Chatbot with Touch and Graphics: An Interaction of Users for Emotional Expression and Turn-taking

Hyeonggeun Yun
Auejin Ham
DGIST
Daegu, South Korea
{yhg8423,ham}@dgist.ac.kr

Jin Kim
Taeyeong Kim
DGIST
Daegu, South Korea
{kjin0830,kt9393}@dgist.ac.kr

Jeongeun Kim
Haechan Lee
DGIST
Daegu, South Korea
{wjddms0114,gocks8}@dgist.ac.kr

Jongrae Park
DGIST
Daegu, South Korea
j-park@dgist.ac.kr

Jinkyu Jang
Yonsei University
Seoul, South Korea
alan.jk.jang@gmail.com

Author Keywords
Chatbot; Emotional expression; Turn-taking; Touch interaction; Graphic interface.

CHI 2020 Extended Abstracts, April 25–30, 2020, Honolulu, HI, USA.

ACM ISBN 978-1-4503-6819-3/20/04.
DOI: https://doi.org/10.1145/3334480.XXXXXXX
*update the above block & DOI per your rightsreview confirmation (provided after acceptance)

Abstract
Use of chatbots for emotional exchange is recently increasing in various domains. However, as existing chatbots have been considered in terms of natural language processing techniques for interaction with text-based chatting, chatbot interaction with users is lacking in terms of considering the emotions of users and managing turn-taking in conversation. This paper suggests an interaction technique having touch interactions with graphic interfaces (TwG) to solve these problems. In the system, users send their emotions and manage turn-taking through TwG technique. We conducted a Wizard of Oz study to evaluate user experience on emotional expression and turn-taking with TwG technique. Results showed that TwG interaction improved emotional expression compared to a traditional text-based chatbot interaction. Furthermore, the results showed that TwG positively affects natural turn-taking of the conversation.

Introduction
Chatbots, also referred to as conversational agents, are used in various fields. For example, there are chatbots for customer care service [21] and teaching knowledge [18]. The goal of interaction with these chatbots primarily focuses on performing tasks, answering questions, and achieving specific purposes [3].

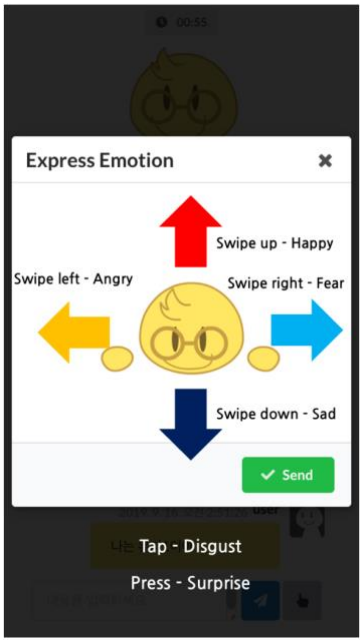

Figure 1: Mapping result between emotions and touch gestures

Figure 2: Facial expressions of character: (a) neutral, (b) happy, (c) angry, (d) fear, (e) sad, (f) disgust, (g) surprise

Accordingly, chatbots try to interact with users in one-and-one interaction that requires pairs of one question and one answer.

Particularly, there are some problems in the interaction between humans and chatbots for emotional exchange, because the context and the flow of conversation between them are different from those between humans. In conversation between humans and chatbots, chatbots struggle to respond to humans as appropriate considering the emotions and characteristics of others [20]. Moreover, people usually send a few short sentences, chunks of one topic, to others in text-based chatting [8], so chatbots are unable to understand the flow of message. Subsequently, there are problems related to understanding emotions and turn-taking (i.e., method to perceive cues for turn transition in the conversation) [4]. To solve these problems, most researchers attempt to apply natural language processing (NLP) techniques to chatbot conversation [1, 11]. However, it is hard to solve the problems with using only NLP techniques because, they may not achieve 100% accuracy and it is hard to understand common sense knowledge of humans.

In this research, we investigate the user experience on context and flow of conversation in human–chatbot interaction with interaction techniques for solving the above problems. We consider human–chatbot conversation in terms of interaction like in human–human conversation with human behavior aspects. Our basic motivation is that chat using text-based interaction giving rise to the above problems makes it difficult for a chatbot to recognize context and nonverbal cues [19, 20]. Humans interact with other people naturally through emoticons or non-verbal cues [5] and these interactions help people to understand the emotions of others and promote conversation between people [5, 12].

Therefore, we suggest touch interaction and graphic interface to address the problems. Touch interaction can express emotion [9] and improve the quality of communication [16]. It is a familiar mode of interaction with people who use mobile devices. A graphic interface can express and analyze the emotions of users with visual cues such as emoticons and color [15]. It can show nonverbal cues of users in online communication [13]. Thus, we expect that touch interaction and graphic interface complement the limitations of the text interaction for emotional expression and turn-taking.

Design and Overview of the System

To manage emotional expression and turn-taking with touch interactions and graphic interfaces, we propose an interaction technique named Touch interactions with Graphic interfaces (TwG) for human–chatbot communication. In TwG with a chatbot, users can talk with the chatbot using text interaction, touch interaction, and a graphic interface. In the conversation, text interaction performs the role of verbal cues, and touch interaction with the graphic interface performs the role of nonverbal cues, such as emotional expression and turn-taking. Furthermore, users can get visual feedback about their touch interaction from the graphic interface. To implement TwG technique, we develop a mobile web-based chatting system. In the system, people can talk and interact with a chatbot in various ways.

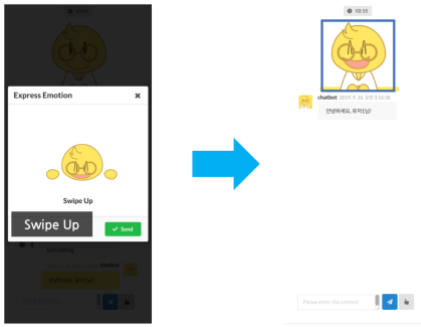

Figure 3: Change of facial expression in accordance with the emotion 'happy' posture

Figure 4: Postures of character: (a) Listening posture, (b) Speaking posture

TwG for emotional expression

To express emotion via TwG, we set up emotions based on Ekman's six emotions [7]. In Ekman's six emotions, there are happy, sad, disgust, fear, surprise, and angry. Consequently, we use Ekman's six emotions and neutral emotion for emotional expression. We map these emotions on touch interactions and graphic interfaces. For emotional expression, we use three single-touch interactions: swipe, tap, and press. We match happy with 'swipe up', angry with 'swipe left', fear with 'swipe right', sad with 'swipe down', disgust with 'tap', and surprise with 'press' (Figure 1). We also match emotions to facial expressions based on previous work (Figure 2) [6]. We designed six facial expressions of the character and match them with six touch gestures. Thus, if users send their emotions with touch, the facial expression of the character is changed in accordance with their emotions (Figure 3). The text below the character briefly tells users what gesture they used (Figure 3). For example, if users swipe up the character, our chatbot system shows the text, 'Swipe up' for one second. Then the character shows a happy facial expression. Therefore, users send their emotions with touch, and they get feedback from facial expressions.

TwG for turn-taking

We use the specific touch interaction 'swipe up' to manage turn-taking with TwG because swipe interaction is used for handing something over, moving some features, or changing current state [2, 17]. In particular, because chatting messages are presented from top to bottom, we chose 'swipe up' interaction for turn-taking. Thus, users pass the turn through swiping up their chat bubble (Figure 5). When they swipe up the chat bubble, the accumulated chat logs steadily go up to the system and disappear (Figure 5). It provides feedback indicating that the existing topic is changing. Furthermore, the posture of the character is changed in conformity with the turn of the conversation (Figure 5).

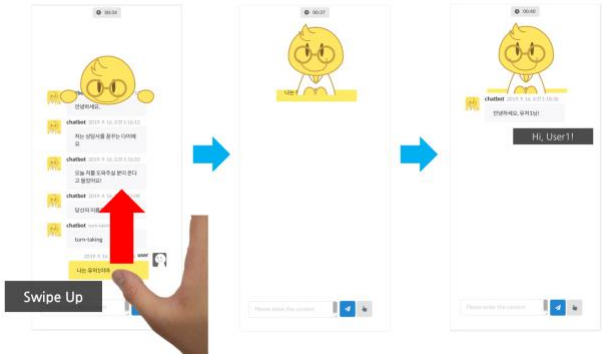

Figure 5: Process of turn-taking through 'swipe-up'

We designed two postures for the character: listening posture and speaking posture (Figure 4). To design the postures, we focus on the behavior of the human listener, especially mimicry. Mimicry is defined as "the effort to imitate facial, vocal, or postural expressions of others with whom we are interacting" [10]. Previous research has shown that mimicry has a positive effect on the pleasantness and naturalness conversation [14]. In the listening posture, to mimic people who see the message when they wait for the response of others, the character bends over and looks down at the message from the top of the system (Figure 4.a). In the speaking posture, to imitate people who see others when they talk with people, the character puts their hands together and looks users straight in the eye (Figure 4.b). Consequently, users do turn-taking by swiping up their chat bubble, and they get feedback from the posture of the character.

Sidebar 1: Conditions of the study

Text-based Interaction (baseline condition): no support of turn-taking gestures (automatic turn-taking after 20 seconds of silence), no support of emotion expression (user can only send text)

Text- and Touch-based Interaction (touch condition): essential use of turn-taking interaction (turn-taking by swiping up chat bubble), touch-based emotion expression (character not presented)

Text- and TwG-based Interaction (TwG condition): essential use of turn-taking interaction (turn-taking by swiping up chat bubble), TwG-based emotion expression (character with facial expressions and changing postures)

Evaluation

Pilot study design and Methods

To evaluate and explore the user experience in human-chatbot interaction with TwG technique, we conducted a Wizard of Oz study. This study evaluated the degree of revealing emotional expression, usefulness of emotional expression, and naturalness of turn-taking via TwG. We recruited 28 participants interested in talking with a chatbot from our university for the study.

The experiment was having a conversation with a chatbot which has three different conditions (Text-based interaction, Text- and touch-based interaction, and Text- and TwG-based interaction) regarding concerns of participants related to three of four topics (learning, career, personality, and interpersonal relationship. See Sidebar 1 for details of conditions.

Before the experiment, participants chose three topics from among the four available. After that, we explained the interaction of each condition and participants could practice the interaction until they were familiar with the interaction. Following the practice, participants communicated with the Oz (the chatbot) for 10 minutes per condition. The Oz chose replies among the predefined set of backchannels and questions. After finishing the conversation, they answered the survey about user experience in communication. The order of conditions and the topic of each condition were set randomly to reduce the effects of the order and topics.

Preliminary Results

To examine the degree of revealing emotional expression, we asked 7-point Likert-scale questions related to emotional connectivity and emotional expression accuracy. The average score of emotional connection in TwG condition (4.39) was higher than in baseline condition (3.51), and touch condition (3.67). Moreover, the average score of emotional expression accuracy in TwG condition (4.86) was higher than in baseline condition (4.24), and touch condition (4.00). Based on the result, we might say that the emotions expressed through TwG revealed the emotions of users better than those expressed through text interaction.

To examine the usefulness of emotional expression, we asked 7-point Likert-scale questions related to ease-of-use in interaction and usability in interaction. The average score of ease-of-use in interaction in baseline condition (5.5) was higher than those in TwG condition (4.71) and touch condition (4.21). This means that text-based interaction is easier than other interactions. The average score of the usability with interaction in TwG condition (4.67) and baseline condition (4.65) were higher than in touch condition (3.9). Therefore, as usability in interaction in TwG condition was not higher than in baseline condition, we might say that TwG condition could be useful considering familiarity of text-based interaction.

To examine the naturalness of turn-taking in conversation, we asked 7-point Likert-scale questions related to positivity in the flow of conversation, and cognition in the flow of conversation. The average score of positivity with the flow in TwG condition (4.94) was higher significantly higher that than in baseline condition (4.29). Furthermore, the average score of cognition with flow in TwG condition (5.49) and touch condition (5.28) were higher than in baseline condition (4.38). Therefore, we might say that turn-taking carried out through TwG is more natural than through text interaction.

Conclusion and Future Work

In this research, we suggested TwG technique to understand emotions and turn-taking cue of users in human-chatbot interaction. To investigate TwG, we conducted Wizard of Oz study, and asked 7-point Likert-scale questions. The result indicated that degree of revealing emotional expression and naturalness of turn-taking was effective in TwG condition, and TwG was also as useful as text-based interaction. Using our interaction technique, users have more ability to send desired emotions and to have a natural turn-taking process in conversation with a chatbot.

However, in our system, the number of expressible emotions is limited to six, and the mapping between touch gestures and emotions is not adequately correlated. Furthermore, the facial expression of the character is not changed if there are no touch gestures, despite changing emotions of users. Therefore, ease of use in TwG interaction is not higher than in text-based interaction due to the limitations of our prototype. Consequently, we plan to extend our work to designing a new way for expressing emotions of users via more relevant touch gestures and designing a chatbot for understanding emotions with text and touch gestures together.

Summary of Interest

A chatbot is one of the most frequently used CUI for various domains. However, the current technologies for chatbot have struggled with several limitations such as sharing emotions and maintaining stable conversational flow. Therefore, we focus on the solution for more human-centered interaction with a chatbot from a more pragmatic view, where users simply indicate those interactions intuitively instead of making the chatbot guess the context.

Acknowledgements

This work was supported by the DGIST Undergraduate Group Research Program (UGRP) grant and the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (NRF- 2017R1C1B2011377). Furthermore, this work was supported by Spirink.

References

[1] Ali Ahmadvand, Jason Ingyu Choi, and Eugene Agichtein. 2019. Contextual Dialogue Act Classification for Open-Domain Conversational Agents. In Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval. ACM, 1273–1276.

[2] Juan Felipe Beltran, Ziqi Huang, Azza Abouzied, and Arnab Nandi. 2017. Don't just swipe left, tell me why: Enhancing gesture-based feedback with reason bins. In Proceedings of the 22nd International Conference on Intelligent User Interfaces. ACM, 469–480.

[3] Joyce Chai, Veronika Horvath, Nicolas Nicolov, Margo Stys, Nanda Kambhatla, Wlodek Zadrozny, and Prem Melville. 2002. Natural language assistant: A dialog system for online product recommendation. AI Magazine 23, 2 (2002), 63–63.

[4] Shammur Absar Chowdhury, Evgeny A Stepanov, and Giuseppe Riccardi. 2016. Predicting User Satisfaction from Turn-Taking in Spoken Conversations. In Interspeech. 2910–2914.

[5] Daantje Derks, Arjan ER Bos, and Jasper Von Grumbkow. 2007. Emoticons and social interaction on the Internet: the importance of social context. *Computers in human behavior* 23, 1 (2007), 842–849.

[6] Shichuan Du, Yong Tao, and Aleix M Martinez. Compound facial expressions of emotion. In *Proceedings*, Vol. 537. National Acad Sciences.

[7] Paul Ekman. 1992. An argument for basic emotions. *Cognition & emotion* 6, 3-4 (1992), 169–200.

[8] Emer Gilmartin and Carl Vogel. 2018. Chat, Chunk and Topic in Casual Conversation. In *Proceedings 14th Joint ACL-ISO Workshop on Interoperable Semantic Annotation*. 45–52.

[9] Matthew J Hertenstein, Rachel Holmes, Margaret McCullough, and Dacher Keltner. 2009. The communication of emotion via touch. *Emotion* 9, 4 (2009), 566.

[10] Ursula Hess and Agneta Fischer. 2014. Emotional mimicry: Why and when we mimic emotions. *Social and Personality Psychology Compass* 8, 2 (2014), 45–57.

[11] Hemkesh Vijaya Kumar, Jeevan Nagaraj, Mohammed Irfan, Nilisha Maheshwari, Pooja Balusani, Pratik Chatterjee, and Gowri Srinivasa. 2018. PESUBot: An Empathetic Goal Oriented Chatbot. In *2018 International Conference on Advances in Computing, Communications and Informatics (ICACCI)*. IEEE, 1083–1089.

[12] Joon Young Lee, Nahi Hong, Soomin Kim, Jonghwan Oh, and Joonhwan Lee. 2016. Smiley face: why we use emoticon stickers in mobile messaging. In *Proceedings of the 18th international conference on human-computer interaction with mobile devices and services adjunct*. ACM, 760–766.

[13] Shao-Kang Lo. 2008. The nonverbal communication functions of emoticons in computer-mediated communication. *CyberPsychology & Behavior* 11, 5 (2008), 595–597.

[14] RM Maatman, Jonathan Gratch, and Stacy Marsella. 2005. Natural behavior of a listening agent. In *International Workshop on Intelligent Virtual Agents*. Springer, 25–36.

[15] KAYA NAz and Helena Epps. 2004. Relationship between color and emotion: A study of college students. *College Student J* 38, 3 (2004), 396.

[16] Young-Woo Park, Kyoung-Min Baek, and Tek-Jin Nam. 2013. The roles of touch during phone conversations: long-distance couples' use of POKE in their homes. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 1679–1688.

[17] Volker Roth and Thea Turner. 2009. Bezel swipe: conflict-free scrolling and multiple selection on mobile touch screen devices. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 1523–1526.

[18] Sherry Ruan, Liwei Jiang, Justin Xu, Bryce Joe-Kun Tham, Zhengneng Qiu, Yeshuang Zhu, Elizabeth L Murnane, Emma Brunskill, and James A Landay. 2019. QuizBot: A Dialogue-based Adaptive Learning System for Factual Knowledge. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. ACM, 357.

[19] Joseph B Walther and Lisa C Tidwell. 1995. Nonverbal cues in computer-mediated communication, and the effect of chronemics on relational communication. *Journal of Organizational Computing and Electronic Commerce* 5, 4 (1995), 355–378.

[20] Ziang Xiao, Michelle X Zhou, and Wat-Tat Fu. 2019. Who should be my teammates: Using a conversational agent to understand individuals and help teaming. In *Proceedings of the 24th International Conference on Intelligent User Interfaces*. ACM, 437–447.

[21] Anbang Xu, Zhe Liu, Yufan Guo, Vibha Sinha, and Rama Akkiraju. 2017. A new chatbot for customer service on social media. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. ACM, 3506–3510.