Analyzing Exploration and Exploitation Patterns in Multimodal Dialogue Games for Preschoolers

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

Vassiliki Kouloumenta
Dept. of Elec. & Comp. Engineering
Technical Univ. of Crete
Chania 73100, Greece
vaskou@telecom.tuc.gr

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

ABSTRACT

We design, implement and evaluate a multimodal dialogue system for preschoolers that contains both story-telling and task-oriented game-play elements. Our main goal is to investigate the exploration and exploitation patterns in gameplay of preschoolers, as well as, the interaction patterns when preschoolers interact with Embodied Conversational Agents (ECAs) using voice. The application contains two modes (story and games) that roughly correspond to exploration and exploitation strategies, and the user is allowed to switch between those modes at will. Results show that younger children tend to use exploratory strategies more often, while older children are more efficient in designing such strategies. Although there is a clear age trend, the exploration vs. exploitation dilemma is shown to be very much child dependent. The implications of this study for designing games for preschoolers are discussed.

Keywords

Child Computer Interaction, Exploration and Exploitation, Multimodal Dialogue Games

1. INTRODUCTION

Many decisions in our everyday lives require an exploration of alternatives before committing to and exploiting the benefits of a particular choice. The dilemma is whether to exploit familiar but possibly suboptimal strategies or explore risky but potentially more profitable ones. This is known as the Exploration vs Exploitation (EvE) dilemma. EvE choices are affected by a number of factors, such as the familiarity of the environment, environment change and perceived efficiency gain (vs time lost) for exploratory behavior [3].

As discussed in [11] there is no general solution to the EvE dilemma. However, optimal policies under constrained circumstances have been proposed in the literature: in a landmark paper, Gittins and Jones developed a straightforward means for calculating the optimal strategy for decision making in Multi Armed Bandits problems (The Gittins Index) [5]. Machine learning methods have also been proposed to determine optimal EvE strategies [17]. The authors in [8] assert that preschool children enjoy a mix of exploration and exploitation when interacting with multimodal dialogue games. Designing games for children and especially for preschoolers can be rather challenging. Such challenges are speech recognition technology for such ages [13] and maintaining the child's attention and engagement. In [18, 1], authors state that a good balance between story and game elements is necessary, but yet not easily achievable. Despite these challenges, notable efforts have been done on the design, implementation and testing of prototype dialogue systems for children, [9, 2, 7, 16, 12, 15, 14, 6, 8]. In [14], about 160 children, played an interactive computer game using multimodal input. Project NICE [6] was another notable effort in multimodal game designing for children, where the interaction with the characters is in a fantasy story-telling environment.

In this context, our primary goal is to examine the preferences and patterns children follow in their interchange among exploration and exploitation possibilities in a multimodal game/application we developed.

2. EXPLORATION AND EXPLOITATION IN COMPUTER GAMES

Our approach towards investigating exploration and exploitation strategies focuses on two interaction scenarios/modes: (1) The **exploitation** or **games** mode. The children have the opportunity to use/exploit their existing fundamental knowledge, while playing five tasks based on popular preschool activities. However, initial interaction with each task also contains exploratory elements (from the perspective that it is still something new for the children). (2) The **exploration** or **story** mode. The children interact multimodally with ECAs. This mode has a high degree of randomness. The child is exploring various conversational possibilities along a plot line. Although, the story-telling mode is mostly exploratory, the degree of exploration (novelty) wears off quickly once the children are acquainted with the ECA characters.

Another important factor here is the novelty of using the speech modality to interact with the ECAs.

The system's modular architecture is as follows: The *Ap-*

plication Manager is responsible for the synchronization and cooperation of the modules. It consists of two parts that follow the client/server architecture. 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 this study, because of the verbal variability children display, 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 allows the wizard to supply the appropriate transcription via a GUI interface.

The selected game tasks were: Animal/Shape/Number recognition, Quantity Comparison and Addition[1]. In our application we used the optimal settings of the Malone [10] factors for fantasy, curiosity and difficulty, as presented in [8] results. In the story-telling mode, the children have the opportunity to engage in conversation and participate in interactive story-telling with ECAs. The main story-telling ECA is a Horse/Unicorn, who presents his story and provides information about the life in his fantasy world. The companion ECA is a Panda that is present in both modes (games and story). Mouse and speech input modalities are supported throughout the interaction, although, the use of mouse in the story-telling mode is rather limited. The interaction for the story-telling mode is different for the first and follow-up playthroughs. The structure for the first playthrough is mostly predefined in order to acquaint the child with the game and characters. In any subsequent playthrough, the story-telling mode has a generic and random structure. The child has the choice of switching freely between modes at certain points in the interaction.

The story-telling is divided into five sections. Each section has a number of alternative dialogue themes that are chosen randomly; alternative plots are thematically correlated so that a certain coherent flow is maintained. The user does not remain passive during story-telling; at the end of each section a question is posed to the user related to the story line. In order to toggle amongst the two modes there are various break-point frames, where the children have to select between games or story (explore vs exploit decision). The final part of the story-telling mode are the *Random Quotes* (RQs). Each RQ is based on a conversational theme that the main character ECA presents to the user.

In terms of spoken dialogue implementation of the application, we have selected a mixed initiative strategy. During story-telling we have implemented turn-based dialogues inspired from the plot of the story. The initiative lies with the ECA that prompts the child to speak by posing various questions. When the children get a basic grasp of how things work they are also prompted at certain points to ask questions themselves (thus surrendering some initiative to the child). Questions posed by the ECAs should be answered within ten seconds by the child, else a time-out occurs. Following a timeout the question is rephrased and reposed by the ECA. Similarly in game playing mode, the initiative lies mostly with the companion ECA that is posing questions to the user. Again the user is given some degree of initiative to change the mode at any time.

[1] Addition task was only available for five and six year olds.

3. EXPERIMENTAL SETUP AND METRICS

Experiments were performed on location at two preschools in Chania, Greece. Fifteen native Greek children, ages four to six participated in this study. The children had no knowledge of the existence and functionality of a wizard (WoZ) and believed that they were interacting with a fully automated system. Each child participated in two sessions and could stop the session at any time. At the end of each session the children participated in an exit interview in the form of a questionnaire elicited by the experimenter.

The following objective metrics are extracted directly from the interaction log-files for each session: session duration, time spent in game and story mode, number of tasks attempted/completed in game mode, interaction time, (actual and animation) inactivity time, number/duration of user voice requests (in domain and out of domain/junk), number of mouse clicks (in domain and out of domain/junk), number of timeouts, number of voice barge-ins. Inactivity time is further divided into actual inactivity and animation inactivity. Actual inactivity time refers to the time that elapses between the end of an ECA's question/prompt and the start of the user's response. Animation inactivity time refers to the time that the user passively views the story unfold (and cannot be interrupted by the user).

4. EVALUATION RESULTS

Objective and subjective evaluation results are presented next for each age group. Results are presented separately for the first and second playthrough, in order to better understand how the EvE dilemma affects game play and user satisfaction, as a function of age.

4.1 Games vs Story-Telling Mode

In Table 1, we present the percent of time spent in story and game mode for each age-group for the first and second playthrough. Time spend transitioning between the two modes is also shown as "Idle & Transition" time. In the first playthrough, four and five year olds show a preference towards story-telling, while for the second playthrough this trend is reversed: five and six year olds spend significantly more time in story-telling than in game playing.

1st Playthrough	**4**	**5**	**6**
Story	59.80%	65.64%	48.93%
Games	29.91%	26.69%	37.36%
Idle & Transition	8.82%	7.75%	9.05%
2nd Playthrough	**4**	**5**	**6**
Story	48.41%	66.96%	62.22%
Games	35.34%	25.36%	29.05%
Idle & Transition	14.38%	7.68%	8.72%

Table 1: Mode duration distribution for the two playthroughs.

In Fig. 1, we present the transition probabilities for a two-state Markov model, where the two states correspond to story-telling 'S' and games 'G' mode. For these statistics we took both playthroughs into consideration. To compute the transition probability matrix we have taken into account all instances in a session where the child is explicitly asked to make a decision between games or story-telling mode. The total probability of transitioning between modes (G to S, S to G) vs staying with the same mode (G to G, S to S) is also given in Table 2 broken down for the first and second playthrough. Six year olds have the highest probability of staying with game mode (once in game mode) $p(G|G) = 0.73$. Also six year olds are more probable to stick to the current mode (G or S) rather than switch modes; this trend is consistent in both the first and second playthrough.

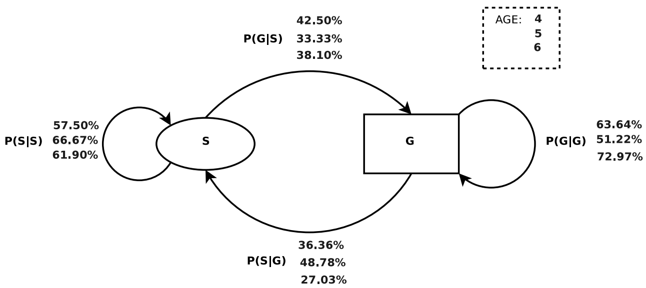

Figure 1: Transition probabilities between game 'G' and story-telling 'S' mode.

	1st Playthrough		2nd Playthrough	
Age	G-G/S-S	G-S/S-G	G-G/S-S	G-S/S-G
4	77.42%	22.58%	53.13%	46.86%
5	67.44%	32.56%	54.24%	45.76%
6	82.5%	17.5%	61.84%	38.16%

Table 2: Total probability of changing modes.

4.2 Objective Metrics

In Table 3 we present the percent of inactivity time for the story-telling mode (broken down in actual and animation inactivity), the average number of mouse clicks and voice requests (in and out of domain), as well as the average duration of voice requests per session. Inactivity time decreases as a function of age; actual inactivity is almost half for six year olds compared to four year olds. The out of domain mouse clicks and voice requests reduce significantly with age, especially for six year olds (compared to four and five year olds). In general, older children interact with the application more effortlessly and know better what to say and when to say it.

Age		4	5	6
Actual Inactivity (story)		10.6%	8.7%	5.6%
Animation Inactivity (story)		18.8%	11.3%	3.8%
Total Inactivity (story)		29.4%	20.0%	9.4%
Avg. # of Mouse Clicks	In Domain	15.3	10.8	10.8
	Out of Domain	17.7	12.5	2.6
Avg. # of Voice Requests	In Domain	57.8	56.8	41.8
	Barge In	2.3	0.5	0.8
	Out of Domain	12.0	11.5	8.5
Avg. Duration of Voice Requests (sec)	In Domain	80.2	50.5	46.0
	Out of Domain	38.8	34.8	7.6

Table 3: Interaction-related objective metrics.

4.3 Subjective Metrics

Our questionnaire included 22 questions that covered a range of topics related to the EvE dilemma (mode preference), the application proper (ECA voices, enjoyment), the interaction modalities and the experimental setup.

For the first playthrough, four and five year olds showed a preference towards story-telling, while for the second playthrough they seemed to equally like the two modes. Six year olds seem to equally like the two modes in both the first and second playthrough.

In terms of engagement, about half of the four and five year olds replied that they were tired or bored at some point in the interaction. Most children replied that they got tired 'towards the end'. None of the six year olds was tired or bored. In general, six year olds where the most enthusiastic users of the application and the most tolerant ones also. Engagement was somewhat higher for the first vs the second playthrough.

5. DISCUSSION

The interaction patterns and mode duration statistics for the first playthrough, as well as the subjective mode preference, show that younger children prefer the story-telling mode, tipping the EvE dilemma more towards exploration. The reversal of mode preferences for the second playthrough are most probably due to the fact that children are now familiar with the story line and the degree of exploration now is significantly lower than in the first playthrough. It is clear from the interaction patterns that although younger children prefer exploration to exploitation somewhat more than older children, older children have much more efficient game playing strategies (including more efficient exploration patterns). A typical strategy in the game mode for six year olds in the first playthrough is to thoroughly explore all games/tasks one by one, and the in the second playthrough quickly revisit the games and select their favourites.

A final important point has to do with the EvE-related iteration patterns for each child between the first and the second playthrough. It is interesting to note that the mode duration and usage patterns for each child are relatively consistent between the first and second playthrough. In fact, as far as the EvE dilemma is concerned, differences between individual children seem to be larger than differences between age groups.

5.1 Effect of speech recognition errors

Speech recognition technology remains fragile for the preschool population; error rates of up to 30% are not uncommon even for a small vocabulary [15]. In order to investigate the effect of speech recognition errors on the user interaction patterns, we ran a pilot experiment with two preschool children where speech recognition errors were simulated by randomly inserting misrecognitions at a 20% error rate in the WoZ scenario.

The typical behavior observed was that one or two misrecognitions in the same game were sufficient to lead children to abandon the game and continue to the next one. This behavior was displayed by children for all tasks with the exception of the animal recognition game, where mouse input was provided as an alternative modality. For the animal recognition game, children switched from speech to mouse input after a misrecognition and completed the game using the mouse. Once children realized that speech recognition errors persisted in game mode, they displayed a preference towards story-telling mode (especially in the second playthrough). This was verified also at the exit interview where children indicated that story mode was more pleasant compared to game mode, because of the speech recognition errors. Overall the results show that children perceived recognition errors as an internal problem of the game mode (rather than the interface), biasing modality usage and the EvE balance. The results are consistent with the literature [14, 4].

6. CONCLUSIONS

In this study, we took a first step towards analyzing the EvE dilemma as it pertains to multimodal dialogue game design for preschoolers. We analyzed the interaction patterns of fifteen children ages four to six interacting with a computer game with two modes emulating exploratory strategies and exploitation strategies. As expected, younger children showed a higher preference to exploration strategies, although, due to the small sample size these results are not statistically significant. It was also shown that in terms of game playing, conversational ability and exploratory strategies six year olds were significantly better than four and five year olds. An important finding is that the EvE dilemma differences were larger among children than among age groups. Overall, this study shows that designing an application with an appropriate mix of exploration and exploitation can be very engaging to preschoolers. This study represents only the first step towards understanding the main factors of the EvE dilemma for preschooler computer games. The challenge remains on how to design such an application that is well suited to the skills of the target preschool population, while maintaining a high level of engagement.

7. REFERENCES

[1] Bura, S. Explicit intent: Shared story ownership between player and game. Proc. AISB (2006).

[2] Cassel, J., and Ryokai, K. Making space for voice: Technologies to support children's fantasy and storytelling. *Personal Technologies 5*, 3 (2001), 203–224.

[3] Cohen, J. D., McClure, S. M., and Yu, A. J. Should I stay or should I go? How the human brain manages the trade-off between exploitation and exploration. *Philosophical Transactions of the Royal Society Biological Sciences 362*, 1481 (2007), 933–942.

[4] Cohen, P., Johnston, M., McGee, D., Oviatt, S., Clow, J., and Smith, I. The efficiency of multimodal interaction: a case study. Proc. of Fifth International Conference on Spoken Language Processing (1998).

[5] Gittins, J. C., and Jones, D. M. A dynamic allocation index for the sequential design of experiments. *Progress in Statistics. Eds.J. Gani et al.* (1974), 241–266.

[6] Gustafson, J., Bell, L., Boye, J., and Wiren, M. The nice fairy-tale game system. Proc. SIGdial (2004).

[7] Hagen, A., Pellom, B., and Cole, R. Children's speech recognition with application to interactive book and tutors. Proc. ASRU Workshop (2003).

[8] Kannetis, T., and Potamianos, A. Towards adapting fantasy, curiosity and challenge in multimodal dialogue systems for preschoolers. Proc. ICMI (2009).

[9] Lee, S., Potamianos, A., and Narayanan, S. Acoustics of children's speech: Developmental changes of temporal and spectral parameters. *Journal of Acoustical Society of America 105*, 3 (1999), 1455–1468.

[10] Malone, T. W. What makes things fun to learn? Proc. Conference on Video Games and Human Development: A Research Agenda for the 80's (1983).

[11] March, J. G. Exploration and exploitation in organizational learning. *Journal of JSTOR 2*, 1 (1991), 71–87.

[12] Mostow, J., Hauptmann, A. G., and Roth, S. F. Demonstration of a reading coach that listens. Proc. ACM Symposium on User Interface Software and Technology (1995).

[13] Narayanan, S., and Potamianos, A. Multimodal systems for children. Proc. European Conf. on Speech Communication and Technology (1999).

[14] Potamianos, A., and Narayanan, S. Spoken dialog systems for children. Proc. International Conf. on Acoustics, Speech and Signal Process (1998).

[15] Russell, M., Brown, B., Skilling, A., Series, R., Wallace, J., Bonham, B., and Barker, B. Applications of automatic speech recognition to speech and language development in young children. Proc. ICSLP (1996).

[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 41* (1993), 5–16.

[17] Sutton, R. S., and Barto, A. G. *Reinforcement Learning*. MIT Press, 1998.

[18] Weib, S. A., and Muller, W. The potential of interactive digital storytelling for the creation of educational computer games. *Lecture Notes in Computer Science (LNCS) 5093* (2008), 475–486.