Eliminating Appearance-Based Discrimination in Communication: How can ML-based Emotion Classifier Facilitate Emotions Transfer in Verbal Communication

Project:
Technology Facilitated Transfer of Emotions and Expressions

Spervisor: Eugene Ch’ng

Members: Danzhao Cheng Yi Song Yifei Yue

Date: 2021-9-7

Acknowledgement

We would like to express the great appreciation to Professor Eugene Ch'ng for his helpful and constructive advice during this research work. Suggestions given by mentors in *Shaping the Connected Museum* have also been a great help. Besides, we are particularly grateful to the assistance offered by our companions in the NVIDIA Joint-lab in UNNC.

Table of Contents

List of Illustrations

1. Abstract

The project explores how machine learning can be used to identify, recognise and transmit facial expressions in relation to verbal communication between users of a remote scenario that maps expressions to simpler, graphical representations of facial features, i.e., animated emojis, simplified virtual avatars. The virtual representation may be in 3D and posited within the virtual environment, augmented space, or as 2D graphical representations on the user interface. The research aims to understand how technology can be used to facilitate the transmission of emotions via machine learning facial expression recognition.

2. Introduction

Verbal communication is essential within our society as it is the main mode for the transfer of information, for sharing cultural identity, imparting knowledge, and reinforcing social bonds. It has been proven that audio storytellings are more cognitively and emotionally engaging at a physiological level, compared to the visual media (Richardson et al., 2020). This may be because listening to a story, rather than watching a video, is a more active process of co-creation. Besides, verbal communication is more efficient than textual communication for voice can contain more emotions than plain text. In other words, compared with other communication methods, verbal communication is worth researching since it can provide more space for imagination, while evoking higher levels of feelings simultaneously. To contribute to this, we believe that verbal communication assisted with emotion transfer techniques may help to convey people's emotions without exposing information related to one's physical appearance.

Regarding the measurement of emotions, relevant research is fraught with methodological limitations, as feeling can have non-discrete, ephemeral, and ineffable qualities. However, there are seven universal emotions (*Universal Emotions*, 2021) that have been identified and well-adopted in multiple research works (Ekman, 1992). The universal emotions include anger, contempt, disgust, enjoyment, fear, sadness, and surprise, which can be used to clarify, or even, quantify people's different emotional states. Besides, with the development of social platforms, emojis are becoming one of the most well-acceptable ways to express emotions, especially in the virtual social communities nowadays. Hence, emotions can be translated into the corresponding emojis (Seconds, 2021). Moreover, the identification of abstract emotions can be assisted by machine learning, which can recognise, and transmit facial expressions into pre-defined emotional classes (Lee et al., 2020). Thus we are wondering whether or how we can utilise machine learning to automatically capture, translate, and transfer people's emotional states elicited in remote verbal communication.

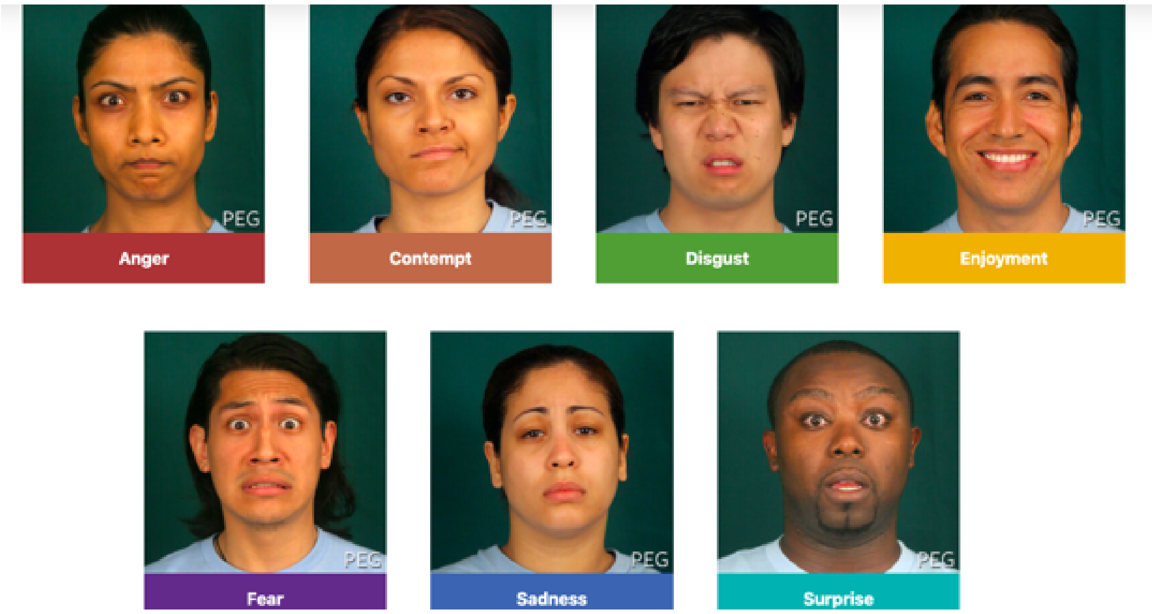

Figure 1. Seven Universal Emotions

To achieve that, we conduct a series of experiments. First, we selected a Chinese folk story called "The Nine-coloured Deer" containing different emotional states, which is suitable to evaluate whether emotions can be better transferred during remote verbal communication via computer-mediated technologies. Next, we recruit participants to read or listen to this story, and examine their emotional changes, through self-report questionnaires and behavioural data. Through those experiments, this project aims to answer the following questions:

1. How are emotional contents understood if verbal communication is accompanied by simplified machine-facilitated expressions (i.e. emoji and virtual avatar)?
2. Does simplified machine-facilitated expressions enhance the accuracy of emotions communicated via voice, emoji, and virtual avatar?
3. How acceptable are machine facilitated expressions towards recipients of emotional messages in terms of their appropriateness and accuracy?

3. Methodology

3.1 Technical Development

In order to test the hypothesis, two systems were built to recognise the facial expression in real-time. One of the systems was implemented mainly by Tensorflow, Keras and Opencv. Opencv is used to open the camera and capture every image/

frame of the user. Tensorflow and Keras were used to analyze the images and predict facial expression. In the meanwhile, the predicted facial expression was replaced by emoji and shown. In addition, the expressions that are constantly predicted were also written into a csv file in chronological order for further analysis. The other system was implemented by Dynamixyz Live Link Plugin and Unity Engine. The Dynamixyz Live Link Plugin provides realtime facial animation for characters and tracks the face of the user in real-time. The Unity Engine was used to build a virtual human model. Then the virtual human can simulate the facial expression of the user in real-time by tracking a face live with Dynamixyz and stream the animation Live in Unity Engine.

Technology identified:

- Physical Hardware
    - Logitech HD 1080p
    - MacBook Pro
    - Lenovo L14
- Software
    - Visual Studio Code
    - Tensorflow
    - Keras
    - Opencv
    - Dynamixyz Live Link Plugin
    - Unity Engine

3.2 Experiment Preparation

We chose a folk story called “The Nine-coloured Deer” to conduct the experiment. The story is based on an ancient painting discovered in cave paintings of the Mogao Caves, which is one of the most famous cultural heritages in China. According to the sequence of the painting, scholars have divided the story into 7 different scenarios, each of which contains significant plot transitions (Ma & Ao, 2008). The nature of this story suits well for our experiment goals since participants can go through different emotional states during the process of reading, listening, or communicating the story.

3.2.1 Story Scenario Division:

Scenario 1: A man called Tiaoda fell into the water. A nine-coloured deer saved his life (enjoyment, fear).

Scenario 2: Tiaoda promised never to tell others where the deer lives (enjoyment, surprise).

Scenario 3: The Queen dreamed of the nine-coloured deer. So the King made an announcement to the country that whoever captured the deer would be rewarded with a big prize (disgust, enjoyment).

Scenario 4: Tiaoda gave in to his greed and came to meet the King (disgust, anger).

Scenario 5: Tiaoda led the King and his soldiers to the place where he saw the deer (sad, fear).

Scenario 6: At that time, the nine-coloured deer was sleeping and didn't know anything (sad, fear).

Scenario 7: Feeling betrayed, the nine-coloured deer told the King the whole story about her saving the man and him breaking the promise (sad, anger, enjoyment).

3.2.2 Identify Emotional States within the Story:

In order to identify the emotions contained in the story objectively, we first recruited a group of volunteers (10 people) to read the story by themselves while highlighting the texts accordingly using seven different colors to indicate seven different emotions (anger, contempt, disgust, enjoyment, fear, sadness, and surprise). After they finished reading, they were asked to fill a questionnaire, which records their self-reports of the emotional change regarding this story, as well as tests that can reflect their comprehensibility of the story content. Next, we compiled all the data and synthesized a final version of the story, which contains sentences associated with different emotional state changes identified using seven universal emotions. It is used to lay a fundamental ground for further measurement and comparison.

3.2.3 Record Storytelling:

Then, we invited a volunteer to be the Reader, who read the story with annotated emotions expressively. We recorded the original video of the Reader's voice and facial expressions during the storytelling. The original video was later classified and translated into corresponding emojis using our machine learning model.

3.3 Experiment Design

For the experiment design, we plan to recruit 3 groups, each of which contains 20 people. For group 2 and 3, the story will be read by a reader, so the communication between the reader and the participant will be developed. The faces of both participants are not seen by each other, only the voice and the emoji or virtual avatar will be shown.

According to the two-factor theory of emotions, an emotion is caused not only by the physical responses of people, but also by their mental interpretation of the body's responses (Nakayama, 2015). As a result, participants' behavioural data, as the measurement of physical responses, and self-report questionnaires, as the measurement of their mental interpretation, will be recorded and evaluated.

Groups:

Test group 1: Listen to the story

Test group 2: Listen to the story while watching the reader's corresponding emotions, which are pre-recorded and transferred to emojis.

Test group 3: Listen to the story while watching a virtual avatar, which is a simplified graphical representation of the reader's facial expressions.

Data Recorded:

For each participants in the groups listed above, we record data as follows:

a. The original video recording facial expressions while listening to the story
b. The changes in emotion (averaged emotions resulting from a time interval $t+t_n$) classified using the machine learning model
c. The physiological measurement using heart rate monitors (indicator serves as an indirect measure of cognitive and emotional engagement - perception vs actual)
d. The self-reported questionnaire rated by the participants
e. The test used to measure their comprehensibility of the content

3.4 Data Processing

As mentioned in 3.1, the expressions that are constantly predicted were written into a csv file in chronological order. Therefore, we read the data from the csv file and plot a line chart for experiment evaluation. In the line chart, we use each sentence of the story as the x-axis, and 7 different emotions as the y-axis. Then we compare each emotion line of the participant with the standard emotion line.

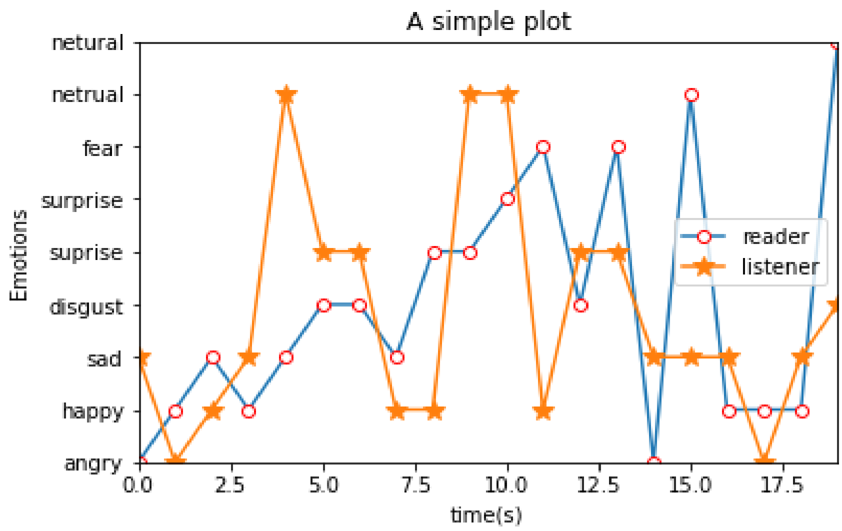

Figure. 2 Results Visualisation

4. Findings

Pre-test:

This adequate size of 60 participants is difficult to recruit during the summer vacation. In consequence, we did a pre-test on control group 1 with 10 people. In addition to recording the required datasets, participants are asked to highlight and label particular sentences of the story that evoke their different emotions. The pre-test aims to examine the effectiveness of the chosen story and provide reference information for future formal experiments.

In the self-report questionnaire, we set three questions that have fixed correct answers, which get an accuracy rate of 93.3%. Anger and enjoyment gain the most percentage of agreement of the emotions of the story at 40%.

Besides, we analysed the sentences with labeled emotions highlighted by participants. Most of the participants marked similar parts of the text with the same kind of emotions, which is adopted as the “standard emotion” of the story. This “standard emotion” will be performed by readers when they read the nine-coloured deer story in experiment 2 and experiment 3. Hence, listeners in control group 2 and control group 3 can listen to the story while watching the emojis of readers corresponding to the emotions of the story.

5. Discussion

Regarding the mixed results of the pre-test in control group 1, we found that the nine-coloured deer story is comprehensible and can evoke different sorts of emotions in readers. The majority of participants agree that the story is easy to understand, rather than being ambiguous. The high accuracy rate of objective multiple-choice questions also proves this result. Besides, participants report that they are deeply involved in the story and their feelings change significantly corresponding to the plot transition of stories. Thus, the nine-coloured deer story is suitable to be used in the following formal experiments.

However, limitations are also shown in the results of the questionnaire. Firstly, few people say that they find the story stimulating though they are infected by the emotions contained in the story. Secondly, the majority of people claim that they would not like to repeat the experience again. It means that the traditional way of communication, such as reading the text-based story, has become a boring experience for modern people. In consequence, after conducting formal experiments on control group 2 and control group 3, the results can help us to find out whether communicating with others with corresponding emojis or avatars can better convey the emotional contents and increase the degree of social presence.

Another difficulty that we encounter is for the reader to match emotions and texts while reading the story. Based on the highlighted sentences with emotions by participants in the pre-test, the reader can read the story with "standard emotion" to listeners. However, performing certain facial features to convey the corresponding emojis of the sentences to listeners is a relevant hard task. Our current solution is to slow down the frame rate, so the machine can capture the emotions expressed by readers more accurately.

Also, the accuracy of the trained model is only 66%, so we changed the calculation method of model accuracy. We found a new dataset --- CK+ dataset, and extracted 6000 labeled pictures with various expressions. Finally, the 6000 pictures were predicted with the model, and the final accuracy reached 93.28%.

6. Conclusion

In conclusion, verbal communication is of great importance as it provides more space for imagination than video chatting and contains more emotions than texting. Our project creates a machine to transfer and transmit the facial emotional expression to graphical representations of facial features, such as emojis and avatars, by using machine learning. A pre-test has been done to provide reference information to the formal experiments.

At this stage, the research is a fundamental experiment to test the effectiveness, accuracy, and appropriateness of technology facilitated transfer of emotions and expressions. After being fully developed, this machine can be used to eliminate bias or discrimination in the museum context. Sociologists (DiMaggio, 1996) argue that aesthetic taste and participation in the arts represent a form of cultural capital that is related to social distinction. Therefore, as the machine can hide information about one's physical appearance or cultural and racial background, thus reducing relevant discrimination, it shows the possibility of making the museum a more democratic public place.

7. References

DiMaggio, P. (1996). Are art-museum visitors different from other people? The relationship between attendance and social and political attitudes in the United States. Poetics, 24(2–4), 161-180. https://doi.org/10.1016/S0304-422X(96)00008-3.

Ekman, P. (1992) An argument for basic emotions. *Cognition and Emotion, 6*(3-4), 169-200. 10.1080/02699939208411068

Lee, J. R. H., Wang, L., & Wong, A. (2020). EmotionNet Nano: An Efficient Deep Convolutional Neural Network Design for Real-time Facial Expression Recognition. Computer Vision and Pattern Recognition. https://arxiv.org/abs/2006.15759

Ma, Q., & Ao, D. (2008). The Narrative Structure Analysis of Nine-Coloured Deer. J. Northern University for Nationalities (Philosophy and Social Science), 4. pp. 79-85.

Nakayama, K., Oshima, C., Higashihara, R., & Machishima, K. (2015). "Mood Induction through emotional prosody modification — Experiments of students reading a folk story scenario," *2015 54th Annual Conference of the Society of Instrument and Control Engineers of Japan (SICE)*. pp. 391-396, doi: 10.1109/SICE.2015.7285406.

Universal Emotions | What are Emotions? (2021, February 16). Paul Ekman Group. https://www.paulekman.com/universal-emotions/

Richardson, D. C., Griffin, N. K., Zaki, L., Stephenson, A., Yan, J., Curry, T., Noble, R., Hogan, J., Skipper, J. I., & Devlin, J. T. (2020). Engagement in video and audio narratives: contrasting self-report and physiological measures. Scientific Reports, 10(1). https://doi.org/10.1038/s41598-020-68253-2

Seconds, S. (2021, April 19). *Plutchik's Wheel of Emotions: Feelings Wheel.* Six Seconds. https://www.6seconds.org/2020/08/11/plutchik-wheel-emotions/