International Journal of Speech-Language Pathology

ISSN: 1754-9507 (Print) 1754-9515 (Online) Journal homepage: https://www.tandfonline.com/loi/iasl20

Speech-driven mobile games for speech therapy: User experiences and feasibility

Beena Ahmed, Penelope Monroe, Adam Hair, Chek Tien Tan, Ricardo Gutierrez-Osuna & Kirrie J. Ballard

To cite this article: Beena Ahmed, Penelope Monroe, Adam Hair, Chek Tien Tan, Ricardo Gutierrez-Osuna & Kirrie J. Ballard (2018) Speech-driven mobile games for speech therapy: User experiences and feasibility, International Journal of Speech-Language Pathology, 20:6, 644-658, DOI: 10.1080/17549507.2018.1513562

To link to this article: https://doi.org/10.1080/17549507.2018.1513562

Published online: 09 Oct 2018.

International Journal of Speech-Language Pathology, 2018; 20: 644–658

Speech-driven mobile games for speech therapy: User experiences and feasibility

BEENA AHMED[1,2], PENELOPE MONROE[3], ADAM HAIR[4], CHEK TIEN TAN[5], RICARDO GUTIERREZ-OSUNA[4] AND KIRRIE J. BALLARD[3]

[1]*School of Electrical Engineering and Telecommunications, University of New South Wales, Sydney, Australia,* [2]*School of Electrical Engineering and Telecommunications, Texas A&M University at Qatar, Doha, Qatar,* [3]*Faculty of Health Sciences, University of Sydney, Sydney, Australia,* [4]*Department of Computer Science and Engineering, Texas A&M University College of Engineering, Texas (TX), USA and,* [5]*Games Studio Department, University of Technology, Sydney, Australia*

Abstract

Purpose: To assist in remote treatment, speech-language pathologists (SLPs) rely on mobile games, which though entertaining, lack feedback mechanisms. Games integrated with automatic speech recognition (ASR) offer a solution where speech productions control gameplay. We therefore performed a feasibility study to assess children's and SLPs' experiences towards speech-controlled games, game feature preferences and ASR accuracy.

Method: Ten children with childhood apraxia of speech (CAS), six typically developing (TD) children and seven SLPs trialled five games and answered questionnaires. Researchers also compared the results of ASR to perceptual judgment.

Result: Children and SLPs found speech-controlled games interesting and fun, despite ASR–human disagreements. They preferred games with rewards, challenge and multiple difficulty levels. Automatic speech recognition–human agreement was higher for SLPs than children, similar between TD and CAS and unaffected by CAS severity (77% TD, 75% CAS – incorrect; 51% TD, 47% CAS, 71% SLP – correct). Manual stop recording yielded higher agreement than automatic. Word length did not influence agreement.

Conclusion: Children's and SLPs' positive responses towards speech-controlled games suggest that they can engage children in higher intensity practice. Our findings can guide future improvements to the ASR, recording methods and game features to improve the user experience and therapy adherence.

KEYWORDS: *speech-controlled games; mobile therapy apps; ASR applications; ASR in games; childhood apraxia of speech*

Introduction

In speech therapy, a child undergoes extended therapy sessions with a trained speech-language pathologist (SLP) in a clinic. Therapy can last from several months to several years depending on the child's level of impairment. To make progress, frequent and regular practice is critical and hence the child can acquire new skills and habits. However, this can be difficult due to the shortage of trained professionals, large distances involved and/or high cost of speech therapy (Ruggero, McCabe, Ballard, & Munro, 2012). Therapy can, in some cases, be complemented with exercises at home under the supervision of a parent or guardian. However, there are two key problems with home-based therapy practice. First, children need to be constantly motivated to perform exercises that are often monotonous and repetitive. Second, children should ideally be monitored by a supervising adult while performing these exercises at home to obtain feedback on their productions.

Increasingly, both SLPs and parents are resolving the first problem of demotivation with mobile-based speech therapy apps. These apps are gaining acceptance as valuable clinical tools, especially because touch-based devices (e.g. tablets and smartphones) are intuitive and engaging thus keeping children motivated during repeated practice of their exercises. Speech apps typically incorporate a substantial library of stimulus images and captivating interfaces in game-like environments (Expressive Solutions, 2018; Speech With Milo, 2017). The American Speech-Language-Hearing Association too advocates for

Correspondence: Dr. Beena Ahmed, Ph.D., School of Electrical Engineering and Telecommunications, University of New South Wales, Sydney 2052, Australia. Email: beena.ahmed@unsw.edu.au

ISSN 1754-9507 print/ISSN 1754-9515 online
Published by Informa UK Limited, trading as Taylor & Francis Group
DOI: 10.1080/17549507.2018.1513562

their use and provides links to such apps (American Speech-Language-Hearing Association, 2017). Studies also have shown that children have higher levels of engagement and make fewer errors with electronic interventions than with traditional therapy (Jamieson, Kranjc, Yu, & Hodgetts, 2004).

Though these apps address the first problem to some extent, they do not address the second problem of lack of external feedback. Few of the speech therapy apps currently available provide the child with immediate feedback on their productions nor offer any remote and/or automated speech assessment. This restricts the child to undertaking practice only in the company of an adult or on their own; potentially leading to lost opportunities for additional practice or limited external feedback. To mitigate this, several tools provide children with indirect feedback mechanisms to assist therapy. A number of apps allow the child to listen to audio models of the exercise (Smarty Ears – Articulate it, 2017; Tactus Therapy Solutions), whereas others allow the child to record their speech and play it back (Little Bee Speech, 2018; Smarty Ears – Apraxiaville, 2017; Smarty Ears – Articulate It, 2017). Some apps provide visualisations of the child's voice (Balbus Speech, 2017; Micro Video Corporation, 2014), animated stimuli (Smarty Ears – Apraxiaville, 2017) or voice-activated characters that move when the child speaks (Laureate Learning Systems, 2014; Tiga, 2011). Critically though, in these apps, progress through the game is not controlled by the child's speech. To maximise the benefit of app-based therapy exercises, feedback needs to be intrinsically integrated into the exercise by linking success in the game to speech performance so children receive timely feedback about their production.

The importance of response-contingent feedback has created interest in integrating automatic speech recognition (ASR) into speech therapy apps. Advancements in speech technologies and mobile computing have meant that speech processing is no longer restricted to offline analysis of the pitch, formants and amplitude of speech productions (Bälter, Engwall, Öster, & Kjellström, 2005; Pratt, Heintzelman, & Deming, 1993) but can now be used to recognise produced words and sentences in real time. Recently, a number of mobile-based voice- and speech-controlled games have been proposed to encourage regular in-home therapy. They allow players to control gameplay in real time by either varying the pitch or amplitude of their vocalisations (Ganzeboom, Yılmaz, Cucchiarini, & Strik, 2016; Lan, Aryal, Ahmed, Ballard, & Gutierrez-Osuna, 2014; Lopes, Magalhães, & Cavaco, 2016) or saying simple words and sentences (Rubin & Kurniawan, 2013; Tan, Johnston, Ferguson, Ballard, & Bluff, 2014). The player's success in these games thus serves as an explicit feedback measure of their productions.

In addition to being more engaging, these voice- and speech-controlled games offer multiple advantages over standard speech therapy games. (1) They can allow for smoother gameplay as they do not depend on a SLP or parent to assess the player's production. (2) Integrating automated feedback on the child's production into the gameplay provides real-time feedback with immediate consequences, i.e. progress in the game. (3) Based on the child's performance, the level of difficulty in the game can be adjusted and allow the child to practice a skill in gradually more challenging activities. Together, these advantages can maximise the benefit the child would gain from practising their therapy exercises at home and potentially enable the child to progress faster in their therapy.

The purpose of this initial study was to survey children's and SLPs' impressions and experiences of prototype speech-controlled games to examine how feasible each game type, with its associated features is for augmenting home-based speech practice. We included children with typically developing speech, as well as children with childhood apraxia of speech (CAS), a group who usually require intensive speech practice over long periods (Murray, McCabe, & Ballard, 2014) and who might benefit from engaging speech therapy games. Collecting their input at this early stage of app development is vital prior to progressing into a resource intensive efficacy trial; it will determine how feasible each game type and their features is for speech therapy delivery as well as guide the development of future versions. This study also tested the functionality and reliability of the ASR system integrated into the games in the context of speech therapy. We tested five different prototype games in this study: WordPop, SpeechWorm, Whack-A-Mole, Asteroids and Memory. All of these games are based on popular computer or mobile games, with the graphics and gameplay designed to be engaging for children. Of these, only Memory is not speech-controlled. SpeechWorm and WordPop are iOS games for use on Apple devices, whereas Whack-A-Mole, Asteroids and Memory are Android games. The games we presented contained two different recording methods. All the games required the screen to be touched to initiate recording but only three of the games required a second action to end the recording (SpeechWorm and Memory required a second touch and WordPop required the release of the initial touch). In the other two games, Whack-A-Mole and Asteroids, the recording ended after a pre-defined time. This varied the difficulty of gameplay and the degree of time pressure on producing a spoken response. We compared children's and SLPs' opinions as well as ASR–human agreement in relation to the differing recording mechanisms. We also looked at how children and SLPs responded to the added feature of response play-back in one of the games. Finally, we populated the games with a set of

30 one- to four-syllable words to provide a range of difficulty, accommodating different speech skills across participants. This also allowed us to explore any word length effects on ASR–human agreement across groups.

We hypothesised that children and SLPs

- would enjoy playing the games, finding them interesting and fun,
- would prefer games with reward systems and more than one level of difficulty,
- would prefer games with ASR
- would be supportive of their use in therapy and
- may disagree on which games they like most or least, given that they may approach the games with different motivations and different levels of prior exposure to apps.

We further hypothesised that ASR–human agreement may,

- decline as CAS severity increases and thus negatively influence response to the games in the more severely impaired children
- be affected by recording method (automatic vs. manual stopping) and
- be higher for longer words as single syllable words tend to have more dense phonological neighbourhoods (e.g. Luce & Pisoni, 1998) and are therefore, more likely to be confused by ASR algorithms.

Method

Participants

A total of 23 Australian-English-speaking participants were recruited for the study. Participants included 10 children with CAS ranging from mild to severe (9 male and 1 female; mean age: 7.9 years; range: 6–11 years), six children reported by parents as typically developing (TD; one male and five female; mean age: 8.7 years; range: 7–11 years) and seven SLPs (7 female; median experience: 12 years; range: 8–28 years). Five of the seven SLPs reported using computer games and apps routinely in their clinical practice. Of these five, all reported that none of these games or apps were voice or speech controlled. The participating CAS children were tested in one author's lab as part of an earlier study to rule out receptive language impairment, dysarthria or structural craniofacial anomalies. They had received a consensus diagnosis of CAS from two to three experienced SLPs; however, the intent in this study was to test the games with speech-disordered children and the diagnosis of CAS was not critical to the study aims. However, severity of speech impairment was estimated using percent phonemes correct (PPC; Shriberg & Kwiatkowski, 1982) for each child's first production of the each of the 30 words used in the games (during gameplay) as well as percent of the polysyllabic words (17/30 words) perceived to have correct lexical stress (PLSC). Percent phonemes correct ranged from 43.0 to 94.5 (Med = 85.5, IQR = 13.0) and PLSC from 35.3 to 100 (Med = 64.3, IQR = 45.8). Inter-rater reliability between two raters (PM and KB) for PPC in CAS children was calculated on a random 20% of words (60 words from 300 [30 words $\times$ 10 children]). Using intra-class correlation (absolute agreement), PPC, calculated for each word, was 0.809 (95%CI 0.696–0.883) for single measures and 0.894 (95%CI 0.820–0.938) for average measures. Reliability on PLSC judgment was calculated for the 31 polysyllabic words in this 60-word set. Using Cohen's Kappa, strength of agreement was "good" (Kappa = 0.668, Asymptotic SE = 0.133, 95%CI = 0.407–0.929). A list of speech error types and words in error for each CAS child is provided in Supplementary Table I. All procedures were approved by the University's Human Research Ethics Committee and all participants provided written informed consent/assent before study commencement.

Procedure

All participants were asked by the researchers to test all of the five prototype games – WordPop and SpeechWorm on Apple's iOS platform and Asteriods, Whack-A-Mole and Memory on the Android platform. The order of Apple and Android platform was randomised and, within platform, the order of games presented to each participant was randomised.

First, participants were given instructions on how to play a given game and offered a brief demonstration by the researcher. It was explained that the games would decide if their word productions were correct or incorrect, but that the games were in development and sometimes they would make mistakes. They were then asked to play the game for about 5 min and answer a set of questions about the game. The researcher provided general motivational feedback during the games (e.g. "you said those words really well", "you've nearly reached 100 points!"). Given the brief exposures to each game, the researcher assisted the child with screen taps for starting and/or stopping recordings when needed. This ensured that a minimum of 15 word recordings per game was targeted; although, this was not reached on three occasions when participants requested to discontinue a game. The questionnaire queried likes, dislikes, ease of use, level of interest in using the game again and suggestions for improvement. A combination of 5-point Likert scales (e.g. 1 = hard to use … 5 = easy to use) and open-ended questions (e.g. "What did you like about the game?") were used. For children, the experimenter read the questions aloud to the child in case of reading difficulty. There was a 5- to 7-min break between games as the researcher presented the user survey for the last played game, setup the next game and conversed with the child. After the fifth game, the participants were shown a screenshot from all five games, and asked which was their least favourite game and why. For open-ended questions, the researcher recorded and transcribed children's verbal responses; SLPs wrote their responses in the questionnaire. To score the tools' success in recognising each participant's word productions, the results from all users including SLPs were compared to the perceptual judgement of the researcher administering the games, who kept a tally during each game.

All participants were tested in a quiet environment in their home or in the University speech clinic. For six children, a headset microphone was used. For the remaining children, who either refused to wear the microphone or continually played with it, the inbuilt tablet microphone was positioned within 30 cm of their mouth. All but one of the SLPs consented to use the headset microphone. To rule out any influence non-compliance with wearing the headset microphone might have had on the ASR, recognition rates were compared for the two groups: headset and tablet microphone. No difference in word recognition rate was observed between headset versus tablet microphone (accuracy – children: headset off = 53.2%, on = 56.3%; SLPs: off = 68.0%, on = 71.8%).

Speech-language pathologist's support was given if the children had difficulty controlling the recording buttons during game play (the researcher pressed the start/stop button when the child could not manage). To minimise the effects of the differing recording mechanisms/window lengths within the games and the varying ability of the participants to coordinate initiating recording and speaking on the ASR, participants were given up to a total of three production attempts per target word if the ASR did not accept their first production. That is, participants were presented with a new word either when the system recognised a production as correct or, the SLP manually progressed them after three unrecognised attempts by the ASR to avoid frustration. Automatic speech recognition–human agreement counted if there was a point-to-point match within those one to three attempts. The examiner counted a production as correct when all phonemes and, for multisyllabic words, the lexical stress were produced correctly.

Game descriptions

All the speech-controlled games tested in this study incorporated the mobile device version of CMU Sphinx speech recognition framework, PocketSphinx (SourceForge, 2018). PocketSphinx is an open source, lightweight speech recognition engine tuned specifically for handheld and mobile devices, thus making it fast and appropriate for use in speech-controlled games. It has been incorporated into a number of applications for children, for example Speech Adventure, a therapy tool for children with cleft speech (Rubin & Kurniawan, 2013). Its desktop version, Sphinx, has been used in literacy tools for children, for example Project LISTEN, an automated reading tutor that has been tested with both native and non-speakers of English. (Mostow & Aist, 2001; Poulsen, Hastings, & Allbritton, 2007). For the iOS games, PocketSphinx was adapted and compiled for use within an iOS environment.

We used the acoustic model in-built into PocketSphinx, trained with a speaker-independent adult American English speech corpus. Though the acoustic model has been trained with adult speech, studies comparing its performance with children at the phoneme production level found good correlations with manual scoring (Xu, Richards, & Gilkerson, 2014). However, when PocketSphinx and other ASRs such as Google speech were tested with a range of different forms of children's speech, their performance was found to decrease dramatically for continuous speech and long sentences; best results were obtained when the dictionary size was limited to single words and short phrases (Kennedy et al., 2017). The ASR dictionary was thus limited to ∼150 single words only. We did not create our own domain-specific acoustic model for use with PocketSphinx due to the lack of a sufficiently-sized children's speech corpus (TD and disordered) to train the model with.

To develop all of our speech-controlled games except Memory, we modified existing, open-source mobile games (IdeaMK, 2017; Minimal Games, 2018; Squadventure, 2018) by incorporating ASR (via PocketSphinx) into the gameplay. We developed Memory, on the other hand, from scratch (Parnandi et al., 2015). While playing these games, the child is prompted with a target word which they need to produce. The ASR is used to pick 10 candidate words from the dictionary that had the highest correctness probability of matching the speech production and score them from high to low probability (i.e. the *n*-best list). The game engine then compares the ASR output to the target word to verify if the production is correct. The child is then provided with real-time, automated feedback on their production by producing immediate consequences during gameplay that impacts their progress in the game.

In all the games, the SLPs can create word lists aligned with the therapy needs of the child; however here, a standard set of 30 words was used for all participants. The words ranged from one to four syllables in length and sampled most phonemes of English. All words were common nouns familiar to children. In the word lists, close phonetic neighbours were avoided to minimise ASR confusion due to the absence of contextual cues extracted from sentences, typically used to improve recognition. The target word presented to the child in each game is randomly selected from this word list and words repeated only after each has been shown to the child.

WordPop. As shown in Figure 1a, WordPop is a single-interaction mobile game akin to game mechanics

Figure 1. WordPop. (a) Instructions appear for how to record a production of the displayed word, (b) the target word appears and (c) the word pops and letters float towards the screen edge when the automatic speech recogniser deems it correct, and a new target word appears. Points are added for each letter leaving the screen.

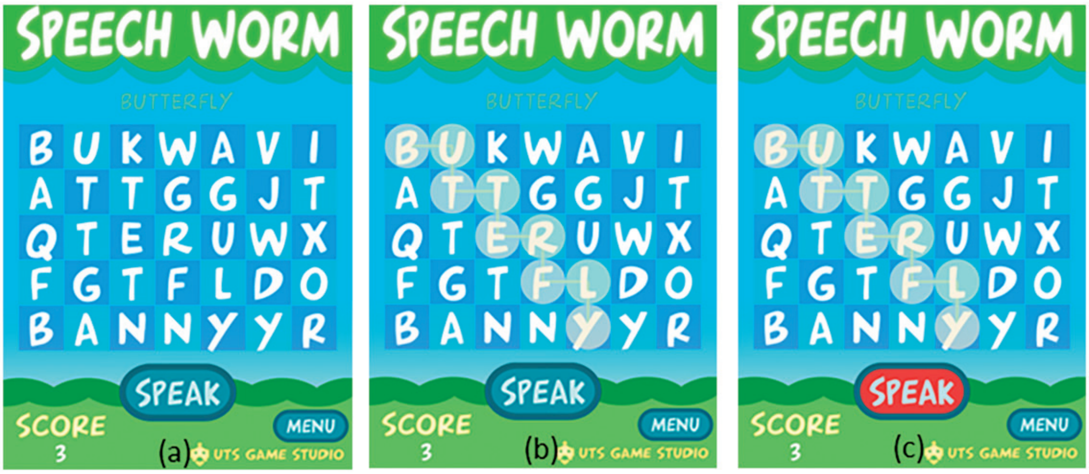

Figure 2. SpeechWorm. (a) A target word is displayed – "butterfly", (b) the child swipes or taps the letters in the grid to make the word and (c) by tapping on the SPEAK button to turn it red, the child can record their production of the word and tap again to stop recording and return the SPEAK button to blue. Points are added for each word recognised.

in popular games like Fruit Ninja. The goal is to get as many points as possible by "popping" the words. A word appears on the screen with letters in colourful bubbles (Figure 1b). Touching the screen starts the recording and ASR, after which the child says the word out loud. The child needs to hold the touch till their production is complete. If the ASR detects that the word was produced correctly in the recording, the bubbles break apart (Figure 1c). As they drift off the screen one point per letter is accrued. There are motivational sounds for the word popping and for the accruing points. If the ASR fails to recognise a word, then it can be attempted again indefinitely or a double tap on the screen to skip to the next word.

SpeechWorm. SpeechWorm is a word-search type game. The goal is to get as many points as possible by finding and saying the words. A word is displayed and, first, the child finds the letter string within the grid on the screen and swipes their finger across the letters in the correct order (Figure 2a). Once the letter string has been highlighted (Figure 2b), the child taps the blue "Speak" button under the grid to change it to red, activating the ASR. The child then records their word production and taps the button again to stop recording (Figure 2c). Points are given for each spoken word that the ASR recognises as correct. A motivating sound plays as the child touches each letter and when the ASR recognises a production. If the ASR fails to recognise a word, the child can attempt again indefinitely or double tap the screen to progress to the next word.

Whack-A-Mole. Whack-A-Mole is similar in style to the carnival game of the same name – two rows each of five cards face down which randomly flip over one at a time. The goal is to get 10 reward stars by saying the words. The flipped cards show pictures that prompt the child to say a specific word. Tapping a card – the whack element of Whack-A-Mole – stops it from flipping back over, displays a timer bar and starts the recording (Figure 3a). The recording is automatically stopped after 2 min. If the ASR detects that the speech production corresponds with the picture, the child is awarded a star and the card flips back over (Figure 3b). If the child does not produce the target correctly, the card still flips back over but they do not get a reward point. However, periodically, a bomb will appear on the flipped over card and tapping this card will cause the child to lose a star (Figure 3c); if a child has no stars, they cannot lose by tapping a bomb.

Asteroids. Asteroids is based on an open-source version of the retro game. The goal is to get as many

Figure 3. Whack-A-Mole. (a) The child has tapped on the target word and the yellow timer bar appeared to show how long the child has to say the word; (b) The target word was correctly spoken, and hence a green checkmark and star appeared; (c) A bomb card. Tapping on this will cause the child to lose a star. Images are from the Nuffield Dyspraxia Programme (NDP3, 2017).

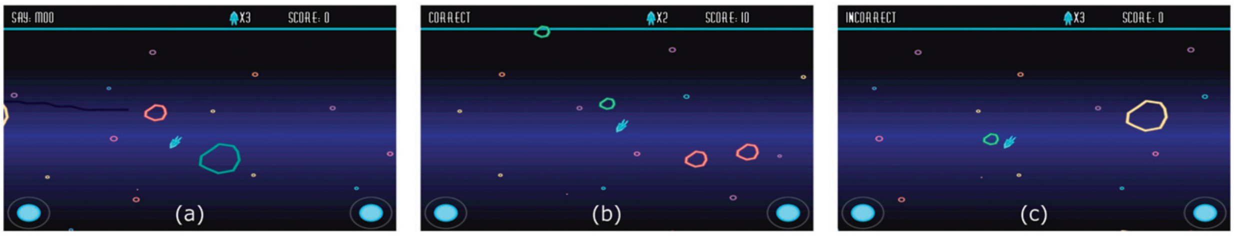

Figure 4. Asteroids. (a) The large yellow asteroid has changed to green to indicate that the game is listening for the target word; (b) The target word was correctly pronounced and the large asteroid broke into two small green asteroids; (c) The target word was spoken incorrectly and the large asteroid changed colour back to yellow.

points as possible and accrue lives by saying the words to break large asteroids up and avoiding being hit by an asteroid. Children move a continually shooting spaceship with the on-screen controls to shoot asteroids and break them up before they hit the ship. When they touch the large yellow asteroids with their fingertip, a target word is displayed and the asteroid changes colour to green (Figure 4a) and starts the recording and ASR. Once the ASR is activated, it times out if no speech is detected within 750 ms but if speech is detected, the recording is stopped after 2 min to allow the child to complete their production. If the ASR recognises that the speech production is the target word, the asteroid breaks into smaller pieces (Figure 4b). If the ASR detects that the speech production is not the target word, the asteroid changes back to the original yellow colour (Figure 4c). Children earn extra lives as they reach pre-determined point tiers; allowing an asteroid to collide with the ship causes the child to lose a life. Once all lives are gone, the game is over. It has also three levels of difficulty, which vary the speed of asteroid movement; only the easy level was trialled here.

Memory. Memory is an interactive version of the card game young children play. In the game, the child is presented with five pairs of images hidden behind bubbles; the goal is to find/match all pairs of stimuli (Figure 5a). On uncovering each bubble, the child has to press the record button that appears and record an utterance before they can uncover the next bubble (Figure 5b). Once the record button is pressed, the stop button shows up on the screen and the record button becomes inactive. After stopping, the play and record buttons are reactivated and hence the child can playback the recorded speech or record a new utterance after uncovering another bubble (Figure 5c). The game provides additional visual and audio cues with a flashing stop button to remind the child to stop as well as audio and text prompts for each image. This game is not integrated with an ASR but the child, parent or SLP can manually score accuracy of each speech production by tapping on stars at the bottom of the screen: a gold star indicates a good production, whereas a silver star a fair production. The child also receives generic written and audio encouragement from an animated character (a tabby cat) via expressions such as *Well done!* and *Excellent!*.

Data analysis

Questionnaire responses were summarised descriptively by question and by group. Non-parametric correlations were run to identify potential relationships between CAS impairment severity (i.e. PPC and PLSC) and responses to questions eliciting numeric ratings. To explore ASR–human agreement, first data was explored for normality using the Shapiro–Wilk test with no violations detected. Next, a one-way ANOVA was used to test for a group effect (SLP, TD and CAS) on the dependent measure of % ASR–human agreement for each participant. We also tested for a correlation between ASR–human agreement (%) and CAS severity, indexed by PCC, using the Spearman nonparametric test due to the variable of PPC being non-normally distributed. Finally, two repeated measures ANOVAs were used to test for the influence of the within-participants factors of (1) recording method (with automatic vs. manual stop, pooling data for games using each recording method) and (2) word length (1, 2 or 3–4 syllables; pooled across games) on ASR–human agreement across the three groups.

Figure 5. Memory. (a) Five pairs of images hidden behind bubbles. (b) The child has uncovered a bubble and tapped the record button, and hence can now record the word associated with the uncovered image. (c) The child has uncovered two pairs of matched images, receiving one gold star and one silver star in the process. Images are from the Nuffield Dyspraxia Programme (NDP3, 2017).

Result

Questionnaire

For rating-scale questions, Figure 6 shows the percentage of participants responding with "Yes", operationalised as a rating of 4 or 5 on the 5-point rating scale. For comment box responses, two researchers reached consensus on common themes identified across each groups' responses.

Children. Table I summarises the children's responses to the questionnaires, including abbreviated comments. Most children liked all of the games and could immediately play them independently or felt they would be able to with some practice. More than 50% of the children felt that they would like to keep playing all the games more, with the exclusion of Memory and Whack-A-Mole for the speech impaired (CAS) children. More than 50% of typically developing (TD) children found Asteroids, Memory and Whack-A-Mole interesting, and found all the games easy to play (rated 4–5 on the 5-point scale). The TD group showed no strong preference for any game, with Asteroids chosen most often as the best game (i.e. 30% of TD children). Children with CAS, however, showed a stronger preference for Asteroids; 70% of CAS children selected Asteroids as the best game. Typically developing children rated Whack-A-Mole, SpeechWorm and Memory most frequently as easy to play by (>80% responded yes to the easy-to-play question), whereas CAS children selected Asteroids, Whack-A-Mole and Memory as most easy to play (>70%). No game was selected as least-liked by more than 50% of children.

The main themes identified in responses of children were that they liked the rewards, the challenge, the game themes and having fun. Typically developing children rarely commented on the speech-controlled aspect of the games, whereas CAS children mentioned this as a positive feature of the four speech-controlled games. Children with childhood apraxia of speech also liked the playback feature, available only in Memory. The most common dislikes were that a game was too hard (e.g. in connection with the more complex gameplay in Asteroids), too easy, quickly became boring, was frustrating when the ASR did not recognise their speech and had no sound effects. They also found the record buttons difficult to manage, particularly in Memory and SpeechWorm games that required multiple screen taps. Multiple female TD children did not like the shooting theme in Asteroids although the predominantly male group of CAS children did like this theme. Children generally did not like the time bomb in the Whack-A-Mole game. They wanted the games to have "levels". Most children with CAS also needed support with reading in the three games using orthographic stimuli – WordPop, SpeechWorm and Asteroids. At least 80% of children thought that they would use these two games one or more times a week for their speech practice.

To explore whether any numeric ratings were correlated with speech impairment severity (i.e. PPC and PLSC) for the CAS children, we performed a series of nonparametric Spearman's correlations. The average rating across the five games was calculated for the questions probing whether the child liked the game, if they would have liked to keep playing, if they found it interesting, how easy it was to play and if they would use it for their speech therapy practice. Only responses to the question about using the games for practice were significantly correlated with PPC; although, this result would not survive Bonferroni correction (games for practice: PPC $\rho = 0.634$, $p = 0.049$, PLSC $\rho = -0.012$, $p = 0.973$; liked the game: PPC $\rho = -0.074$, $p = 0.84$, PLSC $\rho = -0.105$, $p = 0.773$; want to keep playing: PPC $\rho = 0.396$, $p = 0.257$, PLSC $\rho = -0.049$, $p = 0.894$; interesting: PPC $\rho = 0.566$, $p = 0.088$, PLSC $\rho = 0.222$, $p = 0.538$; easy: PPC $\rho = 0.587$, $p = 0.074$, PLSC $\rho = 0.171$, $p = 0.636$). Scatterplots are shown in Supplementary Figures 1 and 2.

Speech-language pathologists. Responses to the questionnaire are summarised in Table II, including abbreviated comments. Most SLPs liked all the games and could play them independently or felt they would with some practice. More than 50% said they would like to keep playing all games, except Asteroids. Most found all games except Asteroids interesting and easy, with most finding the more complex gameplay of Asteroids difficult to master.

The features that SLPs liked were the rewards, the challenge, the game themes, having fun, interactiveness, room for continued development within games, customisable word lists, potential for home practice and capability for modelling, recording and playing

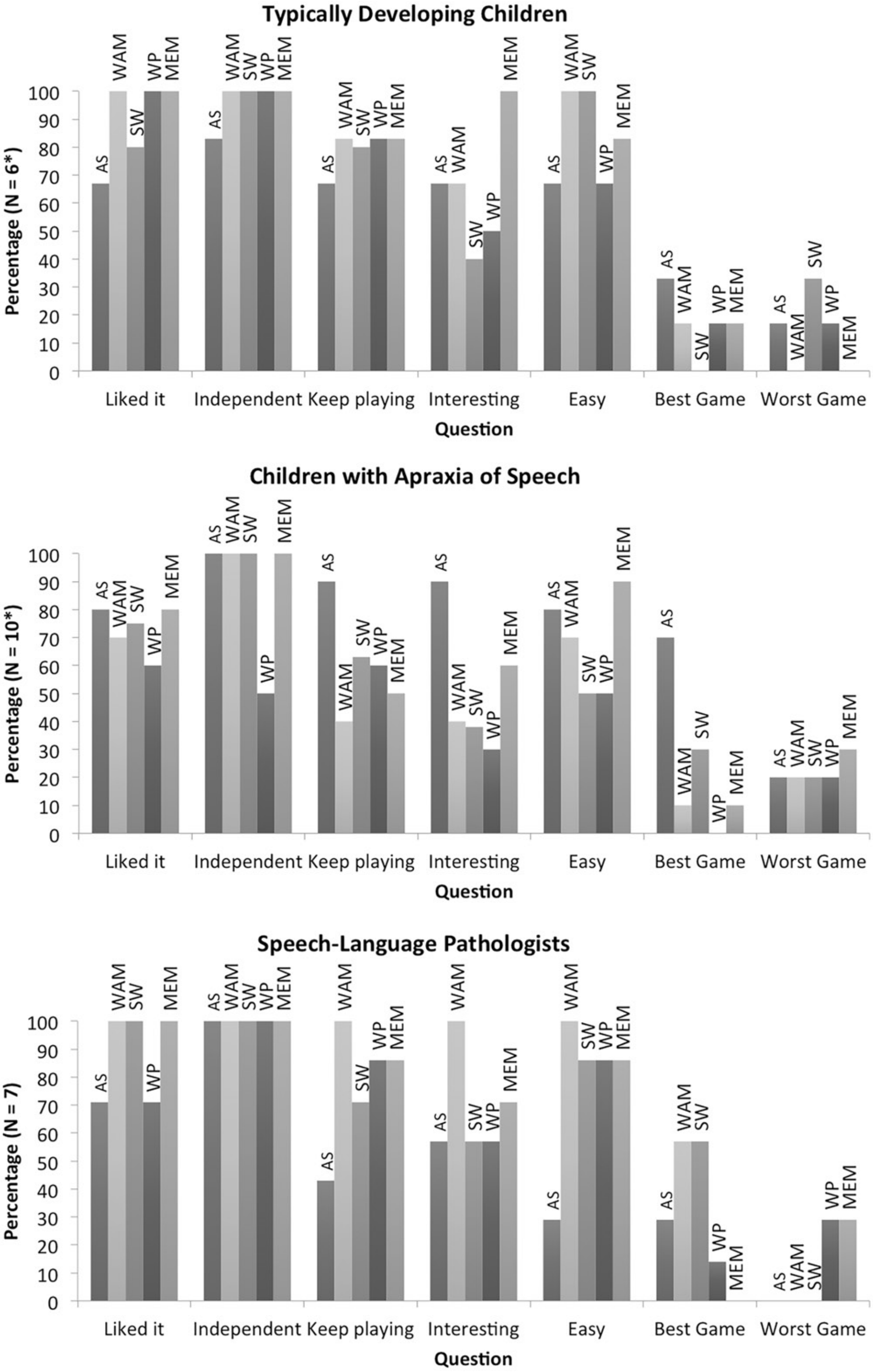

Figure 6. Percentage of TD children, children with CAS and SLP selecting "Yes" or 4 or 5 on a 5-point rating scale for the seven forced-choice questions in the questionnaire for each of the five games: Asteroids (AS), Whack-A-Mole (WAM), SpeechWorm (SW), WordPop (WP) and Memory (MEM). "Liked it" refers to Question 1: Did you enjoy using this game?, "Independent" to Questions 2: Did you need any help to play the game and Question 6: With some more practice, I think I could play the game by myself, "Keep playing" to Question 3: When you played the game, did you ever feel you wanted to play for longer?, "Interesting" to Question 4: How interesting/boring did you find the game?, "Easy" to Question 5: How easy was it to play the game?, "Best game" to Question A: Which game did you like the *most*?, "Worst game" to Question B: Which game did you like the *least*?.

back speech. They commented that picture stimuli, rather than orthographic stimuli, made the games accessible to both younger and older children. They liked the presentation and graphics in the SpeechWorm, WordPop and Whack-A-Mole games and the star chart and verbal praise from Memory. They liked the sense of competitiveness in most games as well as games that were not too difficult for children and could promote home practice with a high number of trials.

Table I. Questionnaire text responses from six TD children and 10 children with CAS regarding experiences with the five speech-controlled games.

	TD ($n=6$)					CAS ($n=10$)				
	Asteroids ($n=6$)	SpeechWorm ($n=5$)	Memory ($n=6$)	Whack-A-Mole ($n=6$)	WordPop ($n=6$)	Asteroids ($n=10$)	SpeechWorm ($n=8$)	Memory ($n=10$)	Whack-A-Mole ($n=10$)	WordPop ($n=10$)
Needed help with	Recording	Recording	Recording		Recording	Reading Recording	Reading Recording	Recording	Recording	Reading Recording
Features liked	Reward Challenge Theme Fun Speech	Theme Challenge Fun	Theme Reward Challenge	Theme Reward Challenge Fun	Theme Reward Challenge Fun Screen taps	Theme Reward Fun 2-Hand ctrl Screen taps	Theme Reward Challenge Fun Speech	Theme Reward Challenge Fun Speech Playback	Theme Reward Challenge Fun Speech Screen taps	Theme Reward Fun Speech
Features disliked	Hard Shooting 2-Hand ctrl No sound ASR	Boring ASR	Rec button Tabby's voice	Speed Bombs ASR No levels Boring	ASR Pop sound Boring	Few lives Speech Touching asteroids No effects Boring	ASR Record button No goal No sound Boring Too easy	Theme Not novel Record button Boring Too easy Too hard	ASR Bomb Boring Too easy Theme	Repetitive ASR Rec button
Would use for speech practice	–	–	–	–	–	80%	60%	33%	38%	43%
Would use for 4+ weeks	–	–	–	–	–	50%	33%	0%	17%	50%
Would use for 1–3 weeks	–	–	–	–	–	38%	67%	100%	67%	50%
Would use for <1 week	–	–	–	–	–	12.5%	0%	0%	17%	0%
Why liked best	Engaging Shooting	–	Easy Fun	Fun	Fun	Challenging Shooting Fun	Familiar Challenging Fun	–	–	–
Why liked least	Too hard Boring	No challenge Boring Not fun No feedback	–	–	Poor ASR	Too easy Needs improving	No challenge Needs improving	Too easy Would get boring	Bombs Poor ASR Tap and say	Too hard Pop slow

Table II. Questionnaire text responses from six SLPs regarding experiences with the five speech-controlled games.

	SLPs ($n=7$)				
	Asteroids ($n=7$)	SpeechWorm ($n=7$)	Memory ($n=7$)	Whack-A-Mole ($n=7$)	WordPop ($n=7$)
Thought children would need help with	Game controls Using both hands	Recording Game task sequence	Recording Game task sequence	Game task sequence	Recording Game rules
Features liked	Theme Reward Challenge Fun Ergonomic controls Room to develop Speech Great for home Custom word lists	Theme Reward Challenge Fun Presentation Good for practice Interactive Sound effects	Reward challenge Fun Praise from tabby Record/playback pictures Star chart	Reward Easy Fun Speech Interactive Not challenging Picture stimuli Range of words Presentation High trial number Time bomb	Reward Challenge Sound effects Easy Fun Presentation competitive Interactive
Features disliked	Hard 2-Hand control Not enough practice Speech feedback unclear Too fast Hard to see points	No chance to repeat when ASR says you're wrong Does not detect spelling errors Speech plus letter Time to trace word Boring No levels	Rec buttons Same pictures as Whack-A-Mole Not challenging Uncompetitive	Name of game Lag in count down box and tick display No modelling ASR accuracy Only 10 reward stars	Cloud image in background Can skip difficult words Sound effects while saying the next word Letter stimuli Recording method No goal
Appropriate for your clients?	86%	100%	100%	100%	100%
Why?	Appealing for boys If able to read 7+ yrs For phonology and articulation cases For milder cases	If able to read words Articulation cases Mild-moderate cases	Younger children (pictures)	Younger children (pictures) If ASR improves If slower speed for more severe cases	If able to read words If used pictures Suitable for younger and older children Phonology and articulation cases
Why liked best		Multimodal but not overwhelming Easy but fun Also develops literacy Fun		Graphics Feedback Easy Fun Unpredictable Good for a range of ages	Easiest Most fun
Why liked least	Hard Boring	Hard	Recording disrupts game play Not competitive or motivating	Name of game Count-down bar frustrating Needs more rewards	Distracting sound effects Not interesting Interface not user-friendly

There were a range of suggestions for improving specific games. Speech-language pathologists found the Asteroid game challenging with its two-hand control and suggested that it potentially would not allow enough practice trials per minute. They preferred games that gave multiple attempts at words when the ASR did not recognise a production and games with clear game goals, reward systems and multiple levels of difficulty or challenge. They did not like game features that allowed a child to play without saying any words at all, as in SpeechWorm and Asteroids, or to skip over words that might be perceived as difficult, as in WordPop, SpeechWorm and Asteroids.

The most preferred games for the SLPs were SpeechWorm and Whack-A-Mole because they perceived them as easy to play and fun. Speech-language pathologists also commented that SpeechWorm could be used to simultaneously work on literacy.

Accuracy of automatic speech recogniser

The results of ASR are summarised in Table III. As discussed earlier, all games except Memory incorporated ASR to enable speech control. The accuracy of the ASR was calculated on a total of 1816 word productions with 543 from all six TD children, 719 from nine of the 10 CAS children and 546 from all seven SLPs. Due to experimenter error, ASR–human agreement data were not available for one child with CAS.

Group effect. Using the percent point-to-point agreement between ASR and human judgement for each individual, regardless of stimulus item/type or recording method, there was a significant group effect on ASR agreement ($F(2,19)=7.293$, $p=0.004$, $\eta^2=0.434$). Least significant difference post-hoc testing showed that ASR–human agreement was significantly higher for SLPs (mean =72.1% agreement, SE =0.04) than both groups of children (SLP vs.TD: $p=0.002$, mean = 51.3%, SE = 0.04; SLP vs. CAS: $p=0.011$, mean = 57.5%, SE = 0.03); TD versu CAS, $p>0.5$. As summarised in Table III, the ASR recognised fewer productions as correct compared to the researchers (TD: 96% correct by human, 51% by ASR; CAS: 67% by human, 40% by ASR; SLP: 100% by human, 71% by ASR). It is worth noting that no significant correlation was detected between overall ASR–human agreement and CAS severity (PPC $\rho=-0.317$, $p=0.406$, PLSC $\rho=-0.567$, $p=0.112$).

Table III. Automatic speech recogniser decision against gold-standard human decision of word production correctness (correct, incorrect).[a]

TD		Human		
		Incorrect	Correct	Total
ASR	Incorrect	10	258	268
	Correct	3	272	275
	Total	13	530	543
		Specificity (%)	Sensitivity (%)	
		77	51	
CAS		Human		
		Incorrect	Correct	Total
	Incorrect	180	254	434
	Correct	59	226	285
	Total	239	480	719
		Specificity (%)	Sensitivity (%)	
Agreement		75	47	
SLP		Human		
		Incorrect	Correct	Total
	Incorrect	0	157	157
	Correct	0	389	389
	Total	0	546	546
		Specificity (%)	Sensitivity (%)	
Agreement		–	71	

[a]Data are pooled across all participants within each group. Shaded values are used in calculating the percent agreement value.

Recording method. Using repeated measures ANOVA, we explored the potential effect of the recording method (automatic vs. manual stop) within the app on ASR–human agreement by group. There was no main effect of group (F(2,19) = 2.563. $p = 0.103$, $\eta^2 = 0.212$). It should be noted that percent ASR–human agreement for each child under each recording context was entered in this analysis, and hence the group analysis here was computed on different values to the group analyses above. There was a highly significant effect of recording method ($F(1,19) = 37.560$, $p < 0.001$, $\eta^2 = 0.664$). The interaction term was not significant (F(2,19) = 1.361, $p = 0.280$, $\eta^2 = 0.125$). The highest ASR–human agreement was observed for the manual-stop recording method used in WordPop and SpeechWorm, with mean ASR–human agreement of 65.3% (SE =3.4) for CAS, 66.7% (SE =4.2) for TD and to 81.1% (SE =3.9) for SLPs. In contrast, the mean accuracy for the automatic-stop recording method of Asteriods and Whack-A-Mole was 47.9% (SE =5.7) for CAS, 37.6% (SE =7.0) for TD and to 47.7% (SE =6.5) for SLPs.

Word length. A second repeated measures ANOVA was used to explore the potential effect of word length, in syllables, on ASR–human agreement accuracy by group. The main effect of group was significant ($F(2,19)$ = 8.316, $p = 0.003$, η^2 = 0.467) with ASR–human agreement higher for SLPs than TD or CAS children ($p = 0.003$ and 0.029, respectively). The main effect of word length and the interaction term were not significant ($F(1,19) = 0.560$, $p = 0.463$, $\eta^2 = 0.029$ and $F(2,19) = 1.825$, $p = 0.188$, $\eta^2 = 0.161$, respectively).

Discussion

This feasibility study surveyed the experiences of speech impaired and TD children and SLPs when using five different prototype speech-controlled games to assess the utility of such games for augmenting home-based speech practice. The development of such games is based on the premise that interactive speech-controlled games will encourage children to practice their speech therapy exercises over the long-term as the games will (1) provide the children with immediate feedback on their productions and (2) better engage children.

As hypothesised, both children and SLPs generally reported finding the games interesting and fun. They specifically liked reward systems (e.g. points), games that were challenging and the option of multiple levels of difficulty. The two most-liked games both had integrated ASR, indicating that ASR was not considered a major negative feature although all participants expressed frustration when they felt it did not recognise their speech attempts. Most SLPs indicated that they would use the games in therapy and most children during home practice and would prefer to use the games over paper-and-pencil activities. As expected, children and SLPs differed on their most favourite game.

Game play

The Asteroids game was reported to be the most liked of the speech-controlled game by children with CAS. It differed from the other games in terms of its continuous and fast-paced gameplay, lacking the frequent pauses in the other games. Its popularity suggests that more complex and fast-paced games than typically preferred by SLPs may be better at engaging and motivating children. This creates a conflict for future game development, however, and the challenge will be in ensuring there is sufficient focus on speech productions and no sacrifice is made to the number of speech production opportunities all the while creating an engaging gameplay experience.

Our results indicate that it is important to provide children with a range of games to suit different individual preferences. For example, boys typically reported that they enjoyed Asteroids, whereas girls tended to prefer non-shooting games. In contrast, age did not appear to be a factor in which games children preferred, with all ages enjoying the most difficult game to play, Asteroids. Anecdotally, and counter to researchers' concerns, all children were able to understand and execute the complex game play of Asteroids (e.g. the two-hand control of the spaceship's position and direction for shooting and avoiding asteroid collisions) with minimal assistance from the researchers. In contrast, two of the SLPs struggled with the gameplay and opted to stop playing before reaching 15 word-production attempts. This difference may reflect the intuitiveness of touch screens and the greater experience children have with these game types over adults. It is possible that the dual hand and speech control for this game will distract children from focussing on the quality of their speech productions. However, this game may be helpful in improving automaticity in production later in therapy, making the speech task more challenging through distractors and time pressures.

Headset versus tablet microphone

Not all participants were willing to wear the headset microphone. A 3% difference in ASR–human agreement for microphone being worn versus not worn was found for both the children and the SLPs. These results are positive, given that children often dislike wearing such headsets. They seem to reflect the increased quality of inbuilt microphones in tablets and iPads now available combined with the quiet testing environment.

Recording method

The apps used two different recording methods: automatic or manual stop of recording. For Asteroids and Whack-A-Mole, a recording window started when the stimulus (asteroid or picture) were touched and ended automatically after a fixed time duration or correct response. For WordPop, the participant touched anywhere on the screen to start recording and held the touch until completion of the production. For SpeechWorm and Memory, the participant touched the record icon to start recording and touched it again to stop recording. This subtle difference is important because it varied the complexity of game controls and the degree of time pressure on responding. If the child had difficulty controlling the recording mechanisms, the possibility was (1) that they would begin speaking simultaneously with the record button being pressed, resulting in the beginning of their production being clipped or (2) that they would pause once the recording window (with an automatic stop) had started, resulting in the end of their production being clipped. In either situation, the ASR would not process the complete production, negatively influencing its performance.

We found that the recording method used by the games had a significant impact on ASR performance. In this study, we unfortunately did not store the tablet recordings for later analysis although the capability for SLPs to access these recordings remotely via a secure server can be built in (Parnandi et al., 2015). Inferior ASR performance was observed for Whack-A-Mole and Asteroids; both of these games required a single touch to initiate recording and ended automatically, whereas WordPop and SpeechWorm required a manual action to both start and stop recording. Based on our prior experience, the lower ASR–human agreement was related to difficulty coordinating the timing of button presses with onset/offset of speech resulting in clipped speech productions. The children reported having difficulty with the recording method in all games.

Though it is possible that sufficient practice could improve the recording quality, ASR performance would be better enhanced by having discrete actions to initiate and stop recording, whether controlled by the child or the supervising adult. However, this method would interfere with gameplay in the Asteroids game where an alternative recording method is suggested: (1) a single screen tap to activate an asteroid, (2) a short delay before the target word is presented, to avoid clipping of speech onset, (3) a longer waiting time before the ASR times out, perhaps 2 s, to accommodate slower response times in speech impaired children and (4) recording until production has ended. Clearly, special consideration needs to be given on how the recording method is integrated into gameplay to minimise the impact on both the ASR–human agreement and flow of the game.

ASR–human agreement

Analyses indicated that the results of ASR were more accurate for SLPs (i.e. adults) than for children, speech impaired or not. This can be partially attributed to the clipping of speech samples for the children as observed earlier. However, ASR also still exhibits higher error rates for children as less progress has been made in recognising speech produced by children (Shivakumar, Potamianos, Lee, & Narayanan, 2014). The performance of PocketSphinx, the specific ASR integrated into our games, has been shown to drop when used with children's speech (Kennedy et al., 2017). This drop in performance is because PocketSphinx and most currently available ASRs are trained with adult speech, which differs from children's speech. Children's speech has been shown to have a higher variability in vocal tract length, formant frequency and pronunciation than adult speech (Fitch & Giedd, 1999; Lee, Potamianos, & Narayanan, 1999). In addition, less work has been done on the processing of children's speech, resulting in a lack of appropriately sized databases of children's speech to develop ASRs specific for children.

However, the focus of this study was to test the feasibility of speech-controlled games incorporating current off-the-shelf ASRs; not assessing the ASR performance in detail. We did not develop our own domain-specific ASR for these games, nor did we use an acoustic model trained on children's speech. To improve ASR–human agreement, the dictionary size was restricted to a small set of words (e.g. we constrained the ASR in SpeechWorm and SpeechPop to a dictionary of 129 words and used the *n*-best list to identify 10 candidate words from the set with the highest correctness probability). Also, close phonological neighbours were avoided in the word list as ASRs have difficulty differentiating minimally different single words (e.g. cat/bat) due to their similar acoustic features and the absence of contextual cues (Goldwater, Jurafsky, & Manning, 2010). Contextual cues, extracted from adjacent words in sentence-based speech, are used by ASRs to differentiate between words with similar acoustic features. It is quite likely that certain types of consonants, stress patterns or phoneme placements, which cause greater ASR confusion, could be making greater contributions to the error rates and biasing the results. Further studies are underway to explore this.

It is worth mentioning that the differences between the ASR and the human judgements could have occurred due to variations in human perceptual judgements. Human judgements could have been affected by factors such as word frequency, whereas the ASR might not have these types of influences. It would also be interesting to test ASR–human agreement with lay people versus SLPs, which would focus more on intelligibility perhaps and the ecological validity of the system.

Ultimately, integration of ASR was not a deciding factor in selecting the most favourite game, given that the two most popular games used ASR and in fact the child with the most severe speech impairment selected an ASR-integrated game. This was true despite the fact that many children commented that they were frustrated by low recognition rates at times. This offers a promising suggestion for the viability of speech-controlled games for speech therapy. In the future, the threshold for ASR–human agreement below which the game no longer becomes acceptable should be determined. It is also worth exploring how to better utilise opportunities afforded by ASR errors in the gameplay to elicit multiple responses from the child without increasing their frustration.

Limitations and future directions

There were several limitations to this study. The participant sample was small, especially within the TD subgroup; however, no new themes arose in survey responses in the final two to three participants, suggesting that saturation was reached for the open-ended responses. A larger sample would have permitted more detailed statistical analyses of game features and their interactions. A larger sample of CAS children would have also allowed a more rigorous analysis of the impact of speech impairment severity on ASR–human agreement and the appropriateness of speech-controlled games for practice across the severity range.

The games tested represented a limited set of variations on game features. With the development of more speech-controlled games, a clearer profile of most desirable features will emerge and we may identify specific game features that best suit specific age and gender groups. Our analysis does reveal that recording method had a substantial effect on ASR performance – updated versions of the games need to address this. We recommend that future games allow SLPs to individualise word lists; sample lists and very clear instructions on what factors affect ASR–human agreement are essential. Then the SLP can decide if tablet-based ASR-based feedback is going to be appropriate for their client(s). We also recommend that games provide the option of ASR/no ASR so that those children (or word sets) with low ASR–human agreement might still have the opportunity to engage in fun high-intensity practice without the risk of inaccurate ASR-based feedback.

Although ASR performance has improved dramatically over the past 10 years, it clearly still performs better for adults than child. Due to the lack of a children's speech corpus (TD and/or disordered) of sufficient size, we did not create an acoustic model for children to integrate into our ASR, PocketSphinx. Further study is required to understand the factors that influence the ASR–human agreement and help improve the word recognition rates for child speech. However, the similarity in word error rates across children with and without speech impairment is encouraging and suggests that applications for speech therapy can utilise ASR algorithms trained on larger, more accessible databases of TD child speech.

Based on the feedback collected in this feasibility study, we are redeveloping our games to incorporate these preliminary findings for use in a longer study. Questions that can be explored include (1) the treatment efficacy of the games over longer periods (with and without ASR); (2) in-depth analysis of factors that influence agreement between ASR and human judgement (e.g. consonant types, stress patterns, word confusions and word repetitions); (3) how the game should be prescribed (with/without ASR, alternating between the two) and (4) the impact of inaccurate and accurate ASR feedback on treatment outcomes. If the ASR–human agreement can reach an acceptable level, it will be feasible to use the ASR-based measures of a child's performance to dynamically adjust game features such as difficulty or speed in an automated fashion.

Conclusion

The findings of our feasibility study suggest that children prefer games that mimic popular gameplay, with challenging tasks, rewards and multiple levels of difficulty to sustain engagement. The children were receptive to speech control of gameplay but expressed some degree of frustration when their speech was not recognised by the game. Careful design of recording functions will likely maximise ASR to acceptable levels. Our findings on ASR–human agreement, as well as evidence from motor-learning studies (Maas et al., 2008; Newell, Carlton, & Antoniou, 1990), suggest that speech-controlled games are likely best for children who are able to achieve some correct responses and who have an internal reference of the speech target that allows them to self-correct errors; particularly when used for home practice. It is also important to recognise that children and SLPs may differ in their opinion of what makes a game fun and engaging.

Declaration of interest

No potential conflict of interest was reported by the authors.

Funding

This work was made possible by NPRP grant #[8-293-2-124] from the Qatar National Research Fund (a member of Qatar Foundation). Ballard was supported by Australian Research Council Future Fellowship FT120100355. The statements made herein are solely the responsibility of the authors.

ORCID

Beena Ahmed http://orcid.org/0000-0002-1240-6572
Chek Tien Tan http://orcid.org/0000-0002-1023-8819
Ricardo Gutierrez-Osuna http://orcid.org/0000-0003-2817-2085
Kirrie J. Ballard http://orcid.org/0000-0002-9917-5390

References

American Speech-Language-Hearing Association. (2017). Apps for speech-language pathology and practice. Retrieved from http://www.asha.org/SLP/schools/Applications-for-Speech-Language-Pathology-Practice/

Balbus Speech. (2017). Speech4Good. [Mobile application software]. Retrieved from http://speech4good.com/features/

Bälter, O., Engwall, O., Öster, A.M., & Kjellström, H. (2005). *Wizard-of-Oz test of ARTUR: A computer-based speech training system with articulation correction.* Paper presented at the Proceedings of the 7th international ACM SIGACCESS conference on Computers and accessibility, Baltimore, MD, USA.

Expressive Solutions. (2018). ArtikPix (version 3.1.1). [Mobile application software]. Retrieved from http://expressive-solutions.com/artikpix/

Fitch, W.T., & Giedd, J. (1999). Morphology and development of the human vocal tract: A study using magnetic resonance imaging. *The Journal of the Acoustical Society of America*, *106*, 1511–1522. doi:10.1121/1.427148

Ganzeboom, M., Yılmaz, E., Cucchiarini, C., & Strik, H. (2016). On the development of an ASR-based multimedia game for speech therapy: Preliminary results. In *Proceedings of the 2016 ACM Workshop on Multimedia for Personal Health and Health Care*. (pp. 3–8). ACM. Retrieved from https://dl.acm.org/citation.cfm?id=2985771

Goldwater, S., Jurafsky, D., & Manning, C.D. (2010). Which words are hard to recognize? Prosodic, lexical, and disfluency factors that increase speech recognition error rates. *Speech Communication*, *52*, 181–200. doi:10.1016/j.specom.2009.10.001

IdeaMK. (2017). Whack A Mole (version 1.8). [Mobile application software]. Retrieved from https://play.google.com/store/apps

Jamieson, D.G., Kranjc, G., Yu, K., & Hodgetts, W.E. (2004). Speech intelligibility of young school-aged children in the presence of real-life classroom noise. *Journal of the American Academy of Audiology*, *15*, 508–517. doi:10.3766/jaaa.15.7.5

Kennedy, J., Lemaignan, S., Montassier, C., Lavalade, P., Irfan, B., Papadopoulos, F., … Belpaeme, T. (2017). Child speech recognition in human-robot interaction: evaluations and recommendations. In*Proceedings of the 2017 ACM/IEEE International Conference on Human-Robot Interaction* (pp. 82–90).ACM. Retrieved from https://dl.acm.org/citation.cfm?id=3020229

Lan, T., Aryal, S., Ahmed, B., Ballard, K., & Gutierrez-Osuna, R. (2014). *Flappy voice: An interactive game for childhood apraxia of speech therapy.* Paper presented at the Proceedings of the first ACM SIGCHI annual symposium on Computer-human interaction in play. Retrieved from https://dl.acm.org/citation.cfm?id=2661305

Laureate Learning Systems. (2014). TalkTime with Tucker. [Mobile application software]. Retrieved from http://www.laureatelearning.com/products/descriptions/talktdesc.html

Lee, S., Potamianos, A., & Narayanan, S. (1999). Acoustics of children's speech: Developmental changes of temporal and spectral parameters. *The Journal of the Acoustical Society of America*, *105*, 1455–1468. doi:10.1121/1.426686

Little Bee Speech. (2018). b Station app for iPad & iPhone. [Mobile application software]. Retrieved from http://littlebeespeech.com/apps.php

Lopes, M., Magalhães, J., & Cavaco, S. (2016). A voice-controlled serious game for the sustained vowel exercise. In *Proceedings of the 13th International Conference on Advances in Computer Entertainment Technology* (p. 32). ACM. Retrieved from https://dl.acm.org/citation.cfm?id=3001807

Luce, P.A., & Pisoni, D.B. (1998). Recognizing spoken words: The neighborhood activation model. *Ear and Hearing*, *19*, 1–36. doi:10.1097/00003446-199802000-00001

Maas, E., Robin, D.A., Hula, S.N.A., Freedman, S.E., Wulf, G., Ballard, K.J., & Schmidt, R.A. (2008). Principles of motor learning in treatment of motor speech disorders. *American Journal of Speech-Language Pathology*, *17*, 277–298. doi:10.1044/1058-0360(2008/025)

Micro Video Corporation. (2014). Video Voice. [Mobile application software]. Retrieved from http://www.videovoice.com/default.htm

Minimal Games. (2018). Asteroids (version 1.7). [Mobile application software]. Retrieved from https://play.google.com/store/apps

Mostow, J., & Aist, G. (2001). Evaluating tutors that listen: An overview of Project LISTEN. In K.D. Forbus & P.J. Feltovich (Eds.), *Smart machines in education* (pp. 169–234). Cambridge, MA: AAAI Press/The MIT Press.

Murray, E., McCabe, P., & Ballard, K.J. (2014). A systematic review of treatment outcomes for children with childhood apraxia of speech. *American Journal of Speech-Language Pathology*, *23*, 486–504. DOI: doi:10.1044/2014_AJSLP-13-0035

NDP3. (2017). Nuffield Dyspraxia Programme. Retrived from https://www.ndp3.org/

Newell, K., Carlton, M., & Antoniou, A. (1990). The interaction of criterion and feedback information in learning a drawing task. *Journal of Motor Behavior*, *22*, 536–552. doi:10.1080/00222895.1990.10735527

Parnandi, A., Karappa, V., Lan, T., Shahin, M., McKechnie, J., Ballard, K., … Gutierrez-Osuna, R. (2015). Development of a remote therapy tool for childhood apraxia of speech. *ACM Transactions on Accessible Computing (TACCESS)*, *7*, 1–23. doi:10.1145/2776895

Poulsen, R., Hastings, P., & Allbritton, D. (2007). Tutoring bilingual students with an automated Reading Tutor that listens. *Journal of Educational Computing Research*, *36*, 191–221. doi:10.2190/A007-367T-5474-8383

Pratt, S.R., Heintzelman, A.T., & Deming, S.E. (1993). The efficacy of using the IBM speech viewer vowel accuracy module to treat young children with hearing impairment. *Journal of Speech and Hearing Research*, *36*, 1063–1074. doi:10.1044/jshr.3605.1063

Rubin, Z., & Kurniawan, S. (2013). Speech adventure: Using speech recognition for cleft speech therapy. In *Proceedings of the 6th International Conference on Pervasive Technologies*

Related to Assistive Environments. (p. 35). ACM. Retrieved from https://dl.acm.org/citation.cfm?id=2504373

Ruggero, L., McCabe, P., Ballard, K., & Munro, N. (2012). Paediatric speech-language pathology service delivery: An exploratory survey of Australian parents. . *International Journal of Speech-Language Pathology*, *14*, 338–350.

Shivakumar, P.G., Potamianos, A., Lee, S., & Narayanan, S. (2014). Improving speech recognition for children using acoustic adaptation and pronunciation modeling. In *WOCCI* (pp. 15–19). Retrieved from https://www.isca-speech.org/archive/wocci_2014/wc14_015.html

Shriberg, L.D., & Kwiatkowski, J. (1982). Phonological disorders III: A procedure for assessing severity of involvement. *Journal of Speech, Language, and Hearing Research*, *47*, 256–270.

Smarty Ears. (2017). Apraxiaville. [Mobile application software]. Retrieved from http://smartyearsapps.com/service/apraxia-ville/

Smarty Ears. (2017). Articulate it. [Mobile application software]. Retrieved from http://smartyearsapps.com/service/articulate-it/

SourceForge. (2018). CMU PocketSphinx. [Mobile application software]. Retrieved from http://sourceforge.net/projects/cmusphinx/

Speech With Milo. (2017). Speech with Milo – Apps for speech therapy. [Mobile application software]. Retrieved from http://www.speechwithmilo.com/

Squadventure. (2018). Word Pop: Endless brain game. [Mobile application software]. Retrieved from http://itunes.apple.com

Tactus Therapy Solutions. Speech Flipbook. [Mobile application software]. Retrieved from http://tactustherapy.com/apps/speechflipbook/

Tan, C.T., Johnston, A., Ferguson, S., Ballard, K., & Bluff, A. (2014). *Retrogaming as visual feedback for speech therapy.* Paper presented at the in Proceedings of SIGGRAPH Asia 2014 MGIA, New York.

Tiga, T. (2011). Tiga talk speech therapy games. [Mobile application software]. Retrieved from http://tigatalk.com/app/

Xu, D., Richards, J.A., & Gilkerson, J. (2014). Automated analysis of child phonetic production using naturalistic recordings. *Journal of Speech, Language, and Hearing Research*, *57*, 1638–1650. doi:10.1044/2014_JSLHR-S-13-0037

Supplementary material available online