# Alexa, Is This A Mood?

## An investigation into the effects voice user interfaces have on mood tracking

CO620

Brandon Okeke

Supervised by Jason Nurse

Words: 5,738

***Abstract* - This report explores the potential of using voice assistant platforms to develop applications that can support an individual's wellbeing. This is measured by analysing development platforms' support for the creation of these types of applications, and if these applications can show an improvement in an individual's emotional self awareness. A prototyped mood journaling application was created in the form of an Alexa Skill and presented to 12 participants. Participants provided feedback on their usage of the mood journal after a week and their usage statistics were analysed. No significant changes in their emotional self-awareness was measured but feedback regarding possible improvements for the application was collected. Currently available development platforms were found to provide the necessary tools to create these applications, but more research into how to do this effectively is still required.**

# 1. Introduction

## 1.1. Background & Motivation

A conversation agent (CA) can be described as "a system that relies on chat, supported by artificial intelligence, to have verbal interactions with end users" (Sciuto et al. 2018). The usage of voice-based conversation agents through the form of voice assistants (VA) has grown rapidly. A 2018 report by Voicebot.ai showed the amount of monthly active users of voice assistants was at 90.1 million via smartphones and 45.7 million for smart speakers. The report also suggested that voice assistants are entering their second phase, through a focus on more complicated interactions and introducing more contexts (Voicebot.ai 2018).

SABORI is as self care application developed by the University of Tokyo with the purpose of providing preventives measures for mental health issues. Suganuma, Sakamoto and Shimoyam's (2018) comparison of SABORI and their updated version, which included conversational features, showed that conversational elements contributed to providing benefits in a wider range of areas outside of depression. Suganuma, Sakamoto and Shimoyam (2018) make referees to further advancement of SABORI, suggesting that its focus on text and pictures could be bridged with a smart speaker, with the potential to improve usability.

A review by Provoost et al. (2018) on embodied conversational agents (ECA) through multiple contexts, has shown the positive effects ECAs can have in terms of keeping users engaged with a task. The review also demonstrated how ECAs can be useful when with in conjunction with convivial methods for treating mental disorders. They acknowledge the complexities in designing an ECA and suggest a "low-tech" approach with the aim to create studies faster. This approach will, most importantly, allow for the connection of which attributes of an ECA make it effective. The study also fails to discuss the technological aspects of the ECAs reviewed. The review focuses on demonstrating the practical use of ECAs in a clinical setting for health practitioners.

Previous research, including the aforementioned studies, into CAs has aimed towards assessing embodied or anthropomorphised agents (Luger and Sellen 2016), which incorporate non-verbal communication in their interactions; features not present in commercial VAs such as Alexa. Available research clearly shows that CAs can be developed, but said research is also void of commercial context (Luger and Sellen 2016). Provoost et al.'s (2018) review on ECAs explicitly removed its focus away from the technical requirements to develop these conversational agents. In addition to this, research into the effects of commercially available CAs tends to have a focus on written communication through the form of chatbots.

Still, these chatbots, such as Woebot, have been shown to be useful when delivering cognitive behavioural therapy. A study by Fitzpatrick, Darcy and Vierhile concluded "those in the Woebot group experienced a significant reduction in depression" (2017), with some participants also personifying the chatbot. Fitzpatrick, Darcy and Vierhile made comments on how "conversational interfaces may be better positioned than visually oriented mobile apps to deliver structured manualized therapies" (2017).

Continued advanced in the technologies that power conversational agents has provided a novel way to for people to engage with machines socially (McLean and Osei-Frimpong 2019). Consumers' growing familiarity with voice assistants, along with evidence to suggest that CAs can be beneficial for mental wellbeing, prompts the question of whether VAs can provide similar benefits. It also invites the discussion of if the development platforms for the VAs provide the tools for developers to easily create these kind of applications

## 1.2. Aims & Contexts

The aim of the project is to develop, and test, a prototyped mood-journaling application through a VA to determine if developing an application to provide basic management tools for one's mental wellbeing is feasible by utilising commercially available VA platforms. To do this we will look to identify how comfortable people are when using VAs to record emotion, if certain contexts affect this and if this medium provides significant positive effects on an individual's mental wellbeing as standard mood journal techniques (Ullrich and Lutgendorf 2002; Utley and Garza 2011) or through applications presented by a graphical user interface. We will combine these findings with analysis done on the tools used to develop the mood journal to create links between the success of specific features and the manner in which they were implemented.

We will make comments on how features may be implemented on a technical level in future iterations or if this is not possible due to the limitations of the VA's platform. Other, more general, discussions regarding the development of the application will also be made.

**Choosing a voice assistant.** When it comes to VAs available commercially, there are several to choose from with each optimised for different areas. Results from respondents published in a Microsoft report tells us that Apple's Siri and Google Assistant were the two most popular voice assistants, each being used by 36% of people, while Amazon's Alexa fell in third place with 25% (Microsoft 2019). Although Siri may be tied for first place, its software development kit, SiriKit, does not provide a platform to create independent voice experiences. SiriKit is tightly integrated into iOS, iPadOS and macOS by providing voice functionality to extend mobile applications and so will be disregarded for this project (Apple Developer n.d.). This positions Google Assistant and Alexa as the two top commercially available VAs that allow for the development of independent voice experiences from scratch through Actions and Skills respectfully. For the purposes of this project, we are also disregarding open source equivalents that could be used to mimic the features of VAs as the project specifically targets commercially supplied platforms. This includes, but is not limited to chatbots and other natural language processing tools. Amazon's Alexa Skills platform was selected as the platform to develop the application on the basis of having physical access to an Amazon Echo Dot, making prototyping easier.

The remainder of the report abides to the following structure. Section 2 describes the requirements gathering and designing of the application to test. Following this, Section 3 discusses the methodology used to carry out the study including how data will be analysed. We detail the results collected from the study in Section 4. Expanding from this, Section 5 includes a discussion on the results gathered, incorporating technical limitations discovered during the development process. Finally, in Section 6, we summarise the discoveries made from carrying out this project and outline possible paths for future work.

# 2. Requirements Analysis & System Design

## 2.1. Modelling the Mood Journal

A review of commercially available mood-tracking apps by Caldeira et al. (2018) concluded that the common feature set found throughout could be modelled under the personal informatics systems:

> "1) the Preparation stage consists of planning that occurs before users start collecting personal data, 2) the Collection stage refers to when users record their data, 3) the Integration stage includes formatting and combining data, 4) the Reflection stage involves making sense of and learning from their personal data, and 5) the Action stage, when users act based on the insights gained through reflection" (Caldeira et al. 2018).

Following this system allows us to determine the minimum set of features required to produce a functioning application. For the purposes of this project, we will be focusing on the Collection and Reflection stages. The Collection stage had various implementations across applications, for example, predefined text boxes, colours or ratings; 21 out of 32 apps allowed users to rate a specific mood. Moreover, data visualisations were common place during the Reflection stage, found in 29 out of the 32 apps included bar graphs or pie chart (Caldeira et al. 2018). As we will be developing an application with no visual elements, this would need to be translated in a way that can be relayed solely through voice. The nature of our application allows us to treat the Collection and Integration stages as one given that integration automatically happens during collection. As the analysis on how participants react to their recorded moods is not within the scope of this project, it will not be implemented. Participants will be briefed on how to use and collect entries before they partake in the study, facilitating the preparation stage.

## 2.2. Requirements Design

With the limitations developed, it is clear that the application would have a simple use case, detailed in Figure 1. To facilitate both the Collection and Reflection stages, participants will need to be able to create a mood entry and subsequently have the ability to view said entry. Participants will be able to list the titles of all previous entries, as well as being able to delete entries to allow for the amendment of mistakes.

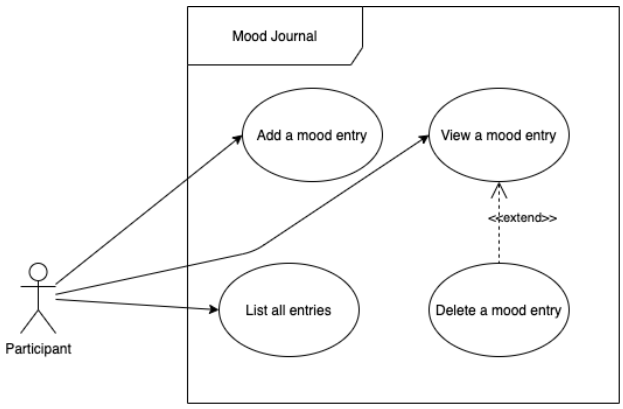



**Figure 1. Use case diagram display the minimum requirements needed to develop the prototype.**

## 2.3. Interaction Design

Defined below are the basic successful flows for the stated use cases.

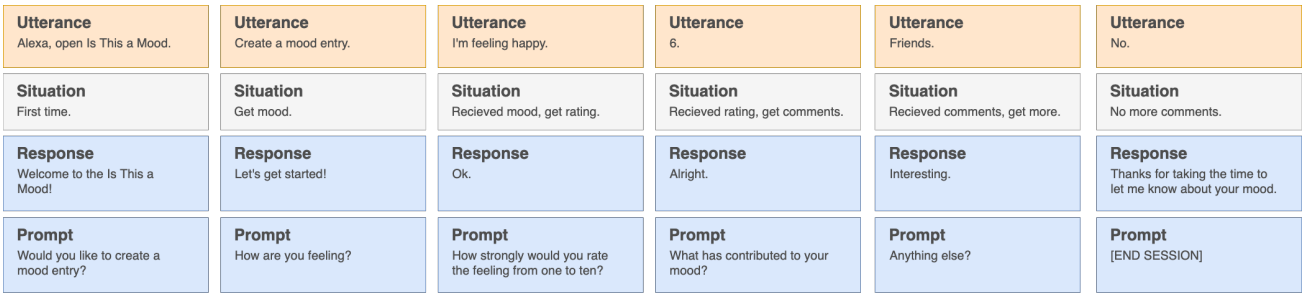



**Figure 2. An example script developed to plan the expected interaction for a user trying to create a new entry.**

*Adding a mood entry:*

i. Participant provides a unique title for the entry.
ii. Participant provides the mood for the entry.
iii. Participant provides a number between 1 and 10 to express the strength of the mood.
iv. Participant given option to provide a list of activities that contributed to that mood.

*Listing all entries:*

i. Participant requests to view all of their stored entries.
ii. The titles of entries stored by a participant are read back to them in a page-based format.
iii. The participant can request to go forward and back between pages.

*Viewing a mood entry:*

i. Participant requests for an entry to be viewed.
ii. Participant provides the title of the entry they would like to view.
iii. The entry's contents is read back to the participant.

*Deleting a mood entry:*

i. Participant requests for an entry to be deleted.
ii. Participant provides the title of the entry they would like to delete.
iii. Participant is asked to confirm the deletion request.

## 2.4. Alexa's Skill Platform

Alexa's skill platform makes use of 3 elements to convert a user's requests into meaningful, dynamic actions. This is done by utilising intents, slots and utterances. An intent can be likened to an action, one that is designed to complete a specific request. Utterances are textual examples of phrases that a user may speak in order to activate a certain intent for example, in order to active an intent design for purchasing an item, a user may ask Alexa to "buy me a drink". Alexa uses sample utterances to train a natural language processing model unique to the application in development. This allows for similar phrases to be used interchangeably whilst still mapping to the appropriate intent. Utterances may include optional slots, slots define variables that can be filled by a user. In the previous example, the we would define the utterance as "buy me a `[item]`". When a user makes the request, the word drink will be assigned to the item variable which can then be used programatically. (Amazon n.d.).

**Table 1 - Planned intents and their purposes.**

| Intent | Purpose |
|---|---|
| **CreateEntry** | Starts the processes for creating a mood entry. |
| **ListEntries** | A request to list the titles of all entries under a participant |
| **ViewEntry** | A request to display the details for a specific entry. |
| **DeleteEntry** | A request to delete a specific entry. This requires a confirmation. |
| **CanUse** | Allows the participant to ask if a keyword can be used as a mood or an activity. |

Amazon recommends that developers create scripts to help visualise possible interactions users may have with an application (Amazon n.d.). These scripts can later be converted to intents with appropriate slots and utterances. Figure 2 displays and example script for the successful path when creating an entry. Table 1 shows a mapping of the intents original create alongside their purpose.

Alexa has a series of predefined intents and slot types, which are prefixed with `AMAZON`. An examples of such an intent is the `AMAZON.YesIntent` for confirmations. Predefined slots types represent common real-world entities such as actors, places and digits, presenting constraints on the values users can provide. There is not a slot type for free-form-text, however. The closest match is the `AMAZON.SearchQuery` slot type, defined as "a search query like you might enter into a search engine, for example for searching a database" (Amazon n.d.). The search query slot type can only be used once per intent and is not intended to be used to capture large amounts of information. In order to allow for participants to provide a unique title for their entries, the `AMAZON.SearchQuery` slot type will be utilised.

It is important to note that by default, there is no slot type designed for collecting a list of items. The use case's requirements means the system must provide a way to attach multiple activities to one entry. In order to achieve this, we will utilise Alexa's dialogue delegation feature.

> "The dialog model identifies the prompts and utterances to collect, validate, and confirm the slot values and intents. When you delegate the dialog to Alexa, Alexa determines the next step in the conversation and uses prompts to ask the user for the information." (Amazon n.d.).

By default, dialogue is delegated to Alexa. This means the process of collecting inputs from a user is handled automatically for us based on settings within our intent. Alexa will only send data to our servers if a dialogue has been completed by successfully filling all of its slots. Once a user has completed the `CreateEntry` intent, we will manually delegate the dialogue, forcing the Alexa to present a new intent of `AddActivityToEntry`. A participant will be continuously asked for the entry they would like to add until they activate the `AMAZON.NoIntent` by using a termination word.

As entries will be encrypted, users are required to set a pin which will be provided each time they launch the application. It is possible for users to go directly into an action by requested to do so. An example utterance in the context of this application would be "Alexa, ask my mood journal to create a new entry". Doing this would skip pin entry. We will use a manual dialogue delegation here by creating a new `SetPin` intent, having any intents be forwarded here if the no pin is detected during that session.

When waiting for a response from a user on Skills not designed for screens, Alexa waits for "a few seconds… [before] Alexa closes the microphone" (Amazon n.d.). Even when the re-prompt option is enabled and no response from the user is given, Alexa will end the session. With this in mind, when a participant asks for a list of their entries, the responses should be limited to three items per page. This should reduce the potential stress on participants who may be overwhelmed with the choices, thus giving them enough time to provide a response.

## 2.5. Application Design

With the intents, slots and utterances planned, the next phase involves the design of the backend system. Amazon recommends developing applications via their serverless computing solution AWS Lambdas. The push to use AWS Lambdas provides a few benefits, with the most prominent being that it reduces the need to manage infrastructure, whilst providing a focus on developing meaningful features. As AWS is owned by Amazon, it is tightly coupled with Alexa's development kit and provides a few conveniences such as simpler state storage; however, a requirement of this project is to ensure that it is securely processed and stored on the University of Kent's servers.

To avoid the reliance on a third party, we opted to develop a self-hosted application using the ASP.NET Core framework, written in C#. A self-hosted Alexa Skill backend in an unsupported programming framework works identically to an AWS-hosted skill, with the exception of needing to manually serialise and deserialise JSON requests correctly. See Appendix A for a data flow diagram detailing the interaction between the three parties that make up the system. A participant's voice request is interpreted by Alexa on Amazon's servers. The interpreted request is forwarded to the backend hosted on the University of Kent's servers. Here, any required authentication takes place, using the pin set by the participant to encrypt or decrypt entries as necessary. Whilst Alexa tracks analytical usage for developers and does so in a way that provides us with richer general data, it does not provide a method to uniquely identify customers. Therefore, when an endpoint on the backend is reached, non-sensitive analytical data, such as if the request was successful, is captured and stored alongside an anonymous participant identifier in an Elasticsearch instance which will be used for producing data visualisations.

# 3. User Study

The study uses a pretest-posttest design. Participants were recruited via social media call-outs on University of Kent accounts, as well as by word of mouth. The inclusion criteria was as follows: a student at the University of Kent; has access to an Android phone running Android 5.0+ or an iOS device running iOS 12 with the ability to download apps from the App Store; has a good understanding of written and verbal English; has, or can create, an Amazon account; is available for two consecutive weeks. As an incentive, participants who completed the study were entered in a draw to win one of three £15 Amazon vouchers. Accepted participants were contacted individually via video conferencing and had their accounts set up. They were shown examples of how to interact with Alexa and where to seek help if required.

## 3.1. Data Collection

The study focuses on collecting three types of data (two quantitative and one qualitative): a level of emotional self-awareness (ESA), usage statistics and usage feedback. An increased level of ESA has been shown to "predict a decrease in depressive symptoms" (Kauer et al. 2012). Participants were asked to fill out a questionnaire to receive a pre-test score on the ESA scale. After doing so, participants were instructed to use the Alexa Skill for a week as and when they felt it appropriate to do so. This was to mimic the behaviours expected when using a mood journaling mobile application. After a week of using the application, participants were asked to fill out the ESA questionnaire again to generate their post-test score. Participants were finally required to fill out a short questionnaire to detail their experiences of using the Alexa Skill, the questions are detailed in Table 2. The questionnaire uses a convergent parallel mixed-methods design in order to collect qualitative and quantitate data simultaneously.

Alongside this, usage data was collected as participants used the Alexa Skill. The data collected was placed into four categories: (1) Completed Intent Usage By Participant, (2) Intent Errors By Participant, (3) Overall Completed Intent Usage and (4) Completed Intent Used During Single Session.

## 3.2. Data Analysis

**ESA Scores.** To determine if the differences in ESA scores before and after using the mood journal are statistically significant, we will be performing a one-tailed, matched pairs t-tests with a significance level of $P < .05$.

**Qualitative & Quantitative feedback.** To perform analysis on qualitative results from the feedback questionnaire, we used an approach outlined by Braun and Clarke (2006) to create thematic maps. Codes will be generated to identify overarching themes in a way that can be quantified. These themes will further be divided into sub-themes where necessary. We will disregard comments that do not relate to the interaction with Alexa, such as feedback relating the implementation of general mood journalling features that were not included in the prototype. These comments may be useful in understanding how to interpret features for voice user interfaces but this is not the primary focus of the study. The generated thematic maps will be combined with quantitative results collected from the server and the usage statistics. This will allow us to find patterns in the participants' usage of the mood journal, or lack thereof. This will also be used to assess whether or not the technical limitations related to the implementation of specific features had an effect on the effectiveness of the application.

**Table 2 - Questions provided to participants for feedback.**

| # | Question |
|---|---|
| 1 | Have you used mood journals before taking on the study? |
| 2 | Did you have experience with Alexa or other voice assistants / chatbots before taking on the study? |
| 3 | In general, did you have trouble using the journal? |
| 4 | Which of the available features, if any, did you find easy to use? Please provide examples of why it was easy. |
| 5 | Which of the available features, if any, did you find difficult to use? Please provide examples of why it was difficult. |
| 6 | Did you feel comfortable using the journal? |
| 7 | Did you feel comfortable with Alexa's responses? |
| 8 | What would you change about Alexa's responses? Be as detailed as possible, using examples if you can. |
| 9 | Did you limit your interactions with Alexa to specific locations? |
| 10 | Why did / did you not limit your interaction to specific locations? Be as detailed as possible, using examples if you can. |
| 11 | Do you feel you have a decent understand of your mood from the week? |
| 12 | Did privacy or security concerns impact your usage of the journal? Please explain why, using examples if you can. |
| 13 | Would you continue to use this type of journal in the real world? |
| 14 | Is there anything you would change about the way in which you used the journal? This can include things such as how you spoke to Alexa, her responses or features. |
| 15 | Please provide any other feedback that wasn't mentioned. Be as detailed as possible, using examples if you can. |

# 4. Results

## 4.1. Participants & ESA Scores.

A total of 14 people registered to participate in the study, two failed to set up their application collect data (N = 12).

There was no significant difference in ESA Scores between the pre-test (M = 61.8442; SD = 11.0305) and the post-test (M = 63.9542; SD = 8.4439); $t(11) = 0.810023$, $p = .21755$.

**Table 3 - ESA Scores before and after using the mood journal.**

| Participant # | Pretest | Possttest |
|---|---|---|
| 1 | 49.67 | 73.71 |
| 2 | 64.62 | 60.50 |
| 3 | 58.21 | 64.26 |
| 4 | 50.76 | 53.02 |
| 5 | 50.29 | 49.74 |
| 6 | 54.14 | 54.57 |
| 7 | 60.12 | 70.17 |
| 8 | 66.62 | 68.62 |
| 9 | 74.10 | 74.43 |
| 10 | 54.98 | 58.74 |
| 11 | 82.50 | 70.43 |
| 12 | 76.12 | 69.26 |
| | | |
| Mean | 61.84 | 63.95 |

## 4.2. Questionnaire.

**Quantitative results.** Participants' previous usage of voice assistants was split equally. More participants had not used mood journals in any capacity before the study (N = 7). More participants said they had trouble using the journal in general (N = 7) than those who did not (N = 5). 8 participants felt comfortable using the mood journal in its current form, as well as 9 participants stating feeling comfortable with Alexa's responses. 2 participants did not limit their usage of the mood journal to specific locations. Most participants felt they had a decent understanding of their mood for the week in which they took part in the study (N = 10). 7 participants said they would continue to use an application, which was similar to the one presented to them, in a real environment.

**Qualitative results.** Thematic analysis of each question is broken down below. See Appendix B for an example of a thematic map generated for question 5.

***Q#4:*** 6 participants stated that creating an entry was the easiest action to perform, as it was considered a low-effort task (N = 2) and was non-judgmental (N = 1). Viewing entries was also considered to be an easy action to perform (N =1). 1 participant noted that no feature was particular easy to use whilst another stated the overall experience of using the journal was easy to grasp.

***Q#5:*** 4 participants stated the creating an entry was the hardest action to perform which led to repeatedly abandoning the process (N = 3). The difficulties in creating entries were attributed to a lack of understanding of the type of words or phrases that Alexa accepted (N =3). 2 participants struggled with naming entires in the space allotted, defaulting to using the date as the name. 2 participants expressed distaste towards having to restart the process again due to a sometimes unnatural conversation flow.

> "The number of attempts was challenging, often meaning having to restart the entry all over again and interrupting the experience of logging moods. [This] also means that the mood could change within that time or I could become more frustrated in trying to log it." — Participant 12.

> "In the middle of asking questions, Alexa would ask for my pin and after giving it to her I'd have to restart answering all the questions which was annoying." — Participant 4.

***Q#8:*** 4 participants stated that they would change aspects of Alexa's responses. Specific changes mentioned are as followed: allowing the response to have more human attributes (N = 1); having Alexa provide relevant responses to an entry to support an individual if needed, for example, signposting when low mood is registered often (N = 2) and providing visualisations of entries entered (N = 6).

***Q#10:*** Concerns regarding personal space as a reason for limiting the usage of the application were raised by 8 participants. 2 participants made reference to specifically restricting their usage to quiet areas so that others would not hear them. Sub-themes surrounding the social acceptability of using Alexa in public also emerged (N = 2). 2 participants did not limit their interactions to specific locations on the basis that they wished to form entries whilst the mood was still fresh in their minds.

***Q#12:*** Participants were generally not concerned about the privacy or security in relation to their mood journals. Of those that were, 3 participants made comments relating to security, with a sub-theme relating to pin entry feeling unsafe due to it having to be spoken out loud.

***Q14:*** More options for moods and activities (N = 2) was a suggested feature. Within that were mentions of adding customised moods and activities as well as including the possibility of adding more contextual detail. 2 participants felt the overall voice user interface could be adapted, with 1 participant noting that it would be easier if there were fewer steps involved. 1 participant discussed the opportunity to include more chances to fail.

> "There was limited vocabulary available at your disposal, so it was difficult to express how you truly felt, so it wasn't useful for monitoring your moods. For example, I wanted to say that the weather made me happy, but this wasn't an option. Going forward I think the app could allow the person to freely dictate how they feel so they can add more detail." — Participant 8.

***Q#15:*** 1 participant expressed interest in a companion application specifically for housing visualisations of their entries. Participants felt that overall, the mood journal was useful and convenient (N = 3). 1 participant made a comment about the reliability of the application when considering those who are multilingual.

> "I wonder if the use of accents would get in the way of using the feature. My accent tends to fluctuate and often times I struggled with Alexa to pick up these subtleties and would have to start all over again. I wonder if I would struggle with this if my accent was stronger and if this means others may struggle with this too." — Participant 12.

## 4.3. Usage Statistics.

One participant's usage statistics was corrupted during collection and has been specifically excluded from this section (4.3).

There were a total of 246 completed actions. 84 of these actions were related to opening the mood journal and providing a pin to allow further actions to take place. Removing launch actions from this figure brings the total of feature-related actions to 162. The most used action was creating an entry, taking up 34% of the overall usage of the mood journal. However, only 42 entries were registered in total, resulting in a success rate of 57%. Adding activities to entries followed in terms of usage at 16%, being performed a total of 34 times. 30 activities were found to be registered to entries, giving it a success rate of 88%. Viewing a specific entry was performed 5 times and was spread over 3 participants. Listing all entry titles and navigating through them had a usage rate of 24%. On average, participants asked if a specific mood or activity could be used 1.8 times in a single session, as well as asking to create an entry and to add activities to entries 1.2 times per session respectfully. Most participants performed one action per session. 59 errors beyond those detected by Alexa were recorded. A total of 4 participants ran into errors whilst using the mood journal, mostly with failing to provide a correct name when recalling an entry that had been created (74.6%). 32 out of the 44 occurrences of this error were performed by one participant. Other errors include: using an incorrect pin (16.9%), performing an unknown request (6.8%) and creating an entry with a name that already existed (1.7%).

**Table 4 - Total registered usage of each action.**

| Actions | Count |
|---|---|
| Launching the skill | 84 |
| Creating an entry | 73 |
| Adding an activity to an entry | 34 |
| Cancelling an action | 21 |
| Asking if a mood / activity is usable | 15 |
| Cancelling an action / explicitly calling for the skill to be closed | 14 |
| Viewing a single entry | 5 |
| Listing all entries | 52 |
| | |
| **Total:** | **298** |

# 5. Discussion

The study looked to understand if a mood-journaling application presented exclusively through a voice user interface can be developed using commercially available platforms and if this mood journal provided benefits to an individual's wellbeing.

## 5.1. Design & Usability

Creating an entry was regarded as one of the easiest actions to use even though it had a success rate of 57%. As a whole, the application we developed was not usable. Additionally, the application was not found to have an effect on ESA. This can be attributed to designing inadequate interactions, evident from the low success rate of creating an entry. Examples of inadequate design are as follows:

i. **Limiting participants' entries to predefined slot values.** The manner in which we collated moods and activities led to a list of values that underrepresented the expectations of participants. Furthermore, participants felt the inclusion of more detail would have made the application more useful. This is also a desired feature for GUI-based mood journals (Caldeira et al. 2018).

ii. **Improper state management.** Participants' expectations of how a conversation should be carried out were not met. Participants received unexpected prompts, such as asking for a pin in the middle of creating an entry. For erroneous situations, their state in the application was not saved, requiring them to begin an entire process from the start. Experiencing this repeatedly led to frustration and ultimately abandoning the application.

iii. **Time and chance limits.** The short amount of time in which participants could provide a response may be an explanation for the low success rates. It requires an individual to prepare for each question in advance, speak in a potentially unnaturally structured way and does not support the retraction of filler speech. Participants also received a maximum of two attempts before Alexa stopped accepting input, which is expressed to the participant solely through a sound effect.

Allowing participants to enter their own moods and activities would solve issues regarding the limited choice. However, not allowing participants to do this was done deliberately for the purposes of the study. We wanted to ensure the interactions were equal between participants and rule out potential errors regarding inconsistencies with Alexa's interpretation of customised slot values. Including more detail through free text is possible, although is not recommended. We have shown that free-form-text is possible using the slot type `AMAZON.SeachQuery` but this is not the intended purpose of this slot type. Each intent can also only include one `AMAZON.SearchQuery` slot type; in order to include multiple free text inputs, intents would need to be mapped to inputs rather than an action. A similar solution was used to achieve adding multiple entries, as detailed in Section 2.4. The incorrect management of the application's state that came as a result of this approach can be attributed to causing the improper matching of participants' expectations. Additionally, Alexa does not listen for input indefinitely. A paragraph of extra detail would need to be inputted as a series of short sentences, with unnatural gaps, to account for Alexa's processing delays.

## 5.2. Solutions

In order to complete this solution, it is recommend to use manual slot delegation instead of automatic slot delegation which was used for this study.

The feedback participants supplied provides insight into the decisions that need to be made to create a more robust application. Despite the inconveniences presented by the application, it is clear that participants were comfortable with using the application, particularly with how Alexa responded, and expressed interest in seeing this type of application in a commercial setting. Further research is required to discover the ideal way in which to convert the interactions present in GUI mood journals to their VUI counterparts. This includes looking at how to best structure natural conversations that are optimised for collecting appropriate data. The effectiveness of VUI-based mood journals can be properly analysed once these interaction models have been discovered.

Having access to a voice assistant such as Alexa, provides a convenient alternative method to take note of simple moods. An ideal solution to combat the issues discovered would be to include a voice interface inside of a GUI. The VUI would be used as the primary method of interacting while the GUI would be used for secondary processes. For

example, the former would be used to create an entry, while the latter would be used to display visualisations.

The mixture of the two interfaces would solve the issue regarding the limits placed on inputting data. When presented with a prompt, participants could choose when to begin inputting data, reducing the stress to provide a quick response. It would also provide users with standard visualisations which have been shown to be an extremely valuable feature for those who use mood journaling applications (Caldeira et al. 2018). Additionally, contextual support regarding what a participant can say and cannot could be presented within the application; this is appreciated amongst users of VUIs (Corbett and Weber 2016). The inclusion of such a feature would aid users in understanding the types of inputs accepted by the VUI. Furthermore, users would not be required to form a new mental model of the interface through trial and error-based learning, reducing potential abandonment.

Further research into comparing the capabilities of different voice-based platforms to provide a mood journaling application is required.

## 5.3. Limitations

It is important to note that the success rates calculated are optimistic as they are based off actions that reached our server. As we utilised automatic dialogue delegation, an intent would be sent to our server only when all slots have been filled. For example, failing to provide a correct mood for an entry multiple times would cause Alexa to suspend the application which would not be captured by our usage statistics. Amazon's analytics do not provide a way to uniquely identify users making it difficult to remove data generated through testing the system.

The onboarding process may have contributed to the lack of understanding of how to navigate the interface. Restraints in physically meeting with participants meant the onboarding process was done online, however different methods had to be used for participants. 7 out of the 12 participants were able to partake in a video call, with a live demonstration of the application but the remaining 4 were sent emails detailing the usage of the application. Proper tutorials are useful for teaching new interaction models (Corbett and Weber 2016).

# 6. Conclusion

As voice assistants and the platforms to extend them develop, research into their ability to provide tools to aid mental wellbeing is needed. This study has presented the questions on whether mood journals accessed through a voice user interface provide similar benefits to their graphical counterparts, and whether the tools to produce such an application exist. We developed and tested a prototype mood journal application on Amazon's Alexa Skills platform. Whilst there is no evidence to suggest the mood journal we developed impacted emotional self awareness, the responses from participants show an interest in using mood journals accessed through voice user interfaces. We recommend that more research is carried out to understand the attributes that make an accessible conversation with Alexa, which would ultimately lead to a reduction in the number of abandonments. Furthermore, research into comparing the technical availability of different developer platforms that could be used to create such an application is required. Many issues found with the prototype, such as the low success rate for completing actions, can be attributed to the technical limitations presented by using Amazon's Alexa Skill kit. This is not to say that with time and careful consideration, it is not possible to use Amazon's platform to create such an application, however. Proper utilisation of Alexa's manual dialogue delegation features should allow for the more complex interaction presented by the mood journal. We propose that a solution to the limitations discovered would be the development of a standalone GUI application. Such an application would allow for the main interaction to be voice based, whilst providing convenient access to secondary actions in a more recognisable fashion. Voice based interactions show great promise, but they may be unable to provide a complete toolkit for aiding mental wellbeing alone.

# 7. Acknowledgments

Thank you to my family, friends and project supervisor for supporting my vision.

Also, I would like to give a big thanks to the Internet for connecting me to those I love, and the knowledge I desire.

# 8. Bibliography

Amazon Alexa Skills Kit Official Site: Build Skills for Voice [Online]. Available at: https://developer.amazon.com/en-GB/alexa/alexa-skills-kit [Accessed: 6 March 2020a].

Amazon Create Intents, Utterances, and Slots [Online]. Available at: https://developer.amazon.com/en-US/docs/alexa/custom-skills/create-intents-utterances-and-slots.html [Accessed: 19 April 2020b].

Amazon Create the Interaction Model for Your Skill [Online]. Available at: https://developer.amazon.com/en-US/docs/alexa/custom-skills/create-the-interaction-model-for-your-skill.html [Accessed: 10 March 2020c].

Amazon Delegate the Dialog to Alexa [Online]. Available at: https://developer.amazon.com/en-US/docs/alexa/custom-skills/delegate-dialog-to-alexa.html [Accessed: 3 May 2020d].

Amazon Write Out a Script with Conversational Turns [Online]. Available at: https://developer.amazon.com/en-US/docs/alexa/alexa-design/script.html [Accessed: 10 March 2020d].

Apple Developer Siri - Apple Developer [Online]. Available at: https://developer.apple.com/siri/ [Accessed: April 2020].

Bickmore, T., Gruber, A. and Picard, R. (2005). Establishing the computer–patient working alliance in automated health behavior change interventions. Patient Education and Counseling [Online] 59:21–30. Available at: 10.1016/j.pec.2004.09.008 [Accessed: 29 February 2020].

Braun, V. and Clarke, V. (2006). Using thematic analysis in psychology. Qualitative Research in Psychology [Online] 3:77–101. Available at: 10.1191/1478088706qp063oa [Accessed: 10 March 2020].

Caldeira, C. et al. (2018). Mobile apps for mood tracking: an analysis of features and user reviews. AMIA Annu Symp Proc. 2017:495–504.

Corbett, E. and Weber, A. (2016). What Can I Say? Addressing User Experience Challenges of a Mobile Voice User Interface for Accessibility. Proceedings of the 18th International Conference on Human-Computer Interaction with Mobile Devices and Services [Online] MobileHCI '16:72–82. Available at: 10.1145/2935334.2935386.

Fitzpatrick, K.K., Darcy, A. and Vierhile, M. (2017). Delivering Cognitive Behavior Therapy to Young Adults With Symptoms of Depression and Anxiety Using a Fully Automated Conversational Agent (Woebot): A Randomized Controlled Trial. JMIR Ment Health [Online] 4:e19. Available at: 10.2196/mental.7785.

Kauer, S.D. et al. (2012). Self-monitoring Using Mobile Phones in the Early Stages of Adolescent Depression: Randomized Controlled Trial. J Med Internet Res [Online] 14:e67. Available at: http://dx.doi.org/10.2196/jmir.1858 [Accessed: 20 April 2020].

Luger, E. and Sellen, A. (2016). "Like Having a Really Bad PA": The Gulf between User Expectation and Experience of Conversational Agents. Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems [Online] CHI '16:5286–5297. Available at: http://dx.doi.org/https://dl.acm.org/doi/10.1145/2858036.2858288 [Accessed: 19 April 2020].

McLean, G. and Osei-Frimpong, K. (2019). Hey Alexa … examine the variables influencing the use of artificial intelligent in-home voice assistants. Computers in Human Behavior [Online] 99:28–37. Available at: 10.1016/j.chb.2019.05.009 [Accessed: 29 February 2020].

Microsoft (2019). Voice Report: Consumer Adoption of Voice Technology and Digital Assistants. [Online]. Available at: https://about.ads.microsoft.com/en-us/insights/2019-voice-report [Accessed: 29 February 2020].

Provoost, S. et al. (2017). Embodied Conversational Agents in Clinical Psychology: A Scoping Review. J Med Internet Res [Online] 19:e151. Available at: 10.2196/jmir.6553 [Accessed: 7 March 2020].

Sciuto, A. et al. (2018). "Hey Alexa, What's Up?": A Mixed-Methods Studies of In-Home Conversational Agent Usage. Proceedings of the 2018 Designing Interactive Systems Conference [Online] DIS 2018:857–868. Available at: 10.1145/3196709.3196772.

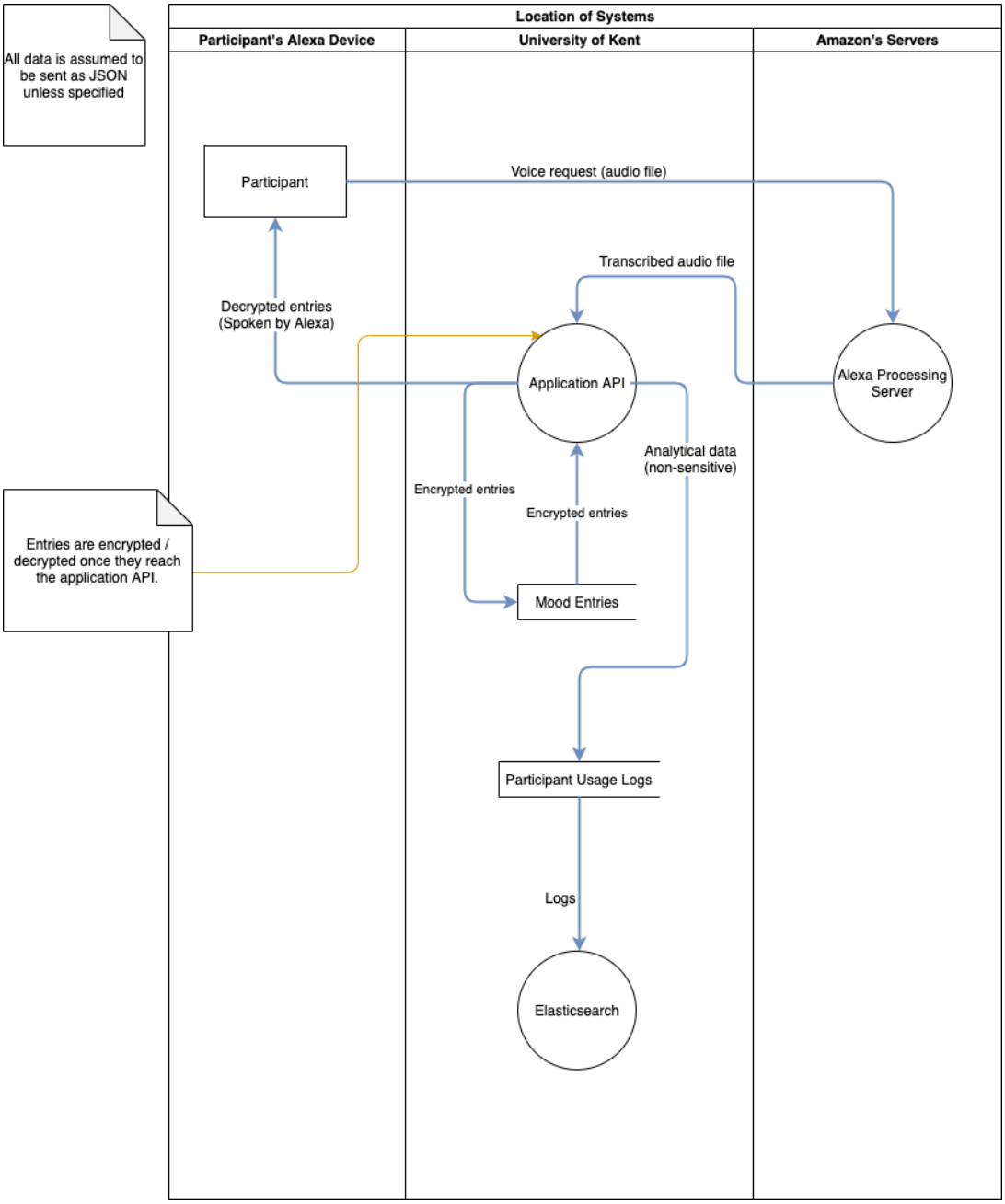
Suganuma, S., Sakamoto, D. and Shimoyama, H. (2018). An Embodied Conversational Agent for Unguided Internet-Based Cognitive Behavior Therapy in Preventative Mental Health: Feasibility and Acceptability Pilot Trial. JMIR Ment Health [Online] 5:e10454. Available at: 10.2196/10454 [Accessed: 28 February 2020].

Ullrich, P.M. and Lutgendorf, S.K. (2002). Journaling about stressful events: Effects of cognitive processing and emotional expression. ann. behav. med. [Online] 24:244–250. Available at: 10.1207/s15324796abm2403_10 [Accessed: 1 May 2020].

Utley, A. and Garza, Y. (2011). The Therapeutic Use of Journaling With Adolescents. Journal of Creativity in Mental Health [Online] 6:29–41. Available at: 10.1080/15401383.2011.557312 [Accessed: 1 May 2020].

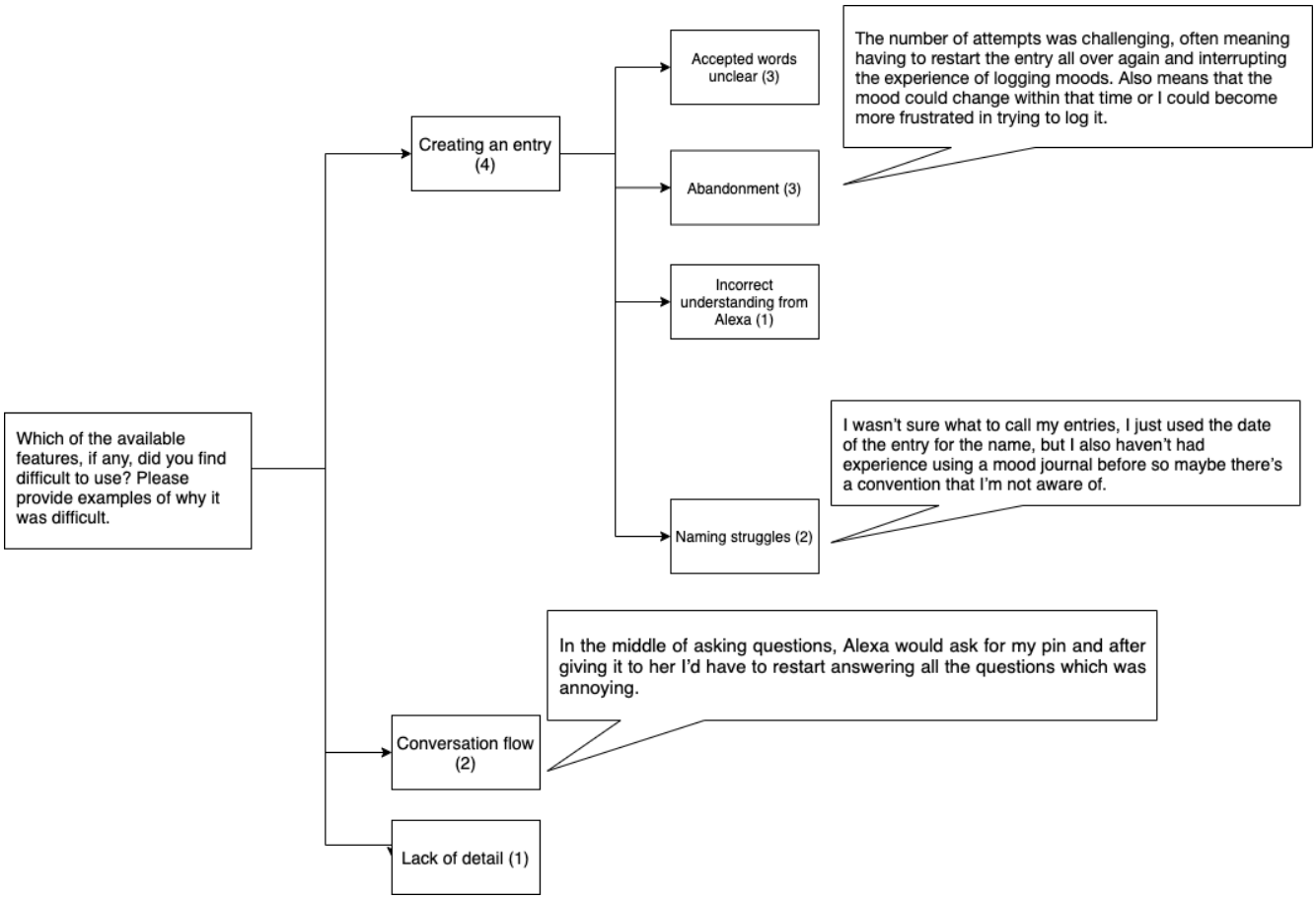
Voicebot.ai (2018). Voice Assistant Consumer Adoption Report. [Online]. Available at: https://voicebot.ai/wp-content/uploads/2019/01/voice-assistant-consumer-adoption-report-2018-voicebot.pdf [Accessed: 29 February 2020].

# Appendix A




**Data flow diagram detailing the interactions between the different systems.**

# Appendix B




**Example of a thematic map generation for question 5.**