Exploring Effects of Conversational Fillers on User Perception of Conversational Agents

Yuin Jeong
Yonsei University
Seoul 03722, Republic of Korea
youin.jeong@gmail.com

Juho Lee
Yonsei University
Seoul 03722, Republic of Korea
juho@yonsei.ac.kr

Younah Kang
Yonsei University
Seoul 03722, Republic of Korea
kang.younah@gmail.com

ABSTRACT

Through technological advancements in various areas of our lives, Conversational Agents progressed in their human-likeness. In the field of HCI, however, the use of conversational fillers (e.g., "um," "uh," etc.) by Conversational Agents have not been fully explored in an experimental setting. We observed the effects on user perceptions of Intelligence, Human-likeness and Likability of Conversational Agents by a 2 x 2 experimental design. From the results of 26 total participants, we concluded that 1) the use of fillers by Conversational Agents are perceived as less intelligent and less likable in task-oriented conversations, 2) and the fillers did not have any statistically significant change in perception of human-likeness. However, further examination showed that users reported filler-speaking agents as more entertaining for social-oriented conversations. With these findings, we discuss design implications for voice-based Conversational Agents.

CHI'19 Extended Abstracts, May 4-9, 2019, Glasgow, Scotland, UK.

ACM ISBN 978-1-4503-5971-9/19/05.
DOI: https://doi.org/10.1145/3290607.3312913

KEYWORDS

Conversational Agent; Voice User Interface; Conversational Fillers; User Perception; Perceived Intelligence; Human-likeness; Likability

1 INTRODUCTION

With advanced technology, voice-based Conversational Agents (CAs) have become much smarter and more 'human-like'. Based on previous studies showing people respond more positively to human-like interface [8, 10], researchers and designers have applied human communication mechanisms to the agent: the humorous 'Siri' and laughing 'Alexa' are well-known examples.

Conversational fillers are widely used in human conversations but are not generally considered as design factors for CAs. Conversational fillers (e.g. "um" or "uh") are one of the most frequent features in everyday speech [13], which serve a variety of linguistic functions in human conversations. For example, fillers help speakers to buy more time especially when faced with abstract and unfamiliar topics [4]. Fillers are also used in managing natural turn-taking, helping listeners predict and prepare faster for the upcoming utterance [14]. Furthermore, fillers affect listeners' impressions of speakers. 'Perceived intelligence' is one of the most debated effects of fillers. According to [1], fillers make speakers seem less prepared. By contrast, there are studies that argue fillers are not a reliable predictor of evaluating intelligence [12].

However, despite their frequent use and important role in human conversations, fillers in CAs have not yet been fully explored. In the HCI field, several studies have examined multimodal interactions with the agents, where they investigated the effects of vocal fillers along with body movements such as an eye gaze or touching of the chin. For example, a study found that vocal and bodily fillers in a robot help moderate awkwardness during silences [9]. Another study also revealed that the perceived humanness and likability of the robot increased when it included fillers in its speech [16]. On the other hand, a study of virtual agents' vocal fillers with eye-gaze behavior reported no significant effects on user satisfaction and naturalness of interactions [11].

Therefore, it is necessary to study the application of fillers to CAs in different situations involving diverse users. Particularly, as voice-based CAs are "becoming increasingly embedded in personal technologies and device" [7] and fillers are frequently used in audio-only human conversations [6], there is a need to further analyze what effects fillers have in CAs when they are a vocal interface without body movement. In addition, although perceived intelligence is being discussed as one of the major factors in researching about fillers in human conversations, the effects have not been the focus of HCI. Given these aforementioned results from the literary review, the research questions are presented below:

RQ1. Do fillers influence the perceived intelligence of CAs?
RQ2. Do fillers influence users' perception on the human-likeness of CAs?
RQ3. Do fillers influence users' likability of CAs?
RQ4. Do fillers influence differently in task- and social-oriented conversations?

2 RESEARCH METHOD

2.1 Overview

A 2 (Use of Filler: With fillers vs. Without fillers) x 2 (Context of conversation: Task-oriented vs.

Table 1: Experimental Design

		Use of Fillers	
		With-Fillers	*Without-Fillers*
Context of Conversation	*Task-Oriented*	A	B
	Social-Oriented	C	D

* Since it was a 2x2 factorial within-subject designed experiment, all participants were exposed to every condition (A, B, C, D).

Table 2: A script displaying experimental condition A (With-fillers, Task oriented)

Speaker	*Contents*
User	"Jane, order a pizza"
Agent	"Hi, do you want to make a new order or order your favorite?"
User	"New order"
Agent	"What would you like to order? You can add toppings if you want"
User	"Medium pizza with sausage"
Agent	"um, would you like to have anything else?"
User	"and a coke"
Agent	"um, now you have one medium pizza and a coke in your order. Would you like to place this order?"
User	"Yes"
Agent	"You will receive 30 % discount for the first order. That will be… um…33 dollars in total. Would you like to place this order?"
User	"Yes, order it."
Agent	"The order has been made. Enjoy your meal."

Social-oriented) factorial within-subject design experiment (Table 1) was conducted in a well-controlled laboratory located in Yonsei University, South Korea. For the purpose of this experiment, CAs were controlled with a Wizard-of-Oz setup on another computer in a separate room.

2.2 Participants

26 participants took part in this within-subject experiment (Mean age = 21.4, SD = 2.71, 9 males). All were undergraduate or graduate students whose first language was English. 10 participants had previous experiences using CAs at home. After the experiment, each participant received $15 as compensation.

2.3 Stimuli

Among conversational fillers, "um", one of the most frequently occurring fillers in English conversation [4], was used in the experiment. Each script consisted of 12 turns of dialogue and 3 "um"s were inserted in the 'with-filler' condition (shown in Table 2). The voice of the agent named "Jane" was designed via Amazon Polly text-to-speech engine and embedded in the Amazon Echo device. The intonation and occurrence of each filler was adjusted to sound as natural as possible and was verified by four native speakers who did not take part in the experiment.

In order to analyze the effects of fillers in different situations, a total of four scenarios were presented; two were task-oriented conversations and the other two were social-oriented. In the task-oriented scenario, users were asked to order a pizza. In the social-oriented scenario, the users were asked to have an open-ended conversation on abstract questions with the agent (e.g. "Jane, what is love?"). In order to design the conversation as authentically as possible, all scripts were written referring to Amazon Alexa skill set scenarios.

2.4 Procedure

First, the participants were asked to have a sample dialogue with the agent to reduce novelty effects before the actual experiment. Following the exercise, participants conducted four different types of conversation according to the different experimental conditions. In order to control for order effects, the scenarios were randomized following the Latin Square Design. The instruction cards for the participants' responses were given in advance and the responses of the CAs were controlled by human operation as in the Wizard of Oz method. After each condition was completed, the participants filled out a survey questionnaire (shown in Table 3) to evaluate their impression. Finally, a semi-structured interview was conducted to gain a deeper understanding of participants' overall responses to the conversational experience, comprised of interview questions such as overall experience, feelings, preference, impression on the agent and its utterance.

3 RESULTS

A two-way repeated measures ANOVA test was used in line with previous works done in the field of HCI, despite the small sample size for an ANOVA analysis [5, 17]. Also, the data acquired from the interviews were analyzed based on thematic analysis and an open coding method [3].

Table 3: Survey Questionnaires

Construct	*Questions*
	Please rate your impression of the agent on these scale *(7-point semantic differential scales)*
Perceived Intelligence (4 items)	Incompetent / Competent Unintelligent / Intelligent Foolish / Sensible Ignorant / Knowledgeable
Human-likeness (4 items)	Artificial / Lifelike Machine-like / Human-like Unconscious / Conscious Fake / Natural
Likability (4 items)	Dislike / Like Unfriendly / Friendly Awful / Nice Unpleasant / Pleasant

*Evaluation questionnaires were consisted of questions to assess users' impression on the agent regarding Perceived intelligence, Human-likeness and Likability. Modifying related literature[2], the study used 7-point semantic differential scales to evaluate the impression towards the agent. 4 items were included to measure each construct.

3.1 Perceived Intelligence

A significant main effect of fillers ($F(1,25)=5.862$, $p=.023$), and a significant interaction effect between fillers and context of conversation ($F(1,25)=8.553$, $p=.007$) on perceived intelligence were revealed. This indicates that the participants perceived the filler-speaking agent as less intelligent, especially in task-oriented conversations. In social-oriented conditions, the perceived intelligence rates of filler-speaking agents were slightly increased (reported in Fig. 1).

The results can be attributed to participants' responses from the interviews. 19 out of 26 participants indicated that the fillers made the agent seem less prepared and unsure of what it is doing in task-oriented conversation: *"It felt a bit weird, because it made her seem less sure about the price or her work"-p13, "She said 'um' which threw me off a little, it seemed less credible"-p19*

In particular, four participants commented that since the agent is a machine, they expected a more precise answer from the agent, not a level of hesitation: *"If it's supposed to be AI, I think it always should be better than human. I mean, it's not allowed to hesitate at any moment"-p8, "It's a little strange to hear a machine do it[um]"-p20*

Lastly, just like in human conversations, the use of "um" by CAs in formal task-oriented situations gave the impression that the CA lacked understanding of the situation and was not providing an appropriate response: *"When I was at high school, I was trained to not use fillers in public speaking. That's why I thought the ums from Jane seemed inappropriate and less intelligent"-p5.*

On the other hand, several participants responded that when the CA used "um" in its response to an open-ended question, it gave them the impression that it was conscious and had its own opinion about the topic, and made the participant appreciate the advance in technology: *"By doing that[um], it seems like the computer is also giving itself more time to find the answer, that's clever"-p10.* Therefore, these results indicate that the users' perceived intelligence differ by context.

3.2 Human-likeness

Although there was no statistical significance, participants in the 'With-filler' condition had a tendency to evaluate the agent as more human-like than those in the 'Without-filler' condition($p=.15$). This result was also consistent in the interviews. 14 out of 26 participants responded that agents using fillers were more human-like, because 1) the agent used more natural speech similar to humans: *"Actually in human conversation we also speak that way..."-p23, "it was very natural because it comes up in human conversation as well"-p1* and 2) the agent felt more conscious: *"It feels like she takes a moment to think before answering the tough questions and that made me think the following answer was 'real'"-p22.* However, there were also some contrasting responses. Since the participant was already aware of the fact that the agent was a machine and did not require more time to think, the fillers used by the agent felt unnatural and fake: *"Because I already know it is something like a robot, when it says um, it seems 'fake'"– p7.*

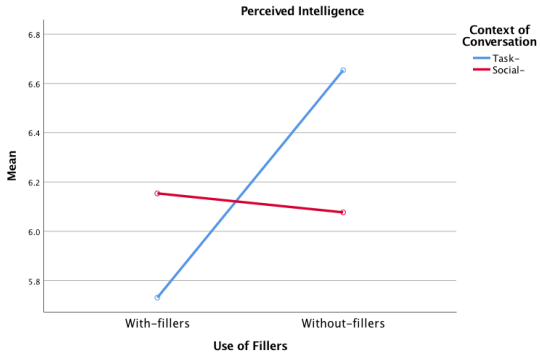

Figure 1: Interaction effect between two variables on Perceived intelligence

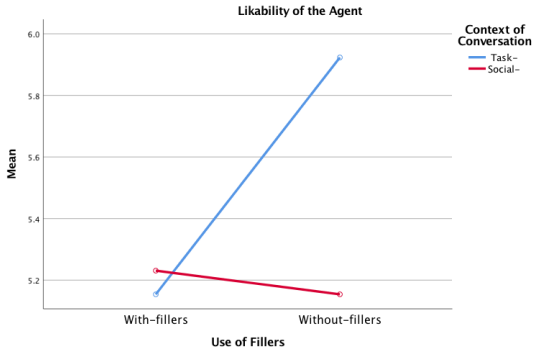

Figure 2: Interaction effect between two variables on Likability

3.3 Likability

There was no statistically significant main effect of fillers ($F(1,25)=3.130$, $p=.089$) on users' likability, but a significant interaction effect between fillers and context of conversation($F(1,25)=4.590$, $p=.042$) was found. This indicates that fillers and users' likability ratings correlated differently depending on context (showed in Fig. 2).

The analysis of the interview also revealed that participants had mixed feelings about interacting with an agent that used fillers. They considered fillers as entertaining and fun in social conversation, regardless of its perceived human-likeness rating: *"It definitely makes the conversation more interesting and entertaining, I like to have a 'thinking agent'"-p26, "I know that's pre-programmed to react with some funny responses, it's still fun though"-p17.*

By contrast, similar to the findings of previous studies [15], 15 participants responded negatively to the use of fillers as they preferred the CA to provide clear information, since it is an assistant performing tasks for the users: *"the reason for using one of these devices is not looking for friends, but looking for something to carry out services for me. So, it should convey information clearly and does not need to be life-like"-p7, "Because it's supposed to give information, it should seem surer of itself instead of being like 'um'"-p18.*

Moreover, individual difference was observed in likability according to the participants' personal speaking habits in using fillers in face-to-face communication. One participant stated that she was more familiar with and tolerant of the use of fillers since she frequently used them in daily conversations. On the other hand, another participant responded that he did not like the use of "um"s by the CA because he was trained not to use them.

4 DESIGN IMPLICATIONS

Based on our findings, we propose implications which can be considered when designing CAs.

(a) Avoid using fillers when dealing with sensitive information: *"When it comes to doing tasks relating to ordering or planning, it needs to be more to the teeth"-p20.* Utterances of CAs should be precise and clearly conveyed especially in handling important information, in which CA's unprecise performance is negatively affecting benefits to users. For example, when calculating the cost of ordering products or checking the time and date for reserving a flight, usage of fillers needs to be restrained since it makes users feel anxious.

(b) Consider fillers as entertaining factors rather than human-like design factors: The application of fillers in CAs has powerful potential in increasing entertaining effects of CAs. Our research showed that the use of fillers in answering the open-ended questions or silly questions increase people's enjoyment in both groups of participants, who perceived the filler-speaking agents as more human-like or those who did not. By making the agent pretend to "think", using fillers increases the possibility for users to engage in social-oriented conversations.

(c) Provide various speech forms of fillers: *"If it would sound different, because to me now, it sounds just like 'a sound'. So, if the ums were different and there were variations, I will buy the one with fillers"-p3.* The length and tone of the fillers serve as central non-lexical markers that convey emotions in conversations. Therefore, when applying fillers in CAs, the intonation and length need to be

ACKNOWLEDGMENTS

This research was supported by Korea Institute for Advancement of Technology (KIAT) grant funded by the Korea Government (MOTIE) (N0001436, The Competency Development Program for Industry Specialist).

considered significantly. In face-to-face conversations, “um” performs different functions according to different lengths and tones, so variations of “um”s need to be adopted in the application to CAs. Moreover, using the same type of “um” repeatedly in a conversation needs to be avoided.

5 CONCLUSIONS AND FUTURE WORKS

In this paper, we explored the effect of fillers in voice-based conversational agents. Overall, fillers have been shown to be perceived differently regarding the context of conversations, showing significant interaction effect on perceived intelligence and the likability of the agent. Future research is required to explore the effects of different types of fillers such as “uh” or “well” with a broader range of participants, including those who frequently use CAs in their everyday lives.

REFERENCES

[1] Arnold, J. E., Kam, C. L. H., & Tanenhaus, M. K. (2007). If you say thee uh you are describing something hard: the on-line attribution of disfluency during reference comprehension. Journal of Experimental Psychology: Learning, Memory, and Cognition, 33(5), 914.

[2] Bartneck, C., Croft, E., & Kulic, D. (2008, March). Measuring the anthropomorphism, animacy, likeability, perceived intelligence and perceived safety of robots. In Metrics for HRI workshop, technical report (Vol. 471, pp. 37-44).

[3] Burnard, P. (1991). A method of analysing interview transcripts in qualitative research. Nurse education today, 11(6), 461-466. https://doi.org/10.1016/0260-6917(91)90009-Y

[4] Clark, H. H., & Tree, J. E. F. (2002). Using uh and um in spontaneous speaking. Cognition, 84(1), 73-111.

[5] Iqbal, S. T., Zheng, X. S., & Bailey, B. P. (2004, April). Task-evoked pupillary response to mental workload in human-computer interaction. In CHI'04 extended abstracts on Human factors in computing systems (pp. 1477-1480). ACM.

[6] Kasl, S. V., & Mahl, G. F. (1965). Relationship of disturbances and hesitations in spontaneous speech to anxiety. Journal of personality and social psychology, 1(5), 425.

[7] Luger, E., & Sellen, A. (2016, May). Like having a really bad PA: the gulf between user expectation and experience of conversational agents. In Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems (pp. 5286-5297). ACM.

[8] Nass, c., Steuer, 1. & Tauber, E. R. (1994), Computers are Social Actors, in B. Adelson, S. Dumais & 1. Olson (eds.), Proceedings of CH1'94: Human Factors in Computing Systems, ACM Press, pp.72-7.

[9] Ohshima, N., Kimijima, K., Yamato, J., & Mukawa, N. (2015, August). A conversational robot with vocal and bodily fillers for recovering from awkward silence at turn-takings. In Robot and Human Interactive Communication (RO-MAN), 2015 24th IEEE International Symposium on (pp. 325-330). IEEE.

[10] Pawel, D., Michal, P., Rafal, R., & Kenji, A. (2009, May). Humoroids: conversational agents that induce positive emotions with humor. In AAMAS'09 Proceedings of The 8th International Conference on Autonomous Agents and Multiagent Systems (Vol. 2, pp. 1171-1172). ACM.

[11] Pfeifer, L. M., & Bickmore, T. (2009). Should agents speak like, um, humans? The use of conversational fillers by virtual agents. In International Workshop on Intelligent Virtual Agents (pp. 460-466). Springer, Berlin, Heidelberg.

[12] Pytko, J. L., & Reese, L. O. (2013). The Effect of Using" Um" and" Uh" on the Perceived Intelligence of a Speaker. College of St. Elizabeth Journal of the Behavioral Sciences.

[13] Rose, R. L. (1998). The communicative value of filled pauses in spontaneous speech. MA Diss., Univ. of Birmingham.

[14] Tree, J. E. F. (1993). Comprehension after speech disfluencies. Stanford University.

[15] Westerman, D., Cross, A. C., & Lindmark, P. G. (2018). I Believe in a Thing Called Bot: Perceptions of the Humanness of “Chatbots”. Communication Studies, 1-18.

[16] Wigdor, N., de Greeff, J., Looije, R., & Neerincx, M. A. (2016, August). How to improve human-robot interaction with Conversational Fillers. In Robot and Human Interactive Communication (RO-MAN), 2016 25th IEEE International Symposium on (pp. 219-224). IEEE.

[17] Zagermann, J., Pfeil, U., & Reiterer, H. (2018, April). Studying Eye Movements as a Basis for Measuring Cognitive Load. In Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems (p. LBW095). ACM.