study in efficiency and modality usage in multimodal form filling systems." *IEEE Trans. on Audio, Speech, and Language Processing*, 16(6):1194-1206, Aug. 2008

[17] Piaget, J., "Science of Education and the Psychology of the Child," *Orion Press*, New York, 1970.

[18] Potamianos, A. and Narayanan, S., "Robust recognition of children's speech," *IEEE Trans. on Speech and Audio Processing*, vol. 11, pp. 603-616, Nov. 2003.

[19] Ruff, H. A. and Lawson, K. R., "Development of sustained, focused attention in young children during free play," *Developmental Psychology,* pp. 85-93, 1990.

[20] Russell, M., Brown, B., Skilling, A., Series, R., Wallace, J., Bonham, B., and Barker, P., "Applications of automatic speech recognition to speech and language development in young children," in *Proc. ICSLP,* Philadelphia, PA, Oct. 1996.

[21] Strommen, E. F. and Frome, F. S., "Talking back to big bird: Preschool users and a simple speech recognition system," *Educational Technology Research and Development*, vol. 41, pp. 5-16, 1993.

[22] Yannakakis, G. N., and Hallam, J., "Evolving Opponents for Interesting Interactive Computer Games," in *Proc. of the 8th International Conference on Simulation of Adaptive Behavior*. The MIT Press, pp. 499-508, 2004.

[23] Yannakakis, G. N., Hallam, J., and Lund, H. H., "Comparative Fun Analysis in the Innovative Playware Game Platform," in *Proc. of the 1st World Conference for Fun n' Games*, pp. 33-37, 2006.

[24] Yannakakis, G. N., Hallam, J., and Lund, H. H., "Capturing Entertainment through Heart-rate Dynamics in the Playware Playground," in *Proc. of the 5th International Conference on Entertainment Computing, Lecture Notes in Computer Science*, vol. 4161, pp. 314-317, Cambridge, UK, Sept. 2006.

[25] Yannakakis, G. N. and Hallam, J., "Entertainment Modeling in Physical Play through Physiology beyond Heart-Rate," in *Proc. of the Int. Conf. on Affective Computing and Intelligent Interaction, Lecture Notes in Computer Science*, vol. 4738, pp. 256-267, Lisbon, Portugal, Sept. 2007.

[26] Yannakakis, G. N. and Hallam, J., "Feature Selection for Capturing the Experience of Fun," in *Proc. of the AIIDE'07 Workshop on Optimizing Player Satisfaction*, AAAI Press Technical Report WS-01-01, pp. 37-42, Stanford, June 2007.

[27] Yildirim, S., Min, C.L., Sungbok, L., Potamianos, A., and Narayanan, S., "Detecting Politeness and frustration state of a child in a conversational computer game," in *Proc. EUROSPEECH*, Lisbon, Portugal, 2005.

[28] Xiao, B., Girand, C., and Oviatt, S., "Multimodal integration patterns in children," in *Proc. ICSLP*, pp. 629-632, 2002.