



Use of Voice Activated Interfaces by People with Intellectual Disability

Saminda Sundeepa Balasuriya,
Laurianne Sitbon, Andrew A. Baylor, Maria Hoogstrate, Margot Brereton
Queensland University of Technology

samindasundeepa.balasuriya@connect.qut.edu.au, l.sitbon@qut.edu.au, a.baylor@qut.edu.au, m2.hoogstrate@qut.edu.au,
m.brereton@qut.edu.au

ABSTRACT

People with intellectual disability are keen users of information technology, but the need for spelling and typing skills often presents a barrier to information and media search and access. The paper presents a study to understand how people with intellectual disabilities can use Voice Activated Interfaces (VAIs) to access information and assist in daily activities. The study involves observations and video analysis of 18 adults with intellectual disability using VAIs and performing 4 tasks: calibrating the VAIs, using voice assistant (Siri or Google) to search images, using voice to query Youtube, and using the voice assistant to perform a daily task (managing calendar, finding directions, etc.). 72% of participants stated that this was their preferred form of input. 50% could perform all four tasks they attempted with successful outcomes, and 55% three of the tasks. We identify the main barriers and opportunities for existing VAIs and suggest future improvements mainly around audio feedback given to participants. Notably, we found that participants' mental model of the VAIs was that of a person, implications for which include the user having to speak in long polite sentences and expecting voice responses and feedback about the state of the device. We suggest ways that VAIs can be adjusted so that they are more inclusive.

CCS CONCEPTS

• **Human-centered computing** → **Human computer interaction (HCI)** → Interaction devices → Sound-based input/output • **Human-centered computing** → Accessibility → Empirical studies in accessibility.

KEYWORDS

Voice activated interfaces, voice assistant, intellectual disability, information access

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from Permissions@acm.org.

OzCHI '18, December 4–7, 2018, Melbourne, VIC, Australia

© 2018 Copyright is held by the owner/author(s). Publication rights licensed to ACM. ACM 978-1-4503-6188-0/18/12...\$15.00

<https://doi.org/10.1145/3292147.3292161>

ACM Reference format:

Saminda S. Balasuriya, Laurianne Sitbon, Andrew A. Baylor, Maria Hoogstrate and Margot Brereton. 2018. Use of voice activated interfaces by people with Intellectual Disability. In *Proceedings of ACM OzCHI'18 conference, Melbourne, Australia*, 10 pages.

1 INTRODUCTION

People with intellectual disabilities are often keen users of information technologies, however the technologies are often not designed to suit their skill sets. Intellectual disability (ID), as defined by a leading authority, the American Association of Intellectual and Developmental Disabilities (AAIDD), is “characterized by significant limitations in both intellectual functioning and in adaptive behaviour, which covers many everyday social and practical skills and originates before the age of 18” [9]. This definition strikes a comparison with neurotypical people, the broader group of people for whom most technology is designed, and suggests that people with ID will find it more challenging to use information and communication technologies (ICTs) because of comparative cognitive, physical and behavioural challenges. Our hope is that rather than focusing on difference and characterizing participation in relation to deficits, technologies can be designed to support people with ID on their own terms, to suit their own skill sets, so that they have much the same access as everybody else to ICTs and the benefits that they have to offer.

A study conducted in 2005 in the USA showed that only 41% of people with disabilities had access to a computer and 25% had access to the internet. In the same year 68% of the total population of America had access to the internet [5]. A more recent study in UK found that people with disabilities were 3 times more likely to have never accessed the internet [5]. In as much as technology use of persons with ID is not at par with general society, research notes that the few persons with ID who use ICTs, usually use them on the web for several activities, including information seeking.

Voice and speech technologies have potential to overcome barriers presented by typing and spelling, however, research exploring how persons with ID access information using interactive voice activated services or search engines is limited [23, 6]. In this study, we explore how voice activated technologies can be used by people with ID to access online information. To do

this, we have conducted 18 semi-structured interviews and observations with persons with ID on accessing information on (the topics that the search was based on) using Google and Siri.

Our findings highlight the abilities of people with ID to use VAIs as well as the limitations of the technology to support people with ID in accessing information using voice and speech technologies. Further to this, our findings contribute an understanding of the perceptions and attitudes of people with ID towards voice and speech technologies.

Our work contributes unique insights into what kinds of voice search interfaces persons with ID prefer and why, and a reflection on design ideas to support effective use of voice activated interfaces by persons with ID.

After introducing related work, we detail our methodology and then present our findings and observations. We finally present a discussion of our results, implications for design and conclusions.

2 RELATED WORK

2.1 Technology demands on working memory

People with ID often find speaking easier than expressing their ideas in written sequences [16]. Writing puts a lot of demands on working memory, which can be challenging for a person with ID [19]. Working memory is the various cognitive means by which individuals maintain and manipulate information while completing a task [26]. It involves the short-term memory (STM) and long-term memory (LTM). A typing task for example requires an individual to maintain the full goal of the message in mind (STM) while recollecting the multiple symbols that are the letters of that specific language (LTM). Limited literacy and reading comprehension also poses problems for people with ID that want to use information presented in written form [15]. A study of 57 participants with moderate ID conducted to observe their ability to use web browsers suggested that reducing the required working memory load and utilizing simpler visual search might be an effective way of improving web sites so that participants with ID had more success in using them [29]. Learning to use computers early in their development can also help children with disabilities use them more competently [10].

2.2 Mainstream technology as assistive devices

It is vital to develop assistive technologies that exploit users' abilities rather than focus on disabilities [27]. Assistive technologies tend to favour such an approach, providing ongoing task-based assistance [1]. Mainstream technologies used as portable assistive devices (handheld computers, PDAs, iPods, audio recorders and iPads) have been proven to increase independence among students with ID [7], mostly through an audio element to help prompt the user to complete tasks. The fact that such devices are also commonly used by the public increases their social acceptance as the users were not embarrassed to use these devices in public. Lee et al. [15] states that devices that use touchscreens, for example smartphones and tablets, are far easier to operate and better suited for people with ID. This is because the user can see a cause and effect relationship when they press the

screen and see a corresponding reaction, like taking the user to the desired web page [15].

2.3 Online information access by people with ID

People with ID are rarely included in decisions relating to the design of website and app interfaces. That is why such interfaces are cognitively inaccessible for this group of people [15]. 13% of internet uses with ID stated that complexity was a barrier to use, [20]. People with ID show great interest and motivation to navigate and learn using the internet [22], however, it is a complex and interactive task that requires the user to realise that specific actions are needed to lead to the desired outcomes [5]. They are required to make a connection between the action that they performed and the corresponding response from the device. For example, if they do not make this connection, they may tap or click multiple times when the page does not load because they do not realise that one click/tap is sufficient [5].

Other research has suggested that layouts need to be simple and minimize distractions and clutter [23], and that participants with ID clicked on image results over text results. Rocha et al [22] found that participants were better guided by cartoonish images than text, as the text was not able to capture their attention, but rather it confused them. Wilson et al [28] found that in their case people with ID preferred real photos to abstract icons.

People with mild to moderate developmental disabilities have common difficulties in entering the correct spelling of the words in the Google search box [11]. Selecting from a large variety of text links in the search results was also a problem for these users [14].

Observational research done in educational facilities for people with ID in Brisbane revealed that supporting emotional, social and visual dimensions of information seeking are vital when designing interactions for people with ID [24]. This study also revealed the competencies to proceed with aspects of information seeking via search engines: most participants knew where the search bar was and how to use the virtual keyboard. Some made spelling errors but could still use the suggested results to find what they were looking for. When the participants made spelling errors they were usually unaware of their mistake.

2.4 Speech Technologies

In the past speech to text programs used to require a lot of training to correctly interpret speech to text, but modern software does not require as much training. Thus, current speech to text software can help compensate for challenges with transcription, spelling, handwriting, punctuation and capitalisation [16]. This is supported by a study conducted on three students with various disabilities who used speech to text software. They found an increase in the total number of words written, a greater number of correct writing sequences using speech to text technology, and preferences for the software over writing passages by hand.

An exploratory study conducted at a special needs school in Colorado examined students preferred methods of searching for information on Google Chrome. The three input methods were typing, voice search with manual microphone control and hands free voice search. Out of the six participants, three preferred

typing and three preferred hands free voice search [18]. Those who liked typing stated that they had issues with voice recognition in the voice search function. The others stated that they preferred hands free voice search because they did not have to worry about spelling and it was a faster way to get the answers [18]. The study concluded that due to the variation in abilities and conditions among individuals with cognitive disabilities, there is no one design that fits all [18].

Rocha et al. [23] found that participants had more success when typing their queries on Google rather than using their voice to search Google. They were told to search for simple things like cat, dog and bread (in Portuguese). It must be noted that they received assistance when typing for spelling and character recognition. Only 25% of the participants completed the task using voice search. The researchers attributed this failure to mispronunciation of words. However, the 25% that did succeed performed better using Google voice search than standard text search. That is, they were able to complete the searches faster. The researchers suggested more research be done on speech-based search in the future as an alternative for text-based search.

It is no longer necessary for people with ID to use special assistive devices as smartphones can serve this purpose [15]. Typing is still the most common form of input on smartphones, however people with ID have some trouble with spelling and remembering icons [11]. VAIs are of interest because they could offer an alternative to typing and provide a more natural hands-free user interface. Research to date has not focused on how people with ID can use this technology to aid them in their daily life, or the potential problems that they might discover.

3 METHODOLOGY

Previous research undertaken with people with ID investigating their use of technology has employed qualitative methods of data collection. Several studies reviewed by Collins & Collet-Klingenberg [7] on the use of portable electronic assistive technology employed qualitative research. The sample size of the studies ranged from three to forty participants, but most of them had between three and five participants [7].

This study involved 18 participants. The researchers observed the participants while they completed tasks using VAIs. The use of VAIs was also videotaped. Semi-structured interviews were conducted afterwards to gather information about the usability, preferences and attitudes towards VAIs.

The research team met to draw together written user observations and field notes. Video was used for further review and analysis of typical and notable interactions using the method of Interaction Analysis [12]. Using an iterative approach we discussed and distilled key themes from the combined data sets in a thematic analysis approach broadly based on [3].

3.1 Participants

Eighteen participants with intellectual disability participated in this study, seven females and eleven males. Seventeen of the participants are supported in one of four sites of Endeavour Foundation (Learning and Lifestyle centres), located in various

suburbs throughout Brisbane. Sites 1, 2 and 4 are places for people with intellectual disability to socialize, access community and learn tasks associated with daily living skills. Site 3 is a supported employment industry for people with a disability. P17 was supported by a similar but different organization, and was known to the research team, because one member of the team had worked for both organisations. This participant was interested to participate in the study. The age group ranged from 18-63. The average age of the participants was 38.7. In the remainder of the paper, we identify as young adult participants those aged 18-34 (the age range used by the Australian Bureau of Statistics) and the older participants as those over the age of 42.

Rather than categorizing the nature of the participant's disability, which was unknown to us and also regarded as irrelevant from a person-centred perspective, the participants are included in the study by virtue of participating at Endeavour Foundation or a related organization supporting people with ID and showing some interest in technology use. What then becomes relevant is the particular interactions that each individual experiences when using the technology. See also Sitbon et al. [25] call for a non-clinical approach to describing people with intellectual disability.

Having worked with Endeavour for several years, we have also developed a good relationship with the support workers. However, the organisation's approach is to encourage participants with ID to speak for themselves and to foster independence. Consent was directly sought from participants, using explanations and an easy-to-read consent form designed for the purpose, with support workers helping to identify and encourage participants who would be interested to take part. In some research, involvement of proxies and support workers is necessary to fill in missing perspectives in design [4, 21]. However, in this case it was not felt necessary to have support workers or families involved in the research, rather it was beneficial for participants to engage by themselves.

The research team travelled to the sites to meet the participants. They were identified by the support workers as having a range of diagnoses that meet the AAIDD definition of intellectual disability. The researchers went to Site 1 and Site 2 three times and the other places only once. The reason for repeat visits was to observe participants a second time to see if the voice search skills that they learned initially had been retained.

Table 1: Participant details

Site	Participant	Gender	Age	Task Success Rate %
1	P1	M	63	0
1	P2	F	44	100
1	P3	M	44	75
1	P4	M	50	50
1	P5	F	42	25
2	P6	M	44	75
2	P7	M	46	75
2	P8	F	19	75

3	P9	M	31	100
3	P10	M	32	100
3	P11	F	22	100
3	P12	F	34	0
4	P13	M	44	0
4	P14	F	23	0
4	P15	M	48	75
4	P16	M	43	100
Other	P17	F	51	100
2	P18	M	18	75
Mean			38.7	62.5

3.1.1 Familiarity with VAI technologies The four sites that we visited provide their users with smart devices like iPads, Samsung tablets and even large touch screens for interactive learning and web browsing. Sites 2 and 4 had dedicated rooms for this technology. All the people who used the services at these centres were accustomed to using smart tablets. They primarily used these devices to play games, watch videos and take photographs. All the participants had used smart devices prior to this research. Some people had their own smartphones. They used them mainly as communication devices to contact their family members and support staff. Some of the younger participants, such as P11 and P8, were using social media like Snapchat on their smart devices.

Participants are not given regular training on how to use these devices to search for information and complete useful daily tasks like setting reminders, getting directions or checking the weather. They are also not taught how to use these devices independently, as the devices are locked with a passcode that is known only to paid staff. Often when they want to watch a video on YouTube the support worker would open up the device and type the search query for them. If they wanted to watch a different video they would look for it on the suggested videos section instead of entering a new search query.

Most participants were unaware of the speech search functions on the devices. Some have family members who were using Siri but have not been taught how to use it themselves. P11 was the only participant who was using Siri prior to our interview. However, even she was not aware that Siri could be used to search for information online. She simply used it for tasks such as setting reminders.

3.2 Interview Settings

The participants signed easy to read consent forms giving permission for us to observe and record their activities with this technology. Most participants still needed help to read the forms, so they were read out to them. During the first meeting with the participants they were asked about their previous experience with technology. The questions started out broad and became narrower to understand the kind of previous exposure they have had with smart devices, searching on web browsers and internet use in general. An example of a broad question was "Have you used a smart device like a tablet or smartphone before?". A more specific

question would be "Have you used Siri to search for [participant's interest] on the internet?".

Most Endeavour Learning and Lifestyle centres had second generation iPads that had Siri. In most of these iPads Siri had been disabled, probably because it tended to initiate unexpectedly and cause confusion as no-one had been taught how to use it. Thus, Siri had to be reactivated and configured to come online when the participant said 'hey Siri'. We incorporated these devices into our research so that participants could practice using them on later occasions. To test Google Assistant an HTC 10 Android device was used. The settings were configured so that when the participant said 'ok Google' the voice assistant would come online. If the participants had their own smart devices we encouraged them to use those devices. If the device was running on iOS we used Siri and if it was an Android device we used Google Assistant. They are examples of voice assistants. Voice search on YouTube is not a voice assistant.

Next, we provided participants with a brief tutorial on how to use VAIs. This tutorial was also given to those who said that they had previous experience with the software. This was done as a refresher, so that they did not have to tax themselves to remember certain procedures. Previous research has showed that systematic instruction can help people with ID learn digital literacy skills like sending emails [6]. We provided systematic instruction and then asked participants to perform the task we had just shown. For example, we would show them how to tap on the mic icon to activate voice recognition and then ask them to do the same. When teaching participants how to use Siri the researcher would open up app, set up voice recognition and use Siri to open other apps like YouTube or search for sports scores. Afterwards we set up Siri to recognise the participant's voice so that they could activate it hands free. Once the voice assistant had been calibrated to recognise the participants they were asked to undertake a list of tasks. They were given time to do the tasks on their own but prompts were given if they appeared to be stuck or asked for help.

The participants were asked to perform the following tasks:

Task 1 - start Siri or Google Assistant using their voice

Task 2 - search for a video of interest. They did this by asking the voice assistant to open the YouTube application. Next, they had to activate the voice search function in YouTube by tapping on the mic icon. Then they had to search for their interest and open up a video that was relevant to that interest.

Task 3 - search for an image of interest on a search engine.

Task 4 - use the VAI to complete a useful daily living task like setting a calendar event, finding directions to go home, checking the next day's weather or to call or text someone.

Participants were asked to search for something they were interested in, so that the task had intrinsic value and they would be personally motivated [5]. Previous research with young adults with ID that explored searching on YouTube used a similar approach where participants were asked to search for videos related to their interests [2, 24].

After completing these tasks, we conducted a semi structured interview to gain more insights from the participants about their experience and their thoughts on VAIs. We asked participants

different questions based on how they performed but had some staple questions such as:

Did you enjoy using your voice to access information?

Was it easier than typing into the search field?

Would you use it in the future?

What problems did you come across when using this application?

How can it be improved for easier use?

3.3 Methods of Analysis

The research team consisted of a senior researcher, research student and research assistant, plus a mentor. The research team analysed all of the notes made during observations and the interview recordings to reach a consensus on the themes.

In our research we observed participants and wrote down our observations in logbooks. The semi structured interviews were carried out after the tasks were completed. They were done to record participants' attitudes and perceptions towards the VAI, the results of using it, whether they are likely going to use it again and suggested improvements. The entire observation and interview process took between 15-30 minutes.

All of the VAI tasks and interviews were video recorded for later analysis. The assurance of having video footage enabled us to focus on observing and assisting the participants when needed. Re-watching the footage helped us verify things that we were unsure of. For example, if the participant reacted favourably or unfavourably to a certain output from the device. Analysing videos helped deduce the success rate of tasks. It was also useful for the researchers to come to a consensus on the themes. By re-watching the footage of the participants using the VAIs we were able to draw up themes and agree or disagree on which ones we thought were relevant to the use of VAIs.

The main devices used to record the interviews were the Theta Ricoh 360 degree camera and the Kodak Pix Pro 360 degree camera.

4 FINDINGS

We first present an account of our observations of how participants approached each of the tasks in section 4.1. We then present some observations across the tasks, organised according to the themes:

Graphical user interfaces issues;

User perceptions and preferences;

Speech pronunciation related barriers;

Conversational expectations of devices.

4.1 Participants' ability to complete the tasks

For each of the four tasks that we invited participants to undertake, we found that except in three cases where the VAI could not parse the participant's pronunciation, and cases where participants did not attempt a task at all, they were able to use VAIs. Overall, 50% of the participants could perform all the tasks they attempted with successful outcomes, and 55% could successfully complete three of the tasks. P1 and P14 interacted through touch with the devices but chose not to attempt any of

the tasks. P8 demonstrated how she could use the device with her voice, including, daily tasks (task 4), but did not respond to requests to perform other tasks. P5 interacted well with the research team but found it difficult to focus on the tasks, and only attempted the third task. The research team identified that P4, P12 and P13 had pronunciation that was less intelligible to the VAI, possibly in relation to speech impediments.

Some tasks required more prompting than others, and we now detail our quantitative and qualitative observations for each task.

4.1.1 Task 1: Calibrating the Voice Assistant. The participant's voice was calibrated to the voice assistant on the device (Siri or Google). This is done by saying "Hey Siri" or "OK Google" three times.

Out of the 18 participants, ten could successfully calibrate on the first attempt and four did not attempt the task. The other four participants had trouble pronouncing the names of the voice assistant. One of them said "Giggle" instead of Google and a few others could not pronounce "Siri". Sometimes they were able to get the correct pronunciation after multiple attempts but were not able to maintain consistency over the three repetitions.

Participants who calibrated the voice assistants were then able to activate them by simply saying "Hey Siri" or "OK Google". They found this to be a lot easier than using their fingers to manually open up the assistants. All of them explicitly stated that they preferred this mode of activation.

4.1.2 Task 2: Asking the voice assistant for an image of interest on the internet. The participants searched for an image of one their interests using the voice assistant.

Nine of the participants were able to find one or more images of interest using their voice. Two of them had to ask twice to get satisfactory results, and three of them were unsuccessful on occasion and decided to ask for something else when that happened. Two of the participants attempted the task with a range of queries but never obtained satisfying images from the system. Seven participants chose not to attempt the task.

Participants would speak to the devices in complete sentences. For example, P6 searched for images of Spiderman by asking Google Assistant to "Show me images of Spiderman". The search results came up with images of Spiderman and a short audio description of the character by Google. P6 was very pleased to hear it as he had trouble reading off the screen. A few participants had to specify in a second attempt that they wanted to see images or pictures, as the search results initially included text links rather than images.

4.1.3 Task 3: Using a Search Engine (searching for videos on YouTube). All the participants who had calibrated their voice were able to open the YouTube app by voice command, others were prompted to locate the icon. We then asked participants to first type in their query in the search bar. This provided us with insights on their typing ability and provided them with an opportunity to formulate their search need in a familiar context. Eight participants succeeded in finding their videos by typing in the search box. Most of them had trouble spelling words and were attempting to spell complete sentences. One individual who was not able to calibrate their voice was able to type and search for his interest on YouTube.

Afterwards participants would repeat their search, using voice input instead of typing. Nine out of the 15 participants who attempted the task could complete it with satisfactory results, but they required a lot of prompting to implement the longer sequence of actions, requiring them to remember not only the sequence but also new icons associated with the sequence: from the search bar, tap the microphone icon, then speak the query; if the query is a reformulation, then clear the previous one with the X icon, then start again. Several participants did not identify the mic icon without prompting and clearing a prior query led to confusion.

4.1.4 Task 4: Using the Voice Assistant for Daily Tasks. For this final task, we prompted participants to use the voice assistant to look for directions, weather, time, and some who were regular users of mobile technologies also suggested to use it for sending text messages or setting calendar entries. Eight of the 12 participants who attempted this task were satisfied, and most participants attempted two or more tasks.

Participants who had previously exhibited difficulties formulating search queries or who had pronunciation that was not intelligible to the VAI were not able to complete these tasks independently. Other participants typically succeeded with repeated instructions and prompting but could not replicate the steps independently. For example, P6 was able to use Siri to set up a calendar reminder to go to the park the following day at noon but was not able to repeat the process. This is because he tried using different phrasing that did not trigger Siri to open up the calendar and set up an event.

P2 was continuously successful at this task. She was able to get directions from her current location to the city and find ballet classes near her location. P17 was able to set up calendar events, enter a researcher's phone contact to her contact list and send her a text message which she dictated to Siri. During a follow-up meeting, P17 told us that she had since continued to use Siri to set up reminders, take calls, send text messages, get directions and check for public transport to get around the city.

4.2 Graphical User Interface (GUI) issues

Older participants had trouble with the small icons, which was attributable to either a level of vision impairment, their ability to remember what to tap in a given scenario, where it was located and when to tap the icon. P16 and P2 were able to recognise which icons to tap in a given scenario without prompting.

We met P2 on three separate occasions during this research. During the first meeting we showed her which icons to tap to activate voice search on YouTube. She was able to identify and use the correct icons in the next two meetings as well. She would use Siri to open the YouTube app and then use the app's voice search feature to search for her favourite music.

To search using YouTube, a user first needs to locate and tap the magnifying glass (Figure 1.a), then the microphone icon has to be tapped to activate voice search on YouTube (Figure 1.b). To clear the previous search the 'X' icon had to be tapped (Figure 1.c). And then tap the mic icon again to activate voice search. This brings up a screen with a pulsating animation (Figure 1.d). P13 and P1 found the voice input screen a major distraction as the

pulsating encouraged them to tap on the icon which would cancel the input screen.

Google Assistant (Figure 1.e) had a 'four dots' animation as circled in red in the screenshot. It moved slightly up and down while waiting for the user to input using their voice. Most participants did not notice this animation. It was not as prominent as Siri's animation (Figure 1.f).

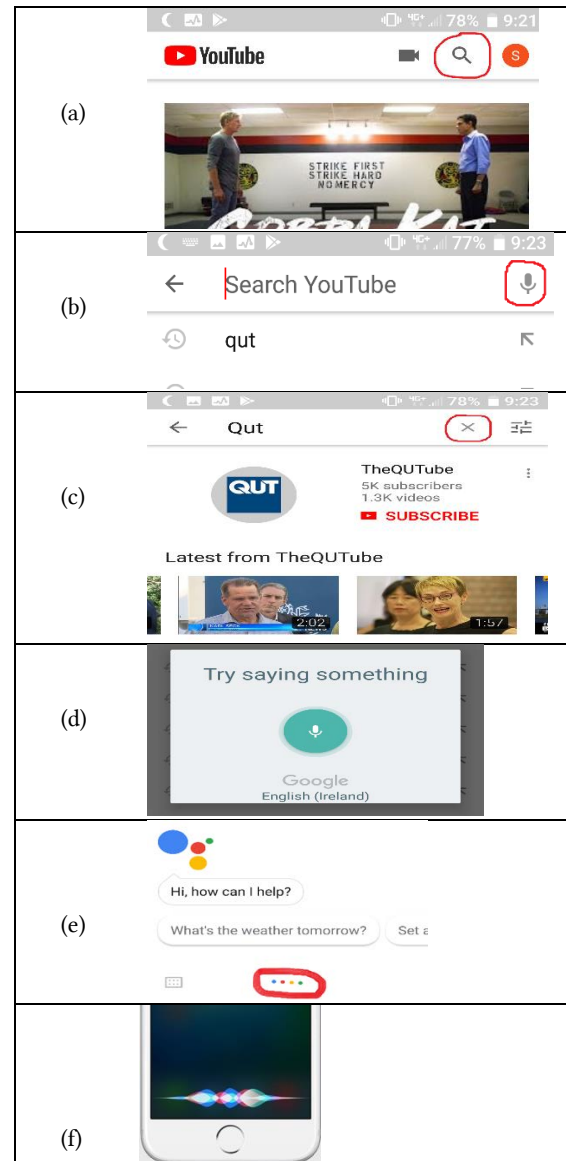


Figure 1: GUI elements

4.3 User perceptions and preferences

At the end of each interview we asked participants how they felt about using VAIs. We asked them if they preferred using their voice rather than their hands to input data. Thirteen out of eighteen (72%) said they preferred using their voice and two said they would rather input using their hands, while three had no preference. The main reason for not preferring speech input was

when speech impediments prevented a smooth user experience. The voice assistants had trouble recognising what they were saying, resulting in incorrect or no search results and leading to frustration.

All those who could speak such that the VAI understood said that they preferred using the VAIs, either because they did not have to worry about spelling, or using their voice to input was much faster than typing, or Siri and Google Assistant gave audio replies, or it was easier to complete tasks like setting calendar events.

Several participants said using VAIs made doing the tasks much simpler than typing. P2 said she preferred using Siri because it gave 'clear answers'. By this she meant that she did not have to read the results and can just ask for image results or have Siri read out the search result. P7 stated that he preferred using speech to search because he did not have to concern himself with spelling words correctly. Many others echoed his sentiments as they also had a hard time remembering the correct spelling of a word.

Out of the 13 that said they enjoyed using VAIs, all of them said that they would like to use it again in the future. P18, who was our youngest participant at the age of 18, stated that he liked it and would ask his sister to set up Siri on his personal phone when he went home that day. P9 was also keen to get Google Assistant set up on his Android device. We set it up for him and he was delighted that he could use it whenever he wanted to. P2 stated that she has asked her parents to buy her an iPad as she enjoyed the interactions. They were clearly left with a positive impression.

4.4 Speech pronunciation related barriers

People with speech impediments had trouble with voice assistants. Siri and Google could not recognise what they were saying and it was not possible to calibrate their voices to use hands free voice search. P1, P12, P13 and P5 could not calibrate their voices and so were not able to use the hands free features. They were shown how to tap the mic icon to use the voice assistant instead. This added to the number of things they had to remember to do. Even though they could access the mic icon, the VAIs had trouble recognising what they were saying. The research team also found these participants' speech to be unintelligible at times and this was possibly due to speech impediments. Some users did not have speech impediments but spoke very softly. Their words would not be picked up by the voice assistant, or only a few of the words would be picked up resulting in an incorrect search query.

In some instances the VAIs had trouble identifying the pronunciation of certain words. When P16 wanted to search for his favourite singer, Max Merritt, Google could not identify his pronunciation of "Merritt" and returned a wrong search but was able to recognise it when P16 repeated the query louder. They could successfully search for things that were easier for them to pronounce. Those who could use the hands free features said it made the process much easier and enjoyable.

Participants who did not have severe speech impediments were able to use handsfree voice search. The issues they did face could be rectified by repeating the search query again or

pronouncing a word slightly differently. Rocha et al. [23] conducted a study that compared the effectiveness of typing a search query in Google versus using Google Voice Search. In their study only 25% could successfully use Google Voice Search. The researchers attributed this low success rate to Google not understanding the participant's pronunciation. Our participants also faced these problems, but the average success rate was 62.5%. six participants completed all four tasks and 6 were able to complete three out of four.

It was also possible to teach participants how to change the way they speak to the voice assistant. When P2 began by saying "Hello Giggle" we were able to teach her the standard way of pronouncing the word so that she could calibrate her voice. We began by slowing down her speech rate and correcting her pronunciation. After she repeated this a few times she was able to calibrate her voice. Speech pathologists can improve the speech skills of people with ID. By providing instructions in understandable terms while demonstrating the correct learning processes in well-structured environments, people with ID can improve their speech output [8].

4.5 Conversations with devices

People with ID conversed with the voice assistants as they would with a person. P18 commented on the 'stupidity' of the device when it did not return the expected result. All the other participants were extremely patient, even when they were not getting the expected results. They were willing to try several times.

Participants' natural language included introductory terms, polite sign offs and long sentences. When P15 wanted to search for a Phil Collins song he would say, "Hello YouTube can you show me Phil Collins music please". P17 was visibly tired after having used Siri for about 20 minutes. Even though P17 was one of our most successful participants when using Siri, she had trouble communicating long complicated sentences and preferred to break her queries to smaller chunks.

Long drawn out sentences also reduced the likelihood of success, as the program only listens to part of the request and stops listening after lengthy pauses for more input. Also, some participants needed time to formulate what they wanted to say, and the device had stopped listening, or shown completely irrelevant search results. When P1 wanted to search for hot cross buns, he took about 10 seconds to think of how he wanted to phrase his query and by the time he started saying it, Siri had stopped listening. He would say the query but would not get any search results or response in return. P1 was not aware that Siri had stopped listening and was confused about not getting a reply. Siri would let the user know if it did not understand a query but would not give any prompts if it timed out.

As the participants engaged in conversations, they were expecting audible responses. However, these were usually inconsistent. Google Assistant would show a knowledge graph of the search results (images and text from multiple sources) and read off the knowledge base only occasionally. Siri would always give an audio reply, except in instances of time out.

Google did not audibly indicate if the query was not understood and the participants were often confused as to why the expected search results were not showing up on the screen. P10, a younger participant, was confused by Google's lack of response. When it was evident that no answer was coming he would ask us what went wrong. P10 stated that it would be helpful if the VAI would prompt him to repeat his query.

In Nour's study [18] only one user found the audio response helpful. However, participants in this study found it very helpful and were not happy when they did not get one. They would smile and engage with the VAI when getting audio replies.

5 DISCUSSION

Our study has shown that VAIs can be beneficial for people with ID, with 72% of participants stated that they preferred using their voice than typing. Only one participant was using a voice assistant prior to this research. All the others stated that they were unaware of this technology. It must be noted that this technology will not be helpful for people who cannot speak or have a significant speech impediment. Previous literature highlighted a lack of research on how people with ID could benefit from using speech technology, despite it being universally accepted that assistive technology has benefits supporting the person's social and emotional independence. Assistive technologies are an effective alternative or a supplement to one on one coaching [7]. VAIs add another pathway to being more independent. The ability to access information without having to use one's hands is a huge advantage for people with motor and intellectual disabilities. It is a step towards natural user interfaces which provides a more natural way for humans to interact with machines using speech, gestures and touch.

5.1 Necessity of voice technology

It is important to consider whether using VAIs is necessary for information access by people with ID. Answering this question will indicate whether it is necessary to invest in training people with ID to use VAIs. Most modern smartphones have voice assistants but many people still prefer typing to find information. However, typing does not present a cognitive challenge to people without ID. Understanding which form of input presents an easier route to accessing information is important.

Most people with ID who were observed for this study had difficulties typing. This ranged from not being able to type or spell, to those who were able to type proficiently and fast, (this group was in the minority). P15 would type very slowly and have many spelling mistakes when attempting to type. P7 had limited mobility in his hands and found tapping tiny icons on a small screen difficult so typing was not an easy option for him. There were also instances when people who did not know how to spell the words they were typing took advantage of the options offered to them by the technology. They occasionally had strategies to overcome their inability to spell, P16 used suggested search options given to him by YouTube after typing only half the sentence. This way he was able to find what he was looking for without having to type the whole thing out. Kumin, Lazar & Feng

[14] also reported how people with Down syndrome would use Google's "auto-suggest" feature to find a website instead of typing the entire url.

Only six participants could type without assistance. It should be noted that five out of seven younger participants (below 34) were able to type on touch screen devices. This may be due to the fact that it is common practice for modern special education schools to teach their students how to use smart devices and computers, whereas the older participants with ID did not have such practices when they were in school.

It is important to note whether or not they made use of the search results they received. Even though they were successful in getting the results they wanted, some participants did not care about the search results. They were simply happy that the device returned what they were looking for. This brings into question how much people with ID can actually learn by using the internet and search engines. Higher functioning participants were more interested in the results than those with more severe ID.

5.2 Participant suggestions for future improvements

At the end of the interviews, we asked participants what improvements they thought could be made to improve current voice assistants and voice search applications. P10 stated that having animations to point towards the positions of the mic icon would be helpful, and this echoes our observation that some participants would forget which icon to tap to initiate the speech input. People with ID preferred search results that were displayed as cartoonish images than hyperlinks [22]. Since images hold their attention better than text, having a cartoon mic icon may help them to identify it better. P10 also stated that instead of just having an audio response, an animated response where the animations mouthed the words would be helpful.

P17 stated that she would like it if Siri could understand what she was saying even when she was speaking softly. Then she would not have to raise her voice in public and bring attention to herself. P17 was a reserved person and speaking loudly to her phone in public made her uncomfortable. Other participants like P2 and P4 spoke softly. The tail end of their sentences would drift off at times and become softer than when they started speaking. So being able to understand softer voices would be beneficial for people with ID.

5.3 Implications for future VAIs

5.3.1 Keyword identification All participants spoke to the voice assistant in full sentences. The sentences included a greeting and a polite sign off like "Hey Siri show me music of ACDC please". They would speak to the device like they speak to a person. This was not a problem with Siri as it was able to filter out the greetings and identify the keywords, but on YouTube when the entire search query was used, it would often give incorrect search results.

Speech recognition software shows a lower accuracy rate when recognizing speech of elderly users. This is because they are optimized to listen for young adult and middle aged people's

input. Elderly people have a slower speech rate with longer inter-syllabic silence length and slightly lower speech intelligibility [13]. From our observations it can be noted that people with ID also have lower speech rate and long duration of silence. There is a strong correlation between the severity of ID and speech intelligibility [8].

P6, P13 and P7 all took a long time to formulate their search and say it to the device. By the time they said it the device had stopped listening for their input. On Siri, the animation would be a good indication on whether it was listening or not, but Google and YouTube made it harder to identify when they were listening. This would cause frustration among the users. A simple work around would be for devices to have an accessibility option for people with disabilities to have longer input times when using speech software. If the software recognises a trigger phrase like "Hey Siri" endpointer modification can extend the wait time and thus prevent early cut-off [17]. By using pre-processing, it was possible to increase the accuracy of elderly speech recognition by 12% [13]. A similar solution would work for the people with ID too. Another solution would be to provide specialised help for those who have speech difficulties.

5.3.2 Simpler Interface While the visuals and tactile interactions are clearly central to participants accessing information, the interface of the smart devices presented a challenge to many of the older participants. Some participants preferred tablets as they had bigger screens and thus bigger icons. Having other distractions on the screen should also be minimised. Rocha et al. [23] found that having distractions like pop-up windows, videos that auto play and advertisements can completely derail the attention of a person with ID. This was the case with YouTube in particular. The search bar has to be activated by tapping the magnifying glass icon and then to use voice search the mic has to be tapped. These icons are found on the top right-hand side of the screen and take up a very small portion of the screen, especially in smartphones. Removing the previous search and entering a new one involves tapping on the 'X' and then tapping the mic icon again. These icons take up a very small part of the screen and the rest is made up of suggested videos and search results which may distract the user.

The requirements of the computer tasks must be matched with the individual's sensory-motor abilities and their cognitive capabilities. Taking this into consideration will enhance the human computer interaction [29]. The process of remembering what icons look like and searching for their position requires working memory. As mentioned previously, people with ID have limitations in working memory. Having icons that perform the same task be visually consistent across a variety of applications and interfaces will be beneficial. They will not have to remember multiple icons. Also having the icon in a similar position across all interfaces means they only have to remember one location on the screen and this will put less strain on their working memory. On Siri the mic icon is at the bottom of the screen but on YouTube it is alongside the search bar. Having the mic icon appear underneath the search bar and towards the bottom would resemble its position on Google Assistant and Siri. It also would be beneficial to have a bigger search section which takes up the

top 25% of the screen. Visual and audio prompts could help identify the icons. A glowing animation around the mic would help bring attention to it. An audio prompt can be given if the participant has not taken an action after a period of time like 10 seconds. The prompt can ask them to type in the search bar or tap the mic icon to use speech input. If not, they can just proceed to type on the search bar.

Siri has an animation that shows when it is listening for input. Google has an animation as well but it is much subtler. P10 was unsure if Google was able to get his input or not because there was no feedback. A livelier recording animation would be helpful to resolve this problem.

Things that start automatically like videos, advertisements and pop-ups will move the attention away from the task at hand and must not be present in the interface in accessibility mode. The search bar and button should be large, easy to identify and in a prominent place. Using buttons with images instead of text will also help.

6 CONCLUSION

This study was undertaken to understand how people with intellectual disabilities can use speech technologies to learn and help them perform daily tasks independently. Although the participants encountered many different challenges, the majority of participants liked and were successful at using speech interfaces. The problems they encountered were different due to the varying levels of intellectual disability, age, motor abilities and digital literacy. One of the principal barriers to use was for people whose speech was not intelligible to VAI's due to the way that they spoke being inconsistent with the speech detection models. This barrier extends beyond people with ID to people with other speech idiosyncrasies that do not have ID. However, most of our participants with ID did not encounter difficulties in producing speech that was intelligible to a machine.

We distilled some common lessons for informing design of speech interfaces to make them more inclusive of people with ID. Many lessons relate to the graphical component of screen based interfaces that utilize speech. We suggest uniformity of icons and icon placement for future interfaces to reduce the strain on the working memory of the user. Some of the participants suggested highlighting the icons using animations. Reduction of other visual distractions is important (such as ads). Better feedback about the state of the device and whether it is listening or has stopped listening would also be very helpful. People with ID had conversational expectations of devices and often talk politely, as if talking to a human, not adjusting their speech as if talking to machine. If they pause, the conversational partner waits and listens, but machines often stop. Design of VAIs could examine how to accommodate this. Also people with ID can benefit by learning how to construct shorter queries.

The majority of participants preferred using their voice to complete the tasks than typing. The interviews with the participants revealed that 72% preferred using speech as input rather than typing. The main reasons for this were; it was a faster form of input than typing, they did not have to worry about spelling, it was possible to get audio replies from the voice

assistant and it helped them to do complicated tasks like set calendar events more easily than when typing.

We found that people with intellectual disabilities can greatly benefit by using speech technologies. During our observations participants were able to operate smart devices without touching them, access information from the internet in the form of videos and images and perform tasks like setting up calendar reminders. Voice assistants and voice search in search engines can help them complete tasks faster and with less help. Being less dependent on support workers and family members' is one of the most important things that this technology can do for people with an intellectual disability, even though this view is not always shared by proxies [4]. Currently people with ID often require someone else to type out search queries for them to watch a video or find something they are interested in on the internet. But by using VAs they can find what they were looking for, even if they have motor disabilities. People with ID spoke to the devices as if they would to a person and were delighted when they heard the device reply to them and even address them by their name. Many people with ID had a positive experience and said they would continue to use voice user interfaces in the future.

ACKNOWLEDGMENTS

This research is supported by the Australian Research Council under grant LP160100800, jointly with Endeavour Foundation. We acknowledge the time and dedication of our participants, as well as the staff of the sites we visited.

REFERENCES

- [1] Kevin M. Ayres, Linda Mechling, and Frank J Sansosti, F. 2013. The use of mobile technologies to assist with life skills/independence of students with moderate/severe intellectual disability and/or autism spectrum disorders: considerations for the future of school psychology. *Psychology In The Schools* 50, 3, 259-271. <http://dx.doi.org/10.1002/pits.21673>
- [2] Andrew A. Baylor, Filip Bircanin, Laurianne Sitbon, Bernd Ploderer and Margot Brereton. 2018. Characterizing Participation across Social Media Sites amongst Young Adults with Intellectual Disability, To appear *Proceedings of the 30th Australian Conference on Computer-Human Interaction*. ACM.
- [3] Virginia Braun and Victoria Clarke. 2006. Using thematic analysis in psychology. *Qualitative Research In Psychology* 3, 2, 77-101. doi: 10.1191/1478088706qp0630a
- [4] Margot Brereton, Laurianne Sitbon, Muhammad Abdullah, Mark Vanderberg and Stewart Koplick. 2015. Design after design to bridge between people living with cognitive or sensory impairments, their friends and proxies. *Codesign* 11, 1(Jan 2015), 4-20. doi: 10.1080/15710882.2015.1009471
- [5] Darren Chadwick, Caroline Wesson and Chris Fullwood. 2013. Internet Access by People with Intellectual Disabilities: Inequalities and Opportunities. *Future Internet* 5, 3, 376-397. <http://dx.doi.org/10.3390/fi5030376>
- [6] David Cihak, Rachel Wright, Cate Smith, Don McMahon and Kelly Kraiss. 2016. Incorporating functional digital literacy skills as part of the curriculum for high school students with intellectual disability. *Education And Training In Autism And Developmental Disabilities* 50, 2, 155-171.
- [7] James Collins and Lana Collet-Klingenberg. 2017. Portable electronic assistive technology to improve vocational task completion in young adults with an intellectual disability. *Journal Of Intellectual Disabilities* 22, 3, 213-231. <http://dx.doi.org/10.1177/1744629516689336>
- [8] MC Coppens-Hofman, H Terband, A.F.M Snik and B.A.M Maassen. 2016. Speech Characteristics and Intelligibility in Adults with Mild and Moderate Intellectual Disabilities. *Folia Phoniatrica Et Logopaedica* 68, 4, 175-182. doi: 10.1159/000450548
- [9] Definition. *Aaidd.org*. Retrieved 6 March 2018, from <http://aaidd.org/intellectual-disability/definition>
- [10] Jinjuan Feng, Jonathan Lazar, Libby Kumin and Ant Ozok. 2008. Computer Usage by Young Individuals with Down Syndrome: An Exploratory Study. In *Assets 08*, 35-42. Halifax, Nova Scotia: ACM.
- [11] Bjorn Harrysson, A. Svensk and G. Johansson. 2004. How people with developmental disabilities navigate the Internet. *British Journal Of Special Education* 31, 3, 138-142. <http://dx.doi.org/10.1111/j.0952-3383.2004.00344.x>
- [12] Brigitte Jordan and Austin Henderson. 1995. Interaction Analysis: Foundations and Practice. *Journal Of The Learning Sciences* 4, 1, 39-103. doi: 10.1207/s15327809jls0401_2
- [13] Soonil Kwon, Sung-Jae Kim and Joon Yeon Choeh. 2016. Preprocessing for elderly speech recognition of smart devices. *Computer Speech & Language* 36, 110-121. doi: 10.1016/j.csl.2015.09.002
- [14] Libby Kumin, Jonathan Lazar and Jinjuan Feng. 2012. Expanding Job Options: Potential Computer-Related Employment for Adults with Down Syndrome. *ACM SIGACCESS Accessibility And Computing* 103, 14-23. doi: 10.1145/2335867.2335869
- [15] Youngsun Lee, Michael L. Wehmeyer, Susan B. Palmer, Kendra Williams-Diehm, Daniel Davies and Steven E. Stock. 2010. The Effect of Student-Directed Transition Planning With a Computer-Based Reading Support Program on the Self-Determination of Students With Disabilities. *The Journal Of Special Education* 45, 2, 104-117. <http://dx.doi.org/10.1177/0022466909358916>
- [16] Dixie McCollum, Steven Nation and Sharon Gunn. 2014. The Effects of a Speech-to-Text Software Application on Written Expression for Students with Various Disabilities. *National Forum Of Special Education Journal* 25, 1, 1-13.
- [17] Asaf Michaely, Frank Zhang, Gabor Simko, Carolina Parada and Petar Aleksic. 2017. Keyword spotting for Google assistant using contextual speech recognition. *Automatic Speech Recognition and Understanding Workshop (ASRU)*, 2017 IEEE (pp. 272-278). IEEE. doi:10.1109/ASRU.2017.8268946
- [18] Redwhan Nour. 2015. Web searching by individuals with cognitive disabilities. *ACM SIGACCESS Accessibility and Computing* 111, 19-25
- [19] Thierry Olive. 2004. Working Memory in Writing: Empirical Evidence From the Dual-Task Technique. *European Psychologist* 9, 1, 32-42. <http://dx.doi.org/10.1027/1016-9040.9.1.32>
- [20] Susan B. Palmer, Michael L. Wehmeyer, Daniel Davies and Steven E. Stock. 2011. Family members' reports of the technology use of family members with intellectual and developmental disabilities. *Journal Of Intellectual Disability Research* 56, 4, 402-414. <http://dx.doi.org/10.1111/j.1365-2788.2011.01489.x>
- [21] Ravihansa Rajapakse, Margot Brereton and Laurianne Sitbon. 2018. Design Artefacts to Support People with a Disability to Build Personal Infrastructures. In *Proceedings of the 2018 on Designing Interactive Systems Conference 2018*, pp. 277-288. ACM, 2018.
- [22] Rocha, T., Bessa, M., Gonçalves, M., Cabral, L., Godinho, F., & Peres, E. 2012. The Recognition of Web Pages' Hyperlinks by People with Intellectual Disabilities: An Evaluation Study. *Journal Of Applied Research In Intellectual Disabilities* 25, 6, 542-552. <http://dx.doi.org/10.1111/j.1468-3148.2012.00700.x>
- [23] Tania Rocha, Diana Carvalho, Maximino Bessa, Sofia Reis and Luis Magalhães. 2016. Usability evaluation of navigation tasks by people with intellectual disabilities: a Google and SAPO comparative study regarding different interaction modalities. *Universal Access In The Information Society* 16, 3, 581-592. <http://dx.doi.org/10.1007/s10209-016-0489-5>
- [24] Laurianne Sitbon, Andrew A. Baylor, Filip Bircanin, Margot Brereton and Stewart Koplick. 2018. An Exploration of How People with Intellectual Disability Engage with Online Information Retrieval. In *CHI'18. Montreal: In Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems*, p. LBW536. ACM, 2018.
- [25] Laurianne Sitbon, Maria Hoogstrate, Stewart Koplick, Filip Bircanin and Margot Brereton M. 2018. A non-clinical approach to describing participants with intellectual disability, To appear *Proceedings of the 30th Australian Conference on Computer-Human Interaction*. ACM.
- [26] Jennifer J. Thistle and Krista M. Wilkinson. 2013. Working Memory Demands of Aided Augmentative and Alternative Communication for

Individuals with Developmental Disabilities. *Augmentative And Alternative Communication* 29, 3, 235-245.

<http://dx.doi.org/10.3109/07434618.2013.815800>

- [27] Thomas Welton, David Brown, Lindsay Evett and Nasser Sherkat. 2014. A brain-computer interface for the Dasher alternative text entry system. *Universal Access In The Information Society* 15, 1, 77-83. <http://dx.doi.org/10.1007/s10209-014-0375-y>
- [28] Cara Wilson, Laurianne Sitbon, Margot Brereton, Daniel Johnson and Stewart Koplick. 2016. Put yourself in the picture': designing for futures with young adults with intellectual disability. In *Proceedings of the 28th Australian Conference on Computer-Human Interaction* (pp. 271-281). ACM.
- [29] Alex Wong, Chetwyn Chan, Cecilia Li-Tsang, C and Chow Lam. 2009. Competence of people with intellectual disabilities on using human-computer interface. *Research In Developmental Disabilities* 30, 1, 107-123. <http://dx.doi.org/10.1016/j.ridd.2008.01.002>