Affective Evaluation of Multimodal Dialogue Games for Preschoolers Using Physiological Signals

Vassiliki Kouloumenta, Manolis Perakakis, Alexandros Potamianos

Dept. of ECE, Technical Univ. of Crete, Chania 73100, Greece
{vaskou,perak,potam}@telecom.tuc.gr

Abstract

In this pilot study, we investigate the differences in the electroencephalography (EEG) signal patterns of children and adults while interacting with a multimodal dialogue computer game. The gaming application is designed for preschoolers, implements five popular learning tasks and has variable levels of difficulty. In this pilot, to simplify the data collection process for young children, we use the NeuroSky MindSet device which is a single forehead dry sensor device. The raw signals and the estimated attention, meditation and arousal signals are analyzed during the interaction and compared for adult and children user populations. Results show consistent variations as a function of modality used (speech vs mouse input), difficulty level and task success. The physiological signal pattern within an interaction turn is also estimated and analyzed. Overall, children and adults demonstrated very similar physiological signal patterns during multimodal interaction.

Index Terms: child-computer interaction, multimodal dialogue systems, electroencephalography, affective analysis.

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).

In human communication affect and emotion play an important role, as they enrich the communication channel between the interacting parties. Recently there has been much research interest in the CHI community aiming at incorporating affective and emotional cues in the human computer interaction loop. These efforts are known collectively as affective computing [11]. Multimodal spoken dialogue systems are traditionally evaluated with objective metrics such as interaction efficiency (turn duration, task completion, time to completion), error rate, modality selection and multimodal synergy [1, 3]. In this work we extend the work in [2] and investigate the use of Electroencephalography (EEG) for studying multimodal interaction in preschoolers. EEG is a rich source of information which is able to reveal hints of both affective and cognitive state during an interaction task. In this work, we investigate the use of EEG elicited affective metrics such as attention and arousal for the evaluation of interactive systems and multimodal dialogue systems in particular. This, not only provides a more qualitative approach to evaluation, it also provides a better understanding of the interaction process from the user perspective. Extracting robust information from such physiological channels is a challenging but also potentially rewarding task, opening new avenues for the emotional and cognitive assessment of multimodal interaction design. To our knowledge, this is the first effort for the affective evaluation of multimodal spoken dialogue systems for children using brain signals.

The remainder of this paper is organized as follows. First we briefly review the state-of-the-art and the multimodal application evaluated in Sections 2 and 3 respectively. Then the EEG device and derived physiological measurements are discussed in Section 4. The experimental procedure and user populations are presented in Section 5. The main results from our experiments on children and adult user populations are shown in Section 6. We conclude with an analysis of the affective evaluation results and propose future work in Section 7.

2. Prior Work

Recently there is a growing interest in the design and development of multimodal dialogue systems for children. As in early ages the learning procedure takes place through gaming it is important to develop educational computer games. Multimodal interactive systems have become increasingly popular, mainly due to the way children express themselves. To be more specific children tend to use more than one modalities in communication with others, (e.g. voice, gestures etc.) [6], [7]. As a result, different input modalities may help in order to increase efficiency in child-computer interaction.

Although children are good adopters of technology, there are several challenges in the design and the implementation of multimodal dialogue systems for children. Automatic speech recognition in such ages is not an easy task because of the linguistic variability children display when talking. So the existence of various modalities should help to overcome such recognition problems.

In recent years lot of research has been done in the design of multimodal dialogue for children. A variety of prototype systems, containing spoken dialogue interfaces and capabilities, have been implemented. In [8], [5] authors describe their efforts in designing and building a prototype multimodal system for children users. The CHildren's Interactive Multimedia Project (agent CHIMP) provides design guidelines for

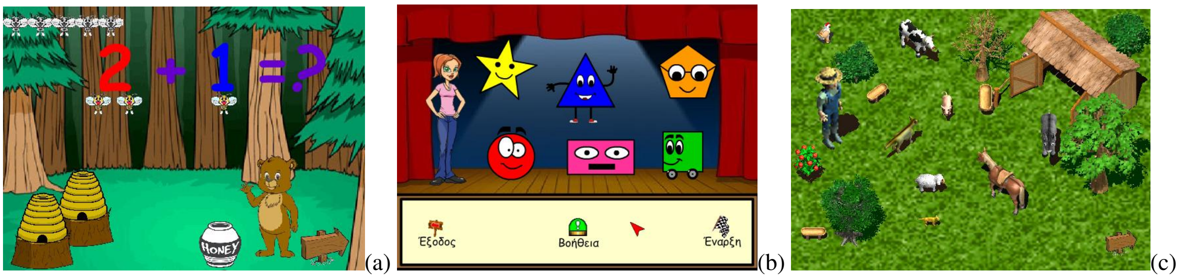

Figure 1: Example screen-shots of the five tasks: (a) addition, (b) shape recognition, and (c) animal recognition.

building successful multimodal-input, multimedia-output applications for children users. An important feature of the CHIMP system involves the integration of multiple input and output modalities such as voice, audio, keyboard, mouse, graphics and animation. In [13] researchers make a comprehensive analysis of children's multimodal integration patterns during interaction with an educational software prototype. Moreover, the NICE Fairy-Tale game system, which is described in [15], allows users to interact with various animated characters in a 3D world.

At ages 4-6, learning and playing are intertwined activities. Thus the main goal of a successful game for preschooler is to provide fun, excitement and engagement. Several theoretical studies have attempted to identify what is "fun" in a game. According to Malone [4] the essential characteristics of a good computer game can be organized into three categories: fantasy, curiosity and challenge. Alternatively, Lazzaro [9] 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 [10], i.e., strong involvement in a task occurs when the skill of an individual meets the challenge of the task. Finally, in the field of entertainment capture, Yannakakis [12] showed that the player-opponent interaction is a major factor in entertainment.

EEG studies more relevant to the domain of HCI and affective computing are those studying emotions and fundamental cognitive processes related to attention and memory. EEG emotion recognition has been an active topic in the last years. There are various representations of emotions such as the wheel of emotions by Plutchik [14] and the mapping of various emotions in the arousal-valence space, one of the most used frameworks in the study of emotions. Arousal is the degree of awakeness and reactivity to stimuli and valence is the positiveness degree of a feeling. According to previous studies, indicative metrics of arousal is the beta/alpha band power ratio in the frontal lobe area. For valence the alpha ratio of frontal electrodes (F3, F4) has been used, as according to [16] there is hemisphere asymmetry in emotions regarding valence e.g. positive emotions are experienced in the left frontal area while negative emotions on the right frontal area. User state estimation based on cognitive attributes related to attention and memory (in addition to emotions) is also of great importance in the context of HCI research. Memory load, an index of cognitive load [17] is an important index of mental effort while carrying out a task. Memory load classification has thus drawn attention from the HCI research community since it can reveal qualitative parameters of an interface. In [18] authors report a classification accuracy of 99% for two and 88% for four different levels of memory load, by exploiting data from the *N-back* experiment [19] for their classifier. They also argue that previous research findings that high memory loads correlate with increase in theta and low-beta(12-15 Hz) bands power in the frontal lobe or the ratio of beta/(alpha+theta) powers may not hold always true.

3. Application Design

The application evaluated is the one described in [20]. It consists of five educational gaming tasks based on popular preschool activities, specifically: (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). The user is guided through these tasks by an animated character based on the research experience in [5, 15], the animated character was (in most cases) assigned the social role of a friend/helper. Sample screen shots of three of the five games are shown in Fig. 1.

The interface was designed to be appealing to the target age group, using a different mix of sounds, good animation and graphics in order to keep the child engaged. Given that younger preschoolers are not always well versed with the use of the mouse, the games were designed with speech as the main interaction modality for input, and graphics and speech the main modalities for output. In addition to speech input, users are able to play the games using mouse input. Given that the target audience (ages 4-6) is at a preliterate level, the use of text as an output modality was avoided. Text output was substituted or complemented with sounds, graphics and animation.

Motivated by the work of Malone [4], the application was implemented with variable levels of fantasy, curiosity and challenge. In the set of experiments that follows only the challenge level was varied so we briefly describe this implementation next: We have implemented three different levels of difficulty for each of the five tasks. For example, for the number recognition game, 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. For details see [20].

A Wizard of Oz (WoZ) setup was used throughout this study, so the effect of speech recognition errors was not studied. For an affective analysis of speech recognition errors using physiological signals on adult users see [2].

4. Physiological Measurements

The usual process to obtain brainwave measurements is a medical procedure supervised and performed by trained personnel, involving several electrodes positioned around the head and attached with conductive gel. This may cause discomfort for the subject and it was deemed an unrealistic data collection method for preschoolers. Instead we used the NeuroSky MindSet [21], an easy to use single electrode EEG device. The Mind-

Set monitors electrical potential between the sensing electrode, positioned on the forehead, and the reference electrodes, positioned on the left earlobe. The single point electrode means that changes in brainwave activity in different parts of the brain cannot be monitored. The rationale for using the MindSet device is that it is easy to setup and convenient for children to wear, compared to clinical EEG systems. Despite being a single electrode EEG device, it is considered adequate for this pilot study because it is known that the frontal cortex is a rich source of both cognitive as well as emotional activity in the brain.

The NeuroSky device provides access to the raw brainwave (EEG) signal values broken down by frequency range, namely alpha (8-12 Hz), beta (12-30 Hz), gamma low (30-40 Hz) and gamma high frequencies (41-100 Hz), as well as delta (0.1-3 Hz) and theta (4-7 Hz), sampled at 512 Hz. In addition the device provides access to "eSense" values that are computed using a proprietary algorithm from the raw signals after removing ambient noise and artifacts. The eSense algorithm provides two output signals sampled at 1Hz and labelled as "attention" and "meditation". These two states of mind are described as:

- **Attention** indicates the intensity of a user's level of mental "focus" as it occurs during intense concentration and directed (but stable) mental activity. Distractions, wandering thoughts, lack of focus, or anxiety may lower the attention meter levels.
- **Meditation** indicates the level of a user's mental "calmness" or "relaxation". Note that meditation is a measure of a person's mental levels. However, for most people in most normal circumstances, relaxing the body often helps the mind to relax as well.

The eSense values range between 1 and 100. On this scale, a value between 40 to 60 at any given moment in time is considered "neutral". A value between 60 to 80 is considered "slightly elevated" and a value between 80 to 100 is considered "elevated", meaning they are strongly indicative of heightened levels of that eSense. Similarly, on the other end of the scale, a value between 20 to 40 indicates "reduced" levels of the eSense, while a value between 1 to 20 indicates "strongly lowered" levels of the eSense. These levels may indicate states of distraction, agitation, or abnormality, according to the opposite of each eSense.

In addition to attention and meditation we also report on an arousal metric estimated directly from the raw α and β waves. **Arousal** is a physiological and psychological state of being awake or reactive to stimuli, and often is referred to be associated with β and α waves [22]. It is known that β brainwaves indicate an alerted state of mind, while α brainwaves indicate a more relaxed state. So arousal is often estimated by the β/α ratio. So $\beta/\alpha > 1$, indicates an active state, while $\beta/\alpha < 1$, indicates a passive state.

5. Experimental Procedure

The experimental procedure took place in a sound attenuated office room, where the subjects sat on a comfortable chair. During the whole process subjects should wear the NeuroSky MindSet, so that we could collect EEG and eSense data during interaction. Before the experiment begins there was a test "period", where the subject should find the most comfortable position in front of the laptop's screen and we tried to best fit the NeuroSky device on the subject's head[1]. The procedure was a slightly different for the two user groups: children and adults.

		Attention			
Modality	Statistic	WP	P	WA	A
Mouse	MEAN	59.3	44.4	45.6	43.8
Voice	MEAN	48.1	49.2	39.7	43.3
		Meditation			
Modality	Statistic	WP	P	WA	A
Mouse	MEAN	45.2	52.2	47.4	54.2
Voice	MEAN	52.3	48.3	48.8	49.5
		Arousal			
Modality	Statistic	WP	P	WA	A
Mouse	MEAN	0.88	0.38	0.93	0.75
Voice	MEAN	0.92	0.47	0.82	0.99

Table 1: Average attention, meditation, arousal of children users broken down by interaction turn segment for mouse/voice.

For children, each user was asked to play the game twice in a single session. During that session we modified only the challenge value from 1 to 2, while the other two factors (fantasy and curiosity) remained constant at level 1. Also, in a separate session, each child was asked to play the game (with all three factors at level 1) twice: the first time by using only voice and the second time by using only the mouse input device. Thus in total each child played the game four times in two sessions. For adults, each user was asked to play only one application setup in a single session, where all the three factors are taken the value 1. Each adult played the game three times: using speech input only, mouse input only or multimodal (speech and mouse) input.

The adult user populations for this pilot study consisted of 14 individuals, 5 females and 9 males ages 20-59. The children user populations consisted of 6 preschool children (5 females and 1 male) aged 5.5 to 6.5 years. A total of 56 interaction session were collected for adults and 24 interactions for children (total of 80).

6. Results

We present mean attention, meditation and arousal values, averaged over all user turns for different population groups (children vs adults), modalities (mouse vs speech), and gaming tasks. In addition, we break down the interaction turn into four segments and report average EEG metric values for each segment, namely: 1) While user waits for the system's prompt: **WP**, 2) The system prompt: **P**, 3) While user is thinking about the answer (user is inactive): **WA**, and 4) When the user speaks or clicks the answer (user is active): **A**.

6.1. Interaction Turn and Modality Patterns

In Fig. 2(a),(b), we show the average attention values broken down by interaction turn segment for mouse vs. voice interaction, for the adult and children user populations. Note that the average interaction turn patterns are consistent for mouse vs voice for both user populations with the exception of the A state where the user is providing input via mouse (this difference could be due to the mouse manipulation skills of six-year olds). Overall, mouse input displays higher attention values than speech, whereas in P state (voice prompt) the opposite is true. This is expected due to modality symmetry, i.e., user pays less attention to voice prompts when using graphical input.

The average values for attention, mediation and arousal per

[1] Acquiring high-quality signal for younger children was a challenge. For this reason we only collected data from five- and six-year olds (the modified NeuroSky device was a better fit for them).

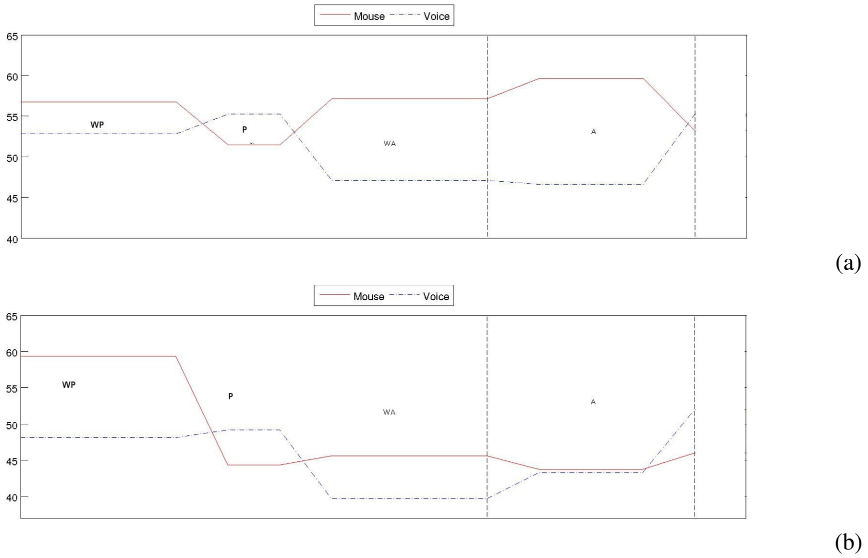

Figure 2: Average attention values per interaction turn broken down into: waiting for prompt (WP), prompt (P), waiting for answer (WA) and answer (A) when using voice (dashed line) or mouse (solid line) input modalities for: (a) adult users and (b) children user populations. Attention values are averaged over all interaction turns, games played and users. WE refers to waiting to exit (at the end of each game). Meditation (now shown here) follows the inverse pattern over an interaction turn.

	cor. coef. (p-value)
Activity time / Arousal	-0.27 (0.001)
Correct-Wrong Ans. / Attention (WA)	0.15 (0.06)
Arousal / Attention	0.28 (0.001)

Table 2: Correlations between EEG and objective metrics.

	Adults		
	Attention	Meditation	Arousal
NUMBERS	43.7	53.2	1.07
MORE-LESS	42.7	56.7	1.22
SHAPES	45.0	56.4	0.87
ADDITION	44.7	59.8	0.97
FARM	49.1	57.0	1.00
	Children		
	Attention	Meditation	Arousal
NUMBERS	46.2	49.9	0.80
MORE-LESS	45.1	53.9	0.78
SHAPES	48.6	53.1	0.85
ADDITION	42.7	51.1	0.89
FARM	47.4	55.0	0.89

Table 3: Mean EEG metrics per game and population.

modality and interaction turn segment are show in Table 1 for children users. Observe that the meditation patterns relative to modality and interaction segment are reversed compared to attention. Arousal follows a pattern similar to attention, for the P state (low values especially for mouse input) and WA state, however, for the A state arousal for speech input is much higher than for mouse input.

6.2. Correlation with Objective Metrics and Difficulty

In Table 2, we show the correlation between selected objective and EEG metrics for children users, as well as correlation among EEG metrics. Activity time (time required to provide input) is negatively correlated with arousal. Attention is also mildly correlated with providing correct answers. As expected arousal and attention are also correlated. We have observed no significant differences between difficulty level 1 and 2 in the average EEG metrics (one-way ANOVA indicated $0.08 < p < 0.8$) probably due to the simplicity of the tasks for six-year olds (not shown here).

6.3. Per Gaming Task

In Table 3, we show average attention, meditation and arousal values per gaming interaction for children and adult user populations. Overall, the patterns are similar between children and adults. There are differences between gaming tasks that are consistent for both adult and children populations, however, these differences are not readily explainable in terms of the interaction type. In general, visual tasks (FARM, SHAPES) have higher attention/meditation values than computational tasks (ADDITION, NUMBERS), probably due to the location of the electrode. Note that the attention values are higher for children than for adults, while the meditation values are higher for adults throughout, possibly indicating lower cognitive load for these simple games for adults.

7. Conclusions

Overall, we have observed very similar patterns for children and adults. Despite the small number of users and sessions we have observed consistent patterns in mouse and speech usage, e.g., higher attention values when waiting for prompt and listening to prompt, higher meditation values when the user is producing an answer. As expected, we have also observed negative correlation between activity time and arousal. This pilot study shows that it is feasible to use simple EEG devices to collect physiological and affective signals from young children. More research is needed on how such signals can be used during game development, game evaluation and gameplay.

8. References

[1] Perakakis, M. and Potamianos, A. “A study in efficiency and modality usage in multimodal form filling systems” Audio, Speech, and Language Processing, IEEE Transactions on, vol. 16, no. 6, pp. 1194-1206, 2008.

[2] Perakakis M. and Potamianos A., “Affective evaluation of multimodal dialogue interaction using physiological signals”, SLT, pp. 43-48, IEEE, 2012.

[3] Perakakis, M. and Potamianos, A. “Multimodal system evaluation using modality efficiency and synergy metrics” Proceedings of the 10th international conference on Multimodal interfaces, ICMI, pp. 9-16, 2008

[4] Malone, T. W, “What make things fun to learn? A study of intrinsically motivating computer games,” In *Proceedings of the 3rd ACM SIGSMALL Symposium and the First SIGPC Symposium on Small Systems*, Palo Alto, California, United States, September, 1980.

[5] Potamianos, A. and Narayanan, S., “Creating conversational interfaces for children” *IEEE Transactions on Speech and Audio Processing*, vol. 10, pp. 65-78, February, 2002

[6] Druin, A., Bederson, B., Boltman, A., Miura, A., Knotts-Callahan, D. and Platt, M., “Children as Our Technology Design Partners”, The Design of Children’s Technology: How We Design, What We Design and Why, Druin, A., Kaufmann, M., 1998.

[7] Druin, A. and Inkpen, K., “When are Personal Technologies for Children?”, Personal and Ubiquitous Computing, vol. 5, pp. 191–194, 2001.

[8] Narayanan, S., Potamianos, A. and Wang, H., “Multimodal Systems For Children: Building a Prototype”, Proc. of EuroSpeech, 1999.

[9] Lazzaro, N., “Why We Play Games: Four Keys to More Emotion Without Story”, *Technical Report*, XEO Design Inc., [Online], Available at: http://www.xeodesign.com, 2004

[10] Csikszentmihalyi, M., “Flow: The Psychology of Optimal Experience”, New York: Harper & Row, 1990.

[11] Picard, R. W., “Affective Computing”, MIT Press, pp. 292, 1997.

[12] 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.

[13] Benfang, X., Girand, C., and Oviatt, S., “Multimodal integration patterns in children”, In *Proc. ICSLP-2002*, pp. 629-632, 2002.

[14] Plutchik, R., “The Nature of Emotions”, American Scientist, vol. 89, no. 4, pp. 344-350, 2001.

[15] Gustafson, J., Bell, L., Boye, J., Lindstrom, A. and Wiren, M., “The NICE Fairy-tale Game System”, *Proceedings of SIGdial 04*, Boston, April, 2004.

[16] Davidson, R. J., “What does the prefrontal cortex “do” in affect: perspectives on frontal EEG asymmetry research”, Biological Psychology, vol. 67, no. 1-2, pp. 219-233, 2004.

[17] Paas, F., Tuovinen, J. E., Tabbers, H. and Van Gerven, P. W. M., “Cognitive Load Measurement as a Means to Advance Cognitive Load Theory”, Educational Psychologist, vol. 38, no. 1, pp. 63-71, Lawrence Erlbaum, 2003.

[18] Grimes, D. ,Tan, D. S., Hudson, S. E., Shenoy, P. and Rao, R. P. N., “Feasibility and pragmatics of classifying working memory load with an electroencephalograph”, Proceeding of the twenty-sixth annual SIGCHI conference on Human factors in computing systems, CHI ’08, Florence, Italy, pp. 835-844, ACM, 2008.

[19] Gevins, A. and Smith, M. E. “Neurophysiological measures of cognitive workload during human-computer interaction”, Theoretical Issues in Ergonomics Science, vol. 4, no. 1, pp. 113-131, 2003.

[20] Kannetis, T. and Potamianos, A., “Towards adapting fantasy,curiosity and challenge in multimodal dialogue systems for preschoolers”, Proc. International Conf. on Multimodal Interaction (ICMI), 2009.

[21] “NeuroSky: Brain Wave Sensors for Every Body”, http://www.neurosky.com

[22] Bos, D. O. “EEG-based Emotion Recognition”, [Online], Available at: http://emi.uwi.utwente.nl/verslagen/capita-selecta/CS-oude Bos-Danny.pdf, 2006.