Speech and Gestures for Smart-Home Control and Interaction for Older Adults

Anbarasan*
University of Glasgow
Singapore
2287546A@student.gla.ac.uk

Jeannie S.A. Lee
Singapore Institute of Technology
Singapore
jeannie.lee@singaporetech.edu.sg

ABSTRACT

Older adults have been encountering difficulties in using modern technological devices to control home appliances as they are lacking in technology literacy and mobility. This led to the usage of remote controllers or requiring assistance from family members, which is not beneficial for older adults since there is less independence. To alleviate this problem, this project aims to develop a prototype system named "Genie" which caters for older adults ranging from 65 to 80 years old, allowing for easy control of smart home appliances through combination of speech and gesture interactions. An experiment was carried out with a total of 20 older adults on the prototype system where the initial results demonstrate a significant increase in usability. Based on the evaluation, such interaction methods show promise to be effective in replacing manual operations of home appliances through the use of simple speech or gesture commands.

KEYWORDS

human interactions; speech; gestures; smart-home; older adults

ACM Reference Format:
Anbarasan and Jeannie S.A. Lee. 2018. Speech and Gestures for Smart-Home Control and Interaction for Older Adults. In *3rd International Workshop on Multimedia for Personal Health and Health Care (HealthMedia'18), October 22, 2018, Seoul, Republic of Korea.* ACM, New York, NY, USA, 9 pages. https://doi.org/10.1145/3264996.3265002

1 INTRODUCTION

The world has been experiencing a rapidly aging population [34]. Data from the World Population Prospects of 2017 Revision has shown that the number of older adults aged 60 years and above is expected to double by 2050, which is a massive rise from 962 million globally in 2017 to 2.1 billion in 2050 [33].

Singapore, being one of the fastest ageing populations in the world [10], is expected to triple to 900,000 by 2030 [14][35]. These older adults are mostly either less literate, do not possess smartphones, have little or zero exposure to technology or face concerns with privacy related issues [31]. According to Kachar [23], "Computers and information technologies offer the potential to improve the quality of life of the elderly, providing them information and useful services to their daily life". In contrast, the next generation of the older adults population is expected to be more knowledgeable, financially stable, and receptive towards technology [31]. As such, there is a potential for technology solutions for ageing-in-place to cater to different older adults' profiles and demographics. Companies such as Amazon and Google have invented technologies such as smart speaker assistants like the Amazon Echo and Google Home as assisted devices in intelligent home automation. These devices enable users to interact with services, both in-house and third-party through integrated voice commands, allowing users to control smart home appliances via Smart Hubs such as the Logitech Harmony Hub [15].

However, these systems mainly target a younger age group, and which may be less suitable for older adults due to language limitations since modern technology such as assisted smart-home devices do not influence a large percentage of the older adults in Singapore [36][12]. This limitation results in older adults resorting to the use of switches, remote controllers or the help of their family members or domestic helpers [31] to control home appliances.

The main aim and objective is to develop a prototype speech and gesture interactive system to assist older adults aged from 65 to 80 years old in Singapore to control home appliances such as lights and fan from a seated location using simple speech commands and gestures. Some objectives include reducing the complex language requirements for issuing a command, and overcoming mobility limitations and the need to reach out for the appliance to control it.

2 RELATED WORK

2.1 Smart Speaker Assistants

A smart speaker assistant is a speaker with voice command device incorporated with an integrated virtual assistant that offers interactive voice actions and hands-free activation. Commercial examples of these are the Amazon Echo [1][4] and Google Home [18][3]. This is done through a verbal "wake word", usually the system name. It is also able to perform audio playback and control home automation systems. Some limitations include compatibility across some services and platforms, and peer-to-peer connection through mesh networking and virtual assistants. Each system has its designated interfaces and features in-house, usually launched or controlled via applications or home automation software [41]. Although exhibiting a conversational speech interface, disadvantages are the length of speech commands, depending on the context and use case. Studies have found age-related declines in short-term memory [19] [40], and older adults may encounter difficulties in

[*] Anbarasan has no last name.

HealthMedia'18, October 22, 2018, Seoul, Republic of Korea
© 2018 Copyright held by the owner/author(s). Publication rights licensed to ACM.
ACM ISBN 978-1-4503-5982-5/18/10...$15.00
https://doi.org/10.1145/3264996.3265002

the memorization of lengthier commands, as demonstrated in the subsequent sections.

2.2 Existing Work

2.2.1 GeeAir: a universal multimodal remote-control device for home appliances. GeeAir is a handheld device designed to control home appliance via a mixed modality of speech, gesture, joystick, button and light naturally [22]. It takes the user input to select a target appliance and then recognises the predefined hand gesture of users to control the device [22]. A three-axis built-in accelerometer is used to capture user's 3-D hand gesture signals for gesture recognition component and a built-microphone for acquiring the user's speech commands for speech recognition [22]. A speaker is built into the system for the user to receive voice feedback.

The main drawback of GeeAir is that it only acts as a remote, which is made using the Nintendo Nunchuk. This requires the user to hold the controller with their hand to control it physically. The remote itself is more complicated for an older person as it has too many input methods such as joystick and buttons [36] [**?**]. If the device is misplaced, there will be no possibility of controlling the system.

2.2.2 Put That There. The system "Put that there" [5], developed by the Massachusetts Institute of Technology Cambridge, focuses on receiving input through gesture and speech for the system to act accordingly. It consists of gesture designs such as pointing to a blank space to draw a shape by a speech command, using a joystick. It was groundbreaking for its time, and due to the latest technology advancements, there is the potential to improve gestures for more representative and direct interaction. This has also resulted in an inspiration for the development of the prototype system.

2.2.3 Smart Home ML: Towards a Domain-Specific Modeling Language for Creating Smart Home Applications. Through the development of Smart Home ML [2] Reykjavik, University of Iceland students have combined the application Amazon Echo with Samsung SmartThings [44]. It uses a domain-specific modelling language for smart home applications which allows users to define new custom skills such as "Off my air conditioner when my room temperature is at 18 degrees". The primary objective of the system is to combine the modelling language to allow both third-party hubs to work with the Amazon Echo. This is done through using custom skills which can be added to a service using this modelling language.

3 REQUIREMENTS GATHERING

Data was gathered from a sample of 25 older adults consisting of both males and females of different ethnicities representative of Singapore, equally divided into two different age groups - 65 to 75 years old and 76 years and above. Each individual took part in a verbal interview and a requirements gathering experiment. From the results gathered, an interaction model was designed.

3.1 Requirements Gathering: Interview

The interview was conducted verbally, and a translator was used to communicate with older adults whose first language is not English. They were asked questions on the difficulties they encountered at home and experiences with smart assistant technology.

In the age group category of 65 to 75 years old, 8 were high school level qualified, with 4 being pre-university graduates and 1 holding a university degree. In the age group category of 76 years and above, 5 were primary level qualified and 6 were high school level qualified. The results showed that most older adults encountered difficulties with English due to a lower qualification attained.

The majority of inverviewees encountered difficulties when staying alone at home. Examples included moving around the home, finding the remote control or frequently walking to the kitchen just to operate the kettle. They relied on the assistance of their family members due to reasons such as restrictions in movement or forgetfulness. It was also observed that most of the interviewees opted for the idea of a smart system to aid them in activities such as operating fans and lights through the use of basic English words.

3.2 Requirements Gathering: Experiment

The experiment was conducted with the same participants, namely older adults who are inexperienced in using smart assistants such as the Amazon Echo and Google Home.

During the experiment, measurements were recorded to obtain the number of times a wrong command was issued, the number of times the system incorrectly recognised a command, and the time taken by the participant to memorise the command. Two commands were performed five times, namely: "Alexa, What time is it ?" and "Alexa, what's the weather in Singapore?".

Based on the experiment results, the first command, "Alexa, what time is it?" was spoken incorrectly for 20 times out of 120 times which revealed a user error rate of 17% and the second command, "Alexa, what's the weather in Singapore?", was spoken incorrectly 40 times out of 120 times which revealed a user error rate of 33%. This showed that most of them faced problems with the second command. Some of the participants also mentioned facing difficulties with pronouncing the word "what's the" and "Alexa" as it has 3 syllables [40].

It was also seen that system failed to recognise the commands occasionally, especially on command two which had an error rate of 27 out of 120 times (22.5%) whereas command one only had an error rate of 9 out 120 times (7.5%).

From the results, it can be seen that the age group category for 65 to 75 years old tend to have better memory skills as compared to 76 years old and above based on the time taken to memorise a command. In age group category for 65 to 75 years old, it took an average time of 7 minutes (range of 6 mins to 9 mins) whereas age group category 76 years old and above took an average time of 12 minutes (range of 11 mins to 13 mins). From the results, it can be inferred that education level and age does affect the time taken to memorise.

In conclusion, older adults encountered difficulty with the second command where they were unable to pronounce "what's the" as compared to the first command which was easier. At the end of the experiment, participants were tasked to answer questions regarding user experience for feedback gathering. From the feedback gathered, most of them could not pronounce command two and the word "Alexa" and suggested using an easier "wake word" will attract them to use the system due to its simplicity.

4 DESIGN

4.1 Hardware Selection

4.1.1 Speech and Gesture Recognition Hardware. The evolution of technology has been vast and rapid over the last decade, creating affordable multi-modal input devices that are small and yet well designed with features that have the hardware capability to support Natural User Interface (NUI) with speech and gesture recognition.

Comparisons between Microsoft Kinect V2 [46], Leap Motion [25] and MYO [24] were done to select the most suitable hardware for the system. Some conditions taken into consideration while selecting the hardware include full body traction, speech recognition, the range of the sensor, and library services provided with the SDK. An important consideration was that the hardware should not be a wearable device which could obstruct the ease of convenience for the older adults [11].

The Microsoft Kinect V2 was chosen as the speech and gesture recognition hardware as it supports full-body tracking with 25 body joints per person and can track a total of 6 people [16]. It is also equipped with a built-in multi-array microphone consisting of four microphones that was designed for beamforming [30], a technique used to amplify sound from one direction and suppress any sound from other directions. The important feature supports gesture of body movements and hand states such as open, close and lasso. Another factor is the longer range of depth ranging from 0.5m to 4.5m compared to the Leap Motion. Kinect V2 also uses an infrared to work under dark area conditions. Eventually, the Kinect V2 was selected as the better choice as compared to the Leap Motion which is only able to trace hands, and the MYO Armband that required the user to wear the device which was inconvenient.

4.1.2 Smart Home Hub. A smart home hub is a device that connects all the smart appliances (e.g. light bulbs, wall outlets) on a home automation network and controls communications among them. Companies which have developed smart home hubs include Samsung [44], Philips [39] and Logitech [28].

During the process of selecting the smart home hubs to be used in the system, some key factors were considered such as the compatibility with both Zigbee protocol and Z-Ware protocol devices. Zigbee protocol and Z-ware protocol are upcoming standard protocols followed in a smart home hub for wireless communication between two or more smart appliances [37]. The smart home hub should consist of Application programming interface (API) services or developers' SDK to write their own API services.

Based on the process of selecting of the hubs, Philips Hue [39], Samsung SmartThings Hub V2 [44] and Logitech Harmony Hub [28] were chosen as comparison products. Philips Hue and Samsung SmartThings were eventually chosen as the smart hub devices for the system.

The Philips Hue can communicate with smart light appliances and also provides various types of lights such as RGB and dimmable lights. The Philips Hue also comes with RESTful API services [38] for controlling the hub.

Samsung SmartThings Hub V2 has the capability of including custom API services by developers [43] and communicating with different brands of smart appliances such as the Philips Hue lights and TP-Link plugs which are not created by Samsung, providing flexibility across various smart appliances. On the contrary, the Logitech Harmony Hub was not opted due to incompatibility issues with Zigbee and Z-ware protocols [45].

4.2 Language Model: Taxonomy of Speech & Gesture Combination Vocabulary

The language model is crucial in a well-developed speech and gesture recognition system [47]. It can be designed using grammar formats such as Speech Recognition Grammar (SRGS) and Java Speech Grammar Format (JSGF). The role of the language model is to design a well-structured prototype system with a custom vocabulary list for human speech and defined gestures.

JSGF was selected as the optimal choice as it provides a general grammar file easily understood by developers interested in using it to develop vocabularies and gestures listed. JSGF can easily be converted to SRGS and is also usable in well-known speech recognition frameworks such as the CMU Sphinx-4.

It is important to create the grammar by limiting the structure and vocabularies of human speech according to the system requirements to improve the recognition accuracy. During the requirements gathering interview and experiment with older adults using the Amazon Echo Dot, it was observed that most older adults tended to seek help from their family members using their name followed by action words and location name and device such as "David, turn on bedroom light". Based on the requirements gathering phase, the full grammar was created with a vocabulary and gesture combination.

A system name is required to activate the system before commands can be taken in, similar to existing semi-related systems such as Google Home's "Ok Google" command and Amazon Echo's "Alexa" command. From the requirements gathering phase, "Genie" was chosen as the system name as it is easy to pronounce and remember for older adults since it has only 2 syllables whereas "Alexa" has 3 syllables. The definition of "Genie" originated from a magical being bound to obey the commands of a mortal possessing its container which is related to this system's objectives and functions. Table 1 shows a snippet of the grammar for system name.

```
#JSGF V1.0;
grammar genie;
public <genie> = <Action>;
<SystemName> = Genie;
```

Table 1: Snippet of Language Model Consisting of System Name

The grammar has selections for location name and devices name to list common home locations such as living room, bedroom and devices such as lights, fan, and heating, ventilation, and air conditioning (HVAC). Under devices, it was designed such that each device may possess a different name. For instance, light may be named either as "main light" or "table light" which allows specific selection of appliance to be acted upon. This selection is shown in Table 2.

```
<Location_Name> = living Room|kitchen
|bedroom|bedroom two|bedroom three
|dining hall|toilet;

<Device_Name> = <lights>|<fans>|<hvac>
|<entertainment>;

<lights> = main light|table light
|ceiling light;
<fans> = table fan| ceiling fan
| wall fan| fan;
<hvac> = aircon|air conditioner|ac;
<entertainment> = tv|television|dvd player;
```

Table 2: Snippet of Location and Device Segments in Language Model

Actions such as turning on, off and dimming of function controllers are required to be stated in a well-structured manner for successful speech recognition. The structure is defined using a system name followed by command and location name and device name. An example is "Genie turn on the living room main light". The reason for this structure was due to the feedback from older adults where 90% of them followed the stated sequence. This selection is shown in Table 3.

```
<turnOn>  = <SystemName>
(turn on the <Location_Name> <Device_Name>)
| <SystemName> (<Location_Name> <Device_Name> turn on)
| <SystemName> (<Location_Name> <Device_Name> on);

<turnOff> = <SystemName>
(turn off the <Location_Name> <Device_Name>)
| <SystemName> (<Location_Name> <Device_Name> turn off)
| <SystemName> (<Location_Name> <Device_Name> off);

<dim> = <SystemName>
(dim the <Location_Name> <lights> <gesture>);
<brighten> = <SystemName>
(dim the <Location_Name> <lights> <gesture>);

<changecolor> = <SystemName>
(change <Location_Name> <lights> color <gesture>
|change <Location_Name> <lights> colour <gesture>);
```

Table 3: Snippet of Action Commands Segment in Language Model

There is a limitation in voice command due to mispronunciation and voice recognition of the older adults. Therefore, some of the action commands require gesture movements and the main objective is to make gesturing simple for older adults to ensure ease of comfort [17]. The gesture model includes daily human interactions such as performing hand movements in up, down, left and right directions where it can recognize a range between neck to hip area as older adults may be less flexible and will opt to use minimal effort for optimal results [9]. Some of the gestures were inspired by existing gestures designs in touch based interaction devices [11]. The gesture segment in the language model and gesture movements are shown in Table 4 and Figure 1 respectively.

```
<gesture> =
Hand_Swipe_Right_To_Left|Hand_Swipe_Left_To_Right
|Hand_Swipe_Up_To_Down|Hand_Swipe_Down_To_Up;
```

Table 4: Snippet of Gesture Segment in Language Model

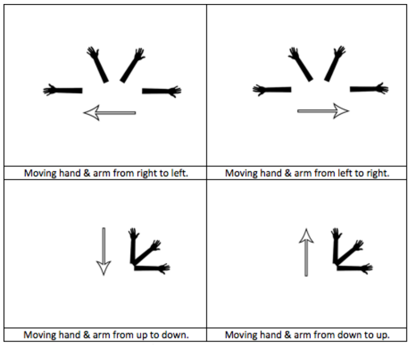

Moving hand & arm from right to left.
Moving hand & arm from left to right.
Moving hand & arm from up to down.
Moving hand & arm from down to up.

Figure 1: Gesture Movements

Some examples of the output based on the grammar is shown in Table 5.

Turn On:	1. Genie turn on the living Room light 2. Genie turn on the kitchen ceiling light 3. Genie bedroom table fan on
Turn Off:	1. Genie turn off the living Room light 2. Genie turn off the kitchen ceiling light 3. Genie bedroom table fan off
Dim:	1. Genie dim the dining hall main light Hand_Swipe_Up_To_Down
Brighten:	1. Genie brighten the dining hall light Hand_Swipe_Down_To_Up
Change Light Color:	1. Genie change living Room ceiling light color Hand_Swipe_Left_To_Right 2. Genie change kitchen main light color Hand_Swipe_Right_To_Left

Table 5: Grammar Output Examples

To assist programs and humans in better understanding the specification of conceptualizations. An ontology model such as Web Ontology Language (OWL DL diagram) was designed for expressing the system's home appliances and home locations [21] and this is shown in Figure 2.

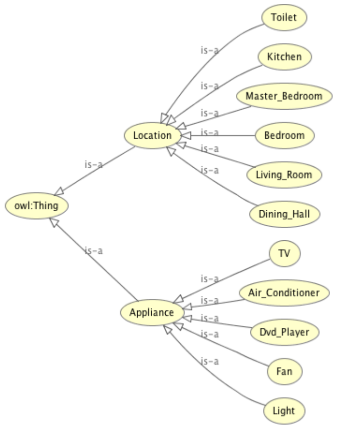


Figure 2: Gesture Movements

4.3 System Architecture

The system consists of a main component which is the Kinect-Enabled Windows Presentation Foundation Application (Genie Prototype Application). It interacts with two other components which are Smart-Hubs (Philips Hue Bridge V2 & Samsung SmartThings Hub V2) and the input device for speech and gesture recognitions (Kinect Sensor V2). A high-level system architecture diagram is shown in Figure 3.

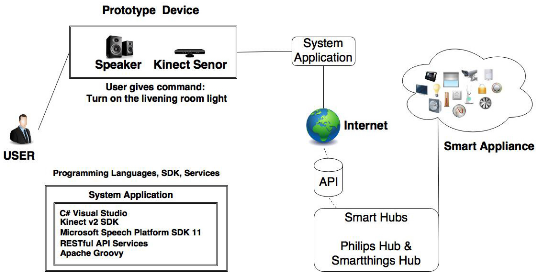


Figure 3: System Architecture

The system's purpose is to recognize speech and gestures from the user to control smart home appliances. It is built using C# and Kinect SDK with Microsoft Speech API and gesture algorithm. A set of vocabulary list is required to perform speech and gesture commands. The system application is designed to return voice commands as responses to users as feedback. It also requires the input device, Kinect, to be connected. The system application will communicate with smart hubs such as Samsung SmartThings Hub and Philips to send commands to the smart appliances. Some example of these appliances includes smart plugs and smart lights. The communication with the hubs is designed using RESTful API and JSON which return data formats to the system application.

5 IMPLEMENTATION

The system is a WPF application consisting of a combination of three systems working together. It runs on a Windows platform where the Kinect Sensor is connected to the computer. The SmartHubs are connected to the home network through a router. The system application connects to the SmartHubs via Internet and LAN connection through the API services for issuing commands. The smart appliances are connected to wall power outlets and wirelessly to the smart hubs. The SmartThings SmartAPP is created to allow SmartThings to provide API services.

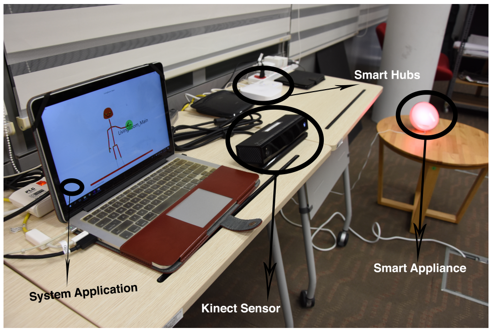


Figure 4: Systen Setup

5.0.1 *Body Frame Capture.*

The user's body and movements are captured using Kinect Sensor Recognition and a skeleton is displayed to mirror the movements in real-time. The gesture recognition was implemented using the Kinect SDK library.

5.0.2 *Building of Grammar.*

By adding a grammar builder, the speech and gesture combination commands can be added to the vocabulary list. The creation of grammar is done by building the grammar structure from a list of sentences with the aid of "Microsoft.Speech.Recognition" library (Microsoft Speech Platform SDK 11) which allows dynamic grammar creation.

5.0.3 *Retrieving Information From SmartHubs: Philips Hue and SmartThings Hub.*

Information of the smart hubs and smart appliances connected to the smart hubs (Philips Hue and Samsung SmartThings Hub V2) is retrieved using GetAsync Function. It will receive a JSON formatted data from the Smart Hub via an API service which will be converted to JSON object. Extraction of information such as location name of the device and appliance names are performed using the JSON object which will be stored into a Device class along with dynamic command sentence for each action such as voiceOn: Location_name + lights on.

5.0.4 *Speech Recognition.*

The implementation of the speech recognition was done using Microsoft library "Microsoft.Speech.-Recognition". The speech is captured from the Kinect V2 hardware and a recognition method is used to check if the phrase starts with the wake word "Genie", followed by any grammar commands such as "living room light on". The command is then acted upon and the light will be switched on automatically. However, commands which have more variations as opposed to "on" and "off", such as "dim living room light", will require a gesture action.

5.0.5 *Gesture Recognition.*

The gesture recognition uses Kinect V2 SDK library which has built-in gesture recognition algorithms for Kinect V2 Sensor to track the body joints (X,Y,Z) coordination. Using the body joints, algorithms and logic, it allows performing of gestures such as swiping up, down, right to left, and left to right. Some gestures were designed for either right or left hand.

5.0.6 *Sending Action Commands To Smart Hubs .*

The system application will send an action command to the smart hubs to control the smart appliance with the use of HttpClient which will send request with PutAsync() to the API request address.

5.0.7 *Samsung SmartThings SmartApp API.*

To retrieve information of the smart appliance connected to the SmartThings Hub, an Apache Groovy Script was created on the web-based IDE for developing SmartThing solutions [42]. The script creates a SmartApp named "Genie" which can be installed on the SmartThings Mobile Application. This allows selection of smart appliance information to be sent to the prototype system via custom API method created on the web-based IDE on the SmartThings Developer Portal.

5.1 Pilot Testing

After the system was developed, a pilot test was conducted with a few test participants to assess the usability and accuracy of the speech and gesture functions.

After the pilot test, changes were made to the two gesture algorithms to limit the range of hand movements to improve the accuracy and detection of the gestures. A test was then conducted on another small set of test participants to re-perform the same commands whereby the gesture accuracy is proved to be outstanding.

6 EVALUATION

The evaluation study objectives were (1) to investigate the usability and accuracy of the prototype system for older adults aged 65 years to 80 years, and (2) to investigate if the prototype system has a higher usability compared to the Google Home and Amazon Echo smart-home speakers in the context of the older adult population in Singapore.

The NASA Task Load Index (NASA-TLX)[32] and System Usability Scale (SUS) [27] were used to measure the cognitive workload and usability, and the system error rate was used to measure the accuracy of the system to benchmark. A within-subjects design was conducted in the experiment, where each participant was exposed to more than one of the conditions using various systems. Statistical methods of analysis such as hypothesis testing (Student's t-test) were used to find the differences.

6.1 Methodology

6.1.1 Participants. The study was conducted on 20 participants that are older adults in the age range of 65 to 80 years old (8 male, 12 female) with different ethnic origin over a period of 2 weeks (14 days). In the experiment, the participants are required to interact with the system, using the specified speech and gesture movements. Participants were recruited by providing an official recruitment and participant information sheet which also includes a consent form for the participant and witness/guardian to receive their signature. The consent form is provided in four different languages: English, Chinese, Malay and Tamil. Participants were given at least 3 days to decide on participation in the study.

6.1.2 Measures. The System Usability Scale (SUS) [27][7] and NASA Task Load Index (NASA-TLX) were used as a subjective assessment tools to measure the system usability and cognitive workload in this study. The SUS and NASA-TLX were carried out at the end of each system condition to identify the usability and workload of each interface.

The SUS is a commonly used scale for evaluating a system's usability, consisting of 10 questions, each with a 5-point scale ranging from "Strongly Disagree" to "Strongly Agree". The results are computed to a 0-100 score range.

The NASA-TLX is a widely used to identify the cognitive workload of a system or task [32][13], consisting of 6 measurements which are Mental Demand, Physical Demand, Temporal Demand, Performance, Effort and Frustration Levels. The measurements are then rated within a 100-point range with 5-point steps where they are combined to compute the workload [20].

In addition, two other measurements were used to determine the system usability for the older adults. The first method was measuring the time taken by the participants to memorise each set of commands in terms of minutes and seconds. The second method was measuring the user performance error rate in terms of number of times an error was committed by the participant.

The results of SUS, NASA-TLX data, user error rate and time taken to memorise was statistically analysed to assess the levels of motivation and feedback when using the prototype system to control the appliances [20]. The results are shown in Section 6.5.

To measure the accuracy of the system, the system error rate was obtained when each participant performed a command and the system failed to recognize it. The result shows the difference between the system's error rates and accuracy.

6.1.3 Procedure. The within-subjects approach was conducted with a total of 20 participants where they had undergone two randomized conditions consisting of the prototype system against Google Home and Amazon Echo. The conditions are namely: Condition A: Issue speech commands to commercial smart speaker devices: The Google Home and the Amazon Echo. Condition B: Issue speech and gesture commands to the prototype system developed.

To mitigate the order effect, counterbalancing was performed, whereby each participant either started with condition A followed by condition B or vice versa depending on what the previous participant started with.

Depending on the assigned condition, the participant had to memorise 3 commands for each system, which was then performed by them to the respective system. A stopwatch was used to time the duration each of them took to memorise each command. Each set of commands shown in Table 6 was performed by the participant for all three systems for 5 times.

A measurement log was used to record the time taken by the participant to memorise each set of commands and was taken down in minutes and seconds. Another measurement log was also used to record the user performance error rate, mainly the number of times the participant has committed an error in the command, the system error rate, as well as the number of the times the system failed to recognise the command. A third measurement log was used to take down the volume of speech level of the participant when performing a command. The measurement was recorded in decibels.

After a condition was performed, the participant was given a 5 minutes break. Simultaneously, they were required to complete two survey forms System Usability Scale and NASA-TLX. After the 5 minutes break, the participant proceeded to memorise another set of commands and used the system in the other condition which is similar to the previous condition.

At the end of the experiment, the participant was asked to fill out the final short survey form which includes participant's satisfaction and usability of the device for both conditions. Some feedback questions were asked upon completion of the short survey. The participant's feedback were written to a section on the form of SUS's participant other comments section.

Command No.	Google Home	Amazon Echo	Prototype System
Command 1	Ok Google, set living-room light to 50% ,	Alexa, dim the living room to 50%	Genie, dim living-room light, swipe up to down or vice versa.
Command 2	Ok Google, turn living-room light to green	Alexa, turn living room to green	Genie, change living room light colour, followed by hand gesture: swipe right hand from right to left or left hand from left to right.
Command 3	Ok Google, turn on the living-room light	Alexa, turn on the living room light	Gesture: point to the light, Genie ON that.

Table 6: Command Table

6.2 Results

Table 6 shows the commands used during the experiment. The time taken to memorise, user and system error rates were analysed using student's t-test to determine the best system.

6.2.1 Time Taken To Memorise The Commands. The mean time taken to memorise the commands by all the participants across the two conditions (Prototype System) and (Google Home & Amazon Echo) are shown in Figure 8. Student's t-test was conducted on the mean time taken. Statistical significant was found for command 1 and command 3 between the prototype system and Google Home (p<0.05) & Amazon Echo (p<0.05). However, no statistical significance was found for command 2 between the prototype system and Google Home (p>0.05) & Amazon Echo (p>0.05). The least mean time taken to memorise was observed in the prototype system and the most mean time taken to memorise with Google Home. From the t-test results, it was observed that the prototype's commands were easier to memorise and this is due to the simple command structure and gesture combinations as compared to the other two system's commands. This proved better usability of commands for older adults and aligns with other research studies [26].

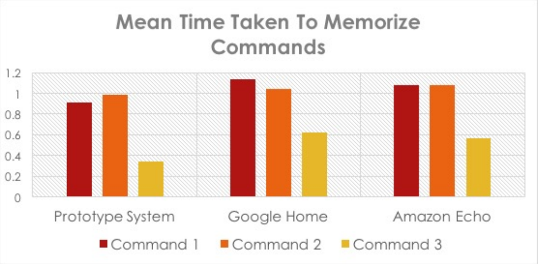


Figure 5: Mean Time Taken To Memorise The Commands

6.2.2 User Error Rate. The mean of the user error rate by all participants across the two conditions is shown in Figure 9 where it is noticeable that the prototype system has a lower user error rate compared to Google Home and Amazon Echo. Similarly, Student's t-test was conducted on the mean user error rate across the systems to determine differences in statistics. Result of the statistical significant was observed for command 3 between the prototype system and Google Home (p<0.05) & Amazon Echo (p<0.05). However, no statistical significant were found for command 1 and command 2 between the prototype system and Google Home (p>0.05) & Amazon Echo (p>0.05). Based on the t-test results, prototype's command 3 had the least error rate. The reason could be due to a more natural way of gesturing to switch on an appliance [29]. In overall, the user error rate is still lower for all the commands of the prototype system compared to other system's user error rate. The accuracy of the prototype system is considered higher than other systems.

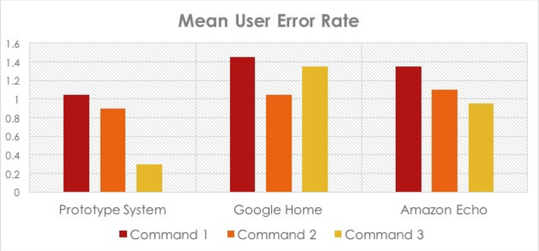


Figure 6: Mean User Error Rate

6.2.3 System Error Rate. Figure 10 shows the mean of the system error rate across all the systems within the two condition. T-test was conducted and statistical significant was observed for all the commands between the prototype system and Google Home (p<0.05). However, significant difference was observed in command 3 between the prototype system and Amazon Echo (p<0.05) only. It is proven through the study that Google Home had an overall high error rate compared to the prototype system as it could be due to command 1 using words such as "50%" whereby the older adults had difficulty pronouncing. Thus, this led to Google Home recognising it wrongly as "30%" [19].

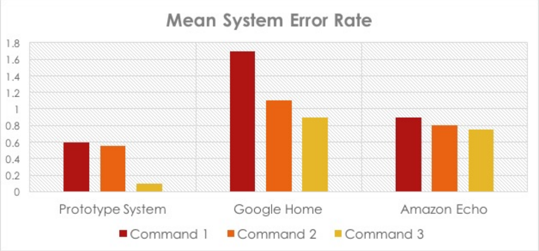


Figure 7: Mean System Error Rate

6.2.4 Speech Level. The results of Mean Speech Level in decibels (dB) is shown in Figure 11 for both prototype system and Google Home & Amazon Echo. There was no statistical significant was found when t-test was conducted on the speech level across all the systems as participant maintained similar speech level. The overall results showed that all across the systems, the speech range from lowest around 60 db to highest about 66 db which was within the range of a normal communication speech [8].

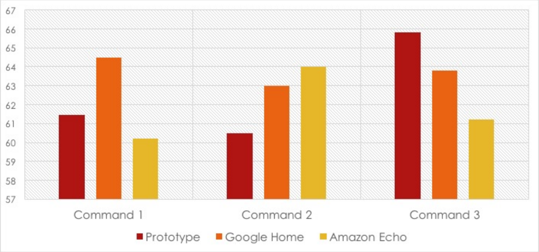


Figure 8: Mean Speech Level

6.2.5 *System Usability Scale Score.* The results of System Usability Scale score (SUS) is shown in Figure 12 for both the prototype system and Google Home & Amazon Echo. The prototype system scored 76.62 out of 100 whereas other systems sccored 57.75 out of 100. Based on the score, it shows that the prototype system has 'Good' usability and the others at 'OK' usability, based on the adjective rating scale shown Figure 6.

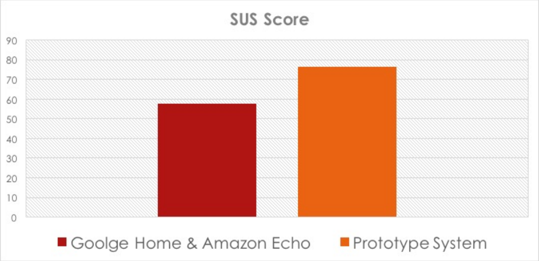


Figure 9: SUS Score

6.2.6 *NASA-TLX Average WorkLoad.* The NASA-TLX Average Workload of the 20 participants for both conditions is shown in Figure 13. Based on the workload, it shows that participants felt more rushed to complete a task with the Google Home and Amazon Echo while the prototype system had a better success rate. In general, the workload of the system was at 41.6% for Google Home and Amazon Echo as compared to the prototype system which was only 25.6%, proving that the system had better usability [20].

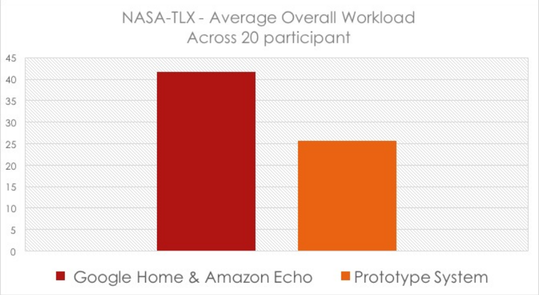
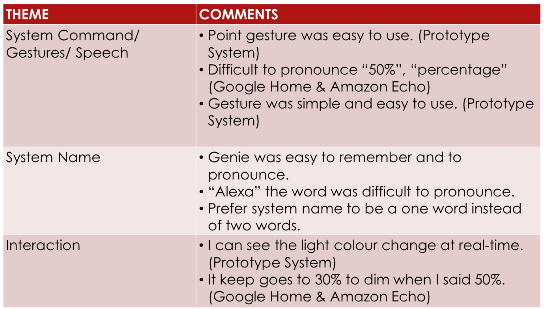


Figure 10: NASA-TLX Average WorkLoad

6.2.7 *Survey & Feedback.* Based on participant feedback, 81.3% of the participants support the prototype system with positive feedback after experiencing the system. Participants felt that the system is user-friendly and the speech and gesture commands are easy to remember, despite using such a system for the first time.

Thematic analysis was used to analyze the feedback given by the participants [6]. A summary of the themes is shown in Figure 12.



THEME	COMMENTS
System Command/ Gestures/ Speech	• Point gesture was easy to use. (Prototype System) • Difficult to pronounce "50%", "percentage" (Google Home & Amazon Echo) • Gesture was simple and easy to use. (Prototype System)
System Name	• Genie was easy to remember and to pronounce. • "Alexa" the word was difficult to pronounce. • Prefer system name to be a one word instead of two words.
Interaction	• I can see the light colour change at real-time. (Prototype System) • It keep goes to 30% to dim when I said 50%. (Google Home & Amazon Echo)

Figure 11: Identified Themes

6.3 Experiment Results

The overall results demonstrate that the prototype system has good usability and accuracy. The older adults had a positive response towards the prototype system after the experiment. Most of them felt that the prototype system was user friendly and the gestures were simple and easy to perform based on the survey results. The prototype system satisfies the criteria to reduce the complex language requirements for issuing a command and the need for physically moving towards an appliance to control it [14].

7 CONCLUSION

A prototype for speech and gesture interactions to assist older adults aged from 65 to 80 in Singapore to control home appliances such as lights and fan from a seated location was developed. This reduces the language complexity for issuing a command and the need for actual physical movement or usage of other controllers.

The initial results demonstrate good usability and accuracy compared with commercial smart assistant devices, and participants had positive feedback and keen interest in such a system, despite using such a system or technology for the first time. The study could be expanded to obtain more results from a larger population. Additional measurements and health outcomes could also be measured, such as stress levels or increased sense of independence, and other clinical psychological measures of wellness.

ACKNOWLEDGMENTS

The authors would like to thank all the participants for their time and effort in taking part in the experiment.

REFERENCES

[1] Amazon. 2015. Amazon.com Help: Set Up Your Amazon Echo. https://www.amazon.com/gp/help/customer/display.html?nodeId=201601770. Accessed: 05.11.2017.

[2] Mohammad Hamdaqa Atli F. Einarsson, Patrekur Patreksson. 2017. SmartHomeML: Towards a Domain-Specific Modeling Language for Creating Smart Home Applications. *2017 IEEE International Congress on Internet of Things (ICIOT)* (2017), 82–88.

[3] Dieter Bohn. 2016. Google Home is smart, loud, and kind of cute. https://www.theverge.com/2016/10/4/13156676/google-home-assistant-speaker-photos-video-device-hands-on. Accessed: 09.11.2017.

[4] Dieter Bohn. 2017. You can finally say 'Computer' to your Echo to command it. https://www.theverge.com/tldr/2017/1/23/14365338/amazon-echo-alexa-computer-wake-word-star-trek. Accessed: 05.11.2017.

[5] Richard A. Bolt. 1980. Put-That-There: Voice and Gesture at the Graphics Interface. *SIGGRAPH Comput. Graph.* 14, 3 (July 1980), 262–270. https://doi.org/10.1145/965105.807503

[6] Virginia Braun and Victoria Clarke. 2006. Using thematic analysis in psychology. *Qualitative Research in Psychology* 3, 2 (2006), 77–101. https://doi.org/10.1191/1478088706qp063oa

[7] John Brooke. 2013. SUS: A Retrospective. , 29 - 40 pages. http://uxpajournal.org/sus-a-retrospective/

[8] Chris Burke. [n. d.]. The Decibel Level of Normal Speech. https://classroom.synonym.com/decibel-level-normal-speech-8599569.html. Accessed: 05.11.2017.

[9] Wendy A. RogersArthur D. FiskSherry E. MeadNeff WalkerElizabeth Fraser Cabrera. 1996. Training older adults to use automatic teller machines. *Human Factors: The Journal of the Human Factors and Ergonomics Society* 38, 3, 425–433.

[10] Juliana Chan. 2017. Ageing isn't a tsunami. In *SMU Office of Research and Tech Transfer.* https://research.smu.edu.sg/news/smuresearch/2016/12/20/ageing-isnt-tsunami

[11] Hartmut WandkeLucienne Christian Stobel and WandkeLucienne Blessing. 2010. Gestural Interfaces for Elderly Users: Help or Hindrance?. In *Gesture in Embodied Communication and Human-Computer Interaction*, Stefan Kopp and Ipke Wachsmuth (Eds.). Springer Berlin Heidelberg, Berlin, Heidelberg, 269–280.

[12] G. Clauser and Brent Butterworth. 2010. Is the Google Home the Voice-Controlled Speaker for You? https://thewirecutter.com/reviews/google-home-voice-controlled-speaker/. Accessed: 01.10.2017.

[13] Lacey Colligan, Henry W.W. Potts, Chelsea T. Finn, and Robert A. Sinkin. 2015. Cognitive workload changes for nurses transitioning from a legacy system with paper documentation to a commercial electronic health record. *International Journal of Medical Informatics* 84, 7 (2015), 469 – 476. https://doi.org/10.1016/j.ijmedinf.2015.03.003

[14] John Donaldson, Catherine J.Smith, Balambigai Balakrishnan, Mumtaz Md. Kadir, and Sanushka Mudaliar. 2015. *Elderly Population in Singapore: Understanding Social, Physical and Financial Needs.*

[15] Matthew Field. 2017. From Amazon Echo to Google Home, the best smart home devices for 2018. https://www.telegraph.co.uk/technology/0/amazon-echo-google-home-best-smart-home-devices-2018/. Accessed on 19.12.2017.

[16] Kinect for Windows Team. 2014. The Kinect for Windows v2 sensor and free SDK 2.0 public preview are here. https://blogs.msdn.microsoft.com/kinectforwindows/2014/07/15/the-kinect-for-windows-v2-sensor-and-free-sdk-2-0-public-preview-are-here/. Accessed: 12.09.2017.

[17] Kathrin Gerling, Ian Livingston, Lennart Nacke, and Regan Mandryk. 2012. Full-Body Motion-Based Game Interaction for Older Adults. In *ACM, Proceedings of the SIGCHI Conference on Human Factors in Computing Systems.*

[18] Google. 2017. Google Home Mini. https://store.google.com/?srp=/product/google_home_mini. Accessed on 10.11.2017.

[19] Susan K. Gordon and W. Crawford Clark. 1974. Adult Age Differences in Word and Nonsense Syllable Recognition Memory and Response Criterion. *A Journal of Gerontology* 29 (1974), 659 – 665.

[20] Sandra G. Hart. 2006. Nasa-Task Load Index (NASA-TLX); 20 Years Later. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* 50, 9 (2006), 904–908. https://doi.org/10.1177/154193120605000909 arXiv:https://doi.org/10.1177/154193120605000909

[21] Vijayan Sugumaran andMarvin Hubl Joerg Leukel. 2016. The Role of Application Domain Knowledge in Using OWL DL Diagrams: A Study of Inference and Problem-Solving Tasks. (2016).

[22] J.Wu, Daqing Zhang, Zhaohui Wu, Yingchun Yang, and Shijian Li. 2010. GeeAir: a universal multimodal remote control device for home appliances. *Personal and Ubiquitous Computing* 14, 8 (2010), 723–735.

[23] Victoria KACHAR. 2003. Terceira idade e informaÌĄtica: aprender revelando potencialidades. Sao Paulo: Cortez.

[24] Thalmic Labs. 2013. MYO. https://www.thalmic.com. Accessed: 05.09.2017.

[25] Leapmotion. 2015. Leap Motion. https://store-world.leapmotion.com/products/leap-motion-controller. Accessed on 05.09.2017.

[26] Chiara Leonardi, Adriano Albertini, Fabio Pianesi, and Massimo Zancanaro. 2010. An Exploratory Study of a Touch-based Gestural Interface for Elderly. In *Proceedings of the 6th Nordic Conference on Human-Computer Interaction: Extending Boundaries (NordiCHI '10).* ACM, New York, NY, USA, 845–850. https://doi.org/10.1145/1868914.1869045

[27] James R. Lewis and Jeff Sauro. 2009. The Factor Structure of the System Usability Scale. In *Human Centered Design*, Masaaki Kurosu (Ed.). Springer Berlin Heidelberg, Berlin, Heidelberg, 94–103.

[28] Logitech. 2017. Logitech Harmony Hub. https://www.logitech.com/en-us/product/harmony-hub. (2017). Accessed on 05.09.2017.

[29] Macedonia and Manuela. 2014. Bringing Back the Body into the Mind: Gestures Enhance Word Learning in Foreign Language. *Frontiers in Psychology* 5 (Dec 2014).

[30] J. Mark R. P. Thomas, Jens Ahrens and Ivan Tashev. 2012. Optimal 3D beamforming using measured microphone directivity patterns. Acoustic Signal Enhancement; Proceedings of IWAENC 2012; International Workshop, Mountain View, CA, USA.

[31] W. Y. Mei and Teo Zhiwei. 2016. *Technologies for Ageing-in-Place: The Singapore Context.*

[32] NASA. 1986. NASA Task Load Index. https://ntrs.nasa.gov/archive/nasa/casi.ntrs.nasa.gov/20000021488.pdf

[33] United Nations. 2017. Department of Economic and Social Affairs, Population Division (2017). In *World Population Ageing 2017 - Highlights (ST/ESA/SER.A/397).*

[34] United Nations. 2017. Global Issues: Ageing. In *Ageing.* http://www.un.org/en/sections/issues-depth/ageing/

[35] Department of Statistics Singapore. 2017. Population Trends. In *Singapore Management Univrsity. Lien Centre for Social Innovation.*

[36] Ollie. 2015. Designing For The Elderly: Ways Older People Use Digital Technology Differently. https://www.smashingmagazine.com/2015/02/designing-digital-technology-for-the-elderly/. Accessed on 10.10.2017.

[37] KEVIN PARRISH. 2017. ZigBee, Z-Wave, Thread and WeMo: What's the Difference? https://www.tomsguide.com/us/smart-home-wireless-network-primer,news-21085.html. Accessed on 10.09.2017.

[38] Philips. [n. d.]. Philips Developers. https://developers.meethue.com. Accessed on 05.09.2017.

[39] Philips. [n. d.]. Philips Hue. https://www2.meethue.com/en-sg. Accessed on 05.09.2017.

[40] Warren JE Rohrer JD, Knight WD. 2008. Word-finding difficulty: a clinical analysis of the progressive aphasias. *BrainâĂŕ: A Journal of Neurology* 131(Pt 1), 7 (2008), 8 – 38. https://doi.org/10.1093/brain/awm251

[41] Margaret Rouse and Matthew Haughn. 2017. Smart speaker. http://whatis.techtarget.com/definition/smart-speaker. Accessed on 05.11.2017.

[42] Samsung. 2017. Samsung Developer Portal. (2017). http://docs.smartthings.com/en/latest/getting-started/overview.html

[43] Samsung. 2017. Samsung Smarthings developer. https://developers.smartthings.com.

[44] Samsung. 2017. Samsung SmartThings. https://www.samsung.com/us/smart-home/smartthings/.

[45] Patrick Sinclair. 2017. Samsung SmartThings Hub vs Logitech Harmony Hub: Which is Best? (2017). https://www.allhomerobotics.com/samsung-smartthings-hub-vs-logitech-harmony-hub/

[46] Kinect's Windows Team. 2014. Set up Kinect for Windows v2 or an Xbox Kinect sensor with Kinect Adapter for Windows. (2014).

[47] Y.Han, Jonghwan Hyun, and Taeyeol Jeong. 2016. A smart home control system based on context and human speech. In *Advanced Communication Technology (ICACT), 2016 18th International Conference on, Pyeongchang, South Korea.*