FitChat: Conversational AI for Active Ageing

Nirmalie Wiratunga, Anjana Wijekoon, Chamath Palihawadana, Kay Cooper, Vanessa Mendham
School of Computing, School of Health Sciences
Robert Gordon University
Aberdeen AB10 7GJ
Scotland, UK
a.wijekoon@rgu.ac.uk

Abstract

Advances in conversational AI are creating novel and engaging experiences for user interaction through AI Chatbots. In this work, we present the voice-based AI chatbot "FitChat" developed to deliver behaviour change interventions that encourage physical activities among older adults. We start by identifying conversation skills or topics necessary to promote physical well-being through co-creation activities with users. For each conversation skill, we further explore the use of Natural Language Understanding (NLU) and Natural Language Generation (NLG) techniques to improve the conversation. We generate personalised conversation, contextualised with the information extracted from user responses. Provisioning educational content from WHO guidelines on physical well-being provided a useful knowledge source for contextualising chatbot responses and a corpus-based approach helped to avoid non-repetitive chatbot responses. We evaluate the prototype using think-aloud sessions where thematic analysis emphasises that voice-based chatbots are a powerful mode of intervention delivery. Analysis of user responses shows the NLU techniques were instrumental in extracting information that is essential to create cohesive and personalised conversations using NLG techniques.

1 Introduction

Presently, the most common method of delivering digital behaviour change interventions for encouraging physical activities is via text-based notifications on mobile phones. Despite the popularity of this approach, there is little evidence to indicate that text notifications are effective at promoting positive behaviour change, particularly long-term impact. The main problem is that text notifications offer only one-way communication, from the device to the user, meaning explicit interaction is not required. Accordingly, text notifications are easily ignored; fewer than 30% of received notifications are typically viewed by users with an average delay close to 3 hours (Morrison et al., 2018). There is clearly a need for an alternative approach.

As a communication medium, conversation appeals to all age groups, but arguably more so towards older adults. This group can have difficulties with new technologies and may be more inclined to appreciate the natural interaction offered by conversational dialogue. It is also noteworthy that older adults in general are not accustomed to text entry with smart phones. With this in mind, we posit that conversation (more specifically, voice-based conversation) presents an opportunity to deliver behaviour change interventions to motivate higher levels of adoption and adherence in older adults when compared with traditional approaches. In addition, the recent popularity of home hubs is considered a positive indication of user-acceptance towards voice based conversational agents for both smartphone and home-hub platforms. This means that our work is well-placed to investigate conversation as an alternative to current text-based intervention methods.

Our aim is to develop an ubiquitous and proactive system that delivers behaviour change interventions in the form of conversation aimed at promoting physical activities in older adults. We start by bringing together end-users from the community through co-creation workshops to help understand what are meaningful conversational interventions and therein develop and design a prototype. In development, we explore the state-of-the-art methods for Natural Language Understanding (NLU) and Natural Language Processing (NLP) to extract information from user responses and integrate these to generate contextually relevant chatbot responses that are non-repetitive. Accordingly we make the following contributions:

- present a co-creation method to identify and design 5 conversational AI skills and associated educational content required to promote physical well-being among older adults;
- develop a personalised response generation strategy that combines information extraction from open ended user responses to contextualise a template-based NLG method;
- create a cloud-based architecture for secure storage and integration with a cross-platform chatbot app; and
- carry out a comprehensive evaluation of the 5 conversation AI skills using a thematic analysis of think aloud sessions.

Rest of the paper is organised as follows; related literature is explored in Section 2 and Methods for identifying conversational skills, extracting information and response generation are explored in Section 3. Next we present the conversation design for each skill in Section 4, followed by the prototype implementation details in Section 5. Qualitative evaluation is presented in Section 6 and concluding remarks are mentioned on Section 7

2 Related Work

Conversational agents have been used as intervention delivery methods in many healthcare application domains including mental health (Morris et al., 2018; Inkster et al., 2018; Suganuma et al., 2018), weight loss and obesity (Stein and Brooks, 2017; Addo et al., 2013), alcoholism treatment (Lisetti et al., 2011, 2013), physical activity and diet (Fadhil et al., 2019; Fadhil and Villafiorita, 2017) and medication adherence (Fadhil, 2018). But this is lacking in applications which target general fitness. Existing smart phone applications restrict user responses to a selection from a number of choices or through free text entry. Initial research has explored the use of web based avatars to integrate voice and emotions into intervention delivery (Lisetti et al., 2013). However voice based conversational interfaces in the form of chat-bots are more naturally intuitive, compared to these web based avatars. Here good conversational coverage is essential to ensure that the learning curve is manageable without requiring the user to memorise key phrases to carry on a dialogue with the tool.

A recent evaluation of Wysa (Inkster et al., 2018), a text/multiple-choice empathetic AI chat-bot for mental well-being, focused on analysing user acceptance of conversational agents. Their findings suggests that a majority of 67% found Wysa to be a “Favourable Experience” compared to 32% who found it to be a “Less Favourable Experience”. Users preferred to respond by clicking on options given by the app when compared to entering free text. Lark [1] is another well-known text/choice based Conversational Agent specialised in diabetes management and Stein and Brooks (2017) evaluates Lark for user acceptability and satisfaction where users rated the app at 7.9 (average) on a 0-10 scale. These studies suggest that in general conversational agents are widely accepted by the users, but they are limited to text or choice based responses. This motivates us to exploit advances in conversational AI and explore conversation as a form of delivering interventions in general fitness applications, specifically for older adults.

Recent literature suggests a corpus-based approach for enforcing empathy into text/choice based conversational bots (Morris et al., 2018). A corpus is curated with empathetic responses that will be used by the conversational agent when responding to a user. They measure the acceptability of empathetic responses presented by the bot compared to responses presented by a peer and found that users accept bot responses 79% of the time. Our work is closely related to this approach, where we also create a response bank. We acknowledge that there is a significant burden on knowledge engineering however we overcome this by using several user co-creation activities with a view to creating a conversational agent that is designed by the users for the users.

3 Method

We identify three main steps to delivering personalised conversation to encourage physical activity. Firstly, understanding the interesting topics of conversation, secondly, extracting information from the user to contextualise the conversation and finally, generating non-repetitive responses to encourage long term engagement. In this section we detail our methods applied to realise these steps for FitChat.

3.1 Identifying Conversational Skills

We adapt co-creation methodology to identify the most effective conversational skills expected in a

[1] https://www.lark.com/outcomes

Table 1: Natural Language Understanding skills

Skill	Information	Conversation Type
Personalisation	name, age, gender, height, weight and location	question-answer
Weekly Goal Setting	daily step goal, activity plan for each day of the week	semi-structured/open-ended
Daily Activity Reporting	number of steps, activity and the duration, the reason for not doing a planned activity	semi-structured/open-ended

conversational AI for encouraging PA among older adults. We followed an iterative refinement process (Augusto et al., 2018) and conducted three workshops where the intended stakeholders from the community were invited to participate. Each workshop was held one-month apart allowing the stakeholders to learn the capabilities of the technology and explore and refine requirements iteratively. It was specifically significant given the novelty of the conversational technology within the intended age group.

Workshop 1 introduced participants to the study and the concept of voice based conversational interventions. We sought their views on skills including goal setting and reporting that were proposed by the research group to "break the ice" and start the conversation. In workshop 2 a participatory method was followed (Leask et al., 2019). Role playing (Matthews et al., 2014) activities among workshop participants helped understand expectations where the aim was to observe the forms of natural dialogue that transpired between pairs. Participants designed conversation interactions that would allow a user to record their daily activities or that would allow a user to set goals for the coming week. Workshop 3 reviewed and refined the conversational interactions discussed in the previous workshop. Together with the research group, stakeholders prioritised the list of conversations skills identified during the workshops and short listed the top five skills that form the final prototype of FitChat. Workshop 3 also encouraged the participants to propose a name for the conversational AI where they selected the title "FitChat", inspired by the Doric term "Fit like?" (Hello, How are you?). Following are the shortlist of five skills or *intents* identified during the co-creation activities.

Personalisation: The goal is to provide a personalised experience throughout the application. This skill is likely to be used only once during the initial on-boarding process.

Weekly Goal Setting: The Goal Setting intent is aimed towards making a conscious commitment to specific physical activity goals that are considered to enforce positive behaviour change (Michie et al., 2013). This skill is expected to be used at the beginning of every week.

Daily Reporting: The Reporting intent is aimed at enabling conversation about daily activities. This conversation can be aligned with the goals set for the day and would be encouraging to the user to out-perform themselves next day. This skill is likely to be used at the end of every day.

Weekly Summary: The Summary intent is aimed at providing the user a retrospective look at the of last week's goal achievement. This skill is likely to be used at the end of every week.

Exercise Coach: The purpose of the Exercise Coach intent is to guide users to perform exercises by providing exercise steps through read-aloud instructions in a conversational format. This skill is likely to be invoked multiple times (minimum of twice) a week, according to WHO physical activity guidelines.

3.2 Natural Language Understanding

Once the conversational skills are identified, we examine Natural Language Understanding (NLU) capabilities required in each skill to maintain a cohesive personalised conversation with the user. We identify that personalisation, weekly goal setting and daily reporting are the main three skills that are focused on extracting information from the user. The role-playing co-creation activity further identified the types of information each skill should extract in order to personalise or contextualise future conversation. Note that conversation here covers question and answer forms that can be are both open or close-ended. We summarise our findings in Table 1.

3.3 Natural Language Generation

Table 2: Information for Natural Language Generation skills

Skill	Information to deliver
Weekly Goal Setting	The WHO recommendations for daily step count and physical activities.
Weekly Summary	Summary of activities and steps recorded during the last seven days
Exercise Coach	exercise steps for three types of exercises, balance, strength and flexibility.

To create contextually relevant responses, firstly we look at what information is required to deliver the response through conversation; and secondly, how we can make that conversation personalised. To address the first concern, we identified three skills essential to deliver information to the user. These are listed with the information they deliver in Table 2. We consider two aspects for personalisation; contextualisation with information extracted at the NLU phase, and generating non-repetitive responses with a corpus or message bank. Two skills stand to benefit from contextualisation as shown in Table 3 and three skills benefit from non-repetitive motivational responses as shown in Table 4.

Table 3: Personalisation of Natural Language Generation skills

Skill	Contextualisation Information
Daily Activity reporting	The planned daily step goal for the week, planned activities for the day
Weekly Goal Setting	age, height, weight and gender for appropriate WHO recommendations
Weekly Summary	The planned daily step goal for the week, planned activities for each day of the past week

Table 4: Non-repetitive Motivational Responses

Skill	types of motivational messages
Daily reporting	Positive message if activities or steps reported, an encouraging message based on their reason if they fail to perform a planned activity
Weekly Goal Setting	Positive message
Weekly Summary	Positive message
Exercise Coach	Positive message after each exercise set

In order to ensure the non-repetitive behaviour, we adapt a similar methods to corpus-based methods used in literature (Morris et al., 2018) to develop a Motivational Message bank organised under three main categories. Firstly we create a bank of general motivational messages for when the user reports on completed activities; for instance a message such as “Well Done! Regular physical activity is really good for your well being” is uttered by the agent at the end of reporting. Secondly a set of messages to be used when a user does not perform a planned activity due to a specific barrier. Messages in the barriers category are grouped under six barriers that are commonly found in literature (these include Family, Support, Tiredness, Work, Time and Weather). The aim is to deliver a personalised and empathetic response when a user is unable to perform an activity. This message bank is integrated with Reporting, Goal Setting and Summary skills. We started with 20 messages and it is updated regularly. We include few examples from the message bank in Table 5.

4 Conversation Design

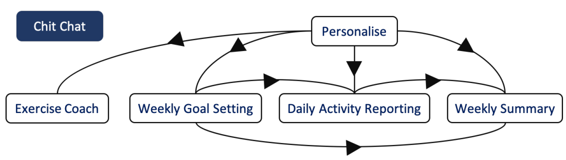


Figure 1: Conversation design pathways

Figure 1 illustrate the 5 skills we identified in Section 3.1. The arrows indicate how information

Table 5: Motivational Message Bank Examples

Achievement	Type	Barrier	Example
Positive	Steps	-	You are doing well with your steps... Keep it up!
Positive	Activity	-	Well done John, You have completed all your planned swimming sessions for this week.
Negative	Steps	-	You are taking less steps compared to last week. Think of more ways to add steps to your daily routine.
Negative	Activity	-	You have completed less exercise sessions than last week. Try to add more sessions to your weekly schedule.
Negative	Activity	Family	Try and exercise with family members. Go for a walk, play tag, put on some music and dance! You'll spend time together and increase your step count
Negative	Activity	Weather	Keep a positive attitude and embrace the weather - walking in the rain can be invigorating and you can go home for a warm shower afterwards!

extracted using NLU will be used to contextualise a conversation skill. For instance, the step goal information extracted during the Weekly Goal Setting skill will be used to contextualise the Daily Activity Reporting skill. In addition to the 5 identified skills we include a chit-chat skill such that the user is able to carry on an informal conversation about generic topics such as weather or news if needed. This will increase the usability of the conversational bot application giving the user more freedom. Next we look at each skill in depth exploring how the conversations are designed and responses are generated.

4.1 Personalisation

Personalising skill is a simple question and answer conversation that is repeated until information required is extracted from the user. The questions are designed such that the goal of each question is to extract a single piece of information. This results in easier extraction of information compared to more open-ended questions and answers. For instance, an open-ended approach would be to ask "Tell me about yourself?" and ask follow-up questions for each missing information. Instead we design a simpler approach; a list of questions such as "What is your name?", "What is your age?", "How tall are you?" to extract information independently. We argue that this approach is suitable here, because the personalising skill is typically used once at the the start.

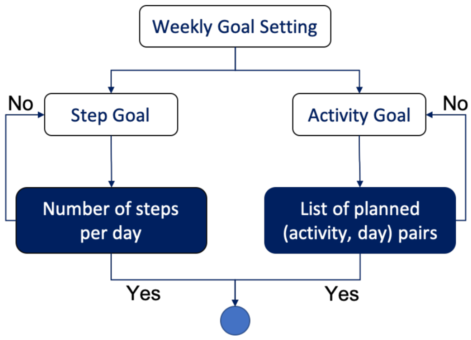


Figure 2: Weekly Goal Setting

4.2 Weekly Goal Setting

During co-creation, participants identified the limitations of fitness apps in recognising physical activities that are beyond ambulatory activities (for instance activities like dancing or tai-chi). Accordingly they proposed two types of goals; steps goal and activity goals. The conversational agent starts the conversation by understanding the type of goal the user wants to set, then guides the user towards providing information required. For a step goal, we extract the number of steps the user plans to complete each day of the week. For an activity goal, we extract one or more activities and the respective day the user plan to perform each activity.

4.3 Daily Activity Reporting

Depending on the type of goal being set for the day the conversational agent initiates a contextually relevant dialogue with the goal of extracting activities the user had undertaken during the day. For this

Table 6: Goal Setting Information Extraction

Goal Type	Agent Utterance	User Response	Information
Step	How many steps do you plan to complete a day?	around 8000 steps	steps per day = 8000
Activity	Tell me about the activities you have planned this week.	I will be swimming on Monday morning. Then some golf on Tuesday and Thursday	Activity List = [(Swimming, Monday), (Golf, Tuesday), (Golf, Thursday)]

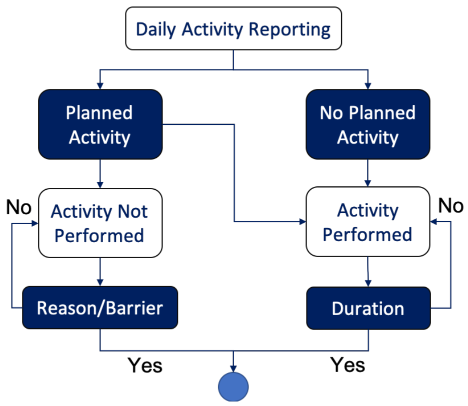


Figure 3: Daily Reporting

purpose the data obtained from the Goal Setting template is retrieved to form the context of the conversation. In Figure we can see that there are two main conversation pathways a user can be directed towards: either the user has performed one or more physical activities and they record them with the Agent, or the user has not performed any physical activities and records a reason (e.g. a barrier). At the end of a conversation pathway, the Agent is designed to respond with an appropriate motivational message from a message bank. These are selected based on the pathway and the reason (i.e. barrier) for when a user has not performed an activity and more details on the message bank. (See examples in Table 5)

4.4 Weekly Summary

Information extracted from the Goal Setting and Reporting templates provide the context of this conversation by highlighting goals achieved and reported physical activities. A motivational message is included at the end of the summary to encourage the user to maintain or improve their performance next week. Here the non-repetitiveness is preserved by diversifying the motivational messages.

"Hello {**name**}! You did {**average number of steps per day**} steps in average per day last week. This is an {**increase/decrease**} from the average from week before, {**motivational message: positive/negative**}. You also did following activities last week; {**activity**} on {**day**} and {**activity**} on {**day**}. {**motivational message: positive**}}!"

4.5 Exercise Coach

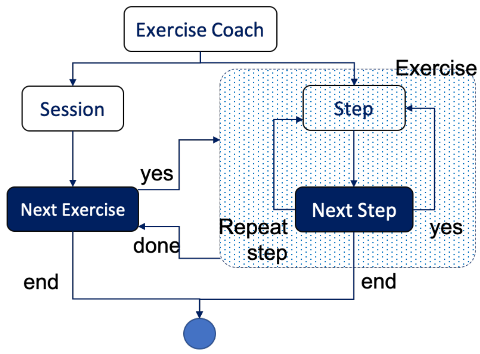


Figure 4: Conversation design pathways

Exercise coach intent can be invoked in two alternatives formats: a single exercise at a time; or a set of exercises curated by a physiotherapist as a single plan to be performed. A detailed view of the conversation flow is illustrated in Figure 4. Parts of this intent can be viewed as an instructional read-aloud function, with the added functionality to enable voice-commands that enable the user to interact in real-time. Example voice commands include: *next* (move to next step or exercise); *repeat* (repeat the current step); and *all steps* (read out the entire exercise).

5 Prototype Implementation

A robust system with minimal maintenance requirements was designed to achieve rapid proto-

Table 7: Daily Activity Reporting Information Extraction

Activity Type	Context	Agent Utterance	User Response	Information
Steps	height, gender	How many steps did you do today?	I did around 2km, not sure about the number of steps	Approximate no of steps:4500, Distance: 2km
Activity	None	Did you do any physical activities today?	Yes. I went dancing with a friend for an hour	Activity: (Dancing, 1 hour)
Activity	planned act.	Did you go swimming today?	Yes I did swim for 2 hours this morning	Activity: (Swimming, 2 hours)
Activity	planned act.	Did you go swimming today?	No, I was not feeling well, may be tomorrow	Barrier: illness

typing. The overall architecture is illustrated in Figure 5. The FitChat mobile application consists of three components: the conversational framework, cloud backend and the smart phone application. DialogFlow implements the conversational intents, while the smart phone application contains the voice based chatbot and the step counting (Traxivity) components. A cloud based micro-services architecture is used to develop the backend enabled by Firebase services. In next sections we will discuss each component in detail.

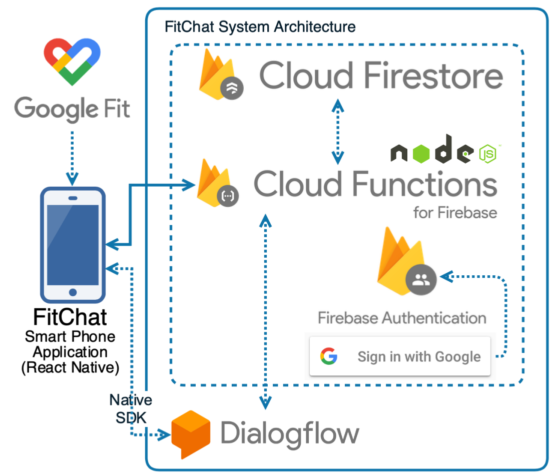


Figure 5: System Architecture

5.1 Intent Implementation

A comparative study of conversational AI frameworks was conducted before choosing one for FitChat intent implementation. We considered three frameworks; Amazon Lex, Google DialogFlow and the Open-source framework Rasa. Both DialogFlow and Lex frameworks offer the flexibility of using additional backend services and mobile integration support that enable building an end to end solution. Rasa, as a open-source platform enables the flexibility to customise conversational techniques and algorithms. We chose DialogFlow as it provided better integration flexibility with mobile applications and was more versatile due to its in-built general conversational intent library. It seamlessly linked with other Google services and most importantly the text to speech functionality was the best among the alternatives. DialogFlow intents are created by adding appropriate training phrases and responses. Each intent calls for custom "Entities" to implement Natural Language Understanding. For instance, goal setting intent defines Entity "ActivitySchedule" that extracts a list of day and an activity pairs from the user response as in Figure 6. Here a combination of regular expressions and response construct structures are used to extract information by matching against the user utterances.

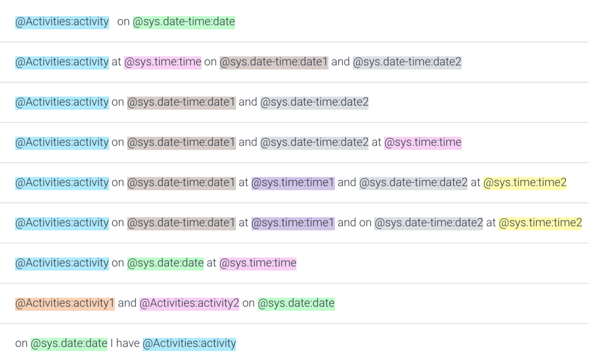


Figure 6: ActivitySchedule Entity for Goal Setting Intent

5.2 Smart Phone Application

FitChat was made available to the end-users via a smart phone application. An analysis of development support for conversational AI concluded that a cross-platform development framework such as React-native to be more suitable compared to native platforms such as Java for Android or Swift for iOS. Low time consumption and minimal development overhead were two deciding factors that emerged in our analysis in favour of React-Native. We made use of the data obtained through Google Fit APIs to record step counts, distance travelled and calorie information for physical activity for Traxivity.

User interface and user experience design of the system is crucial, specifically for the target audience of older adults; accordingly, the application was designed and iteratively refined based on feedback from co-creation workshops (Figure 7). The home screen of the application is set to the FitChat voice bot and the second tab contains the Traxivity component for physical activity tracking. Preferences and manual goal setting can be navigated through the side menu. Additionally, the FitChat bot can be customised for voice speed, pitch and the voice persona according to the user preferences.

5.3 Cloud Backend

We use Firebase which is a Backend-as-a-Service solution by Google for business logic, authentication and data storage implementation. Firebase Cloud Functions (FCF) service is used to run the backend logic which is developed using NodeJS. FCF service also facilitates Firebase Firestore database read/write access which is a flexible NoSQL database where we store user data for contextualising the conversation. The FitChat application uses the Firebase Authentication service along with Google Sign-In. Google Sign-In enables a seamless on-boarding process for the end-user eliminating the need for filling registration forms and increases security by eliminating the need to manage application specific user login details. The cloud backend is developed to be scalable and is server-less with no infrastructure maintenance.

6 Qualitative Evaluation

Think Aloud sessions were planned to evaluate the first prototype of FitChat that allowed generating real-time feedback on the intervention. Think aloud methods are frequently used for usability testing of e-Health applications (Maramba et al., 2019), and involve participants thinking out aloud whilst they perform a task, or immediately afterwards (Eccles and Arsal, 2017).

6.1 Evaluation Protocol

Seven participants took part in five think aloud sessions. Only four participants of the seven had previously participated in co-creation workshops. During the sessions, participants explored the features of the application with minimal input from the researchers. The participants were asked to discover skills and only if needed a keyword is given to initialise the conversations. The think aloud sessions were audio recorded and the feedback was arranged in to themes. Given the data privacy issues, personalisation skill was not included in the evaluation protocol, instead the users were asked to use an existing profile. Exercise coach feature was also excluded from the study given the ethics requirements surrounding study participants performing impromptu exercises without evaluating the physical readiness. Accordingly, we evaluated the efficacy of three conversational skills and observed the general feedback on the FitChat application as a whole.

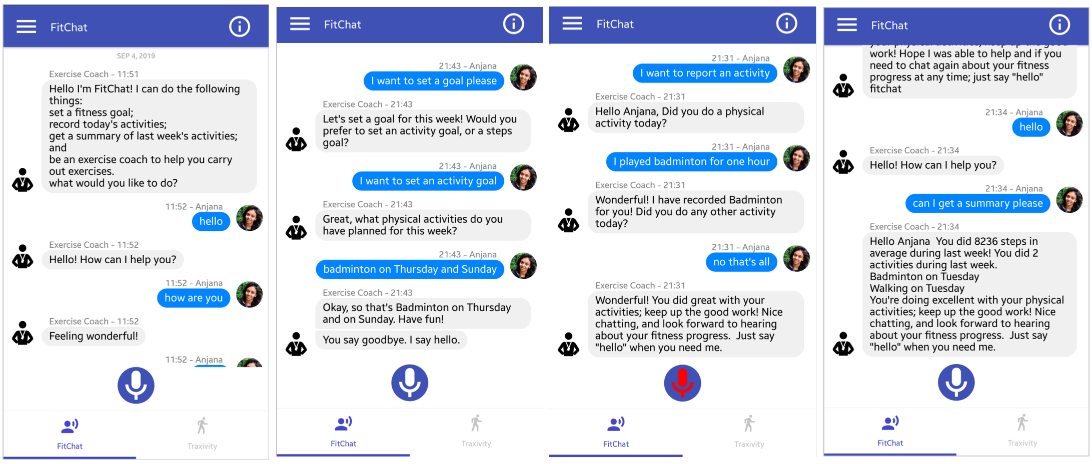
6.2 Thematic Analysis

The thematic analysis arranged the feedback from the users in to five themes; Weekly Goal Setting, Daily Activity Reporting, Weekly Summary, General Feedback and Conversation. These results highlight that all three conversational skills that were included in the this evaluation are identified as core components of the FitChat application. This further validates the selection of these skills over many other that were suggested. The participants of the study had many positive feedback for these features as well as some pointers for improvement as below.

6.2.1 Weekly Goal Setting

Participants generally responded positively to the goal setting feature. However, they commonly said "I want to set an activity/steps goal", rather than "I want to set a goal", and then selecting an activity goal, as instructed in the step-by-step dialogue. Participants commonly suggested that the goal setting feature could be improved if their goals could be stored for longer than one-week:

> *P5: "[be]cause its easier setting up like that and then cancelling"*



(a) FitChat (b) Weekly Goal Setting (c) Daily Activity Reporting (d) Weekly Summary

Figure 7: FitChat Android Application

> *P6: "yes.. would be useful to set this up for a longer period"*

Another suggestion was to include a reminder feature using the goals that are created:

> *P1: "I've got Pilates on a Tuesday, if it could remember every Tuesday and it would remind me"*

6.2.2 Daily Activity Reporting

Participants responded well to the reporting feature and were particularly impressed that the app could link their reported activities with their activity goals, and the motivational feedback:

> *P1: "I like the 'well done...' I think that's important to say well done, you've achieved that"*
> *P2: "I think the very fact that you have to report what you have or haven't done if you set a goal would kind of make me get out of the sofa"*

Most participants suggested that further detail should be added to the activities that are reported to the FitChat application such as calories burned or intensity:

> *P2: "slightly more detail, in that was it a leisurely swim, I don't know how you're going to phrase it, or was it a power swim?"*
> *P2: "we could walk 10,000 steps but strolling does nothing for us, so does that come into it?"*

6.2.3 Weekly Summary

Feedback was overwhelmingly positive for this feature, with several participants outlining its motivational aspect. It was also suggested that it would be useful if this feature included a comparison of summaries in order for participants to reflect on differences in physical activity:

> *P6: "it might say well done last week you did 40 minutes, the week before you did whatever"*
> *P7: "I think that would motivate me that I see I've done that"*
> *P6: "by doing that you can sort of maybe set a different goal for yourself, oh I'll have to beat that next week sort of thing"*

6.2.4 General Feedback

Identification of effective skills to minimise complexity is important usability aspect for conversational bots. A complex solution will introduce a learning overhead to the user which is not desirable specially among the older adults. At times, participants did not intuitively converse with the expected phrases/terms required to interact with the application; however, they quickly learned the terminology during the think aloud sessions. They acknowledged that the app was easy to use once they were familiar with the terminology but expressed that a more complex system would discourage them:

> *P1: "you would get into this lingo because it's obviously got lingo that you have to tap into"*
> *P3: "I think we would learn very quickly"*
> *P4: "it takes a wee while to know the*

tricks like"
P6: "its easy, simple to use, it's just because at the moment its very word specific and restrictive"

6.2.5 Conversation

With regard to the conversational component of this app, the feedback was largely positive:

> *P1: "I think it's one stage up from a Fitbit definitely because you can interact with it"*
> *P3: "it's quite powerful the speaking bit"*
> *P4: "because you have to speak and listen , aye, you're almost admitted to yourself, it's a bit more, you take it more to heart than just clicking a button"*
> *P6: "talking is a lot simpler I think, certainly with older people"*
> *P7: "conversation is far more motivating"*

Some participants suggested that improvements to the conversation could include a greater variety of responses. Some also expressed that they would prefer more informal language:

> *P4: "[be]cause you'll pay attention and kind of look forward to it's going to be a different form of praise every week"*
> *P5: "just a small criticism, eh when I first read my summary, that feedback, I got all these fancy words"*

7 Conclusion

In conclusion, we have identified that conversation has great potential to deliver effective Digital Behaviour Change Interventions to encourage physical activity in older adults. In this work we harness recent chatbot technology advances to build the voice based Conversational AI bot "FitChat". We identify the essential features of such an intervention with older adults from the community through co-creation workshops and implement conversational skills using the state-of-the-art NLP and NLG techniques available for prototyping. We evaluated the first prototype through think aloud sessions. Thematic analysis of the think aloud session outcomes suggests that voice is a powerful mode of delivering motivational content and simplicity is the key to a successful deployment. Generating authentic non-repetitive responses remains a real challenge for chat based digital interventions. Involving uses in the curation of a corpus to address this challenge is an initial solution; however we need methods to evolve and enhance this corpus through system usage; and how to do this efficiently remains an open problem. In future we plan to integrate more customised AI algorithms for NLP and NLG and further improve the quality of the conversation.

Acknowledgments

This project is funded by the GetAMoveOn Network+ (funded by the Engineering and Physical Sciences Research Council, UK (EPSRC) under the grant number: EP/N027299/1) and the SelfBACK Project (funded by the European Union's H2020 research and innovation programme under grant agreement No. 689043). We like to thank everyone involved in the project in many capacities.

References

Ivor D Addo, Sheikh I Ahamed, and William C Chu. 2013. Toward collective intelligence for fighting obesity. In *2013 IEEE 37th Annual Computer Software and Applications Conference*, pages 690–695. IEEE.

J Augusto, Dean Kramer, Unai Alegre, Alexandra Covaci, and Aditya Santokhee. 2018. The user-centred intelligent environments development process as a guide to co-create smart technology for people with special needs. *Universal Access in the Information Society*, 17(1):115–130.

David W Eccles and Güler Arsal. 2017. The think aloud method: what is it and how do i use it? *Qualitative Research in Sport, Exercise and Health*, 9(4):514–531.

Ahmed Fadhil. 2018. A conversational interface to improve medication adherence: Towards ai support in patient's treatment. *arXiv preprint arXiv:1803.09844*.

Ahmed Fadhil and Adolfo Villafiorita. 2017. An adaptive learning with gamification & conversational uis: The rise of cibopolibot. In *Adjunct publication of the 25th conference on user modeling, adaptation and personalization*, pages 408–412. ACM.

Ahmed Fadhil, Yunlong Wang, and Harald Reiterer. 2019. Assistive conversational agent for health coaching: A validation study. *Methods of information in medicine*.

Becky Inkster, Shubhankar Sarda, and Vinod Subramanian. 2018. An empathy-driven, conversational artificial intelligence agent (wysa) for digital mental well-being: real-world data evaluation mixed-methods study. *JMIR mHealth and uHealth*, 6(11):e12106.

Calum F Leask, Marlene Sandlund, Dawn A Skelton, Teatske M Altenburg, Greet Cardon, Mai JM Chinapaw, Ilse De Bourdeaudhuij, Maite Verloigne, and Sebastien FM Chastin. 2019. Framework, principles and recommendations for utilising participatory methodologies in the co-creation and evaluation of public health interventions. *Research involvement and engagement*, 5(1):2.

Christine Lisetti, Reza Amini, Ugan Yasavur, and Naphtali Rishe. 2013. I can help you change! an empathic virtual agent delivers behavior change health interventions. *ACM Transactions on Management Information Systems (TMIS)*, 4(4):19.

Christine L Lisetti, Ugan Yasavur, Ubbo Visser, and Naphtali Rishe. 2011. Toward conducting motivational interviewing with an on-demand clinician avatar for tailored health behavior change interventions. In *2011 5th International Conference on Pervasive Computing Technologies for Healthcare (PervasiveHealth) and Workshops*, pages 246–249. IEEE.

Inocencio Maramba, Arunangsu Chatterjee, and Craig Newman. 2019. Methods of usability testing in the development of ehealth applications: A scoping review. *International journal of medical informatics*.

Mark Matthews, Geri Gay, and Gavin Doherty. 2014. Taking part: Role-play in the design of therapeutic systems. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '14, pages 643–652, New York, NY, USA. ACM.

Susan Michie, Michelle Richardson, Marie Johnston, Charles Abraham, Jill Francis, Wendy Hardeman, Martin P Eccles, James Cane, and Caroline E Wood. 2013. The behavior change technique taxonomy (v1) of 93 hierarchically clustered techniques: building an international consensus for the reporting of behavior change interventions. *Annals of behavioral medicine*, 46(1):81–95.

Robert R Morris, Kareem Kouddous, Rohan Kshirsagar, and Stephen M Schueller. 2018. Towards an artificially empathic conversational agent for mental health applications: system design and user perceptions. *Journal of medical Internet research*, 20(6):e10148.

Leanne G. Morrison, Charlie Hargood, Veljko Pejovic, Adam W. A. Geraghty, Scott Lloyd, Natalie Goodman, Danius T. Michaelides, Anna Weston, Mirco Musolesi, Mark J. Weal, and Lucy Yardley. 2018. The effect of timing and frequency of push notifications on usage of a smartphone-based stress management intervention: An exploratory trial. *PLoS ONE*, 13.

Natalie Stein and Kevin Brooks. 2017. A fully automated conversational artificial intelligence for weight loss: longitudinal observational study among overweight and obese adults. *JMIR diabetes*, 2(2):e28.

Shinichiro Suganuma, Daisuke Sakamoto, and Haruhiko Shimoyama. 2018. An embodied conversational agent for unguided internet-based cognitive behavior therapy in preventative mental health: feasibility and acceptability pilot trial. *JMIR mental health*, 5(3):e10454.