

Research Report

Smart speaker devices can improve speech intelligibility in adults with intellectual disability

Elizabeth Smith[†], Petroc Sumner[†], Craig Hedge^{†‡} and Georgina Powell[†] 

[†]School of Psychology, Cardiff University, Cardiff, UK

[‡]Cardiff University Brain Research Imaging Centre (CUBRIC), Cardiff University, Cardiff, UK

(Received September 2020; accepted February 2021)

Abstract

Background: Successful communication is vital to quality of life. One group commonly facing speech and communication difficulties is individuals with intellectual disability (ID). A novel route to encourage clear speech is offered by mainstream smart speakers (e.g., Amazon Alexa and Google Home). Smart speakers offer four factors important for learning: reward immediacy, spaced practice, autonomy/intrinsic motivation and reduced social barriers. Yet the potential of smart speakers to improve speech intelligibility has not been explored before.

Aims: To determine whether providing individuals with intellectual disabilities with smart speaker devices improved ratings of speech intelligibility for (1) phrases related to device use and (2) unrelated words via a semi-randomized controlled trial.

Methods & Procedures: In a semi-randomized controlled trial, an intervention group of adults with ID ($N = 21$) received smart speakers, while a control group ($N = 22$) did not. Before and after about 12 weeks, participants were recorded saying smart speaker-related phrases and unrelated words. Naïve participants then rated the intelligibility of the speech recordings.

Outcomes & Results: The group that received smart speakers made significantly larger intelligibility gains than the control group. Although the effect size was modest, this difference was found for both smart speaker-related phrases and unrelated words.

Conclusions & Implications: While the mechanism of action remains to be determined, the presence of smart speakers in the home had a demonstrable impact on ratings of speech intelligibility, and could provide cost-effective inclusive support for speech and communication improvement, improving the quality of life of vulnerable populations.

Keywords: smart speakers, intelligent personal assistants, speech intelligibility, intellectual disability, communication.

What this paper adds

What is already known on the subject

- Speech intelligibility is a key obstacle for social relationships and quality of life across several vulnerable populations (children with speech difficulties, older adults with dementia, individuals with ID). Anecdotal reports suggest mainstream smart speakers (e.g., Amazon Alexa, Google Home), could improve speech intelligibility.

What this paper adds to existing knowledge

- We used a semi-randomized controlled trial to show that using a smart speaker for about 12 weeks could improve ratings of speech intelligibility in adults with ID for both smart speaker-related phrases and unrelated words.

Address correspondence to: Georgina Powell, Cardiff University, Cardiff CF10 3AT, UK; email: PowellG7@cardiff.ac.uk

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

What are the potential or actual clinical implications of this work?

- These initial findings suggest that smart speaker technology could be a novel, and inclusive, route to improving speech intelligibility in vulnerable populations.

Introduction

Intellectual disability (ID) is characterized by limitations in intellectual and cognitive functioning and adaptive behaviours (Harris 2006). Between 50% and 90% of people with ID have communication or speech difficulties (Coppens-Hofman *et al.* 2016, Royal College of Speech and Language Therapists (RCLST) n.d.), which compound their other challenges (Cooper *et al.* 2015, Jansen *et al.* 2004, Smiley *et al.* 2005). We report a new approach to improve speech intelligibility in this population—mainstream smart speakers (e.g., Amazon Alexa and Google Home)—which we hypothesized would combine elements essential for learning, while also being engaging in a naturalistic setting.

Speech impairments in adults with ID

Speech impairments are common in adults with ID, with triple the rate seen in the general population (Harris 2006). They are a key issue in this population because of their impact on social interactions, employment and quality of life (Hitchcock *et al.* 2015, Law *et al.* 2009, McCormack *et al.* 2009). Furthermore, despite difficulties, speech remains the primary method of communication for adults with mild to moderate ID (e.g., Bradshaw 2001, McConkey *et al.* 1999).

There is a large range in the source and severity of speech impairments, just as there is in the degree of ID (Shriberg and Widder 1990, Icht 2019). Some individuals fail to develop any speech, while others have minor speech impairments. Articulation errors are particularly common (Shriberg and Widder 1990). Speech intelligibility impairments in adults with ID may stem from high-level cognitive difficulties involved in speech motor control and planning, rather than the development of the phonemic and syllabic repertoire *per se* (Coppens-Hofman *et al.* 2016). Therefore, targeted speech and language therapy may need to focus on repetition and continuous, understandable feedback so that utterances can become more automated and less reliant on higher level cognitive processes.

Targeted speech therapy for adults with ID is often limited or difficult to access (Graves 2007, Terband *et al.* 2017, Scottish Government 2012). Terband *et al.* (2017) suggested that was a perception that speech therapy will not be effective post-childhood, but recent research has suggested that targeted and engaging speech

therapy can be effective in improving intelligibility and clarity in adults with ID (see also Icht 2019). Speech and language interventions for individuals with ID are often based around one-to-one sessions with a speech and language therapist. These sessions might focus on both listening and articulation, and consist of activities such as practising exercises that increase in difficulty as the sessions progress (e.g., Terband *et al.* 2017). There is evidence for the effectiveness of these interventions in improving speech intelligibility and clarity in individuals with ID (Terband *et al.* 2017). However, they are also resource and time intensive, requiring both the therapist and the individual to meet for regular sessions, and the motivation to practise and carry out speech-and-language exercises can sometimes be low, which is a key factor in users abandoning speech therapy (Johnson *et al.* 2006, Koegel *et al.* 1998).

Learning is strongly influenced by an individual's sense of autonomy, which helps drive intrinsic motivation and persistence (Deci *et al.* 1981, McCrocklin 2016, Ryan and Deci 2006). Many adults with ID are somewhat resistant to help, preferring to emphasize their agency, and noting that they want to be 'treated as an adult' (Abbott and McConkey 2006, Klassen 2002, Smith *et al.* 2020). Individuals with ID often respond best to speech therapy when it is individually adapted to their interests (Terband *et al.* 2017). Icht (2019) developed a novel technique based on beat-boxing and rhythm ('Beataik'), which was successful in fostering engagement and interest and overcame the problem surrounding motivation and enjoyment in speech therapy. However, this intervention still required participants and instructors to attend regular sessions, and therefore there is scope to explore complementary alternatives based around low-pressure, highly rewarding and also distributed practice in the home environment.

The aim of the current study is to explore whether speech intelligibility improvement can occur naturally in the home, without individuals attending specific therapy sessions, via interactions with smart speaker devices.

Smart speakers to improve speech intelligibility

Recent anecdotal and unpublished reports have indicated that smart speakers could provide a novel route to support speech intelligibility improvement among individuals with speech impairments. Interacting with a



Figure 1. Features of a smart speaker. [Colour figure can be viewed at wileyonlinelibrary.com]

mainstream smart speaker to access functions via verbal commands (figure 1) may unite key requisites for learning: immediacy of rewards, spaced repetition, intrinsic motivation/autonomy and reduced anxiety/social barriers. In a parallel study, using both qualitative and quantitative data, we established that verbal individuals with mild to moderate ID are able to interact with smart speaker devices, with around 80% reporting that they enjoy using them and that they enable them to be more independent (Smith *et al.* 2020). Only 9.5% of participants with an ID in this study reported not using the device to access any features.

With smart speakers, the reward (such as entertainment) is immediate, contingent upon speaking clearly enough for the device to recognize the command. Immediate reward leads to a stronger association with the behaviour (Bermudez and Schultz 2014, Woolley and Fishbach 2018).

Second, successful interactions with smart speakers are likely to be spaced across days and weeks (e.g., asking for a favourite song; see below for a discussion of unsuccessful interactions). In contrast, speech and language therapists tend to be available for restricted time periods. It is a long-standing finding that spaced (distributed) practice is more effective than massed practice (for reviews, see Dempster 1996 and Kang 2016). For example, participants receiving the same number of practice sessions for a new motor task show superior learning outcomes when the sessions are spread across days rather than within the same day (Shea *et al.* 2000).

Third, the motivation for smart speaker engagement is the outcome (e.g., entertainment) rather than

as an exercise to improve speech per se. As noted above, intrinsic motivation and engagement is critical for any therapeutic approach in an ID population, and the entertaining and rewarding nature of smart speakers may encourage individuals to engage frequently and independently, without prompting or schedule. In a parallel study (Smith *et al.* 2020), we found that the majority of people with ID reported that they enjoyed using smart speaker devices, including when they were alone. We found that participants reported that the smart speakers provided social companionship, easy access to entertainment and information, and helped them to carry out tasks independently (e.g., reminders for medication). Some participants reported challenges with the smart speakers understanding their commands and were not aware of all the available features. However, perseverance was generally high due to motivation to engage with the smart speaker and access its unique features.

Fourth, failing to be understood by a device does not present the same issues of social awkwardness that may occur when having to repeat oneself to another person. Social factors such as fear of judgment and performance anxiety limit learning, especially among people with ID where anxiety is common (Polvinen and Dionne 2019). Individuals with ID can also feel self-conscious about speech difficulties (Shessel and Reiff 1999). When using smart speakers, it is normal that any individual may have to repeat themselves to be understood. The device provides immediate feedback (e.g., Alexa responds, 'I'm sorry, I didn't catch that'), which typically prompts users to repeat the command, potentially more slowly and clearly. Moreover, mainstream devices are affordable and inclusive, rather than labelling someone as needing the assistance of a 'bespoke device'.

Unhelpful features of smart speakers

It is also possible that smart speakers could preclude or limit learning. Off-the-shelf devices have a higher threshold for speech intelligibility than humans (e.g., Ballati *et al.* 2018), who can often guess the intention of unclear speech or recognize non-verbal communication. The device's more consistent threshold might facilitate learning (Hulac *et al.* 2016), but it is just as likely to hinder learning. Smart speakers lack the ability to deploy 'shaping', where the criterion for reward starts low and is dynamically adjusted as behaviour more closely approximates the desired behaviour (if anything the criterion gets slightly lower with use, since the software can learn to recognize pronunciation by users). The high initial threshold demanded by the device could lead to frustration and demotivation, and potentially preclude learning experiences altogether.

Second, the range of words used to command the smart speaker will be much narrower than those needed to converse with people. Therefore, if the smart speaker is to have any utility as an intervention to improve speech intelligibility in general, it is important to assess whether any learning generalizes beyond the specific words and phrases used to command the device.

Indications that smart speakers might improve speech intelligibility

In reviews of Amazon Alexa on the Amazon website, Pradhan *et al.* (2018) observed that users with speech impairments discussed speech recognition as a challenge, but there were also reports that some individuals learnt to speak more slowly and clearly. A recent study found that speech production in a second language (English) improved among four Japanese students while using Amazon Alexa over a 10-week period (Dizon 2020 note there was no control group). In addition, a small unpublished study (Denman and Jones 2019) found that when talking to a smart speaker, students with dysarthria appeared to speak more clearly according to their speech and language therapists; this generalized with individuals trying to speak as clearly to others as they do when speaking to the smart speaker. Innovate Trust (a supported living provider) also offers anecdotal reports that some individuals with ID and speech impairment gradually improved their speech due to a desire to engage with speech-based technology (Vass 2018). Therefore, although there are some initial suggestions—either anecdotal reports or small studies in different populations—that smart speakers might improve speech intelligibility, there has been no controlled evaluation in an ID population.

Summary

Smart speakers contain several features that could support speech intelligibility learning. However, communication difficulties could also preclude device use and learning opportunities. Even if learning occurs, it may not generalize beyond the specific commands needed for smart speakers. This study aimed to determine whether providing individuals with intellectual disabilities with smart speaker devices improved ratings of speech intelligibility for (1) phrases related to device use and (2) unrelated words via a semi-randomized controlled trial.

Method

The study was undertaken in collaboration with a local charity, Innovate Trust, which provides supported ac-

commodation to individuals with intellectual disabilities living in Cardiff, Wales, UK. At the outset of the study, the charity was in the process of introducing off-the-shelf smart speakers to supported living houses. The ongoing roll-out provided the opportunity to run a semi-randomized controlled trial for individuals in houses yet to receive devices (as detailed in the participants section below).

The study was entirely naturalistic: individuals in the intervention group received the device from the charity. The smart speaker devices were a mixture of Amazon and Google models, and were purchased in 2018. Instructions on how to use the devices were given verbally by the member of staff at the charity who set up the device, and involved basic demonstrations of different uses and features (e.g., playing music, setting reminders) and the answering of questions. Beyond this, individuals were free to use the device as and when they desired. This approach translates directly to how devices would typically be used beyond the study setting. Voice profiles were not set up by the supported living charity as they were still relatively new features at the time the study was conducted.

Where a suitable communal space existed, the charity provided a household with one device for all individuals to use; if there was no appropriate communal space, individuals received separate devices in their own rooms. No routine or protocol was imposed on participants and device usage was not recorded or monitored (for ethical reasons, as it would require listening to recordings to determine who initiated each command, including potential use by house members, staff or visitors not consented in the study). All participants in the control group received devices in the same way at the end of the study period.

Participants

All individuals with mild to moderate ID yet to receive devices were given the opportunity to participate in the study. Individuals with severe ID who could not provide informed consent were excluded. Participants who were considered unable to learn to communicate with the devices (e.g., due to cognitive or speech impairments that were too severe) were not included in the study, as judged by the supported living charity. Sample size was therefore determined by the maximum number of eligible participants available within the specific population.

A total of 48 individuals with ID were recruited to participate, living in 27 households (flats, houses or buildings with shared living space). These participants were allocated to either an intervention group, receiving a device for the study period (median = 12 weeks, range = 8–20 weeks), or a control group, who received

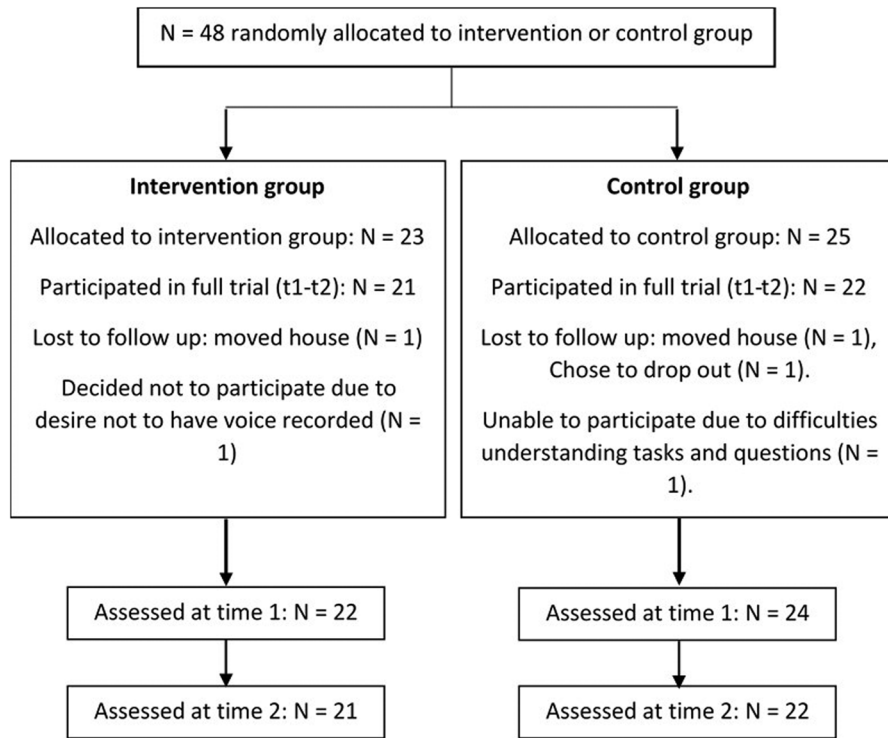


Figure 2. Participant allocation and progress through the study.

Table 1. Descriptive statistics (mean, standard deviation (SD) and range) for chronological age and each IQ/abilities assessment for each group

	<i>N</i>	Mean ± SD	Range
<i>Chronological age (years)</i>			
Intervention group	22	45.3 ± 13.7	22–69
Control group	22	48.6 ± 16.9	22–82
<i>WAIS-IV verbal</i>			
Intervention group	20	59.05 ± 8.54	50–76
Control group	16	57.31 ± 6.80	50–72
<i>PPVT</i>			
Intervention group	2	32.50 ± 12.02	24–41
Control group	6	43.33 ± 23.29	20–77
<i>WAIS-IV Matrix reasoning</i>			
Intervention group	19	3.11 ± 1.29	1–5
Control group	19	3.47 ± 2.32	1–9

Note: WAIS-IV verbal composite index: norm of 100, SD = 15; WAIS-IV Matrix reasoning subtest: norm of 10, SD = 3; and PPVT: norm of 100, SD = 15.

devices after the study period. Groups were allocated using a semi-randomized design; individuals within the same household were allocated to the same group to avoid exposure to the intervention conditions for those in the control group (intervention group *N* = 14 households, *N* = 23 participants; control group *N* = 13 households, *N* = 25 participants). Households were randomized by the supported living charity in Microsoft Excel, and the researchers were not involved in this process. The final sample size and number of exclusions and drop-outs in each group are shown in figure 2.

The ages of participants in each group are shown in table 1.

Design

The study used a 2 × 2 design. The within-participants variable of time had two levels: time 1 (pre-intervention/control) and time 2 (post-intervention/control). The between-participants variable of group had two levels: intervention group and control group. The dependent measure was speech intelligibility.

Measures

Participants were given a detailed description of what each assessment and task (described in detail below) would involve, but were not told the purpose of the study so as not to influence the results.

IQ assessments

Verbal IQ was assessed via the Wechsler Adult Intelligence Scale—IV (WAIS-IV; Wechsler 2008), and non-verbal ability was assessed via the Matrix Reasoning subtest of the WAIS-IV. A subset of participants with more limited verbal abilities completed the Peabody Picture Vocabulary Test—Fourth Edition (PPVT-4; Dunn and Dunn 2007). An overview of the number of participants completing each of these assessments and the IQ for each group is presented in table 1.

Intelligibility assessment

Recordings. A bespoke intelligibility measure was designed for this study. Each participant with ID was recorded saying five phrases related to the smart speaker (e.g., turn on the light) at times 1 and 2. The experimenter read out the phrase and asked the participant to repeat it. This procedure was used as it does not require reading skills. Participants were also recorded saying 17 words: five pictures of common items (e.g., hat) and five colours were shown to the participant on a piece of paper, and they were asked to name them one at a time. This meant that participants did not have the opportunity to copy the experimenter's pronunciation of the words. The remaining seven items were included to cover a 'wide variety of vocalic nucleus types' (Weismer *et al.* 1988: 1282). These seven words were less common and therefore repeated after the experimenter. Words and phrases were consistent across participants and across sessions and were always read out by the same experimenter (who also conducted the rest of the testing sessions).

Participants were recorded using a Samson Go USB condenser microphone (resolution 16-bit), attached to a laptop, placed on a table in front of them at a distance of approximately 50 cm. Audio was recorded using the software Audacity. The volume settings were kept consistent and the default sample rate (44,100 Hz) in Audacity was used. The microphone was set to Cardioid mode in order to record vocals directly in front of the microphone. Recordings took place in a separate, quiet room in the supported living house, or at a day centre, so as to minimize background noise.

Ratings. At the end of the trial, 24 psychology students served as blind raters (unrelated to the project);

they rated the intelligibility of each phrase from 1 to 7, where 1 = completely unintelligible and 7 = completely intelligible. The raters heard the phrases/words via headphones (Sennheiser HD 201). The volume was pre-set by the experimenter; raters had the opportunity to adjust the volume to a comfortable level before beginning the experiment, as determined during the practice trials. All raters found the volume comfortable and no adjustments were made.

The order of items (each recorded phrase across each time point) was randomized for each rater and for each verbal phrase the corresponding label was written on the screen (see the example in supplementary material A in the supplemental data online). It took approximately 25 min to rate the phrases. The recordings of 11 of the individual words were rated by a second group of 24 blind raters, following similar randomized procedures (35–40 min). Due to rater time constraints, we had to exclude the words repeated after the experimenter, except 'wax', which contains a critical phonological component of 'Alexa' that was absent in the five colours and five picture names.

We note that the rating method was more consistent with a speech clarity design rather than an intelligibility design, because raters were shown the target word or phrase during the rating. We took this decision because each of the words/phrases was repeated several times (because they were the same across participants with ID) and we wanted to avoid order effects, where ratings might increase based on how familiar the words/phrases had become to raters. However, we were clear in our verbal instructions to raters that we were interested primarily in intelligibility rather than clarity. For example, they should disregard any accent or perceived norms and rate the recordings based on how easy to understand they found them. We found that raters were happy to make these ratings and did not report confusion with the instructions.

Data analyses

Data were analysed with linear mixed effects models using the lme4 (Bates *et al.* 2015) package in R (R Core Team, 2016). To obtain *p*-values for the main effects and interactions analogous to traditional repeated-measures analysis of variance (ANOVA), denominator degrees of freedom were calculated using Satterthwaite's approximation implemented in the lmerTest package (Kuznetsova *et al.* 2016). Post-hoc comparisons were therefore analogous to a *t*-test.

Results

Ratings of participants' intelligibility at times 1 and 2 were assessed using a multilevel linear mixed effect

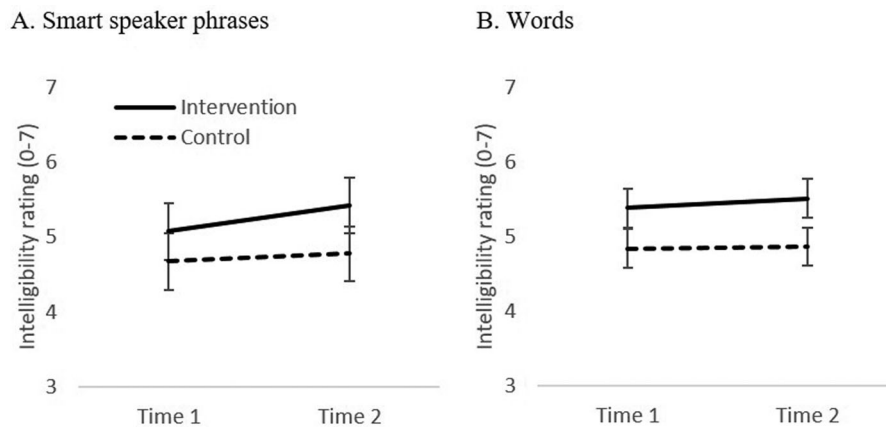


Figure 3. (A) Mean intelligibility ratings by independent raters of phrases spoken by individuals with intellectual disability (ID) at each time point for each group; and (B) mean intelligibility ratings of spoken words (colours, object naming and 'wax') at each time point for each group. These plots are based on the fixed effect estimates from the model (i.e., the standard error is associated with the raters and variance across the individuals with ID). The rating scale has a range from 1 to 7.

model in order to account for the shared variance both within the individuals with ID (i.e., it is a repeated measure) and the student participants who rated the phrases for intelligibility. Participant houses were also initially included as a random effect, however, there were too few instances of participants sharing houses to allow the model to calculate the variance associated with this factor, and it was therefore not necessary to include it in the final models. The model included three fixed effects: group (experimental versus control, between subjects), time (time 1 versus time 2, within subjects) and the group \times time interaction. Individual raters and individuals with ID were treated as random effects (intercepts). Mean scores across groups and conditions, and full post-hoc comparisons are shown in supplementary material 2 in the supplemental data online.

In the first model, we assessed intelligibility ratings of phrases directly related to the smart speaker (e.g., 'play some music'). There was a significant interaction of time \times group (figure 3A) ($F(1, 1996) = 23.91, p < 0.001$), with post-hoc comparison showing a greater increase in intelligibility ratings for the intervention group (mean = 0.36, $t(1996) = 9.83, p < 0.001$) relative to the control group (mean = 0.11, $t(1996) = 3.07, p < 0.01$). The main effect of time was also significant ($F(1, 1996) = 84.23, p < 0.001$).

While there was a numerical difference between the control and intervention group at baseline, it was not significant ($t(41) = -0.78, p = 0.44$), and there was no significant main effect of group ($F(1, 41) = 1.04, p = 0.31$). Note that time was within subject, while group was between subject and the ID population had wide heterogeneity. We could not pairwise control for speech intelligibility ratings; households of participants were randomly allocated to either group. Reassuringly, there was also no significant difference between groups on the

IQ measures, verbal or otherwise. There were numerically more participants in the control group who did not take the verbal IQ test, indicating possible speech difficulties, but their picture vocabulary (PPVT) was not lower than the interventional group (table 1).

The finding of importance here is the interaction, indicating that despite heterogeneity, a significantly different effect of time could be detected. Due to the significant interaction in the first model, we carried out a second (identical) model to assess if improvements were replicated in words not directly associated with smart speaker commands (colours, object naming and 'wax'). For these unrelated words we also found that the interaction of time \times group was significant (figure 3B) ($F(1, 1996) = 6.93, p < 0.01$), with a post-hoc comparison again showing that the increase in intelligibility ratings was greater for the intervention group (mean = 0.13, $t(1996) = 4.50, p < 0.001$) relative to the control group (mean = 0.02, $t(1996) = 0.84, p = 0.40$). The main effect of time was also significant ($F(1, 1996) = 14.46, p < 0.001$) and the main effect of group was not ($F(1, 41) = 3.60, p = 0.06$). Again, there was no significant group difference at baseline ($t(41) = -1.72, p = 0.09$).

Together, these analyses show that access to smart speakers led to improvements in speech intelligibility ratings not only for phrases related to smart speakers but also more generally for unrelated words.

Effect of intervention duration

The median duration of the intervention period was 12 weeks; however, there was some variance around this due to logistical constraints (range = 8–20 weeks). We were therefore interested in whether the improvement in speech intelligibility ratings was related to the

duration of the intervention period (i.e., how long the individual had access to the smart speaker). However, we did not find evidence for a correlation between the duration of the intervention period (in weeks) and improvement in intelligibility ratings for either words ($r = 0.27, p = 0.25$) or phrases ($r = -0.23, p = 0.3$).

Reliability of the ratings

To ensure our results were underpinned by good reliability of the intelligibility ratings, we analysed agreement between the raters using a two-way random effects intraclass correlation coefficient (ICC) for absolute agreement. We report both the single and average rater ICCs as we average across the raters for our main analysis. For the phrases, this showed good agreement for the single rater, $ICC(2,1) = 0.72$, 95% confidence interval (CI) [0.69, 0.75], and excellent agreement for the average rater, $ICC(2,k) = 0.98$, 95% CI [0.98, 0.99]. For the words, this showed moderate agreement for the single rater, $ICC(2,1) = 0.48$, 95% CI [0.44, 0.51] and excellent agreement for the average rater, $ICC(2,k) = 0.96$, 95% CI [0.95, 0.96]. These results therefore suggest good rater reliability.

Discussion

We found improved ratings of speech intelligibility for individuals with mild to moderate ID who had the opportunity to use smart speakers over about a 12-week intervention period. Although the exact mechanism of effect is still undetermined, this is the first controlled trial to examine, and find, improved speech intelligibility ratings following smart speaker use. The findings are consistent with previous anecdotal reports (Denman and Jones 2019, Dizon 2020, Pradhan et al. 2018, Vass 2018), and potentially open up an exciting new avenue for speech intelligibility therapy within the field of ID and beyond.

Generalization beyond specific smart speaker phrases

Speech improvement was found for phrases that were related to device use (e.g., *play some music*), and the effect was replicated, although a little smaller, for unrelated words (e.g., colours), thus going beyond specific practiced phrases. This finding is particularly significant given that the smart speakers were not actively provided as a speech training tool, but rather to improve quality of life with entertainment functions. No protocols or practice regimes were imposed to improve speech intelligibility. The intelligibility gains were a by-product of unstructured use of smart speakers for desired functions (e.g., playing music).

Reduced speech intelligibility is common among individuals with mild to moderate intellectual disabilities and has a major impact on social interaction (Hitchcock et al. 2015, McCormack et al. 2009). Given the importance of social interaction for quality of life among this population (Nota et al. 2007), any increase in intelligibility could have important implications for well-being.

This study raises the question of whether benefits would also occur for the many other individuals with speech and language difficulties (Law et al. 2007), where waiting lists for therapy can be long (Bercow 2018). The motivating features of smart speakers could similarly support second language learning, as implicated in previous (uncontrolled) research (e.g., Dizon 2020, McCrocklin 2016).

What drives the effect of smart speakers on speech intelligibility ratings?

The intention in this study was to determine outcomes in a typical autonomous setting, rather than through a specific regime. Therefore, it does not establish the means by which smart speakers helped improve intelligibility. However, we expect four factors to have played a role: immediacy, spaced repetition, autonomous motivation and removing social barriers.

Immediate feedback about whether phrases have been understood, and immediate reward if the intelligibility threshold is exceeded, are expected to be key for learning (Bermudez and Schultz 2014, Woolley and Fishbach 2018). Smart speakers also support spaced repetition (Dempster 1996). One of five 'good communication standards' outlined by the RCSLT (2014) is giving individuals frequent chances to communicate. Smart speakers are always present and will always respond.

In our study learning was autonomous, self-motivated and naturalistic, which is reported to be particularly effective for improving speech intelligibility (Camarata 1993, Koegel et al. 1998), and motivation more generally (Guay et al. 2008, McCrocklin 2016, Ryan and Deci 2006). There is limited research on intelligibility training for adults with ID, and previous research indicates that compliance and motivation levels can be low for some types of speech therapy (van Leer and Connor 2010). Providing options for home learning has been shown to be effective in the past, for example, giving individuals digital tools for home practice can lead to improved outcomes in vocal training (van Leer and Connor 2012).

Attrition to an intervention is also likely to reduce when the intervention is enjoyable and engaging. Icht (2019) reported positive outcomes following the design of a fun beat-boxing intervention for adults with ID.

In our study, smart speakers are more than just a route to practise verbal communication: they also unlock a number of enjoyable and useful features that individuals with ID may not be able to access without the smart speaker. In a parallel study to this one, where we conducted questionnaires and qualitative interviews about device use and experiences, we found that some individuals with ID reported strong motivation to persevere with their smart speaker despite the challenge of being understood (Smith *et al.* 2020). In that study, we also found that individuals with ID were positive about the devices in general, reporting that they felt they gave them more independence and that they particularly enjoyed accessing entertainment features. Therefore, because there are strong motivations to continue to use the devices, improvements in speech intelligibility are really a secondary, 'hidden', outcome, and are not the main reason that someone would use the device.

We further suspect that smart speakers might help to remove social barriers (e.g., social anxiety or concerns of embarrassment when repeating questions or requests), and avoid the social stigma and negative aesthetic often associated with bespoke support tools (Judge *et al.* 2009).

It is also possible that improvements are not due to interacting with the smart speaker per se, but are driven by indirect factors related to the presence of the smart speaker. For example, the smart speaker might prompt more interactions with staff members, family or housemates (e.g., instructions on how to use it, commenting on features, offering encouragement or feedback). Therefore, the device could have served to increase the participants' interactions with other people, and this could in turn underlie speech improvements. However, we found in a previous study that many of the participants with ID enjoyed using the smart speaker device when alone (Smith *et al.* 2020) and only 27% reported that they needed help using it, which suggests that this is unlikely to be the only credible explanation.

We cannot distinguish between direct and indirect effects in the present study, although it would be useful and interesting to do so in future work. However, even if the smart speaker's influence on improvement is partly indirect, it is still a relatively simple and cost-effective solution. Voice and video calling from the smart speakers was unfortunately not set-up for some participants in the study, but in future it would be interesting to see if these features might prompt more social interactions that might, in turn, lead to improvements in speech and language.

Limitations of smart speakers

It is worth noting that although improvements were found following the use of smart speakers, the effect

sizes were relatively modest. It is possible that formal and targeted speech therapy might produce larger treatment effects in the same time frame, although we did not test this comparison in the present study. In our parallel study (Smith *et al.* 2020) we found that a minority of individuals did not use the devices very frequently because they had not received sufficient training, and that some features had not been set up (e.g., voice calling). However, only 9.5% of participants in this study did not access any features of the smart speakers, and around 80% of individuals reported that they enjoyed using the devices and that they increased their independence. Qualitative data from both participants with an ID and their support workers also provided evidence that the devices were used and enjoyed by the majority of participants. Qualitative data also suggested that there was a high level of perseverance and that participants with ID would repeat commands until the smart speaker registered them.

We had limited control over these factors because the study was naturalistic and opportunistic. In future, we would look to increase support for device use and tailor features to individual needs. This could lead to greater device use and interaction, and associated gains in speech improvement. For example, many smart speakers now have a 'voice profile' option, which allows them to adapt to individual voices, although for the purposes of speech intelligibility training, it may be useful for the device to have a relatively high speech recognition threshold as this would encourage greater intelligibility. The duration of the study was also relatively short (approximately 12 weeks), and it is possible that greater gains might be observed with longer treatment exposure. We did not find a correlation between intervention duration (i.e., the number of weeks that participants access the devices) and improvement in speech intelligibility, but the variance in duration was limited and not a design feature of the study. In future research exploring exposure duration, it would be important to distinguish between intervention duration (i.e., in weeks) and usage duration (i.e., number and length of interactions).

It is also important to acknowledge that poor speech recognition, at least in the initial stages of device use, can cause frustration for a minority of individuals (Smith *et al.* 2020). We found that some individuals showed strong perseverance, for example, practising alone, but a small minority (9.5%) did not use their device or learn what it could do. Pradhan *et al.* (2018) also reported speech challenges for smart speaker use among individuals with speech difficulties (as well as improvements). While the mainstream nature of smart speakers means that they are inclusive, individuals with more severe speech and language difficulties may nonetheless be excluded from using devices due to the verbal nature of

the technology. Smart speaker companies are continually improving the speech recognition of devices. In the face of these potential developments, from a speech intelligibility intervention viewpoint it may be important to consider the possibility of setting specific thresholds of voice recognition to ensure that speech intelligibility is encouraged at an achievable level. There are also privacy issues to keep in mind, as with all web-linked technology.

Study limitations

A limitation of the study was that we were not able to record the number of times individuals used and interacted with the smart speakers and the quality of these interactions. These data were not recorded for ethical reasons as it might require us to listen to recordings from people not consented in the study, such as visitors. Therefore, we cannot measure any 'dose-dependent' effects or assess the different types of utterances that were directed towards the device and whether these related to specific improvements in speech. However, in associated research we have shown that individuals did engage with the devices in general and enjoy using them, so it was not the case that participants in the intervention group did not interact with the smart speakers at all (Smith *et al.* 2020). This study was an important first step in establishing a potential intervention effect, and future research could explore the mechanism of this effect in greater detail.

Conclusions

Access to smart speakers led to significantly improved speech intelligibility ratings among adults with ID. This improvement extended beyond the specific phrases needed to control the device. These findings have potential relevance for other groups experiencing speech and language difficulties, and could viably offer a simple-to-deliver, affordable route to supplement speech and language therapy at scale. Future research should further investigate the mechanisms behind these improvements.

Acknowledgements

This research was funded by a Health and Care Research Wales Fellowship (grant number SCF-18-1504). The authors thank Innovate Trust, Cardiff, UK, and all participants for supporting this research.

Data availability statement

The data that support the findings of this study are openly available in Open Science Framework at:

https://osf.io/n6r5d/?view_only=682a91826916430388f1a724734cee17

References

- ABBOTT, S. and MCCONKEY, R., 2006, The barriers to social inclusion as perceived by people with intellectual disabilities. *Journal of Intellectual Disabilities*, **10**(3), 275–287.
- BALLATI, F., CORNO, F. and DE RUSSIS, L., 2018, Hey Siri, Do You Understand Me?": Virtual Assistants and Dysarthria. In *Intelligent Environments (Workshops)* (pp. 557–566).
- BATES, D., MAECHLER, M., BOLKE, B. and WALKER, S., 2015, Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, **67**(1), 1–48, <https://doi.org/10.18637/jss.v067.i01>.
- BERCOW, J., 2018, *Bercow: Ten Years On: An independent review of provision for children and young people with speech, language and communication needs in England*. (London: ICan charity). Accessed June 2020.
- BERMUDEZ, M. A., SCHULTZ, W., 2014, Timing in reward and decision processes. *Phil. Trans. R. Soc. B*, **369**, 20120468. <http://doi.org/10.1098/rstb.2012.046>
- BRADSHAW, J., 2001, Complexity of staff communication and reported level of understanding skills in adults with intellectual disability. *Journal of Intellectual Disability Research*, **45**(3), 233–243.
- CAMARATA, S., 1993, The application of naturalistic conversation training to speech production in children with speech disabilities. *Journal of Applied Behavior Analysis*, **26**(2), 173–182.
- COOPER, S. A., MCLEAN, G., GUTHRIE, B., MCCONNACHIE, A., MERCER, S., SULLIVAN, F. and MORRISON, J., 2015, Multiple physical and mental health comorbidity in adults with intellectual disabilities: population-based cross-sectional analysis. *BMC Family Practice*, **16**(1), 110.
- COPPENS-HOFMAN, M. C., TERBAND, H., SNIK, A. F. and MAASSEN, B. A., 2016, Speech characteristics and intelligibility in adults with mild and moderate intellectual disabilities. *Folia Phoniatrica et Logopaedica*, **68**(4), 175–182.
- DECI, E. L., SCHWARTZ, A. J., SHEINMAN, L. and RYAN, R. M., 1981, An instrument to assess adults' orientations toward control versus autonomy with children: reflections on intrinsic motivation and perceived competence. *Journal of Educational Psychology*, **73**, 642–650.
- DEMPSTER, F. N., 1996, Distributing and managing the conditions of encoding and practice. *Memory*, 317–344.
- DENMAN, A. and JONES, V., 2019, *Hey Google: can using a smart speaker improve intelligibility?* (Poster presentation). Royal College of Speech and Language Therapists Conference, Nottingham, UK.
- DIZON, G., 2020, Evaluating Intelligent Personal Assistants for L2 Listening and Speaking Development. *Language Learning & Technology*, **24**(1), 16–26.
- DUNN, L. M. and DUNN, D. M., 2007, Peabody Picture Vocabulary Test: PPVT-4A. NCS Pearson.
- GRAVES, J., 2007, Factors influencing indirect speech and language therapy interventions for adults with learning disabilities: the perceptions of carers and therapists. *International Journal of Language & Communication Disorders*, **42**, 103–121.
- GUAY, F., RATELLE, C. F. and CHANAL, J., 2008, Optimal learning in optimal contexts: the role of self-determination in education. *Canadian Psychology/Psychologie canadienne*, **49**(3), 233.
- HARRIS, J. C., 2006, *Intellectual disability: Understanding its development, causes, classification, evaluation, and treatment*. (Oxford University Press).

- HITCHCOCK, E. R., HAREL, D. and BYUN, T. M., 2015, Social, emotional, and academic impact of residual speech errors in school-aged children: a survey study. In *Seminars in Speech and Language* (Vol. 36, No. 04, 283–294). (Thieme Medical Publishers).
- HULAC, D., BENSON, N., NESMITH, M. C. and WOLLERSHEIM SHERVEY, S., 2016, Using variable interval reinforcement schedules to support students in the classroom: an introduction with illustrative examples. *Journal of Educational Research and Practice*, **6**(1), 6.
- ICHT, M., 2019, Introducing the Beataalk technique: using beat-box sounds and rhythms to improve speech characteristics of adults with intellectual disability. *International journal of language & communication disorders*, **54**(3), 401–416.
- JANSEN, D. E., KROL, B., GROOTHOFF, J. W. and POST, D., 2004, People with intellectual disability and their health problems: a review of comparative studies. *Journal of Intellectual Disability Research*, **48**(2), 93–102.
- JOHNSON, J. M., INGLEBRET, E., JONES, C. and RAY, J., 2006, Perspectives of speech language pathologists regarding success versus abandonment of AAC. *Augmentative and Alternative Communication*, **22**(2), 85–99.
- JUDGE, S., ROBERTSON, Z., HAWLEY, M. and ENDERBY, P., 2009, Speech-driven environmental control systems—a qualitative analysis of users’ perceptions. *Disability and Rehabilitation: Assistive Technology*, **4**(3), 151–157.
- KANG, S. H., 2016, Spaced repetition promotes efficient and effective learning: policy implications for instruction. *Policy Insights from the Behavioral and Brain Sciences*, **3**(1), 12–19.
- KLASSEN, R., 2002, A question of calibration: a review of the self-efficacy beliefs of students with learning disabilities. *Learning Disability Quarterly*, **25**(2), 88–102.
- KOEGEL, R. L., CAMARATA, S., KOEGEL, L. K., BEN-TALL, A. and SMITH, A. E., 1998, Increasing speech intelligibility in children with autism. *Journal of Autism and Developmental Disorders*, **28**(3), 241–251.
- KUZNETSOVA, A., BROCKHOF, P. B. and CHRISTENSEN, R. H. B., 2016, lmerTest: Tests in linear mixed effects models. R package version 2.0-33. <https://CRAN.R-project.org/package=lmerTest>.
- LAW, J., GAAG, A., HARDCASTLE, W. J., BECKETT, D. J., MACGREGOR, A. and PLUNKETT, C., 2007, Communication Support Needs: A Review of the Literature. Scottish Executive.
- LAW, J., RUSH, R., SCHOON, I. and PARSONS, S., 2009, Modeling developmental language difficulties from school entry into adulthood: literacy, mental health, and employment outcomes. *Journal of Speech, Language, and Hearing Research*.
- MCCONKEY, R., MORRIS, I. and PURCELL, M., 1999, Communications between staff and adults with intellectual disabilities in naturally occurring settings. *Journal of Intellectual Disability Research*, **43**(3), 194–205.
- MCCORMACK, J., MCLEOD, S., MCALLISTER, L. and HARRISON, L. J., 2009, A systematic review of the association between childhood speech impairment and participation across the lifespan. *International Journal of Speech–Language Pathology*, **11**(2), 155–170.
- MCCROCKLIN, S. M., 2016, Pronunciation learner autonomy: the potential of automatic speech recognition. *System*, **57**, 25–42.
- NOTA, L., FERRARI, L., SORESI, S. and WEHMEYER, M., 2007, Self-determination, social abilities and the quality of life of people with intellectual disability. *Journal of Intellectual Disability Research*, **51**(11), 850–850–865.
- POLVINEN, L. and DIONNE, H., 2019, *The Relationship between Language-Based Learning Disabilities (LBD) & Anxiety*. LD Online. <http://www.ldonline.org/article/65699/>
- PRADHAN, A., MEHTA, K. and FINDLATER, L., 2018, Accessibility came by accident: use of voice-controlled intelligent personal assistants by people with disabilities. In Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (459). ACM.
- RYAN, R. M. and DECI, E. L., 2006, Self-regulation and the problem of human autonomy: does psychology need choice, self-determination, and will? *Journal of Personality*, **74**(6), 1557–1586.
- ROYAL COLLEGE OF SPEECH AND LANGUAGE THERAPISTS (RCSLT), 2014, Five good communication standards easy read version. [RCSLT.org/www.rcslt.org/members/docs/5_good_comms_standards_easy_read](https://www.rcslt.org/members/docs/5_good_comms_standards_easy_read)
- ROYAL COLLEGE OF SPEECH AND LANGUAGE THERAPISTS (RCSLT) (n.d.). Supporting people with learning disabilities [Fact Sheet]. RCSLT.org.
- SHEA, C. H., LAI, Q., BLACK, C. and PARK, J. H., 2000, Spacing practice sessions across days benefits the learning of motor skills. *Human Movement Science*, **19**(5), 737–760.
- SHEssel, I. and REIFF, H. B., 1999, Experiences of adults with learning disabilities: positive and negative impacts and outcomes. *Learning Disability Quarterly*, **22**(4), 305–316.
- SHRIBERG, L. D. and WIDDER, C. J., 1990, Speech and prosody characteristics of adults with mental retardation. *Journal of Speech, Language, and Hearing Research*, **33**(4), 627–653.
- SMILEY, E., 2005, Epidemiology of mental health problems in adults with learning disability: an update. *Advances in Psychiatric Treatment*, **11**(3), 214–222.
- SMITH, E., SUMNER, P., HEDGE, C. and POWELL, G., 2020, Smart-speaker technology and intellectual disabilities: agency and wellbeing. *Disability and Rehabilitation: Assistive Technology*, 1–11.
- TERBAND, H., COPPENS-HOFMAN, M. C., REFFELTRATH, M. and MAASSEN, B. A., 2018, Effectiveness of speech therapy in adults with intellectual disabilities. *Journal of Applied Research in Intellectual Disabilities*, **31**(2), 236–248.
- SCOTTISH GOVERNMENT, 2012, The same as you? A review of services for people with learning disabilities: Consultation report. Gov.scot.
- VAN LEER, E. and CONNOR, N. P., 2010, Patient perceptions of voice therapy adherence. *Journal of Voice*, **24**(4), 458–469.
- VAN LEER, E. and CONNOR, N. P., 2012, Use of Portable Digital Media Players Increases Patient Motivation and Practice in Voice Therapy. *Journal of Voice*, **26**(4), 447–453.
- VASS, K., 2018, Research Report: Smart Speakers for Supported Living: Findings, Recommendations and Evaluations. Innovate Trust.
- WECHSLER, D., 2008, *Wechsler adult intelligence scale—Fourth Edition (WAIS-IV)*. (San Antonio, Texas: Pearson)
- WEISMER, G., KENT, R. D., HODGE, M. and MARTIN, R., 1988, The acoustic signature for intelligibility test words. *The Journal of the Acoustical Society of America*, **84**(4), 1281–1291.
- WOOLLEY, K. and FISHBACH, A., 2018, It’s about time: earlier rewards increase intrinsic motivation. *Journal of Personality and Social Psychology*, **114**(6), 877.

Supporting Information

Additional supporting information may be found online in the Supporting Information section at the end of the article.

Supporting information.