Apraxia World: A Speech Therapy Game for Children with Speech Sound Disorders

Adam Hair*
Texas A&M University
College Station, TX, USA
adamhair@tamu.edu

Penelope Monroe*
University of Sydney
Sydney, NSW, Australia
penelope.monroe@sydney.edu.au

Beena Ahmed
University of New South Wales
Sydney, NSW, Australia
Texas A&M University at Qatar
Doha, Qatar
beena.ahmed@unsw.edu.au

Kirrie J. Ballard
University of Sydney
Sydney, NSW, Australia
kirrie.ballard@sydney.edu.au

Ricardo Gutierrez-Osuna
Texas A&M University
College Station, TX, USA
rgutier@cse.tamu.edu

ABSTRACT
This paper presents Apraxia World, a remote therapy tool for speech sound disorders that integrates speech exercises into an engaging platformer-style game. In Apraxia World, the player controls the avatar with virtual buttons/joystick, whereas speech input is associated with assets needed to advance from one level to the next. We tested performance and child preference of two strategies for delivering speech exercises: *during* each level, and *after* it. Most children indicated that doing exercises after completing each level was less disruptive and preferable to doing exercises scattered through the level. We also found that children liked having perceived control over the game (character appearance, exercise behavior). Our results indicate that (i) a familiar style of game successfully engages children, (ii) speech exercises function well when decoupled from game control, and (iii) children are willing to complete required speech exercises while playing a game they enjoy.

Author Keywords
Games for health; speech therapy; speech sound disorders; childhood apraxia of speech.

ACM Classification Keywords
- **Social and professional topics~People with disabilities**
- Social and professional topics~Children
- **Applied computing~Consumer health**
- Human-centered computing~Natural language interfaces

IDC '18, June 19–22, 2018, Trondheim, Norway

ACM ISBN 978-1-4503-5152-2/18/06…$15.00
https://doi.org/10.1145/3202185.3202733
*A. Hair and P. Monroe contributed equally to this work.

INTRODUCTION
Speech sound disorders (SSDs) can affect language production and speech articulation in children, leading to serious communicative disabilities [4]. Estimates for the prevalence of SSDs in children vary; some suggest between 2% and 25% of children aged 5-7 years may be affected [21], while others estimate values closer to 1% of the primary-school-aged population [26]. Regardless of their exact prevalence, SSDs can have potentially devastating effects on a child's communication development [10]. Fortunately, children can reduce symptoms and improve speech skills by working closely with a speech language pathologist (SLP) [4]. To be effective, these treatments must be "frequent, high-intensity, individualized, and naturalistic" [25]. However, scheduling appointments with SLPs can be difficult, especially for children who live far from clinics [35, 44, 45]. Thus, clinic-based intervention typically must be supplemented with considerable home practice. Previous work indicates that remote digital sessions can be as effective as clinic-based sessions [6]. To alleviate the repetitive nature of frequent intense practice, however, these computerized therapies must be engaging.

A promising strategy to increase engagement is to deliver the speech exercises through mobile games. Accordingly, a number of game-like applications for speech therapy have been developed (e.g., Apraxiaville [40], Tiga Talk [46], Tabby Talks [31, 32], Articulation Station [23], ArtikPix [7], Pocket SLP [33]), though few provide feedback on speech productions. Among those that do, Tabby Talks [31, 32] combines (i) a mobile game that embeds speech exercises into a "memory/concentration" game where the goal is to find all pairs of identical cards in a deck, and (ii) an automatic speech recognition (ASR) engine running on a remote server that scores each individual production from the child [38]. In a pilot study [31], Tabby Talks was well received by parents, SLPs, and the children themselves, though feedback also suggested that the intervention needed more game-like features to increase the player's interest,

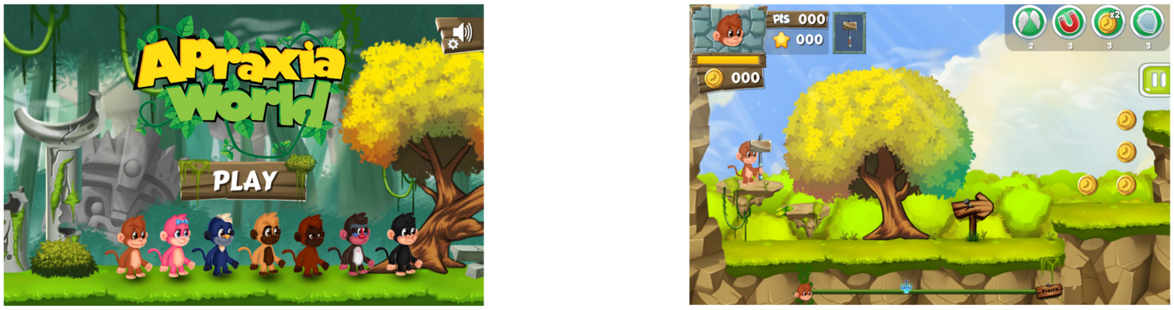

Figure 1 (a) Start screen showing all of the available characters. Players start with the monkey on the far left as the default (b) On-screen information shown to players: collectibles and health in the top left, available power-ups in the top right, and a progress bar in the lower center

especially for younger children. A second area for improvement in Tabby Talks was in terms of providing real-time feedback on productions, which was not possible with the remote ASR engine due to transmission and computation delays. To address these concerns, we have developed Apraxia World, a speech-therapy game constructed on top of a full-fledged, two-dimensional platformer game, which will later be coupled with a mobile ASR engine capable of providing real-time feedback on productions. In Apraxia World, the player guides an avatar (the cheerful monkey character shown in Figure 1) through a multi-level world where the goal is to collect assets while avoiding enemies and traversing an obstacle course.

This paper describes the gaming and therapy elements of Apraxia World, with special emphasis on how to integrate speech production into the game[1]. In Apraxia World, the player controls the avatar with standard inputs (virtual buttons and joystick), and speech input is tied to assets that the player must collect in order to advance from one level to the next. By associating speech production with the assets, players are able to anticipate and control when speech exercises appear, and the speech exercises do not detract from the gameplay or interrupt the player while executing complicated moves.

We validated Apraxia World through a pilot study with 14 children with SSDs (4-12 years old) and 7 typically-developing (TD) children (5-12 years old). This diverse population allowed us to gather feedback from children with varying exposure to speech therapy and their perception of how the speech exercises impacted gameplay. Specifically, we examined two strategies for integrating the speech exercises into the game, a *during-game* condition where the exercises were distributed throughout each level, and an *after-game* condition where the exercises were presented after finishing a level. Each child played both versions of Apraxia World and answered corresponding questionnaires on enjoyability, preference, improvements, etc. We also examined child engagement with Apraxia World based on qualitative questionnaire responses.

The remaining parts of this paper are organized as follows. First, we provide background information on SSDs and review related work on speech-driven games and game-based therapy. Next, we describe the game, the integration of speech exercises, and the (manual) assessment of productions. We then outline the experimental methods, including participant recruitment and study protocol, followed by the results from a pilot study with SSD children and TD children. The paper concludes with a discussion and directions for future work.

BACKGROUND

The term speech sound disorder (SSD) describes a collection of difficulties with perception and/or production of individual speech sounds that affect a person's ability to produce intelligible speech [4]. SSDs that affect the production of the correct form of sounds are associated with motor-based or structural disorders (e.g., childhood apraxia of speech (CAS) or cleft palate, respectively) and are considered to be articulation disorders. SSDs that affect the functional employment of sounds (i.e., when the sounds produced are correct in form but not in usage, for example, a person may say 'dar' for 'car') are considered to be disorders involving the individual's phonological representation of sounds and/or speech segments. These speech difficulties are often overcome with regular and frequent practice [4], the repetitive nature of which makes speech therapy an excellent candidate for game integration.

Speech-driven games

In the context of gaming, speech input has been used to improve accessibility [29], novel interaction [13], physical therapy [20], speech therapy [19], and social skill development [27]. In previous speech-integrated games, the player's voice [5, 11, 29] or vocal features [19, 24, 41] have been used for game control. However, this model limits the choice of game to those slow enough for the player to produce the correct voice command, and impedes gameplay

[1]As will be discussed in the Game Design section, speech assessments in the present study were conducted by an SLP during gameplay rather than by a mobile ASR engine. This allowed us to isolate the game aspects of Apraxia World from issues pertaining to mobile ASR performance, which will be addressed in a separate publication.

if players struggle to produce command words. Furthermore, once a word-to-input mapping has been established, it is difficult to change the word without causing confusion or increasing cognitive load by making players keep track of new command words.

Cai et al. [5] took a different approach for using voice within an arcade-style game. The authors implemented a version of Tetris where voice commands unlock Tetromino (Tetris piece) rotation, rather than using the speech to directly move the piece; this allowed words to be reinforced without slow speech dramatically hindering gameplay. Researchers have also examined non-verbal features as inputs for games; common features include pitch changes or vowel sounds. Sporka et al. [41] designed a version of Tetris where players moved and rotated the Tetromino with pre-defined pitch patterns. They later extended their study of pitch as an input by comparing verbal and non-verbal commands for driving a radio-controlled car [42]. In both studies, users preferred the non-verbal commands due to ease of use and quick response.

The Vocal Joystick [14] maps pitch, power, and vowel quality to computer mouse movements. In tests, users quickly learned how to use the Vocal Joystick and found it less frustrating than using command words. In later work, Harada et al. [15] used non-verbal inputs for four different games, where game-specific commands were mapped to vowels and pitch intensities. They found that system processing time was significantly shorter for non-verbal commands, which is ideal for quick arcade-style games. House et al. [16] further expanded upon the idea of the Vocal Joystick by implementing a 5 degrees-of-freedom control mechanism for a robotic arm moving in three-dimensional space. Vowel sounds have also been used to control retro-style games [43].

Automatic speech recognizers (ASR) are often found in speech-input systems, but they tend to struggle with children's speech. When ASR frameworks are tested with different forms of children's speech, performance decreases dramatically for continuous speech and long sentences as compared to adult speech, and the best results come from limiting the dictionary to single words and short phrases [18]. Speech patterns are typically harder to identify in children's speech due to large variations in vocal tract length, formant frequency and pronunciation quality [9, 22]. Additionally, even when ASR systems perform well with TD speech, they struggle with SSD speech [36].

Game-based therapy

Games have been evaluated for a variety of therapy applications across many disciplines. For example, in a recent IDC paper, Alessandrini et al. [2] developed a collaborative storytelling application to engage children with autism alongside their therapist, and found that the application helped fixate the child's attention on the activity. In another IDC study, Ferri et al. [8] conducted a research-through-design study of games for cognitive behavioral therapy. They prototyped three games to help children improve self-reflection and emotional analysis skills. These games were either non-competitive or gently competitive, without real loss scenarios. After surveying 18 physical therapists, Annema et al. [3] provided three implications for therapy game design: (i) configuration and setup should be simple and quick for the therapist; (ii) games should support the child and therapist by supporting on-the-fly changes and easy pausing or level ending; and (iii) games should report and log child performance to give an overview or report across multiple therapy sessions. While simple games work well for infrequent events, such as a single clinical evaluation [30], arcade-style games may not be the most appropriate for long-term therapy, as gameplay can quickly grow stale [1, 34].

Previous applications for mobile speech therapy, such as Tabby Talks [32, 38], were developed as a proof-of-concept for remote speech therapy with a simple prompt interface. Similarly, Vocaliza [48] is a speech recognition system to help children with phonological, semantic, and syntactic therapy that shows progress over time. Research suggests that children engage better in and make fewer errors with these types of electronic interventions than with traditional therapy [17].

Speech interventions have also been incorporated into casual games. Ganzeboom et al. [11] developed a multiplayer speech therapy game based on feedback from individuals with dysarthria. Players give each other verbal instructions through the interface – the game extracts loudness and pitch from the speech to provide feedback to help the player stay within a certain range. Umanski et al. [47] developed a game that helps children practice syllabic production rhythms. The game is a downhill slalom competition where the player makes their skier turn by producing the syllable at the correct time, with more accurate timing resulting in a tighter turn. Flappy Voice [19] is a modified clone of the popular game Flappy Bird where vocal loudness and pitch are mapped to the bird's position along the vertical axis. Players can use any verbal or non-verbal utterance to guide the bird through openings in the pipes, so long as pitch and loudness patterns can be extracted from the utterances. Lopes et al. [24] developed a game to practice sustained vowel sounds. A bird flies from one branch to another if a vowel is produced with consistent intensity for a set duration, otherwise, it falls and the game resets. A more novel approach is demonstrated by Shtern et al. [39], where the speech articulators (i.e., tongue) are examined rather than the produced speech. In their game, the player uses tongue movements to control a flying bee.

GAME DESIGN

Game development

We developed Apraxia World atop a full-featured, multi-world game project available for the Unity Game Engine. The game (Ekume Engine 2D) is a colorful adventure game where the player controls a monkey character. It comes

(a) (b) (c)

Figure 2 (a) The clothing store offers different pieces to fully dress up the character (b) The weapons store offers four types of weapons with increasing power (c) The power-up store offers uses of power-ups and increases to power duration

with 48 levels divided into 8 worlds, multiple characters, and an in-game store for clothing and power-ups. All of the characters are shown in the start screen; see Figure 1a. Gameplay is linear – players must work their way towards the goal line at the right side of each level by navigating platforms, caverns, and other obstacles while trying to collect assets and avoid or eliminate enemies. Players control their character with a directional pad and two buttons, all overlaid on the tablet screen. Level and character information is shown in a heads-up display; see Figure 1b. The game offers two types of assets to collect: coins and stars. Coins are plentifully dispersed throughout the levels and are used to purchase items in the store. Stars originally served as a secondary challenge where a player could try to collect all stars within a level before finishing; this is similar to other games where players try to find all items of an object class. Each level contains a checkpoint (represented by a large anchor icon) around halfway through – if the player dies before reaching this point, they lose the assets (coins and stars) collected so far in that level. However, if they die after reaching the checkpoint, they keep the assets and restart at the checkpoint.

The in-game store sells clothing/costumes, weapons, and power-ups. The store uses in-game currency, either collected in the levels or awarded for doing exercises. The prices for store items range from 50 to 6,000 coins. The clothing store is shown in Figure 2a, where the player can see how the different items look on their character. The weapons (Figure 2b) vary in power and striking distance (e.g. slingshots can shoot far but swords and hammers are close-proximity weapons)[2]. Power-ups (Figure 2c) include coin value duplication, flight, invincibility, and coin magnets, all of which last for a short duration that can be lengthened by upgrading the power-up in the store.

We left the core gameplay unchanged, and instead modified the role of the stars. In our modified game, a player must collect a predetermined number of stars to complete a level, each star in turn requiring the player to complete a number of speech exercises. The game delivers these speech exercises either *during* or *after* gameplay; the delivery method is explained in the next section. We associated speech production with the stars so that players would be able to anticipate and control when speech exercises would appear. Additionally, we needed a "safe" time to display the exercise that would not detract from the gameplay or interrupt the player while executing complicated moves.

As well as adding the speech exercise, we also edited the levels to make them age-appropriate and increased the number of stars to 7-10 per level. In addition, we set stars to regenerate in the same place 10 seconds after being collected. We wanted a surplus of stars in different locations throughout the level to encourage players to gather extra and complete additional speech exercises if they so desired.

Speech exercises

The SLP can set how many exercises must be completed for each level, as well as provide a customized list of words per level, according to each child's therapy needs. In what follows, let E denote the number of speech exercises (i.e., word prompts) that must be completed per star, S denote the number of stars per level, and C denote the value of each star (in coins), all as defined by the SLP. Prompts are randomly selected from the word list such that they do not repeat until all words have been prompted.

The game delivers exercises in two ways: *during-game* or *after-game*. In the *<u>during-game</u>* mode, an exercise popup (see Figure 3a) appears when the player attempts to collect a star, at which point the player must complete E prompts. Correctly producing the target word triggers the game to either load the next prompt, or dismiss the popup if enough prompts have been completed. Incorrectly producing the target word causes a "Try again!" message to display briefly before the word prompt is displayed again. When the child has completed E prompts, the popup window disappears, a star is awarded, and C coins are also awarded. Players can collect as many stars as they like, each star yielding C coins. If the player attempts to complete the level before collecting S stars, a text banner prompts them to turn around and collect additional stars – see Figure 3b.

[2]Although the game contains weapons and some combat, it is very mild in terms of violence. There are neither blood nor death animations – characters and enemies simply fall over and then disappear.

(a) (b) (c) (d)

Figure 3 (a) Speech exercise popup in the *during-game* condition contains both a pictorial and text cue (b) The game displays a warning message when a player tries to finish the level before collecting enough stars (c) Speech exercise popup in the *after-game* condition. An awarded star count has been added to help children know how far along they are in the exercises (d) Speech exercise popup in the *after-game* condition once the minimum numbers of exercises have been completed. The message tells the player that they can either complete more exercises for a bonus or press the button to continue to the next level

Once the child collects at least S stars and crosses the goal line, the level ends.

In contrast, the *after-game* condition allows children to play the game as normal until they attempt to cross the goal line, at which point they must complete $S \times E$ exercises – the same number as the *during-game* condition. Before attempting to cross the goal line, the player is allowed to collect as many stars dispersed through the game as they want, but these stars do not award any bonus coins nor do they trigger speech exercises. If players so choose, they can collect no stars and go straight for the goal line. Once the player reaches the goal, the exercise popup appears; this popup (Figure 3c) is identical to the one in the *during-game* condition, except that it has a Star Counter so that the player knows their exercise progress. After each correct utterance, the game loads the next prompt. The same brief "Try again!" message as in the *during-game* condition appears if the child incorrectly produced the target word. Every E prompts, the game awards C coins and one star; this reward is reflected in the Star Counter. Once $S \times E$ exercises have been completed, two text banners and a continue button appear (Figure 3d); the banners inform the child that they can continue producing speech to gain additional coins or they can press the continue button to end the level. Once the child presses the continue button, the popup disappears and the level ends.

The speech exercises (i.e., word prompts) are based on the Nuffield Dyspraxia Programme (NDP3), an intervention program for young children with severe SSDs, including CAS [49]. NDP3 is designed to address specific effects of CAS, such as single consonant and vowel articulation, sequencing sounds together, and maintaining accurate prosody. We selected NDP3 because it comes with a 750 image set representing CV, CVC, CVCV, and multisyllabic words, which can easily be displayed in the exercise popup. Furthermore, NDP3 shows good treatment and generalization gains when delivered intensively [28]. Nonetheless, Apraxia World can be extended to other practice materials beyond (or instead of) the NDP3 set.

Speech assessment

Previous mobile speech therapy applications have used some form of automatic speech recognition (ASR), such as Pocketsphinx [19, 31] or custom approaches [37]. However, ASR on mobile devices either produces poor recognition rates with disordered speech or requires an internet connection such that a server can process the audio (e.g., Google Speech, Apple's Siri). Additionally, ASR performs especially poor on speech from children [18]. Therefore, for the present study, we decided to isolate the game aspects of Apraxia World from issues pertaining to mobile ASR performance. Accordingly, we used a Wizard of Oz design where speech was evaluated manually by an SLP via a Bluetooth keyboard that allowed them to indicate (while the child played the game) whether or not each word had been produced correctly. While ASR will be used in future iterations of the game, using the human evaluator gave us the children's best-case impression of the game and speech exercise integration, without any frustration from ASR errors.

We designed the keyboard input to mimic a binary decision: the SLP marks a speech production either as correct or as incorrect. We implemented rules to reduce the number of incorrect attempts on a single word and minimize reinforcing the wrong pronunciation; 4 consecutive incorrect pronunciations will cause a new prompt to come up (i.e., skip the problematic prompt) and 3 skipped prompts during an attempt at collecting a star (i.e., 3 prompts were skipped before 2 prompts were said correctly and a star was awarded) causes the exercise popup to disappear without awarding a star. These rules were put in place now, so that the exercise logic will be the same between the current and future versions when ASR is enabled.

METHODS

We evaluated Apraxia World in a within-subject study where children played two versions of the game, where speech exercises were delivered either *during* or *after* gameplay. In the process, we surveyed the children's impressions of this style of game in terms of enjoyability, ease of play, likes, dislikes, suggestions for improvement; we queried preference for game version; and we analyzed meta-data to identify differences across versions in amount of speech practice completed.

Participants

Twenty-one English speakers took part in the study. Participants included 14 children with diagnosed SSDs ranging from mild to severe (7 motor-speech and 7 phonological impairments; 13 male and 1 female; mean age: 7.4 years, range: 4-12 years old), and 7 children reported by parents to be TD (4 male and 3 female; mean age: 8.7 years; range: 5-12 years old). The children with SSDs had all been formally assessed and diagnosed as having a speech sound disorder by a qualified SLP and at the time of participation had no other developmental diagnosis (e.g., autism spectrum disorder or cognitive impairment). All procedures were approved by the University's Human Research Ethics Committee and all children and guardians provided written informed assent/consent, respectively, before participating in the study.

Selection and Participation of Children

Participating families self-referred in response to flyers and advertisements placed within the University's Speech Clinic, sent out by email, posted on social media, and posted in a local magazine. They were then selected for participation on the basis of SSD diagnosis occurring without other developmental diagnosis or no speech or developmental diagnosis (i.e., TD). Children and parents were asked if they would like to participate in a study looking into the development of tablet-based games to help children with their speech therapy exercises. Children were told that they would be shown two versions of the same game and asked some questions to help the research team continue to develop the game. They were told they could stop playing/discontinue participation at any time.

Procedure

All children were asked to test both versions of the game (*during-game* and *after-game* conditions). The order of presentation of the two game versions was randomized. Audio was recorded during the exercises for later analysis and debugging. The SLP sat beside the child and evaluated speech in real time. Exercise parameters were fixed for all children ($E = 2$, $S = 10$, $C = 25$), such that each child had to correctly produce at least 20 words before completing a level.

Two individualized word lists of approximately equal complexity were created for each child, one for each version of the game. The words were chosen by the accompanying parent and both lists contained (i) five words the parent believed the child should have no difficulty producing and (ii) five that they believed the child would have some difficulty producing[3]. This was done in order to mimic a home-practice setting where some "easy" words are included to ensure some success. Each child's ability to say the words chosen for them was checked before they began playing the game.

The children were first provided a description of the game, its aim (to collect coins and stars to buy things for the character as progression is made through the levels), and instructions on how to play. A brief demonstration of how to use the controls was also provided. The children were not explicitly told that their word productions would be judged as correct/incorrect by the SLP conducting the study. They were asked to play each version for as long as they wanted, up to a maximum of 15 minutes per version. The children were then given the game to play on a Samsung Tab A 10.1-inch tablet (Android 6.0 Marshmallow). All children started with a training level that had no exercises, no enemies, and no chance of falling off the platform. The purpose of this training level was for children to learn the game mechanics. Each child progressed from the training level into Level 1 of the full game in the same way as they transitioned between other levels of the game. Once a child had played the first version of the game (for as long as they wanted to, up to 15 minutes), they were asked to complete a questionnaire about the game before being presented with the second version. On average, the questionnaire took 5 minutes to complete. The child was then again given the game for as long as they wanted to play (up to 15 minutes). After playing the second version, they were asked a series of follow-up questions before being asked (i) which version they preferred and which version they would now like to play again, and (ii) if they would like to play again. The

[3]e.g., for one child, "watch" and "witch" were hard words, while "rabbit" and "peach" were easy words. The same words may not work for different children.

questionnaire focused on game enjoyability, ease of play, likes, dislikes, and suggestions for improvement. It contained a combination of 5-point Likert-scales and open-ended questions; see Figure 4. The questions were read to all children and all responses were written down by the SLP. After answering the questionnaire, the child was allowed to play their version of choice again, if desired.

During gameplay, each child's behavior was also observed to monitor for signs of reduced concentration or signs of frustration, such as fidgeting. Were such signs observed, the child was reminded that they could cease gameplay at any time. Observations on each child's approach to gameplay were also collected, including a willingness to collect additional stars in either condition; focus on collecting all the available coins; a desire to explore the levels or to try to progress through the levels as fast as possible; and use of coins collected to purchase items from the store.

We logged the number of levels completed, strategy of gameplay (e.g., focus on completing the level vs acquiring assets), and number of exercises completed (i.e. words produced) for each child per level. This allowed us to explore whether the two game versions facilitated different amounts of practice.

RESULTS

Feedback from children

Figure 4 summarizes responses to questions that used numeric ratings via boxplots. Four of the children did not answer all questions on the questionnaire, but their available responses are included in the analysis.

Nineteen of the 21 children found the game enjoyable and said they would like to play it again. All 19 would have continued playing beyond the 15-minute time cap had they not been stopped by the researcher. The other two children (one SSD and one TD) requested to discontinue during the allotted testing time because they were not engaged with the game and said that they would probably not play it again. However, they did play both versions and their data are included in all analyses.

The younger children (4-5 years) conflated the question "How difficult was the game?" with ease of control manipulation; for example, some children who struggled to complete a level still rated the game as easy to play. The older children were better able to dissociate ease of game control and gameplay, and their answers as to how easy they found the game more closely reflected their game progression.

Responses to whether the *during-game* condition made the game harder were varied and depended, in part, on whether the child liked having the speech exercises during or after gameplay. Responses included: "[…] *because I liked the game and wanted to concentrate on it*" and *"[the exercises] keep on popping and almost killing you."* Most children agreed that the *after-game* condition did not make the game harder.

When asked which version of the game they preferred playing, 13/21 of the children selected the *after-game* condition (eight preferred *during-game*). The reasons for this preference included: "*the words at the end of the game didn't interrupt your game,*" "*instead of collecting stars you can just say them,*" and "*playing [the* during-game *condition] made the words harder.*" One of the children who liked the *during-game* condition said that they "*liked the exercises popping up*." A child said that although they liked the *during-game* condition, they "*would play [the* after-game *condition] again because of the risk of dying while doing exercises in [the* during-game *condition]*" (some children struggled to navigate immediately after the game un-paused following the exercises). Some children offered alternatives to the two conditions we included: one said they "*would like exercises before the level*" and another said they "*would choose neither – would like the words during the game and then again at the end of the level so that you can practice them and get extra points*."

Other verbal responses surveyed the child's likes, dislikes, and suggestions for improvement. When asked what they liked, children mentioned the monkey characters and fighting the enemy characters (e.g., "*bashing monkeys*," "*the monkey and hitting the monsters*," "*fighting the monkeys*," the "*bashing hammer*," and "*hitting enemies*"). One child said the game structure "*reminds me of Donkey Kong [and I] like that it was hard.*" Two other children also commented that they liked that the game increased in difficulty, saying: "*it gets harder*" and "*it takes work/skill to play*." Other likes included: "*[there are] not a million things to remember*" and "*all the super powers.*"

The children were also asked what aspects of the game they liked the least. The most common comments were about dying, restarting, and losing stars collected (if they died before the checkpoint) (e.g., "*keeping dying*," "*restarting when you die*," and "*losing stars when I die before the checkpoint*"). Although some children enjoyed that difficulty level increased quickly, others cited it as an issue (e.g., "*it got hard pretty quick*").

Suggestions for changes were varied and reflected that the children had engaged well enough with the game to imagine modifications for both individualization and development. Some suggested ensuring that the items for purchase were more varied and matched the characters, or combined with the superpowers (e.g., boots that allow you to fly). One child said they "*would rather princesses and unicorns*" than monkeys. Three children commented that they would like the game more if it had a storyline (i.e., a reason for their character's progression through the islands). For example, one said that they would like the island to have villages so that they could then be the hero who has to save their village. Other comments reflected the same idea of fleshing out the virtual world: *"collect[ing] an army to* *kill the bad guys,"* having *"different types of bad guys,"* and *"buy[ing] pets to help you survive."*

Observations on strategy, gameplay, and engagement

Gameplay data were available for all 14 SSD children. Data for two of the seven TD children were lost due to software malfunction.

All of the children, except the two who asked to discontinue play, were observed to concentrate well during both gameplay and exercise completion. Minor frustration was observed solely in relation to the child's character dying and/or loss of stars collected. This was, however, accepted by all children as a negative, but unavoidable, part of the game. The smaller children were observed to have difficulty holding the tablet and those less familiar with tablet-based games appeared to have difficulty managing the two-handed controls. One child's suggestion for easing these difficulties was to include an option for an external joystick. The double jump maneuver proved difficult for some children, who struggled with the button timing.

Approach to gameplay appeared to be linked to interest in asset collection. Sixteen children rated buying items for their characters highly (*"it made it like a quest to earn cash and buy your accessories"*). They were observed to spend more time collecting coins than the remaining five children, who said that buying items for their character did not interest them. The older children demonstrated a clear understanding of the relationship between completing exercises and asset collection, whereas the younger children did not. For example, three older children (10–12 years) purposefully undertook more than the minimum required exercises per level, with the express intent of purchasing items from the store.

Figure 5 shows the total number of speech exercises completed by the children per finished level. Speech exercises completed in unfinished or restarted levels are not included. Regardless of order of delivery, 14/19 of the children for which we have gameplay data finished more levels in the *after-game* condition; three children finished the same number of levels in both conditions and one child finished more levels in the *during-game* condition. This imbalance is due to two primary causes, (i) levels take longer to finish in the *during-game* condition because the player must spend time looking for stars or waiting for them to regenerate, and (ii) the gameplay data include levels completed in the brief free-play portion after the test. These data were left in because the free-play more closely approximates home-practice (less evaluative pressure on the child).

SSD children in total finished *l=31* levels in the *after-game* condition (median exercises per level: 20, range: 20-36) and *l=16* levels in the *during-game* condition (median exercises per level: 22, range: 20-26). TD children in total finished *l=11* levels in the *after-game* condition (median exercises per level: 22, range: 20-82) and *l=5* levels in the *during-*

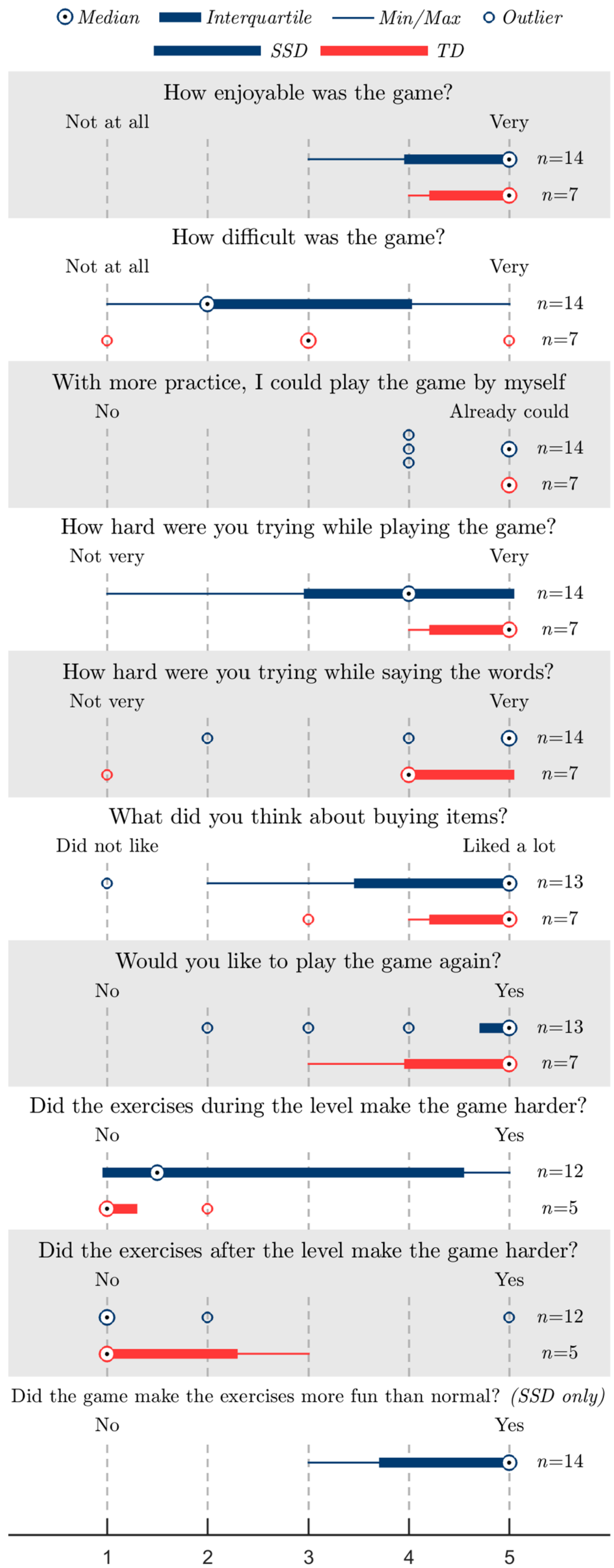

Figure 4 Boxplots for survey responses from all children (some children did not answer all questions)

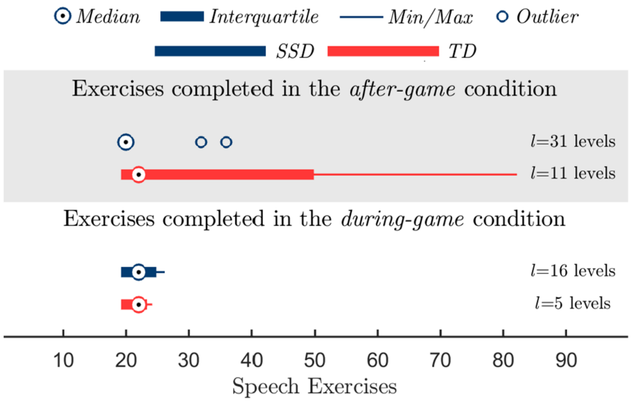

Figure 5 Boxplot of exercises completed per finished level (unfinished and restarted levels excluded)

game condition (median exercises per level: 22, range: 20-24). In general, SSD children completed the minimum number (20) of speech exercises per level in the *after-game* condition; TD children completed more exercises in the *after-game* condition due to a lack of perceived risk (any attempted exercise essentially guarantees a reward). All children completed close to the minimum number of exercises in the *during-game* condition. Across all participants, the median exercises completed in each condition (20 vs 22 – SSD; 22 vs 22 – TD) indicate that children are unlikely to complete large quantities of speech exercises beyond a specified minimum. As such, the choice of exercise delivery method may be more important as a per-player customizable element rather than a way to ensure maximal exercise completion; we further expand upon customization below.

DISCUSSION

This paper presents a novel approach for providing intensive and often tedious speech exercises to children with SSDs in a more engaging manner. We have developed two versions of a platformer game in which speech exercises are integrated and linked to asset collection, wherein the exercises can be presented either during or after gameplay. We surveyed children's impressions of the overall approach and version preference, and also examined meta-data for potential influence of version on the amount of speech practice undertaken.

Overall, the children (13/21) preferred the *after-game* condition for two main reasons: (i) they did not like having their gameplay disrupted, preferring to do the exercises separately and (ii) they did not like losing collected stars in the *during-game* condition when they died before reaching the checkpoint. Although the stars were placed in locations that should have been minimally disruptive to gameplay, the children still reported worrying about controlling their character immediately after the game un-paused when in a potentially difficult position (e.g., if an enemy is close by, if they are jumping over a platform gap). Losing stars upon dying was more discouraging to players than had been anticipated. One child compensated by strategizing: they prioritized reaching the checkpoint before collecting any stars. Stars collected before the player reached a checkpoint were intentionally not saved to encourage additional speech production. However, even though this led to all children completing many more exercises in the *during-game condition* than they did in the *after-game condition*, it also proved to reduce their motivation. Losing stars was judged as being more frustrating than repeatedly dying – the children completed an average of 25 exercises while playing in the *during-game* condition that were not saved due to restarting a level. In future versions of Apraxia World, this could be remedied by, for example, allowing players to keep all coins collected from the exercises (but not stars, still ensuring extra speech production) if their character dies before the checkpoint, or by allowing them to keep all stars and coins collected.

The eight children who preferred the *during-game condition* demonstrate that the preference for one version over the other was not unanimous. These children enjoyed having their speech exercises distributed during gameplay, with one stating that *"it seemed like I had to earn less stars [in the* during-game *condition]."* It could therefore be argued that providing future players access to both game versions would ensure that individual preferences will be met.

One important consideration, regardless of version, is the ratio of speech exercises to gameplay. Although the *after-game* condition was preferred by the majority, if a player struggles to make progress in the game, it becomes non-optimal in terms of number of exercises completed during gameplay, which undermines the major goal of the intervention tool. Most children reached a point, for the younger players (4-5 yrs.) in the first level, where they had to make multiple attempts to reach the end of the level. In the *after-game* condition, this resulted in a lot of gameplay without speech exercises. Similarly, some children seemed to like exploring the level and were in no hurry to move onto the next one, which again increased playtime without speech exercises. This could be remedied with a *before-game* condition, in which players would have to complete exercises whenever starting or restarting a level (from the beginning or checkpoint). Exercises could alternatively be presented at certain time intervals throughout the level. This would ensure that children could still experience uninterrupted gameplay time, while also ensuring that the necessary ratio of gameplay to speech exercises to maintain therapeutic utility would be upheld. An alternative solution may be to add an "energy" level that decays over time and must be replenished by completing exercises; in this fashion, players would be required to complete exercises regularly, but at a time of their choosing.

Providing tiers of game difficulty to cross a broader range of age, physical ability, or SSD severity may be beneficial in future versions. Child age and prior gaming experience were observed to affect player success. Similarly, the children had varying success with the game controls. Even though the controls used are standard for tablet games, some children had trouble with button combinations that required more careful timing. Again, the children who had limited prior experience with tablet-based games were observed to find the dual-handed controls difficult. A subset of children with movement-based speech disorders, such as CAS, have limb coordination difficulties; some children during the study were observed to have difficulty with game controls, extraneous limb movements, and rapidly timed double clicks. Compensatory strategies for these factors, such as an external joystick, need to be addressed in subsequent versions of the game.

The current study highlighted that built-in flexibility in a speech therapy tool is necessary. The subtle complexities in creating and presenting such a tool lie in matching both child and SLP expectations by balancing gameplay and child engagement against the provision of therapeutic levels of speech practice. Providing the user (SLP/child) with the ability to modify parameters such as exercises *before*, *during*, or *after* gameplay will help ensure the functionality and utility of the game as a therapeutic tool; this aligns with the implications for design put forth by Annema et al. [3] for therapy games. One of the aims of Apraxia World is to provide the child with a sense of autonomy during speech practice. Negotiation with their parent or SLP as to when they do exercises during gameplay would provide the child with a sense of control over their speech practice. However, to ensure that this negotiation does not lead to exercise avoidance, all game conditions need adjusting to ensure the ratio of gameplay to speech exercises is carefully balanced.

Similar to traditional gameplay, children undertaking gamified speech therapy want customizability in their game experience. The children generally liked the concept of buying items for their character. They purchased costumes, new weapons, and extra character power-ups. Children were motivated by a desire to customize their game character, and having items to purchase inspired them to collect coins and stars. Character customization is another method to help the child create an individualized gameplay experience, potentially helping them further engage with Apraxia World as a therapeutic tool. Maintaining a child's motivation to use the game and engagement in speech practice over the long-term is vital for the success of Apraxia World. Both character customization and choice over when the speech exercises appear are flexible elements of Apraxia World aimed at supporting this. However, limitations in the inventory of items available were highlighted during the current study. One child commented that the costume items available did not match well and another highlighted that there were no girl clothes. The suggestion of being able to pay to change the name of their character was also made. Developing the range of items available for purchase in subsequent versions of Apraxia World would ensure a rich gameplay experience for the child, helping to maintain motivation and engagement.

This study was limited by the population demographics – only 4 of the 21 participants were female, and only one of them was in the SSD group. Although up to 2.85 times more males than females have a SSD [26], our sex ratio approaches neither that of SSD nor general populations. Seeking a better demographic balance in future studies will help to make sure Apraxia World appeals to a wide audience.

For this study, we focused on the engagement and usability aspects of Apraxia World, which serve as the foundation for ambulatory studies we plan to conduct later in 2018. Direct SLP input will not be available during gameplay; as such, we will automate the speech evaluation through ASR. While mobile ASR engines (e.g., PocketSphinx) lack the capabilities of server-based solutions [38], recent findings [12] suggest that running the ASR engine in "forced-alignment" mode can be used to assess pronunciation. While this is generally not an option for general applications of speech recognition, in the context of speech therapy, the (target) spoken word is known in advance. Alongside the ASR, we plan to develop a therapist portal for managing the remote therapy application. Given that some children found Apraxia World too difficult, future versions will include more graduated level difficulty and adaptive difficulty, such that the game stays at an engaging level of difficulty as players' skills improve. Additionally, we will evaluate a *before-game* condition in the next version of Apraxia World, as the *during-game* and *after-game* conditions both had their own drawbacks.

CONCLUSION

In this paper, we presented Apraxia World, a mobile speech therapy tool built atop a full-fledged, multi-world platformer game. Apraxia World decouples speech production and game control to avoid limiting the type and variety of speech input; players complete speech exercises to make progress, but speech does not control character movement, which requires fine motor control. We conducted a user study to validate game functionality and evaluate how enjoyable children found gameplay alongside speech exercises. Overall, the children showed enthusiasm and engagement with Apraxia World and the novel mode of speech exercise delivery. Most of the children preferred to do exercises in the *after-game* condition, however, this was not unanimous; this indicates that future versions of the game should continue to offer flexibility in how players can do their speech exercises. The results of the study support the feasibility of Apraxia World as an augmentation to traditional clinic-based speech therapy.

ACKNOWLEDGEMENT

We want to thank the children and families who gave us their time and valuable feedback. This work was made possible by NPRP Grant # [8-293-2-124] from the Qatar National Research Fund (a member of Qatar Foundation). The statements made herein are solely the responsibility of the authors.

REFERENCES

1. Beena Ahmed, Penelope Monroe, Adam Hair, Chek Tien Tan, Ricardo Gutierrez-Osuna and Kirrie J. Ballard. Speech-driven mobile games for speech therapy. *International Journal of Speech-Language Pathology*, Under review.
2. Andrea Alessandrini, Victor Loux, Gabriel Ferreira Serra and Cormac Murray. 2016. Designing ReduCat: Audio-Augmented Paper Drawings Tangible Interface in Educational Intervention for High-Functioning Autistic Children. in *Proceedings of the The 15th International Conference on Interaction Design and Children*, Manchester, United Kingdom, ACM, 463-472. 10.1145/2930674.2930675
3. Jan–Henk Annema, Mathijs Verstraete, Vero Vanden Abeele, Stef Desmet and David Geerts. 2012. Video games in therapy: a therapist's perspective. *International Journal of Arts and Technology*, 6 (1). 106-122.
4. ASHA Adhoc Commitee on CAS. 2007. Childhood Apraxia of Speech, American Speech-Language-Hearing Association.
5. Carrie J Cai, Robert C Miller and Stephanie Seneff. 2013. Enhancing speech recognition in fast-paced educational games using contextual cues. in *SLaTE*, 54-59.
6. Gabriella A Constantinescu, Deborah G Theodoros, Trevor G Russell, Elizabeth C Ward, Stephen J Wilson and Richard Wootton. 2010. Home-based speech treatment for Parkinson's disease delivered remotely: a case report. *Journal of Telemedicine and Telecare*, 16 (2). 100-104. 10.1258/jtt.2009.090306
7. Expressive Solutions. 2018. ArtikPix. Retrieved from http://expressive-solutions.com/artikpix/
8. Gabriele Ferri, Wouter Sluis-Thiescheffer, Dries Booten and Ben Schouten. 2016. Playful Cognitive Behavioral Therapy Apps: Design Concepts and Tactics for Engaging Young Patients *Proceedings of the The 15th International Conference on Interaction Design and Children*, ACM, Manchester, United Kingdom, 486-498.
9. W Tecumseh Fitch and Jay Giedd. 1999. Morphology and development of the human vocal tract: A study using magnetic resonance imaging. *The Journal of the Acoustical Society of America*, 106 (3). 1511-1522.
10. Karen Forrest. 2003. Diagnostic criteria of developmental apraxia of speech used by clinical speech-language pathologists. *American Journal of Speech-Language Pathology*, 12 (3). 376-380.
11. Mario Ganzeboom, Emre Yılmaz, Catia Cucchiarini and Helmer Strik. 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*, ACM, 3-8.
12. Yuan Gao, Brij Mohan Lal Srivastava and James Salsman. 2017. Spoken English Intelligibility Remediation with PocketSphinx Alignment and Feature Extraction Improves Substantially over the State of the Art. *arXiv preprint arXiv:1709.01713*.
13. Hayrettin Gürkök, Gido Hakvoort, Mannes Poel and Anton Nijholt. 2011. User expectations and experiences of a speech and thought controlled computer game *Proceedings of the 8th International Conference on Advances in Computer Entertainment Technology*, ACM, Lisbon, Portugal, 1-6.
14. Susumu Harada, James A Landay, Jonathan Malkin, Xiao Li and Jeff A Bilmes. 2006. The vocal joystick: evaluation of voice-based cursor control techniques. in *Proceedings of the 8th international ACM SIGACCESS conference on Computers and accessibility*, ACM, 197-204.
15. Susumu Harada, Jacob O Wobbrock and James A Landay. 2011. Voice games: investigation into the use of non-speech voice input for making computer games more accessible. in *IFIP Conference on Human-Computer Interaction*, Springer, 11-29.
16. Brandi House, Jonathan Malkin and Jeff Bilmes. 2009. The VoiceBot: a Voice Controlled Robot Arm *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ACM, Boston, MA, USA, 183-192.
17. Donald G Jamieson, Garry Kranjc, Karen Yu and William E Hodgetts. 2004. Speech intelligibility of young school-aged children in the presence of real-life classroom noise. *Journal of the American Academy of Audiology*, 15 (7). 508-517.
18. James Kennedy, Séverin Lemaignan, Caroline Montassier, Pauline Lavalade, Bahar Irfan, Fotios Papadopoulos, Emmanuel Senft and Tony Belpaeme. 2017. Child speech recognition in human-robot interaction: evaluations and recommendations. in *Proceedings of the 2017 ACM/IEEE International Conference on Human-Robot Interaction*, ACM, 82-90.
19. Tian Lan, Sandesh Aryal, Beena Ahmed, Kirrie Ballard and Ricardo Gutierrez-Osuna. 2014. Flappy voice: an interactive game for childhood apraxia of speech therapy. in *Proceedings of the first ACM SIGCHI annual symposium on Computer-human interaction in play*, ACM, 429-430.
20. B. Lange, S. Flynn and A. Rizzo. 2009. Initial usability assessment of off-the-shelf video game consoles for clinical game-based motor rehabilitation. *Physical*

Therapy Reviews, 14 (5). 355-363. 10.1179/108331909X12488667117258

21. James Law, James Boyle, Frances Harris, Avril Harkness and Chad Nye. 2000. Prevalence and natural history of primary speech and language delay: findings from a systematic review of the literature. *International Journal of Language and Communication Disorders*, 35. 165-188.
22. Sungbok Lee, Alexandros Potamianos and Shrikanth Narayanan. 1999. Acoustics of children's speech: Developmental changes of temporal and spectral parameters. *The Journal of the Acoustical Society of America*, 105 (3). 1455-1468.
23. Little Bee Speech. 2018. Articulation Station. Retrieved from http://littlebeespeech.com/articulation_station.php
24. Marta Lopes, João Magalhães and Sofia Cavaco. 2016. A voice-controlled serious game for the sustained vowel exercise. in *Proceedings of the 13th International Conference on Advances in Computer Entertainment Technology*, ACM, 32.
25. Edwin Maas, CE Gildersleeve-Neumann, Kathy J Jakielski and Ruth Stoeckel. 2014. Motor-based intervention protocols in treatment of childhood apraxia of speech (CAS). *Current developmental disorders reports*, 1 (3). 197-206.
26. David H McKinnon, Sharynne McLeod and Sheena Reilly. 2007. The prevalence of stuttering, voice, and speech-sound disorders in primary school students in Australia. *Language, Speech, and Hearing Services in Schools*, 38 (1). 5-15.
27. Joan Mora-Guiard, Ciera Crowell, Narcis Pares and Pamela Heaton. 2016. Lands of Fog: Helping Children with Autism in Social Interaction through a Full-Body Interactive Experience *Proceedings of the The 15th International Conference on Interaction Design and Children*, ACM, Manchester, United Kingdom, 262-274.
28. Elizabeth Murray, Patricia McCabe and Kirrie J Ballard. 2015. A Randomized Controlled Trial for Children With Childhood Apraxia of Speech Comparing Rapid Syllable Transition Treatment and the Nuffield Dyspraxia Programme–Third Edition. *Journal of Speech, Language, and Hearing Research*, 58 (3). 669-686.
29. Moyen Mohammad Mustaquim. 2013. Automatic speech recognition- an approach for designing inclusive games. *Multimedia Tools and Applications*, 66 (1). 131-146. 10.1007/s11042-011-0918-7
30. LinhChi Nguyen, Weiquan Lu, Ellen Yi-Luen Do, Audrey Chia and Yuan Wang. 2014. Using digital game as clinical screening test to detect color deficiency in young children *Proceedings of the 2014 conference on Interaction design and children*, ACM, Aarhus, Denmark, 337-340.
31. Avinash Parnandi, Virendra Karappa, Tian Lan, Mostafa Shahin, Jacqueline McKechnie, Kirrie Ballard, Beena Ahmed and Ricardo Gutierrez-Osuna. 2015. Development of a Remote Therapy Tool for Childhood Apraxia of Speech. *ACM Transactions on Accessible Computing (TACCESS)*, 7 (3). 10.
32. Avinash Parnandi, Virendra Karappa, Youngpyo Son, Mostafa Shahin, Jacqueline McKechnie, Kirrie Ballard, Beena Ahmed and Ricardo Gutierrez-Osuna. 2013. Architecture of an automated therapy tool for childhood apraxia of speech. in *Proceedings of the 15th International ACM SIGACCESS Conference on Computers and Accessibility*, ACM, 5.
33. Pocket SLP. 2018. Pocket SLP. Retrieved from http://pocketslp.com
34. Zak Rubin, Sri Kurniawan, Taylor Gotfrid and Annie Pugliese. 2016. Motivating Individuals with Spastic Cerebral Palsy to Speak Using Mobile Speech Recognition. in *Proceedings of the 18th International ACM SIGACCESS Conference on Computers and Accessibility*, ACM, 325-326.
35. Leanne Ruggero, Patricia McCabe, Kirrie J. Ballard and Natalie Munro. 2012. Paediatric speech-language pathology service delivery: An exploratory survey of Australian parents. *International Journal of Speech-Language Pathology*, 14 (4). 338-350. 10.3109/17549507.2011.650213
36. Oscar Saz, Shou-Chun Yin, Eduardo Lleida, Richard Rose, Carlos Vaquero and William R Rodríguez. 2009. Tools and technologies for computer-aided speech and language therapy. *Speech Communication*, 51 (10). 948-967.
37. Mostafa Shahin, Beena Ahmed, Jacqueline McKechnie, Kirrie Ballard and Ricardo Gutierrez-Osuna. 2014. A Comparison of GMM-HMM and DNN-HMM Based Pronunciation Verification Techniques for Use in the Assessment of Childhood Apraxia of Speech. in *Fifteenth Annual Conference of the International Speech Communication Association*.
38. Mostafa Shahin, Beena Ahmed, Avinash Parnandi, Virendra Karappa, Jacqueline McKechnie, Kirrie J Ballard and Ricardo Gutierrez-Osuna. 2015. Tabby Talks: An automated tool for the assessment of childhood apraxia of speech. *Speech Communication*, 70. 49-64.
39. Mark Shtern, M. Brandon Haworth, Yana Yunusova, Melanie Baljko and Petros Faloutsos. 2012. A Game System for Speech Rehabilitation. in Kallmann, M. and Bekris, K. eds. *Motion in Games: 5th International*

Conference, MIG 2012, Rennes, France, November 15-17, 2012. Proceedings, Springer Berlin Heidelberg, Berlin, Heidelberg, 43-54.

40. Smarty Ears Apps. 2017. Apraxiaville. Retrieved from http://smartyearsapps.com/apraxia-ville/

41. Adam J Sporka, Sri H Kurniawan, Murni Mahmud and Pavel Slavík. 2006. Non-speech input and speech recognition for real-time control of computer games. in *Proceedings of the 8th international ACM SIGACCESS conference on Computers and accessibility*, ACM, 213-220.

42. Adam J Sporka and Pavel Slavík. 2008. Vocal Control of a Radio-Controlled Car. *ACM SIGACCESS Accessibility and Computing* (91). 3-8.

43. Chek Tien Tan, Andrew Johnston, Andrew Bluff, Samuel Ferguson and Kirrie J Ballard. 2014. Retrogaming as visual feedback for speech therapy. in *SIGGRAPH Asia 2014 Mobile Graphics and Interactive Applications*, ACM, 4.

44. Deborah G Theodoros. 2008. Telerehabilitation for service delivery in speech-language pathology. *Journal of Telemedicine and Telecare*, 14 (5). 221-224.

45. Deborah Theodoros, Trevor Russell and R Latifi. 2008. Telerehabilitation: current perspectives. *Studies in health technology and informatics*, 131. 191-210.

46. Tiga Talk. 2011. Tiga Talk Speech Therapy Games. Retrieved from http://tigatalk.com/app/

47. D Umanski, D Kogovšek, M Ozbič and NO Schiller. 2010. Development of a voice-based rhythm game for training speech motor skills of children with speech disorders. in *Proceedings 8th International Conference Disability, Virtual Reality and Associated Technologies*.

48. Carlos Vaquero, Oscar Saz, Eduardo Lleida, J Marcos, César Canalís and Colegio Público De Educación. 2006. VOCALIZA: An application for computer-aided speech therapy in Spanish language. *IV Jornadas en Tecnología del Habla*. 321-326.

49. Pamela Williams and Hilary Stephens. 2004. *Nuffield Centre Dyspraxia Programme 2004*. Miracle Factory, for the Speech & Language Therapy Department, Nuffield Hearing and Speech Centre, Royal National Throat, Nose and Ear Hospital.