OYaYa: A Desktop Robot Enabling Multimodal Interaction with Emotions

Yucheng Jin
jinyc2@lenovo.com
Lenovo Research
Beijing, China

Yuanyuan Deng
dengyy1@lenovo.com
Lenovo Research
Beijing, China

Jiangtao Gong
gongjt1@lenovo.com
Lenovo Research
Beijing, China

Xi Wan
wanxi2@lenovo.com
Lenovo Research
Beijing, China

Ge Gao
gaoge1@lenovo.com
Lenovo Research
Beijing, China

Qianying Wang
wangqya@lenovo.com
Lenovo Research
Beijing, China

Figure 1: Multimodal Interaction supported by OYaYa

ABSTRACT

We demonstrate a desktop robot OYaYa that imitates users' emotional facial expressions and helps users manage emotions. Multiple equipped sensors in OYaYa enable multimodal interaction; for example, it recognizes users' emotions from facial expressions and speeches. Besides, a dashboard illustrates how users interact with OYaYa and how their emotions change. We expect that OYaYa allows users to manage their emotions in a fun way.

CCS CONCEPTS

• **Human-centered computing** → **Personal digital assistants**; **Systems and tools for interaction design**.

KEYWORDS

small Robots, emotional interaction, multimodal interaction, emotion management

ACM Reference Format:
Yucheng Jin, Yuanyuan Deng, Jiangtao Gong, Xi Wan, Ge Gao, and Qianying Wang. 2021. OYaYa: A Desktop Robot Enabling Multimodal Interaction with Emotions. In *26th International Conference on Intelligent User Interfaces (IUI '21 Companion), April 14–17, 2021, College Station, TX, USA.* ACM, New York, NY, USA, 3 pages. https://doi.org/10.1145/3397482.3450729

IUI '21 Companion, April 14–17, 2021, College Station, TX, USA

ACM ISBN 978-1-4503-8018-8/21/04.
https://doi.org/10.1145/3397482.3450729

1 INTRODUCTION

The accurate expression and effective regulation of emotion in oneself is an essential part of one's emotional intelligence [12]. Especially in the workplace, managing emotions efficiently has always been an important research topic for psychologists and sociologists [1]. Several existing studies [8, 13] have investigated how people regulate and manage their emotions using computer technologies such as robots. In addition, several early works [3, 10] have investigated the techniques that enable emotional reaction with the visual and audio modes in human-robot interaction (HRI).

On the one hand, people want to get emotional feedback and experience while interacting with the robots [6, 7]. On the other hand, people expect computer technologies to help them better understand, express, and manage emotions [4, 8, 9].

This paper demonstrates a desktop robot that enables people to express emotions through facial expressions, verbal commands, and physical gestures. Toward each recognized emotion, OYaYa provides emotional feedback by showing animated emoticons and speaking our designed language to resonate with people. Furthermore, we developed a dashboard to illustrate how a user can view her emotional changes in a period. By demonstrating the above features, OYaYa aims to provide users with **a fun way to interact and monitor their emotions**.

2 OYAYA

We design OYaYa as a desktop robot with a size of 11*11*17cm (see the left side of Figure 2). It supports multimodal interaction with users' emotions. OYaYa automatically records a user's emotions and generates an emotion diary, and we visualize all the data through

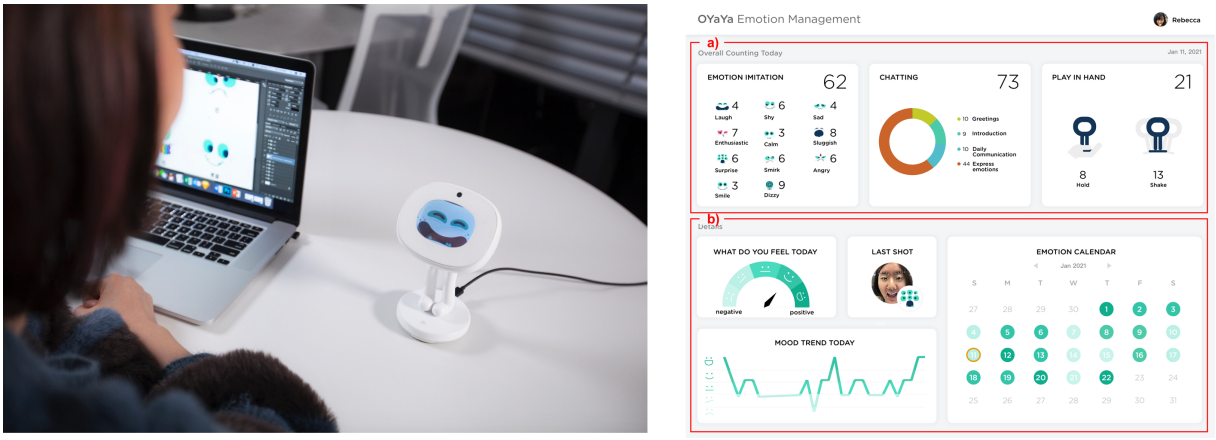

Figure 2: A scenario of using OYaYa in the office (left). A dashboard showing emotion data (right)

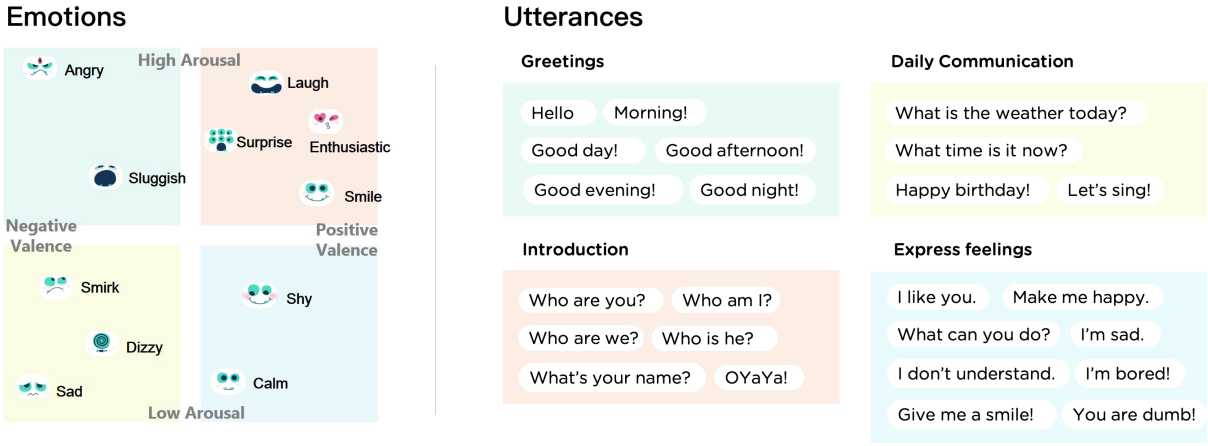

Figure 3: Emotions and utterances that OYaYa can recognize and imitate

a web-based dashboard. By following the proactive emotion regulation strategy, the dashboard visualizes the times of using each modality (the facial, the vocal, and the physical) to express emotions in a day (see the right side of Figure 2 b) and shows the detailed view of a user's emotions (see the right side of Figure 2 b) such as today's emotions, last shot, daily emotion trend, and emotion calendar.

2.1 Facial Interaction

Previous studies [5, 11] have shown the close relationship between emotional mimicry and empathy. We intend to create empathy by asking OYaYa to imitate the emotional facial expressions of users. E.g., when a user shows a smile, in return, OYaYa will play a smiling emoticon. According to the arousal-valence model, in total, we designed 11 animated emoticons to represent 11 expressions (see the Emotions in Figure 3). However, it seems that only in-group members are more likely to mimic negative emotions [2].

2.2 Vocal Interaction

Additionally, a user can activate voice recognition by saying "Hi OYaYa". We have defined four types of utterances, "Greetings", "Introduction","Daily communication", and "Feelings" (see Figure 3), which can be understood by OYaYa. To increase enjoyment, we design a new language for OYaYa that combines several languages' pronunciation and intonation, such as Mandarin, Korean, Shanghainese, Japanese, and Spanish. OYaYa speaks the new language and plays an animated emoticon when it recognizes users' intents.

2.3 Physical Interaction

Besides, we defined several physical gestures to trigger some animated emoticons and vocal responses that match the gestures' meaning. For example, OYaYa shows an animation of feeling dizzy when shaking or rotating its body and an animation of being angry when patting its head (see the right side of Figure 1).

3 HARDWARE

As shown in Figure 5, the hardware of OYaYa consists of six key components: a) 5-megapixel Zero Spy Camera (8.5mm x 11.3mm),

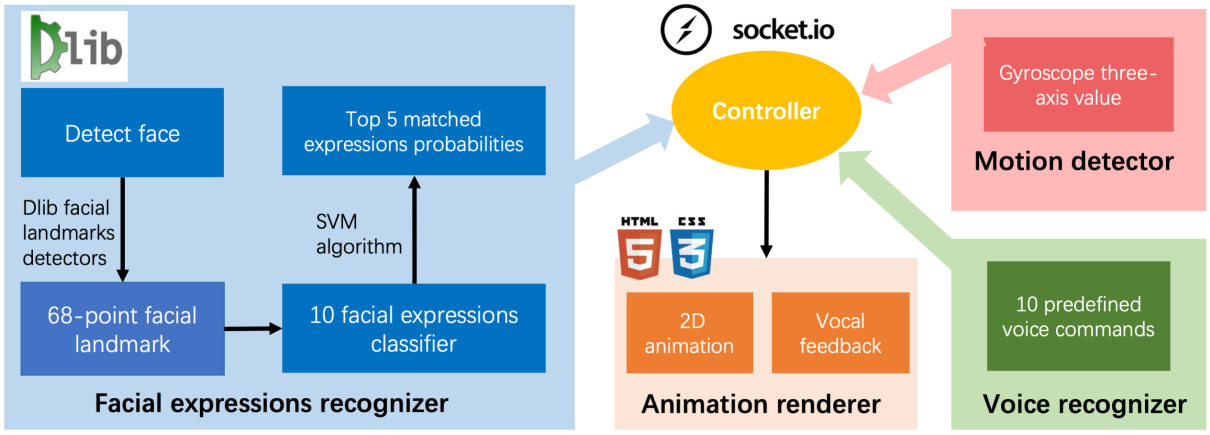

Figure 4: Software architecture of OYaYa.

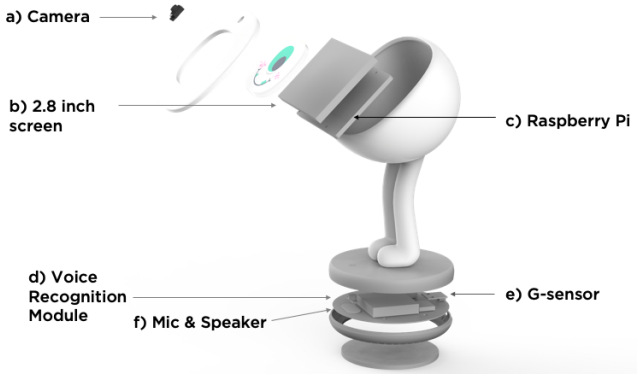

Figure 5: Hardware components of OYaYa.

b) a 2.8-inch screen (320 x 480), c) a Raspberry Pi 3B+ (1.4 GHz), d) a WEGASUN-M6 voice recognition module (VRM), e) a G-sensor, and f) Mic & Speaker. A 5V battery powers OYaYa via a USB2.0 cable.

4 SOFTWARE

We developed the software of OYaYa based on a client-server architecture (see Figure 4). We implemented a controller to process the data collected from a camera, VRM, and G-sensor on the server-side. In specific, we applied a widely used facial landmark detection library Dlib [1] to extract 68 key points of facial landmarks. We recruited 40 users to make all 11 facial expressions five times, and then we trained a classifier for 11 facial expressions using the SVM algorithm. The accuracy is 82.3%. The VRM comes with an application that allows us to import our defined utterances that OYaYa can recognize. By reading the values of acceleration on the three-axis, the controller can recognize the physical gestures. The real-time communication between server and client relies on the socket.io [2]. We render the animated emotions of OYaYa in a web browser.

[1] http://dlib.net/face_recognition.py.html
[2] https://socket.io/

5 AUDIENCE AND RELEVANCE

The demonstration is suitable for any audience: ordinary office workers, students, researchers. We want to discuss the emotional mimicry by desktop robots for managing users' emotions. By demonstrating OYaYa, we would like to know how multimodal interaction can facilitate emotional mimicry to improve negative emotions such as stress, depression, and anger. By trying the system first-hand, attendees will reflect on the topic and experience a relaxation session. We expect OYaYa to foster discussions about issues related to affective computing applications, such as an AI companion.

REFERENCES

[1] Sharon C Bolton. 2004. *Emotion management in the workplace.* Macmillan International Higher Education.
[2] Patrick Bourgeois and Ursula Hess. 2008. The impact of social context on mimicry. *Biological psychology* 77, 3 (2008), 343–352.
[3] Cynthia Breazeal and Brian Scassellati. 1998. Infant-like social interactions between a robot and a human caretaker. *Adaptive Behavior* 8, 1 (1998).
[4] Mariam Hassib, Max Pfeiffer, Stefan Schneegass, Michael Rohs, and Florian Alt. 2017. Emotion actuator: Embodied emotional feedback through electroencephalography and electrical muscle stimulation. In *Proc. of CHI 2017.* ACM, 6133–6146.
[5] Ursula Hess and Agneta Fischer. 2014. Emotional mimicry: Why and when we mimic emotions. *Social and Personality Psychology Compass* 8, 2 (2014), 45–57.
[6] Sooyeon Jeong, Cynthia Breazeal, Deirdre Logan, and Peter Weinstock. 2018. Huggable: The Impact of Embodiment on Promoting Socio-emotional Interactions for Young Pediatric Inpatients. In *Proc. of CHI 2018.* ACM, 495.
[7] Hajin Lim, Dan Cosley, and Susan R Fussell. 2018. Beyond Translation: Design and Evaluation of an Emotional and Contextual Knowledge Interface for Foreign Language Social Media Posts. In *Proc. of CHI 2018.* ACM, 217.
[8] Zhe Liu, Anbang Xu, Yufan Guo, Jalal U Mahmud, Haibin Liu, and Rama Akkiraju. 2018. Seemo: A Computational Approach to See Emotions. In *Proc. of CHI 2018.* ACM, 364.
[9] Adam Lobel, Marientina Gotsis, Erin Reynolds, Michael Annetta, Rutger CME Engels, and Isabela Granic. 2016. Designing and utilizing biofeedback games for emotion regulation: The case of nevermind. In *Proc. of CHI 2016.* ACM, 1945–1951.
[10] Luis-Alberto Perez-Gaspar, Santiago-Omar Caballero-Morales, and Felipe Trujillo-Romero. 2016. Multimodal emotion recognition with evolutionary computation for human-robot interaction. *Expert Systems with Applications* 66 (2016), 42–61.
[11] Eliska Prochazkova and Mariska E Kret. 2017. Connecting minds and sharing emotions through mimicry: A neurocognitive model of emotional contagion. *Neuroscience & Biobehavioral Reviews* 80 (2017), 99–114.
[12] Peter Salovey and John D Mayer. 1990. Emotional intelligence. *Imagination, cognition and personality* 9, 3 (1990), 185–211.
[13] Minna Vuorela and Lauri Nummenmaa. 2004. Experienced emotions, emotion regulation and student activity in a web-based learning environment. *European Journal of Psychology of Education* 19, 4 (2004), 423–436.