Enhancing the Perceived Emotional Intelligence of Conversational Agents through Acoustic Cues

Jiaxiong Hu
Academy of Arts & Design, Tsinghua University
The Future Lab, Tsinghua University
Beijing
hujx19@mails.tsinghua.edu.cn

Yun Huang
University of Illinois Urbana-Champaign
Urbana and Champaign, Illinois
yunhuang@illinois.edu

Xiaozhu Hu
Academy of Arts & Design, Tsinghua University
Beijing
huxz19@mails.tsinghua.edu.cn

Yingqing Xu
Academy of Arts & Design, Tsinghua University
The Future Lab, Tsinghua University
Beijing
yqxu@mail.tsinghua.edu.cn

ABSTRACT

The perceived emotional intelligence of a conversational agent (CA) can significantly impact people's interaction with the CA. Prior research applies text-based sentiment analysis and emotional response generation to improve CAs' emotional intelligence. However, acoustic features in speech containing rich contexts are underexploited. In this work, we designed and implemented an emotionally aware CA, called HUE (Heard yoUr Emotion) that stylized responses with emotion regulation strategies and empathetic interjections. We conducted a user study with 75 participants to evaluate their perceived emotional intelligence (PEI) of HUE by having them observe conversations between people and HUE in different emotional scenarios. Our results show that participants' PEI was significantly higher with the acoustic features than without.

CCS CONCEPTS

• **Human-centered computing → Human computer interaction (HCI)**.

KEYWORDS

conversational agent; voice assistant; emotion; chatbot; emotional intelligence

ACM Reference Format:
Jiaxiong Hu, Yun Huang, Xiaozhu Hu, and Yingqing Xu. 2021. Enhancing the Perceived Emotional Intelligence of Conversational Agents through Acoustic Cues. In *CHI Conference on Human Factors in Computing Systems Extended Abstracts (CHI '21 Extended Abstracts), May 8–13, 2021, Yokohama, Japan.* ACM, New York, NY, USA, 7 pages. https://doi.org/10.1145/3411763.3451660

CHI '21 Extended Abstracts, May 8–13, 2021, Yokohama, Japan

ACM ISBN 978-1-4503-8095-9/21/05…$15.00
https://doi.org/10.1145/3411763.3451660

1 INTRODUCTION

Conversational agents (CAs) are increasingly used in people's daily lives. To enable natural interaction with CAs, recent research extends the focus to emotional intelligence [3, 15, 44]. It is also suggested that improving the PEI of CAs has manifold benefits, including but not limited to: enriching interpersonal relationships, increasing engagement, and enhancing user experience [3]. The original psychological definition of emotional intelligence (EI) involves human's ability to appraise and express emotions, regulate emotions, and utilize emotions [31]. In the context of Human Computer Interaction, the perceived emotional intelligence (PEI) of a CA is evaluated based on the CA's ability to perceive user emotions (e.g., detecting and deciphering emotions from words and voice), utilize user emotions (e.g., leveraging the emotions to support cognitive tasks), understand emotions (e.g., comprehending emotions and knowing their triggers), and manage the emotions (e.g., regulating emotions) [21, 44].

To improve the PEI of CAs, different techniques are developed for both text-based and voice-based CAs [4, 13, 30]. For text-based CAs, sentiment analysis is often applied to detect and understand users' emotions from their text input [4, 30], and to help users regulate emotions [33]. Text sequence-to-sequence models [13, 45] are also created to help CAs generate emotionally appropriate responses by taking tone information into account [13]. Varied chatting styles are further designed to improve users' perceived PEI, e.g., a self-disclosing chatbot was found to improve people's perceived intimacy [19]. Compared to text-based CAs, voice-based CAs deliver rich emotional information [16, 40]. For example, several speech emotion recognition (SER) algorithms are developed to detect users' emotions from acoustical cues, by labeling a voice input as a certain emotional category, e.g., happy, sad or angry [1, 22, 46], or by predicting the valence and arousal of a voice input [5]. To respond to users' emotions, voice-based CAs have been relying on users' self-reported emotions, e.g., [23]. During human-to-human interaction, people often express emotions such as empathy–the ability to comprehend other's feelings and to re-experience them oneself–using emotive interjections (e.g., "WoW!") [9, 39]; because of this, such interjections and fillers are also inserted in voice-based CAs to improve users' PEI [6]. However, *prior research has not systematically evaluated the effectiveness of leveraging SER and interjections to generate emotional responses in improving the PEI of chatbots.*

To address this research gap, we design, implement and evaluate a voice-based CA, called HUE (Heard yoUr Emotion). More specifically, we embedded speech emotion recognition (SER) [1] in the design to better perceive users' emotions, and we emotionalized the CA's responses by using interjections and sentiment adjustment strategies including praising, distracting, and reappraising to create a sense of empathy. We conducted a user study with 75 participants to evaluate the PEI of HUE, providing different chatting styles. The participants evaluated HUE as they listened to audio clips that presented four emotion scenarios where HUE conversed with human characters in different chatting styles. Our work provided empirical evidence that the PEI of a CA was influenced by how it responded to different emotion contexts. We found that the CA with the HUE design was evaluated as having higher PEI. Specifically, using both empathetic interjections and sentiment adjustment can enhance the PEI robustly in different emotion contexts.

2 RELATED WORK

2.1 Emotional Intelligence of Conversational Agents

Most of today's conversational agents can be identified as text-based or voice-based conversational agents. Text-based conversational agents are also known as chatbots. They serve people as counselors or assistants in different scenarios, e.g., the personal fashion counselor [24], the survey chatbot [18, 42], the office assistant [20, 35], the quiz chatbot [37] and so on. The importance of affective experience with CAs has been argued [43]. Ma et al. adapted the emotional intelligence model of psychology to human-agent interaction as perceived emotional intelligence, which entails perceiving, using, understanding, and managing emotions, providing a standard of evaluating how the user perceives the emotional intelligence of conversational agents [21, 32].

For text-based CAs, sentiment analysis is often applied. For example, based on a deep learning model, a conversational agent [13] can identify eight major tones from the input text of customers, including: empathetic, passionate, satisfied, and so on. Agents are usually embedded with machine learning models to detect user's emotions from text input [13, 14, 33]. However, for voice-based CAs, speech emotion recognition (SER) solutions are developed. In the following section, we will focus on reviewing SER technologies.

2.2 Technologies for Speech Emotion Recognition (SER)

Acoustic signals contain rich emotional information [16, 40]. Thus, to build a conversational agent with emotional intelligence, SER is considered for emotion perceiving. Neural networks including the convolutional neural network and the long-short term memory recurrent neural network are more effective for emotion recognition [1, 5, 46]. We employed the neural network [1] for SER and trained the model with Voxceleb [25] in this work. With SER but without emotional feedback, the CA cannot perform emotional intelligence, because the user receives no signal that the CA can perceive or express any emotions nor affect others' emotions. Therefore, to improve the PEI, CAs need to generate appropriate responses to address users' emotions. In the next section, we will review technologies that are developed to respond to users with emotional intelligence.

2.3 Technologies for Responding with Emotional Intelligence

To generate a response with an appropriate emotion, Zhou, et al. and Song, et al. built emotional chatting machines with sequence-to-sequence models [34, 45]. By learning many pairs of requests and responses from natural corpus databases, the emotional chatting machines can generate a response with the most possible emotion. Empathy is the key point to connect emotional appraisal and expression [31]. CAs that respond with empathetic utterances achieve better user satisfaction and engagement [8]. Participants in a previous study regarded the perceived empathy of the therapist chatbots as the best part of the experience [10]. To better explore the effects of CA's empathetic responses to emotions on PEI in a user study, we implemented our proposed HUE with practical strategies of response generation as follows.

To adjust the sentiment of the CA's utterances, we refer to computer praise as a strategy for positive emotions since it is effective in improving user motivation and engagement [24, 36]. Then, to deal with negative emotions, two typical emotion regulation strategies of humans are referenced: distraction and reappraisal [7, 17, 28, 29]. Meanwhile, nonverbal cues are also considered in this work. For conversational CAs, using interjections in responses to express emotions can provoke empathy [6, 27]. Therefore, we choose interjections as the nonverbal cues for emotion expression in this work. Moreover, we utilize SER to select appropriate interjections. Here, we define interjections as short nonverbal words or expressions that express a spontaneous feeling, e.g., an exclamation (wow!) or hesitation (um).

The above literature review suggests that SER technologies for identifying emotions and technologies for responding with empathy are promising to improve the PEI of a CA, however, to the best of our knowledge, they have been investigated separately. Little is known regarding the effect of integrating them on people's perceived emotional intelligence of a CA.

3 THE DESIGN OF HUE: A CA WITH SER-ENABLED EMPATHETIC RESPONSES

In this work, we investigate the effectiveness of integrating SER and empathetic response generation. We aimed to answer the research question: *How will SER-enabled empathetic responses affect people's PEI of a voice-based conversational agent?* To this end, we designed HUE (Heard yoUr Emotion), a conversational agent that combines speech emotion recognition and strategy-based emotional response generation. HUE can provide SER-enabled empathetic responses, namely, perceiving and responding to a user's emotion in speech empathetically.

3.1 Different Responding Strategies

We used a neural network model [1] trained with Voxceleb [25] for SER to get emotional input based on acoustic features in speech,

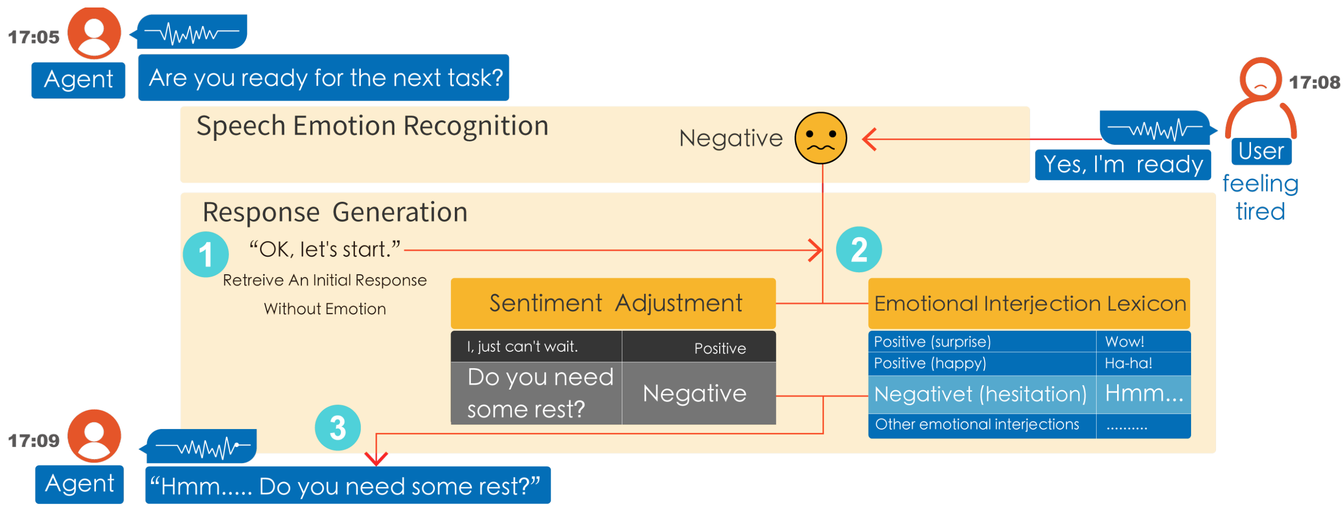

Figure 1: An example dialogue of HUE. A user is collaborating with an agent on tasks. HUE would have responded "OK, let's start" to proceed with the tasks. However, with the SER module recognizing the negative emotional state of the user in speech, HUE inserted an interjection at the beginning to express empathy and adjusted the response content. Eventually, HUE asked the user "Hmm... Do you need some rest?"

e.g., MFCCs. The model was widely applied and proved effective in classifying a speech segment as a discrete emotion label, such as happy [1]. We reconfigured the SER to classify the speech as positive, negative, or neutral. Thus, the response generation script would not be too complicated. We developed two major response strategies: (1) inserting a short interjection in the original response content and (2) adjusting the original response contents with empathetic strategies. The two strategies can be applied simultaneously or individually, as illustrated in Fig 1.

3.1.1 With Interjections. According to the semantic meaning of the user's speech, the CA has an initial response for the neutral emotional input. When the SER detects a positive input, HUE inserts a positive interjection at the beginning of the initial response, e.g., "Ha-ha", and alternatively; when the emotional input is negative, an interjection of hesitation, e.g., "Hmm...". HUE reciprocates the user's emotion by expressing the same emotion, because mimicry is a way for humans to express empathy [12].

3.1.2 Sentiment Adjustments. For the sentiment adjustments, three strategies are considered. When the user performs well and feels positive, praising improves engagement and helps the user persist in tasks [24, 36]. Sometimes, when the user is frustrated, two strategies are available to help the user regulate emotion: distracting from the current task and reappraising the negative stimuli guides the user to a more positive emotional state [7, 17, 28, 29].

3.2 Study Settings

CAs in four conditions were implemented (see in Table 1). Three HUEs use different emotional response strategies: With Interjection (*WI*), Sentiment Adjustment (*SA*), and With Interjection + Sentiment Adjustment (*WI+SA*), while the control CA only provides neutral responses. During the study, participants listened to audio clips of human-agent conversations, as the prior work conducted a video study where participants observed human-agent conversations [21]. The CA information was hidden from participants, so that participants could only infer the characteristics of the CAs based on the presented interaction.

Table 1: For emotion perceiving, *WI*, *SA*, and *WI+SA* enable the SER to recognize user emotion in speech while the control condition doesn't detect any emotional signals. For emotion expressing, the control condition only generates neutral responses; *WI* inserts emotional interjections at the beginning of the responses; *SA* applies emotion regulatory strategies to adjust the sentiment of the responses; *WI+SA* combines both strategies of *WI* and *SA*.

Conditions	SER	Response Strategy
Control	Without SER	Neutral Response (without emotional response strategies)
WI	With SER	Neutral Response + Interjections (e.g., "wow", "ha-ha", "um...")
SA	With SER	Neutral Response + Sentiment Adjustment (praising, distraction and reappraisal)
WI+SA	With SER	Neutral Response + Interjections + Sentiment Adjustment

Table 2: We scripted four emotion scenarios of human-agent conversations. For each scenario, four audio clips were recorded (four conditions). Though the duration of each audio clip was short, the total duration of all audio clips was long, and the comparison among conditions helped participants differ the PEI ratings.

Emotion	Scenario	Audio Duration
Happy	A human character is watching a funny video and asks HUE to share the video to his friends.	7.0s
Sad	The user who just went through a break up in relationship asks the CA for chitchat.	8.4s
Frustrated	The user fails in controlling the lights through voice commands to the CA.	5.0s
Angry	Irritated by the traffic, the user asks the CA to play music.	5.6s

4 USER STUDY: EVALUATING THE PEI OF HUE

4.1 Design of a Controlled Experiment

Simulating daily use scenarios and provoking natural emotions in conversations between the participant and the CA is difficult in the lab. So, we prepared audio clips presenting daily dyadic conversations between human characters and CAs. Listening to the conversations in the audio clips and comparing the CAs in different conditions, participants rated the PEI of the CAs one by one. This was a within-subject design, because each participant was treated with all the conditions. In this way, the user study aims to address the research question by testing in different emotion scenarios, whether people evaluated HUEs (*WI*, *SA*, and *WI+SA*) with higher PEI than traditional CAs (the control condition). It took each participant approximately 25 minutes to complete, including listening to all the audio clips and filling out the questionnaire.

We prepared 16 audio clips (four scenarios and four conditions), presenting the CAs' conversations with human characters. For each scenario, four audio clips were recorded in which the human character conversed with the CA in four conditions. The human characters acted as in four emotional states, i.e., the four scenarios: frustrated, angry, sad, and happy (see Table 2). These four emotions are experienced in daily life [26, 41]. For example, in the sad scenario, the human character just went through a break-up in a relationship and was experiencing deep sadness. He said to the CA, "Would you like to chat?" CAs in different conditions responded differently. The control condition CA didn't perceive the speaker's emotion and it said "What would you like to talk about?" *WI* recognized the emotion and used interjections to express empathy. It said "Um... What would you like to talk about?" *SA* applied sentiment adjustment strategy, and said "Sure, I'm here for you." *WI+SA* combined the above two strategies for a response, said, "Um... Sure, I'm here for you." The context of the conversations was presented as subtitles on the display for participants at the beginning of each audio clip. To guard against the order effects, we used the Fisher-Yates shuffle algorithm to shuffle the playlist of the audio clips for each participant.

Then the participant rated the PEI of the CA in each audio segment with a questionnaire that evaluated the EI in four aspects: perceiving emotions, using emotions, understanding emotions, and managing emotions. Since the participant needed to fill out the questionnaire for each CA in 16 audio clips, it would be annoying if the questionnaire was too long. Hence, we adapted the Perceived Emotional Intelligence Questionnaire [21] as follows. Each question was answered with a 5-point Likert scale (1=totally disagree). This questionnaire demonstrated good internal consistency with an overall Cronbach's coefficient alpha of 0.92. The corrected-item total correlation coefficients of each item with the total of the remaining items ranged from 0.76-0.85.

Questions of the perceived emotional intelligence questionnaire: (The agent is able to...)

- Convey a sense that it listens openly to participants' emotions. (Perceiving Emotion)
- Convey a sense that the agent can feel what the user is feeling. (Using Emotions)
- Respond empathetically to the user. (Understanding Emotions)
- Help the user regulate emotions, reduce negative emotions or keep positive emotions. (Managing Emotions)

4.1.1 Participants. A total of 75 participants (aged 19-73 years, M=24.5, SD=7.78), including 45 females, were recruited for the study. We posted recruitment messages on social media and all the participants registered voluntarily. They are all Chinese native speakers, so the CAs used in the study produce spoken Chinese. Please note that the examples of conversations are all translated into English throughout this paper (the original Chinese contents are in the supplementary material). We surveyed participants on their usage of conversational agents. 69% of participants have used a conversational agent (a voice assistant) with 52% of them using their CA at least once a week. All the participants in this study were recruited to the lab. All the audio clips were presented through a 13' laptop with its speaker. After each clip, participants filled out the PEI questionnaire.

5 RESULTS

Compared to the CA that did not respond to emotion, HUE was rated as having higher PEI. It demonstrated that the SER-enabled empathetic responses increased *bystanders'* PEI.

First, we checked the normality with the Shapiro-Wilk test. The results did not confirm the normality of the data ($p<0.05$), which rejected the use of ANOVA. Thus, we used non-parametric tests to analyze the PEI results. As this was a within-subject design, the Friedman test was used for the main effect analysis and the Wilcoxon signed-rank test with the Holm Bonferroni correction was for the post-hoc analysis.

As shown in Figure 2, the test results showed that the HUE design had a main effect on the PEI ratings of perceiving, using, understanding, and managing emotions (for each aspect $p<0.001$).

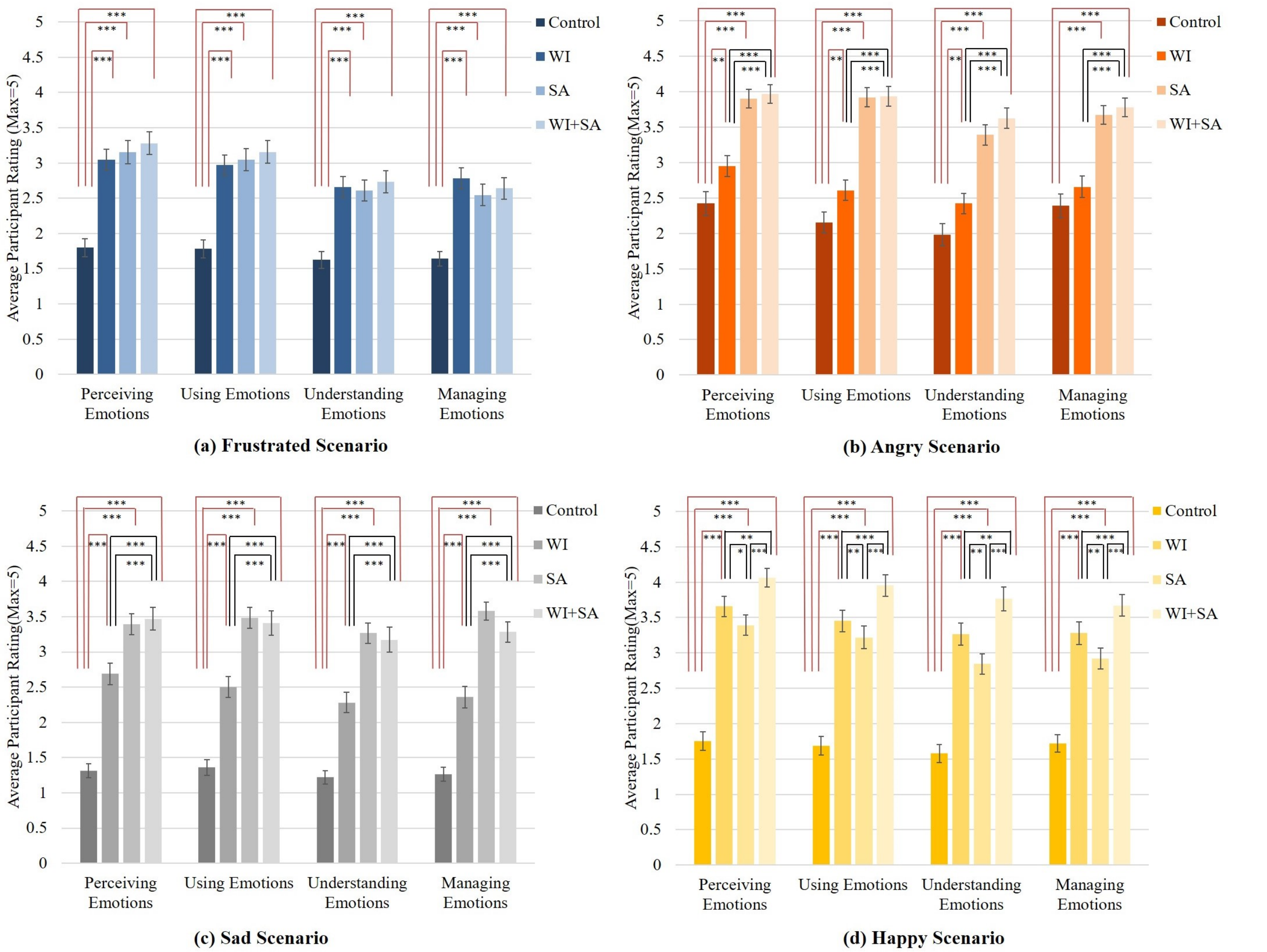

Figure 2: The average participant PEI ratings of the three HUEs and the control CA (in audio clips) in terms of perceiving, using, understanding, and managing emotions. *p<.05, **p<.01, ***p<.001; with standard error.

Further, according to the post-hoc analysis, the average PEI ratings of condition *WI*, *SA*, and *WI+SA* were significantly higher than the control condition across all scenarios (p<0.001, see Figure 2). Though significant improvements were commonly found, the average ratings of HUE varied among different emotion scenarios. For example, the ratings of all HUEs in the frustrated scenario only achieved around 3 (neutral), while *SA* and *WI+SA* were rated around 4 (agree) in the angry scenario.

The comparison among *WI*, *SA*, and *WI+SA* was based on the post-hoc analysis. In the frustrated scenario, the results showed that there was no significant difference among them (see Figure 2). In both sad and angry scenarios, *WI* had lower average ratings than both *SA* and *WI+SA* (p<0.001, see Figure 2 (b) and (c)). But *WI+SA* had no significant difference with *SA*. In the happy scenario, the result was as expected: *WI+SA* performed as the most emotionally intelligent among all conditions (see Figure 2 (d)).

The results showed that SER-enabled empathetic responses significantly improved *bystanders*' perceived emotional intelligence of HUE. The responding strategies–using interjections and sentiment adjustment simultaneously–improved the PEI in different emotional contexts. However, the impact of different strategies varied by the emotional contexts. In particular, using interjections appeared to be more effective in the happy scenario, while the sentiment adjustment was more effective in the angry scenario and the sad scenario.

6 DISCUSSION

6.1 Factors Affecting the Perceived Emotional Intelligence of a Voice-based CA

Most commercial voice-based CAs choose to use a neutral-to-slightly-positive tone, so that it can be accepted by a large number of customers. Our results showed that the empathetic responses improved participants' overall PEI and perceived positive impacts on helping them ease negative feelings.

Using interjections in the CA's responses helped improve the PEI. From the qualitative feedback, we found two possible explanations. 1) Interjections in HUE's responses were quite noticeable. Currently, few speech synthesis systems used interjections [6]. Thus, when hearing interjections from a CA, participants changed their impression of the CA as more human-like. 2) The interjection was considered as confirmation that the CA correctly comprehended the user's emotion. Considering the results showed that using interjections was more effective to improve PEI in positive emotional contexts than negative, we addressed two possible explanations: 1) some interjections like "wow" are beneficial for expressing enthusiasm and strong emotion in human speech [38]; 2) according to broaden-and-build theory of positive emotions [11], under the influence of positive emotions, people have wider perceptual access and semantic reach [11]. Nonverbal information like interjections were smaller details compared to the verbal contents. This suggested that positive emotions helped the interjections access people's minds. Regarding the results, *WI+SA* was rated a better PEI in most of the circumstances. The interjection is possibly too short to draw enough attention, but the sentiment adjustment worked well. So, using both of them simultaneously has the potential to achieve better robustness on improving PEI.

6.2 Design Implications

Our findings have several design implications for improving user interaction with voice-based CA. Neural network based speech emotion recognition algorithms were evaluated as effective in prior work [1, 5, 46]. Prior work also indicated that empathy is the principle for the CA when generating responses [8, 10]. Our findings showed that generating empathetic responses based on the SER results in the CA's conversations improved PEI. This suggested that the CA should take the SER results into consideration to generate empathetic responses. Prior work says that conversational CAs using interjections in responses to express emotions can provoke empathy [6, 27]. Our findings showed that the CA using empathetic interjections achieved increased PEI. This suggested that using interjections is an effective way to respond to user emotion for the CA, and future voice-based CAs may prepare natural interjection records with the professional voice actor/actress [6]. Prior work says reappraisal and distraction are typical human emotion regulation strategies [7, 17, 28, 29], which were proven to be helpful for participants to regulate their emotional states. Our findings showed that the CA which applied these strategies to support user emotion regulation can also have improved PEI. This suggests that more conversational strategies that evolve user emotion may have the potential to enhance the PEI of the CA, and future work may explore this.

7 LIMITATIONS AND FUTURE WORK

This study evaluated people's perceived intelligence of the CA as observers. Future work may explore the effects of direct interaction with HUE. Also, longitudinal study where users interact with the agent such as Lee et al.'s work [19] will yield deeper and richer understandings. Additionally, multi-modal feedback design including audio, facial, physical touch and gesture is also potentially beneficial to our framework [2, 44].

8 CONCLUSION

We proposed HUE, the conversational agent with acoustically emotion awareness and responding, and presented evidence to support that when observing a CA that can perceive and respond to user emotions, participants rate the CA with higher perceived emotional intelligence. Then we discussed the factors affecting PEI and people's expected experience with HUE. Next, we provided the design implication of the combination of speech emotion recognition and emotional response generation. In particular, we offered evidence from the user study that emotional response generation with SER, a) using empathetic and emotional interjections in responses and b) adjusting responses to praise the user, distract or reappraise negative stimuli, can both effectively contribute to the emotional intelligence of the agents. As such, we advocate considering speech emotion recognition and emotional response generation as an acoustically emotion awareness and responding design for conversational agents.

ACKNOWLEDGMENTS

This work was supported by the National Key Research and Development Plan under Grant No.2019YFF0302902 and Tsinghua University-Alibaba Joint Research Laboratory for Natural Interaction Experience under Grant No.20182911173. Assistance provided by Yi Feng, Jincheng Liu, Yuekang Teng, Kexin Quan, Yiting Cheng, Jiahong Sun was greatly appreciated.

REFERENCES

[1] Samuel Albanie, Arsha Nagrani, Andrea Vedaldi, and Andrew Zisserman. 2018. Emotion Recognition in Speech Using Cross-Modal Transfer in the Wild. In *Proceedings of the 26th ACM International Conference on Multimedia* (Seoul, Republic of Korea) *(MM '18)*. Association for Computing Machinery, New York, NY, USA, 292–301. https://doi.org/10.1145/3240508.3240578

[2] T. W. Bickmore, R. Fernando, L. Ring, and D. Schulman. 2010. Empathic Touch by Relational Agents. *IEEE Transactions on Affective Computing* 1, 1 (2010), 60–71. https://doi.org/10.1109/T-AFFC.2010.4

[3] Ana Paula Chaves and Marco Aurélio Gerosa. 2019. How should my chatbot interact? A survey on human-chatbot interaction design. *CoRR* abs/1904.02743 (2019). arXiv:1904.02743

[4] Huimin Chen, Maosong Sun, Cunchao Tu, Yankai Lin, and Zhiyuan Liu. 2016. Neural sentiment classification with user and product attention. In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*. 1650–1659.

[5] Shizhe Chen, Qin Jin, Jinming Zhao, and Shuai Wang. 2017. Multimodal multi-task learning for dimensional and continuous emotion recognition. In *Proceedings of the 7th Annual Workshop on Audio/Visual Emotion Challenge*. ACM, 19–26.

[6] Michelle Cohn, Chun-Yen Chen, and Zhou Yu. 2019. A Large-Scale User Study of an Alexa Prize Chatbot: Effect of TTS Dynamism on Perceived Quality of Social Dialog. In *Proceedings of the 20th Annual SIGdial Meeting on Discourse and Dialogue*. 293–306.

[7] Michelle G Craske, Linda Street, and David H Barlow. 1989. Instructions to focus upon or distract from internal cues during exposure treatment of agoraphobic avoidance. *Behaviour research and therapy* 27, 6 (1989), 663–672.

[8] Kohji Dohsaka, Ryota Asai, Ryuichiro Higashinaka, Yasuhiro Minami, and Eisaku Maeda. 2009. Effects of Conversational Agents on Human Communication in Thought-Evoking Multi-Party Dialogues. In *Proceedings of the SIGDIAL 2009 Conference: The 10th Annual Meeting of the Special Interest Group on Discourse and Dialogue* (London, United Kingdom) *(SIGDIAL '09)*. Association for Computational Linguistics, USA, 217–224.

[9] Martina Drescher. 1997. French interjections and their use in discourse. *The Language of Emotions* (1997), 233–246.

[10] Kathleen Kara Fitzpatrick, Alison Darcy, and Molly Vierhile. 2017. Delivering cognitive behavior therapy to young adults with symptoms of depression and anxiety using a fully automated conversational agent (Woebot): a randomized controlled trial. *JMIR mental health* 4, 2 (2017), e19.

[11] Barbara L. Fredrickson. 2013. Positive Emotions Broaden and Build. Advances in Experimental Social Psychology, Vol. 47. Academic Press, 1 – 53. https://doi.org/10.1016/B978-0-12-407236-7.00001-2

[12] Béatrice S Hasler, Gilad Hirschberger, Tal Shani-Sherman, and Doron A Friedman. 2014. Virtual peacemakers: Mimicry increases empathy in simulated contact with virtual outgroup members. *Cyberpsychology, Behavior, and Social Networking* 17, 12 (2014), 766–771.

[13] Tianran Hu, Anbang Xu, Zhe Liu, Quanzeng You, Yufan Guo, Vibha Sinha, Jiebo Luo, and Rama Akkiraju. 2018. Touch Your Heart: A Tone-Aware Chatbot for Customer Care on Social Media. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (Montreal QC, Canada) *(CHI '18)*. Association for Computing Machinery, New York, NY, USA, 1–12. https://doi.org/10.1145/3173574.3173989

[14] Jing Huang, Qi Li, Yuanyuan Xue, Taoran Cheng, Shuangqing Xu, Jia Jia, and Ling Feng. 2015. Teenchat: a chatterbot system for sensing and releasing adolescents' stress. In *International Conference on Health Information Science*. Springer, 133–145.

[15] Mirjana Ivanović, Miloš Radovanović, Zoran Budimac, Dejan Mitrović, Vladimir Kurbalija, Weihui Dai, and Weidong Zhao. 2014. Emotional Intelligence and Agents: Survey and Possible Applications. In *International Conference on Web Intelligence*.

[16] Tom Johnstone and Klaus R Scherer. 1999. The effects of emotions on voice quality. In *Proceedings of the XIVth international congress of phonetic sciences*. Citeseer, 2029–2032.

[17] Philipp Kanske, Janine Heissler, Sandra Schönfelder, André Bongers, and Michele Wessa. 2011. How to regulate emotion? Neural networks for reappraisal and distraction. *Cerebral Cortex* 21, 6 (2011), 1379–1388.

[18] Soomin Kim, Joonhwan Lee, and Gahgene Gweon. 2019. Comparing data from chatbot and web surveys: Effects of platform and conversational style on survey response quality. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. 1–12.

[19] Yi-Chieh Lee, Naomi Yamashita, Yun Huang, and Wai Fu. 2020. "I Hear You, I Feel You": Encouraging Deep Self-Disclosure through a Chatbot. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '20)*. Association for Computing Machinery, New York, NY, USA, 1–12. https://doi.org/10.1145/3313831.3376175

[20] Q Vera Liao, Muhammed Mas-ud Hussain, Praveen Chandar, Matthew Davis, Yasaman Khazaeni, Marco Patricio Crasso, Dakuo Wang, Michael Muller, N Sadat Shami, and Werner Geyer. 2018. All Work and No Play?. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. 1–13.

[21] Xiaojuan Ma, Emily Yang, and Pascale Fung. 2019. Exploring Perceived Emotional Intelligence of Personality-Driven Virtual Agents in Handling User Challenges. In *The World Wide Web Conference*. ACM, 1222–1233.

[22] Seyedmahdad Mirsamadi, Emad Barsoum, and Cha Zhang. 2017. Automatic speech emotion recognition using recurrent neural networks with local attention. In *2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2227–2231.

[23] Christos N Moridis and Anastasios A Economides. 2012. Affective learning: Empathetic agents with emotional facial and tone of voice expressions. *IEEE Transactions on Affective Computing* 3, 3 (2012), 260–272.

[24] Jonathan Mumm and Bilge Mutlu. 2011. Designing motivational agents: The role of praise, social comparison, and embodiment in computer feedback. *Computers in Human Behavior* 27, 5 (2011), 1643–1650.

[25] Arsha Nagrani, Joon Son Chung, and Andrew Zisserman. 2017. VoxCeleb: A Large-Scale Speaker Identification Dataset. In *Proc. Interspeech 2017*. 2616–2620. https://doi.org/10.21437/Interspeech.2017-950

[26] John B Nezlek, Kristof Vansteelandt, Iven Van Mechelen, and Peter Kuppens. 2008. Appraisal-emotion relationships in daily life. *Emotion* 8, 1 (2008), 145.

[27] A.I. Niculescu, S.S. Ge, Elisabeth M.A.G. van Dijk, Antinus Nijholt, Haizhou Li, and Swan Lan See. 2013. Making social robots more attractive: the effects of voice pitch, humor and empathy. *International journal of social robotics* 5, 2 (21 4 2013), 171–191. https://doi.org/10.1007/s12369-012-0171-x eemcs-eprint-22397.

[28] Kevin N Ochsner and James J Gross. 2005. The cognitive control of emotion. *Trends in cognitive sciences* 9, 5 (2005), 242–249.

[29] Kevin N Ochsner and James J Gross. 2008. Cognitive emotion regulation: Insights from social cognitive and affective neuroscience. *Current directions in psychological science* 17, 2 (2008), 153–158.

[30] Qiao Qian, Minlie Huang, Jinhao Lei, and Xiaoyan Zhu. 2017. Linguistically Regularized LSTM for Sentiment Classification. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. Association for Computational Linguistics, Vancouver, Canada, 1679–1689. https://doi.org/10.18653/v1/P17-1154

[31] Peter Salovey and John D Mayer. 1990. Emotional intelligence. *Imagination, cognition and personality* 9, 3 (1990), 185–211.

[32] Peter Ed Salovey and David J Sluyter. 1997. *Emotional development and emotional intelligence: Educational implications.* Basic Books.

[33] Heung-Yeung Shum, Xiao-dong He, and Di Li. 2018. From Eliza to XiaoIce: challenges and opportunities with social chatbots. *Frontiers of Information Technology & Electronic Engineering* 19, 1 (2018), 10–26.

[34] Zhenqiao Song, Xiaoqing Zheng, Lu Liu, Mu Xu, and Xuan-Jing Huang. 2019. Generating Responses with a Specific Emotion in Dialog. In *Proceedings of the 57th Conference of the Association for Computational Linguistics*. 3685–3695.

[35] Carlos Toxtli, Andrés Monroy-Hernández, and Justin Cranshaw. 2018. Understanding chatbot-mediated task management. In *Proceedings of the 2018 CHI conference on human factors in computing systems*. 1–6.

[36] Jeng-Yi Tzeng and Cheng-Te Chen. 2012. Computer praise, attributional orientations, and games: A reexamination of the CASA theory relative to children. *Computers in Human Behavior* 28, 6 (2012), 2420–2430.

[37] Justin D Weisz, Mohit Jain, Narendra Nath Joshi, James Johnson, and Ingrid Lange. 2019. BigBlueBot: teaching strategies for successful human-agent interactions. In *Proceedings of the 24th International Conference on Intelligent User Interfaces*. 448–459.

[38] Wierzbicka and Anna. 1992. The semantics of interjection. *Journal of Pragmatics* 18, 2-3 (1992), 159–192.

[39] Anna Wierzbicka. 1999. *Emotions across Languages and Cultures: Diversity and Universals*. Cambridge University Press. https://doi.org/10.1017/CBO9780511521256

[40] Carl E Williams and Kenneth N Stevens. 1972. Emotions and speech: Some acoustical correlates. *The Journal of the Acoustical Society of America* 52, 4B (1972), 1238–1250.

[41] Emily C Willroth, Jayde AM Flett, and Iris B Mauss. 2020. Depressive symptoms and deficits in stress-reactive negative, positive, and within-emotion-category differentiation: A daily diary study. *Journal of personality* 88, 2 (2020), 174–184.

[42] Ziang Xiao, Michelle X. Zhou, Q. Vera Liao, Gloria Mark, Changyan Chi, Wenxi Chen, and Huahai Yang. 2020. Tell Me About Yourself: Using an AI-Powered Chatbot to Conduct Conversational Surveys with Open-Ended Questions. *ACM Trans. Comput.-Hum. Interact.* 27, 3, Article 15 (June 2020), 37 pages. https://doi.org/10.1145/3381804

[43] Xi Yang, Marco Aurisicchio, and Weston Baxter. 2019. Understanding Affective Experiences with Conversational Agents. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland Uk) *(CHI '19)*. Association for Computing Machinery, New York, NY, USA, 1–12. https://doi.org/10.1145/3290605.3300772

[44] Yang Yang, Xiaojuan Ma, and Pascale Fung. 2017. Perceived emotional intelligence in virtual agents. In *Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems*. ACM, 2255–2262.

[45] Hao Zhou, Minlie Huang, Tianyang Zhang, Xiaoyan Zhu, and Bing Liu. 2018. Emotional chatting machine: Emotional conversation generation with internal and external memory. In *Thirty-Second AAAI Conference on Artificial Intelligence*.

[46] Suping Zhou, Jia Jia, Qi Wang, Yufei Dong, Yufeng Yin, and Kehua Lei. 2018. Inferring emotion from conversational voice data: A semi-supervised multi-path generative neural network approach. In *Thirty-Second AAAI Conference on Artificial Intelligence*.