
Can Alexa be your Therapist? How Back-Channeling Transforms Smart-Speakers to be Active Listeners

Nasim Motalebi

Pennsylvania State University
University Park, PA, USA
nfm5140@psu.edu

Eugene Cho

Pennsylvania State University
University Park, PA, USA
exc75@psu.edu

S. Shyam Sundar

Pennsylvania State University
University Park, PA, USA
sss12@psu.edu

Saeed Abdullah

Pennsylvania State University
University Park, PA, USA
saeed@psu.edu

ABSTRACT

Smart-speakers such as Amazon Alexa are becoming increasingly popular among the general population. These devices support a wide range of user-initiated tasks. However, their current interaction capabilities are limited to quick turn-takings and short dialogues, which leads to limited user engagement. In this project, we aim to enhance the user engagement and interaction abilities of a smart-speaker by transforming it into an *active listener*. Specifically, we explored how providing random back-channelling (i.e., verbal continuers including “hm”, “uhum”, “aha”, “yeah”) can result in longer interactions and more sustained user engagement. The findings of the study have implications

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).

CSCW '19 Companion, November 9–13, 2019, Austin, TX, USA

© 2019 Copyright held by the owner/author(s).

ACM ISBN 978-1-4503-6692-2/19/11.

<https://doi.org/10.1145/3311957.3359502>

for usability and future design of smart-speakers. We also explored how the enhanced interactions in smart-speakers through back-channeling can be used for self-centered therapy leading to better emotional and social support.

CCS CONCEPTS

• **Human-centered computing** → **Sound-based input / output**; *Empirical studies in interaction design*; Ubiquitous and mobile computing systems and tools.

KEYWORDS

Conversational Agents; Smart-speakers; Back-channeling; Active Listening; Wellbeing

ACM Reference Format:

Nasim Motalebi, Eugene Cho, S. Shyam Sundar, and Saeed Abdullah. 2019. Can Alexa be your Therapist? How Back-Channeling Transforms Smart-Speakers to be Active Listeners. In *2019 Computer Supported Cooperative Work and Social Computing Companion Publication (CSCW '19 Companion)*, November 9–13, 2019, Austin, TX, USA. ACM, New York, NY, USA, 5 pages. <https://doi.org/10.1145/3311957.3359502>

INTRODUCTION

The adoption rate of smart-speakers including Amazon Alexa and Google Home has seen significant rise in recent years. Approximately 21% of adults (53 million) in the United States now own one of these devices [9]. These devices use voice UI (VUI) to interact with users. For example, users can tell these devices to perform search queries or control smart appliances. Use of voice as an input modality has resulted in improved usability for these devices.

However, the current VUIs supported by these devices are optimized for short commands and quick turn-takings. This limits the types of interactions and engagements users can have with these devices. Specifically, interactions that span over multiple turns and keep users engaged over longer period of time are often infeasible. Furthermore, the short command-and-response interaction model used in these devices makes it difficult to provide social or emotional support to users. In this project, we explore how to address these issues and extend the capabilities of smart-speakers to sustain longer interactions. One of our key goals is to transform these devices into “active listener” [4] — devices that can help users to engage in *troubles talk* and disclosure.

Disclosure can have positive effect on wellbeing. Pennebaker et al. [5] introduced *expressive writing* in which participants disclosed information about traumatic events or emotional upheavals in written forms. Expressive writing has been effective in emotion control and improving wellbeing in clinical populations with different symptoms including PTSD [3], anxiety [5], depression [7], and mood disorder [1]. It has been shown to have positive health effects for non-clinical populations as well [2]. Expressive writing requires iterative practices. However, sustaining longitudinal engagement with

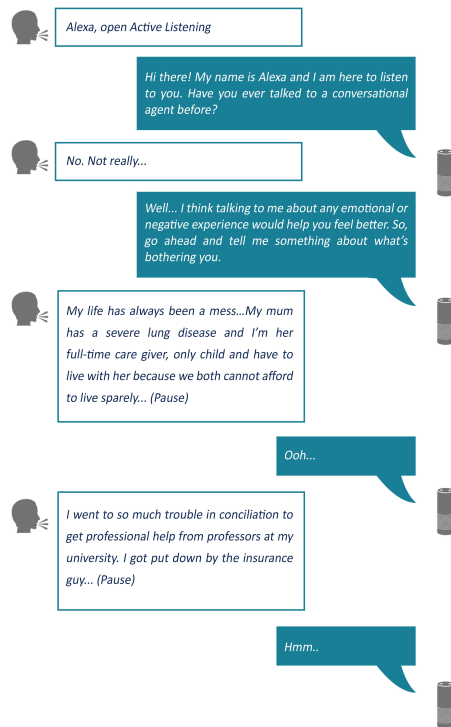


Figure 1: Scripted interactions with the Active Listening skill

expressive writing practices can be difficult. We argue that smart-speakers as active listeners can elicit similarly useful disclosure (we coin this as *effective speaking*). If successful, this therapeutic process can be made available in a scalable manner to a large population given the recent popularity and adoption of these devices. To transform these devices into active listener, we focused on back-channeling and explored how the use of verbal continuers impact user engagement.

BACK-CHANNELING IN ALEXA

Verbal continuers (e.g., “hm”, “uhum”, “aha”, “yeah”) can enhance perceived human-likeness of conversational agents [10, 11]. However, no previous study has looked into the effect of verbal continuers in enhancing user engagement, levels of information disclosure, disclosure intimacy, and emotional well-being. To assess the efficacy of verbal continuers in the perceived level of active listening, we developed an Alexa Skill (application) called *Active Listening*.

“Alexa, open Active Listening”

The user invokes the skill by saying “Alexa, open Active Listening”. After invoking the skill, it delivers a specific prompt or question (e.g., asking to talk about a negative experience). As the user continues, the skill provides verbal back-channelling cues such as “hm”, “aha”, and “oh”. These cues are randomly selected and delivered only when the user take a pause. This helps to minimize interrupting the flow of the conversation. Figure 1 shows an example interaction with the skill. The user can stop or ask Alexa to quit the skill at any time. If the user stays silent for too long, the skill asks the user if there is anything else s/he would like to share. If not, it closes the session with the final message saying “I hope talking to me helped you feel better. Let me know if you would like to talk to me more”.

STUDY DESIGN

To explore the impact of back-channeling, we conducted a pilot study with 4 participants. All participants were graduate students in a large public university in the USA. During the study, the participants interacted with the skill in two stages. In the first stage, they used a pre-written script to interact with the skill. In the second stage, the skill asked participants to talk about a negative emotional experience. Having a pre-written script helped participants to become familiar with the skill as well as identifying any potential usability issues (e.g., accents). The non-scripted second stage enabled free-flowing interactions between the participants and the skill.

The goal of the pilot study was to understand potential issues and identify future design directions. As such, we were particularly interested in assessing the perceived usability and efficacy of the skill as an active listener. The description of the surveys are available in the side bar. We also conducted open-ended questionnaires to better understand the users’ perceptions and identify any potential issues.

Surveys:

The study used the following surveys and open-ended questionnaires:

- Measures of usability and interaction quality by using a 7-point Likert Scale following Gearhart et al. [6] and Lala et al [8].
- Measures of active listening using a 10-point Likert Scale developed by Oertel et al. [10, 11].
- Open ended questions on user expectations, satisfaction, and acceptability towards the idea of Alexa as a potential counsellor including:
 - 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?

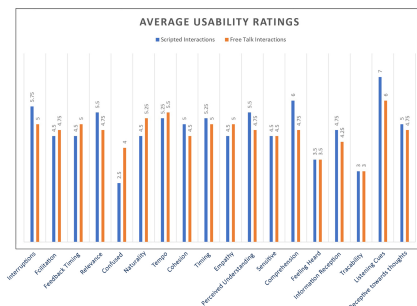


Figure 2: Average usability ratings for both scripted and non-scripted interactions

RESULTS

Based on the average ratings, the skill received relatively high ratings for its performance in providing back-channeling cues. The survey measures for both scripted and non-scripted interactions indicate that the skill has performed adequately (above the mid-point for most Likert scales) as shown in Figure 2. In addition, the perceived active listening scores are also relatively high (average score for scripted and non-scripted sessions are 6 and 5.25 from a 10 point likert scale). This indicates the potential of using back-channeling cues to transform Alexa into an active listener.

Using the open-ended questionnaires, we also asked the participants about their initial expectations of the skill as an active listener, their overall satisfaction from their interaction, and their suggestions for future development. We also explored their opinions regarding the use of Alexa as a potential counsellor. Participants' responses can be summarized as follow:

Satisfaction: Almost all participants acknowledged that the timing of verbal cues from Alexa was reasonable. However, subject 3 did not perceive it as an agent with empathy. Subject 2 was disappointed with the skill since it only provided back-channeling cues without asking or answering any other questions.

Further development: Overall, participants asked for a more empathetic agent. Furthermore, they also suggested going beyond just back-channeling and verbal continuers. Specifically, they recommended tailored responses and follow-up questions to make interactions more engaging, elicit in-depth disclosure, and effectively provide emotional support through these devices.

Alexa as a counselor: Most participants agreed that they would not talk to Alexa as a counselor at this stage. However, they were willing to consider using Alexa for therapeutic purposes in future provided it meets their expectations.

CONCLUSION

In this project, we aimed to support active listening using smart-speakers. Specifically, we explored the use of back-channeling and verbal continuers to maintain longer interactions and sustain user engagement. For this, we developed an Amazon Alexa skill that delivers random back-channeling while users interact with it. We also evaluated the skill in a small pilot study. Our initial findings show that most participants were open to receive such verbal cues. However, they did not consider the skill as an agent with empathy, which might negatively impact its perceived ability for active listening. They suggested going beyond just back-channeling and providing appropriate tailored responses as well to improve its performance as an active listener. Future work will leverage these findings to expand the current interaction model for providing emotional and social support through smart-speakers.

REFERENCES

- [1] Karen A Baikie and Kay Wilhelm. 2005. Emotional and physical health benefits of expressive writing. *Advances in psychiatric treatment* 11, 5 (2005), 338–346.
- [2] Mark Bond and James W Pennebaker. 2012. Automated computer-based feedback in expressive writing. *Computers in Human Behavior* 28, 3 (2012), 1014–1018.
- [3] Alison Bugg, Graham Turpin, Suzanne Mason, and Cathy Scholes. 2009. 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 (2009), 6–12.
- [4] Pamela Fitzgerald and Ivan Leudar. 2010. On active listening in person-centred, solution-focused psychotherapy. *Journal of Pragmatics* 42, 12 (2010), 3188–3198.
- [5] Martha E Francis and James W Pennebaker. 1992. Putting stress into words: The impact of writing on physiological, absentee, and self-reported emotional well-being measures. *American Journal of Health Promotion* 6, 4 (1992), 280–287.
- [6] Christopher C Gearhart and Graham D Bodie. 2011. Active-empathic listening as a general social skill: Evidence from bivariate and canonical correlations. *Communication Reports* 24, 2 (2011), 86–98.
- [7] Katherine M Krpan, Ethan Kross, Marc G Berman, Patricia J Deldin, Mary K Askren, and John Jonides. 2013. An everyday activity as a treatment for depression: the benefits of expressive writing for people diagnosed with major depressive disorder. *Journal of affective disorders* 150, 3 (2013), 1148–1151.
- [8] Divesh Lala, Pierrick Milhorat, Koji Inoue, Masanari Ishida, Katsuya Takanashi, and Tatsuya Kawahara. 2017. Attentive listening system with backchanneling, response generation and flexible turn-taking. In *Proceedings of the 18th Annual SIGdial Meeting on Discourse and Dialogue*. 127–136.
- [9] NPR and Edison Research. 2019. The Smart Audio Report (Spring 2019). Available at https://www.nationalpublicmedia.com/wp-content/uploads/2019/06/The_Smart_Audio_Report_Spring_2019.pdf.
- [10] Catharine Oertel, Joakim Gustafson, and Alan W Black. 2016. Towards Building an Attentive Artificial Listener: On the Perception of Attentiveness in Feedback Utterances.. In *INTERSPEECH*. 2915–2919.
- [11] Catharine Oertel, José Lopes, Yu Yu, Kenneth A Funes Mora, Joakim Gustafson, Alan W Black, and Jean-Marc Odobez. 2016. Towards building an attentive artificial listener: On the perception of attentiveness in audio-visual feedback tokens. In *Proceedings of the 18th ACM International Conference on Multimodal Interaction*. ACM, 21–28.