Spokelt: Building a Mobile Speech Therapy Experience

Jared Duval, Zachary Rubin, Elena Márquez Segura, Natalie Friedman, Milla Zlatanov, Louise Yang, Sri Kurniawan

University of California Santa Cruz, Santa Cruz, US

{jduval@ucsc.edu, zarubin@ucsc.edu, elena.marquez@ucsc.edu, nvfriedm@ucsc.edu, mzlatano@ucsc.edu, lbyang@ucsc.edu, skurnia@ucsc.edu}

ABSTRACT

Spokelt is a mobile serious game for health designed to support speech articulation therapy. Here, we present Spokelt as well as 2 preceding speech therapy prototypes we built, all of which use a novel offline critical speech recognition system capable of providing feedback in real-time. We detail key design motivations behind each of them and report on their potential to help adults with speech impairment co-occurring with developmental disabilities. We conducted a qualitative within-subject comparative study on 5 adults within this target group, who played all 3 prototypes. This study yielded refined functional requirements based on user feedback, relevant reward systems to implement based on user interest, and insights on the preferred hybrid game structure, which can be useful to others designing mobile games for speech articulation therapy for a similar target group.

Author Keywords

Speech Therapy; User-centered design; Serious Game for Health; Spokelt; Developmental Disabilities

ACM Classification Keywords

H.5. Information interfaces and presentation (e.g., HCI); H.5.2. Voice I/O; I/O J.3. Life and Medical Sciences: Health

INTRODUCTION

Speech is a crucial skill for effective communication, expression, and sense of self-efficacy. Speech impairments often co-occur with developmental disabilities such as Autism Spectrum Disorder [44], Cerebral Palsy [9], and Down Syndrome [8]. The prevalence of speech impairments in individuals with developmental disabilities has been as high as 51% [38]. Each of these developmental disabilities exhibit symptoms of an articulation disorder. An articulation disorder is categorized as having difficulty producing speech sounds that constitute the fundamental components of a language [54]. Many individuals with speech impairments experience depression, social isolation, and a lower quality

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

MobileHCI '18, September 3–6, 2018, Barcelona, Spain

© 2018 Association for Computing Machinery.

ACM ISBN 978-1-4503-5898-9/18/09...\$15.00

<https://doi.org/10.1145/3229434.3229484>

of life [27]. Speech problems can negatively impact a person's employment status [30], and their ability to receive proper healthcare. This includes receiving wrong diagnosis, inappropriate medication, and access to service [33]. The rate of arrests and convictions was higher for boys with language impairments [5]. In 2012, approximately 10% of the U.S. adult population experienced a speech, language, or voice problem [30].

Speech is a skill that can often be improved with individualized therapy and practice [39,45]. Access to Speech Language Pathologists (SLPs) is crucial to improving speech, but up to 70% of SLPs have waiting lists indicating a shortage in the workforce and disrupted access to therapy [47]. As a result, many non-professional therapists are being trained by SLPs to deliver speech therapy outside of the office [24,51]. This is not an ideal situation because the SLP must take the time to train the non-professional speech therapy facilitator, the individual's therapy schedule then relies on the facilitator's schedule, and these facilitators may not be as effective at delivering a speech curriculum [24]. Even worse, many untrained facilitators attempt to deliver speech curriculums reporting a general low sense of competence in assisting people with disabilities in their assigned curriculums [33].

Mobile speech therapy games could help people practice articulation anywhere without the need to be facilitated, which may potentially expedite their speech therapy progress [40]. The pervasiveness of mobile hardware makes it an ideal platform for delivering speech therapy to those who may not have access to a speech therapist or a facilitator. Many SLPs design games and activities to engage their clients. Games and play have been widely recognized as a valid motivator for otherwise jaded individuals [4]. We expect there are many benefits to using a mobile speech therapy game, including the ability to practice anywhere, collect fine-grained speech data, track the frequency and time individuals spend practicing, track performance over time, and create dynamic, custom therapies to each individual. This has presumably motivated the appearance of many mobile speech therapy apps with different features and function. Yet, they tend to require a facilitator to evaluate speech. Speech recognition has been successfully used to facilitate speech therapy [2,3,6,10,31,32,46,49,52], but not in a mobile context focusing on articulation.

In this paper, we describe key implementation details of the underlying mobile offline real-time critical speech

recognition system. We present the designs of 2 articulation therapy game prototypes that culminated in the creation of SpokeIt. We share results from our comparative within-subject study, which included 5 adults with disabilities co-occurring with speech impairment who played each of the 3 designs. Finally, we discuss lessons learned about these designs, reward systems our participants were interested in, and promising future work.

RELATED WORK

We use this section to a) introduce the target group in our studies and discuss the type of articulation therapy that is generally appropriate, b) motivate our choice to build a serious game for health for speech therapy, and c) argue that a gap exists for mobile articulation therapy games utilizing critical speech recognition systems.

Articulation Disorders Co-occurring with Developmental Disabilities

Adults with developmental disabilities co-occurring with speech impairment would benefit from speech therapy [38]. There is a trend towards helping children with speech impairments [6,8,12,41,43,49,51], but adults with speech impairments need support as well [44,45]. This support can come in the form of a speech recognition system, discussed below.

Autism Spectrum Disorder (ASD)

ASD is one of the most common developmental disabilities, affecting approximately 400,000 individuals in the United States [55]. A follow-up study was conducted on children with ASD and communication problems when they reached early adulthood showing that the group continued to show significant stereotyped behavior patterns, problems in relationships, troubles with employment, and lack of independence [17]. A person with ASD may have monotonic (machine-like) intonation, deficits in pitch, vocal quality, volume, and articulation distortion errors [44].

Cerebral Palsy

A person with Cerebral Palsy and dysarthria (difficult or unclear articulation of speech that is otherwise linguistically normal) may include anterior lingual place (front of the tongue) inaccuracy, reduced precision of fricative (consonant requiring breath through a small opening) and affricative (Plosive followed by fricative sound like *j* as in *jam*) manners, slowness of speech, and indistinctness of speech [9,35].

Down Syndrome

Many people with Down Syndrome have muscle hypotonia [8]. Muscle hypotonia may cause abnormal movement of the cheek, lips, and tongue, resulting in articulation errors. Many people with Down Syndrome also speak at a fast and fluctuating rate [48] known as cluttering [16].

Speech Recognition to Support Speech Therapy

Each of the aforementioned developmental disabilities have one general symptom in common—an articulation disorder. For this reason, we focused on the development of a critical speech recognition system capable of distinguishing between

correct and incorrect pronunciations. We discuss these implementation details thoroughly in the *Creating a Novel Speech Mechanic* section. There are many more technologies that focus on improving speech skills such as fluency, pitch, rhythm, dialect, and aphasia [3,6,31,31,46,49,52], but we do not focus on these.

Serious Games for Health

Games can be used as effective educational interventions [1,28]. Games have the ability to teach while providing a motivating and interactive environment [50], and can be as effective as face-to-face instruction [37]. They can create opportunities to train new skills in a safe and engaging environment, improving perceived self-efficacy, a key aspect in health promotion interventions [22]

Multiple serious games for health are documented to be effective for diverse platforms, health outcomes, and target populations. Some examples that populate this space range from an exergame to help blind children with balance [29] to embodied persuasive games for adults in wheelchair [15] to mobile games for motivating tobacco-free life in early adolescence [34].

Games are such a powerful motivator that non-game activities are often designed to look like games (i.e. gamified systems). This attracts user attention and entices engagement [11], which is particularly useful for tedious and otherwise non-interesting activities [7]. By adding game design elements to a pre-existing activity, designers manage to engage people more with this activity [23]. However, traditional gamification approaches have been widely criticized [25]. Ethics apart, simply adding superficial game-looking elements to an otherwise tedious activity does not work in the long run. Both Nicholson and McGonigal [25,26] have pointed out that extrinsic rewards like those typically used in gamification can decrease intrinsic motivation and engagement after the initial novelty effect. McGonigal suggests a more fruitful approach by making the activity intrinsically motivating and the rewards meaningful to the player [25]. These relevant considerations prompted us to study the types of rewards that interest our players.

Speech Therapy Applications

We present a comparison of some existing speech therapy systems, shown in Figure 1. The survey of speech therapy applications we present was compiled from interviews with 3 speech language pathologists, a literature review, and current offerings on the App Store. The features we focus on are whether the solutions run on mobile hardware and whether or not they use speech recognition. We choose these features to illustrate the gap SpokeIt is intended to fill. Running on mobile hardware is important because of the pervasiveness and convenience of mobile hardware [36]. A critical speech recognition system is important for a speech therapy game because it removes the requirement to be facilitated by an often unreliable third party, opens up opportunities to record fine-grained speech data, track progress, and listen for both correct and incorrect pronunciations.

Of all the applications surveyed, we include Sayin' it Sam, the only other mobile game with an identical underlying speech recognition library, to illustrate that similar technology can serve very different purposes based on the target population. We found no mobile games that use a critical speech recognition system for articulation therapy. We include non-mobile applications with speech recognition systems because they offer novel speech therapy experiences and features we wish were available in a mobile context. We include a set of mobile speech therapy games without speech recognition systems because they have many noteworthy features and designs—which could be further improved if they implemented a critical speech recognition system.

Figure 1: Venn Diagram depicting dichotomy between therapy systems that use speech recognition and those that run on mobile hardware

Sayin' It Sam is the only other mobile application featuring a speech recognition system identical to the library SpokeIt's critical recognition is built on. Sayin' it Sam is primarily focused on motivating non-verbal children to speak and is therefore trained to be very forgiving. SpokeIt, however, is built to listen critically to speech and to be used as an articulation therapy tool.

Other researchers have integrated speech recognition into non-mobile interactive game environments to improve literacy. Project LISTEN is an automated reading tutor aimed at helping children learn pronunciation and proper speech when reading aloud. It does this by analyzing various aspects such as pitch, speed, and pauses. Researchers have tested the system in India for assisting children learning English as a second language, as well as in Canada for children looking to improve their speaking skills [31].

Project ALEX is a non-mobile application that has proposed a very robust application for language learners of any age. Project ALEX focuses on a large dictionary with text-to-speech functionality. Most importantly, Project ALEX included pronunciation practice and used speech recognition to check if the user says the word correctly using Microsoft SAPI [32]. Project ALEX is focused on studying the cultural differences in speech and dialect.

Articulate it! is a unique multi-player mobile application created by a SLP specifically to help children improve their

speech sound production. Articulate it! employs over 1000 images selected for working on English consonant sounds at the word and phrase level and has the ability to store data for multiple patients. Articulate It! has multiple game modes such as a phonemes mode where the facilitator can select target sounds and a mode where the facilitator can focus on words with a specified number of syllables. Once a mode is selected, the facilitator has the option to customize the dictionary of target words by removing unwanted ones. Modes can be switched without ending a session and speech can be recorded for later comparisons and for the player to listen back to their speech. This app requires a facilitator to score speech.

Articulation Station [43] is a novel mobile speech therapy app that allows SLPs to customize target sounds and sound placement for patients of all ages to practice. For example, SLPs can make the app focus on /k/ sounds that occur at the beginning, middle, or end of a word, such as cat, pickle, or tick. The app has three levels of difficulty where users must say target words, sentences, or full stories. Like Articulate it!, Articulation Station requires a SLP or facilitator. Here, they grade speech, which is recorded in the app and available for reporting. The app has pre-recorded samples of correct speech for all target words.

Articulation Games is a comprehensive, flexible, and fun speech-therapy app for iPads that was created by a certified speech and language pathologist for children to practice the pronunciation of over forty English phonemes, organized according to placement of articulation. It includes thousands of real-life flashcards, accompanied by professional audio recordings and ability to record audio. Players practice phonemes through activities like memory games and flashcards. Articulation Games requires a facilitator to grade speech. Auditory Workout, Articulation Vacation, and Real Vocabulary all offer similar features and experiences as Articulations Games, Articulation Station, and Articulate It!.

Many apps have more features than what is presented, such as data collection capabilities for progress reports, the ability to record player's voices so they can listen back, and challenges that progress from single sounds and words to sentences and finally free natural speech. In the full release of SpokeIt, we plan to include many of these features.

METHOD

We started our work by conducting 3 semi-structured interviews with medical speech experts, which lead to the creation of our functional requirements. We used these functional requirements to inform the development of our speech recognition system. Following an iterative user-centered design process, we designed and implemented 3 prototypes that used this speech recognition system. Each of the designs explored a unique game format and core speech mechanic, which we detail and motivate in the following. These prototypes were developed in sequence, so design knowledge carried over from one prototype to the next.

We conducted a within-subject comparative study on the 3 prototypes with 5 adults with developmental disabilities co-occurring with speech impairment. We were concerned with participants' interest in using a speech therapy game to improve their speech, their opinion on each of the 3 prototypes, their preferred game structure, and relevant reward systems they would be interested in.

We present here the main insights from this study, which directly influenced the further development of SpokeIt. We think they can be useful for others designing mobile games for speech articulation therapy for a similar target group.

DESIGN

Creating a Novel Speech Mechanic

Our work began with semi-structured interviews with medical speech experts. Researchers asked what technology was currently used for speech therapy, what benefits and drawbacks they see to using technology for speech therapy, and what functionality must exist. The only technology our experts used during speech therapy sessions where iPads displaying images of speech targets for diagnostic purposes. Experts suspected that their patients were practicing little to none outside of the office, even though they recommended 10 minutes of practice per day. They were hopeful that a speech therapy game would motivate their patients to practice outside of the office. They stipulated that the system must critically listen to pronunciation and were mildly worried that a speech therapy game could condition bad speech practices if the recognition was not accurate. The inclusion of examples of correct speech mitigated these concerns. Finally, they expressed concerns that many of their patients from lower socio-economic statuses may not have access to the internet, so our game must be functional offline. We budgeted for iPads in our grant so that these populations can keep the iPads with the software installed after the development and evaluation of SpokeIt has ended. We chose iPads because they are categorized as medical devices by the United States Government.

While there are many novel and interesting mobile applications available that can improve the speech therapy experience, we found none that provided a critical speech recognition system in a game for on-the-go or at-home use. Before developing any game prototypes, it was necessary to ensure that it was feasible to listen for both correct and incorrect speech. Many speech recognition systems exist today such as those used in personal assistants like Cortana, Siri, Google Assistant, and Amazon Alexa. We chose not to use these services for multiple reasons:

- We needed a solution that is offline because we are dealing with sensitive speech data. In addition, not every home has access to the internet. Also, online speech recognition systems often have lag and usage caps that would hinder real-time game play.
- Digital assistants like the ones listed above are designed to best guess speech, not listen to it critically. We needed to

be able to fine-tune the recognition to listen for incorrect speech as well as correct utterances.

- We did not want to discard the possibility of using and recognizing non-existent words. Having the freedom to play with silly nonsensical words is a tactic many SLPs use to target specific sound production.

With these requirements in mind, we began searching online for mobile speech recognition libraries that are highly customizable and do not require an internet connection. The library we chose is Pocketsphinx, an offline speech recognition system for handheld devices from Carnegie Mellon [19]. A speech therapy game must be able to listen to speech critically so that the intervention will promote correct speech. Pocketsphinx uses customizable dictionaries that allow developers to customize the targets that can be recognized [18]. The dictionaries that Pocketsphinx employs use ARPabet, a set of phonetic transcription codes, to map speech sounds to English words [14]. ARPabet can be used to construct any sequence of phonetic sounds to a word—even words that do not exist. Any set of sounds that an English speaker can produce can be mapped to an ARPabet representation. We can make new “words” that map to common miss-pronunciations of correct words. Providing both correct ARPabet codes and ARPabet codes that represent miss-pronunciations give us the power to distinguish between correct and incorrect speech. Table 1 shows ARPabet codes that represent both correct and incorrect ways to say the word balloon.

Common pronunciations of “Balloon”	ARPabet Code
Balloon	B AH L UW N
Walloon	W AH L UW N
Walloo	W AH L UW
Bawoon	B AH UW N
Baloo	B AH L UW
Bawoo	B AH UW
Alloon	AH L UW N
Loon	L UW N

Table 1: Pronunciations of the word “Balloon” including correct pronunciation (first) followed by common miss-pronunciations and their corresponding ARPabet Code

Pocketsphinx uses acoustic models to map sound data to targets in the dictionary. These acoustic models are hot-swappable and can be altered for better accuracy [18]. This feature creates the potential to alter acoustic models for specific populations, allowing a more accurate model that can listen to adults with developmental disabilities, or even one specifically for children with cleft speech. OpenEars is a free open-source framework that brings the power of Pocketsphinx to iOS devices in native objective-c language for speed and reliability. RapidEars is a paid plugin for OpenEars that gives Pocketsphinx the ability to listen to

speech in real-time, which is important for a responsive game. The ability to customize acoustic models, to customize dictionaries, to run offline, and to listen in real-time motivated our choice to use RapidEars for our speech therapy game prototypes.

Design and Development of prototypes

Our primary target population for the final release of SpokeIt was children, and hence the 3 prototypes were designed with children in mind. However, our primary IRB and physical location limited our access to children with speech impairments at the time. For accessibility reasons, we tested our prototypes on adults with developmental disabilities co-occurring with speech impairment—who can also benefit from the kinds of games we were creating. This could be seen as a limitation of our study, yet as we will further discuss later, the acceptance and enjoyment of our adult participants towards our designs indicated they had good potential to engage population older than our original target group.

To initiate the design process and after interviews with a SLP, we drew inspiration from media that successfully elicits speech from children, such as the popular children's program Dora the Explorer. The program inspired us because it integrates learning and vocal participation in a storybook style setting with intermittent learning challenges that children love.

Speech Adventure

The first prototype that was developed was Speech Adventure [42]. Speech Adventure, shown in Figure 2, is a storybook style game that employs an off-screen narrator to give directives on how to help Sam the Slug complete tasks. Visual cues in the form of glowing blue outlines inform players which parts of the scene can be interacted with and touched. Once touched, the corresponding target phrase is announced by the off-screen narrator.

To make progress in the game, the player must repeat the target phrase that was just announced. The target words are displayed at the bottom of the screen and are tightly tied to the speech recognition system. Words turn green as they are said correctly. To make progress, all words must be green. Words can be said out of order and words that were missed anywhere in the phrase can be repeated.

The green ear in the upper left corner of the screen indicates that the system is actively listening. When the off-screen narrator is speaking, the green ear turns white to signal the system is not listening. We found it important to suspend recognition during these moments so that the in-game audio didn't trigger game events.

The story of Speech Adventure starts with dressing Sam in boots and a hat. Once Sam is dressed, the player must say "Open the door" to journey outside. Once outside the player must pop 3 balloons that are blocking a bridge by saying "Pop a balloon" before continuing on the journey. These phrases were provided by SLPs.

Figure 2: Speech Adventure cabin scene where player must help Sam the Slug get dressed before going outside

A challenging design aspect of this game was that each target phrase had to be carefully crafted to fit into the narrative of the game. Development of new scenes that incorporated target words proved to be very time consuming, which could result in minimal content. We worried that Speech Adventure would lose its novelty after a first play-through. From conversations with the SLP, we realized re-playability was an important feature for a game that should be played for 10 minutes per day.

The majority of the allotted 10 minutes of daily practice should be spent producing target words and practicing speech. Yet, the narrative nature of our storybook-style design limited the number of utterances that could be produced in 10 minutes, which then caused concerns within the design team about the therapeutic value of our game.

The nature of hand-crafted narratives limited our ability to dynamically swap target words. According to a SLP we interviewed, a speech therapy game would benefit from the ability to customize targets dynamically based on the types of speech therapy each individual player needs.

Although preliminary play tests indicated that players love the storybook-style Speech Adventure game [40], the considerations above prompted us to rethink how a speech therapy game should be structured.

Speech with Sam

To maximize the number of target words produced in a 10-minute period, we hypothesized that single-word utterances that had an immediate effect on gameplay would yield more speaking time. Removing narrative between utterances reduces the amount of time that the recognition system is suspended. This resulted in the development of Speech with Sam, a series of speech controlled mini-games, shown in Figure 3. The mini-game structure allowed quicker development time, which yielded more content to play. Because targets are not directly linked to a narrative, implementing a diverse range of dynamic targets would be more feasible.

Figure 3: Speech with Sam rocket mini-game where rockets are set off by saying the appropriate target

Speech with Sam played a series of mini-games ordered randomly for a specified amount of time. In the rocket mini-game example above, the player must tap one of the three rockets. Touching a rocket reveals an in-game prompt that specifies a rocket's trigger word. Once the trigger word is said, the rocket blasts off the screen and is replaced by a new rocket with a new random trigger word.

For every rocket launched, the player score is increased by 1 and displayed in the green rounded rectangle in the upper-left corner of the display. This score is recorded at the end of every mini-game to keep a record of high scores and to track trends about player scores.

The number in the blue rounded rectangle, next to the score is a countdown timer until the next mini-game is played. After the timer reaches 0, the score is recorded, and the player is presented with a different randomly chosen mini-game. All mini-games run for a standard amount of time. It is possible for players to play the same mini-game in one session, but not twice in a row.

The green ear in the upper-right corner of the screen works identical to Speech Adventure in that when it is green, the speech system is listening, and when white, the speech system is suspended so that in-game audio does not trigger events. The written prompt was moved to the top to join the other heads up display elements. When players say the target word, the text turns green and the game immediately responds. In Speech Adventure, targets were often multi-word phrases, whereas in Speech with Sam, the targets are single words.

In a preliminary study [40], we found that Speech with Sam was successful in increasing the words per minute from participants, meaning Speech with Sam has the potential to be a more effective speech therapy solution. In that same study, we found that participants were unenthusiastic when presented with a mini-game they had played before. Our hypothesis about mini-game re-playability being high was not necessarily true.

SpokeIt

SpokeIt, shown in Figure 4, is storybook-style mini-game hybrid. By adding a story around the mini-games that fit into

an overarching plot, we have the potential to both produce high output of words per minute and keep users engaged.

Figure 4: SpokeIt card coloring mini-game

SpokeIt is the first speech therapy game to both demonstrate correct pronunciation with pre-recorded audio and lip animations. One of the medical experts we interviewed later suggested that showing correct speech is important as well as hearing correct speech. We did not want to break immersion by displaying a realistic mouth in an animated 2D environment. We found a solution that allowed us to sync the audio that exemplified correct speech with animated mouth transitions on our characters. The lip animation effects were achieved using Adobe Character Animator. Each phoneme mouth shape was crafted in Adobe Photoshop with three frames per transition, resulting in smooth transitions between mouth shapes. Adobe Character Animator's Lip Sync feature and our frame transitions automatically map our voice actor's speech performances to the appropriate mouth shapes. The motivation behind this work is to give players visual cues on how a word is said. Adobe Character Animator's abilities to synchronize lip animations and replicate actor's facial expressions is shown in Figure 5.

The SpokeIt prototype, shown in Figure 4, is meant to blend the positive aspects of the Speech Adventure and Speech with Sam experiences. Therefore, the beginning of the game starts with a narration from Sam the Slug and their desires to go and visit a close friend. To get to the friend, the player must partake on a long journey filled with new experiences and challenges. In the example above, Sam meets a friendly creature named Red who is struggling to learn colors that start with the letter "B"—or in the future, colors that start with any letter that player needs to practice. Sam knows the player is great with colors and asks them to teach Red colors that start with "B" by saying the color on numbered cards.

Unlike Speech with Sam and Speech Adventure, an on-screen character narrates the game. SpokeIt is the first prototype to include completely animated characters with mouth transitions. Unlike Speech with Sam, but similar to Speech Adventure, SpokeIt demonstrates how each target

should be pronounced. SpokeIt is the only prototype that automatically moves on if a player is struggling for over ten seconds.

Figure 5. Top: Character demonstrating /V/ sound, 2nd from Top: Character demonstrating /L/ Sound, Bottom: Characters showing sad and disgusted expressions

Instead of words lighting up green as they are spoken correctly, SpokeIt uses a “Heard” element that displays exactly what speech was recognized—correct pronunciations and incorrect pronunciations. The target word or phrase is displayed in the upper-left corner of the screen. SpokeIt has word, phrase, and sentence targets. The ear in the upper-right is crossed out when the system pauses.

Unlike both predecessors, SpokeIt is completely touch-free. To simplify game directives and required interactions, the only form of input that makes progress in the game is speech. We use the touch-screen to aid players. When an element is tapped, Sam says that word aloud to help players know their targets, which is important to players who cannot read.

Design Overview

Table 2 summarizes some key differences between the three prototypes we developed, namely the game structure style and how the players are prompted on game targets. We hypothesized players would enjoy the hybrid mini-game storybook style SpokeIt provides because it includes fast paced game play surrounded by narrative. We also hypothesized a main character who demonstrates speech with mouth animations would increase usability and therapeutic value.

Prototype	Game Style	Instruction
Speech Adventure	Storybook	Off-Screen Narrator
Speech with Sam	Minigames	Text prompts
SpokeIt	Hybrid	Main character

Table 2: Key characteristics of each prototype

STUDY

After development of the third prototype had been completed, researchers wanted to ensure progress on the design was moving in the correct direction. We wanted to ensure our game was usable and learn what future features or rewards would keep players engaged.

Protocol

We began by administering a preliminary survey to collect demographics, interest in speech therapy games, and general game use.

We then conducted a within-subject comparative study where each participant played each of the three prototypes in a random order. Researchers were present to facilitate, answer questions, and change prototypes when necessary.

Following, a post interview was conducted, where we collected rankings of each of the designs, usability feedback, and core mechanic feedback, asking a mixed set of targeted questions to explore positive and negative characteristics from each of the three prototypes. We also discussed the kinds of reward systems our participants would be interested in. Each participant was asked the same set of questions. Facilitators wrote down answers to each of these questions and also jotted down any quotes or observations they had. Our study was video recorded.

We concluded the study with a 3-question 5-point Likert survey to receive feedback on how well-received our speech recognition system and speech mechanics were. We were interested in its perceived accuracy, responsiveness, and mechanic. Participants were asked if 1) the game accurately heard them, 2) The game responded quickly to speech, and 3) They would want to play at home.

Participants

Our research lab has an on-going relationship with a local day program for adults with developmental disabilities. We asked the program staff to provide us with facilities to conduct our study and to provide us with participants with speech impairments who are legally able to provide consent.

Regarding the demographic information collected, all individuals who attend this day program are adults. Two participants were not comfortable sharing their age but seemed to be similar in age to the other participants. Table 3 below outlines basic demographic information of our 5 participants. One participant had Cerebral Palsy, one had Down Syndrome, one had ASD, and two were diagnosed with mental retardation co-occurring with articulation disorders.

Participant	Age	Sex	Game Play Frequency
P1	27	Male	5 Days/Week
P2	31	Female	2 Days/Week
P3	24	Male	7 Days/Week
P4		Male	7 Days/Week
P5		Male	2 Days/Week

Table 3: Participant Demographics

Facility

We were allowed to use two medium rooms. Because neither of the two rooms was big enough to accommodate the entire group and we had to work within the daily schedule, we split participants between the two rooms to run the study in parallel with all participants. The situation was not ideal, but to remove as much bias as possible, we asked each question to each individual in a random order—meaning P3 might answer question 1 first, then P2, but P1 answered question 2 first followed by P3. Each participant had the opportunity to answer each question before the group moved on to the next question. The facility was also not ideal for the within-subject play of the games. The speech recognition works best in a quiet environment, but this was not possible given the constraints of our facility.

Equipment

We brought enough iPads for each participant and a few extra in case of technical problems, two laptops with webcams to record each of the rooms, surveys, scripts, consent forms, and note-taking materials.

RESULTS

We use handwritten notes from researchers containing participant's responses to questions about relevant reward systems and opinions on each of the 3 prototypes. We use the results of our 3-question Likert survey and participant quotes about the speech recognition to report insights about its use. We use video recordings of the participants playing our 3 game prototypes to identify usability issues, and player reactions.

For our analysis, two researchers created codes and themes while three independent coders analyzed all videos for our qualitative analysis. We use BORIS to analyze the videos using our codes and themes. Three researchers independently analyzed all videos using our BORIS file. The emerging insights include preferred game styles, reward systems, and usability concerns.

All participants play games two or more times per week and would be interested in using games to improve their speech. Participants report that the games they play most commonly are car games, racing games, solitaire, bowling, and NFL sports games. Participants report they have difficulty speaking loudly, have fluency difficulties, and are sometimes unsure of what to say when speaking.

In the following sections, we organize our findings by the specific prototype they relate to. We then report general insights into reward systems our participants are interested in and their experience using the speech mechanics.

Speech with Sam

Researchers observed that two participants were laughing while playing Speech with Sam. This may be because of humorous phrases that are present in the narrative such as, "Slugs don't wear boots!"

Due to software updates, some features of the game became unresponsive, which understandably frustrated many of our players. It was not always obvious to the players that they needed to touch flashing objects to progress in the game. All instructions in the game were displayed as text, but many of our participants could not read, so researchers aided players by dictating the instructions. Players wanted better feedback when interacting because at times, they were unsure if the game accepted or rejected their responses. They were also unsure when they had to repeat themselves. Participants enjoyed the pace of Speech with Sam.

Speech Adventure

The mini-games in speech adventure, particularly in the rocket scene, gave immediate feedback when the player correctly interacted with the game. Many player's visceral reactions to the sounds and animations were very positive. They enjoyed that fireworks celebrated their success with a satisfying pop sound. Speech Adventure kept track of player scores and displayed them to the participants. Two players reported that seeing their score was satisfying and marked their progress.

Players who could not read also struggled with this prototype because the instructions were written out as text and needed the researcher's help to navigate through the game objectives. Many of the mini-games rely on the players to speak in the correct rhythm, but this was extremely challenging for some. The scenes automatically progress after an allotted amount of time. Players found this happened too fast—just as they were beginning to understand the objectives and mechanics, the next game would be displayed. Some of the game objectives were too complicated and several players never learned how to complete objectives in the allotted time. In general, Speech Adventure needs to be slowed down drastically and the instructions need to be clearer.

Spokelt

Three out of five participants preferred Spokelt to the rest of the prototypes. They especially appreciated that all the instructions were spoken allowed by the main character and

displayed as text on the screen. Participants found the interaction objectives much simpler because SpokeIt never requires a player to touch the screen. Many participants found SpokeIt to be incredibly aesthetically pleasing. They loved the colors, graphics, and animations. Participants preferred the highly animated main character in SpokeIt because it represented a more responsive and lively element. One participant was very interested in bringing SpokeIt home with her.

One participant commented that he would like SpokeIt to repeat the instructions because he did not always remember what he was supposed to say. Most users seemed most enthusiastic about playing SpokeIt again, indicating the hybrid structure may improve re-playability of mini-games because they fit within an overall narrative.

Speech Recognition and Mechanics

We report that users are neutral about the accuracy of our speech recognition system (Q1), but found it responded quickly (Q2). They found the speech mechanic was rewarding and enjoyable, and they think the games are suitable systems to promote practicing speech at home (Q3).

	Likert Results		
	Q1	Q2	Q3
Values	4	4	4
	4	4	4
	5	4	5
	2	4	4
	2	4	4
Average	3.2	4	4.2

Table 2: Speech Recognition and Speech Mechanics Likert Survey Results

Rewards

We are interested in rewarding players for practicing speech in a meaningful way to them. Hence, we brainstormed a few ideas with our participants and asked them to vote on which reward would be most interesting to them. Our ideas included:

- Hats and clothes to accessorize a character or avatar after completing sessions
- Using scored points to spend in virtual store. Points could be spent to buy items for a virtual garden or furnishing the character's house.
- Reducing total time needed to practice speech in the future. If a player does really well in a 10-minute session, then tomorrow they only need to play for 8 minutes.
- Out-of-game rewards (Stickers, candy, other physical reward)

Overwhelmingly, our participants were interested in out-of-game rewards. They were extremely excited about the idea of receiving candy when they do well in the game.

Usability Considerations

Watching players use our systems was very informative. We identify 3 main issues that must be addressed in speech therapy systems:

- Many participants cannot read: the game must be very clear and be designed around to accommodate players that cannot read. Objectives should be spoken aloud and be repeated if necessary. If participants are struggling, the game should either change the objectives or move forward with the plot.
- To make progress in the game, players must use their speech. Touch should be used to support players, provide clues, or demonstrate correct pronunciation. These mechanics should not be mixed.
- More feedback for correct and incorrect interactions must be given to make progress clear. If the player pronounces a target incorrectly, the game should support that player in saying it correctly.
- Many users put the iPads to their heads because they had trouble hearing—the game volume must be louder, or headphones must be provided, especially in noisy environments.

Structure

We found that players prefer the hybrid structure because mini-games were given context in an overarching plot. Mini-games that are played out of a narrative context seem to lose their novelty as soon as they are repeated. Our users seemed to care a lot about aesthetics and animations indicating that high levels of polish are important.

Storybook-style games require much more work to generate content and narrative consistency. Speech therapy games should be available and fresh for as long as the individual needs to practice speech. Narrative content that surrounds mini-games is an effective balance of development time, re-playability, and diversity of speech targets.

DISCUSSION

Speech Recognition Accuracy

In the future, we would like to explore improving the accuracy of the speech recognition system. This is obviously an important feature of a speech therapy game, especially for when we explore the clinical validity of the game. There are two controllable factors that we consider—the physical hardware, and the acoustic models. We would like to compare the accuracy of the system when using a noise cancelling microphone and when using the built-in microphone. We would also like to collect speech data from our specific target populations to train the acoustic model and compare it to the default model. Collecting speech data from vulnerable populations is a challenge due to HIPAA regulations in the United States. In this context, speech is

categorized as medical data and must be approved before it can be stored safely. Another challenge is the tedious process of hand-coding the speech data in a way that allows acoustic models to be trained using machine learning techniques.

Rewards

Our brainstorm session about preferred reward systems indicated that our users were not interested in badges, stars, and points—they wanted tangible real-world rewards and commemoration from their mentors. We do not want to simply gamify [25] speech therapy with banners, badges, ribbons, and stars. In the future, we are interested in creating a rewarding experience that integrates into current practices. Our serious game for health should be one component of a broader holistic therapy experience. We envision the final product to have a companion app for speech therapists to assign individualized curriculum goals and receive reports of patient progress. If patients meet these goals, an SLP could reward patients with tangible prizes.

Working with Children with Speech Impairments

As previously stated in the *Design* section, SpokeIt is intended for children with speech impairments, but regular access to this population is difficult, so out of convenience, we conducted our study with adults with developmental disabilities co-occurring with speech impairment. Using a childish design on an adult population is a limitation of this study, but our participants truly enjoyed the experiences and never indicated that they felt patronized. Also, it is worth reiterating that this adult population has the potential to benefit from this work, as they also have speech goals and articulation disorders.

Speech and language development is important for children's future ability to live independently and to participate fully in society [20]. In 2012, nearly 8% of children aged 3-17 in the United States had a communication disorder and younger children, boys, and non-Hispanic white children were more likely than other children to receive an intervention service for their disorder [53]. Children with speech impairments such as Cleft Lip Palate have high risks of behavioral problems and increased symptoms of depression [20]. They show more deficits in social and academic competencies, score higher for social problems [13], and are more likely to be teased in social settings [20]. Even those who undergo a corrective surgery tend to display a delay in scholarship, have a lower income, marry later in life and become independent from their parents significantly later [21]. Clearly, further exploration of using SpokeIt to help children improve their speech is worth exploring.

Methodological considerations

Our users struggled with Likert style questions, so we needed to adapt how we conducted them onsite, which we detail here. This can serve to other researchers working with a similar target group (adults with developmental disabilities).

We first asked whether they agreed with the statement, disagreed with the statement, or did not know. If they did not know, we marked down a 3. If they said they agreed, we

asked if they agreed a lot or a little. If they said they agreed a lot, we put a 5. If they said they agreed a little, we marked down a 4. We followed the same process if they disagreed. If facilitators felt a participant was answering just to please us, we would ask the same questions in the opposite way and remind users that we want them to be authentic. Some participants changed their answers, which lead us to believe our results may not be completely representative of our population. Asking Likert style questions in this way was cumbersome and may result in data that does not represent the population.

ACKNOWLEDGMENTS

This material is based in part upon work supported by the National Science Foundation under Grant number #1617253. We also thank Doctor Travis Tollefson and SLP Christina Roth for their aid in conducting user evaluations. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

REFERENCES

1. Clark C Abt. 1987. *Serious games*. University press of America.
2. Frank R Adams, Hubert Crepy, David Jameson, and J Thatcher. 1989. IBM products for persons with disabilities. In *Global Telecommunications Conference and Exhibition 'Communications Technology for the 1990s and Beyond' (GLOBECOM), 1989. IEEE*, 980–984.
3. Olle Bälter, Olov Engwall, Anne-Marie Öster, and Hedvig Kjellström. 2005. Wizard-of-Oz test of ARTUR: a computer-based speech training system with articulation correction. In *Proceedings of the 7th international ACM SIGACCESS conference on Computers and accessibility*, 36–43.
4. Elizabeth Boyle, Thomas M Connolly, and Thomas Hailey. 2011. The role of psychology in understanding the impact of computer games. *Entertainment Computing* 2, 2: 69–74.
5. EB Brownlie, Joseph H Beitchman, Michael Escobar, Arlene Young, Leslie Atkinson, Carla Johnson, Beth Wilson, and Lori Douglas. 2004. Early language impairment and young adult delinquent and aggressive behavior. *Journal of abnormal child psychology* 32, 4: 453–467.
6. H Timothy Bunnell, Debra M Yarrington, and James B Polikoff. 2000. STAR: articulation training for young children. In *Sixth International Conference on Spoken Language Processing*.
7. Biran Burke. 2016. *Gamify: How gamification motivates people to do extraordinary things*. Routledge.
8. Kerstin Carlstedt, Gunilla Henningsson, and Göran Dahllöf. 2003. A four-year longitudinal study of palatal plate therapy in children with Down syndrome: effects on oral motor function, articulation and communication preferences. *Acta Odontologica Scandinavica* 61, 1: 39–46.

9. Mary Clement and Thomas E Twitchell. 1959. Dysarthria in cerebral palsy. *Journal of Speech and Hearing Disorders* 24, 2: 118–122.
10. Colette Coleman and Lawrence Meyers. 1991. Computer recognition of the speech of adults with cerebral palsy and dysarthria. *Augmentative and Alternative Communication* 7, 1: 34–42.
11. Sebastian Deterding, Staffan L. Björk, Lennart E. Nacke, Dan Dixon, and Elizabeth Lawley. 2013. Designing gamification: creating gameful and playful experiences. In *CHI'13 Extended Abstracts on Human Factors in Computing Systems*, 3263–3266.
12. Jared Duval, Zachary Rubin, Elizabeth Goldman, Nick Antrilli, Yu Zhang, Su-Hua Wang, and Sri Kurniawan. 2017. Designing Towards Maximum Motivation and Engagement in an Interactive Speech Therapy Game. In *Proceedings of the 2017 Conference on Interaction Design and Children*, 589–594.
13. Kristin Billaud Feragen, Ingela L Kvale, Nichola Rumsey, and Anne IH Borge. 2010. Adolescents with and without a facial difference: the role of friendships and social acceptance in perceptions of appearance and emotional resilience. *Body Image* 7, 4: 271–279.
14. Javier Franco-Pedroso and Joaquin Gonzalez-Rodriguez. 2016. Linguistically-constrained formant-based i-vectors for automatic speaker recognition. *Speech Communication* 76: 61–81. <https://doi.org/10.1016/j.specom.2015.11.002>
15. Kathrin Maria Gerling, Regan L Mandryk, Max Valentin Birk, Matthew Miller, and Rita Orji. 2014. The effects of embodied persuasive games on player attitudes toward people using wheelchairs. In *Proceedings of the 32nd annual ACM conference on Human factors in computing systems*, 3413–3422.
16. Donna M Hanson, Alfred W Jackson, Randi J Hagerman, John M Opitz, and James F Reynolds. 1986. Speech disturbances (cluttering) in mildly impaired males with the Martin-Bell/fragile X syndrome. *American Journal of Medical Genetics Part A* 23, 1–2: 195–206.
17. Patricia Howlin, Lynn Mawhood, and Michael Rutter. 2000. Autism and developmental receptive language disorder—A follow-up comparison in early adult life. II: Social, behavioural, and psychiatric outcomes. *The Journal of Child Psychology and Psychiatry and Allied Disciplines* 41, 5: 561–578.
18. D. Huggins-Daines, M. Kumar, A. Chan, A. W. Black, M. Ravishankar, and A. I. Rudnick. 2006. Pocketsphinx: A Free, Real-Time Continuous Speech Recognition System for Hand-Held Devices. In *2006 IEEE International Conference on Acoustics Speech and Signal Processing Proceedings*, I–I. <https://doi.org/10.1109/ICASSP.2006.1659988>
19. David Huggins-Daines, Mohit Kumar, Arthur Chan, Alan W Black, Mosur Ravishankar, and Alexander I Rudnick. 2006. Pocketsphinx: A free, real-time continuous speech recognition system for hand-held devices. In *Acoustics, Speech and Signal Processing, 2006. ICASSP 2006 Proceedings. 2006 IEEE International Conference on*, I—I.
20. Dr Orlagh Hunt, Dr Donald Burden, Dr Peter Hepper, Dr Mike Stevenson, and Dr Chris Johnston. 2007. Parent Reports of the Psychosocial Functioning of Children with Cleft Lip and/or Palate. *The Cleft Palate-Craniofacial Journal* 44, 3: 304–311. <https://doi.org/10.1597/05-205>
21. Mercy Larnyoh. 2015. Determining social challenges of children with cleft lip and or palate as perceived by parents or caretakers at Komfo Anokye Teaching Hospital in Kumasi Metropolis in Ashanti region, Ghana.
22. Debra A Lieberman. 1997. Interactive video games for health promotion: Effects on knowledge, self-efficacy, social support, and health. *Health promotion and interactive technology: Theoretical applications and future directions*: 103–120.
23. Elena Márquez Segura, Annika Waern, Luis Márquez Segura, and David López Recio. 2016. Playification: The PhySeEar case. 376–388. <https://doi.org/10.1145/2967934.2968099>
24. Robert C Marshall, Robert T Wertz, David G Weiss, James L Aten, Robert H Brookshire, Luis Garcia-Bunuel, Audrey L Holland, John F Kurtzke, Leonard L LaPointe, Franklin J Milianti, and others. 1989. Home treatment for aphasic patients by trained nonprofessionals. *Journal of Speech and Hearing Disorders* 54, 3: 462–470.
25. Jane McGonigal. 2011. We Don't Need No Stinkin' Badges: How to Re-invent Reality Without Gamification. Retrieved December 18, 2015 from <http://www.gdcvault.com/play/1014576/We-Don-t-Need-No>
26. Jane McGonigal. 2011. *Reality is broken: Why games make us better and how they can change the world*. Penguin.
27. Ray M Merrill, Nelson Roy, and Jessica Lowe. 2013. Voice-related symptoms and their effects on quality of life. *Annals of Otology, Rhinology & Laryngology* 122, 6: 404–411.
28. David R Michael and Sandra L Chen. 2005. *Serious games: Games that educate, train, and inform*. Muska & Lipman/Premier-Trade.
29. Tony Morelli, Lauren Lieberman, John Foley, and Eelke Folmer. 2014. An exergame to improve balance in children who are blind. In *FDG*.
30. Megan A Morris, Sarah K Meier, Joan M Griffin, Megan E Branda, and Sean M Phelan. 2016. Prevalence and etiologies of adult communication disabilities in the United States: Results from the 2012 National Health Interview Survey. *Disability and health journal* 9, 1: 140–144.
31. Jack Mostow and others. 2001. Evaluating tutors that listen: An overview of Project LISTEN. In *Smart machines in education*, 169–234.

32. Cosmin Munteanu, Joanna Lumsden, Hélène Fournier, Rock Leung, Danny D'Amours, Daniel McDonald, and Julie Maitland. 2010. ALEX: mobile language assistant for low-literacy adults. In *Proceedings of the 12th international conference on Human computer interaction with mobile devices and services*, 427–430.
33. Joan Murphy. 2006. Perceptions of communication between people with communication disability and general practice staff. *Health Expectations* 9, 1: 49–59.
34. Heidi Parisod, Anni Pakarinen, Anna Axelin, Riitta Danielsson-Ojala, Jouni Smed, and Sanna Salanterä. 2017. Designing a Health-Game Intervention Supporting Health Literacy and a Tobacco-Free Life in Early Adolescence. *Games for Health Journal*.
35. Larry J Platt, Gavin Andrews, Margrette Young, and Peter T Quinn. 1980. Dysarthria of adult cerebral palsy: I. Intelligibility and articulatory impairment. *Journal of Speech, Language, and Hearing Research* 23, 1: 28–40.
36. Alessandra Preziosa, Alessandra Grassi, Andrea Gaggioli, and Giuseppe Riva. 2009. Therapeutic applications of the mobile phone. *British Journal of Guidance & Counselling* 37, 3: 313–325.
37. Josephine M Randel, Barbara A Morris, C Douglas Wetzel, and Betty V Whitehill. 1992. The effectiveness of games for educational purposes: A review of recent research. *Simulation & gaming* 23, 3: 261–276.
38. William M Reynolds and Susan Reynolds. 1979. Prevalence of speech and hearing impairment of noninstitutionalized mentally retarded adults. *American journal of mental deficiency*.
39. John C Rosenbek, Margaret L Lemme, Margery B Ahern, Elizabeth H Harris, and Robert T Wertz. 1973. A treatment for apraxia of speech in adults. *Journal of Speech and Hearing Disorders* 38, 4: 462–472.
40. Zachary Rubin. 2017. Development and evaluation of software tools for speech therapy. University of California, Santa Cruz.
41. Zachary Rubin, Sri Kurniawan, and Travis Tollefson. 2014. Results from using automatic speech recognition in cleft speech therapy with children. In *International Conference on Computers for Handicapped Persons*, 283–286.
42. Zak Rubin and Sri Kurniawan. 2013. Speech Adventure: Using Speech Recognition for Cleft Speech Therapy. In *Proceedings of the 6th International Conference on Pervasive Technologies Related to Assistive Environments (PETRA '13)*, 35:1–35:4. <https://doi.org/10.1145/2504335.2504373>
43. Ellen Sciuto. 2013. The iPad: Using new technology for teaching reading, language, and speech for children with hearing loss.
44. Lawrence D Shriberg, Rhea Paul, Jane L McSweeny, Ami Klin, Donald J Cohen, and Fred R Volkmar. 2001. Speech and prosody characteristics of adolescents and adults with high-functioning autism and Asperger syndrome. *Journal of Speech, Language, and Hearing Research* 44, 5: 1097–1115.
45. Ann Bosma Smit. 2004. *Articulation and phonology resource guide for school-age children and adults*. Cengage Learning.
46. Ali JA Soleymani, Martin J McCutcheon, and MH Southwood. 1997. Design of speech illumina mentor (SIM) for teaching speech to the hearing impaired. In *Biomedical Engineering Conference, 1997., Proceedings of the 1997 Sixteenth Southern*, 425–428.
47. Jonathan M Sykes and Travis T Tollefson. 2005. Management of the cleft lip deformity. *Facial plastic surgery clinics of North America* 13, 1: 157–167.
48. John Van Borsel and An Vandermeulen. 2008. Cluttering in Down syndrome. *Folia Phoniatrica et Logopaedica* 60, 6: 312–317.
49. Klara Vicsi, Peter Roach, A Öster, Zdravko Kacic, Peter Barczikay, Andras Tantos, Ferenc Csátári, Zs Bakcsi, and Anna Sfakianaki. 2000. A multimedia, multilingual teaching and training system for children with speech disorders. *International Journal of speech technology* 3, 3–4: 289–300.
50. Maria Virvou, George Katsionis, and Konstantinos Manos. 2005. Combining software games with education: Evaluation of its educational effectiveness. *Educational Technology & Society* 8, 2: 54–65.
51. Laurie A Vismara, Costanza Colombi, and Sally J Rogers. 2009. Can one hour per week of therapy lead to lasting changes in young children with autism? *Autism* 13, 1: 93–115.
52. Charles S Watson, Daniel J Reed, Diane Kewley-Port, and Daniel Maki. 1989. The Indiana Speech Training Aid (ISTRA) I: Comparisons between human and computer-based evaluation of speech quality. *Journal of Speech, Language, and Hearing Research* 32, 2: 245–251.
53. 2015. *National Center for Health Statistics*. Centers for Disease Control and Prevention. Retrieved from <https://www.cdc.gov/nchs/products/databriefs/db205.htm>
54. 2016. *Statistics on Voice, Speech, and Language*. U.S. Department of Health and Human Services. Retrieved from <https://www.nidcd.nih.gov/health/statistics/statistics-voice-speech-and-language#5>
55. *Quick Statistics About Voice, Speech, Language*. National Institute on Deafness and Other Communication Disorders. Retrieved from <https://www.nidcd.nih.gov/health/statistics/quick-statistics-voice-speech-language>