A Narrative Speech, Gaze and Gesturing Robot Accessing to Human Emotion and Memory Which is Investigated by Using a Simultaneous Recording of Electroencephalogram and Eye-Tracker System

Aravind Muthusamy¹, Gagandeep Kaur¹, Rohit Gupta¹, Vihang Agarwal¹, Bishakh Bhattacharya¹, Sayaka Nakayama², Mayu Ichiki², Jisha Maniamma² and Hiroaki Wagatsuma^{2, 3, 4}

> ¹ Department of Mechanical Engineering Indian Institute of Technology, Kanpur Uttar Pradesh, India -208016 <u>bishakh@iitk.ac.in</u>

 ² Graduate School of Life Science and Systems Engineering Kyushu Institute of Technology
2-4 Hibikino, Wakamatsu-Ku, Kitakyushu 808-0196, Japan

³ RIKEN Brain Science Institute 2-1 Hirosawa, Wako-shi, Saitama 351-0198, Japan

⁴Artifical Intelligence Research Center, AIST 2-3-26 Aomi, Koto-Ku, Tokyo 135-0064, Japan waga@brain.kyutech.ac.jp

ABSTRACT. The effect of robot behavior in a passive interaction scenario, where the human does not make any tangible reaction to robot behavior is being studied here. A desktop biped robot is used to induce mental situational comprehension in the subject and the brain dynamics is analyzed during this interaction. The Robot does a recitation of a math table and makes synchronized motion during narration. The effect of the aural and visual behavior on the subject is studied during each phase using EEG and Eye tracker information acquired during this exercise. The changes in attention and its memory maintenance is discussed. The effect of anthropomorphism on episodic memory is also explored.

Keywords: Episodic Memory, Human Robot Interaction, Mediation, Social Robot, EEG

1. **Introduction.** Technical advancements have bolstered the dreams of researchers and designers to bring robots from the shop floor and remote locations like Martian terrain to the domestic household. These possibilities have given rise to many critical questions. The reason is that, the robot as a machine that is designed to execute a predetermined job with precision is being thought of as a social being [1]. The primordial factor that dictates this endeavor is the behavior of the social robot. The behavior is again governed by the destined role it is designed to play. Various avenues of introduction of these social robots are being explored to define a possible role for the robots in social space. Robots as personal

assistants, as peers, pets and toys for children, as surveillance agents in the household, as companions for old people are some of the possible roles that are being designed. In a broad sense, a social robot is defined as a robot designed to work in the social space of humans, capable of interacting with humans in a natural way [1]. Being a social species, humans invariably employ social models of evaluation to predict, understand and explain the behavior of non-living things of sufficient complexity. If, form was to follow function, anthropomorphism plays a vital role in augmenting the designed behavior of the social robot. The robots also exhibit certain traits of autonomy which facilitates the assimilation of the robot in the social space. Based on these factors, the social robots can be classified into the following classes [2].

As socially evocative, the robots have the singular task of evoking tendencies to anthropomorphize in humans. They are not intended to deliver any further goals. A robotic toy is a typical example of this class of robots. As social interfaces, a robot can be an interface for a human in a social setting. They would need to have capabilities of gesticulating and speech. Socially receptive robots are those who can be trained to understand social cues like gaze based direction, gesture based commands. This would require cognitive models of a particular modality to be active in the robot that can effect deep changes in the structure of behavior of the robot. Learning a gesture effecting deep changes in the motor actions of the robot is an example. These robots are only receptive and do not exhibit any autonomy of their own. Social robots, on the other hand participate in social interactions and exhibit internal goals and motivations. They not only interact with people to enhance and enrich the interaction, the robots also learn from these interactions to model their behavior.

In all the cases, the material exchange that happens is the content of interaction and the robot facilitates this interaction between people or between people and itself through mediation. The mediation aspect of the robot is the core in social robotic studies [3]. After each of these interactions, the participants of the interaction, whether it be the human or the robot, change. And this history plays a role in influencing further interactions. There is memory, emotion and learning involved in these interactions [4].

The interactions that humans have with their environment get stored as episodic memory or autobiographical memory [5]. These experiences have an important role in forming semantics of the world around us [6]. This paper makes an attempt to enquire human robot interactions through the lens of episodes of interaction and the possible changes and markers that can be noticed in the human participant as a result of these interactions.

1.1 Frontal Midline Theta. Electroencephalogram (EEG) technique is a long held effective non-invasive tool to understand brain dynamics in a temporal scale. Of the various EEG signals, Frontal midline theta (Fm Θ) rhythm (4-8 Hz) had been the cynosure of a majority of researchers, studying memory encoding and maintenance. Even though, Theta rhythm is attributed to internal allocation of resources for attention [7], theta band has been associated with episodic memory maintenance as well in the same attentive span [8,9]. Fm Θ is also seen as a representative of working memory operations. The limbic system network which is responsible for emotion, behavior, motivation and long term memory is supposedly to be the source of theta rhythms. The hippocampus plays a major role in the limbic system and is seen to be responsible for episodic memory operations. The frontal midline theta is mostly associated with the meaning of a stimulus and episodic memory related operations of encoding and retrieval. Parietal Theta rhythms are not related to habituation with stimulus and are still being researched [10].

1.2 Experiment rationale. Fm Θ is seen to be prominent in the Fz electrode site when subjects are administered a mental arithmetic task [11]. This is an indication of memory operations being employed for the task in hand. As had been discussed before, social robots act as mediators in interaction resulting in fleeting as well as permanent changes in the participants. The experiment attempts to understand the changes that are effected in the subject through their interaction with the robot. The aim is to induce an intrinsic mental arithmetic thought in the subject through passive interaction with the robot. By passive, we mean that the subject does not react to the robot's actions. A small sized desktop bipedal robot (https://plen.jp/about/) is employed for this experiment. The experiment is designed such that there are inter-stimulus delays during which intrinsic mental arithmetic thoughts can be invoked by the subjects. The robot is employed just to narrate a math multiplication table of a particular number that the subject is unfamiliar with. The narration is up to 10 times that number. The robot narrates the number in standing position facing the subject. For the last five numbers the Robot, is made to exhibit a forward and back swaying motion in synch with the narration frequency. During the first half of the narrating phase the robot was just emitting aural stimulus, for the latter half, it exhibits visual action (stimulus) along with aural inputs. The narration session is preceded and followed by a period of rest in which the subject just sits idle in repose. The EEG signals acquired during this phase helps in acting as baseline for the experiment.

2. Details of test setup. An 8 Electrode EEG signal acquisition system is used to acquire the EEG signal. A microcomputer board (Raspberry pi) is used to play the table narration. The same board sends triggers to the EEG signal acquisition system to time-stamp the EEG signals when the narration is played. A trigger is sent before every utterance of a multiplication statement. A trigger is sent before and after the rest states as well. A small desktop biped robot is used to play the role of the narrator. A mobile eye-tracker, ETG from SMI (https://www.smivision.com/eye-tracking/product/eye-tracking-glasses) is used to acquire pupillometry data and gaze pattern to analyze the visual engagement that the subject has with the robot.(Figure 1).



FIGURE 1: Test Setup

2.1 EEG Equipment. An unipolar, active, 8 Electrode wireless EEG signal acquisition

system (gTech's Mobilab Plus) is used. The AgCL, wet electrodes are placed at Fz,Cz,Pz,P1,P2,Oz,O1 and O2 positions mounted on a 10-20 standard cap. The ground is at nasion the reference is left mastoid. The sampling rate is set at 256 Hz.

2.2 Stimulus set up. A desktop biped robot, PLEN2 (https://plen.jp/about/), by PLEN Project Company, is used. PLEN2 is a wireless controllable robot. The robot has 18 freely moveable servomotor powered joints which can be programmed to achieve a desired motion. PLEN2 is programmed to oscillate its upper body forward and backward utterance of math tables. rhythmically with the the Raspberry Pi 3 (https://www.raspberrypi.org/products/raspberry-pi-3-model-b/) has been used for programming. The Trigger signals are sent to the EEG system through a microcontroller (Figure 2).



FIGURE 2: Trigger setup

2.2.1 Audio stimulus and Trigger setup. Bash scripts are used to reciting math tables. It generates the utterance of the math table in a female voice at the rate of 150 word/minute. The utterances are stored as audio files in '.wav' format. (Note: The utterance of the table is divided into 10 lines and each line's utterance is stored in different file.) Another script plays these files along with some other announcements (which are also in the same voice and same word rate).

2.3 Stimulus Design:



FIGURE 3: to and fro motion of the robot

The stimulus is designed in such a way the subject would be induced to do math arithmetic in anticipation. The desktop robot is placed in standing position facing the subject. The test starts with a resting state period of 30,000 ms where in the subject is in idle repose. Resting state EEG potentials are acquired during this state. A math table of a particular number that

the subject is unfamiliar with is narrated. The words are narrated at the rate of 25 Hz (150 words / minute). Each narration takes about 900 ms. This is followed by an inter stimulus delay of 4000 ms. The subject, even though, not instructed to calculate, is bound to calculate the next multiplication number in anticipation in the 4000 ms period. The robot starts swinging when the table reaches number 5. The to and fro motion of the robot is matched with the narration of the table (Figure 3). Once the number ten is reached, the narration and the swinging stops, followed by a period of 30 second rest (Figure 4).

3. EEG signal analysis. On raw EEG, a wavelet packet decomposition using Daubechies wavelet of order 8 was applied to extract theta (4- 8 Hz) and alpha (8- 12 Hz) signals. To witness Event Related Potentials (ERP), trials expected to contain ERPs were ensemble averaged. The Power spectral density was computed to get absolute power of Theta and Alpha rhythms at Fz and Pz during resting, Narration and movement phases of stimulus.

3.1 Observed Theta characteristics. At Fz, the theta power during narration increased as compared to rest, while during synchronized movement along with narration a dip in theta power was observed (Figure 5). At Pz, the difference in theta power across the various phases was not remarkable. The average theta power between Fz and Pz was relatively comparable for the different phases of stimuli. In addition to Fm Θ , frontal alpha has also been established as a marker for task based attention. At Fz, alpha rhythm during narration is noticed to have the highest power as compared to rest and narration synched with movement. However, at Pz, alpha rhythm at rest is highest, followed by narration and narration synched with movement in that order.





FIGURE 5: Power spectral density of Alpha and Theta at Fz and Pz respectively.



ERP for trials with Narration+Movement and underlying Alpha activity at Pz







FIGURE 6: ERP activity at Pz while attending to narration and narration synchronized with movement. In the alpha plots, the activity can be seen to acquire higher amplitude once the ERP transience ended as seen in Fz. In the theta plots, it can be witnessed that the theta gained higher amplitudes during ERP transience as in Fz. The overall ERP is lesser prominent at Pz as compared to Fz. IT is more significant during the narration phase as compared to the narration + movement phase.



FIGURE 7: ERP activity at Fz while attending to narration and narration synchronized with movement. In the alpha plots, the activity can be seen to acquire higher amplitude once the ERP transience ended. In the theta plots, it can be witnessed that the theta gained higher amplitudes during ERP transience. The ERP can be seen to be more prominent in the narration phase as compared to the narration + movement phase.

3.2 Observed ERP characteristics. ERP was observed at Fz and Pz during narration and narration with synchronized movements only The ERP at Fz (Figure 6). was more prominent as compared to the one at Pz. At Fz, a surge in theta power coordinated with occurrence of ERP and it eventually decreased. Theta power before and after the ERP was relatively less. Similar theta power characteristics are noticed at Pz as well. The alpha at Fz can be seen to surge after the occurrence of ERP. During and before it, Alpha can be seen to at a relatively lower power. At Pz (Figure 7). Similar alpha characteristics in relation to ERP are noticed.

4. Pupillometry data analysis. The average pupil diameter was calculated by averaging the average of left and right pupil diameters. The information is sampled at 30 Hz by the eye tracker. The average pupil dilation was 4.48 mm when the robot was in the narration synched with movement phase. It was at 4.1 mm at narration state. At rest the average dilation was at 3.75 mm (Figure 8). Clearly, the robot movement elicited the most dilation.

5. Discussion. The pupil dilates as a function of mental effort. It is supposed to be a representation of cognitive control, attention and cognitive processing [12]. In our case, the pupil dilation is seen to have a marked difference between resting state and stimulus state which rightfully indicates these claims. When the robot just narrates or when it narrates along with movement, it does seem to invoke attention as compared to resting state. However, there is a difference in pupil dilation within the 2 stimulus states of narration and movement + narration. The pupil seems to have more dilation in the case of narration synched with movement as compared to pure narration. Increased Fm Θ power is also seen as a case of increased attention as well as memory load and is noticed to have increased in a case of narration state as compared to rest state. This corroborates well with increased pupil dilation.



FIGURE 8: Pupillometry Information graph

A decrease in alpha at Fz during stimulus phase also indicates increased attention. further increase in dilation of pupil during the narration and movement can be attributed to increase in cognitive load since both aural and visual stimuli need to be processed. In the same phase, theta power at Fz decreased.



FIGURE 9: Evident Anthropomorphism

The movement would have played as a distractor and would have brought down focused attention to aural stimulus. All the subjects reported the same after the test. The higher theta during narration along with increased pupil dilation, also certifies increased memory processing, since all the subjects reported doing mental calculation of the fore coming mathematical table statement in the 4000 ms inter stimulus window. It has also been established that pupil dilation is an indicator of episodic memory retrieval and creation and there is a marked change in dilation between familiar and new stimulus [13]. The increased pupil dilation during the last phase of narration and movement is a novel condition as compared to narration state only and can be attributed to episodic memory maintenance at this stage. Reduced theta power at this stage as compared to narration can be attributed to episodic memory maintenance at the stage.

Subjects were tested for remembrance of math tables periodically after their exposure to the stimulus. They were tested for remembrance of the math tables immediately after the test, after 24 hours and after 4 days. they were asked for numbers that fell in the narration only and narration plus movement stages of the stimulus. Almost all the subjects were able to recall the answers immediately post the experiment. Only one subject was able to recollect without mental calculation the day after. The rest who had reported that the robot movement was distracting could not do immediate recall of the numbers and attempted calculating them again. However, all of them termed it as the "robot action test" remembering the robot action more than the content of the narration. It was also noticed that all the subjects maintained eye contact with the robot and were tracking its face when it swayed (Figure 9).

6.Conclusion. Based on the content of recollection, it can be claimed that this exercise had played a huge impact on the creation of episodes as compared with creation of factual memory. The fact that eye contact was maintained during robot motion would indicate the aspect of anthropomorphism setting in [14]. This is bound to influence the episodic memory creation in a very effective way. As is evident, human like features of talking and movement coupled with physical appearance effectively influence the storing of instances of interaction in humans. Since, the robot are just human like and not real living beings, there is room for making more extraordinary actions that can play a role in creating long standing episodes in subjects. In that way, the robot can be effectively employed as a mediator of information fusion. It can also be safely said, that understanding brain dynamics during human robot interaction opens up new avenues of understanding the paradigm better. Children acquire information through daily experiences. These experiences act as effective scaffolds for further learning [15]. The tests, as done in this paper should be extended to children and understanding their dynamics will lead to more insights into the effect of using robots as effective mediators of learning in children.

3. Acknowledgments. Our thanks to LSID lab, IIT Kanpur and Wagatsuma Lab, KIT for their earnest support in carrying out this study. The authors also want to acknowledge that the collaboration between the two labs got strengthened by the SAKURA exchange program initiated by JST. This work was partly supported by JSPS 16H01616 and the New Energy and Industrial Technology Development Organization (NEDO) "Next-generation Artificial Intelligence: (3) Base technology for next-generation AI."

REFERENCES

- [1] B. Duffy, Anthropomorphism and the social robot, *Robotics and Autonomous Systems*, vol. 42, no. 3-4, pp. 177-190, 2003.
- [2] C. Breazeal, Toward sociable robots, *Robotics and Autonomous Systems*, vol. 42, no. 3-4, pp. 167-175, 2003.
- [3] Breazeal, Cynthia L. Designing sociable robots. *MIT press*, 2004.
- [4] C. Breazeal, Emotion and sociable humanoid robots, *International Journal of Human-Computer Studies*, vol. 59, no. 1-2, pp. 119-155, 2003.
- [5] Tulving, Endel. Elements of episodic memory, 1985.
- [6] E. Tulving, Episodic Memory: From Mind to Brain, *Annual Review of Psychology*, vol. 53, no. 1, pp. 1-25, 2002.
- [7] L. Aftanas and S. Golocheikine, Human anterior and frontal midline theta and lower alpha reflect emotionally positive state and internalized attention: high-resolution EEG investigation of meditation, *Neuroscience Letters*, vol. 310, no. 1, pp. 57-60, 2001.
- [8] W. Klimesch, M. Doppelmayr, H. Russegger and T. Pachinger, Theta band power in the human scalp EEG and the encoding of new information, *NeuroReport*, vol. 7, no. 7, pp. 1235-1240, 1996.
- [9] L. Hsieh and C. Ranganath, "Frontal midline theta oscillations during working memory maintenance and episodic encoding and retrieval", NeuroImage, vol. 85, pp. 721-729, 2014.
- [10] Kropotov, Juri D. Quantitative EEG, event-related potentials and neurotherapy. *Academic Press*, 2010.
- [11] R. Ishii, K. Shinosaki, S. Ukai, T. Inouye, T. Ishihara, T. Yoshimine, N. Hirabuki, H. Asada, T. Kihara, S. Robinson and M. Takeda, Medial prefrontal cortex generates frontal midline theta rhythm, *NeuroReport*, vol. 10, no. 4, pp. 675-679, 1999.
- [12] S. Wierda, H. van Rijn, N. Taatgen and S. Martens, Pupil dilation deconvolution reveals the dynamics of attention at high temporal resolution, *Proceedings of the National Academy of Sciences*, vol. 109, no. 22, pp. 8456-8460, 2012.
- [13] K. Hellmer, H. Söderlund and G. Gredebäck, The eye of the retriever: developing episodic memory mechanisms in preverbal infants assessed through pupil dilation, *Developmental Science*, 2016.
- [14] B. Duffy, Anthropomorphism and the social robot, *Robotics and Autonomous Systems*, vol. 42, no. 3-4, pp. 177-190, 2003.
- [15] Lev Vygotsky "Mind in society"