Co-Designing Voice User Interfaces with Teenagers in the Context of Smart Homes

Daniel Fitton DBFitton@UCLan.ac.uk
Janet C Read JCRead@UCLan.ac.uk
Gavin Sim GRSim@UCLan.ac.uk
Brendan Cassidy BCassidy1@UCLan.ac.uk
ChiCI Research Group,
University of Central Lancashire,
Preston, UK.

ABSTRACT
This paper explores the design of voice user interfaces for smart homes with teenagers. The work was motivated by two research questions: How can we co-design voice interfaces with teenagers? and What ideas and expectations do teenagers have in relation to voice interfaces in a smart home? A design process was used which involved the participants initially scripting exchanges with a smart home on paper then prototyping at a higher level of fidelity using a tablet app with speech output. The study was carried out in a high school in the UK with 55 pupils in Year 10 (14-15 years old). This work is the first of its kind to explore the co-design of VUIs with teenagers. The key contribution of this paper is the design method that was used which proved successful and gave insights into the use of dual prototyping fidelities and the impact of scaffolding on the designs produced. Other contributions include the themes which emerged from the designs and a set of four themes related to teenagers' expectation of smart homes. The wide range of findings reported in the paper also bring insights that are valuable to those wishing to design and develop VUIs with and for younger users.

Author Keywords
Design; Co-design; Teenagers; Adolescents; Voice User Interface; Smart Home

ACM Classification Keywords
CCS → Human-centered computing → Human computer interaction (HCI) → HCI design and evaluation methods

INTRODUCTION
Voice User Interfaces (VUIs) have seen a recent resurgence since the release of Apple's Siri on phones and tablets in 2011. More recent VUI products such as the Amazon Echo and Google Home, released in 2015 and 2016 respectively, are specifically intended for use in the home. These modern VUIs typically integrate with existing services to enable voice interaction where an app/web site would previously be used. This model is most obvious with the Amazon Echo where specific 'skills' can be enabled which often correspond to integration with specific services (e.g. travel, music, news, weather). This integration also extends to commercially available 'smart' home technologies (such as Philips Hue light bulbs, Hive heating controls etc.). Such integration has been specifically used in the marketing strategy of Amazon's Alexa device range. While smart homes with VUIs become increasingly popular, very little work has been done on the design or co-design of voice interactions in this context. Additionally, while new technology products are naturally targeted at adults, who have the spending power to purchase them and their associated services, once they are installed in the home they are likely to be adopted and/or appropriated by younger family members. In this work we sought to explore the design of new ideas for the use of VUIs within the context of a smart home. Teenagers in particular were chosen as the participants in this work as they are often early adopters and prolific users of technology [6]. The key questions addressed in this work are:

- RQ1. How can we co-design voice interfaces with teenagers?
- RQ2. What ideas and expectations do teenagers have in relation to voice interfaces in a smart home?

In order to explore these questions a co-design study, with design tasks incorporating two levels of fidelity, was run in a school in the UK with 55 teen participants. Scaffolding was used to enable participants to rapidly engage in the design activity with minimal instruction [19]. After the design task was completed a questionnaire was used to gain insights into individuals' understandings of voice interfaces and their potential uses.

The main contribution of this paper is the design method, the first of its kind to be used with teenagers. Use of the design method enabled insights into the use of dual prototyping fidelities and the impact of scaffolding on the designs produced, which are also reported in the paper. Other

IDC '18, June 19–22, 2018, Trondheim, Norway
© 2018 Association for Computing Machinery.
ACM ISBN 978-1-4503-5152-2/18/06…$15.00

https://doi.org/10.1145/3202185.3202744

contributions arose from the analysis of the study data which, after the scaffolding prompts used in the design method, gave 10 categories for potential new application areas along with themes relating to the expectations the teen participants had for a smart home.

RELATED WORK

The notion of conversing with technology can be traced back to the 1950's and the Turing Test [25]. Since then a wide range of conversational agents have been explored and voice has become a popular way of interacting with the Virtual Private Assistants (VPAs) provided by modern technologies such as Microsoft's Cortana, Google Now/Google Assistant, Amazon Alexa, and Samsung S Voice [9]. These VPAs provide a voice interface to a conversational agent which mimics human conversation and, despite a previously uncertain relationship between HCI and Speech-based interfaces [1], they have begun to become the focus of recent research within the CHI community. For example, Luger and Sellen [14] have studied the mismatch between the expectation and experience for 14 adults who claimed to 'regularly' use Siri and Google Now. The findings were primarily related to the shortcomings of the technology (errors in speech recognition and failures in task completion) which negatively impacted trust and led to only simple tasks being carried out using simplified vocabulary. Additionally, participants without technical knowledge had higher expectations of capability (compared to those with technical knowledge) and were more likely to blame themselves (rather than the technology) when tasks failed. Other findings were that speech-interaction was useful for multi-tasking, voice being a means to carry out peripheral tasks hands-free, and that almost all (13 out of 14) participants had been introduced to the speech interface through play/novelty.

Porcheron et al. [16] have explored the use of VPAs in multi-party social situations where face-to-face conversations are deliberately interrupted to pose a question or request to the VPA. An ethnomethodological approach is used in this work and key findings related to the need to repeat and refine queries, the humanlike qualities attributed to the VPA, and the collaborative opportunities for interaction which emerged when using a personal device as a VPA in a social situation. Purington [17] explored the personification of Amazon's Alexa through analysis of user reviews posted on the Amazon web site, finding high levels of personification and evidence that personification increased satisfaction (in terms of product reviews of the device).

Research focussing on the design of speech-based interfaces is a relatively sparsely populated space. In 1989 Rudnicky [20] published work on designing a 'voice-driven interface' for a spreadsheet system through sampling the language used in a simulated task. The work considered issues of 'fluency' and voice recognition performance. In 2001, Cassell et al. [2] reported on the design of an embodied conversational agent and presented a set of design requirements targeted at enabling the agent to converse like a human along with details of the underpinning software architecture they developed. Related work such as [3] followed an iterative design process (involving sketching and Wizard of Oz studies) using Speech Act Theory to explore the design of speech interfaces in the context of an automated home with users with physical and speech impairments.

In the CHI research community, it is common to take a user-centred approach to creating technology [5]. However, very few examples of research work with children, or teenagers, exist that consider the design of voice interfaces specifically. The most closely related work in this domain has studied how younger children interact with contemporary VPAs to explore how these voice interfaces are used. In the study by Druga et al. [4] 26 children aged between 3 and 10 interacted with four different conversational agents (Google Home, Amazon Alexa, Anki's Cozmo and Julie Chatbot) for 15 minutes each then completed a questionnaire activity exploring trust, intelligence, identity attribution, personality, and engagement. Key findings were that all four agents were perceived as friendly and trustworthy, that exploratory and playful questions were posed, and questions were often misunderstood by the agents meaning help was needed from adults/peers to refine or re-phrase.

In other research [13] an online survey was used to gather data from 118 parents on their child's use of 'voice input systems' (31 of which had used voice input) along with content analysis of 40 You Tube videos of children interacting with Apple's Siri. The work focused on children aged up to 7 years and the system used, as reported by parents, was almost exclusively Siri (23/31 – 74%). The authors identify five key categories of use; exploratory (e.g. playful interactions), information seeking, functional (sending messages, making phone calls etc.), abuse (threats and profanities directed at the technology), and accidental (these latter two making up the smallest number of occurrences). In common with [4], a key finding from the authors of [13] was the challenge the young children encountered in successfully interacting with the technology.

In contrast with much of this previous work, we seek to explore design ideas and possibilities for VPAs for smart homes and wish to avoid grounding design activities in these specific technologies, where voice recognition still often proves problematic.

THE STUDY

In order to explore our first research question (How can we co-design voice interfaces with teenagers?) a design activity was planned involving prototypes of two different fidelities. The use of prototypes with different fidelities is often used in the design of new technologies (e.g. [12] and [23]) primarily as a means to enable evaluation. In this work we use prototypes as a means to enable design exploration in order to investigate our second research question (What ideas and expectations do teenagers have in relation to voice interfaces in a smart home?). Using Lim et al.'s prototyping framework [11], the fidelity of the prototyping techniques

was deliberately kept low to filter out the typical problems found with recognition and VPA technology, and the material manifestation used in prototyping was changed between the first and second design activities. In this dual prototyping approach the participants firstly used paper sheets to script imagined voice interactions between themselves and the smart home. Secondly, participants input text intended to be spoken by the VUI into an app running on an Android tablet (as text) which could then be spoken out loud when an appropriate button was pressed. This allowed participants to revise and revisit ideas on paper and then act out their ideas, experiencing the imagined interactions at a higher level of fidelity. Following these two tasks every participant completed an individual questionnaire form. Data collected for analysis included the paper scripts, logs from the Android app, and questionnaires, it was not possible to audio record the sessions.

Participants

The study was carried out with 55 Year 10 pupils at a High School in the UK. Two classes were selected for participation by teachers. The researchers worked with half of a class at a time over four 50 minute sessions on the same day, each group ranged from 12 to 19 pupils. The majority of the pupils were 15 years old with five pupils aged 14 and eight pupils not recording their age. In terms of gender breakdown: 30 pupils were female, 19 male and six pupils did not record their gender. The pupils completed the tasks in self-selected groups of two or three. The study was covered by institutional ethical approval with information and consent sheets being provided by the research team which were distributed by the school. Participants were told on the day that they could withdraw their data by specifying this on the evaluation sheet completed at the end of the study (no pupils chose to do this). All data was collected anonymously.

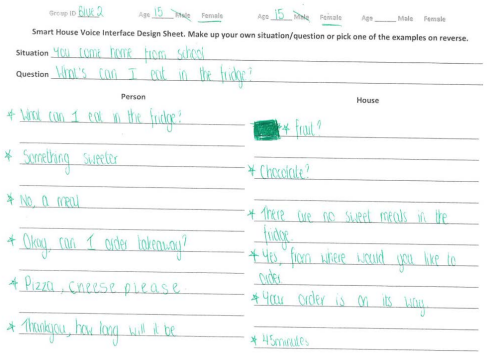

Figure 1. Example of Completed Paper Design Sheet

Study Materials

For the paper design activity, pre-printed sheets were provided (see Figure 1 for a completed example) which asked for the situation, initial question posed to the house, and space to write down the script of the subsequent exchange. Each group was provided with multiple sheets and pens. In the second part of the activity the participants were given an Android tablet (one tablet per group) running an application created by the research team which included a box to enter the initial question posed to the house and 20 text boxes with corresponding 'Play' buttons which when pressed spoke out loud the text that had been input (the interface can be seen in Figure 2). The application used the built-in Android speech synthesis engine and the synthesis engine settings were left at their defaults. The application logged every action (changes to text fields, text spoken out loud etc.) for later analysis. The design of the application was deliberately simple and there were no technical problems with the tablets, or application, during the study. Following the design tasks a single-page questionnaire was given to each participant to complete individually. This included questions asking the participants to rate their understanding of VUIs and smart homes (with 5-point Likert scale answers), as well as open questions about ideas for VUIs. The former questions were included as these two concepts are central to the design studies, the latter questions were used to capture ideas which may not have been shared in the group design activities or may have occurred during them.

Figure 2. Tablet App in Use During Design Session

Procedure

The introduction used a script to ensure consistency between the four design session. The script explained the purpose of the session using a Google Home as a prop when mentioning voice interfaces, then provided a pragmatic introduction to the concept of a smart home (actuation, sensing, access to data and services etc). The script included six example questions that could be asked of a smart house, chosen by the research team as potentially being able to resonate with the teenagers, to help scaffold understanding of the design context. The use of example scenarios likely to be typical in the lives of the participants was partly inspired through its use by others carrying out similar work with speech interfaces (e.g. [27]), and partly to help the participants ground the VUIs in their daily lives. The questions were selected to cover a range of application possibilities and underlying technical complexities within a smart home context, whilst also highlighting potential privacy concerns. The prompting questions span a range of location awareness

possibilities including food, personal possessions, siblings and parents. They also include examples of information seeking behaviour where it is implied that the house is able to understand and respond to a relatively nuanced requests (e.g. 'Is anyone out tonight?'). Through this construction of the prompting questions it was hoped that the participants would include similar (if not more) complex features in their designs. The use of these questions was considered a trade-off in quickly conveying a clear understanding of the design context vs. potentially influencing the designs produced. Only the opening question posed to the smart house was given so that even if the initial part of a design was biased the following dialog would be constructed by the participants themselves. The questions were set in the context of returning home from school as this was deemed to be an almost universally applicable and relatively well understood situation. The prompting questions were:

- What can I eat?
- What has he been up to all day? (in the context of an older sibling)
- Anything good on Netflix?
- Where's mum/dad?
- Is anyone out tonight?
- Where are my swimming goggles?

The researcher introducing the sessions made the point that a smart house may have access to data that could provide answers to these kinds of questions, and that in this session the participants would be constructing their own questions around ideas that would be useful to themselves, and scripting the exchanges between themselves and the smart house. The researcher then explained how to use the paper design sheet and the participants then worked in groups for 15 minutes. Next, the tablet app was introduced and demonstrated to the group. The participants were asked to test (act out) the paper designs they had created using the app, refining the initial ideas and exploring new ideas if they had time. The groups worked with the app for another 15 minutes, then were asked to fill in the evaluation forms individually. If there was sufficient time after completing the forms, each group acted out their favourite idea to the other groups, and before leaving the room each group was thanked for their participation. Two researchers were present for the study and circulated around the room during the design activity answering any questions, and speaking to the groups about their ideas. A teacher was present during the studies but was seated and working in a different part of the room.

Analysis

Analysis was conducted in two stages, with stage one having two phases and stage two having four phases. The first stage was concerned with the questionnaire, the second with the designs created.

In the first phase of stage one, the two Likert questions 'I now understand more about voice interfaces than I did yesterday and I understand what a smart home is', were coded on a scale of 1 to 5 (1 = strongly disagree and 5 = strongly agree). In the second phase content analysis was used to cluster the data derived from the question 'List two things that you think a smart home (with a voice interface) could help you with'. The content analysis began with one of the researchers examining all the statements to derive themes. For example 'Finding', which related to searching for a mobile phone or locating other objects in the house was considered a main theme. In total 10 themes were derived. Three statements that were impossible to decipher were left un-coded. To ensure reliability of the coding, 20 responses were selected, two from each category and placed in a random order in a spreadsheet. Two researchers were given the themes and asked to code this data. The results of the interrater analysis using Cohen's Kappa show a high level of agreement between the two researchers ($k=0.944$ with $p<0.001$).

The second stage of analysis involved four phases. The first phase was to investigate the relationship between the paper designs and those in the Android app. The data from the app logs was processed to extract the questions posed, responses input, and responses spoken by the app. The timestamping of the log entries was used to match the app data with the paper designs from the four sessions. For each participant group the designs on paper and designs from the app were compared, with the relationship between them coded as none, facsimile, evolution or revolution. The code 'facsimile' was used when designs were judged to have a high degree of similarity, 'evolution' was used when there was similarity but also key differences, and 'revolution' when there a small degree of similarity and a high degree of difference. The second phases of analysis determined whether the six example questions posed in the session introduction script appeared in each design set (paper and app) using the same codes (none, facsimile, evolution or revolution). The third phase of analysis was to code the purpose of the underlying functionality evident in the design sets using the 10 themes from stage one of analysis. In the fourth and final phase, the coders were asked to rate the expectations of the smart house capability implied by the design sets on a four-point scale of none (no smartness), using currently available technology, using near future technology, and requiring far future technology.

In stage two all the coding was done independently and in parallel by three coders. The coders initially coded two example design sets together to familiarize themselves with the coding process and clarify their interpretations of the codes. Multiple codes for the same design were allowed to account for designs where an exchange included multiple related tasks. The decision was made that if there was agreement between at least two coders this would be used as a consensus, and where a consensus could not be reached the designs would be discussed by all three coders.

Theme	Freq.	Description
Finding	27	The house is able to locate missing physical objects and people e.g. mobile phone, parent.
Food	19	Assisting the teenager in finding snacks and meals e.g. what is in the fridge, if there are sweets.
Controlling	18	Control household appliances and features e.g. heating and lighting.
Calendar/Events	7	Maintaining personal diaries and enabling organisation and planning of activities.
Homework/ Information	6	The house would assist you with knowledge acquisition, for the teenagers it includes finding information on a subject or helping with specific homework activities.
Security	6	This involves personal security within the home. This may include phoning the emergency services, information about the security home or who is in the home.
Message / Call	6	The house would be able to send text messages to individuals on the users behalf and/or make phone call from anywhere the home.
Advice	4	The house is able to offer personal advice, for example what to wear.
Entertainment	3	The house will entertain the teenager, examples include playing music and telling jokes.
Ordering	2	Placing online orders e.g. takeaway food delivery or goods from web sites.

Table 1: Themes Emerging from the Post-Task Questionnaire

Theme Category	Number
Finding	11
Food	9
Advice	6
Entertainment	4
Calendar/Events	3
Controlling	2
Homework/ Information	2
Security	1
Message / Call	1
Ordering	1
Food AND Controlling	1
Advice AND Entertainment	1

Table 2. Results from Coding of the Purpose of the Designs Created During the Design Activities.

FINDINGS

In this section, we firstly explore RQ2 (What ideas and expectations do teenagers have in relation to voice interfaces in a smart home?) by considering the design ideas and the expectations evident within them. RQ1 (How can we co-design voice interfaces with teenagers?) is then explored through examining several facets of the design method used including the scaffolding questions and dual fidelities.

Design Ideas

In this section, we firstly present results from the first stage of analysis of the 55 post task questionnaires then discuss the stage two analysis of the designs from the 23 groups created on paper and in the app.

Design Ideas from the Questionnaire

In total 10 themes emerged from the 98 design ideas in response to the open question *List two things that you think a smart home (with a voice interface) could help you with*. The themes are presented in Table 1 showing frequencies and example statements. The most popular theme, evident in 28% of the designs, was that of 'Finding'. This was followed by the 'Food' theme, evident in 19% of the designs, which most often related to the participant acquiring a suitable snack or meal. 'Controlling' individual appliances or aspects of the home environment (such as heating and lighting), was evident in 18% of the designs. The remaining 35% of designs aligned with the seven remaining themes shown in Table 1.

Design Ideas from the Design Activities

Across all 23 groups 42 unique designs in total were collected from the scripts created in the design activities (on paper and in the app). Table 2 shows the results from coding the purpose of each of all 42 designs against the themes from Table 1. As outlined earlier, it was possible to code a single design to more than one theme. No new themes were needed during the coding process and only two designs were assigned to multiple categories (as shown in Table 2).

From Table 1 and Table 2 is it clear that the most common theme is 'Finding', with 27% (27 out of 98) of the ideas from the post-study questionnaire and 26% (11 out of 42) of the designs relating to this. The second most common category was 'Food', with 19% (19 out of 98) of the ideas from the post-study questionnaire relating to this, and a slightly higher 21% (9 out of 42) in the case of the paper designs. The results then diverge, for example with the third most popular category being 'Controlling' (18%, 18 out of 98) in the case of the post-study questionnaire, but being 'Advice' in the case of the designs (14%, 6 out of 42). Given the low numbers beyond the most popular themes, the order in which items appear in the lower part of both Tables 1 and 2 is arbitrary and is not worth scrutiny.

Expectations

In this work we define expectations as assumptions evident in the designs which concern the behaviour and capabilities of the VUI and smart home. When rating the expectations of the smart home capability implied by the designs, 32 were rated as being possible with current technology and 10 coded

as requiring near future technology to implement. No designs were coded as having no smartness and no designs were coded as and requiring far future technology. In the proceedings subsections excerpts from design ideas are shown for each of the key themes, following these emergent issues in relation to expectations are discussed.

Finding In the paper design scripts 'Finding' focussed on both people and objects, the objects were typically mobile phones, money and other personal items such as bags and trainers, people were typically the mother or parents. The scripts usually had a sequence of exchanges narrowing down the location and even in one case stating the distance between the user and bag as they were looking for it. This design is from a group that did not give their gender or ages, the situation was having lost a phone:

User: "Where is my phone?"
House: "The last place you had your phone was in the kitchen"
User: "Do you know where I left it in the kitchen?"
House: "Possibly on the kitchen table"
User: "I still can't find it will you call the phone"
House: "Calling phone now"

Food This typically related to what was in the house and around what could be eaten straight away, most often this related to chocolate. In several cases the teenager interrogated the house to find chocolate or the house attempted to dissuade the teenagers from eating chocolate because it was unhealthy. In only two cases did 'food' relate to preparing an actual meal. This design was created by two 15-year-old females, the situation was coming home from school:

User: "What can I eat in the Fridge?"
House: "Fruit?"
User: "Something sweeter"
House: "Chocolate?"
User: "No, a meal"
House: "There are no sweet meals in the fridge"
User: "Okay, can I order takeaway?"
House: "Yes, from where would you like to order?"

Advice A very common query was on what clothes could be worn or what clothes should be worn for a specific event. The house then gave options or made choices and justified them. This design was created by three 15-year-old females, who were getting ready to go out and needed an outfit:

User: "What can I wear to go to a party?"
House: "Here are the options in your wardrobe"
User: "What are the online options?"
House: "Here are some options I've chosen to suit your style"
User: "Please order the red dress"
House: "Any shoes or accessories to suit this dress?"
User: "Yes, please order a new necklace and earrings but I will use my own shoes"
House: "Okay, your dress and accessories will arrive at 5 o'clock and your total is £85"

Entertainment In this category it was often clear that the teenager was bored and wanted entertainment or help making a decision about what to do for entertainment. Queries centred around decided whether to go out or stay in, or what games to play on a computer/console. In this example, the teenager had no one to talk to:

User: "Up for a chat?
House: "What do you want to talk about?"
User: "Omgs! Angie's so annoying she cancelled on me again!"
House: "Which one's Angie?"
User: "The one who broke the door"
House: "Oh! I hate her"
User: "I know!"

Calendar/Events This category was primarily related to checking what events had been organized previously, both school classes and social events. This design was created by two 15-year-old females checking calendars in a morning:

User: "What am I doing on Friday"
House: "Christine's wedding from 12-9pm"
User: "Ok cancel plans with Adam"
House: "Plans with Adam removed do you want to send an email to tell him?"
User: "No, I'll text him myself"

Controlling This category was used for controlling appliances in the home. This group did not give their gender or ages, the situation was them leaving the house and wanting to know if anything electrical has been left on:

User: "What devices are turned on in the house?"
House: "Just checking… You have 17 electronical devices turned on, would you like them to be turned off?"
User: "Yes, I would like to turn off all devices in the house".
House: "OK, turning them off now".

Homework/ Information This category related to use of the home to retrieve information. This design was created by two 15-year-old females and concerned letters about very high bills that had been received in the post:

User: "Why are my bills so expensive?"
House: "One second – let me have a look"
User: "OK thank you"
House: "It's my pleasure"

Security This category related to aspects of personal security within the home. This group did not give their gender or ages, the situation was that they had locked themselves out of their home:

User: "I'm locked out?"
House: "I'll need a quick quiz"
User: "Ok"
House: "Name?"
User: "Charlotte"

House: "Chosen Password?"

Message / Call In this category the purpose of the interaction was to communicate with others outside of the home. This design was created by two 15-year-old females, who wanted the house to leave a note for a parent:

User: "Please will you tell my parents I have gone shopping? thanks"
House: "Who are you going with?"
User: "My friends"
House: "Okay, what time will you be back?"
User: "Around 7:30 today"
House: "Do you need anything else?"
User: "Ask my mum to make my tea"

Ordering The designs in this category related specifically to the ordering of food or goods such as clothes and games. This design was created by three 15-year-old males, and centred around them playing games in their bedroom:

User: "Will you go to PS-Store and buy Minecraft use my credit card details to buy it"
House: "Minecraft has been bought for £31.99 by your credit card"
User: "Check how much money I have left in my bank please?"
House: "You have £293.83 in your bank"
User: "Download Minecraft Please?"
House: "Minecraft is being downloaded as we speak"

The examples given above show that it was very common for the teenagers to design in a very conversational way. In each case the word house could quite easily be replaced with the word 'mum' and the designs would still make sense. In the following sections, the expectations around the conversations with the house are teased out along four themes.

The Role of the House

The role of the house was most often servile, carrying out the instructions of the user without emotion. Occasionally pleasantries were exchanged in a relatively formal way (e.g. please/thanks). In these example, the house usually acted as an interface to existing services, as is the case with many uses of current VUI technology. Colloquialisms and slang words were very rarely used when addressing the house. In some designs, it was clear the house was acting in the way that a parent might, for example asking probing questions with safety in mind (e.g. "Who are you going with?" / "What time will you be back?"), or encouraging the user to act for themselves by giving a partial or uncertain answer (e.g. "Possibly on the kitchen table"). Overwhelmingly the house was compliant and on the side of the user, in some cases deliberately seeming to be ingratiating (e.g. "Oh! I hate her").

Tracking

It was often assumed that the house had accurate knowledge of food contained in the fridge and kitchen, knowledge of a user's wardrobe (including shoes and accessories), and the ability to track the location of portable items, such as phones, with high degree of accuracy. It was assumed that the house could track the location of parents with accuracy on demand. There was no evidence of the movements/activities of the user themselves or their peers (or siblings) being tracked despite the fact that one scaffolding question deliberately encouraged the participants to think out surveillance possibilities ("What has he been up to all day?" in the context of a sibling). It is unclear if this is something they specifically wished to avoid or was a feature they did not consider useful.

Reliable Recommendations

It was expected that the house could make reliable recommendations based on the users' preferences and knowledge of the upcoming event/activity. There were no examples of decisions made by the house being challenged by the users and there was apparent desire to offload decision making processes. In some cases, the house was allowed to make decisions independently (e.g. "Yes, please order a new necklace"), the implied level of trust and expertise here is an interesting area for future work.

Technical Perfection

Design scripts typically had a number of exchanges where additional detail was sought by the user or additional helpful information was provided by the house. The house always had a helpful and correct answer for the user, there were no examples of failure or uncertainty in the responses from the house in any of the designs. Prior work has highlighted the practical problems of using voice interaction for VPA contemporary technologies ([4],[13],[14]) and it is perhaps the unfamiliarity of the participants with these technologies that has led to this omission. Alternatively, if the prototype was of higher fidelity incorporating speech recognition the possibility of failure might have been evident and explored more.

The Design Method

A conventional approach of designing scripts on paper was adopted, followed by a higher-fidelity technique using an app to help act out the scripts. Scaffolding was used to assist the participants in understanding the capabilities and possibilities of a voice interface to a smart home through examples. The design study ran without issue and the participants encountered no problems completing the paper design task and using the app. In the following sub-sections we discuss findings related to understanding of key concepts, use of multiple fidelities in the design task, followed by findings related to the scaffolding questions.

Understanding of Key Concepts

A key concern that the research team had was whether the participants would gain sufficient understanding of the concept of VUIs and smart homes to design effectively. The post-task questionnaire specifically explored participant understanding of these two key concepts and the overwhelmingly positive responses shown in Table 3 were encouraging that our approach had been successful.

Concept	Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree
1. VUI	0	1	6	33	15
2. Smart Home	0	1	3	36	15

Table 3: Frequency of Responses on a 5-point Likert Scale to Statements in the Questionnaire

For the first statement (*I now understand more about voice interfaces than I did yesterday)* the mean score was M=4.13 SD=0.668. This would suggest that participating in the study helped the teen participants learn about voice interfaces, which is encouraging given that the majority had past experience using them. The questionnaire also asked (*Before today have you used voice interfaces?)* and results showed only 5 (9%) had never used a voice interface previously, 38 (69%) claimed to have used Apple's Siri, and the remaining 12 (22%) had used technologies such as Amazon's Echo, Google Voice, Microsoft Cortana, and voice commands on Sony's PS4 console. The second question (*I understand what a smart home is)* had had a mean score of M=4.18, SD=0.61 from which can be inferred that the participants were generally confident they had sufficient understanding of this concept within the context of the design tasks.

Multiple Fidelities

The stage two analysis focused on coding the sets of designs (paper and app) created by the groups. In the designs analysed there was agreement between at least two (of the three) coders for all of the data. Across all 23 groups 42 unique designs in total were collected. In terms of the relationship between the paper designs and app designs, 21 (50%) of the designs input into the app were facsimiles of the previous paper designs, three (7%) were evolutions of the paper designs, and 18 (43%) designs appeared on paper but not in the app logs. These results are what would be expected given the design task which was; firstly to design on paper, then to test the design with the app, then to refine existing ideas or explore new ones.

The three designs input in the app that were coded as evolutions of an initial design, included asking the house to recommend a game (an evolution of 'What can I eat?'), a situation where something was lost outside of the house and the house helped track it down (an evolution of 'Where are my goggles?'), and asking the house for the fastest route to a destination (an evolution of 'What can I eat?'). Only a single example was seen of exploring an entirely new idea in the app, with the question posed being "What is the meaning of life?" and the answer provided being "42" (this was not included in the analysis as it was not possible to obtain approval to voice record the study and so only the part of the design spoken by the house was collected).

As mentioned earlier, the majority of groups selected just one of their initial ideas to explore using the app. From the log files it is clear that each group carefully input the responses from the house into the app user interface and, once they had tested and refined the spoken responses to their satisfaction, assumed the task to be complete. This was only apparent through analysis of the log files as every group appeared to the facilitators to be working intently for the full duration of the second design session, rather than stopping after a design was input successfully. The design of the app was intentionally simple, to avoid becoming an obstacle within the design session, and it provided a scrollable list of up to 20 voice responses in the hope of enabling it to support multiple design scripts. This highlights a trade-off between the paper and digital approaches, in the paper approach the group could simply set aside their sheet and begin a fresh one, whereas with the app some way to save/load the scripts or manage multiple scripts within the app would have been required to achieve similar results. This also highlights a draw-back in using intentionally simple design tools, in this case giving each group multiple tablet devices running the app may have resulted in them prototyping more of their designs at a higher fidelity.

It was expected that the voice synthesis in the app would make it less embarrassing for the participants to act out their voice designs. While empirical data on this was not gathered, the facilitators noticed that during the paper scripting design activity the room was very quiet, with participants speaking very quietly and infrequently, while in the second part of the session the room was much noisier with participants typically having no hesitation in talking to their tablet devices. The app seemed to be acting as both a prompt and an enabler for the participants to overcome their self-consciousness and speak out loud within the session to test their designs.

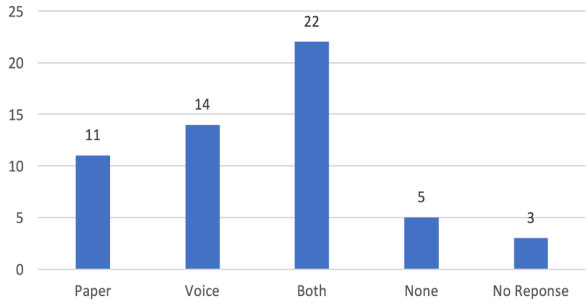

Figure 3. Preference for Expressing Design Ideas

To conclude, the dual prototyping approach appeared to work synergistically and this was recognised by many of the participants. In order to explore the participant preference for the two fidelities the question *Which design method enabled you to easily express your ideas?* was included. Results can be seen in Figure 3, (22, 40%) felt that both paper and 'voice' (the speaking Android app) were needed to express their ideas (the category with the highest number of responses). The difference between preferences for paper and voice separately was negligible.

The initial task of inputting a previous paper design scripts into the app worked well in keep the participants 'on task' as

they were keen to appropriate the app for more playful purposes when the opportunity arose, such as telling jokes, speaking song lyrics, and insulting peers (this was seen extensively in the app log files after the study was complete and the groups were allowed to play with the tablets until the school bell rang). The selection of ideas to prototype at a higher fidelity (from multiple ideas that had been generated previously) provides a useful opportunity for convergence in the design process where specific criteria could have been used, or decision-making processes explored, in the selection of a design ideas to take forwards.

While the designs on paper were unambiguous to analyse (aside from handwriting issues) the app logs proved slightly more challenging. The app logged every change to text on the interface and every piece of text spoken (synthesized) along with timestamps. The main aim of analysis of the logs was to determine the text spoken by the 'house' but the teens in the groups made large numbers of minor edits to the text such as correcting spelling mistakes, adding punctuation, minor re-phrasing etc. both before and after using the app to speak the text, during the entire of the second design session. This made it difficult to determine the 'final' version of the text that should be analysed, and required automated and manual processing of the logs to identify text input that that was left unchanged (i.e. the final version). It is not clear whether this continual editing and refinement was inadvertent or an intentional form of procrastination, but it was not seen in the paper designs and potentially explains why mostly just single designs were explored with the app.

Scaffolding in the Design Task

The scaffolding questions were carefully chosen to represent scenarios that may be meaningful to teenagers in their homes and so elements of these were expected to be evident in the designs produced, the coding carried out in relation to the lineage of design ideas was intended to provide specific insights into the extent of this. The coding showed that 11 (26%) had no identifiable relationship with the scaffolding questions and 12% (5 out of 42) were judged to be facsimiles of an initial question, implying that while this approach (of providing examples) should be used with care its impact when used with this age group (14-15 years) is limited. Within the five designs judged to be facsimiles, two were of "what can I eat?" and three of "where's mum/dad?". In these facsimiles, the example questions would be found in the scripts verbatim or with only minor alterations. The designs with no identifiable relationship with the scaffolding questions included the opening questions:

1. "Can you turn the lights off downstairs at 10:30pm?"
2. "What classes do I have on Wednesday?"
3. "Can you find out why my bills are so expensive?"
4. "Who's Sexy?"
5. "Up for a chat?"
6. "I'm locked out?"
7. "Where's the draft coming from?"
8. "Please will you tell my parents I've gone shopping?"
9. "What am I doing on Friday?"
10. "Will you go to PS Store and buy Minecraft?"
11. "What electronic devices are turned on in the house?"

The majority of designs (26, 62%) were evolutions of the six initial question ideas, with 11 being evolutions of "What can I eat?", eight being evolutions of "Where are my swimming goggles?", five being evolutions of "Anything good on Netflix?", one being an evolution of "Is anyone out tonight?", and one being related to both "What can I eat?" and "Where are my swimming goggles?" (which were counted separately from the two individual categories). These were coded as 'evolutions' as they had similarities with the initial ideas but also key differences, for example the question 'What shall I do tonight?' was unanimously coded as being an evolution of 'Anything good on Netflix?' as both are seeking a recommendation for entertainment from the house. These findings indicate that these initial questions may have been providing a scaffold for participants to think about design opportunities or simply that the questions covered a wide range of features that the participants considered useful.

If we consider the 10 themes which emerged from the designs (from Table 2) only four of these are evident in the six scaffolding questions: "What can I eat?" (Food), "What has he been up to all day?" (none), "Anything good on Netflix?" (Entertainment), "Where's mum/dad?" (Finding), "Is anyone out tonight?" (Calendar/events), "Where are my swimming goggles?" (Finding). Again, this implies that the impact of the scaffolding in terms of biasing designs was minimal. Overall, none of the designs were related to "What has he been up to all day?" (siblings), only one was related to "Is anyone out tonight?" (socialising), three were related to "Where's mum/dad?" (parents), eight being related to "Where are my swimming goggles?" (personal belongings), 13 being related to "What can I eat?" (food), and one being related to both these latter two categories (11 having no relationship). The designs are likely to have been influenced by the developmental stage and priorities of the participants, in addition to dynamics within home environments.

DISCUSSION AND FUTURE WORK

The problem of finding lost objects, the most common theme in the designs, is not constrained to teenagers and has been a common topic of exploration with the domain of Ubicomp and related fields (e.g. [15]). However, memory is one of the many cognitive systems which develops during adolescence [24] and this may contribute to the propensity of children and teenagers for losing things. The second most popular theme of food emerging in the designs aligns with clinical research showing food intake increases during puberty, particularly in the case of boys [21]. In this study, we did not examine any gender differences in the design, but this is a potential area for future research. Other studies have shown that, for the age group participating in the design study, EAH (Eating in the Absence of Hunger) in response to palatable foods can be a particular problem [22]. While some of the designs

captured searches for 'palatable' snack food, it is encouraging that others attempted to dissuade users from eating chocolate. Adolescence is a time when there is a particular risk of developing eating disorders but from a dialog about available food with a house it may be possible to infer the onset of destructive behaviour such as loss of control (LOC) eating [26]. The servile role expected of the smart home and evident in the design scripts has been found in research exploring children's social relationships with robots, much work to this end has been carried in this area by Kahn and colleagues [8], where a key concern is the potentially negative impact compliant robots may have on moral and social development. There is a tension between a manufacturer wishing to provide high levels of positive user experience in their VUI/smart products and its developmental impact on younger users. However, it is encouraging that some designs assumed parent-like roles indicating that, for some participants at least, more positive roles were not unacceptable.

A fascinating area for future exploration is the interlinked issues of privacy and personalisation of services though sensed data collected about the teenagers, which also extends as far as tracking and surveillance. While the participants wanted personalised services, no designs utilised the potentially more invasive data that would have been gathered about themselves related to their locations and activities. Past work within the IDC community has found that younger children responded negatively when a robot used knowledge of their actions not explicitly given [10] and a similar study with adolescent users, exploring this topic and the trade-offs involved in enabling personalised services would likely yield interesting findings.

The scaffolding within the design activity proved valuable but can also be seen a limitation of this work; if other examples had been used, or even no examples at all, would different designs have been produced? Without comparative studies this is a challenging area to understand, and is a trade-off for the designer to consider. Without scaffolding the participants in this work may have relied upon their prior experiences of using VUIs to inspire their design ideas, which in this case would have primarily been Apple's Siri on a mobile device. Design methods intended to convey requirements without the risk of biasing designs, such as Obstructed Theatre [18], could potentially be adapted in the design of VUIs to overcome this limitation of scaffolding.

A novel area for future work is that of evaluating VUIs in the context of younger users, both design prototypes and finished products, as they become more prevalent in homes. Past work such as Home's work on SASSI (Subjective Assessment of Speech System Interfaces) [7] provides a strong foundation within the context of adult users which could be adapted and explored within the context of younger users and contemporary technologies.

The more general insights arising from this work we feel other designers conducting similar work should consider are:

- The low fidelity design activities were valuable in avoiding the recognition problems typically found with current VUIs.
- The 'speaking' app was helpful to enable teenagers to overcome the embarrassment of talking out loud in the activities and explore their designs at a higher fidelity.
- The use of dual fidelities with teenagers may produce few additional design insights but has the potential to enhance engagement in the design activity.
- Scaffolding prompts that resonate with the participants and appear relevant to the activity can be a powerful tool to help participants in a design activity rapidly understand an unfamiliar and futuristic design context.
- When examining participants' design ideas is important to consider the scaffolding prompts that were used.
- The popular themes of food, finding and controlling are seemingly important to teenagers and therefore valuable to consider when wish to engage this user group in design activities.

CONCLUSION

In response to the question of 'How can we co-design voice interfaces with teenagers?' this paper explored the design of VUIs for smart homes with 55 participants aged 14 - 15. The participants worked in groups and generated 42 unique designs using a dual fidelity process which initially required scripting of a voice exchange on paper followed by higher-fidelity prototyping using an Android app with speech output, and a post-study questionnaire (where further design ideas were captured). The design method used proved successful in that a range of useful ideas were collected and no issues were encountered during the session. While the higher-fidelity prototyping activity proved valuable it did not capture any additional design ideas. The use of scaffolding, particularly example uses of a VUIs, within the design activity ensured understanding of key concepts (VUIs and smart homes) and an exploration of the impact of the scaffolding on the designs revealed that only a small number of designs were facsimiles of the examples used. Analysis of the designs created was used to explore the second research question of 'What ideas and expectations do teenagers have in relation to voice interfaces in a smart home?'. Coding the purpose of the designs produced ten distinct categories, the most commonly occurring being related to 'Finding' (objects and parents) and the second most commonly occurring being related to 'Food' (often acquiring palatable snacks and meals). From considering the expectations implied in the designs four key themes emerged: The Role of the House, Tracking, Reliable Recommendations, and Technical Perfection. The design method used prompting questions to help scaffold understanding of the design task and this must be borne in mind when considering the designs produced and associated expectations. This work contributes to the underexplored space of co-designing voice interfaces with young people and we hope the insights provided here will be valuable to others working in similar areas and with VUIs more generally.

SELECTION AND PARTICIPATION OF CHILDREN

In this work children were selected for participation in the study by their school teachers, the study was carried out during a normal school day. The school was provided with information and consent sheets which the school distributed and collected prior to the study. The children were given an opportunity to withdraw their consent at the end of the study, by ticking a check box at the end of the questionnaire. The research team also used a School Participation Agreement outlining their commitment to the school (to adhere to Data Protection legislation, to follow all instructions given by school staff etc.), the expectations of the school (to have appropriate Risk Assessments in place, to make the team aware of school policies etc), and making clear issues such as Intellectual Property ownership.

REFERENCES

1. Matthew P. Aylett, Per Ola Kristensson, Steve Whittaker, and Yolanda Vazquez-Alvarez. 2014. None of a CHInd: relationship counselling for HCI and speech technology. In *CHI '14 Extended Abstracts on Human Factors in Computing Systems* (CHI EA '14). ACM, New York, NY, USA, 749-760.
2. Justine Cassell, Tim Bickmore, Lee Campbell, Hannes Vilhjálmsson, and Hao Yan. 2001. Human conversation as a system framework: designing embodied conversational agents. *Embodied conversational agents.* MIT Press, Cambridge, MA, USA, 29-63.
3. Jan Derboven, Jonathan Huyghe, and Dirk De Grooff. 2014. Designing voice interaction for people with physical and speech impairments. In *Proceedings of the 8th Nordic Conference on Human-Computer Interaction: Fun, Fast, Foundational* (NordiCHI '14). ACM, New York, NY, USA, 217-226.
4. Stefania Druga, Randi Williams, Cynthia Breazeal, and Mitchel Resnick. 2017. "Hey Google is it OK if I eat you?": Initial Explorations in Child-Agent Interaction. In *Proceedings of the 2017 Conference on Interaction Design and Children* (IDC '17). ACM, New York, NY, USA, 595-600.
5. Allison Druin. 2002. The Role of Children in the Design of New Technology. *Behaviour and Information Technology*, 21, 1: 1-25.
6. Daniel Fitton and Beth Bell. 2014. September. Working with teenagers within HCI research: understanding teen-computer interaction. In *Proceedings of the 28th International BCS Human Computer Interaction Conference on HCI 2014-Sand, Sea and Sky-Holiday HCI*, BCS, 201-206.
7. Kate S. Hone and Robert Graham. 2000. Towards a tool for the Subjective Assessment of Speech System Interfaces (SASSI). *Nat. Lang. Eng.* 6, 3-4 (September 2000), 287-303.
8. Peter H. Kahn, Heather E. Gary, and Solace Shen. 2013. Children's Social Relationships With Current and Near-Future Robots. *Child Dev Perspect,* 7: 32–37.
9. Lorenz C. Klopfenstein, Saverio Delpriori, Silvia Malatini, and Alessandro Bogliolo. 2017. The Rise of Bots: A Survey of Conversational Interfaces, Patterns, and Paradigms. In *Proceedings of the 2017 Conference on Designing Interactive Systems* (DIS '17). ACM, New York, NY, USA, 555-565.
10. Iolanda Leite and Jill Fain Lehman. 2016. The Robot Who Knew Too Much: Toward Understanding the Privacy/Personalization Trade-Off in Child-Robot Conversation. In *Proceedings of the The 15th International Conference on Interaction Design and Children* (IDC '16). ACM, New York, NY, USA, 379-387.
11. Youn-Kyung Lim, Erik Stolterman, and Josh Tenenberg. 2008. The anatomy of prototypes: Prototypes as filters, prototypes as manifestations of design ideas. *ACM Trans. Comput.-Hum. Interact.* 15, 2, Article 7 (July 2008).
12. Linchuan Liu and Peter Khooshabeh. 2003. Paper or interactive?: a study of prototyping techniques for ubiquitous computing environments. In *CHI '03 Extended Abstracts on Human Factors in Computing Systems* (CHI EA '03). ACM, New York, NY, USA, 1030-1031.
13. Silvia Lovato and Anne Marie Piper. 2015. "Siri, is this you?": Understanding young children's interactions with voice input systems. In *Proceedings of the 14th International Conference on Interaction Design and Children* (IDC '15). ACM, New York, NY, USA, 335-338.
14. Ewa Luger and Abigail Sellen. 2016. "Like Having a Really Bad PA": The Gulf between User Expectation and Experience of Conversational Agents. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (CHI '16). ACM, New York, NY, USA, 5286-5297.
15. Jens Nickels, Pascal Knierim, Bastian Könings, Florian Schaub, Björn Wiedersheim, Steffen Musiol, and Michael Weber. 2013. Find my stuff: supporting physical objects search with relative positioning. In *Proceedings of the 2013 ACM international joint conference on Pervasive and ubiquitous computing* (UbiComp '13). ACM, New York, NY, USA, 325-334.
16. Martin Porcheron, Joel E. Fischer, and Sarah Sharples. 2017. "Do Animals Have Accents?": Talking with Agents in Multi-Party Conversation. In *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing* (CSCW '17). ACM, New York, NY, USA, 207-219.
17. Amanda Purington, Jessie G. Taft, Shruti Sannon, Natalya N. Bazarova, and Samuel Hardman Taylor.

2017. "Alexa is my new BFF": Social Roles, User Satisfaction, and Personification of the Amazon Echo. In *Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems* (CHI EA '17). ACM, New York, NY, USA, 2853-2859.

18. Janet C. Read, Daniel Fitton, and Emanuela Mazzone. 2010. Using obstructed theatre with child designers to convey requirements. In *CHI '10 Extended Abstracts on Human Factors in Computing Systems* (CHI EA '10). ACM, New York, NY, USA, 4063-4068.
19. Mary Beth Rosson, John M. Carrol, and Rachel K. E. Bellamy. 1990. Smalltalk scaffolding: a case study of minimalist instruction. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (CHI '90), Jane Carrasco Chew and John Whiteside (Eds.). ACM, New York, NY, USA, 423-430.
20. Alexander I. Rudnicky. 1989. The design of voice-driven interfaces. In *Proceedings of the workshop on Speech and Natural Language* (HLT '89). Association for Computational Linguistics, Stroudsburg, PA, USA, 120-124.
21. Lauren B Shomaker, Marian Tanofsky-Kraff, David M Savastano, Merel Kozlosky, Kelli M Columbo, Laura E Wolkoff, Jaclyn M Zocca, Sheila M Brady, Susan Z Yanovski, Melissa K Crocker, Asem Ali, and Jack A Yanovski. 2010. Puberty and observed energy intake: boy, can they eat! *Am J Clin Nutr,* July 2010, Vol. 92 No. 1, 123-129.
22. Lauren B Shomaker, Marian Tanofsky-Kraff, Jaclyn M Zocca, Amber Courville, Merel Kozlosky, Kelli M Columbo, Laura E Wolkoff, Sheila M Brady, Melissa K Crocker, Asem H Ali, Susan Z Yanovski, and Jack A Yanovski. 1992. Eating in the absence of hunger in adolescents: intake after a large-array meal compared with that after a standardized meal, *Am J Clin Nutr*. 2010 Oct; 92(4): 697–703.
23. Iris Soute, Susanne Lagerström, and Panos Markopoulos. 2013. Rapid prototyping of outdoor games for children in an iterative design process. In *Proceedings of the 12th International Conference on Interaction Design and Children* (IDC '13). ACM, New York, NY, USA, 74-83.
24. Laurence Steinberg, Cognitive and affective development in adolescence, *Trends in Cognitive Sciences*, Volume 9, Issue 2, 2005, Pages 69-74, ISSN 1364-6613.
25. Alan M. Turing. 1950. Computing machinery and intelligence. *Mind*, 59, 236: 433-460.
26. Anna Vannucci, Marian Tanofsky-Kraff, Lisa M. Ranzenhofer, Nichole R. Kelly, Louise M. Hannallah, Katherine Pickworth, Mariya V. Grygorenko, Sheila M. Brady, Tania A. Condarco, Merel Kozlosky, Andrew P. Demidowich, Susan Z. Yanovski, Lauren B. Shomaker, and Jack A. Yanovski. 2014. Puberty and the manifestations of loss of control eating in children and adolescents. *Int. J. Eat. Disord*., 47: 738–747.
27. Linda Wulf, Markus Garschall, Julia Himmelsbach, and Manfred Tscheligi. 2014. Hands free - care free: elderly people taking advantage of speech-only interaction. In *Proceedings of the 8th Nordic Conference on Human-Computer Interaction: Fun, Fast, Foundational* (NordiCHI '14). ACM, New York, NY, USA, 203-206.