Proceedings of the Ninth AAAI Conference on Human Computation and Crowdsourcing (HCOMP 2021)

Making Time Fly: Using Fillers to Improve Perceived Latency in Crowd-Powered Conversational Systems

Tahir Abbas[1,2], Ujwal Gadiraju[3], Vassilis-Javed Khan[4], Panos Markopoulos[1]
[1] Eindhoven University of Technology, Eindhoven, Netherlands
[2] Software Engineering Dept., Mirpur University of Science & Technology, Mirpur AJK, Pakistan
[3]Web Information Systems, Delft University of Technology, Netherlands
[4]Digital Transformation, Sappi Europe, Brussels, Belgium
[t.abbas, p.markopoulos]@tue.nl, u.k.gadiraju@tudelft.nl, javed.khan@sappi.com

Abstract

Crowd-Powered Conversational Systems (CPCS) are gaining traction due to their potential utility in a range of application fields where automated conversational interfaces are still inadequate. Currently, long response times negatively impact CPCSs, limiting their potential application as conversational partners. Related research has focused on developing algorithms for swiftly hiring workers and synchronous crowd coordination techniques to ensure high-quality work. Evaluation studies typically concern system reaction times and performance measurements, but have so far not examined the effects of extended wait times on users. The goal of this study, based on time perception models, is to explore how effective different time fillers are at reducing the negative impacts of waiting in CPCSs. To this end, we conducted a rigorous simulation-based between-subjects ($N = 930$) study on the Prolific crowdsourcing platform to assess the influence of different filler types across three levels of delay (8, 16 & 32s) for Information Retrieval (IR) and stress management tasks. Our results show that asking users to perform secondary tasks (e.g., microtasks or breathing exercises) while waiting for longer periods of time helped divert their attention away from timekeeping, increased their engagement, and resulted in shorter perceived waiting times. For shorter delays, conversational fillers generated more intense immersion and contributed to shorten the perception of time.

Introduction

Despite great progress, current artificial intelligence (AI) and natural language processing techniques are incapable of dealing with the full complexity of free-form dialogues, often resulting in conversation breakdowns (Ashktorab et al. 2019). As a solution to AI's flaws, crowd-powered conversational systems (CPCS) (Lasecki et al. 2013b; Huang, Chang, and Bigham 2018) have been proposed. CPCSs are a natural extension of Wizard of Oz (WoZ) setup (Riek 2012), with the following key distinctions; a dynamic group of crowd workers is recruited on-demand to offer responses and vote for the best ones; workers are motivated to submit responses quickly and accurately through a game theoretic reward system (Lasecki et al. 2013b). Potentially, CPCSs could be more robust than current AI in terms of handling interaction with users in a fluid, multi-turn conversation. A pioneering example of a CPCS is Chorus (Lasecki et al. 2013b), which is a text-based conversational agent that assists end-users in information retrieval tasks by conversing with an online group of workers in real time. These CPCSs rely (in part or entirely) on human operation which causes response delays that are likely to annoy end-users and jeopardize the overall user experience.

Earlier research in crowd-powered systems has developed two strategies to achieve fast and reliable answers from the crowd: on-demand recruitment and synchronous crowds. In on-demand recruiting, workers' arrival latency is reduced by hiring them in advance and engaging them in other micro-tasks or simply paying them to wait (Bigham et al. 2010; Bernstein et al. 2011). However, this strategy only helps to solve the "pre-task" latency, which is the time until a crowd worker accepts a newly posted task (Haas et al. 2015). An alternative strategy to reduce the response time is to hire multiple, synchronous workers simultaneously (Huang et al. 2016; Lasecki et al. 2013a; Abbas et al. 2020a; Lasecki et al. 2013b) and then build coordination schemes to aggregate answers or utilize *first-response* strategy to select the first available answer (Bigham et al. 2010; Abbas et al. 2020a). However, hiring several workers does not necessarily ensure a faster response due to a variety of factors, such as poor Internet connection, workers' slow typing speed, and fatigue. Another problem is the higher operating budget associated with hiring multiple workers or paying them for simply waiting (Huang et al. 2016). Furthermore, response delays are inevitable for some type of tasks, such as Information Retrieval (IR) (e.g., Chorus), in which workers have to forge different information sources on the web to find relevant information. Evaluation studies in this area have focused primarily on reporting system response times (Lasecki et al. 2013b; Huang et al. 2016; Huang, Chang, and Bigham 2018; Lasecki et al. 2013a; Huang, Lasecki, and Bigham 2015) and other performance measures but have yet to explore the adverse effects of long waiting on end-users.

The goal of this study is to mitigate the problem of long waiting times in CPCS by time fillers; a filler can be any in-

Copyright © 2021, Association for the Advancement of Artificial Intelligence (www.aaai.org). All rights reserved.

tervention that helps to divert users' attention from waiting, as a result, reducing their perception of elapsed time and/or the frustration that waiting for a system response may cause. There is an ample evidence that the subjective sense of waiting is even more crucial than measurable time itself (Hornik 1984; Tognazzini 1993) because non-temporal characteristics of the event may modify the actual duration. As a result, the advantages of using time fillers that can alter the actual duration can often be achieved immediately and at no cost (Harrison, Yeo, and Hudson 2010). Based on time perception theories, we used seven different types of fillers, divided into four categories: *activity, animation, conversational*, and *relaxation* fillers. *Activity* fillers let users to "do something" while they wait, such as doing secondary tasks or practice breathing exercises; *animation* fillers show progress bars or three-animated dots; *conversational* fillers include short remarks to acknowledge the person about the delay; and finally, *relaxation* fillers show nature or urban videos to sidetrack users. These fillers were employed in two CPCSs for IR and stress management tasks with pre-programmed delays of 8, 16, and 32 seconds. We conducted a rigorous simulation-based between subjects experiment on Prolific platform, involving 930 unique workers across 48 different experimental conditions (2 task types x 3 delay levels x 8 filler types including no-filler) to address the following overarching research question:

RQ: How do different types of fillers effect a user's perception of waiting in IR and stress management-based conversations in the context of a CPCS with response delays?

Our findings show that:

- When compared to other fillers, a microtask filler that allowed users to perform simple tasks while waiting resulted in a shorter perceived time, especially in the information retrieval tasks.
- Conversational fillers were successful in shifting users' attention away from waiting, improving engagement, and making the time seem shorter for short (8s) delays, mindful breathing and microtask were equally useful for moderate (16s) delays, whereas for lengthier delays (32s), microtask filler was more effective.
- Mindful breathing and progress fillers were found to be effective in increasing users' engagement (focused immersion) in both tasks.
- Overall, mindful breathing filler was more enjoyable than the other fillers.

Background and Related Work

Response Delays in CPCS

Recently, researchers have developed methods to recruit workers on demand from crowdsourcing platforms and have created interventions that enable responses in a few seconds, rather than a matter of hours or days as in the past. Such methods enable researchers to build interactive crowd-powered systems (Lasecki, Homan, and Bigham 2014). VizWiz (Bigham et al. 2010) is one of the earliest real-time crowd-powered system which helps blind users answer visual questions about their surroundings by sending photos of objects and audio questions from their phones to crowd workers.

Regarding CPCS, a pioneering example is Chorus (Lasecki et al. 2013b), which supports IR tasks. Chorus lets workers propose responses and vote for one another's responses to support a consistent dialogue with end-users. Field deployments of Chorus show that 25% percent of the conversations obtained a first answer within 30 seconds (Huang et al. 2016). Chorus:view (Lasecki et al. 2013a) is a conversational application, which was built on top of Chorus. Through the use of pictorial questions, it helps visually impaired users converse with crowd workers about their environment. In a trial with blind users, Chorus:View was able to accomplish various tasks with an average response time of 295s, 351.2s and 182.3s for product detail, information finding and navigational tasks, respectively. Evorus (Huang, Chang, and Bigham 2018) is an automated version of Chorus that integrates existing chatbots via REST APIs and learns to select high-quality responses from chatbots based on past workers' evaluations of chatbot's responses. Guardian (Huang, Lasecki, and Bigham 2015) is a crowd powered spoken dialogue system that combines web APIs with crowdsourcing to enhance the scope of open dialogue systems. In their experiments, the average time required to receive a response from the Yelp search API was $\sim$110s. InstructableCrowd (Huang et al. 2019) is a conversational system that enables end users to converse with a group of workers about their requirements for IF-THEN rules. Crowd workers aid end users by developing IF-THEN rules that are then executed on their mobile devices. In their experiment, the average time the system took to create rules was $\sim$5 minutes. Crowd of Oz (CoZ) (Abbas et al. 2020a) crowdsources conversational tasks to a synchronous group of workers for social robots. CoZ transcribes user speech and sends it to crowd workers with Audio-video (AV) feed. The crowd's message is then conveyed by Pepper in real time. Their experiment showed response delays of 8.82s, 6.79s, 6.79s and 4.12s with one, two, four and eight workers, respectively.

In summary, while a considerable body of research in CPCS examines how to reduce the actual system delays or the workers' arrival time, little attention has been paid on how to mitigate the negative effects of extended waiting by augmenting CPCSs with different filler-interventions. Our study is aimed at filling this important gap in the literature.

Time Perception Models and Time Fillers

Psychologists have proposed several theoretical models about how humans estimate time intervals. The attentional allocation model (Zakay 1989) argues that temporal processing requires attentional resources. If more attention is devoted to time passage while waiting, then the perception of time is lengthened because it enhances the accumulation of temporal cues in a cognitive timer mechanism. One way to reduce the perception of time is to introduce non-temporal tasks during waiting, which help to draw cognitive processing resources away from timekeeping. As a result, when less resources are allocated to monitoring time, the accumulation of cues in the timer is distrupted, impeding one's ability to accurately perceive the time – thus leading to shorter perceived time. In this view, (Agarwal and Karahanna 2000) introduced the notion of Cognitive Absorption (CA), which is a state of deep involvement or a holistic experience of a person with the software. It comprises of five dimensions, which are firmly linked with the processing of time: temporal dissociation, focused immersion, heightened enjoyment, control, and curiosity. For instance, when an external non-temporal cue is presented to users while waiting, they cannot accurately register time passage (temporal dissociation), and their attention is shifted away from timekeeping (focused immersion), leading to shorter perceived time.

Long waiting times without any feedback can result in harmful emotions, such as frustration and stress (Schleifer and Amick III 1989). Therefore, HCI researchers have investigated the efficacy of different time fillers in situations when response delays are inevitable. In chatbots, the most popular strategy to inform that a response is in the making, is to make use of typing indicators (e.g., three animated dots or the message "X is typing") (Gnewuch et al. 2018). A closely related strategy is to acknowledge the person with a short text message ("let me think") before presenting the actual answer. These utterances – also known as conversational fillers—have already been shown to moderate the negative effects of waiting (Shiwa et al. 2008). Furthermore, progress indicators do not only help to lower the user's uncertainty about waiting (Osuna 1985) but also comfort long waiting times (Harrison, Yeo, and Hudson 2010; Kurusathianpong and Tangmanee 2018); their efficacy has also been studied for a highly dynamic and roughly estimated duration (Harrison et al. 2007) – as is the case with nonlinear delays in CPCS. Furthermore, during a waiting period when people cannot switch context to perform other tasks, one can introduce secondary tasks in those idle times to alleviate the user's frustration (Vaish et al. 2014). For instance, when people were asked to perform some non-temporal task while waiting, they perceived the time to be shorter (Hohenstein et al. 2016). Other researchers have also investigated the effects of short films about nature and urban settings on time perceptions; in this view, (Davydenko and Peetz 2017) found that participants who watched films showing manmade settings perceived the time shorter.

Most of the research in HCI regarding time perceptions mainly concentrated on the web (Nah 2004), mobile (Wang et al. 2021) or speech applications (Asthana, Singh, and Gupta 2015). Within human-chatbot interaction, we are only aware of one study that explicitly addressed long waiting times through conveying "remaining waiting time" and "queue status" information when a handover request is made to a human service representative (Wintersberger, Klotz, and Riener 2020). Within a CPCS context, the need to study the effects of time fillers is even more crucial due to the nonlinear and uncertain nature of delays.

Method

We conducted a between-subjects experiment on the Prolific crowdsourcing platform to study the influence of different filler types across three levels of delay (8, 16 and 32 s) and two task types (stress mitigation and IR) on the waiting experience. Specifically, we compared 48 different conditions: 2 x (*task types*) × 3 (*delay levels*) × 8 (*filler types*). We designed a chatbot for each task type. To better understand the implications of filler types, we simulated response delays and workers' responses, as well as restricted users to just ask prepared questions. The front-end of these bots was developed using the TickTalkTurk library (Qiu, Gadiraju, and Bozzon 2020) while the server application was developed using Flask. The procedures provided in this research were authorized by the Eindhoven University of Technology's ethics board.

Development of Two Simulated CPCSs

This study focuses on both stress management and information retrieval due to the prevalence of these tasks in CPCSs. For instance, Panoply (Morris 2015) and CoZ (Abbas et al. 2020b) provide on-demand emotional assistance whereas Chorus (Lasecki et al. 2013b) and Evorus (Huang, Chang, and Bigham 2018) provide help with IR-related tasks. The two tasks differ in terms of breadth and depth of focus, duration, and speed (Grudin and Jacques 2019). We set different levels of response delays based on a geometric sequence (of 8, 16 and 32s) that was also used in a classic study on user interface response delays (Butler 1983).

Ustad Bot We created the *Ustad* bot based on the Rogerian principle of active listening (Rogers and Farson 1957) to simulate stress management tasks with CPCS. It enables users to chat about their issues without giving specific advice or solutions for dealing with stress, and it has a set of prepared questions it can ask them. Ustad's responses were created using data from a recent project to construct a computerized motivational interview for stress management (Park et al. 2019). The dataset contains 220 statements that were evaluated by therapists using the Motivational Interviewing Skills Code (MISC) (Miller et al. 2003). Each templated response of Ustad was comprised of three short statements; a) an empathetic response to the user's problem; b) a high-level reflection to the user's problem; c) and a follow-up open question. To control for confounding factors, such as complexity of user questions and variability in response rates, we requested participants to role play the fictional character as precisely as possible when conversing with Ustad. The fictional character exemplifies an undergraduate university student who is constantly worried about her school performance. We built this character based on a YouTube video clip[1] that depicts a motivational interviewing session between a client and a counselor.

Talash Bot To simulate IR tasks with CPCS, we built the *Talash* bot, which was designed to answer IR queries. To control the complexity of user input, we provided custom keyboard options or quick replies that adhered to the cognitive complexity framework's recommendations (Kelly et al. 2015). We explain further how the Talash bot works in the Procedure section. The cognitive complexity framework is based on Bloom's taxonomy which distinguishes between

[1]https://bit.ly/2TtvW5x

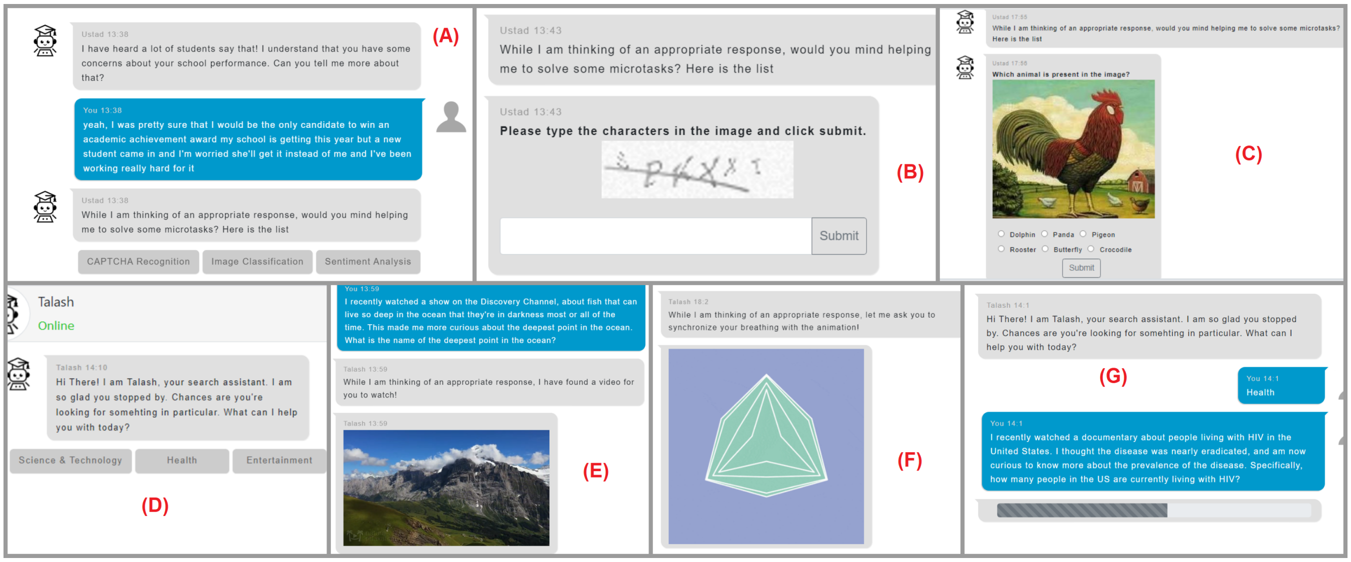

Figure 1: In this Figure we depict different filler types that we used. (A-C): This depicts an example of a microtask. While Ustad (bot) is thinking of an answer, it offers users to solve some microtask in three different categories (Fig. A). Fig. B depicts an example of a Captcha recognition task while Figure C depicts an example of an image classification task. (D-G) Talash (bot) initiates a discussion by offering users to choose pre-defined search tasks (Fig. D). After selecting a search task, some text is automatically generated based on the cognitive complexity framework (Kelly et al. 2015) and displayed in the text field (omitted from the figure); The user then reads and submits the text. Fig. E depicts an example of a nature filler, Fig. F depicts an example of a mindful breathing filler and finally, Fig. G represents an example of a progress bar. Other fillers are omitted.

six types of cognitive processes: *remember, understand, apply, analyze, evaluate*, and *create*. These cognitive processes are ordered based on the amount of effort required to execute them. For instance, remembering tasks are the simplest in IR, requiring the user to just identify or recognize a fact from an information source (e.g., “how many people in the US are currently living with HIV?”). We adopted simplest IR tasks because the emphasis was on confronting the negative users’ waiting experience with filler types; thus, we keep the complexity of user questions consistent for this study. On the basis of the dataset of reusable search tasks (Kelly et al. 2015), we chose three distinct remember tasks, one for each of the three domains: *Health, Science* & *Technology* and *Entertainment*. We meticulously created high quality responses of Talash based on the criteria of appropriateness and helpfulness (Xu et al. 2017). Full transcripts of a few sample dialogues of Ustad and Talash along with other project related details can be found at this link: **https://bit.ly/3DmQcbp**

Rationale for Choosing Different Filler Types

Activity Fillers The activity fillers allows end users to perform some activities when they are waiting for an answer. We included two types of activity fillers: Microtask filler and mindful breathing filler. Microtask fillers, as the name implies, enable end users to complete simple microtasks (e.g., image classification) while waiting for responses from the CPCS. We adopted this idea based on the notion of *low-effort crowdsourcing*, which leverages peoples’ peripheral attention to complete a secondary task that requires less cognitive demands while they are engaged with the primary task (Vaish et al. 2014). This notion is quite pertinent for CPCS where end-users sometimes have to wait a significant amount of time for an answer from the crowd. Thus, microtasks can be integrated into such periods of waiting to elicit relevant crowdsourced data while also acting as a remedy to mitigate the negative consequences of waiting. After a user enters a response and waits for an answer, the chatbot requests assistance in addressing some microtasks classified into three categories: CAPTCHA recognition, image classification and sentiment analysis (Fig. 1.A-C). The chatbot only asks this question at the beginning of each new turn and then it displays related microtasks in succession until an answer is received. For each new turn, the user is allowed to change type of microtask.

In the mindful breathing fillers, we show different breathing animations at each new turn and ask users to synchronize their breath to the rhythm of these animated breathing graphics (Fig. 1.F). Although breathing is an unconscious activity, research shows that if a person consciously controls his or her breathing, then it helps to curb the effects of stress (Russo, Santarelli, and O’Rourke 2017), which is very beneficial for the stress mitigation task. Nevertheless, mindful breathing can be equally useful for IR tasks where precise time to receive an answer is even more ambiguous, and thus can cause stress and frustration. When a user enters a response in this condition, the chatbot displays the breathing animation for the duration of the time delay specified.

Relaxation Fillers Relaxation fillers divert users’ attention away from the wait by playing relaxing movies. They include two types of videos: nature videos and urban videos.

Landscapes, mountains, woods, marine life, animals, and birds, as well as any other places containing aspects of living systems, are included in the nature videos. (Bratman, Hamilton, and Daily 2012) – Fig. 1.E. A number of studies have shown that exposure to nature triggers calming effects, stimulates positive mood and decreases stress (Davydenko and Peetz 2017; Moreno et al. 2018). Thus, the reason we included these videos as fillers was to determine whether a positive mood emerged from watching nature videos can help to moderate a negative user experience caused by long waiting in CPCS. In contrast, videos concerning urban or man-made settings show roads, buildings, bridges, traffic, and peoples. Recent research has shown that "time feels faster" when people are exposed to urban settings (Davydenko and Peetz 2017). When a user composes and delivers a response, the chatbot (Ustad or Talash) displays a nature or urban video for the given time delay. When the chatbot responds, the video fades out.

Animation Fillers Animation fillers are animated visual features that help to draw people's attention and provide aesthetic satisfaction. In this category, we chose (a) three animated ellipses (or three animated dots [...]) and (b) a ribbed progress bar. The concept of a typing indicator was established in human-chatbot conversation to support turn taking and increase the perceived social presence of chatbots (Gnewuch et al. 2018). The use of typing indicators ensures that the response is in the making and thus makes the interaction appear more natural with the chatbot (Cameron et al. 2018; Klopfenstein et al. 2017). This is especially true in CPCS where humans create and curate the response. Typing indications give users the sense of a real-time chat and can reduce waiting time.

A ribbed progress bar (Fig. 1.G) has a backward moving decelerating pattern that gives users the impression that the bar is moving faster than its actual speed (Harrison, Yeo, and Hudson 2010; Kurusathianpong and Tangmanee 2018). This type of progress bar has been shown to decrease the perceived waiting time when compared to other types of progress bars (Harrison, Yeo, and Hudson 2010; Kurusathianpong and Tangmanee 2018). Although the use of a progress bar is not ideal in scenarios where an exact time to receive a response is indefinite, researchers have studied the efficacy of different nonlinear progress bar behaviors (e.g., acceleration or deceleration) for a highly dynamic and roughly estimated durations (Harrison et al. 2007). The most favored nonlinear behavior was found to be the one where a progress bar was initiated gradually but accelerated towards the end of an operation to compensate the preliminary effects of waiting.

Conversational Fillers Conversational fillers or holding messages are fairly common in human discussions, such as "hm", "uh", or the longer "please give me a moment to think on what you have said" (Drummond and Hopper 1993). They demonstrate to the interlocutor that she is not being ignored, and that a response is likely to follow shortly. We compiled a list of conversational fillers and displayed them following the user's response. Conversational fillers have been successfully employed to mitigate user's frustration due to inevitable longer system response time in situations when actual humans are involved in generating a response (Shiwa et al. 2008).

No-Filler In this control condition, we did not show any filler but the rest of the settings were identical.

Participants

We recruited 931 workers (∼20 for each condition) from the Prolific.ac crowdsourcing platform. We restricted the experiment to only US and UK workers since our task required proficiency in English. Out of 930 unique workers, 65% were female, 33.9% were male, and 0.2% did not disclose their gender. 77.3% of the workers were from the UK and the rest were from USA. Their average age was 31.8 years old (SD=10.5). We estimated and paid £2.0 (£8.00/h) and £1.25 (£7.50/h) for the stress and IR task, respectively. The average reward per hour was 9.610 (SD = 1.089) for stress task and 9.663 (SD = 0.925) for IR task.

Procedure

(1) Consent Upon accepting the task, participants were asked to read and sign a consent form. This form explained that the chatbot is powered by hybrid intelligence, which means that answers are generated by a combination of computer and human input. For ethical reasons (Boden et al. 2017), participants were informed at the end of experiment that the chatbot's responses were scripted.

(2) Instructions Workers who read and accepted the consent form were redirected to a detailed instructions page. Workers from the stress management tasks were presented with the description of a fictional character that they have to role-play in the conversation. In case of IR tasks, workers were presented with instructions about how to interact with Talash along with the descriptions of the IR tasks.

(3) Conversational Task Workers were directed to the Ustad for the stress management task. The Ustad bot initiated a conversation by greeting users and inviting them to share their struggles. The user might then discuss their ostensible problems with Ustad, based on the fictional character's description. During the conversational task, workers could refer to the fictional character's description for clarification. After a user sent a response, a filler intervention was presented until a response was received (Fig. 1.A-D). Finally, after 4-5 dialogue exchanges, Ustad concluded the discussion with a piece of advice and thanking the users.

The Talash bot initiated the discussion by greeting the user and allowing them to seek information in three distinct domains, as previously described (Figure 1.D). Following this, users were allowed to select one domain out of the given three domains (*Health, Science & Technology and Entertainment*). After choosing the domain, the search task was automatically created and displayed to the users in the text area based on the dataset provided by (Kelly et al. 2015). Afterwards, users could read and send text. Then a filler intervention was shown during a preset delay. The link[2] to Ustad and Talash's demo videos is given for interested readers.

[2]https://bit.ly/3xkvl5a

(4) Exit Surveys Workers were then requested to complete an exit survey. The measures and hypotheses are discussed in the next section.

Measures & Hypothesis

Cognitive Absorption Stemming from the seminal flow literature (Czikszentmihalyi 1990), cognitive absorption (Agarwal and Karahanna 2000) is a well-grounded and reliable theory for understanding "a state of deep involvement with software" (p. 665). This theory has been successfully deployed in the past to predict and understand people's IT acceptance behaviors (Agarwal and Karahanna 2000). It comprises of five dimensions: (1) temporal dissociation, (2) focused attention, (3) heightened enjoyment, (4) control, and (5) curiosity. *Temporal dissociation* refers to an individual's inability to register the passage of time while interacting with the software. This dissociation can be achieved with an external distractor that can distract users from the waiting process and thus impedes their ability to follow the passage of time. *Focused immersion* is the experience of total engagement. If a person performing some activity is fully immersed with the technology, then the cognitive burden to accomplish the task is minimized. Thus, if a filler intervention provides some non-temporal cues to users, then less attention is paid to the waiting process and in turn their perception of time is reduced (Lee, Chen, and Ilie 2012). *Heightened enjoyment* captures the pleasurable aspects of an interaction with the software. When people experience enjoyment in performing an activity, they perceived the waiting time to be shorter (Lee, Chen, and Ilie 2012; Lee, Chen, and Hess 2017). *Control* represents a user's perception of being in charge of the interaction. Since CPCS waits are unpredictable and nonlinear in nature, augmenting CPCSs with time fillers can help users to feel in control. *Curiosity* is the extent with which the experience arouses an individual's sensory and cognitive curiosity. All constructs were measured on a seven-point scale (1: "strongly disagree", 2: "strongly agree"). The composite reliability coefficients of the five cognitive absorption dimensions are sufficiently reliable ranging from .83 to .93. Based on these considerations, we form the following hypotheses:

> **H1**: During waiting in a CPCS with a filler, users perceive more temporal dissociation (H1a), focused immersion (H1b), Heightened enjoyment (H1c), control (H1d) and curiosity (H1e) than a CPCS without a filler.

Furthermore, we hypothesize that allowing users to perform some non-temporal work, such as exercising mindful breathing or performing microtasks, creates more cognitive absorption with CPCS as compared to filler types where users merely observe something on the screen – watching a video or animation. We can name these fillers as 'passive' fillers. For instance, (Hohenstein et al. 2016) tested the efficacy of different loading screens on the perceived waiting time. They found that an interactive animation where users were asked to "swing the cradle" while waiting resulted in shorter perceived waiting time and greater satisfaction than passive animations. Thus, we propose the following:

> **H2**: During waiting in a CPCS with an activity filler, users perceive more temporal dissociation (H2a), focused immersion (H2b), Heightened enjoyment (H2c), control (H2d) and curiosity (H2e) than CPCSs with a passive filler type.

Perceived Waiting Time (PWT) We measured perceived waiting time using both an open-ended question and using cognitive appraisal of the wait. We asked users to give an estimate of the total time (in seconds) they had spent waiting between responding to the system and receiving the bot's response. We measured cognitive appraisal based on the perception of time spent in terms of long or short judgement (Pruyn and Smidts 1998). It is measured on a five-point scale (1: 'very short' , 5: 'very long'). It can be expected that a CPCS with filler intervention will lead to a shorter waiting time. As a result, we hypothesize that:

> **H3**: During waiting in a CPCS with a filler, users perceive the waiting time shorter than a CPCS without a filler.

Furthermore, a number of scholars (Lee, Chen, and Ilie 2012; Lee and Chen 2019; Lee, Chen, and Hess 2017) provided empirical evidence concerning the inverse relationship between the constructs of cognitive absorption (e.g., temporal dissociation, focused immersion etc.) and perceived waiting time. If we are able to find a similar relationship, then this study can inform a stopping criterion for the upper limit of an acceptable waiting time when augmenting CPCS with a particular filler type. Thus, we hypothesize that:

> **H4**: More temporal dissociation (H4a), focused immersion (H4b), Heightened enjoyment (H4c) and control (H4d) leads to shorter perceived time in CPCS.

Analyses and Results

A three-way MANOVA was run with three independent variables – *filler type, delay, and task type* – and dependent variables concerning *Perceived Waiting Time* (time in seconds and short/long judgment) and *Cognitive Absorption* (five variables). To control for Type-I error inflation in our multiple comparisons, we use the Bonferroni correction for family-wise error rate (FWER), at the significance level of $p < .05$. We found a statistically significant interaction effect between *filler type* and *task type* on the combined dependent variables, $F(56, 4825.034) = 1.52, p = .008, Wilks'\Lambda = 0.910, \eta_p^2 = .013$. We also found a statistically significant interaction effect between *filler type* and *delay* on the combined dependent variables, $F(112, 6287.891) = 1.51, p < .001, Wilks'\Lambda = 0.830, \eta_p^2 = .023$. Follow up univariate two-way ANOVAs were run. There was a statistically significant interaction effect between *filler type* and *task type* for *focused immersion* score, $F(7, 902) = 2.492, p = .015, \eta_p^2 = .019$, and for *cognitive* score, $F(7, 902) = 4.304, p < .001, \eta_p^2 = .032$. We also found a statistically significant interaction effect between *filler type* and *delay* for *focused immersion* score, $F(14, 902) = 2.894, p < .001, \eta_p^2 = .043$, and for *temporal dissociation* score, $F(14, 902) = 2.244, p = .005, \eta_p^2 = .034$.

Ref.	Dependent variable	Task Type	Filler A	Filler B	Mean A	Mean B	Mean Difference	Std. Error	Sig.	95% Confidence Interval	
										Lower bound	Upper bound
Fig.2.B	Cognitive	IR	Microtask	Convo. Filler	1.97	2.61	-.642	.166	.003	-1.16	-.122
				Animated ellipses	1.97	2.55	-.583	.167	.014	-1.10	-.061
				Progress	1.97	3.03	-1.06	.167	.000	-1.59	-.543
				Nature	1.97	2.93	-.960	.168	.000	-1.49	-.433
				Mindful breathing	1.97	2.52	-.550	.167	.029	-1.07	-.026
				No-filler	1.97	2.82	-.850	.167	.000	-1.37	-.326
Fig.2.C	Focused Immersion	Stress	Mindful breathing	Nature	4.60	3.20	1.40	.312	.000	.428	2.38
				Manmade	4.60	2.46	2.14	.313	.000	1.16	3.12
				Microtask	4.60	3.57	1.03	.307	.023	.070	1.99
			Convo. Filler	Nature	4.28	3.20	1.08	.314	.017	.096	2.07
				Manmade	4.28	2.46	1.82	.315	.000	.830	2.81
			Microtask	Manmade	3.57	2.46	1.11	.310	.010	.140	2.08
			Animated ellipses	Manmade	4.08	2.46	1.62	.313	.000	.639	2.60
			Progress	Manmade	4.18	2.46	1.72	.315	.000	.733	2.71
			No-filler	Nature	4.27	3.20	1.07	.339	.048	.005	2.13
				Manmade	4.27	2.46	1.81	.340	.000	.739	2.87
Fig.2.D	Focused Immersion	IR	Microtask	Manmade	4.09	3.01	1.08	.319	.021	.081	2.08
			Progress	Manmade	4.15	3.01	1.14	.319	.010	.141	2.14
			Mindful breathing	Manmade	4.17	3.01	1.16	.319	.008	.164	2.16

Table 1: Pairwise comparisons between different filler types across two task types

Next, we examined the simple main effects for *filler type* at each level of *task type* for *cognitive* and *focused immersion* scores. There was a statistically significant difference between different filler types for the IR task on the cognitive score, $F(7, 902) = 8.615, p < .001, \eta_p^2 = .063$, but not for the stress task, $F(7, 902) = .292, p = .957, \eta_p^2 = .002$. Note that low scores corresponding to the cognitive delay suggest that the response delay was perceived to be shorter. When compared to other fillers, using a microtask filler resulted in much less perceived delay (Cf. Fig.2.B). The pairwise comparisons are listed in Table 1 (1st row). Marginal means across the two task types for the PWT can be seen in the Figure 3. In case of microtask filler, the mean PWT score for 32s condition was only 18.5 and 9.3 for the stress and IR task, respectively.

Regarding the focused immersion score, we found a statistically significant difference between different filler types for the stress task, $F(7, 902) = 10.053, p < .001, \eta_p^2 = .072$, and also for the IR task, $F(7, 902) = 3.802, p < .001, \eta_p^2 = .029$. For the stress task, mindful breathing filler resulted in greater focused immersion than other fillers (Cf. Fig.2.C); The pairwise comparisons are listed in Table 1 (2nd row). In case of the IR task, microtask, progress and mindful breathing fillers caused more focused immersion than manmade filler (Fig.2.D). Table 1 (3rd row) shows the pairwise comparisons. Although, mindful breathing filler created more focused immersion, surprisingly, the PWT did not compress too much for 32s delay; the mean PWT was 25.8 and 35.1 for the stress and IR tasks, respectively (Fig. 3). Nevertheless, for 16s delay, it helped to reduce the PWT down to 8.7s in the IR task.

Next, we examined the simple main effects for *filler type* at each level of *delay* for the focused immersion and temporal dissociation. There was a statistically significant difference between different filler types for the 8s delay on the focused immersion, $F(7, 902) = 6.427, p < .001, \eta_p^2 = .048$, for 16s delay, $F(7, 902) = 5.852, p < .001, \eta_p^2 = .043$ and for 32s delay on the focused immersion, $F(7, 902) = 4.946, p < .001, \eta_p^2 = .037$. Simple comparisons revealed that for an 8s delay, conversational, animated ellipses and no-filler were comparatively more effective for focused immersion (Cf. Fig.4.A1). Pairwise comparisons can be seen in Table 2. Concerning PWT, the marginal means (Fig. 3) revealed that only mindful breathing and no-filler fillers were able to reduce the PWT time convincingly across two task types for an 8s delay while animated ellipses and conversational fillers were effective for IR and stress tasks, respectively. For a 16s delay, conversational filler, progress bar, mindful breathing filler and microtask filler were comparatively more helpful for stimulating focused immersion (Cf. Fig.4.A2). For pairwise comparisons, see Table 2. For 32s delay, progress bar, microtask and mindful breathing fillers created more focused immersion (Cf. Fig.4.A3 & Table 2).

Next, we examined the simple main effects for the *filler type* at each level of *delay* for the temporal dissociation. There was a statistically significant difference between different filler types for the 8s delay, $F(7, 902) = 2.435, p = .018, \eta_p^2 = .019$, for 16s delay, $F(7, 902) = 2.876, p = .006, \eta_p^2 = .022$ and for 32s delay on the temporal dissociation, $F(7, 902) = 3.779, p < .001, \eta_p^2 = .028$. In the 8s condition, conversational filler and nature filler caused more temporal dissociation (Cf. Fig.4.B1 & Table 2). However, regarding PWT, the conversational filler was only effective for the stress task (Fig.3). In the 16s condition, microtask filler and mindful breathing filler instigated more temporal dissociation while conversational filler created smaller temporal dissociation (Cf. Fig.4.B2 & Table 2). In the 32s condition, microtask filler caused more temporal dissociation than other fillers (Cf. Fig.4.B3 & Table 2).

Although, we did not find a difference in the Heightened Enjoyment (HE) scores across conditions from MANOVA test, simple three-factorial ANOVA revealed simple main

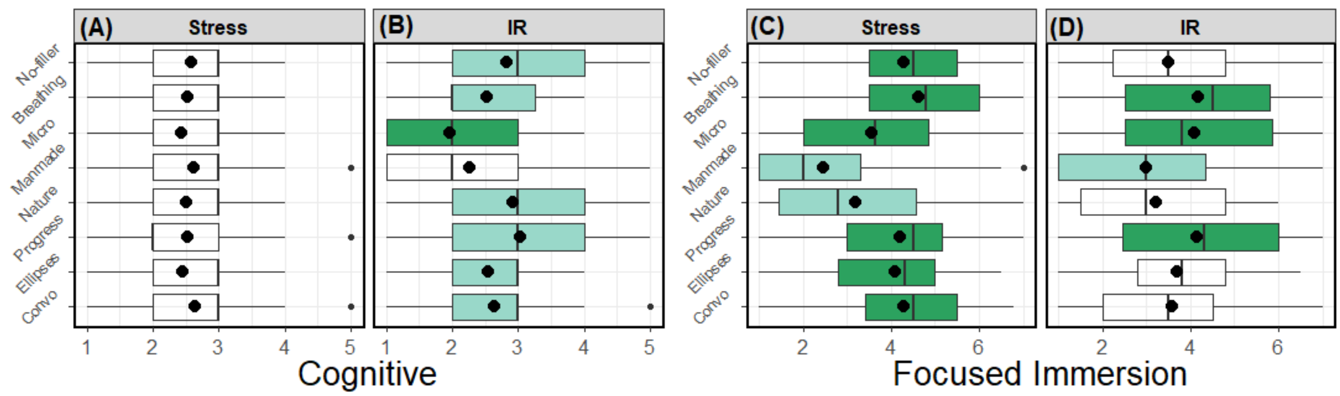

Figure 2: **(A-B)**: These figures depict an interaction effect between *task type* and *filler type* for the cognitive score (short/long wait judgement). Results indicate that for an IR task (B), participants in the microtask condition significantly perceived waiting time to be shorter than other fillers. **(C-D)**: These figures depict an interaction effect between *task type* and *filler type* for the focused immersion scores. In the stress task (C), mindful breathing filler resulted in more focused immersion than other fillers, while progress, microtask and mindful breathing filler resulted in more focused immersion in the IR task (D). Dark green fillers (or dark gray in black and white) function substantially better than light green fillers (or light gray in black and white). The mean scores are represented by black dots on the box plots.

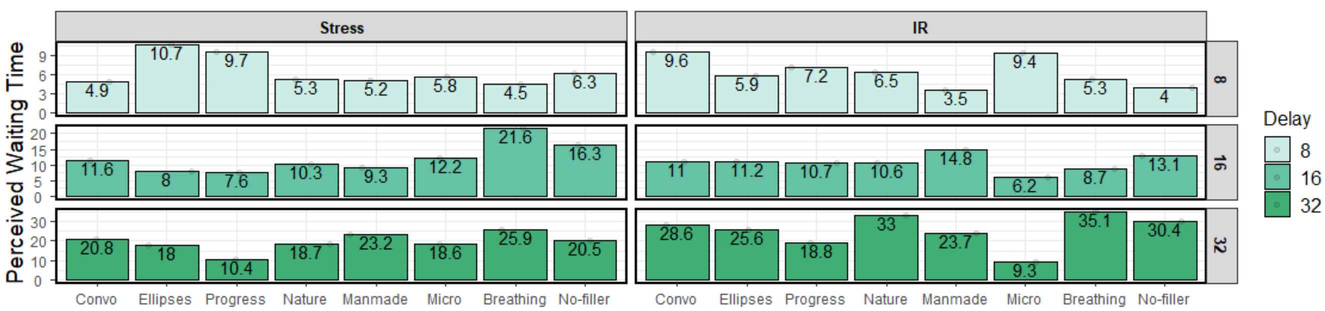

Figure 3: Marginal Means for perceived waiting time across two task types for three levels of delay

effects of *Filler type* for the heightened enjoyment scores ($F(7,910) = 6.283, p < .001, \eta_p^2 = .046$). Mindful breathing filler caused more enjoyment than any other fillers. Simple pairwise comparisons revealed that there was a significant difference in the mean HE scores between mindful breathing and conversational, Progress, nature, manmade and microtask filler – Fig 5.

Regarding **H4**, a linear regression established that only the temporal dissociation (H4a) could statistically predict the perceived waiting time, $F(5,944) = 3.300, p = .006$, which accounted for 1.7% of the explained variability in perceived waiting time. This relationship was inverse as depicted by linear equation ($PWT = 19.28 + (-1.68xTD)$). The overview of the results can be seen in Table 3.

Discussion, Implications & Conclusions

The main focus of a majority of prior work relevant to CPCSs has been to demonstrate their feasibility, while assessment studies have largely concentrated on optimizing response latency of CPCSs without exploring how lengthy delays can impact the experience of users. The aim of this study was to alleviate potentially unpleasant user experiences due to long waits in CPCSs, by employing different low-cost and freely accessible filler-interventions.

Our findings suggest that using a microtask filler in CPCS, particularly in task-oriented conversations, can be an effective way to mitigate the negative effects of waiting. A potential explanation for the effectiveness of microtask fillers in the IR tasks is their relatively less mentally demanding nature, when compared to engaging in a multi-turn and continuous conversation with an agent to relieve stress. As a result, Talash's users may have been able to divert their focus away from the primary task and fully immerse themselves in the microtasks while waiting, resulting in a shorter perceived duration (Zakay 1989). We also found that rhythmic breathing and progress fillers were both beneficial in engaging users, as observed from the focused immersion scores across both task types. This could be attributed to the hedonic appeal of these fillers (Agarwal and Karahanna 2000). The breathing filler however, proved to be ineffective in reducing the feeling of elapsed time during excessively lengthy delays in task focused conversations. Thus, when responses are expected to arrive later in an IR conversation,

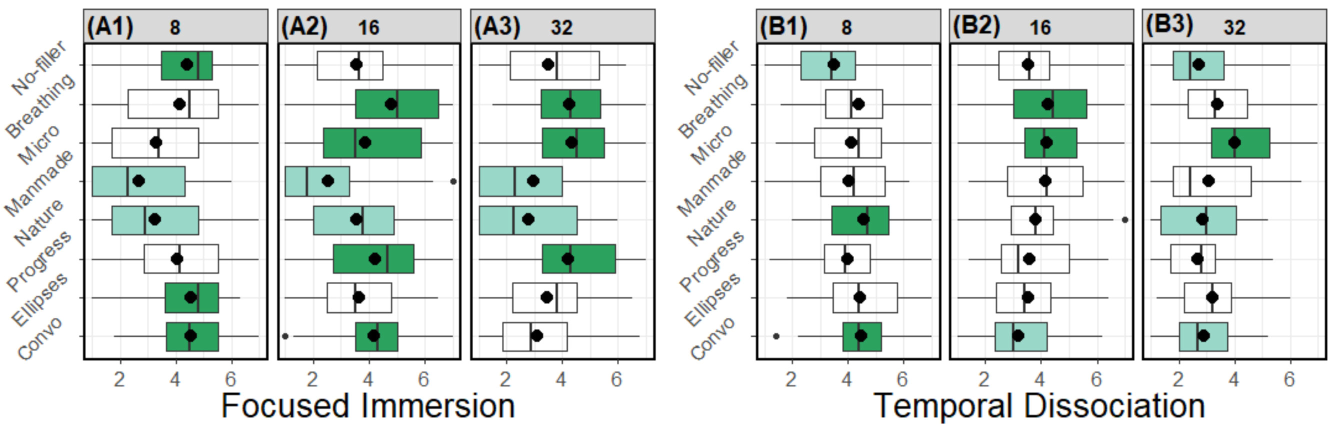

Figure 4: **(A1-A3)**: These figures depict an interaction effect between *filler type* and *delay* for the focused immersion scores. For the 8s condition (A1), conversational, animated ellipses, and no-filler provided more focused immersion than other fillers; for the 16s condition (A2), conversational, progress, microtask, and mindful breathing provided more focused immersion; and for the 32s condition (A3), progress, microtask, and mindful breathing provided more focused immersion. **(B1-B3)**: These figures depict an interaction effect between *filler type* and *delay* for the temporal dissociation scores. For the 8s condition (B1), conversational and nature fillers caused more temporal dissociation than no-filler; for the 16s condition (B2), microtask and mindful breathing fillers caused more temporal dissociation than conversational filler; and for the 32s condition (B3), microtask caused more temporal dissociation than other fillers. Dark green fillers (or dark gray in black and white) function substantially better than light green fillers (or light gray in black and white). The mean scores are represented by black dots on the box plots.

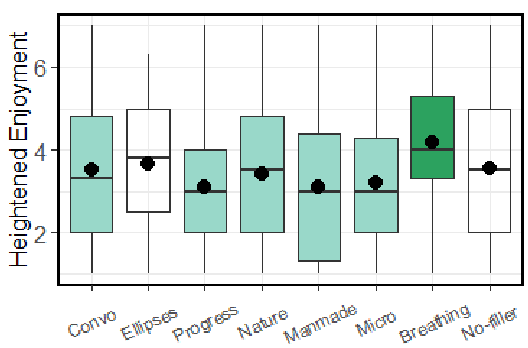

Figure 5: Simple main effects of *Filler type* for the Heightened Enjoyment scores. Mindful breathing filler caused more enjoyment than other fillers.

such as in complex search tasks, the use of breathing fillers should be avoided.

This research also suggests that the use of conversational fillers could be effective in the event of a brief inquiry, when responses are expected to arrive swiftly from the crowd (within 8s). This finding is in line with a prior study (Shiwa et al. 2008) that investigated the usefulness of conversational fillers in a communication robot controlled by human operation. When there are moderate delays, such as during a therapeutic conversational task, a breathing filler can help to divert users and reduce their sense of time. For longer delays, such as in case of complex search tasks, both breathing or microtask fillers can be more effective. However, using a microtask filler for a stress task should be handled with caution, since it can have negative consequences, especially if the conversational task is cognitively taxing or if a person seeking help is under a lot of stress. To address this, one possible solution is to include relevant microtasks, such as asking users to recognize and classify their own stressful situation as either maladaptive or real (also known as cognitive reappraisal (Morris, Schueller, and Picard 2015)), or asking them to match their own inquiry to a set of predefined intents. In this way, we can acquire vital crowdsourcing data for model training to benefit future hybrid intelligent CPCSs, while simultaneously diverting users' attention away from the delay. Additionally, in a stress task, fillers, such as animated ellipses and conversational fillers, produced similar results in terms of focused immersion. One reason for this could be that people opted to watch passive fillers in the stress task so that they can focus only on alleviating their stress rather than undertaking secondary tasks. Finally, both man-made and nature fillers performed worse regarding increasing focused immersion in both tasks. One possible reason is that watching movies while waiting in CPCS may be perceived as a complete distraction and should thus be avoided.

While we studied these fillers in a carefully controlled environment with fixed delays, they may be equally suitable for dynamic delays encountered in real-world scenarios. For example, the anticipated arrival time of a message from a particular worker can be computationally calculated using her typing speed, average response time, average response length, total number of responses, number of her responses up-voted/accepted by other workers, the complexity of the user's inquiry, among other factors (Burlutskiy et al. 2015). Thus, when the message's arrival time is short, more traditional fillers, such as conversational fillers or animated el-

Ref.	Dependent variable	Delay Level	Filler A	Filler B	Mean A	Mean B	Mean Difference	Std. Error	Sig.	95% Confidence Interval	
										Lower bound	Upper bound
Fig.4.A1-A3	Focused Immersion	8s	Convo. Filler	Nature	4.54	3.26	1.28	.385	.025	.076	2.49
				Manmade	4.54	2.70	1.84	.388	.000	.627	3.06
				Microtask	4.54	3.31	1.24	.377	.030	.055	2.42
			Animated Ellipses	Nature	4.52	3.26	1.26	.378	.025	.076	2.45
				Manmade	4.52	2.70	1.82	.381	.000	.627	3.01
				Microtask	4.52	3.31	1.22	.370	.030	.055	2.38
			No-filler	Manmade	4.39	2.70	1.69	.385	.000	.478	2.89
		16s	Convo. Filler	Manmade	4.16	2.52	1.64	.385	.001	.436	2.85
			Progress	Manmade	4.23	2.52	1.71	.385	.000	.506	2.92
			Microtask	Manmade	3.86	2.52	1.34	.385	.015	.133	2.55
			Mindful breathing	Manmade	4.79	2.52	2.27	.383	.000	1.07	3.47
				Nature	4.79	3.55	1.24	.383	.036	.036	2.44
		32s	Progress	Nature	4.22	2.81	1.41	.386	.008	.203	2.62
			Microtask	Manmade	4.33	2.99	1.34	.391	.018	.111	2.56
				Nature	4.33	2.81	1.52	.383	.002	.318	2.72
			Mindful breathing	Manmade	4.26	2.99	1.27	.391	.036	.039	2.49
				Nature	4.26	2.81	1.45	.383	.005	.245	2.65
Fig.4.B1-B3	Temporal Dissociation	8s	Convo. Filler	No-filler	4.51	3.49	1.02	.320	.044	.013	2.02
			Nature	No-filler	4.55	3.49	1.06	.318	.026	.059	2.05
		16s	Microtask	Convo. Filler	4.22	3.16	1.06	.318	.025	.064	2.06
			Mindful breathing	Convo. Filler	4.22	3.16	1.06	.316	.023	.069	2.05
		32s	Microtask	Convo. Filler	4.01	2.94	1.07	.314	.020	.081	2.05
				Progress	4.01	2.68	1.33	.320	.001	.323	2.33
				Nature	4.01	2.86	1.16	.318	.008	.158	2.15
				No-filler	4.01	2.71	1.30	.320	.001	.299	2.31

Table 2: Pairwise comparisons between different filler types within each delay level

Hypos.	Summary	Ref.
H1: During waiting in a CPCS with a filler, users perceive more temporal dissociation (H1a), focused immersion (H1b), Heightened enjoyment (H1c), control (H1d) and curiosity (H1e) than a CPCS without a filler.	**H1a**, **H1b** and **H1c** were supported for different filler types. For both task types, **mindful breathing** and **progress** fillers caused more focused immersion. More pleasure was also induced by the **mindful breathing** filler.	Fig.2.(C-D)
	When considering Focused Immersion (FI) and Temporal Dissociation (TD), and Perceived Waiting Time (PWT) collectively, **H1a** and **H1b** were supported for different levels of delays. For instance, **conversational** fillers were effective for FI, TD and PWT; for 16s delays, **microtask**, **mindful breathing*** were more useful, whilst for 32s delay, **microtask** filler performed well.	Fig.4.(A1-A3), Fig.4.(B1-B3)
H2: During waiting in a CPCS with an activity filler, users perceive more temporal dissociation (H2a), focused immersion (H2b), Heightened enjoyment (H2c), control (H2d) and curiosity (H2e) than CPCSs with a passive filler type.	**H2a**, **H2b** and **H2c** were supported. Considering FI, TD and PWT together, both **mindful breathing** and **microtask** fillers were effective for moderate (16s) and long delays (32s).	Fig.4 (A2-A3, B2-B3), Fig.2 (C-D)
H3: During waiting in a CPCS with a filler, users perceive the waiting time shorter than a CPCS without a filler.	It was partially supported for a **microtask** filler for IR task.	Fig.2.B
H4: More temporal dissociation (H4a), focused immersion (H4b), Heightened enjoyment (H4c) and control (H4d) leads to shorter perceived time in CPCS.	Only **H4a** was supported	

Table 3: Summary of hypotheses & Findings

lipses can be used. In the event of extended delays, the CPCS can be augmented with activity fillers to help mitigate the consequences of prolonged waiting. Additionally, given the well-established relationship between shorter wait times and increased satisfaction (Hoxmeier and DiCesare 2000), service providers can combine our concept with existing computational techniques (e.g., pre-recruiting, queuing etc.) to further increase satisfaction.

Limitations & Future Work

A limitation of the study concerns the nature of user responses. These were restricted to choosing between template-based utterances rather than actually responding. While this approach is adopted for reasons of ease of use in many chatbot applications (Jain et al. 2018), it is also not universally applicable especially in the cases where a free form conversation is intended, as is the case in cases of dealing with stress or other personal problems. Given the restricted nature of text entry in our experiments, our results may not easily generalize to cases of extended free form text input, where perhaps a largely varying time to compose an input to the system or even the content of the system response may influence the expectations of the user regarding the response delay. Future studies could address these and especially if the type of time-filler selected for different cases is considered suitable in different context and for different types of conversation.

References

Abbas, T.; Khan, V.-J.; Gadiraju, U.; Barakova, E.; and Markopoulos, P. 2020a. Crowd of Oz: A Crowd-Powered Social Robotics System for Stress Management. *Sensors* 20(2). ISSN 1424-8220. doi:10.3390/s20020569. URL https://www.mdpi.com/1424-8220/20/2/569.

Abbas, T.; Khan, V.-J.; Gadiraju, U.; and Markopoulos, P. 2020b. Trainbot: A Conversational Interface to Train Crowd Workers for Delivering On-Demand Therapy. *Proceedings of the AAAI Conference on Human Computation and Crowdsourcing* 8(1): 3–12. URL https://ojs.aaai.org/index.php/HCOMP/article/view/7458.

Agarwal, R.; and Karahanna, E. 2000. Time Flies When You're Having Fun: Cognitive Absorption and Beliefs about Information Technology Usage. *MIS Quarterly* 24(4): 665–694. ISSN 02767783. URL http://www.jstor.org/stable/3250951.

Ashktorab, Z.; Jain, M.; Liao, Q. V.; and Weisz, J. D. 2019. Resilient Chatbots: Repair Strategy Preferences for Conversational Breakdowns. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, CHI '19, 1–12. New York, NY, USA: Association for Computing Machinery. ISBN 9781450359702. doi:10.1145/3290605.3300484. URL https://doi.org/10.1145/3290605.3300484.

Asthana, S.; Singh, P.; and Gupta, P. 2015. Survival Analysis: Objective Assessment of Wait Time in HCI. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, CHI '15, 367–376. New York, NY, USA: Association for Computing Machinery. ISBN 9781450331456. doi:10.1145/2702123.2702428. URL https://doi.org/10.1145/2702123.2702428.

Bernstein, M. S.; Brandt, J.; Miller, R. C.; and Karger, D. R. 2011. Crowds in Two Seconds: Enabling Realtime Crowd-Powered Interfaces. In *Proceedings of the 24th Annual ACM Symposium on User Interface Software and Technology*, UIST '11, 33–42. New York, NY, USA: Association for Computing Machinery. ISBN 9781450307161. doi:10.1145/2047196.2047201. URL https://doi.org/10.1145/2047196.2047201.

Bigham, J. P.; Jayant, C.; Ji, H.; Little, G.; Miller, A.; Miller, R. C.; Miller, R.; Tatarowicz, A.; White, B.; White, S.; and Yeh, T. 2010. VizWiz: Nearly Real-Time Answers to Visual Questions. In *Proceedings of the 23nd Annual ACM Symposium on User Interface Software and Technology*, UIST '10, 333–342. New York, NY, USA: Association for Computing Machinery. ISBN 9781450302715. doi:10.1145/1866029.1866080. URL https://doi.org/10.1145/1866029.1866080.

Boden, M.; Bryson, J.; Caldwell, D.; Dautenhahn, K.; Edwards, L.; Kember, S.; Newman, P.; Parry, V.; Pegman, G.; Rodden, T.; Sorrell, T.; Wallis, M.; Whitby, B.; and Winfield, A. 2017. Principles of robotics: regulating robots in the real world. *Connection Science* 29(2): 124–129. doi:10.1080/09540091.2016.1271400. URL https://doi.org/10.1080/09540091.2016.1271400.

Bratman, G. N.; Hamilton, J. P.; and Daily, G. C. 2012. The impacts of nature experience on human cognitive function and mental health. *Annals of the New York Academy of Sciences* 1249(1): 118–136. doi:https://doi.org/10.1111/j.1749-6632.2011.06400.x. URL https://nyaspubs.onlinelibrary.wiley.com/doi/abs/10.1111/j.1749-6632.2011.06400.x.

Burlutskiy, N.; Fish, A.; Ali, N.; and Petridis, M. 2015. Prediction of Users' Response Time in Q amp;A Communities. In *2015 IEEE 14th International Conference on Machine Learning and Applications (ICMLA)*, 618–623. doi:10.1109/ICMLA.2015.190.

Butler, T. W. 1983. Computer response time and user performance. In *Proceedings of the SIGCHI conference on Human Factors in Computing Systems*, 58–62.

Cameron, G.; Cameron, D.; Megaw, G.; Bond, R.; Mulvenna, M.; O'Neill, S.; Armour, C.; and McTear, M. 2018. Best practices for designing chatbots in mental healthcare–A case study on iHelpr. In *Proceedings of the 32nd International BCS Human Computer Interaction Conference 32*, 1–5.

Czikszentmihalyi, M. 1990. *Flow: The psychology of optimal experience*. New York: Harper & Row.

Davydenko, M.; and Peetz, J. 2017. Time grows on trees: The effect of nature settings on time perception. *Journal of Environmental Psychology* 54: 20–26. ISSN 0272-4944. doi:https://doi.org/10.1016/j.jenvp.2017.09.003. URL https://www.sciencedirect.com/science/article/pii/S0272494417301081.

Drummond, K.; and Hopper, R. 1993. Acknowledgment tokens in series. *Communication Reports* 6(1): 47–53. doi:10.1080/08934219309367561. URL https://doi.org/10.1080/08934219309367561.

Gnewuch, U.; Morana, S.; Adam, M.; and Maedche, A. 2018. "The Chatbot is typing ..." – The Role of Typing Indicators in Human-Chatbot Interaction. In *SIGHCI - The 17th Annual Pre-ICIS Workshop on HCI Research in MIS*.

Grudin, J.; and Jacques, R. 2019. Chatbots, Humbots, and the Quest for Artificial General Intelligence. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, CHI '19, 1–11. New York, NY, USA: Association for Computing Machinery. ISBN 9781450359702. doi:10.1145/3290605.3300439. URL https://doi.org/10.1145/3290605.3300439.

Haas, D.; Wang, J.; Wu, E.; and Franklin, M. J. 2015. CLAMShell: Speeding up Crowds for Low-Latency Data Labeling. *Proc. VLDB Endow.* 9(4): 372–383. ISSN 2150-8097. doi:10.14778/2856318.2856331. URL https://doi.org/10.14778/2856318.2856331.

Harrison, C.; Amento, B.; Kuznetsov, S.; and Bell, R. 2007. Rethinking the Progress Bar. In *Proceedings of the 20th Annual ACM Symposium on User Interface Software and Technology*, UIST '07, 115–118. New York, NY, USA: Association for Computing Machinery. ISBN 9781595936790. doi:10.1145/1294211.1294231. URL https://doi.org/10.1145/1294211.1294231.

Harrison, C.; Yeo, Z.; and Hudson, S. E. 2010. Faster Progress Bars: Manipulating Perceived Duration with Visual Augmentations. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '10, 1545–1548. New York, NY, USA: Association for Computing Machinery. ISBN 9781605589299. doi:10.1145/1753326.1753556. URL https://doi.org/10.1145/1753326.1753556.

Hohenstein, J.; Khan, H.; Canfield, K.; Tung, S.; and Perez Cano, R. 2016. Shorter Wait Times: The Effects of Various Loading Screens on Perceived Performance. In *Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems*, CHI EA '16, 3084–3090. New York, NY, USA: Association for Computing Machinery. ISBN 9781450340823. doi:10.1145/2851581.2892308. URL https://doi.org/10.1145/2851581.2892308.

Hornik, J. 1984. Subjective vs. objective time measures: A note on the perception of time in consumer behavior. *Journal of consumer research* 11(1): 615–618.

Hoxmeier, J. A.; and DiCesare, C. 2000. System response time and user satisfaction: An experimental study of browser-based applications. *AMCIS 2000 Proceedings* 347.

Huang, T.-H.; Lasecki, W.; Azaria, A.; and Bigham, J. 2016. "Is There Anything Else I Can Help You With?": Challenges in Deploying an On-Demand Crowd-Powered Conversational Agent. *Proceedings of the AAAI Conference on Human Computation and Crowdsourcing* 4(1): 79–88. URL https://ojs.aaai.org/index.php/HCOMP/article/view/13292.

Huang, T.-H.; Lasecki, W.; and Bigham, J. 2015. Guardian: A Crowd-Powered Spoken Dialog System for Web APIs. *Proceedings of the AAAI Conference on Human Computation and Crowdsourcing* 3(1). URL https://ojs.aaai.org/index.php/HCOMP/article/view/13237.

Huang, T.-H. K.; Chang, J. C.; and Bigham, J. P. 2018. Evorus: A Crowd-Powered Conversational Assistant Built to Automate Itself Over Time. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, CHI '18, 1–13. New York, NY, USA: Association for Computing Machinery. ISBN 9781450356206. doi:10.1145/3173574.3173869. URL https://doi.org/10.1145/3173574.3173869.

Huang, T. K.; Azaria, A.; Romero, O. J.; and Bigham, J. P. 2019. InstructableCrowd: Creating IF-THEN Rules for Smartphones via Conversations with the Crowd. *Hum. Comput.* 6: 113–146. doi:10.15346/hc.v6i1.7. URL https://doi.org/10.15346/hc.v6i1.7.

Jain, M.; Kumar, P.; Kota, R.; and Patel, S. N. 2018. Evaluating and Informing the Design of Chatbots. In *Proceedings of the 2018 Designing Interactive Systems Conference*, DIS '18, 895–906. New York, NY, USA: Association for Computing Machinery. ISBN 9781450351980. doi:10.1145/3196709.3196735. URL https://doi.org/10.1145/3196709.3196735.

Kelly, D.; Arguello, J.; Edwards, A.; and Wu, W.-c. 2015. Development and Evaluation of Search Tasks for IIR Experiments Using a Cognitive Complexity Framework. In *Proceedings of the 2015 International Conference on The Theory of Information Retrieval*, ICTIR '15, 101–110. New York, NY, USA: Association for Computing Machinery. ISBN 9781450338332. doi:10.1145/2808194.2809465. URL https://doi.org/10.1145/2808194.2809465.

Klopfenstein, L. C.; Delpriori, S.; Malatini, S.; and Bogliolo, A. 2017. The Rise of Bots: A Survey of Conversational Interfaces, Patterns, and Paradigms. In *Proceedings of the 2017 Conference on Designing Interactive Systems*, DIS '17, 555–565. New York, NY, USA: Association for Computing Machinery. ISBN 9781450349222. doi:10.1145/3064663.3064672. URL https://doi.org/10.1145/3064663.3064672.

Kurusathianpong, P.; and Tangmanee, C. 2018. Comparison of Perceived Waiting Time Between Two Lengths of Progress Indicator and Two Styles of Graphics Animation With Perceived Uncertainty as a Covariate. In *2018 Seventh ICT International Student Project Conference (ICT-ISPC)*, 1–6. doi:10.1109/ICT-ISPC.2018.8523993.

Lasecki, W. S.; Homan, C.; and Bigham, J. P. 2014. Architecting real-time crowd-powered systems. *Human Computation* 1(1).

Lasecki, W. S.; Thiha, P.; Zhong, Y.; Brady, E.; and Bigham, J. P. 2013a. Answering Visual Questions with Conversational Crowd Assistants. In *Proceedings of the 15th International ACM SIGACCESS Conference on Computers and Accessibility*, ASSETS '13. New York, NY, USA: Association for Computing Machinery. ISBN 9781450324052. doi:10.1145/2513383.2517033. URL https://doi.org/10.1145/2513383.2517033.

Lasecki, W. S.; Wesley, R.; Nichols, J.; Kulkarni, A.; Allen, J. F.; and Bigham, J. P. 2013b. Chorus: A Crowd-Powered Conversational Assistant. In *Proceedings of the 26th Annual ACM Symposium on User Interface Software and Technology*, UIST '13, 151–162. New York, NY, USA: Association for Computing Machinery. ISBN 9781450322683. doi:10.1145/2501988.2502057. URL https://doi.org/10.1145/2501988.2502057.

Lee, Y.; and Chen, A. N. 2019. The effects of progress cues and gender on online wait. *Decision Support Systems* 123: 113070. ISSN 0167-9236. doi:https://doi.org/10.1016/j.dss.2019.05.010. URL https://www.sciencedirect.com/science/article/pii/S0167923619300995.

Lee, Y.; Chen, A. N. K.; and Ilie, V. 2012. Can Online Wait Be Managed? The Effect of Filler Interfaces and Presentation Modes on Perceived Waiting Time Online. *MIS Quarterly* 36(2): 365–394. ISSN 02767783. URL http://www.jstor.org/stable/41703460.

Lee, Y. G.; Chen, A. N.; and Hess, T. 2017. The online waiting experience: Using temporal information and distractors to make online waits feel shorter. *Journal of the Association for Information Systems* 18(3): 1.

Miller, W. R.; Moyers, T. B.; Ernst, D.; and Amrhein, P. 2003. Manual for the motivational interviewing skill code (MISC). *Unpublished manuscript. Albuquerque: Center on Alcoholism, Substance Abuse and Addictions, University of New Mexico* .

Moreno, A.; Baker, S.; Varey, K.; and Hinze-Pifer, R. 2018. Bringing attention restoration theory to the classroom: A tablet app using nature videos to replenish effortful cognition. *Trends in Neuroscience and Education* 12: 7–21. ISSN 2211-9493. doi:https://doi.org/10.1016/j.tine.2018.07.002. URL https://www.sciencedirect.com/science/article/pii/S2211949318300036.

Morris, R. R.; Schueller, S. M.; and Picard, R. W. 2015. Efficacy of a Web-Based, Crowdsourced Peer-To-Peer Cognitive Reappraisal Platform for Depression: Randomized Controlled Trial. *J Med Internet Res* 17(3): e72. ISSN 1438-8871. doi:10.2196/jmir.4167. URL https://doi.org/10.2196/jmir.4167.

Morris, R. R. R. 2015. *Crowdsourcing mental health and emotional well-being*. Ph.D. thesis, Massachusetts Institute of Technology.

Nah, F. F.-H. 2004. A study on tolerable waiting time: how long are Web users willing to wait? *Behaviour & Information Technology* 23(3): 153–163. doi:10.1080/01449290410001669914. URL https://doi.org/10.1080/01449290410001669914.

Osuna, E. E. 1985. The psychological cost of waiting. *Journal of Mathematical Psychology* 29(1): 82–105. ISSN 0022-2496. doi:https://doi.org/10.1016/0022-2496(85)90020-3. URL https://www.sciencedirect.com/science/article/pii/0022249685900203.

Park, S.; Choi, J.; Lee, S.; Oh, C.; Kim, C.; La, S.; Lee, J.; and Suh, B. 2019. Designing a chatbot for a brief motivational interview on stress management: Qualitative case study. *Journal of medical Internet research* 21(4): e12231.

Pruyn, A.; and Smidts, A. 1998. Effects of waiting on the satisfaction with the service: Beyond objective time measures1Both authors contributed equally to this article.1. *International Journal of Research in Marketing* 15(4): 321 – 334. ISSN 0167-8116. doi:https://doi.org/10.1016/S0167-8116(98)00008-1. URL http://www.sciencedirect.com/science/article/pii/S0167811698000081.

Qiu, S.; Gadiraju, U.; and Bozzon, A. 2020. TickTalkTurk: Conversational Crowdsourcing Made Easy. In *Conference Companion Publication of the 2020 on Computer Supported Cooperative Work and Social Computing*, 1–5.

Riek, L. D. 2012. Wizard of Oz Studies in HRI: A Systematic Review and New Reporting Guidelines. *J. Hum.-Robot Interact.* 1(1): 119–136. doi:10.5898/JHRI.1.1.Riek. URL https://doi.org/10.5898/JHRI.1.1.Riek.

Rogers, C. R.; and Farson, R. E. 1957. *Active listening*. Chicago, IL.

Russo, M. A.; Santarelli, D. M.; and O'Rourke, D. 2017. The physiological effects of slow breathing in the healthy human. *Breathe* 13(4): 298–309. ISSN 1810-6838. doi:10.1183/20734735.009817. URL https://breathe.ersjournals.com/content/13/4/298.

Schleifer, L. M.; and Amick III, B. C. 1989. System response time and method of pay: Stress effects in computer-based tasks. *International Journal of Human-Computer Interaction* 1(1): 23–39.

Shiwa, T.; Kanda, T.; Imai, M.; Ishiguro, H.; and Hagita, N. 2008. How quickly should communication robots respond? In *2008 3rd ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, 153–160. doi:10.1145/1349822.1349843.

Tognazzini, B. 1993. Principles, Techniques, and Ethics of Stage Magic and Their Application to Human Interface Design. In *Proceedings of the INTERACT '93 and CHI '93 Conference on Human Factors in Computing Systems*, CHI '93, 355–362. New York, NY, USA: Association for Computing Machinery. ISBN 0897915755. doi:10.1145/169059.169284. URL https://doi.org/10.1145/169059.169284.

Vaish, R.; Organisciak, P.; Hara, K.; Bigham, J.; and Zhang, H. 2014. Low Effort Crowdsourcing: Leveraging Peripheral Attention for Crowd Work. *Proceedings of the AAAI Conference on Human Computation and Crowdsourcing* 2(1). URL https://ojs.aaai.org/index.php/HCOMP/article/view/13191.

Wang, Y.; Huang, Y.; Li, J.; and Zhang, J. 2021. The effect of mobile applications' initial loading pages on users' mental state and behavior. *Displays* 68: 102007. ISSN 0141-9382. doi:https://doi.org/10.1016/j.displa.2021.102007. URL https://www.sciencedirect.com/science/article/pii/S0141938221000214.

Wintersberger, P.; Klotz, T.; and Riener, A. 2020. Tell Me More: Transparency and Time-Fillers to Optimize Chatbots' Waiting Time Experience. In *Proceedings of the 11th Nordic Conference on Human-Computer Interaction: Shaping Experiences, Shaping Society*, NordiCHI '20. New York, NY, USA: Association for Computing Machinery. ISBN 9781450375795. doi:10.1145/3419249.3420170. URL https://doi.org/10.1145/3419249.3420170.

Xu, A.; Liu, Z.; Guo, Y.; Sinha, V.; and Akkiraju, R. 2017. A New Chatbot for Customer Service on Social Media. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, CHI '17, 3506–3510. New York, NY, USA: Association for Computing Machinery. ISBN 9781450346559. doi:10.1145/3025453.3025496. URL https://doi.org/10.1145/3025453.3025496.

Zakay, D. 1989. Subjective Time and Attentional Resource Allocation: An Integrated Model of Time Estimation. In Levin, I.; and Zakay, D., eds., *Time and Human Cognition: A Life-Span Perspective*, volume 59 of *Advances in Psychology*, 365–397. North-Holland. doi:https://doi.org/10.1016/S0166-4115(08)61047-X. URL https://www.sciencedirect.com/science/article/pii/S016641150861047X.