The Pennsylvania State University

The Graduate School

HM AND YES? HOW BACK-CHANNELING TRANSFORMS SMART-HOME DEVICES FROM PASSIVE AGENTS TO BE ACTIVE LISTENERS.

A Thesis in

Information Sciences and Technology

by

Nasim Motalebi

Submitted in Partial Fulfillment
of the Requirements
for the Degree of

Master of Science

December 2023

The thesis of Nasim Motalebi was reviewed and approved by the following:

Saeed Abdullah
Assistant Professor in Information Sciences and Technology Thesis Advisor

Benjamin Hanrahan
Assistant Professor in Information Sciences and Technology

Carleen Maitland
Professor in Information Sciences and Technology

Jeffrey Bardzell
Professor in Information Sciences and Technology
Associate Dean for Faculty and Graduate Affairs

ABSTRACT

Smart Speakers assistants such as Amazon Alexa have become pervasive tools that support a variety of tasks including information search and controlling smart home devices. However, current use of smart-speakers is limited to episodic interactions (e.g., answering single query) that do not span more than a couple of user turn-takings. Moreover, user's turn in talk is limited to short commands (mostly less than a minute); forcing the user to be positioned as a listener in most interactions. Such limitations narrow the use of smart-speakers to command driven interactions which reduces users' turn in talk and as a result, reduces conversational engagement with smart-speakers.

In this project, we aim to extend the capabilities of smart-speakers to support more engaging interactions with users. Towards this goal, we propose transforming smart-speakers to "Active-listeners". More specifically, we have integrated the use of back-channeling (e.g., 'hm', 'yes', 'umm') in smart-speakers to extend user's turn in talk and to sustain interactions that contain multiple-turns with arbitrary durations. We believe the use of back-channeling will result in users perceiving these devices as active listeners and hence improving overall engagement and disclosure intimacy. Such improvement in engagement has potential impact across a number of application domains including effective and scalable delivery of mental health support such as self-centered therapy.

Keywords

Conversational Agents, Human-Computer Interaction, Active Listening, Emotional Well-being.

TABLE OF CONTENTS

LIST OF FIGURES

ACKNOWLEDGEMENTS

This thesis remains evidence to my learning and commitment to human-centered design and research. I thank Dr. Saeed Abdullah for his patience, understanding, and teaching. My Masters education was a learning curve with many contributions to my future professional career and choices.

I also contribute this thesis to my family and friends, near and far, who encouraged me to take steps in unknown scientific territories.

Chapter 1

Introduction

Introduced by James W Pennebaker in 1988, expressive writing is a method similar to talk-therapy in which disclosing information about traumatic events or emotional upheavals improves patients' physiological and psychological wellbeing (Pennebaker, 1988). Expressive writing has had modest health effects, evaluated through objective behavior and physiological measures over several weeks, especially in non-clinical populations (Bond & Pennebaker, 2012). Recently there has been an abundant of studies that demonstrate the benefits of expressive writing on emotion control and mental wellbeing in populations with different symptoms such as PTSD (Bugg, Turpin, Mason, & Scholes, 2009), anxiety (Pennebaker, 1992), depression disorder (Krpin et al., 2013), and mood disorder (Baikie et al., 2012). Nevertheless, expressive writing requires iterative practice; otherwise the observable change in patient's behavior will be insignificant. Hence the presence of a therapist is mostly required to make sure the writing is maintained and is being effective. Yet the logistical barriers (e.g. location, costs, time, lack of resource and therapists, and etc.) and social stigma (e.g fear of being judged) against having a therapist reduces the impact of therapeutic practices such as expressive-writing. Therefore, researchers in the mental health community are in pursuit of finding alternative solutions for distributed mental healthcare that would minimalize the role of a therapist, but keeps the patient engaged in maintaining therapy.

Emerging technologies such as web interfaces, mobile technologies, and conversational agents (chatbots, smart home speakers, virtual assistants, etc.) have recently been used as tools to deliver distributed healthcare among larger populations. So far, such technologies have enabled

personalized therapy in shape of psycho-education, skill development, just in time intervention, tracking patient's progress, and help patients maintain therapy in long term.

Woebot is one of the most recent chatbots designed to deliver therapeutic practices for those suffering from depression and anxiety. Focused on psycho-education, Woebot offers a more engaging interface by incorporating decision trees to maintain conversations with the user. However, decision trees are mostly designed to lead conversations towards a certain predicted end-goal within a pre-defined structure. Such structures do not allow unpredicted user responses, limiting them to provide short answers to specific questions. Such systems do not allow users to freely talk, nor do they support active listening to their users. In this study, I argue for conversational agents as opportunities to create artificial active-listeners that motivate self-disclosure- what I coin as "expressive speaking". Transforming existing conversational agents such as Amazon Alexa to become active-listeners for expressive talking could potentially eliminate obstacles such as the fear of judgment in therapeutic practices.

Chapter 2

Alexa As An Active Listener

Back-channeling being one of the major components of human-human conversation, has been proven to improve the perception of conversational agents as attentive listeners. Specifically, the use of verbal continuers such as *hm*, *uhum*, *aha*, *yeah*, and so on in an agent's response has demonstrated to enhance perceived human-likeness of human-agent conversations. In this study we have specifically defined an active listener as follow:

However, the effects of back-channeling in enhancing user-agent engagement, levels of information disclosure, disclosure intimacy, and emotional well-being is yet to be studied. Only then we can argue for the effectiveness of conversational agents in enhancing levels of information disclosure, disclosure intimacy, and emotional wellbeing.

In this study I have developed and evaluated a custom designed Alexa Skill that supports active listening through simple verbal acknowledgments- back-channeling cues. "Active Listening"- is a Skill designed to transform Alexa to an active listener. To evaluate Alexa as an active listener and users' expressive speaking behavior towards Alexa, I have conducted a pilot study in which Alexa's usability and effectiveness as an active listener is measured through closed ended surveys and open-ended questionnaires. In the surveys, I have asked the participants to rate their perception of Alexa as an active listener based on the definition by Oertel et al [8]:

a) An active listener is someone who pays attention, listens carefully, and is observant.

b) Is careful to fulfill the needs or wants of the speaker; is considerate about the speaker.

Further on in this paper I will explain the study design which includes an overview of the Active Listening, the pilot study, results, discussion, and future steps.

Chapter 3

Study Design

Currently commercialized conversational agents such as Amazon Alexa rarely support any back-channeling feedback. Even if they do, it is not used in the context of free talk and expressive speaking. Therefore, the advantages or disadvantages of using back-channeling to support expressive speaking in the context of smart-home devices is unknown. To better investigate this issue, I have developed an Alexa Skill prototype called "*Active Listening*" which is accessible through any Alexa device. The Active Listening Skill is used to understand the effectiveness of back-channeling in promoting expressive speaking. In a pilot study with 4 participants, the performance and usability of Alexa as an active listener is evaluated. The results of this pilot study can inform future experiments that would help us understand the feasibility of using Alexa as a therapeutic agent that can promote user's disclosure intimacy and consequently, emotional well-being.

Developing an Alexa Skill

The Skill development process includes defining the front end and the back end of the Alexa Skill. The front end of the skill pertains the user's requests and the user-agent interaction schema. Amazon Skills Kit (ASK) services is provided by Amazon's Developer Console to computationally define the front end. A skill developer must know user's intents and potential requests and define Alexa responses appropriately; therefore, designing an interaction model prior to developing the skill is essential. In order to define appropriate responses from Alexa, one needs to be familiar with Alexa's Lambda function which is accessible through Amazon Web Services

(AWS). The Lambda function connects the Alexa Skill to the back-end where all the language and speech processes take place. By using the Lambda function, the Skill designer can define Alexa's actions and appropriate speech responses for user's specific requests. Lambda currently supports a variety of coding languages such as node.js and python. Due to my familiarity with Java based languages, I decided to use node.js to implement the Active Listening Skill in Lambda.

The Lambda function responds to three major Intents; *LaunchRequest*, *StartIntent*, *TalkIntent*.

LaunchRequest: The user would ask Alexa to launch the skill. Alexa would open the skill by speaking out the welcome message.

StartIntent: The user will confirm to continue with the skill. In the case of active listening, the user would acknowledge that they would like to talk to Alexa. Alexa would prompt the user to talk by telling them that she will be listening.

TalkIntent: The user will talk to Alexa. In return Alexa would detect the speech and randomly provide back-channeling feedback to prompt users to talk more.

CloseIntent: The user can stop talking or ask alexa to stop listening. Alexa Would speak out a closing sentence and shut down.

"Alexa, open Active Listening"

To activate an Alexa Skill, the skill must have an invocation name. The invocation name for the *Active Listening* skill is simply "Active Listening". For example, in order to start the skill, the user must ask Alexa "*Alexa, open Active Listening*". Using the invocation name, the user will activate Alexa to listen while they talk about literally anything. In this particular skill, Alexa would prompt users to talk more by providing verbal cues such as *hm*, *aha, oh,* and *hah.* A screenshot of the Active Listening Skill from the Alexa app is shown in figure 1.

By initiating the skill, Alexa would introduce herself and would ask the user whether they would like to talk to her about something? This introduction is based on the interaction model designed for the Active Listening Skill, and its purposed towards self-centered therapy. Based on a hypothetical user response, the initial launch request and interaction script is as follow:

User: *Alexa, open Active Listening.*

Alexa: *Hi there! My name is Alexa and I am here to listen to you. Have you ever talked to a conversational agent before?*

User: *No, not really…*

Alexa: *Well... I think talking to me about any emotional or negative experience would help you feel better. So, go ahead and tell me something about what's bothering you.*

[…]

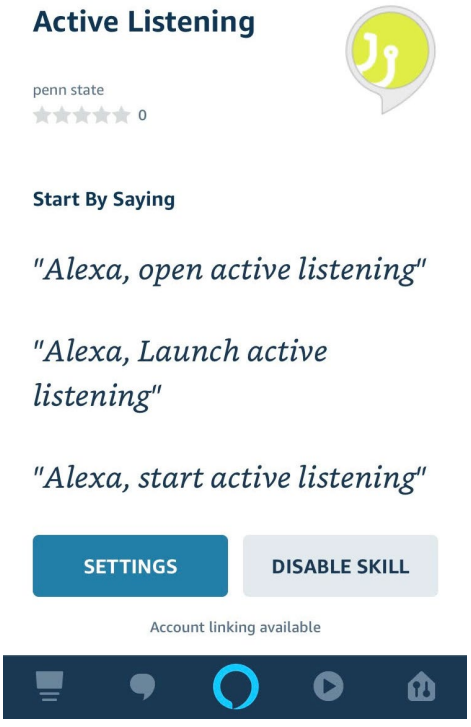

Figure 1 Active Listening Skill from the Alexa app

From this point and beyond, the skill is designed so that it prompts the user to talk more by providing verbal back-channeling cues. In this case, Alexa's responses are totally randomized to pull out a back-channeling response from an array of potential back-channeling phrases, including moments of silence to make the interaction more natural. Alexa would then only prompt the user when she recognized a pause in user's speech. Therefore, any interruption in user's turn in talk is avoided to assure conversation flow with no attempt in stealing the floor from the user. To better understand this dynamic, a hypothetical conversation with Alexa is as follow; which is a continuum to the prior launch request:

User: *My life has always been a mess…My mum has a sever lung disease and I'm her full-time care giver, only child and have to live with her because we both cannot afford to live sparely… (Pause)*

Alexa: *ooh…*

User: *I went to so much trouble in conciliation to get professional help from professors at my university. I got put down by the insurance guy... (Pause)*

Alexa: *hm…*

[…]

The user can stop or ask Alexa to quit the skill at any time. If the user stays silent for too long, Alexa would ask the user if there is anything else s/he would like to share. If not, Alexa would close up the session with a final prompt:

[…]

Alexa: *I hope talking to me helped you feel better. Let me know if you would like to talk to me more.*

Having a prototype ready, I deployed a user-centered design approach to test and evaluate the performance of the skill in order to better perfect the skill for later experiments on disclosure intimacy and emotional well-being. This first design iteration and experiment was meant to better understand user's experience in interacting with Alexa, evaluate conversational or technical

challenges, user's expectations and satisfaction with the skill, and their perception of Alexa as an Active listener. In the next section the details of the pilot study is demonstrated.

Pilot Study

Based on the qualitative nature of the pilot study, a small population would have sufficed to obtain design and performance feedback. The small number of participants allows a thorough analysis of user's experience and their input on how to advance the design of the system and the experiment at large. 4 participants were recruited through convenient sampling method; 4 IST graduate students volunteered to participate in the study. All but one student majors in Human Computer Interaction. Prior to the experiment, the participants were familiarized with the study procedure and were informed that they will be recorded by the researcher but not by Alexa.

Scripted vs Free Talk Interactions

One of the major challenges in laboratory experiments for human-computer interaction is to create a close to real experience. Talking about emotional upheavals or personal challenges would not come naturally in an experiment setting. In this particular experiment, Alexa would ask participants to talk about a negative emotional experience. The challenge is not only that the participants are not familiar with the interaction, but they are not comfortable to instantaneously share an emotional experience or to think of a situation to talk about. To address these challenges, each participant was asked to interact with the Active Listening Skill in two ways; a) Use a pre-written script and follow the script to talk to Alexa. b) freely talk to Alexa about whatever they would like to share or talk about. The scripted context was meant to not only familiarize them with the skill and set participant's expectations from the Skill, but also help them to think about

personal challenges or everyday problems to talk about. Moreover, the scripted interaction could provide a baseline to evaluate the skill's pure technical performance. Both interactions were recorded per participant and the participants were asked to fill out a survey on Alexa's performance as an active listener and their experience for both scripted and free talk interactions.

Survey and Measures

To collect participant's feedback on their experience with the Active Listening Skill, they were asked to fill out a survey that included both closed and open-ended questions. The questions were meant to collect measurable data through Likert Scale ratings and qualitative data through descriptive questionnaires. The survey was constructed to address:

a) Measures of usability and interaction quality by using a 7-point Likert Scale; some of which were inspired by the work of Gearhart et al. and Lala et al.

b) Measures of active listening using a 10-point Likert Scale; defined by Oertel et al. [8].

c) Open ended questions on user expectations, satisfaction, and acceptability towards the idea of Alexa as a potential counselor:

- What did you expect from Alexa during your conversation?
- Did Alexa meet your expectations?
- How can Alexa be a better listener in your opinion?
- Would you talk to Alexa as a counselor? Why or why not?

The survey included a total of 22 questions. However, the open-ended questions were only asked once after the entire experiment was completed by each participant. The 18 remaining questions were asked to be filled out separately after each interaction experience (scripted and open ended).

Chapter 4

Results

The results form individual participants and the average ratings for measures of usability (from a 7-point Likert scale) and active listening (from a 10-point Likert scale) are graphed for both the scripted and free talk interactions. The survey questions are summarized to single items for conciseness. For example, the item "*Interruptions*" stands for Alexa's performance on being least interruptive in the conversation. A higher rate for this item specifically means that participants felt less interrupted during their conversation with Alexa.

In the results section, I have included boxplots that demonstrate the average (Figure 2) and subject specific rankings (Figure 3) for usability items. The reason to include subject specific ranking is to outline differences in user's interactions. Some subjects had experiences specific to them that made the entire interaction less pleasant, hence lower rankings. For example, Subject 2 encountered a network connection issue with Alexa that made Alexa's performance less accurate and more erroneous. Therefore subject 2's satisfaction with the skill is minimal. Due to the small population size, the rankings from subject 2 have impacted the average performance rate of the Skill. On the other hand, subject 4 had a very fluid interaction experience with the skill which led to higher ratings. Thus, the average ratings may not represent the true nature of the interactions. Therefore, a qualitative analysis from the questionnaires and studying the interactions per subjects is necessary.

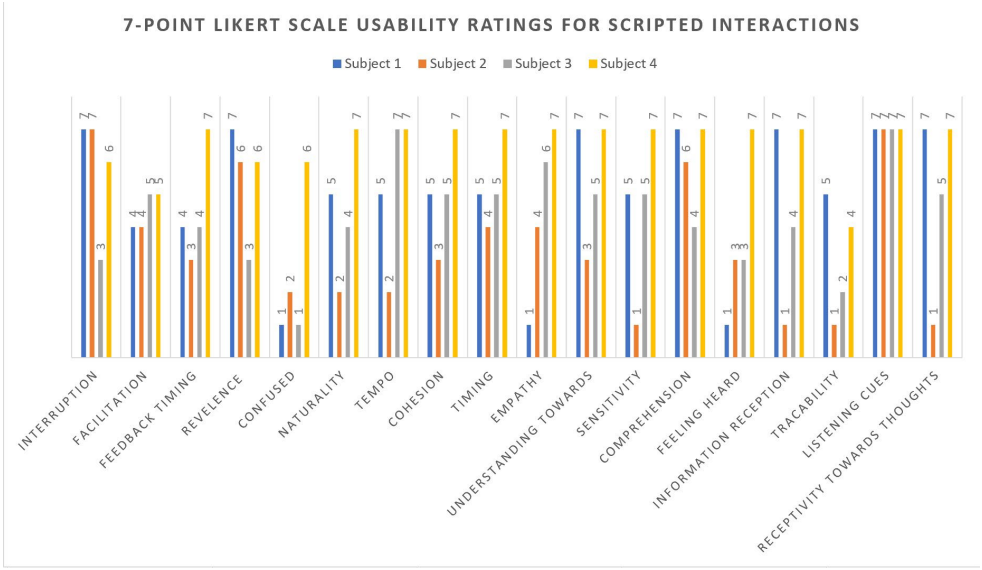

Figure 2Usability ratings for scripted interactions per subject

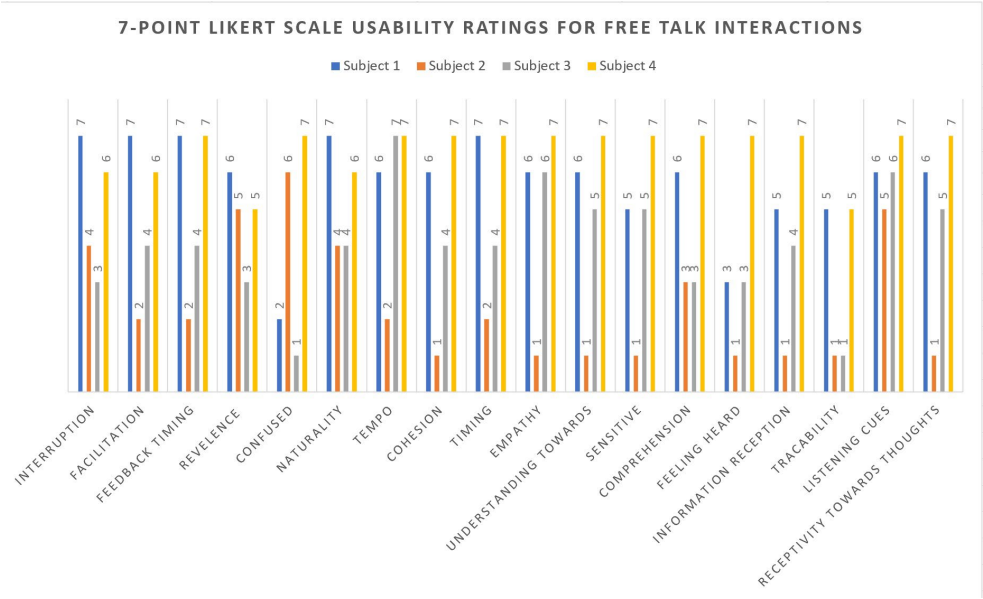

Figure 3 Usability ratings for free talk interactions per person

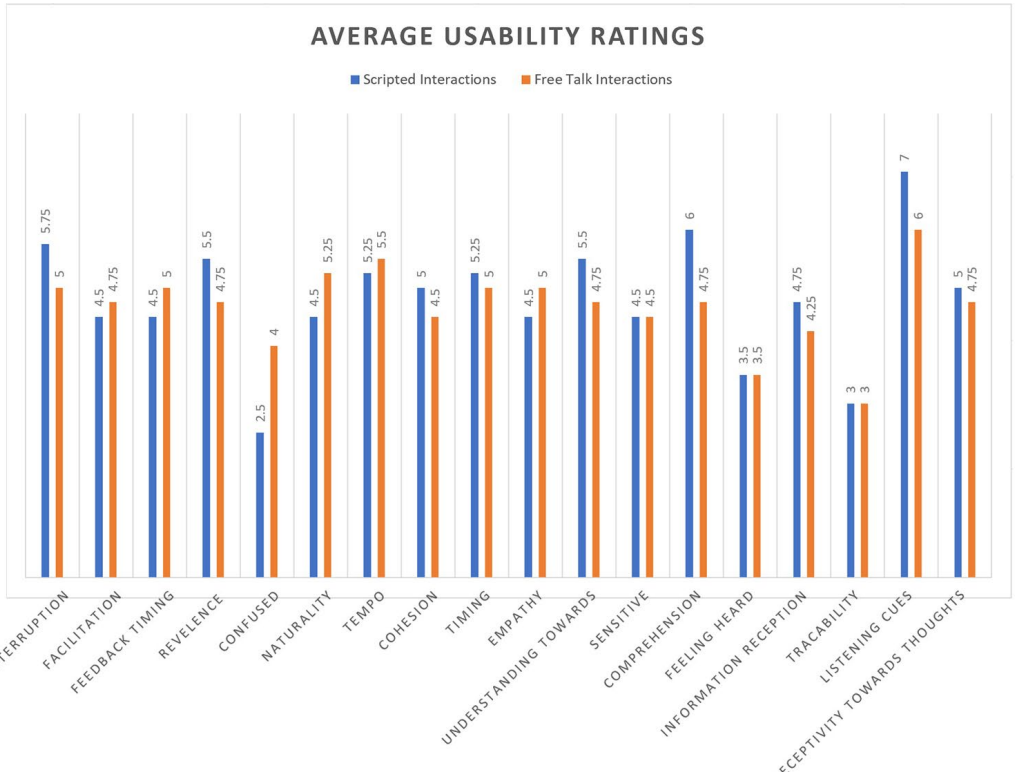

Figure 4 Average usability rankings, comparing scripted to free talk interactions.

Additionally, participants were asked to rate their perception of Alexa as an active listener. Figure 5 demonstrates their ratings per subject and on average. The average ranking for perceived active listening is above 5 out of 10 which overall is a promising first step.

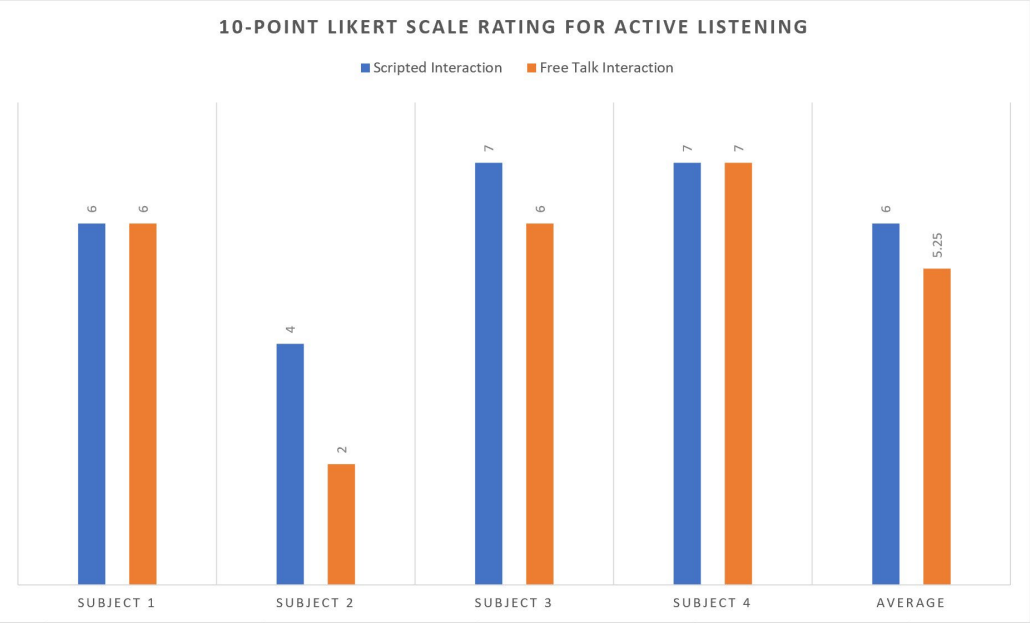

Figure 5 Ratings for perceived active listening per subject and on average for scripted and free talk interactions.

In the descriptive questionnaires, the participants were asked to state their initial expectations from Alexa as a conversational agent, their overall satisfaction from their interaction, their opinion for future development of Alexa, and Alexa as a potential counselor. Participants' responses can be summarized as follow:

Expectations: being empathetic, providing well-timed responses, ask questions and be more interactive rather than just acknowledging them, Tailored responses, giving solutions or suggestions, make them feel better.

Satisfaction: overall participants were somewhat satisfied with their interactions based on their expectations. Almost all participants acknowledged the well-timed verbal feedbacks from Alexa. However, Subject 3 did not see Alexa as an empathetic agent and Subject 2 was disappointed with Alexa since she didn't ask any questions, nor did she answer them.

Further development: Overall participants asked for a more empathetic agent who would not just acknowledge what they say but to provide tailored responses and verbally

sympathize with them. All participants agreed that Alexa could not understand what they are saying and they wished Alexa could provide responses relevant to what they were talking about.

Alexa as a counselor: Most participants agreed that they would not talk to Alexa as a counselor at this stage. However, if Alexa was to be designed to meet their expectations, they could see Alexa to become someone they could talk to.

The input received from the questionnaires lets us better interpret the measured performance and user experience ratings. The results can provide us with valuable insight onto Alexa's performance quality and strategies to advance Alexa to become an active listener. In the discussion section I have pointed out relevant outcomes that can inform future design iterations.

Chapter 5

Discussions

One of the major challenges is to minimize technical malfunctioning that may be caused by weak internet connection.

Based on the average ratings, Alexa received relatively high ratings on her performance in providing back-channeling cues. Demonstrated in figure 4, the following measures for both scripted (S) and free-talk (FT) interaction indicate that Alexa has performed mostly above average in responding with appropriate listening cues and supporting verbal continuers of back-channeling:

- No interruptions (S: 5.75, FT:5)
- Timing of acknowledgments (S:4.5, FT:5)
- Relevance of the cues (S:5.5, FT:4.75)
- Naturality of the cues (S:4.5, FT:5.25)
- Appropriate speech tempo (S:5.25, FT:5.5)
- Cohesion of the conversation (S: 5, FT: 4.5)
- Timing of the responses (S: 5.25, FT: 5)
- Being Understanding (S: 5.5, FT: 4.75)
- Comprehension of Alexa's responses (S: 6, FT: 4.75)
- Appropriate Listening cues (S: 7, FT: 6)

In addition, the rate for perceived active listening is above average which is a positive indication that using back-channeling cues would be effective to transform Alexa to an active listener:

- Perceived Active Listening (S:6, FT: 5.25)

Not surprising, the scripted interactions have higher scores as the participants were instructed to pause at appropriate times; allowing Alexa to provide an appropriate response at the right time. Also, it appears that the participants felt the back-channeling cues were more contextually appropriate in the scripted interaction. However, this could be due to a bias; participants being separated from the content would allow a less biased judgment versus when

they are sharing about their own emotional experiences. In addition, participants may have thought that the scripted interactions include scripted responses from Alexa. However, Alexa would generate random responses for both interaction methods.

Overall, subject 2's experience could be seen as an outlier due to technical difficulties with the internet network. Therefore, the ratings may not be truthful to the skill's ability in providing appropriate back-channeling cues. If we consider Subject 2's interaction as an outlier, the average measured outcomes would turn to become drastically higher; leaving us to believe the Active Listening skill has performed sufficiently in providing appropriate feedbacks. However, Alexa is certainly not perceived as an understanding, nor an empathetic agent towards what the participant's shared. This may affect user's expressive speaking behavior, therefore may reduce their level of disclosure intimacy and information disclosure. Therefore, it is suggested to design Alexa's responses to be more tailored to sound emotionally relevant to what user's are talking of.

Chapter 6

Conclusions and Future Work

Based on my results, the Active Listening Skill has succeeded in delivering neutral back-channeling cues and prompting expressive-speaking. Based on the interactions, participants rarely stopped or paused when talking to Alexa; which could be considered as a positive indication of the back-channeling feedbacks.

However, there were times of which the participants were confused due to Alexa's elongated response time. The long pause moments could be a result of the information processing time that Alexa needs to generate a response. Alexa is designed to handle short requests from users; which does not require heavy information processing. This is a major challenge to think of. Since Amazon does not provide much flexibility in term of adjusting the back-end code, the next step would be finding alternative solutions that would reduce perceived waiting time for the participants. One way to address this challenge is to conduct experiments in a close to real environment. Naturally users would not be sitting in front of Alexa to talk to it; rather they may be engaged in multiple activities such as cooking, doodling, scrolling the web, etc. Engaging users in multiple activities may distract them from the long waiting times.

Another issue rose from the lack of tailored responses from Alexa. Potentially predicting potential user responses and phrases such as "sad", "stressed", "happy", and so on can help us defined more tailored Alexa responses for predictable emotional situations. This could be done through ASK and AWS.

The ultimate goal for the Active Listening Skill is to enhance levels of disclosure intimacy and as a result emotional well-being through self-centered therapy. To evaluate Alexa's

effectiveness to disseminate self-centered therapy, first we have to make sure that the skill will not encounter any performance issues and that it can be perceived as an active listener. This pilot study is of significance as it informs the future design iterations; which is hoped to be pursued in Spring 2019.

Bibliography

1. Lopatovska, I., Rink, K., Knight, I., Raines, K., Cosenza, K., Williams, H., & Martinez, A. (2018). Talk to me : Exploring user interactions with the Amazon Alexa. https://doi.org/10.1177/0961000618759414
2. Ho, A., Hancock, J., & Miner, A. S. (2018). Psychological , Relational , and Emotional Effects of Self-Disclosure After Conversations With a Chatbot, 68(August), 712–733. https://doi.org/10.1093/joc/jqy026
3. Man, C. B. (1966). And Machine.
4. Bugg, A., Turpin, G., Mason, S., & Scholes, C. (2009). Behaviour Research and Therapy A randomised controlled trial of the effectiveness of writing as a self-help intervention for traumatic injury patients at risk of developing post-traumatic stress disorder. Behaviour Research and Therapy, 47(1), 6–12. https://doi.org/10.1016/j.brat.2008.10.006
5. Essay, I. (1993). INVITED ESSAY PUTTING STRESS INTO WORDS : HEAL TH , LINGUISTIC , AND THERAPEUTIC IMPLICATIONS, 31(6), 539–548.
6. Travagin, G., Margola, D., & Revenson, T. A. (2015). Clinical Psychology Review How effective are expressive writing interventions for adolescents ? A meta-analytic review. Clinical Psychology Review, 36, 42–55. https://doi.org/10.1016/j.cpr.2015.01.003
7. Bond, M., & Pennebaker, J. W. (2012). Computers in Human Behavior Automated computer-based feedback in expressive writing. Computers in Human Behavior, 28(3), 1014–1018. https://doi.org/10.1016/j.chb.2012.01.003
8. Oertel, C., Gustafson, J., & Black, A. W. (2016). Towards Building an Attentive Artificial Listener : On the Perception of Attentiveness in Feedback Utterances, 2915–2919.
9. Kawahara, T., Uesato, M., Yoshino, K., & Takanashi, K. (n.d.). Toward Adaptive Generation of Backchannels, 1–10.
10. Lee, N., & Lee, S. (n.d.). The Effect of Back-Channeling Cues on Motivation to Continue Human-Machine Textual Interaction.
11. Oertel, C., & Black, A. W. (n.d.). On Data Driven Parametric Backchannel Synthesis for Expressing Attentiveness in Conversational Agents.
12. Oertel, C., Mora, K. A. F., & Black, A. W. (n.d.). Towards Building an Attentive Artificial Listener On the Perception of Attentiveness in Audio-Visual Feedback Tokens, 21–28.
13. Buschmeier, H., & Kopp, S. (2017). Conversational Agents Need to Be ‘ Attentive Speakers ’ to Receive Conversational Feedback from Human Interlocutors.
14. Lala, D., Milhorat, P., & Inoue, K. (2017). Attentive listening system with backchanneling , response generation and flexible turn-taking, (August), 127–136.
15. Danby, S., Butler, C. W., & Emmison, M. (n.d.). When ‘ listeners can ’ t talk ’ sequences of telephone and online counselling, 91–114.
16. Gearhart, C. C., & Bodie, G. D. (2011). Active-Empathic Listening as a General Social Skill : Evidence from Bivariate and Canonical Correlations, 24(2), 86–98. https://doi.org/10.1080/08934215.2011.610731