ToonWorld: A Mid-Air Medium to Playfully Augment Human Expression

Aditya Retnanto*
aretnanto@uchicago.edu
University of Chicago
Chicago, Il, USA

Ramarko Bhattacharya**
ramarkob@uchicago.edu
University of Chicago
Chicago, Il, USA

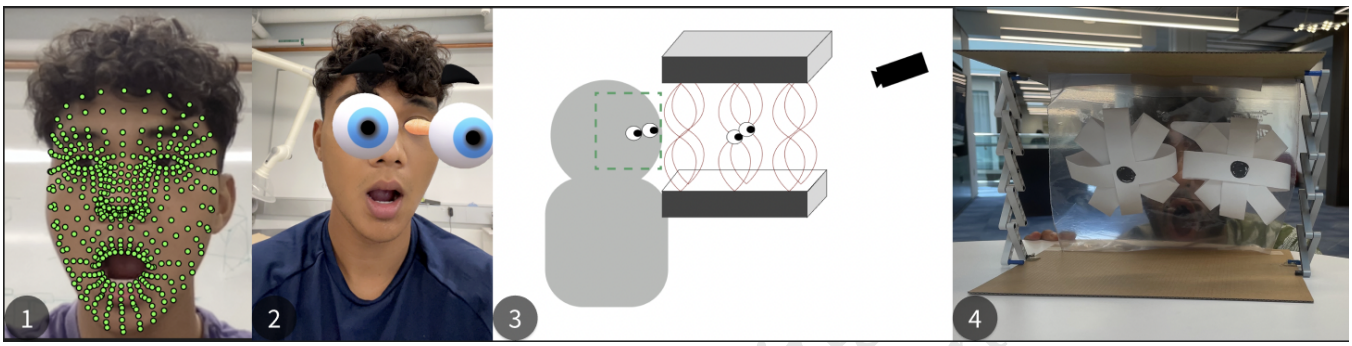

Figure 1: (1) Facial landmark detection using MediaPipe, (2) Snapchat representation of shocked user, (3) Interaction diagram of shocked user and creation of levitating beads, (4) Mock-up of face architecture consisting of beads in between the transducers

ABSTRACT

We propose ToonWorld, a mid-air medium to tangibly augment facial expression through cartoon-like reactions. Using facial expressions as an input modality, we construct dynamically arranged levitating primitives such that they appear to extend from the user. The exaggerated reactions from a user would provide clarity of a user's intent while providing awareness from others.

CCS CONCEPTS

• **Human-centered computing** → **Interaction devices**; **Interaction paradigms**; **Human computer interaction (HCI)**.

KEYWORDS

Augmented Facial Expressions, Shape-Changing Interfaces, Tangible User Interfaces

ACM Reference Format:
Aditya Retnanto and Ramarko Bhattacharya. 2022. ToonWorld: A Mid-Air Medium to Playfully Augment Human Expression. In *Proceedings of UIST '22.* ACM, New York, NY, USA, 3 pages. https://doi.org/XXXXXXX.XXXXXXX

*Both authors contributed equally to this research.

UIST '22, Oct 29-Nov 02 2022, Bend, OR

ACM ISBN 978-x-xxxx-xxxx-x/YY/MM…$15.00
https://doi.org/XXXXXXX.XXXXXXX

1 INTRODUCTION

Devoid of verbal communication, slapstick comedians exaggerate their reactions so viewers can easily process their intent. From live-action TV series such as Mr. Bean to cartoons such as One Punch Man, viewers can easily process the expressiveness of the characters through the tangible interplay in the physicality of reactions. While prominently used digitally in applications such as Snapchat and Zoom, filters are a post-processed pixel representation of ourselves. We view facial expressions as an output modality and speculate a world in which we can all express amplified cartoon-like emotions. An on-air medium proximal to a user's face creates a space that can embody physical extensions to a user. These tangible bits [3] represent an extension of self and enhance facial expression capabilities. For the UIST Student Innovation Contest, we playfully explore these concepts by creating a photo booth of physical filters such that embodied extensions are an extension of self.

2 RELATED WORKS

Technology Mediated Human Augmentation. There are visions in the HCI community to blur the physical and the digital through augmenting humans' physical, intellectual and social capabilities [7]. On-face output augmentations have primarily focused on input sensory capabilities such as sight [4], yet research on augmenting expression is sparse. One recent approach is the wearable Morphace[9]. Inspired by prosthetic make-up, silicone patches are placed on the face with the potential to change color, pattern, medium, and shape. Our approach utilizes the levitating medium to extend our facial output modalities through playful locomotion and embodiment of the familiar yet distinct.

AR/VR Facial Expression. Major commercial camera-facing applications involve the use of filters to create playful outputs of a face.

Rios et al. explore user selection on Snapchat filters where selection tends to skew towards filters that make one more attractive or look funny [8]. In the digital, there is a clear separation between the user and the filters. By physicalizing the filters, users can seamlessly extend their emotive spectrum of expressivity and allow others to ease the cognitive processing of the embodied emotions.

3 INTERACTION

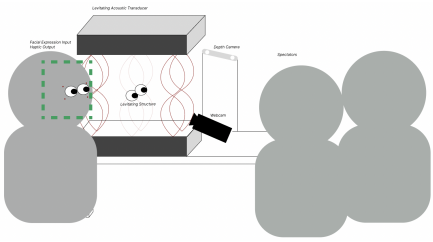

Figure 2: System diagram and general set up of interaction

Figure 2 introduces the general setup of the system. In this section, we explain the interaction concepts for our proposal.

UIST Demo. The aim of the live demo is to showcase the playful side of the exaggerating expressions. The conference should be a joyous occasion, and one common interaction found in these events is photo booths. Upon entry to our booths, participants can select a physical augmentation(akin to props), dynamically interact with it and get a photo souvenir of their experience. A sample input/output combination of props is shown in the figure 3. To 'feel'

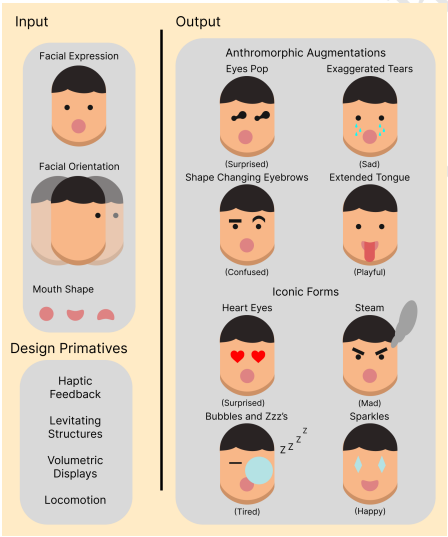

Figure 3: System input, output ,and animation properties

the physical extensions, users will get haptic feedback proximal to the specific body part. While it may deter from the experience, for safety concerns, users will place their head on a chin-strap and haptic feedback will be directed to non-sensitive parts of the face such as the cheek.

Interaction Mode. Ideally, a closed-loop system would be developed where user's can switch between expression. Due to the constraints of the competition we will develop a filter switching mode where users can manually switch between filters through a GUI input.

User Perspective. The mid-air tangible anthropomorphic augmentations are proximal to the user's face such that they perceive it to be an extension of self. In an ideal interaction we aim to make the user feel as if these parts are leaving their body through haptic feedback and feel a sense of ownership for these props akin to the Rubber Hand experiment [1]. While iconic forms may not provide the embodied affordances of the anthropomorphic, we believe that they still ease the cognitive loading for spectators to understand the user's intent.

4 TECHNICAL IMPLEMENTATION

Expanding on the prior section, we detail the technical implementation behind figure 2.

Hardware. Ultrasound acoustic transducers will be top-down oriented to optimize levitation and serve as an external actuator for objects within the space [2]. We borrow beads, threads, and cloths terminology as building blocks for the primitive building blocks from Fender [2]. To demonstrate feasibility, we discuss in detail the build of a complex structure such as an eye. The explored approach is to stack 4 clothes forming a combination of a '+'. A bead will be placed in the center and 4 beads will be placed on the edges of the '+'. When the 4 beads are in the same position, the resulting shape will be a sphere thus meeting Fender's limitation of 12 concurrent levitating beads [2]. Additionally, a depth camera will be used to track the bead position and a webcam for the user's face.

Software. A camera will be used to capture the user's facial expression using the MediaPipe FaceMesh model [5]. We will train a machine learning model to classify the gestures in our design space. During the interaction, the scene of facial landmarks will be classified in real-time and output a string should a facial gesture be detected. The Unity API receives the classification through serial communication and moves the particles to the desired locations.

5 FUTURE WORK AND IMPLICATION

We envision future works for our project into two possible categories.

Live Theatre. Rather than constraining actors to biological capabilities, the air-space of the particles levitating can enable them to have greater control and amplitude of their emotions. Spectators from a distance and odd angles can better view the show to immerse themselves in the experience.

Communication. 55% of communication is done non-verbally [6]. By amplifying and expanding our capability's physical expressions, we envision our interaction as a medium for people to communicate cross-culturally. If we can interact with the levitating props tangibly, we believe this may open up more playful back and forths in communication.

REFERENCES

[1] H. Henrik Ehrsson, Nicholas P. Holmes, and Richard E. Passingham. 2005. Touching a Rubber Hand: Feeling of Body Ownership Is Associated with Activity in Multisensory Brain Areas. *The Journal of Neuroscience* 25, 45 (11 2005), 10564.

[2] Andreas Rene Fender, Diego Martinez Plasencia, and Sriram Subramanian. 2021. ArticuLev: An Integrated Self-Assembly Pipeline for Articulated Multi-Bead Levitation Primitives. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (Yokohama, Japan) *(CHI '21)*. Association for Computing Machinery, New York, NY, USA, Article 422, 12 pages. https://doi.org/10.1145/3411764.3445342

[3] Hiroshi Ishii and Brygg Ullmer. 1997. Tangible Bits: Towards Seamless Interfaces between People, Bits and Atoms. In *Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems* (Atlanta, Georgia, USA) *(CHI '97)*. Association for Computing Machinery, New York, NY, USA, 234–241. https://doi.org/10.1145/258549.258715

[4] Lik-Hang Lee and Pan Hui. 2017. Interaction Methods for Smart Glasses. https://doi.org/10.48550/ARXIV.1707.09728

[5] Camillo Lugaresi, Jiuqiang Tang, Hadon Nash, Chris McClanahan, Esha Uboweja, Michael Hays, Fan Zhang, Chuo-Ling Chang, Ming Guang Yong, Juhyun Lee, Wan-Teh Chang, Wei Hua, Manfred Georg, and Matthias Grundmann. 2019. MediaPipe: A Framework for Building Perception Pipelines. https://doi.org/10.48550/ARXIV.1906.08172

[6] Albert Mehrabian and Suzanne Ferris. 1967. Inference of attitudes from nonverbal communication in two channels. *Journal of consulting psychology* 31 3 (1967), 248–52.

[7] Jun Rekimoto. 2013. From augmented reality to augmented human. In *2013 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. 1–1. https://doi.org/10.1109/ISMAR.2013.6671755

[8] Juan Sebastian Rios, Daniel John Ketterer, and Donghee Yvette Wohn. 2018. How Users Choose a Face Lens on Snapchat. In *Companion of the 2018 ACM Conference on Computer Supported Cooperative Work and Social Computing* (Jersey City, NJ, USA) *(CSCW '18)*. Association for Computing Machinery, New York, NY, USA, 321–324. https://doi.org/10.1145/3272973.3274087

[9] Sijia Wang, Cathy Mengying Fang, Yiyao Yang, Kexin Lu, Maria Vlachostergiou, and Lining Yao. 2022. Morphace: An Integrated Approach for Designing Customizable and Transformative Facial Prosthetic Makeup. In *Augmented Humans 2022* (Kashiwa, Chiba, Japan) *(AHs 2022)*. Association for Computing Machinery, New York, NY, USA, 58–67. https://doi.org/10.1145/3519391.3519406