Exploring Smart Speaker User Experience for People Who Stammer

ANNA BLEAKLEY, University College Dublin, Ireland
DANIEL ROUGH, University of Dundee, Scotland, UK
ABI ROPER, City, University of London, England, UK
STEPHEN LINDSAY, University of Glagow, Scotland, UK
MARTIN PORCHERON, Swansea University, Wales, UK
MINHA LEE, Eindhoven University of Technology, Netherlands
STUART NICHOLSON, Dyson Institute of Engineering and Technology, England, UK
BENJAMIN COWAN, University College Dublin, Ireland
LEIGH CLARK, Swansea University, Wales, UK

Speech-enabled smart speakers are common devices used for numerous tasks in everyday life. While speech-enabled technologies are widespread, using one's voice as a computing modality introduces new accessibility challenges for people with speech disfluencies such as stammering (also known as stuttering). This paper investigates the smart speaker user experiences of people who stammer over three weeks. We conducted diary studies and semi-structured interviews with 11 individuals to identify their daily routines, difficulties with successful interactions, and strategies to overcome these barriers. Our analysis demonstrates key factors such as device location, its affordances, and the structure of commands had a strong impact on user experience. Participants highlighted different linguistic strategies to try and overcome interaction difficulties and discussed the potential of using smart speakers for speech and language therapy. We emphasise the need to further understand the experiences of people who stammer in smart speaker design to increase their accessibility.

CCS Concepts: • **Human-centered computing → Empirical studies in accessibility**; **User studies**; **Empirical studies in HCI**.

Additional Key Words and Phrases: stammer, stutter, inclusivity, accessibility, speech interface, voice interface, conversational user interface

ACM Reference Format:
Anna Bleakley, Daniel Rough, Abi Roper, Stephen Lindsay, Martin Porcheron, Minha Lee, Stuart Nicholson, Benjamin Cowan, and Leigh Clark. 2022. Exploring Smart Speaker User Experience for People Who Stammer. In *The 24th International ACM SIGACCESS Conference on Computers and Accessibility (ASSETS '22), October 23–26, 2022, Athens, Greece.* ACM, New York, NY, USA, 17 pages. https://doi.org/10.1145/3517428.3544823

1 INTRODUCTION

The rapid evolution of speech technology has led to increasing use of speech as a modality to engage with applications through devices such as smart speakers, smartphone voice assistants, and wearables [45]. Smart speakers in particular have become a popular way of accessing applications and services, being owned by over 95 million people in the US alone, with 50% of device owners stating that they use these devices on a daily basis [28]. There are concerns about how users with diverse speech patterns like stammering may successfully interact with speech-enabled devices [11]. These users may find interaction with voice-only systems such as smart speakers challenging in their current form, and at worst they may be potentially excluded from smart speaker use altogether. Recent research efforts have highlighted that these technologies create accessibility challenges for others, including older adult users [43], those who are deaf or hard of hearing [7], and people who are blind or have visual impairments [2, 3].

Stammering (also known as stuttering) is classified as a neurological condition that impacts the rhythmic flow of speech [39]. Disruptions to speech can include repetition, prolongation or blocking of specific sounds and words. Blocking describes audible or silent moments when people are unable to produce a specific sound or word despite intending to produce them [34]. Recent estimations suggest 8% of children and 2% of adults experience some form of stammer [53]. Given the global increase of smart speaker interactions and speech interface use more generally, there is a risk that people who stammer may not have successful interactions with these devices, and their experiences are not commonly included in the design and testing of such interactions [36]. While technical developments in improving speech recognition are ongoing (e.g. [31, 49]), an understanding of how people who stammer engage with these devices is still limited [11].

To address this research gap, this paper explores how 11 people who stammer interacted with commercially available smart speakers over a three-week period in their homes. Through diary studies and semi-structured interviews, we investigated: 1) how smart speakers were used and implemented into daily lives and routines, and 2) the opportunities and challenges that emerged in these interactions. Our findings show common themes in participants regularly using devices as part of their routines, yet these were hindered by the presence of other people during interactions leading to some participants preferring private smart speaker use. The study also highlights the potential of using smart speakers to practice difficult speech patterns in a controlled environment and the possibility of using these devices to access speech and language therapy. However, our findings also showed a number of participants had interaction difficulties with specific sounds and words (including device wake words), and a lack of appropriate error recovery measures. These were compounded by fatigue, anxiety, and past negative experiences. Based on these findings, we emphasise the importance of examining speech interface experiences for people who stammer, and including them in the design process, to complement technical developments.

2 RELATED WORK

2.1 Speech as an Accessible Modality

Prior research on speech as a modality in human-computer interaction (HCI) has demonstrated its benefits in improving access to interactive systems [12]. Through speech, people with limited hand dexterity can gain hands-free access to mobile devices [13], and users with motor difficulties can employ voice as an alternative to mice or touch-pad cursor control[22]. Speech interfaces, in particular those that produce synthetic speech output, can also support people who cannot speak or have temporarily lost the ability to speak, affording them the opportunity to engage in conversation with others through assistive dialogue interfaces [4]. Additionally, speech interfaces have also been shown to offer support to deaf users to privately engage with medical professionals when in the presence of an interpreter [40].

Speech is a critical access modality for blind and low vision people. A review of speech technologies demonstrated the wide array of speech-enabled devices that can be used to improve accessibility for healthcare, educational tools, communication, and daily living [17]. Speech-enabled screen readers such as Jobs Access With Speech (JAWS) [51] are publicly available for people including those who are blind or have low vision, although there is current debate about whether these meet necessary standards for web content accessibility [58]. Speech commands have also been used to help improve the accessibility of using visual content like images and emojis [61].

While speech as a modality allows access to technology previously constrained to motor input devices, improving the accessibility of speech interfaces themselves is a burgeoning research area. Human-centered design (HCD) approaches [16] have been implemented to examine how people who are deaf or hard of hearing can better access and use smart speakers [7]. Additional visual modalities, such as the use of light and customisation of smart assistant voices to match individual needs can improve accessibility. HCD approaches have also been used by Glasser et al. [19] and Mande et al. [35] to understand the challenges and preferences of smart speaker use by those who are deaf and hard-of-hearing (DHH). Their work highlights preferences for waking up devices, as well as the potential for sign language interaction, but also notes the very limited regular use, or even attempts to use, smart speakers by DHH users.

Work on older adults' use of speech interfaces also identifies ongoing challenges and barriers to interaction, how the systems should speak and the potential benefits and drawbacks of anthropomorphism [50]. While the spoken nature of interactions may initially lower the barrier to using technology more generally for older adults [43], there remain concerns about the reliability of both the device and the information it provides. Despite the ongoing challenges, existing work highlights the importance of understanding people's needs and wishes as a means of improving their interactions with speech interfaces.

2.2 Stammering and Speech Interfaces

While work focusing on the wider user experience is scarce, current work on speech impairments and speech interfaces, such as stammering or dysarthria [37], have presented technical improvements in the use of automatic speech recognition (ASR). Progress has been made by the development of appropriate datasets. For instance, The SEP-28k dataset [31] has been shown to improve stammering event detection performance by tagging speech events of blocks, prolongations, sound repetitions, word repetitions, and interjections. Other approaches include University College London's Archive of Stuttered Speech (UCLASS) [25], the LibriStutter dataset of synthesised stuttered speech[29], and the FluencyBank [44]. While improvements to speech recognition are observed when using these datasets, they are still relatively small [52], and do not include those systems that are not made publicly or commercially available [31].

Adaptations for diverse speech patterns like stammering have started to emerge in commercial smart speakers. For example, the Amazon Alexa application has recently introduced a feature to enable longer waiting times for Alexa when people are speaking [6]. Google have also introduced a communication tool to improve speech comprehension for a range of abilities [10], albeit in a beta testing stage.

While advancements in stammering datasets and speech interface design are promising, they are often restricted to closed-source systems that limit the wider coverage of accessibility they can provide [11]. More significantly, there is a lack of research where people who stammer are included in the design and testing process of technology that can support them [36]. As such, there is an opportunity to draw inspiration from existing HCD approaches, in exploring how people who stammer use speech interfaces through devices such as smart speakers and how they may be improved to better accommodate such users.

2.3 Research Questions

Following similar research on home-based smart speaker deployments [42], this study deployed smart speakers in the homes of people who stammer, with an aim to answer two questions:

(1) How do people who stammer interact with smart speakers in their daily routines?
(2) What opportunities and challenges do smart speakers provide for people who stammer?

In doing so, our goal was to create a foundation of knowledge in an under-explored demographic and present actionable insights for the wider research and design community. We achieved this through qualitative analysis using a combination of diary studies and semi-structured interviews.

3 METHOD

3.1 Recruitment and Participants

12 participants were recruited to participate in the study through the STAMMA charity based in the United Kingdom, via social media channels (Facebook, Twitter), emails and the STAMMA charity website. After an introductory session, one participant dropped out of the study, leaving 11 participants (N=11, M=7, W=4; Mean age=33 yrs; SD=10.3 yrs). Of these, 5 were daily users of smart speakers, 1 used them multiple times a week, 2 used them once a month, 1 had only used them once or twice, and the remaining 2 had no prior usage. 64% (N=7) of participants in the study owned a smart speaker, with most (N=5) owning an Amazon Echo device. When asked about previous experiences using smart speakers, almost all participants (N=9) mentioned difficulties interacting with these devices. Stammering severity was not collected as this is typically done using clinical instruments [47].

3.2 Diary & Optional Interview

For our home-based deployment, we asked participants to undertake a diary study in two phases, lasting 11 days and 10 days respectively. The diary study method is an established method in HCI research, being a data gathering practice that prioritises daily logs from participants' perspectives [46].

In the first phase of the diary study (D1; Mean entries: 8; SD: 2.23 entries) participants were asked the following questions to reflect on their barriers to interaction:

1) What experiences did you have interacting with your smart speaker (Google Nest) today?

2) What difficulties have you had interacting with your smart speaker?

3) How did you overcome these difficulties?

For the second phase (D2; Mean entries: 6.8; SD: 5 entries), questions were tailored to understand how these devices were now integrated into participants' daily lives:

1) How were your interactions with your Google Nest today? What did you find easy and what did you find difficult?

2) If you had difficulties, what aspects of your speech might have affected this?

3) Have you used new strategies to overcome difficulties interacting with the Google Nest? If you have, please elaborate on this below.

For both phases, participants were encouraged to log at least one diary entry a day in a Microsoft Form[1] that was sent to them via email between 8:00 AM and 10:00 AM. Midway through the diary study (day 11), a set of unmoderated tasks (UT; N=10) were also introduced as a probe to delineate differences between different task interactions. Participants were asked to complete three tasks: 1) add an event to their calendar; 2) participate in an interactive game of their choice (e.g. quiz); and 3) add items to a shopping list.

[1]https://forms.office.com/

At the end of phase two of the diary study, an optional reflective interview invited participants to expand on their experiences. In these interviews, participants were asked about their overall experiences, difficulties and strategies when interacting with the smart speaker, and any improvements they believed could be implemented in future smart speaker design. Questions were tailored to individual participants' activities they had logged during the diary studies. Those that participated (either in written form or verbally) in this Reflective Interview (RF; N=9) demonstrated a mix of cumulative negative and positive responses and they additionally received a €25 honorarium.

3.3 Procedure

Th study received ethical clearance from Swansea University's ethics board. As mentioned, participants were recruited through a charity based in the United Kingdom as well as via social media channels (Facebook, Twitter), emails and the charity website. The study was conducted between the months of July - September 2021. Upon signing up for the study, and before the diary study phases commenced, participants received a Google Nest Mini [21] to install in their home. One participant expressed extreme difficulties in speaking, and so received a Google Nest Hub [20] that allows for tactile and gestural input along with visual output via a display. The researchers made this decision to ensure a more inclusive and diverse set of experiences could be captured.

Upon commencing phase one of the diary study, participants were given a 20-minute introductory information session via Zoom, giving participants the opportunity to ask questions and allowing the researchers to further explain any aspects that were unclear about the study structure and tasks. In this introductory session, participants were provided any help and advice they required to set up their smart speakers, and how to complete the diary studies. Beyond confirming the smart speakers were set up, actual interactions with the device were left to participants after this session. As audio recordings were not collected as part of this study, participants were also sent material detailing how to delete audio recordings the smart speakers stored online, if they wished to do so. During this session, participants also filled out a demographic questionnaire and answered questions relating to their past experiences with smart speakers. Participants had the option to answer these questions verbally or in written form. Only one participant declined to answer these questions but did participate in the diary studies (II; N=10). At the end of the study, participants were then thanked for taking part and were debriefed as to the aims of the work. They were then paid an honorarium for participation and were also given the Google Nest devices permanently as a thank you for taking part in the study. This brand was selected due to ongoing popularity and an expected largest smart speaker share by 2025 [27].

3.4 Data Analysis

After data collection was completed and optional interviews were transcribed by a member of the research team, the same researcher conducted inductive thematic analysis on the data [8] using NVivo 12.6.0 [32]. Codes were then exported to the Miro[2] platform, where an affinity diagram [24] was used to outline the initial themes along with their definitions. These themes were then discussed among members of the project team in relation to overlap and divergences between them and adjusted following reflection.

4 FINDINGS

We found three distinct themes in our collected study data which highlight the experience of smart speaker usage among people who stammer: 1) "Using Devices in Daily Routines & Speech Therapy Practice"; 2) "Interplay Between

[2] https://miro.com/

Contextual and Technological Barriers"; 3) "Adapting to Device Limitations". These themes are discussed in the following subsections.

4.1 Using Devices in Daily Routines & Speech Therapy Practice

In this section, we discuss how our participants integrated the Google Nest devices in their day-to-day lives, including how participants envisioned the use of these devices to improve their stammer in professional speech therapy contexts. The different activities undertaken by participants and their frequency can be seen in Table 1. Such usage is closely in line with recent work on typical behaviour of smart speaker users, both generally [5], with older users [26] and with DHH users [7]."

Table 1. Participant activities with their devices and the frequency in which these activities were conducted.

Activity	Frequency
Information Queries	13%
Music	15%
Weather	11%
Time	8%
News	8%
Home Appliances	4%
Travel & Commuting Queries	4%
Miscellaneous	37%

4.1.1 Device Activity & Routines. Participants' Google Nest devices quickly integrated into their daily lives, assisting them in morning routines by allowing them to ask for the weather, listen to the news, and interact with home appliances.

> *"It got integrated into not so much my routine but my life in the sense that it became one of the main devices I would use to play music."* [P10 - RI]

The routines participants developed were influenced by the location of the device in their home. Participants that placed the device in their home office [P4, P12, P11, P2] integrated the Nest into their work routines - asking the device to assist with work-related queries, play music during breaks, and turn on the kettle to begin their morning routine:

> *"I used it through the day so like to play music while I'm working or on my lunch break or something. I also have to do like some maths calculation type things."* [P12 - RI]

> *"And we have a plug in the kitchen and the kettle on with working from home at the moment. So I generally would do that three times in the morning to make a cup of coffee because that's my general routine."* [P2 - RI]

Some participants used the device to set alarms or prepare for their commute by checking the weather; these behaviors were again work related.

> *"I use it to set the forecast in the morning and through the day, particularly if I was going into work that day because I work in [a capital city] so it's good to know, you know if I need a coat."* [P12 - RI]

> *"Helping me find the right time to leave with traffic, or whether I should take a coat or umbrella."* [P1 - RI]

Unsurprisingly, opportunities to interact with the smart speaker were determined by its proximity, which thus determined the routines it became part of. Interactions with devices placed in office environments complemented work routines in the home, whereas those in communal spaces such as kitchens assisted their users during the "leaving for work" routine.

4.1.2 Leveraging Device Affordances to Support Speech Therapy.

During the introductory interview, two participants mentioned that they intended to use the Google Nest in a *"therapeutic way"* [P10 - II] and *"to essentially practice and essentially encouraging [...] communicative experiences"* [P8 - II].

> *"going to be about just putting myself out there in terms of difficult sounds."* [P8 - II]

Exploring this theme in the reflective interviews, we found that the majority of participants considered smart speakers to be potentially beneficial to their speech therapy. P2 suggested their place in building speech confidence:

> *"if you were at a point where you were at your stammering journey where you wanted to get some confidence [...] it could ask you to say certain words that you know you find are difficult."* [P2 - RI]

P11 and P9 further suggested how smart speaker applications, informed by speech therapists and people who stammer, could be used to complement certain techniques taught in therapy clinics and reduce reliance on in-person sessions:

> *"I would actually love that. Especially in the [local medical service] it's really hard to have a speech language therapist."* [P11 - RI]

> *"there are techniques they'll teach us and it wouldn't be difficult to code that and a device to assist the person and because these devices are cheap as well it's good because anyone could get one."* [P9 - RI]

Indeed, interacting with the device provided opportunities for participants to practise in the home, a *"liminal space between talking to yourself and talking to a human being"* [P10-RI]. Participants described themselves as *"more anxious"* [P11 - RI] when talking to a person, citing that the device *"wouldn't judge"* - [P10 - RI], which in turn made them less likely to stammer. A further advantage of interacting in a controlled environment was avoiding eye contact, [P9-RI, P8-RI] - a pressing concern during in-person communication:

> *"active effort to maintain eye contact when I'm having a conversation in person, over the telephone that doesn't matter and obviously speaking to the device."* [P8 - RI]

Overall, participants found benefits in interacting with their devices, citing the opportunities for such interactions to improve aspects of their speech and practise techniques learned in therapy. The controlled space the device afforded in their homes allowed them to interact in contexts where certain barriers to in-person interaction could be lowered.

4.2 Interplay Between Contextual and Technological Barriers

Users' varying speech patterns meant that difficulties varied in segments of the interaction. They had to overcome barriers in their speech as well as learn the unwritten rules of interaction to create successful utterances for the device.

4.2.1 Social Pressure, Privacy & Device Locations.

Although the Nest allowed participants to easily engage with the device in a controlled environment, the presence of other people determined whether users could interact without pressure. Difficulties in interaction were attributed to other people in the room, as users stammered more on commands that were said fluently when others were not around. These events occurred in communal areas such as the kitchen or living rooms.

> *"like since before the study, I find it oddly easy to talk to assistant devices, especially when there aren't other people around."* [P10 - D2]

> *"with the speaker because it was in my room, I stammer less than when I'm by myself [...] we actually have a Home Pod in our kitchen, so every time I try to use that, I stammer a lot more even if I were to say the same command."* [P6 - RI]

> *"I don't stammer because of a device, it's because there's added pressure sort of urgency from the other people."* [P9 - RI]

This has resulted in some users avoiding using the device in communal spaces in the home, stating that it was *"less of a problem"* [P6 - RI] if the command was unsuccessful. Although privacy was related to the presence of other people, one participant expressed concerns about how the device collected data, and the potential impact on their stammer.

> *"I am a little apprehensive that the anxiety that I have around privacy is going to affect whether I stammer with the Google Nest."* [P10 - II]

4.2.2 Mood: Stress, Fatigue & Anxiety. Participants reported that moods such as stress [P12-D2, P8-RI, P5-D1, P4-D2], anxiety [P12-RI, P1-D2, P6-RI], and fatigue [P12-RI, P10-D1] negatively impacted their fluency when interacting with the device.

> *"when I was very tired, I definitely did stammer more with it."* [P10 - RI]

These factors often changed on a day-to-day basis as participants reported *"good speech days"* [P12-D1, P2-D1] where they are less likely to stammer frequently and *"bad or worse speech days"* [P12-RI] where their stammer (e.g. a block) would be more prominent. These were often clearly defined external factors, e.g., *"Stress at work making me block more."* [P4 - D2], but sometimes there was no discernible reason:

> *"So it can be specific things like if I'm stressed or if I'm tired or if I'm quite anxious or nervous maybe, but sometimes there isn't a particular reason."* [P12 - RI]

4.2.3 Interaction Length & Interactivity. Participants reported increased likelihood of stammering as utterance length increased; *"short and simple"* [P3 - D1] interactions were more successful. We further explored this in our diary study probe, where participants were assigned preset tasks such as a quiz to complete.

> *"That was easier because I could just answer A,B, and C. So I would just answer one word."* [P6 - RI]

Upon reflection, some participants reported that they stammered more when completing tasks in the diary study probe [P6 - RI, P10 - RI], mentioning that they often had to resort to repetition due to the *"longer phrases"* [P6 - RI] that were required. Others stated that shorter phrases and the specificity of words added additional *"pressure"* [P8-RI, P9-RI] that was not present in the other daily interactions they had with the device.

> *"It does add an element of pressure because you have to say certain words as [opposed to] on a day-to-day basis."* [P9 - RI]

4.2.4 Difficult Sounds & Words. Participants exhibited varied difficulties with different sounds and words, particularly the wake word that initiates interaction with the smart speakers. For example, one participant reported less difficulty saying "Hey Google" in comparison to "Alexa" due to the starting sound of each wake word [P2]. Conversely, another participant had increased difficulties due to the starting sounds for "h" and "g" [P12]. This was particularly problematic, as all commands began using the wake word.

> *"Wake word - getting the initial sound out to get Google to interact with me."* [P1 - D1]

These patterns also persisted in other words in a command. Differences between soft, hard sounds [P6 - RI], hard stop letters [P9 - D1], and harsh [P12 - RI] sounds were used to explain these difficulties.

> *"The letters O, K and G are hard starts and can be difficult to say."* [P9 - D1]

Difficulties relating to sounds also revolved around delays and prolongations of certain letters. Although devices would sometimes recognise users' requests in these contexts, this experience was inconsistent.

> *"It seems to recognise certain elongation of words e.g if I elongated "s" at the start of a word it mostly seems to understand the word. Whereas if I elongated "m" at the start of word, it wouldn't recognise the word."* [P12 - D1]

> *"I think I asked it again and I could say the 'b', but it did actually wait a while for me to do that."* [P6 - RI]

When commands to the device were not registered, users became frustrated, which created negative interaction experiences, in turn increasing the likelihood of the device not recognising them. Additional pressure was added as users repeated commands, analogous to interaction experiences with other people.

> *"Whatever when you're out and about and you actually think you've said that really well and you're really happy and you're like "yeah", but then the person will say like "sorry, I didn't hear that" [...] that kind of pressure when you have to say it again and it intensifies making it worse"* [P2 - RI]

Being forced to give up after such an increasingly negative series of interactions was described as a *"debilitating"* [P8-RI] experience.

4.2.5 Past Negative Experiences. Prior experiences with the Nest influenced users' propensity to stammer. As participants began their day with the device in the diary study, their first communicative experience often influenced subsequent interactions throughout the day. Moreover, past negative experiences with the Nest determined users' speech patterns, with some mentioning a loop wherein thoughts about past stammering-related interactions with the Google Nest would create difficulties on words they did not previously stammer on [P2-RI]. This happened as they prepared themselves for the lack of recognition from the device, expecting to be in a *" difficult situation"* [P10-RI].

> *"I think the main motivation was the awareness that I have a fairly severe speech problem and that I must do something to counteract it. On occasion this led to a feedback loop of enhanced awareness causing a more severe speech block."* [P3 - RI]

> *"If you're having those negative thoughts then it would lead to more block even on words that you wouldn't necessarily expect on an average day to block on."* [P2 - RI]

4.2.6 Timing Out. The device timing out was a key issue, with many participants mentioning that longer waiting times would have improved their interactions with the device. However, a few expressed that waiting times with the Nest were longer compared to other devices [P12 - D1, P9 - RI], creating positive interaction experiences. During unsuccessful interactions, users often struggled to continue due to device interruption as they felt additional pressure to produce utterances. Participants that reported blocking also mentioned the device timing out before the following utterance or the device ignoring a repetitive command after a block. Similarities were reported among those that elongated sounds (also mentioned in 4.2.4).

> *"When blocking, either no sound or a repetitive sound comes out so the nest doesn't always identify what you are saying or thinks you have finished speaking."* [P12 - D2]

> *"I get stuck saying OK Google as I end up saying "OOOOOOOOOOOOOOOOOOOK Google" but it stops listening during my prolonged "O"."* [P1 - RI]

Users stated that one of the main disadvantages of communicating with the devices is the lack of awareness it had to their stammer. Unlike human-human interactions where interlocutors would adjust their turns to accommodate their speaking patterns, devices would cut users off. Moreover, participants mentioned not being able to disclose their stammer, which provides a level of ease and comfort in human-human interactions not available with smart speakers.

> *"People don't stop listening when I can't get a word out, so just taking my time with people is fine."* [P1 - RI]

> *"if there's a way to detect that the user stammers [...] Then maybe it can recognize that or maybe give the user more time to actually finish their command to say people that don't stammer."* [P6 - RI]

4.2.7 Limited Scaffolding for Error Recovery. Recognition was one of the biggest barriers in interacting with the Google Nest. Often, the device did not assist with users' error recovery, leaving them unsure how to correct utterances - *"it just takes a long time to realize or know what the Nest wants"* [P10 - RI] and that they often *"wasn't/not sure why"* [P12 - D2, P9 - D1, P2 - D1] their initial command was not recognized:

> *"Mostly frustrating - Didn't recognise a request I made - had to ask it 3 times . Think fluency was the same each time so not sure why !."* [P2 - D1]

> *"It was just all of a sudden hit a wall and it no longer recognizing and I thought we were friends – I thought we were close."* [P8 - RI]

One participant [P10] explicated error recovery in detail, mentioning that as the device repeated the perceived utterance back to them, they would be able to understand strategies to repair segments of the phrase. This is in contrast to situations where the device would execute a command based on what it perceived the user said.

> *"if you asked it to - where's the best place to buy chips and it said "here's the best place to buy sheep", you would know that the word it didn't understand correctly was chips, you would make [...] some kind of effort to make the Nest understand that word the next time"* [P10 - RI]

4.3 Adapting to Device Limitations

As users tried to overcome these barriers, they deployed strategies both old and new. However, issues such as error recovery limited the strategies deployed in addition to defining the qualities of interaction among users who stammer.

4.3.1 Repetition. To overcome errors they encountered with the Nest devices, users would often repeat their commands, due to the lack of alternative error recovery methods available. Although at times repetition allowed them to eventually execute their commands successfully, when unsuccessful, users would not be able to delineate the cause of the error, due to lack of device feedback (see 4.2.7).

> *"Over time, I found ways to overcome my initial challenge of not being able to say the wake word, mainly through repetition..."* [P1 - RI]

> *"I suppose that's what happened when I put something in the calendar. I had to repeat it even though I originally thought it went well..."* [P2 - RI]

In these situations, participants either persisted until they were successful, or simply gave up using these difficult commands. Indeed, most users mentioned that their patterns of use remained unchanged, often sticking to the same commands for the duration of the diary study to avoid negative stammering experiences.

"That's the command I know I won't stammer and I know I can say." [P6 - RI]

"I am trying to be more aware of the sounds I struggle with as I do avoid then habitually now." [P8 - D1]

Repetition became characteristic of participants' interactions with the device. One participant explained how other household members had *"fewer interactions"* [P3 - RI] with the device, whereas they had difficulties *"being understood the first or even second time* [P3 - RI]. Other participants described *"persistence"* [P1 - D1,P5 - D1] and having *"persevered"* [P2 - D1] in relation to how they overcame these challenges. As users repeated commands to the device, they also employed methods to improve certain aspects of their speech. These are discussed in more detail in the following subsections.

4.3.2 Planning Utterances. As users reformulated utterances, they became more aware of the limitations in recognition that the Nest had. They often tried to opt for other words to overcome blocking or stammering, paralleling strategies used in in-person interaction.

"I think once I did actually I did have a way around that command. Saying "what events are there?" or something. Because one of these words are much easier for me to say, because events begins with 'e'." [P6 - RI]

However, at times, this paradoxically caused them to stammer, due to previously formed expectations.

"I thought about day 1 more, and I definitely had difficulty with thinking that I was about to stutter and planning for it in the way that meant I actually meant I ended up messing up the sentence and confusing the device...." [P10]

4.3.3 Controlled Speaking. The most commonly reported strategy [P1, P12, P3, P9] was speaking in a *"conscious"* [P3], or *"contrived"* [P1] manner. This meant that users were sometimes less relaxed speaking to the device than when speaking to another person, yet this strategy did not always result in successful interactions.

"Again, I tried to talk in a controlled and not very relaxed way. However it didn't always work..." [P1 - D1]

Speaking slowly allowed participants to fluently produce words that previously resulted in errors, and flow into words they previously stammered with [P12]. Additionally, controlling breathing or taking *"deep breaths"* [P1-D2] avoided issues such as timing out or device interruption.

"I just found that speaking at a much more measured pace never really gave it the opportunity to stop you [...]Once I kind of, not mastered that but once I implemented that it was much easier, much more successful" [P8-RI]

At times, participants also mentioned increasing the volume in which they spoke to ensure the device could pick up their commands, while also moving closer to the device to increase recognition [P10 - D1 & D2, P9 - D1 & D2]. This became difficult in less private and noisier environments in the home.

"It helps if I'm on my own when speaking to the device the quieter the room the better as noise can cause a direction and make speaking different....." [P9 - D1]

"repeating, or going closer to the speaker...." [P10 - D1]

5 DISCUSSION

Smart speakers and other intelligent personal assistants (IPAs) are becoming more ubiquitous with a strong presence in the home and other environments. Yet, users that stammer have not been explicitly accounted for in the design of these interfaces. In this study, we highlight three distinct themes that reflect the experiences of users that stammer. We build on previous work [11] highlighting the lack of accessibility in IPA use through smart speakers for users with diverse speech patterns.

5.1 Navigating Difficult Interactions

While participants had some success in interacting with the smart speakers, and even embedded them into their daily routines, there were frequent indications that the devices' design was insufficient to accommodate their stammering. This included an inability to initiate the wake word required to start interactions, difficulties with specific sounds or phrases, and the device timing out for participants, for example when blocking occurred.

Participants engaged in strategies in attempts to overcome difficulties with the devices understanding and correctly interpreting their speech. These included repetition, planning utterances, speaking in a more controlled manner, and adjusting pitch. However, participants encountered mixed results. Repeating commands, for example, would sometimes create additional pressure on producing speech, which mirrored some participants' experiences of talking to other people. Unsuccessful use of strategies that already attempted to navigate system recognition errors could leave participants wanting to avoid interacting with the devices again. Such challenges to regular adoption are evident in related work with DHH users [19, 35] who feel excluded from the speech-centred design process. First impressions are vital in the subsequent adoption of a technology, and we recommend that some future ASR efforts focus on maximising the success of those commands that are most frequently used in this and previous work.

While using and possibly promoting these strategies may be useful in developing more successful interactions, relying on them rather than being able to use one's own speech may be detrimental. Prior work has identified that openly stammering and being comfortable in speaking is a critical value that people who stammer would like to see recognised in the tools they use [9], and it is likely that smart speakers are no different.

Additionally, the variable success of the above strategies can be compounded by how people's stammers were manifesting at the point of interaction. Daily life stress, anxiety, and fatigue had an impact on people's speech, which in turn could make successful interactions with the smart speakers more problematic. Additionally, participants felt further pressure from the presence of others, resulting in less successful device interactions. Prior work has demonstrated the reluctance of people to engage with speech interfaces in public settings [14, 33], due to a sense of social embarrassment or awkwardness [1, 14]. Although groups such as families may engage collaboratively with smart speakers in home settings [41], the increased potential of stammering because of others being co-located with smart speakers may impact on this embedding of devices in collaborative interactions and daily routines.

5.2 Fostering Error Recovery

A lack of error recovery support in the design of smart speakers was a major issue for participants.One suggestion for improving error recovery was to provide feedback on what the device had interpreted from people's speech, which does not currently occur transparently. Similar findings were observed in studies with L2 English speakers and IPAs, in which using visual feedback (e.g. through smartphones) allowed people to understand the cause of communication breakdowns [60]. While this may be one solution, it is not always possible in hands-busy/eyes-busy scenarios where people do not have the ability to access a screen [11]. While device logs are available that indicate what a device has interpreted, these are not readily accessible because they require people to actively look up this information online via their personal accounts.

Similarly, error recovery may need to more clearly indicate why an error has occurred. It was not always clear to participants why difficulties were emerging. Participants were sometimes left wondering if something about their speech pattern at the time of interaction was creating difficulties, or whether it was the nature of the query itself. This echoes findings from [7] with DHH users who were unable to understand why an error had occurred. In addition to the visual 'training wheels' recommended in this previous work, we recommend a "what went wrong?" command that would allow any user to know why their previous command caused an error. This would further help to bridge the "gulf of expectation" [11] that is especially pertinent to users who stammer, who may consider their stammer a "speech error" [36] and incorrectly attribute the error to themselves, rather than the limitations of smart speakers.

While this lack of clarity is true irrespective of fluency [60], the participants in this study were more inclined to blame their own speech rather than the device's performance, resulting in negative self-perception. There is advice on how to talk to people who stammer (e.g. [54]), although it is unclear how this may feasibly be transferred to smart speaker interactions [11]. While improvements may need to draw on existing research on how to make smart speaker interactions progress [15], this may not address the inability to conduct the initial interactions for people who stammer. Improvements could also draw on web accessibility guidance, including creating additional time to speak (as introduced with Amazon Alexa [6]), capturing and re-using smaller parts of speech, and the use of additional modalities when appropriate [48].

5.3 The potential for smart speakers to support speech and language therapy

Our study highlights the opportunities that smart speakers afford people who stammer. Over the course of the study, some participants utilised their devices to practice speaking techniques and sounds they know they have difficulty with, to build confidence for future social interactions. The lack of perceived judgement and the nature of the controlled environment of interaction reduced speaking anxiety for some participants. Additionally, there was potential seen in using smart speakers as a means of delivering speech and language therapy. This echoes the prior use of smart speakers to deliver other forms of therapy or coaching in the past [23] and applied to help build confidence for public speaking [55]. Additionally, this resonates with speech development services like *withVR* [56] that provide customisable virtual reality speaking environments, where the speaker is in full control of the situation. Similar approaches may be adoptable in commercial smart speakers to provide cheaper and more readily available speaking practice. The potential for using voice assistant technologies is shared by recent work on how speech and language therapists (SLTs) use or want to use these devices as part of their practice [30]. The authors indicated some SLTs were already using commercial devices to help their clients with daily tasks and improving and practising speech. Smart speaker integration into daily routine and their use of voice aligns them with the emerging discipline of healthcare coaching, which often uses regular, short conversations to educate and motivate to deliver its interventions [38]. However, as discussed in [31], existing algorithms that "smooth over" stammering support ASR, but not detection of stammering events. Explicating such events is necessary for both meaningful error recovery and potential speech therapy uses (discussed below). We recommend a focus on improving the accuracy of event detection and development of smart speaker-specific datasets, collected through user-centred studies such as this, or building on the work of [29] by synthesising common stammering events in existing smart speaker datasets.

Further work should also focus on exploring the experiences and expectations of both SLTs and people who use their services with IPAs in this context. Such work could then inform professional standards on their use, as well as ways to improve the design of the technology to support speech and language therapy. By learning from the experiences of people who stammer, along with healthcare practitioners, we can develop more inclusive design guidelines for usage of smart speakers by people with specific communicative needs, and further explore their use in SLT practice. Indeed, work by authors of the StammerApp [36] shows support for real-time feedback on speech from users who stammer. While the use of augmented reality (AR) technologies such as Google Glass were suggested for this, we see clear potential for such feedback in smart speakers. We thus urge the third-party development of 'Skills' (Alexa) or 'Actions' (Google Assistant) that provide such feedback to assist in identifying areas for further practice that are not currently available.

5.4 Limitations and Future Work

This paper identifies some of the opportunities and challenges that people who stammer face when interacting with commercially available smart speakers in their homes over a three-week period. We emphasise that our findings are primarily limited to experiences of speech only smart speakers. Future research should seek to examine how people's experiences differ when using other speech-capable devices in and outside of the home environment, as well as how multi-modal forms of smart speakers impact the findings presented. Additionally, further work would benefit from exploring how people who stammer in other languages interact with smart speakers, and how stammering may also interact non-L1 English speakers' experiences [57, 59]. While the participants in our study were all adults, a higher number of children experience some form of stammer [53]. Further research in the developing area of younger users of smart speakers [18] may be required to create a more comprehensive insight into the experiences of people who stammer, alongside other diverse speech patterns. Further quantitative research would also create additional insights by understanding how many people who stammer own and use smart speakers, and how frequently recognition errors emerge when interacting with them.

6 CONCLUSION

In this paper, we examined the use of smart speakers in the home by people who stammer though three-week diary studies and semi-structured interviews. We undertook our work to focus on how we could explore the user experience of these technologies and make them more inclusive. Our participants used smart speakers for a range of features, placing the devices in different places within the home and accordingly using them as part of different day-to-day routines. Although our participants found benefits of using their devices and highlighted the potential of using them for speech and language therapy, external social factors such as the co-presence of others negatively impacted their experiences. Crucial for considering smart speaker design, our findings identified how the length and characteristics of interaction (including specific sounds), timeouts, and insufficient scaffolding for error recovery, impeded device use which our participants overcame through greater planning, repetition, and consciously controlling their speaking. Stammering is a diverse phenomenon that impacts people differently, thus while we propose improvements for smart speaker devices, these must be applied carefully to avoid exclusion or negative side effects.

ACKNOWLEDGMENTS

We would like to thank STAMMA for their help in developing this study. This work was supported by the Engineering and Physical Sciences Research Council grant EP/M022722/1.

REFERENCES

[1] Ali Abdolrahmani, Ravi Kuber, and Stacy M Branham. 2018. " Siri Talks at You" An Empirical Investigation of Voice-Activated Personal Assistant (VAPA) Usage by Individuals Who Are Blind. In *Proceedings of the 20th International ACM SIGACCESS Conference on Computers and Accessibility*. 249–258.

[2] Ali Abdolrahmani, Kevin M Storer, Antony Rishin Mukkath Roy, Ravi Kuber, and Stacy M Branham. 2020. Blind leading the sighted: drawing design insights from blind users towards more productivity-oriented voice interfaces. *ACM Transactions on Accessible Computing (TACCESS)* 12, 4 (2020), 1–35.

[3] Dragan Ahmetovic, Cole Gleason, Chengxiong Ruan, Kris Kitani, Hironobu Takagi, and Chieko Asakawa. 2016. NavCog: a navigational cognitive assistant for the blind. In *Proceedings of the 18th International Conference on Human-Computer Interaction with Mobile Devices and Services*. 90–99.

[4] Norman Alm, John Todman, Leona Elder, and Alan F Newell. 1993. Computer aided conversation for severely physically impaired non-speaking people. In *Proceedings of the INTERACT'93 and CHI'93 Conference on Human Factors in Computing Systems*. 236–241.

[5] Tawfiq Ammari, Jofish Kaye, Janice Y Tsai, and Frank Bentley. 2019. Music, Search, and IoT: How People (Really) Use Voice Assistants. *ACM Trans. Comput. Hum. Interact.* 26, 3 (2019), 17–1.

[6] Steven Aquino. 2021. Exclusive: Amazon Adds New Speech Setting To Alexa App To Help Stutterers Finish Commands, Queries. https://www.forbes.com/sites/stevenaquino/2021/10/05/exclusive-amazon-adds-new-speech-setting-to-alexa-app-to-help-stutterers-finish-commands-queries/ Accessed 2nd April 2022.

[7] Johnna Blair and Saeed Abdullah. 2020. It Didn't Sound Good with My Cochlear Implants: Understanding the Challenges of Using Smart Assistants for Deaf and Hard of Hearing Users. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 4, 4 (2020), 1–27.

[8] Virginia Braun and Victoria Clarke. 2006. Using thematic analysis in psychology. *Qualitative research in psychology* 3, 2 (2006), 77–101.

[9] Robin N Brewer and Vaishnav Kameswaran. 2018. Understanding the power of control in autonomous vehicles for people with vision impairment. In *Proceedings of the 20th International ACM SIGACCESS Conference on Computers and Accessibility*. 185–197.

[10] Julie Cattiau. 2021. A communication tool for people with speech impairments. https://blog.google/outreach-initiatives/accessibility/project-relate/ Accessed 2nd April 2022.

[11] Leigh Clark, Benjamin R Cowan, Abi Roper, Stephen Lindsay, and Owen Sheers. 2020. Speech diversity and speech interfaces: Considering an inclusive future through stammering. In *Proceedings of the 2nd Conference on Conversational User Interfaces*. 1–3.

[12] Leigh Clark, Philip Doyle, Diego Garaialde, Emer Gilmartin, Stephan Schlögl, Jens Edlund, Matthew Aylett, João Cabral, Cosmin Munteanu, Justin Edwards, and Benjamin R Cowan. 2019. The State of Speech in HCI: Trends, Themes and Challenges. *Interacting with Computers* 31, 4 (09 2019), 349–371. https://doi.org/10.1093/iwc/iwz016 arXiv:https://academic.oup.com/iwc/article-pdf/31/4/349/33525046/iwz016.pdf

[13] Eric Corbett and Astrid Weber. 2016. What can I say? addressing user experience challenges of a mobile voice user interface for accessibility. In *Proceedings of the 18th international conference on human-computer interaction with mobile devices and services*. 72–82.

[14] Benjamin R. Cowan, Nadia Pantidi, David Coyle, Kellie Morrissey, Peter Clarke, Sara Al-Shehri, David Earley, and Natasha Bandeira. 2017. "What Can i Help You with?": Infrequent Users' Experiences of Intelligent Personal Assistants. In *Proceedings of the 19th International Conference on Human-Computer Interaction with Mobile Devices and Services* (Vienna, Austria) *(MobileHCI '17)*. Association for Computing Machinery, New York, NY, USA, Article 43, 12 pages. https://doi.org/10.1145/3098279.3098539

[15] Joel E. Fischer, Stuart Reeves, Martin Porcheron, and Rein Ove Sikveland. 2019. Progressivity for Voice Interface Design. In *Proceedings of the 1st International Conference on Conversational User Interfaces* (Dublin, Ireland) *(CUI '19)*. Association for Computing Machinery, New York, NY, USA, Article 26, 8 pages. https://doi.org/10.1145/3342775.3342788

[16] International Organization for Standardization. 2010. *Ergonomics of Human-system Interaction: Part 210: Human-centred Design for Interactive Systems*. ISO. https://www.iso.org/standard/77520.html

[17] Diamantino Freitas and Georgios Kouroupetroglou. 2008. Speech technologies for blind and low vision persons. *Technology and Disability* 20, 2 (2008), 135–156.

[18] Radhika Garg, Hua Cui, Bo Zhang, Spencer Selingson, Martin Porcheron, Leigh Clark, Benjamin R Cowan, and Erin Beneteau. 2022. The Last Decade of HCI Research on Children and Voice-based Conversational Agents. In *CHI Conference on Human Factors in Computing Systems* (New Orleans, LA, USA) *(CHI '22)*. Association for Computing Machinery, New York, NY, USA, 19 pages. https://doi.org/10.1145/3491102.3502016

[19] Abraham Glasser, Vaishnavi Mande, and Matt Huenerfauth. 2021. Understanding deaf and hard-of-hearing users' interest in sign-language interaction with personal-assistant devices. In *Proceedings of the 18th International Web for All Conference*. 1–11.

[20] Google. 2022. Google Nest Hub (2nd Gen). https://store.google.com/gb/product/nest_hub_2nd_gen?hl=en-GB Accessed 4th April 2022.

[21] Google. 2022. Google Nest Mini - A Smart Speaker for Any Room. https://store.google.com/gb/product/google_nest_mini?pli=1&hl=en-GB Accessed 4th April 2022.

[22] Susumu Harada, Jacob O Wobbrock, Jonathan Malkin, Jeff A Bilmes, and James A Landay. 2009. Longitudinal study of people learning to use continuous voice-based cursor control. In *Proceedings of the SIGCHI conference on Human Factors in Computing Systems*. 347–356.

[23] Ahmed Hassoon, Jennifer Schrack, Daniel Naiman, Dina Lansey, Yasmin Baig, Vered Stearns, David Celentano, Seth Martin, Lawrence Appel, et al. 2018. Increasing physical activity amongst overweight and obese cancer survivors using an alexa-based intelligent agent for patient coaching: protocol for the physical activity by technology help (PATH) trial. *JMIR research protocols* 7, 2 (2018), e9096.

[24] Karen Holtzblatt and Hugh Beyer. 1997. *Contextual design: defining customer-centered systems*. Elsevier.

[25] Peter Howell, Stephen Davis, and Jon Bartrip. 2009. The University College London Archive of Stuttered Speech (UCLASS). *Journal of Speech, Language, and Hearing Research* 52, 2 (2009), 556.

[26] Sunyoung Kim and Abhishek Choudhury. 2021. Exploring older adults' perception and use of smart speaker-based voice assistants: A longitudinal study. *Computers in Human Behavior* 124 (2021), 106914.

[27] Brett Kinsella. 2019. Loup Ventures Says 75% of U.S. Households Will Have Smart Speakers by 2025, Google to Surpass Amazon in Market Share. https://voicebot.ai/2019/06/18/loup-ventures-says-75-of-u-s-households-will-have-smart-speakers-by-2025-google-to-surpass-amazon-in-market-share

[28] Brett Kinsella. 2022. The Rise and Stall of the U.S. Smart Speaker Market – New Report. https://voicebot.ai/2022/03/02/the-rise-and-stall-of-the-u-s-smart-speaker-market-new-report/

[29] Tedd Kourkounakis, Amirhossein Hajavi, and Ali Etemad. 2020. FluentNet: end-to-end detection of speech disfluency with deep learning. *arXiv preprint arXiv:2009.11394* (2020).

[30] Pranav Kulkarni, Orla Duffy, Jonathan Synnott, W George Kernohan, Roisin McNaney, et al. 2022. Speech and Language Practitioners' Experiences of Commercially Available Voice-Assisted Technology: Web-Based Survey Study. *JMIR Rehabilitation and Assistive Technologies* 9, 1 (2022), e29249.

[31] Colin Lea, Vikramjit Mitra, Aparna Joshi, Sachin Kajarekar, and Jeffrey P Bigham. 2021. Sep-28k: A dataset for stuttering event detection from podcasts with people who stutter. In *ICASSP 2021-2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 6798–6802.

[32] QSR International Pty Ltd. 2018. Nvivo (Version 12). https://www.qsrinternational.com/nvivo-qualitative-data-analysis-software/try-nvivo

[33] Ewa Luger and Abigail Sellen. 2016. "Like Having a Really Bad PA": The Gulf between User Expectation and Experience of Conversational Agents. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, New York, NY, USA, 5286–5297. https://doi.org/10.1145/2858036.2858288

[34] Gerald A Maguire, Christopher Y Yeh, and Brandon S Ito. 2012. Overview of the diagnosis and treatment of stuttering. *Journal of Experimental & Clinical Medicine* 4, 2 (2012), 92–97.

[35] Vaishnavi Mande, Abraham Glasser, Becca Dingman, and Matt Huenerfauth. 2021. Deaf Users' Preferences Among Wake-Up Approaches during Sign-Language Interaction with Personal Assistant Devices. In *Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems*. 1–6.

[36] Roisin McNaney, Christopher Bull, Lynne Mackie, Floriane Dahman, Helen Stringer, Dan Richardson, and Daniel Welsh. 2018. StammerApp: Designing a Mobile Application to Support Self-Reflection and Goal Setting for People Who Stammer. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. 1–12.

[37] NHS. 2019. Dysarthria (difficulty speaking). https://www.nhs.uk/conditions/dysarthria/ Accessed 2nd April 2022.

[38] Jeanette M Olsen and Bonnie J Nesbitt. 2010. Health coaching to improve healthy lifestyle behaviors: an integrative review. *American Journal of Health Promotion* 25, 1 (2010), e1–e12.

[39] World Health Organization. 2010. ICD-10 Version:2010. http://apps.who.int/classifications/icd10/browse/2010/en#/F98.5 Accessed 23rd Feb 2020.

[40] Anne Marie Piper and James D Hollan. 2008. Supporting medical conversations between deaf and hearing individuals with tabletop displays. In *Proceedings of the 2008 ACM conference on Computer supported cooperative work*. 147–156.

[41] Martin Porcheron, Joel E Fischer, Stuart Reeves, and Sarah Sharples. 2018. Voice Interfaces in Everyday Life. In *Proceedings of the 2018 ACM Conference on Human Factors in Computing Systems (CHI '18)*. Association for Computing Machinery, New York, NY, USA, Article 640, 12 pages. https://doi.org/10.1145/3173574.3174214

[42] Alisha Pradhan, Leah Findlater, and Amanda Lazar. 2019. " Phantom Friend" or" Just a Box with Information" Personification and Ontological Categorization of Smart Speaker-based Voice Assistants by Older Adults. *Proceedings of the ACM on Human-Computer Interaction* 3, CSCW (2019), 1–21.

[43] Alisha Pradhan, Amanda Lazar, and Leah Findlater. 2020. Use of intelligent voice assistants by older adults with low technology use. *ACM Transactions on Computer-Human Interaction (TOCHI)* 27, 4 (2020), 1–27.

[44] Nan Bernstein Ratner and Brian MacWhinney. 2018. Fluency Bank: A new resource for fluency research and practice. *Journal of fluency disorders* 56 (2018), 69–80.

[45] Juniper Research. 2018. Digital Voice Assistants in Use to Triple to 8 Billion by 2023, Driven by Smart Home Devices. shorturl.at/dgoGL. Accessed 22nd Feb 2020.

[46] John Rieman. 1993. The diary study: a workplace-oriented research tool to guide laboratory efforts. In *Proceedings of the INTERACT'93 and CHI'93 conference on Human factors in computing systems*. 321–326.

[47] Glyndon Riley and Klaas Bakker. 2009. *SSI-4: Stuttering severity instrument*. PRO-ED, an International Publisher.

[48] Abi Roper, Stephanie Wilson, Timothy Neate, and Jane Marshall. 2019. Speech and Language. In *Web Accessibility*. Springer, 121–131.

[49] Frank Rudzicz, Aravind Kumar Namasivayam, and Talya Wolff. 2012. The TORGO database of acoustic and articulatory speech from speakers with dysarthria. *Language Resources and Evaluation* 46, 4 (2012), 523–541.

[50] Sergio Sayago, Barbara Barbosa Neves, and Benjamin R Cowan. 2019. Voice assistants and older people: some open issues. In *Proceedings of the 1st International Conference on Conversational User Interfaces*. 1–3.

[51] Freedom Scientific. 2022. Jaws - Freedom Scientific. https://www.freedomscientific.com/products/software/jaws Accessed 23rd Feb 2022.

[52] Shakeel Ahmad Sheikh, Md Sahidullah, Fabrice Hirsch, and Slim Ouni. 2021. Machine Learning for Stuttering Identification: Review, Challenges & Future Directions. *arXiv preprint arXiv:2107.04057* (2021).

[53] Stamma. 2020. Stammer in the Population. https://stamma.org/news-features/stammering-population Accessed 23rd Feb 2020.

[54] Stamma. 2020. Talking With Someone Who Stammers. https://stamma.org/about-stammering/talking-someone-who-stammers Accessed 23rd Feb 2020.

[55] Jinping Wang, Hyun Yang, Ruosi Shao, Saeed Abdullah, and S Shyam Sundar. 2020. Alexa as coach: Leveraging smart speakers to build social agents that reduce public speaking anxiety. In *Proceedings of the 2020 CHI conference on human factors in computing systems*. 1–13.

[56] withVR. 2022. withVR - A safe space to speak. https://withvr.app Accessed 2nd April 2022.

[57] Yunhan Wu, Justin Edwards, Orla Cooney, Anna Bleakley, Philip R Doyle, Leigh Clark, Daniel Rough, and Benjamin R Cowan. 2020. Mental workload and language production in non-native speaker IPA interaction. In *Proceedings of the 2nd Conference on Conversational User Interfaces*. 1–8.

[58] Yan Wu, Stephen Lindsay, Leighton Evans, Martin Porcheron, Leigh Clark, and Rhys Jones. 2021. *Enabling Digital First: A Case Study of Sight-Impaired Users in Wales.* Technical Report. Swansea University, Swansea, UK. https://cronfa.swan.ac.uk/Record/cronfa58304/

[59] Yunhan Wu, Martin Porcheron, Philip Doyle, Justin Edwards, Daniel Rough, Orla Cooney, Anna Bleakley, Leigh Clark, and Benjamin R. Cowan. 2022. Comparing Command Construction in Native and Non-Native Speaker IPA Interaction through Conversation Analysis. In *Proceedings of the 4th International Conference on Conversational User Interfaces.*

[60] Yunhan Wu, Daniel Rough, Anna Bleakley, Justin Edwards, Orla Cooney, Philip R Doyle, Leigh Clark, and Benjamin R Cowan. 2020. See what I'm saying? Comparing intelligent personal assistant use for native and non-native language speakers. In *22nd international conference on human-computer interaction with mobile devices and services*. 1–9.

[61] Mingrui Ray Zhang, Ruolin Wang, Xuhai Xu, Qisheng Li, Ather Sharif, and Jacob O Wobbrock. 2021. Voicemoji: Emoji Entry Using Voice for Visually Impaired People. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. 1–18.