

Fantasy, Curiosity and Challenge as Adaptation Indicators in Multimodal Dialogue Systems for Preschoolers

Theofanis Kannetis¹, Alexandros Potamianos¹, and Georgios N. Yannakakis²

¹Dept. of Elec. & Comp. Engineering, Technical Univ. of Crete, Chania 73100, Greece

²Center for Computer Games Research, IT University of Copenhagen, Copenhagen, Denmark

{thkannetis, potam}@telecom.tuc.gr yannakakis@itu.dk

ABSTRACT

In this paper, 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 five task oriented games suitable for preschoolers, with varying levels of fantasy and curiosity elements, as well as, variable difficulty levels. Nine preschool children were asked to play these games in different configurations and choose the application setup that they enjoyed most. 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, a variety of objective metrics (task completion, interaction time, wrong answers), audio features and emotional state have been investigated as potential features that can predict optimal levels of fantasy, curiosity and difficulty for each child. Emotional state recognition results are also reported.

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

Keywords

Multimodal Interfaces, Graphical user interfaces, Speech interfaces, Children Computer Interaction, Evaluation

1. INTRODUCTION

In recent years, computers have become increasingly popular learning and playing tools, especially for young children.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

ICMI-MLMI'09 Workshop on Child, Computer and Interaction November 5, 2009, Cambridge, MA, USA

Copyright 2009 ACM 978-1-60558-690-8/09/11 ...\$10.00.

At ages 4-6, learning and playing are intertwined activities. 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 [11, 9, 5]. Malone [11] has organized the essential characteristics of a good computer game into three categories: fantasy, curiosity and challenge.

In previous work with children (ages 8-10) playing on Playware game platform [17], it is shown that fantasy is correlated with entertainment but curiosity and difficulty¹ depends on each child's preferences. However, it is unclear if these results hold for younger children interacting using a multimodal dialogue interface. 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, and how these factors can be adapted to increase the entertainment value of the game.

In order to create games that adapt to the children's preferred level of fantasy, curiosity and difficulty it is important to investigate the correlation between various objective metrics and these factors. In addition, information from other sources such as voice, video and physiological measurements could be used as features. In [18, 19, 20], 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 [4], physiological measurements are used in order to adapt difficulty for a Tetris game.

There have 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 [16], aids for reading [12] and pronunciation tutoring [14]. Recently a number of systems with advanced spoken dialogue interfaces, multimodal interaction capabilities and/or embodied conversational characters have been implemented [10, 7, 3, 6]. However, almost all of these systems have focused in the age group 6-15. More recently there has been interest in emotion recognition and modeling of children's mood in spoken dialogue and gaming applications [21]. Emotions are an important part of the gaming experience. Identifying negative emotions can help identify hot-spots in the interaction. Audio, linguistic, pragmatic

¹Challenge and difficulty are terms used interchangeably in this paper.

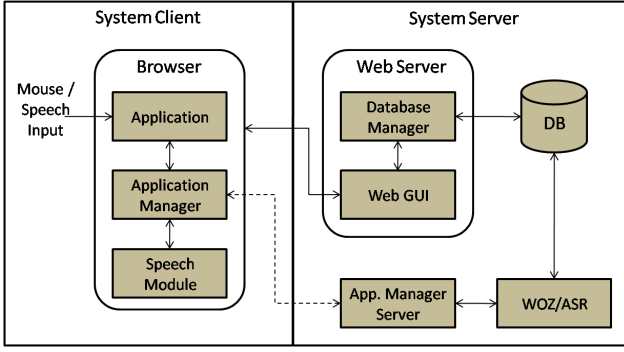


Figure 1: The modular architecture of the platform.

and visual information can be combined to obtain a good prediction of the child’s emotional state [2]. Emotion can also serve as a predictor of the optimal levels of fantasy, curiosity and challenge.

Towards this goal, we have designed and implemented an on-line (web-based) multimodal platform. Using this platform one may quickly prototype, deploy, collect data and evaluate multimodal dialogue systems for preschoolers. Specifically, we have designed five games that use speech and mouse as input modalities, with varying levels of fantasy and curiosity elements, as well as, variable difficulty levels. A detailed description of the system and baseline evaluation can be found in [8]. In this paper, we focus on identifying factors (objective metrics, audio features, emotional state) that can predict the optimal level of the three Malone factors, namely, fantasy, curiosity and challenge. The proposed features are combined into a classifier and preliminary results for factor adaptation are presented. In addition, results from emotional state classification are reported.

The rest of this paper is organized as follows. First a brief overview of the system is presented in Section 2. Then, an overview of the main findings of the system evaluation are discussed in Section 3. Next, results from emotional state classification (two states: negative and non-negative) are presented in Section 4. Finally, we attempt to build classifiers that predict the optimal levels of fantasy, curiosity and challenge in Section 5. We conclude with Section 6.

2. SYSTEM OVERVIEW

The system follows a modular architecture, shown in Fig. 1. Since this is a web-based platform it is (by nature) distributed. The *Application Manager* is responsible for the 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. 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 next. 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.

2.1 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. 2. 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; for more details see [8].

2.2 Fantasy, Curiosity and Challenge

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 [11], 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 2(c)), 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.

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 the 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. Some of the graphics that appear on stage (e.g., answer bar items) are now selected randomly.

Finally, 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.

The implementation of the three fantasy, curiosity and difficulty levels is shown in Table 1 for each task.



Figure 2: Example screen-shots for three of the five tasks: (a) addition, (b) quantity comparison, (c) number recognition

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.

3. SYSTEM EVALUATION

Nine native Greek speakers, ages four to six, participated in the study by playing different versions of the application (at different values of fantasy, curiosity and difficulty). All subjects believed that they were interacting with an automated system, i.e., they had no knowledge of the existence of a wizard. 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 the end of each session each child selected the application setup that he/she enjoyed most. The children also participated in an exit interview. The following objective criteria are also measured: average response time (separated in inactivity and interaction time), task completion, correct/wrong answers and input modality usage.

3.1 Evaluation Results

A summary of the results follows; for more details see [8]. Four year-olds displayed higher average response time (by 1 sec) than five and six year-olds. Also five and six year-old children have significantly better task completion statistics (around 97%), while 4 year-olds are close to 90%. Older children are usually more persistent and insist until they complete the task at hand. Task completion percentage for girls is significantly higher than that of boys (97.62% and 90.32% respectively). 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. However, speech remains their main input modality for all age groups. Modality usage is similar for boys and girls.

The correlation between various objective metrics was investigated. The main findings were that:

- There is positive correlation between response time and wrong answers, i.e., children become increasingly cautious when they make mistakes.
- There is positive correlation between speech usage and wrong answers, i.e., when interacting by speech, children are more spontaneous.
- There is positive correlation between speech usage and task completion, i.e., when interacting by speech, children are more motivated to finish the task.

Next we evaluate how fantasy, curiosity and difficulty/challenge affect the user experience. As shown in Fig. 3(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. 3(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

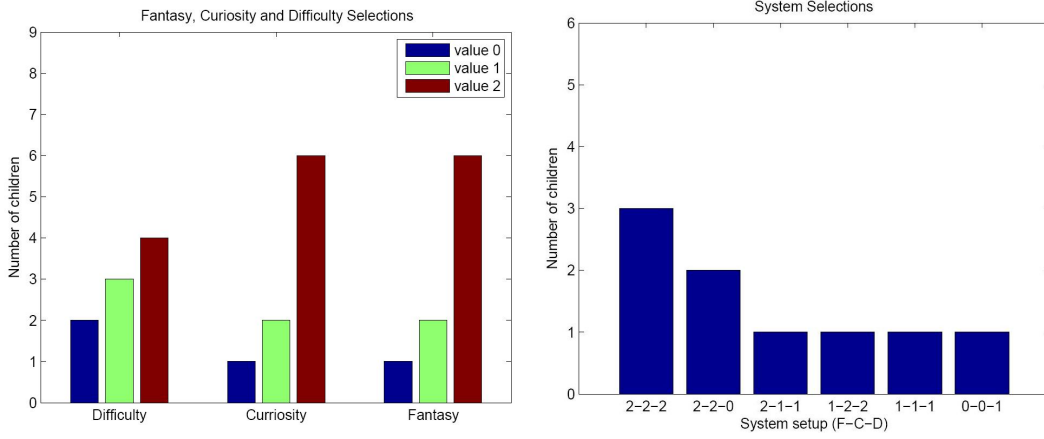


Figure 3: (a) Histogram of subjective optimal levels of fantasy, curiosity and difficulty. (b) Histogram of system picked best overall (by the user), e.g., 2-1-0 corresponds to fantasy level 2, curiosity 1, difficulty 0.

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 2: Correlation between entertainment and the three factors.

(the one picked by the child) is labeled “entertaining” and the other two “not entertaining”. Table 2 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 easy difficulty, three medium and four children selected the games with high difficulty level. As a result the correlation between difficulty and entertainment is modest².

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 3: Correlation between the three factors and objective metrics.

In Table 3, the correlation between the three Malone factors and various objective metrics is shown. The results show a 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

²It 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.

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 increased cognitive load and/or reduced 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 showed that most children enjoyed interacting with the system using speech, liked the graphics, sounds and animation of the games, and enjoyed the underlying story. Also, most children would like to interact again with the application in the future.

4. EMOTION CLASSIFICATION

Three graduate students labeled the collected data and assigned them into five different emotion categories, namely: angry, happy, neutral, bored and sad. In our experiments, we used only the data that all the three libelers agreed on. Due to the fact that for some categories there were not enough data, the labeled data were organized into two categories: non-negative emotions (neutral, happy) and negative emotions (angry, sad, bored). The resulting data consisted of a total of 234 utterances: 189 non-negative and 45 negative.

Thirteen different pitch- and energy-based audio features were extracted from the collected audio data using the Praat toolkit [1]. These features were: minimum, maximum, mean, median and standard deviation of the fundamental frequency (F0) and intensity, as well as, energy, F0 points and duration.

Three different classifiers were trained and evaluated using leave one out cross validation, namely, naive Bayes (Bayes), nearest neighbor (NNR-3) and Bayes with linear discriminant analysis (LDA), i.e., LDA was applied to the feature set before applying them to the Bayes classifier. The results are shown in Table 4. Note that the classification results are shown for the whole data set, as nine-fold cross validation was used to train the models, i.e., models were trained on data from eight children and tested on the ninth child.

The naive Bayes classifier managed to classify 90.03% of the non-negative and 93.42% of the negative samples correctly. The LDA classifier achieved the higher accuracy

Classifier	Non-Negative	Negative	Total
Bayes (%)	90.03	93.42	89.82
NNR-3 (%)	93.17	30.41	87.6
LDA (%)	93.56	93.71	93.62

Table 4: Non-Negative/Negative emotion classification results (% correct).

with 93.56% for the non-negative samples and 93.71% for the negative ones. Finally, NNR-3 achieved very low accuracy (30.41%) for negative samples and high (93.17%) for non-negative. Compared to older children (ages 7-15) and adults [2, 21, 15], the accuracy of emotion classification for younger children is higher. This is probably due to the fact that younger children are more spontaneous, and as a result it is easier to identify their emotional state. It is interesting that the performance of emotion recognition does not follow the same age-trend as that of speech recognition, which is known to degrade for younger children, e.g., see [13].

Factor pair	Correlation	p-value
Negt. Emotion/Inact. time	0.2265	0.0662
Negt. Emotion/Wrong ans.	0.2404	0.0483

Table 5: Correlation between emotional state of the child, inactivity time and wrong answers.

Next, the relationship between emotion and objective metrics was investigated. In Table 5, the correlation between negative emotional state (the percentage of utterances per session labeled as negative), inactivity time and wrong answers is shown. In order to compute the percentage of positive and negative utterances, we labeled each sample using the output from the LDA classifier. The results show that there is positive correlation between inactivity time and negative emotional state of the children. Thus, long inactivity time is an indicator of negative emotions while short inactivity time is indicator of positive (or neutral) emotions. Also positive correlation exists between wrong answers and negative emotions, i.e., wrong answers cause negative feelings to the children. There were no statistically significant correlations between emotional states and the three Malone factors.

5. FACTOR CLASSIFICATION AND ADAPTATION

In this section, we attempt to build classifiers that can predict the optimal levels of fantasy, curiosity and challenge for each user. For this purpose we investigate a variety of features that can be extracted from the interaction patterns of the child-machine interaction, as well as, audio features extracted from the child’s spoken language input. 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., correlation values between pitch statistics (average session pitch minus average speaker pitch) and fantasy was 0.1621.

For each of the three Malone factors a neural network with a single hidden layer (three neurons) is trained. The vector of all the objective metrics and/or audio features estimated for the three sessions that each child played are

given as input, along with the level for the factor (0, 1 or 2). Overall the feature vector consists of: average session time, average response time, average interaction time, speech usage (%), correct answers (%), task completion (%), reported entertainment, fantasy level, curiosity level, difficulty level, positive emotion(%), negative emotion(%), age, and gender. The audio features used are the (session averages of the) ones used for emotion classification (see Section 4). The system is trained to estimate the optimal value of the factor as supplied by the child in the exit interview. The results presented are from 9-fold cross validation, i.e., a separate classifier is trained for each factor from the data of eight children and tested on the ninth child. In addition to neural networks, support vector machine (SVM) and naive Bayes classifiers were also trained, for comparison purposes. Results are reported for a total of 81 sessions.

Neural Network			
	Objective	Audio	Both
Fantasy	35.80%	23.45%	29.62%
Curiosity	38.27%	25.92%	32.09%
Difficulty	37.03%	22.22%	30.86%
Naive Bayes			
	Objective	Audio	Both
Fantasy	35.43%	22.34%	28.52%
Curiosity	37.87%	25.72%	31.58%
Difficulty	37.23%	22.36%	31.06%
Support Vector Machine			
	Objective	Audio	Both
Fantasy	19.75%	16.04%	18.51%
Curiosity	19.75%	14.81%	17.28%
Difficulty	29.62%	22.22%	27.16%

Table 6: Optimal Factor Classification Accuracy for a Neural Network, Naive Bayes and SVM classifier.

The results are shown in Table 6. Classification accuracy is poor; only for the neural network and naive Bayes classifiers for the objective metrics feature set provide results better than chance (33%). Overall, the objective criteria outperform the audio features for this task for all three classification tasks (fantasy, curiosity, difficulty). No gain is obtained when the objective metrics and audio feature sets are combined; this could be due to the simplistic feature-level fusion strategy. The results (especially the poor performance of SVMs) indicate that the amount of data available is not sufficient to accurately estimate the model parameters. In any case, predicting the subjective opinion of the child using objective criteria is a hard task, and more sophisticated feature extraction and classification algorithms should be investigated in the future to improve performance.

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 and investigated how fantasy, curiosity and challenge contribute to the user experience in multimodal dialogue computer games for preschool children. The evaluation results showed that fantasy and curiosity are positively correlated with children’s entertainment, while the level of difficulty seems to depend on each child’s individual

preferences. Next, we investigated the importance of emotional state in multimodal interaction design for preschoolers. Emotion classification experiments showed that a negative emotional state (non-negative vs negative classification) of the child can be identified automatically with 95% accuracy. Negative emotions were correlated with wrong answers and increased response time by the children. Finally, we investigated if objective metrics and audio features can be used to predict the optimal levels for fantasy, curiosity and challenge for each child. Results showed only moderate success when the objective metrics extracted from the interaction were used as features, while audio features provided no improvement. 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.

7. ACKNOWLEDGMENTS

The authors wish to thank 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 and evaluation.

8. REFERENCES

- [1] Boersma, P. and Weenink, D., “Praat: doing phonetics by computer (Version 4.6)”. Available at: <http://www.praat.org/>.
- [2] Busso, C. *et al*, “Analysis of emotion recognition using facial expressions, speech and multimodal information,” in *Proc. of ICMI*, Oct. 2004.
- [3] Cassell, J. and Ryokai, K., “Making Space for Voice: Technologies to Support Children’s Fantasy and Storytelling,” *Personal Technologies*, vol. 5, 2001.
- [4] 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.
- [5] Csikszentmihalyi, M., “Flow: The Psychology of Optimal Experience,” New York: Harper & Row, 1990.
- [6] 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.
- [7] Hagen, A., Pellom, B., and Cole, R., “Children’s speech recognition with application to interactive book and tutors,” in *Proc. ASRU Workshop*, 2003.
- [8] Kannetis, T. and Potamianos, A., “Towards adapting fantasy, curiosity and challenge in multimodal dialogue systems for preschoolers,” in *Proc. ICMI*, Boston, MA, Nov. 2009.
- [9] 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>.
- [10] Narayanan, S. and Potamianos, A., “Creating conversational interfaces for children,” *IEEE Trans. on Speech and Audio Processing*, vol. 10, pp. 65-78, Feb. 2002.
- [11] 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.
- [12] 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.
- [13] 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.
- [14] 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.
- [15] Steidl, S., *Automatic Classification of Emotion-Related User States in Spontaneous Children’s Speech*. PhD thesis, Friedrich-Alexander University, Erlangen, Germany, 2009.
- [16] 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.
- [17] 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.
- [18] 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.
- [19] 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.
- [20] 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.
- [21] 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.