

The Eyes as an Indicator of the Mind—A Key Element of Heart-Touching-AI

Shigeto Furukawa, Makoto Yoneya, Hsin-I Liao, and Makio Kashino

Abstract

Technology for *reading* the mind is an important element of Heart-Touching-AI, which aims to understand and support the human mind and body. The well-known saying “The eyes are the windows to the soul” expresses the fact that pupil response and eye movement can reveal information about a person’s mental state. Research is underway at NTT Communication Science Laboratories to find the principles of understanding a person’s mental state from his or her eyes in order to develop basic technologies that can apply such a capability to real-world problems.

Keywords: eye movement, pupil, mind-reading

1. Introduction

Artificial intelligence (AI) continues to advance and is being applied in numerous fields with various objectives, including with the aim of achieving an enriching society that fulfills the needs of mind and body. Accordingly, AI technologies have recently gone beyond machine intelligence for making logical inferences and are now being applied to understand human intellect, sensitivities, and emotions—all aspects related to the human mind. Thanks to advances in machine learning and other technical fields, there are now robots that can recognize human emotions by analyzing facial expressions and vocal patterns.

However, human emotions and psychological states are not always expressed by such explicit information as a person’s facial expressions or voice. A machine would not surpass human performance in recognizing another’s emotions as long as it uses only explicit information. Consequently, by adding the ability to measure unconscious physiological responses such as eye movement and heartbeat, NTT aims to achieve AI that can *read* a person’s implicit mental state in a

way that only a computer could recognize and communicate with humans more effectively than even humans can.

The brain-computer interface (BCI) based on electroencephalographic measurements has progressed as a basic technology for learning about a person’s implicit mental state. This technology, however, has significant restrictions and suffers from high noise and low robustness, and as a result, its use has largely limited to certain fields such as medical care. Consequently, BCI is not considered to be ready for use in everyday scenarios.

NTT Communication Science Laboratories has developed a basic technology for evaluating a person’s mental state by examining the person’s eyes, as an alternative to electroencephalographic measurements. This technology has practical advantages in that it enables eye-related information to be obtained by a camera or other device in a non-invasive and non-contact manner and enables data to be measured without interfering with people’s natural movements and sensations.

2. Information expressed by the eye

What kind of information can we obtain from the human eyes? The pupil, the black hole surrounded by the iris, adjusts the amount of light admitted to the retina. The light reflex property of the pupil is well known; the pupil constricts in a bright environment (miosis) and dilates (mydriasis) in a dark environment. A number of studies have been published that suggest a relationship exists between pupil diameter and various cognitive processes such as target detection, perception, learning, memory, decision making (e.g., [1]). The factor that determines the pupil diameter is the balance between the mutually antagonist actions of the pupil's sphincter muscle and dilator muscle, which are respectively governed by the parasympathetic nervous system and the sympathetic nervous system. The pupil cannot be controlled consciously (voluntarily).

The eyeball, meanwhile, moves under the control of the external eye muscle governed by the oculomotor nerve. Eye movement can be a voluntary process as when turning one's gaze toward something that one wants to look at. However, it can also be involuntary in the case when one's line of sight moves unconsciously toward a novel or conspicuous stimulus, or when small, jerk-like eye movements (microsaccades) occur in a state of visual fixation. Such involuntary eye movements have been said to reflect a state of caution induced by external stimuli or a state of latent caution (not accompanied with a change of gaze) [2]. Many evaluations of microsaccades have been done to measure their frequency of occurrence, and we as well have taken up a detailed evaluation of their dynamic characteristics and have been achieving results.

A variety of neural pathways contribute in a complex way to the control mechanisms governing pupil diameter and eye movement. This system includes nerve nuclei that are thought to have a deep relationship with cognitive processes. Nuclei whose functions are of particular interest include the locus coeruleus and the superior colliculus in the brain stem.

The locus coeruleus appears to play an important role in alertness, anxiety, stress, attention, and decision-making. These activities interact with various regions of the brain via noradrenaline pathways, and the pupils dilate through actions of the sympathetic nervous system [1, 3].

The superior colliculus, meanwhile, has been said to play a role in spatial attention and in controlling the direction of sight, but recent research has revealed

that it can also interact with the pupils via the parasympathetic and sympathetic nervous systems [1, 2]. Neurons inside the superior colliculus react to stimuli associated with visual, somatic, and auditory sensations, and there is a *spatial map* in which the positions of such neurons correspond to different points in space where the stimuli appear. The superior colliculus not only receives input from lower-level sensory systems that it transmits to higher-level sensory areas in the cortex, but it also receives input from those higher-level sensory and associated brain areas via multiple pathways. It also receives projections from the basal ganglia and locus coeruleus as well. Reading the mind via pupil response and eye movement is based on inferences of the cognitive processes that are supported by this complex nervous system.

3. NTT activities in Heart-Touching-AI

We introduce here our recent efforts in developing Heart-Touching-AI. A number of experiments have confirmed that the pupil and eye movements react to the conspicuousness (saliency) of sound and to the degree that the sound differs from what was expected (surprise). Some of these preliminary findings have been presented in previous articles in this journal [4, 5].

In the case of music, the saliency and surprise of tone and melody are thought to be important factors that determine a person's affinity and preference for a particular song or piece of music. We have created a model for estimating a person's affinity and preference for a piece of music by extracting appropriate multidimensional features from pupil response and eye movement and using machine learning technology (**Fig. 1**). We found that with this model, it is possible to estimate a listener's subjective music rating score with a certain level of accuracy [4].

A similar approach can be applied to evaluation of stimuli other than music, for example, fondness (preference) for a face. In an experiment, we presented images of faces to the observers while measuring their pupil responses. At this time, miosis was found to occur when the brightness of those images was changed. However, it was also found that the extent of miosis differed between the presented images even though the change in brightness was the same for all images (**Fig. 2**). We asked the observers to subjectively evaluate facial-image preference and found that the extent of miosis corresponded to the observer's subjective evaluation [6]. Furthermore, using a model similar to the one described above, we confirmed that

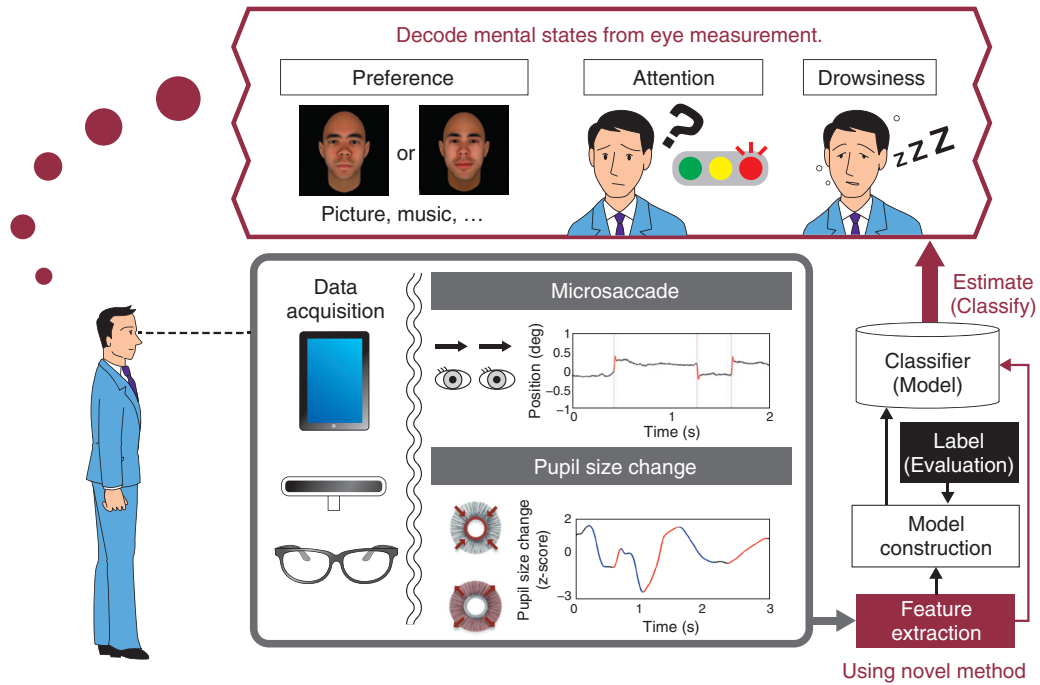


Fig. 1. Model that estimates a person’s latent state of mind based on eyeball movements and pupillary reflex.

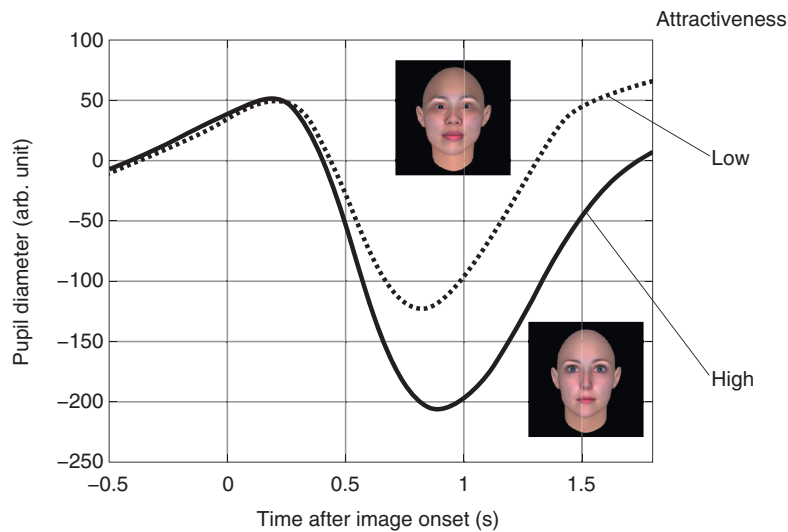


Fig. 2. Example of pupillary response to facial images.

an individual’s evaluation with respect to facial preferences could be predicted with high accuracy by combining a variety of features including those of pupil response and eye movement.

It would be natural to say that this correlation between pupil diameter and facial-image preference

obtained in the above experiments shows how the cognitive process associated with facial-image preference appears as a change in pupil diameter. However, we also consider the possibility of an inverse cause-and-effect relationship, in which differences in certain features in individual images give rise to

differences in miosis, which in turn acts on the cognitive process.

To examine this possibility, we conducted an experiment on manipulating the amount of miosis by introducing a change in brightness to the background picture instead of to the facial images themselves [6]. We found that the subjective evaluation of facial-image preferences changed in conjunction with the amount of empirically introduced miosis. This finding has important implications. It can be said that the pupil not only expresses the state of mind but that it is also a window that will enable AI to *act* on a person's mind or mood.

4. Future outlook

In research up to now, the emphasis has been on music preferences and facial preferences as targets of estimation. However, targets that can be estimated through eye measurements are not limited to preferences. As explained at the beginning of this article, the pupil response and eye movement are governed by the complex nervous system. Therefore, it is inferred that the eyes latently reflect various cognitive processes such as pleasure/displeasure, drowsiness, tension, and fear. The framework of feature extraction and machine learning at the core of our technologies is generic enough to be applied to any of those cognitive aspects (leaving aside the accuracy of the learning outcome).

However, a necessary condition here for the machine learning is that *correct data* are *correctly* obtained. For example, to learn what types of eye movements occur when people are sleepy, the eye movements of a person in an actual state of drowsiness must be recorded. Up to now, it has been possible to obtain preferences related to targets such as music and facial images through questionnaires administered in a laboratory, so it has been relatively easy to obtain correct data. However, the estimation of feelings such as fear and anger is difficult to record in a laboratory because of ethical issues, and it is also difficult to quantify subjective emotional states, even by the observers themselves.

Another issue here is that pupil response and eye movement can be affected by a variety of external and internal factors, and it is therefore difficult to find simple correspondence with specific cognitive processes or neural activity in real-world situations. In future research, sophisticated psychological and behavioral experiments and independent biological reaction measurements will be necessary to accu-

rately evaluate unconscious cognition. Given that many evaluation methods themselves have not yet been established, the only way to solve these problems is to employ a variety of measurement and analysis techniques while keeping their underlying mechanisms and limitations in mind, thereby enhancing our understanding a little at a time. At NTT, we are addressing these issues through basic research that spans diverse fields from physiology and psychophysics to machine learning technologies and through mutual interaction among those fields.

Of course, development in measurement technology itself will also be necessary. The most common approach to measuring pupil and eye movements is to apply image processing to images captured by camera. Today, the smartphones, tablets, and personal computers that we have all come to use are equipped with web cameras, so we can consider that platforms capable of measuring eye activity are expanding all around us. The frame rate of a typical web camera, however, is only several dozen frames per second (fps). If high-speed image capture becomes possible, we can envision the catching of latent reactions of short duration such as the dynamic characteristics of microsaccades. In our research, we are capturing images at 1000 fps using specialized eye-measurement devices, so one key to applying this achievement to general devices will be the development of high-speed cameras.

Provided that we accumulate basic knowledge, collect substantial data via sound experiments, and upgrade measurement technologies, we feel that the day in which AI will be able to read a person's mind (or act on a person's mind) by looking at that person's eyes is not that far off.

References

- [1] C. A. Wang and D. P. Munoz, "A Circuit for Pupil Orienting Responses: Implications for Cognitive Modulation of Pupil Size," *Current Opinion in Neurobiology*, Vol. 33, pp. 134–140, 2015.
- [2] S. Martinez-Conde, S. L. Macknik, X. G. Troncoso, and D. H. Hubel, "Microsaccades: a Neurophysiological Analysis," *Trends in Neuroscience*, Vol. 32, No. 9, pp. 463–475, 2009.
- [3] G. Aston-Jones and J. D. Cohen, "An Integrative Theory of Locus Coeruleus-norepinephrine Function: Adaptive Gain and Optimal Performance," *Annual Review of Neuroscience*, Vol. 28, pp. 403–450, 2005.
- [4] M. Kashino, M. Yoneya, H.-I. Liao, and S. Furukawa, "Reading the Implicit Mind from the Body," *NTT Technical Review*, Vol. 12, No. 11, 2014.
<https://www.ntt-review.jp/archive/ntttechnical.php?contents=ntr201411fa6.html>
- [5] S. Furukawa, S. Yamagishi, H.-I. Liao, M. Yoneya, S. Otsuka, and M. Kashino, "Biological Measures that Reflect Auditory Perception," *NTT Technical Review*, Vol. 13, No. 11, 2015.

<https://www.ntt-review.jp/archive/ntttechnical.php?contents=ntr201511fa3.html>

- [6] H.-I. Liao, S. Shimojo, and M. Kashino, "Correspondence between Pupillary Response and Facial Attractiveness," Oral presentation at the 11th Asia-Pacific Conference on Vision, Singapore, 2015.



Shigeto Furukawa

Senior Research Scientist, Supervisor, Group Leader of Sensory Resonance Research Group, Human Information Science Laboratory, NTT Communication Science Laboratories.

He received a B.E. and M.E. in environmental and sanitary engineering from Kyoto University in 1991 and 1993, and a Ph.D. in auditory perception from University of Cambridge, UK, in 1996. He was a postdoctoral associate at Kresge Hearing Research Institute at the University of Michigan, USA, from 1996 to 2001, where he conducted electrophysiological studies on sound localization, specifically the representation of auditory space in the auditory cortex. He joined NTT Communication Science Laboratories in 2001. Since then, he has been involved in studies on auditory-space representation in the brainstem, assessing basic hearing functions, and the salience of auditory objects or events. In addition, as the group leader of the Sensory Resonance Research Group, he is managing various projects exploring mechanisms that underlie explicit and implicit communication between individuals. He was the principal investigator of a Ministry of Internal Affairs and Communications SCOPE (Strategic Information and Communications R&D Promotion Program) commissioned research project on auditory salience. He is a member of the Acoustical Society of America, the Acoustical Society of Japan (ASJ) (member of the Executive Council), the Association for Research in Otolaryngology (ARO), and the Japan Neuroscience Society (JNSS).



Makoto Yoneya

Researcher, Sensory Resonance Research Group, Human Information Science Laboratory, NTT Communication Science Laboratories.

He received a B.E. and M.Sc. in engineering from the University of Tokyo in 2010 and 2012. He joined NTT Communication Science Laboratories in 2012 and has been studying biological signal processing, especially of eye movements. He is also interested in decoding people's thoughts based on brain or neural activity using machine learning methods and has researched decoding of the 'internal voice' using magnetoencephalography signals and multi-class SVM (support vector machine). He is also studying auditory signal processing and is currently developing a mathematical model of auditory salience. He received the 2011 Best Presentation Award from the Vision Society of Japan. He is a member of ASJ, JNSS, and ARO.



Hsin-I Liao

Senior Research Scientist, Sensory Resonance Research Group, Human Information Science Laboratory, NTT Communication Science Laboratories.

She received a B.S. and Ph.D. in psychology from National Taiwan University in 2002 and 2009. She joined NTT Communication Science Laboratories in 2012 and has been studying auditory salience, music preference, and preference of visual images. She has also explored the use of pupillary response recordings to correlate human cognitive functions such as auditory salience and preference decisions. During 2007–2008, she was a visiting student at California Institute of Technology, USA, where she studied visual preference using recorded eye movements and visual awareness using transcranial magnetic stimulation. She received a Best Student Poster Prize of the Asia-Pacific Conference on Vision (APCV) in 2008, a Travel Award of the Association for the Scientific Study of Consciousness (ASSC) in 2011, and a Registration Fee Exemption Award of the International Multisensory Research Forum (IMRF) in 2011. She is a member of the Vision Sciences Society, ARO, and JNSS.



Makio Kashino

Senior Distinguished Scientist/Executive Manager of Human Information Science Laboratory, NTT Communication Science Laboratories.

He received a B.A., M.A., and Ph.D. in psychophysics from the University of Tokyo in 1987, 1989, and 2000. He joined NTT in 1989. From 1992 to 1993, he was a visiting scientist at the University of Wisconsin (Prof. Richard Warren's laboratory), USA. Currently, he is a visiting professor in the Department of Information Processing, Tokyo Institute of Technology (2006–), and PI (principal investigator) of a JST CREST project on implicit interpersonal information (2009–2015). He has been investigating functional and neural mechanisms of human cognition, especially auditory perception, cross-modal and sensorimotor interaction, and interpersonal communication through the use of psychophysical experiments, neuroimaging, physiological recordings, and computational modeling.