Emotional-perception-control Technology for Estimating and Leading Human Emotions

Rika Mochizuki, Motohiro Makiguchi, Masahiro Kohjima, Masanori Yokoyama, and Ryuji Yamamoto

Abstract

The number of people who participate in live music and sporting events in remote environments has been increasing; however, participants must sacrifice the emotional experiences that can only be obtained on site, such as enthusiasm, a sense of unity, and contagious enthusiasm of the audience. Toward the creation of a world where people's emotions are actively amplified and resonate even in remote environments, we introduce an emotional-perception-control technology for generating a personal virtual site where participants can experience the unique pleasures of virtual reality through estimating their emotional-expression characteristics and leading (guiding) their emotions based on these characteristics.

Keywords: emotion modeling, emotion leading, personal virtual site

1. Emotional-perception-control technology

Online participation in live music, sports, and other events is expected to become essential along with on-site participation as a means of easily gathering people from around the world without worrying about location or distance. Various efforts are being made to improve the sense of presence, such as increasing the resolution of live images, creating multiple viewpoints by arranging multiple cameras, and increasing the wide viewing angle by using 360-degree cameras. However, the emotional experiences that can only be experienced on site, such as enthusiasm, a sense of unity, and contagion of enthusiasm felt in stadiums and live venues, are being lost. Therefore, we aim to create a world in which people's emotions are actively amplified and resonated even in remote environments by generating a personal virtual site optimized for each person's unique way of enjoying themselves, such as feeling a sense of unity in conjunction with other spectators or being absorbed in a space alone, by estimating each person's emotional expression characteristics and leading (guiding) their emotions based on these characteristics.

Emotional-perception-control technology leads to desirable emotions for users, such as enhancing enthusiasm and a sense of unity, through two core technologies: emotion estimation for estimating and understanding human emotional characteristics through sensing and data analysis and emotion leading for leading emotions through perceptual stimuli tailored to human emotional characteristics. By combining the estimation and leading of emotions including those of individuals, crowds, and interactions between the two, we can generate a personal virtual site for providing an optimal experience for each person (**Fig. 1**).

Emotion-estimation technology is used to quantitatively understand and model the emotions of individuals and crowds on the basis of sensed biometric, image, sound, and content data. Emotion-leading technology is an interaction technology for inducing
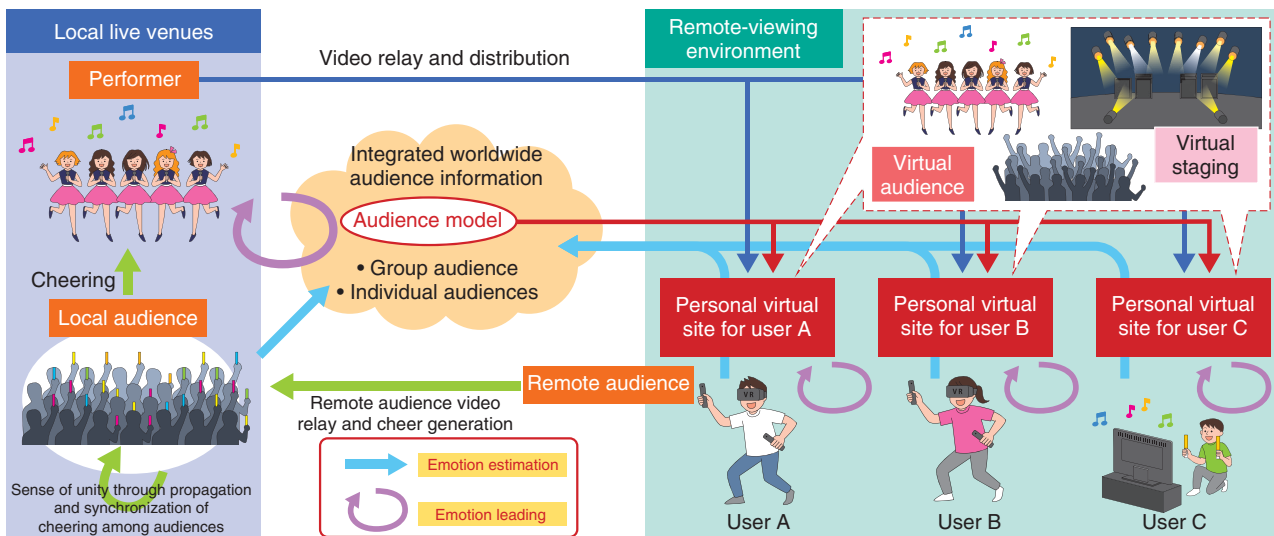
Fig. 1.   Personal virtual sites generated with emotion-estimation and emotion-leading technologies.

Fig. 2.   Image of a personal virtual live site.

the same or even more emotions in a remote environment as in a local venue using knowledge of perception, cognitive psychology, and human-computer interaction. Emotion-estimation technology estimates and models the states of local and remote audiences and reflects them in the virtual audience, while emotion-leading technology optimizes the behavior of the virtual audience and the staging of the venue according to the characteristics and emotional state of each remote audience member, as shown in Fig. 1.

This article introduces an emotional-perception-

control technology for generating a personal virtual live site (**Fig. 2**) for events such as concerts. By optimizing the presentation and interaction methods of the virtual audience for each remote audience member, the emotional experience is enhanced by perceptually enhancing the sense of unity and enthusiasm caused by the propagation and synchronization of cheering that occurs among audiences at different locations during live concerts.
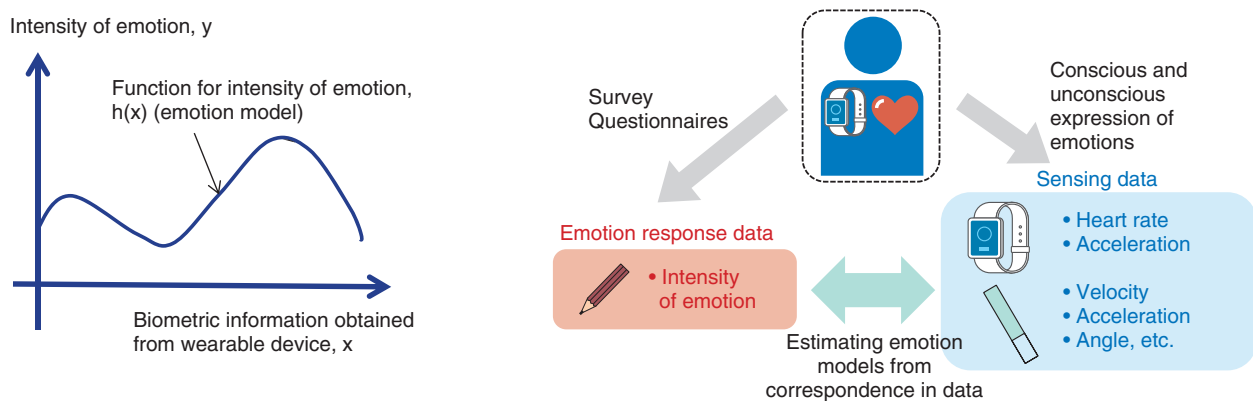
Fig. 3. Emotion-model estimation.

2. Emotion-estimation technology

In live music and sporting events, the point of excitement differs depending on the individual/entire audience's preference for music and staging of live events and their knowledge of sports. Suppose we can estimate the emotion of individual spectators and entire audiences without explicitly asking them. The magnitude of the effects of the music, staging, and camera work on emotional responses can then be measured, which helps evaluate and improve the music and staging. Estimated emotion can also be used for generating a personal virtual live site, as illustrated in Fig. 2. We can adaptively change the type of venue (such as a house and arena), staging, and movements of the surrounding virtual audiences for each site to enhance or suppress emotions. With these application scenarios in mind, we are developing methods for estimating (i) the emotions of individual audience members using wearable devices and (ii) the emotions (collective characteristics) of the entire audience using live videos of the actual event venue.

2.1 Estimating the emotions of individual audience members using biometric signals

This is a method that uses artificial intelligence (AI) to estimate the intensity of emotions, such as pleasure, discomfort, high arousal, and low arousal, and types of emotions, such as joy and sadness, felt by the user. This involves using biometric information obtained from wearable devices such as smartwatches and hitoe™*, which have become widespread, as input. The advantage of this method is that it can be applied to any environment where there are

no cameras or microphones, such as the living room or outside the home, because it uses a wearable device that the user wears in daily life. The method requires the training of AI, i.e., estimating an unknown function (*emotion model*), the input of which is sensed biometric information and output is the intensity or classification result of an emotion from the user's biometric information and an emotion response (e.g., a five-point rating of pleasure) at a certain point in time (**Fig. 3**). This training requires the user's subjective responses since emotions that are not observable are the model's output. However, such a simple five-point rating format has problems such as differences in interpretation of each point between participants [1], making it more difficult than the standard setup for training AI using non-subjective data such as whether a certain image is a cat or dog, and collection of large amounts of data. We are currently developing a method for handling such difficulties, including a data-collection format suitable for handling subjective data and an AI training method tailored to that format [2, 3].

2.2 Estimating the collective characteristics of the entire audience using live video from an actual event venue

This method estimates the collective characteristics of a group based on the observed group behavior (such as shouting, waving penlights, clapping, and hand-signing) in the live video of the actual event venue, including the audience seats. For example,

---

*   hitoe™: A functional fabric that is made by coating a conductive polymer on a cutting-edge fiber material called nanofiber fabric. A clothing-type vital sensor using hitoe™ can measure biosignals with high sensitivity without burden on the wearer.

how the audience's behavior is affected (on average) when the performer shouts out to excite them or how well the audience as a whole behaves in a synchronized and united manner. Unlike the above method involving wearable devices, this method does not require users to own a device and can be used at any event venue. However, the estimation target is not the emotion of an individual but the characteristics of the group as a whole. By using the estimated group characteristics, it is possible to understand which songs in a music concert generate a unified behavior of the entire audience. We can also enable the virtual audience in the personal virtual live site behave similar to that of the audience at the actual venue.

3. Emotion-leading technology

Emotion-leading technology is an interaction technology that naturally leads remote audience members to the desired emotional state when they participate in an event online. In many cases, online participation, such as live music streaming, differs from on-site participation in many ways such as participation from a computer or smartphone at one's home. Therefore, we are conducting research to enhance the emotional experience from two aspects: psychological and behavioral [4], to provide similar or better emotional experience in a remote environment as in a local environment. An example of the psychological aspect is the excitement from the enthusiasm of the surrounding audience, and an example of the behavioral aspect is the excitement of cheering, clapping, and waving penlights.

To enhance the emotional experience from the psychological aspect, we are researching an audience-presentation-optimization method for optimizing the presentation of audience members other than oneself to remote audience members.

As a concrete example of optimization, we can consider reducing the display of audiences that behave too differently from the target audience or are not excited at all and present audiences that behave and are excited in the same way as the target audience. Another example would be to highlight the behavior of audiences who are similarly cheering for a band member that the target audience likes. To make such optimization possible, it is first necessary to obtain data on the behavior of the audience. We focused on penlights as a typical item at music events and developed a sensing penlight equipped with multiple sensors such as acceleration (**Fig. 4**).

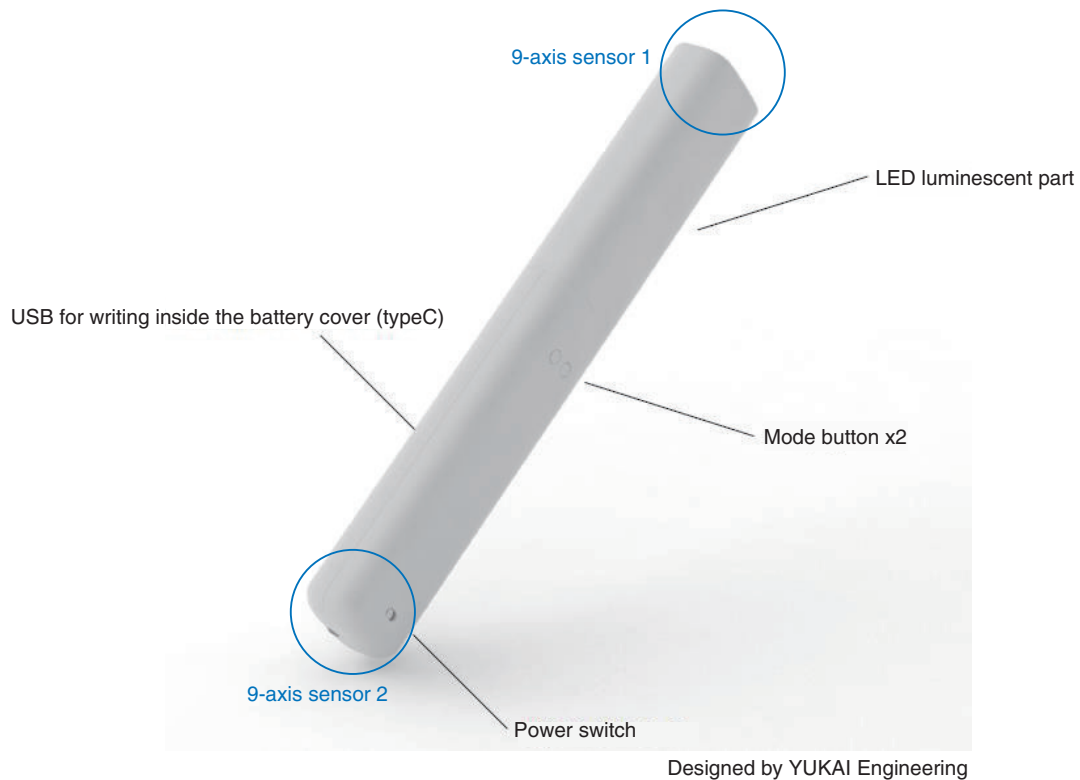The sensing penlight can detect the presence or absence of swing, swing scale, swing cycle, color conversion, and other data. The data obtained can be used to generate the optimal audience-presentation pattern for each remote audience. The behavior of the audience in response to the generated pattern is obtained again to tune the presentation pattern, thus improving optimization accuracy.

Our approach to enhancing the emotional experience from a behavioral aspect is based on the fact that in many online viewing environments, such as at home, it is difficult for the audience to cheer loudly or move as much as they would at a local venue. We are investigating a method of adding multimodal feedback stimuli to give the remote audience the illusion that their actions are larger than they actually are. This method aims to provide an experience in which the remote audiences feel as if they are performing the same physical actions as in the actual venue. By studying the methodologies for incorporating an on-site experience into the remote viewing experience, multiple emotion-leading methods and application patterns of such methods for each user can be developed. The selection scheme, intensity, and timing of application of emotion-leading methods can be optimized on the basis of the emotions of individual users or groups of users using emotion-modeling methods.

By combining emotion-leading technology and emotion-estimation technology, it will be possible to design a personalized virtual space and venue and generate a personalized virtual live site where each individual remote audience member can share their enthusiasm as well as be influenced by others'.

4. Future work

Hybrid events held in both real and remote environments have become more common, and it is important to enhance the experience that can only be had in a remote environment. Our emotional-perception-control technology aims to provide new experiences through the generation of a personalized virtual site, where each user can enjoy the experience of being involved in a space optimized for him or her by taking advantage of the effects of online performances that are impossible in reality, such as unlimited seating for audience members from around the world, freely changing seating arrangements, and the spread of one's actions throughout the venue. We will also consider extending this technology to fields other than live music performances.

9-axis sensor 1

LED luminescent part

USB for writing inside the battery cover (typeC)

Mode button x2

9-axis sensor 2

Power switch

Designed by YUKAI Engineering

LED: light-emitting diode
USB: Universal Serial Bus

Fig. 4.   Developed sensing penlight.

References

[1]   G. N. Yannakakis and H. P. Martinez, "Ratings Are Overrated!", Frontiers in ICT, Vol. 2, No. 13, 2015.
[2]   Y. Nanbu, M. Kohjima, and R. Yamamoto, "Learning of Regression Models from Pairwise Data and Target Variable Quantiles," Transactions on Mathematical Modeling and its Applications (TOM), Vol. 15, No. 3, pp. 19–28, July 2022.
[3]   M. Kohjima, "Uncoupled Nonnegative Matrix Factorization with Pairwise Comparison Data," Proc. of 2022 ACM SIGIR International Conference on the Theory of Information Retrieval (ICTIR), pp. 103–107, Madrid, Spain, July 2022.
[4]   D. Yamashiro, M. Aihara, C. Ono, H. Kanemura, K. Aoyagi, Y. Goto, Y. Iwadare, and S. Nakazawa, "Sympathetic Skin Response and Emotional Changes of Visual Stimuli," No to Hattatsu (Brain and Development), Vol. 36, No. 5, pp. 372–377, 2004 (in Japanese).

Rika Mochizuki
Senior Research Engineer, Symbiotic Intelligence Laboratory, NTT Human Informatics Laboratories.
She received an M.E. and Ph.D. in engineering from Chuo University, Tokyo, in 2009 and 2011. She joined NTT in 2009, and her research interests include color vision compensation, life-log processing, and their applications for communication support. She received the Minister of Education, Culture, Sports, Science and Technology Award of the 23th Grand Prize for Frontier Technology: Pave a New Way for Creativity, and the Best Paper Award from the Institute of Electronics, Information and Communication Engineers (IEICE) in 2009. She is a member of IEICE, Human Interface Society, the Computer Software Association of Japan, and the Information Processing Society of Japan (IPSJ).

Motohiro Makiguchi
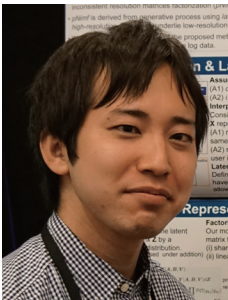Research Engineer, Symbiotic Intelligence Laboratory, NTT Human Informatics Laboratories.
He received an M.E. and Ph.D. in information science from Hokkaido University in 2012 and 2020. He joined NTT in 2012 and has been researching three-dimensional display technology and virtual-reality technology.

Masahiro Kohjima
Research Engineer, Symbiotic Intelligence Laboratory, NTT Human Informatics Laboratories.
He received a B.E., M.E., and Ph.D. in computer science from Tokyo Institute of Technology in 2009, 2012, and 2019. He joined NTT in 2012, and his research interests lie in the area of machine learning with emphasis on probabilistic models, Bayesian methods, and reinforcement learning. He is a member of IPSJ.

Masanori Yokoyama
Research Engineer, Symbiotic Intelligence Laboratory, NTT Human Informatics Laboratories.
He received an M.Sc. from Interdisciplinary Graduate School of Science and Engineering, Tokyo Institute of Technology, Kanagawa, in 2010. He joined NTT in 2010, and his research interests include nonverbal media, telepresence interface, and their applications for communication support. He received the 3rd Best Paper Award of ACM Augmented Human 2016, and Young Researcher Award of the HIS Symposium on Human Interface 2017. He is a member of IEICE and IPSJ.

Ryuji Yamamoto
Senior Research Engineer, Supervisor of Symbiotic Intelligence Laboratory, NTT Human Informatics Laboratories.
He received an M.E. from Kyushu Institute of Technology, Fukuoka, in 1998 and joined NTT the same year. His research interests include media processing and human-computer interaction.