Qualitative Model for Neuro-Functional Mechanism Inducing Human Error Detection to Humanlike Agents Based on Cortico-Cerebellar Function

Yoshimasa Tawatsuji[1(✉)] and Tatsunori Matsui[2]

[1] Graduate School of Human Sciences, Waseda University, Saitama, Japan
wats-kkoreverfay@akane.waseda.jp
[2] Faculty of Human Sciences, Waseda University, Saitama, Japan
matsui-t@waseda.jp

Abstract. The uncanny valley is a major phenomenon that occurs during the implementation of facial expressions in a humanlike agent. Recent studies have explained that the negative impression of the agent's expressions is a "prediction error." We hypothesized that the prediction error stemmed from the differential output of dual processing, - that is, emotional and rational processing, - and it negatively affected a person's impression of the facial expression of the agent, which was similar to that when perceiving an eerie sensation. A qualitative computational model was proposed to clarify this mechanism according to cortico-cerebellar function to compute the prediction of the humanlike agent's facial movement. Thus, rapid emotional processing was indispensable with the thalamus playing an essential role in the processing.

Keywords: Uncanny valley · Brain functional connection · Cerebellum

1 Introduction

1.1 Background

The improvement of 3D computer graphics and robotics has gradually affected our lives by introducing humanlike agents, such as androids. In the field of human computer interaction or human agent interaction, it has been pointed out that these agents require the implementation of human (or user)-centered or human-friendly functions. Therefore, to familiarize these agents with human beings, they must be replicated into a genuine person. That is, along with their appearance, their functions should mimic those of a human. One of their important functions is the "facial expression," which conveys much information, such as emotional, social and health. However, it has been criticized that extraordinary resemblance of agents to humans elicits negative impressions by humans (ref. "the uncanny valley" [1]). One critical factor of such negative impression against agents is the complex muscle system of the facial expressions in humans. In this paper, we focused on the slight differences in the agent's facial expression that cause the negative emotional response from humans.

C. Stephanidis (Ed.): HCII 2016 Posters, Part II, CCIS 618, pp. 194–200, 2016.
DOI: 10.1007/978-3-319-40542-1_31

1.2 Previous Study Related to Inconsistencies with the Prediction

Previous studies have implied the importance of the "prediction error" framework. Here, "prediction" implies the fast emotional response toward the visual information of humanlike agent imprecisely captured by the retina as if it were a human. This predictive response can be modified by slow and detailed processing of the information because "it is not a human." The inconsistency switches on the negative emotional processing (e.g. the eerie feeling).

According to Seyama et al. [2], the extraordinary expansion of the humanlike agent's eye-size induced negative impression on an observer. Contrary to the observer's prediction of balance between facial feature and its component (eye) size, the extraordinary eye-size negatively affected the impression. Saygin et al. [3] conducted experiments in which the participants observed the actions, such as waving of a hand, by three types of agents – a person, humanlike android, and mechanical android. The participants' reactions were scanned through fMRI to analyze the physiological (blood oxygenation level dependent: BOLD) responses to each agent. The results indicated that compared with the actions of a person or a mechanical android, the BOLD response increased when observing the actions of the humanlike android. The results were explained as follows: the humanlike android which would typically not move mechanically is not predictable and then an error signal occurs as a prediction error [3].

As suggested in the previous study, the prediction error framework must contribute to our understanding of the mechanism inducing the uncanny valley phenomenon. Thus, we adapted the framework to account for elicitation of the negative emotional response against the facial expression of humanlike agents.

1.3 Purpose

In this paper, we propose a functional brain-model of the mechanism that generates negative emotions against the facial expressions of humanlike agents. We focused on the facial expression "smile," involving the zygomatic muscle movement.

Most importantly, we modeled the functional brain model not on the biological level but the systematic level. The biological model helped understand the details of the neuronal connection and the way in which the ionic transition is achieved. However, this model cannot explain more than how each region is connected and what architecture it has. Many brain areas should be engaged in the process of generating an eerie feeling in order for us create a bran model in the appropriate systematic level. We thus constructed the brain model through qualitative description of functional connectivity of each system.

2 Previous Model

In this section, we introduce and briefly describe some tasks of the previous model [4]. The previous proposed model provided an abstract neural-based explanation for the occurrence of the uncanny valley. We hypothesized that information processing resulted in emotional and cognitive responses until the occurrence of the eerie feeling, well known as the uncanny valley. This basic concept of dual parallel information processing is based on the "dual pathway of emotion" [5], and the differential processing results signal an alert [4] (see also: [6]). In this concept, the emotional information processing which is rapid and imprecise, generates an immediate response toward the humanlike agent as if it were a person. Contrastively, the cognitive information processing that is processed more slowly and in detail, generates a rational response toward the agents. The emotional information processing undergoes without the cortical modulation yet, and the cognitive information processing resulted through cortical networks such as visual pathways. This dual parallel information processing model is expected to fit the "prediction error" framework.

The model has a limitation in that the input of the model is defined as the degree of abnormality of the ad hoc facial component (eyes, nose or mouth) features, which in turn, reflects the omission of the computational process for the abnormality in the model. Here, the abnormality should be the result of error calculation between prediction and perception. Therefore, in this study, we explain the error detection process. Figure 1 depicts our entire model.

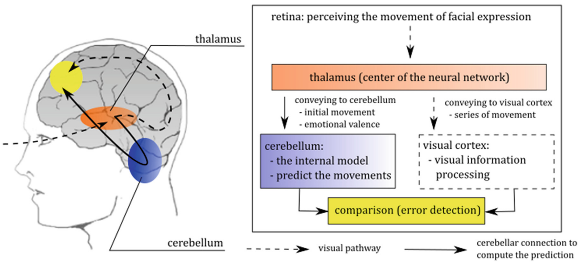

Fig. 1. Brain area contributing to prediction and error detection (Color figure online)

3 Error Detection Mechanism Subserved by Cerebellar Function

3.1 Calculation of Error Detection

First, we formulated the calculation method of the degree of abnormality. Generally, abnormality should be perceived after the error for an event is detected. The error detection can be calculated using the following two methods. The first method realizes error detection by pre-learning of the association of a certain event (stimuli itself) with a response (e.g. fear conditioning). The other method realizes error detection by detecting the deviation from a possible event (prediction): this method should correspond to predicted error detection. Therefore, we adopted the second calculation method.

The calculation of error detection requires a deviation between the predictive and actual perceived zygomatic muscle movements, that is, $\boldsymbol{m}^*$ and $\boldsymbol{m}$, respectively. The degree of abnormality is qualitatively proportional to the deviation d, and is defined as

$$d(\boldsymbol{m}^*,\ \boldsymbol{m})(t) = \frac{\|\boldsymbol{m}^*(t) - \boldsymbol{m}(t)\|}{\|\boldsymbol{m}^*(t)\|}. \tag{1}$$

3.2 Neural System Subserving Calculation for Generating Facial Movement

Before proposing the computational neural system for predicting how a humanlike android's facial muscles move, we introduce the qualitative model representing facial movement in humans. In this model, the main functional brain system subserves the internal model preserved in the cerebellum (to predict the muscle movement) through the functional connectivity between the cortex and the cerebellum via pontine nuclei.

Both the inverse and forward models should be a prerequisite for predicting how a person moves his or her facial muscle. Such a model is expected to be implemented in the cerebellum [7]. Gomi [8] clarified that "the cerebellum (the burst of Purkinje cell) should code the dynamical components, such as acceleration and velocity of voluntary eye movement" (original in Japanese).

Researchers have argued about the roles of the cerebellum on emotional expressions in the pseudobulbar affect, in which the cortico-pontine-cerebellar connectivity plays an important role in modulating the emotional expressions [9]. According to Ahmed et al. [9], the pseudobulbar affect is "characterized by uncontrolled crying or laughing which may be disproportionate or inappropriate to the social context." The cerebellum modulated the motor information relayed by the motor cortex in consistent with the emotional information relayed by frontal and temporal cortex to output. In addition, the facial expressions comprised the various combinations of facial muscles. The facial nucleus to which facial muscles belong is located at the caudal pontine [10]. Thus the facial expression is executed as follows: (i) the cerebral cortex determines the emotional states, (ii) the cerebellum modulates the motor commands appropriately to a social context, and (iii) the pontine receives the modular commands to move facial muscles.

In study, we modeled the cortico-cerebellar function to modulate the motor commands in accordance with the social context by using the following qualitative differential equation:

$$\frac{d^2\boldsymbol{m}}{dt^2} + \lambda_1(\boldsymbol{x})\frac{d\boldsymbol{m}}{dt} + \lambda_2(\boldsymbol{x})\boldsymbol{m} = \boldsymbol{0} \tag{2}$$

where the parameters λ_1 and λ_2 represent the modulating velocity and acceleration of facial muscle movement, dependent on emotional states $\boldsymbol{x}$. Gomi [8] experimentally demonstrated that the inverse dynamics model with a second order differential equation for determining the eye direction was persuasive in representing the burst patterns of Purkinje cells.

Define $\boldsymbol{y}$ and $\boldsymbol{\xi}$ as satisfying $\boldsymbol{y} = \frac{d\boldsymbol{m}}{dt}$, $\boldsymbol{\xi} = (\boldsymbol{m}\ \boldsymbol{y})^T$ (where $\boldsymbol{x}^T$ denotes for transpose of vector $\boldsymbol{x}$), then Eq. (2) can be transformed into

$$\frac{d\boldsymbol{\xi}}{dt} = \boldsymbol{\Lambda}\boldsymbol{\xi}, \text{ where } \boldsymbol{\Lambda} = \begin{pmatrix} 0 & 1 \\ -\lambda_2(\boldsymbol{x}) & -\lambda_1(\boldsymbol{x}) \end{pmatrix}. \tag{3}$$

Let z be an output representing the state of facial muscles, then Eq. (3) and $z = \boldsymbol{m}$ can be regarded as the equation of states in the framework of control theory, where $\boldsymbol{\Lambda}$ represents the internal model generating the series of facial movements. Hence, we introduced the model generating the emotional facial movement according to the cortico-cerebellar function.

3.3 Proposal for the Mechanism to Predict the Facial Movement of Other's

Prediction processing requires immediate processes for how the agent should move the muscles, especially before the movement is perceived. Let us assume that self-internal model $\boldsymbol{\Lambda}$ should be applied to the other's internal model $\boldsymbol{\Lambda}^*$, that is, $\boldsymbol{\Lambda}^* = \boldsymbol{\Lambda}$. Then, for perceiving the agent's facial movements at the initial momentary duration $\boldsymbol{\xi}^*(0) = (\boldsymbol{m}(0) \quad \boldsymbol{y}(0))^T$, the state of facial muscles at the arbitrary time can be easily calculated analytically by using Eq. (3). In this case, this application means the prediction.

4 Qualitative Model Representing the Transition of Emotional States

The dynamical system framework is an elaborate approach to model the transitioning of emotional states, especially considering the implementation of emotional states into the robot [11]. We introduce a model not only represents the transitioning of emotional states but also comprising valid neuropsychological knowledge.

The transition of emotional states is mathematically written as

$$\frac{d\boldsymbol{x}}{dt} = f(\boldsymbol{x}) + \boldsymbol{u}(t). \tag{4}$$

As is shown in the equation, emotional state $\boldsymbol{x}(t)$ transits in accordance with the unknown function and the given input $\boldsymbol{u}(t)$. In this study, the emotional states represents the physiological state determined by the states of certain neural systems.

Posner et al. [12] proposed the neural systems of the circumplex model proposed by Russell. In their proposal, a mesolimbic dopamine system plays a role in emotional valence, and reticular formation in arousal. These neural structures are an efficient clue in characterizing the function f in Eq. (4).

5 Discussion

On the basis of the qualitative model for prediction, Eq. (3) requires rapid information processing of emotions, as $\boldsymbol{\Lambda}$ is dependent on variable $\boldsymbol{x}$. This rapid emotional processing is expected to be relayed through "low-road" route [4]. Here the thalamus, which is a transit point between sensory receptor and the amygdala [4], should play an important role in processing emotional information.

We focused on the thalamus because of two main reasons. First, the absence of neural projections between the amygdala and cerebellum causes the more latency to convey emotional information to the cerebellum. Second reason is the thalamus is connected to various regions such as amygdala, cerebellum, basal ganglia and cortical area [10]. Pessoa et al. discussed the role that the pulvinar plays in human visual emotional processing [13]. The pulvinar is the part of the thalamus connected to the amygdala. In particular, the thalamus rapidly processes the emotional information to determine the internal model as the functional center of the entire brain networks.

6 Conclusion

In this paper, we proposed the functional brain model for the mechanism generating negative emotion in response to the facial expressions of the humanlike agents. The model was constructed based on the internal model of the cerebellum. The model implied the existence of rapid emotional processing and the role of the thalamus in such emotional processing.

Acknowledgement. We would like to thank Editage (www.editage.jp) for English language editing.

References

1. Mori, M.: The uncanny valley (2012). http://www.getrobo.com/. Accessed 16 Feb 2013
2. Seyama, J., Negayama, R.S.: The uncanny valley: effect of realism on the impression of artificial human face. Presence Teleoperators Vis. Environ. **16**(4), 337–351 (2007)
3. Saygin, A.P., Chaminade, T., Ishiguro, H., Driver, H.J., Frith, C.: The thing that should not be: predictive coding and the uncanny valley in perceiving human and humanoid robot actions. Cogn. Affect. Neurosci. **7**(4), 413–422 (2012)
4. LeDoux, J.: The Emotional Brain: The Mysterious Underpinnings of Emotional Life. Simon & Schuster Paperbacks, New York (1996)
5. Tawatsuji, Y., Muramatsu, K., Matsui, T.: Qualitative description representing brain functional connection for human emotional states toward human-like agents. Trans. Jpn. Soc. Artifi. Intell. **30**(5), 626–638 (2015)
6. Shimada, M., Minato, T., Itakura, S., Ishiguro, H.: Uncanny valley of androids and its lateral inhibition hypothesis, Robots and Human Interactive Communication. In: The 16th IEEE International Symposium on RO-MAN 2007, pp. 374–379 (2007)
7. Kawato, M.: Nou no Keisan Riron [Computational Theory of Brain]. Sangyo Tosho, Tokyo (1996)
8. Gomi, H.: The cerebellar role and computational model for reflective eye movement. Vis. J. Vis. Soc. Jpn. **6**(4), 153–160 (1994). (in Japanese)
9. Ahmed, A., Simmons, Z.: Pseudobulbar affect: prevalence and management. Ther. Clin. Risk Manag. **2013**(9), 483–489 (2013)

10. Fujiyama, F., Nishimura, Y., Yamao, Y., Shibata, J., Usami, K., Matsumoto, R. Trans.: In Principles of Neural Science Fifth Edition, Medical Science International, pp. 335–365 & 999–1015 (2014)
11. Miwa, H., Itoh, K., Takanobu, H., Takanishi, A.: Development of a human-like head robot for emotional communication with human: 3rd Report, Introduction of the equations of emotion and robot personality. Trans. Jpn. Soc. Mech. Eng. C **70**(699), 3244–3251 (2004)
12. Posner, J., Russell, J.A., Peterson, B.S.: The circumplex model of affect: An integrative approach to affective neuroscience, cognitive development, and psychology. Dev. Psychopathol. **13**(3), 715–734 (2005)
13. Pessoa, L., Adolphs, R.: Emotion processing and the amygdala: from a 'low-road' to 'many roads' of evaluating biological significance. Nat. Rev. Neurosci. **11**(11), 773–783 (2010)