Architecture of an automated therapy tool for childhood apraxia of speech

Avinash Parnandi[1], Virendra Karappa[1], Youngpyo Son[1], Mostafa Shahin[2], Jacqueline McKechnie[3], Kirrie Ballard[3], Beena Ahmed[2], and Ricardo Gutierrez-Osuna[1]

[1]Department of Computer Science and Engineering, Texas A&M University, College Station, TX 77843, USA
[2]Department of Electrical and Computer Engineering, Texas A&M University, Doha 23874, Qatar
[3]Faculty of Health Sciences, The University of Sydney, Sydney, NSW 2141, Australia

[1]{parnandi,vkarappa1,yson,rgutier}@tamu.edu, [2]{mostafa.shahin,beena.ahmed}@qatar.tamu.edu, [3]{jacqueline.mckechnie, kirrie.ballard}@sydney.edu.au

ABSTRACT

We present a multi-tier system for the remote administration of speech therapy to children with apraxia of speech. The system uses a client-server architecture model and facilitates task-oriented remote therapeutic training in both in-home and clinical settings. Namely, the system allows a speech therapist to remotely assign speech production exercises to each child through a web interface, and the child to practice these exercises on a mobile device. The mobile app records the child's utterances and streams them to a back-end server for automated scoring by a speech-analysis engine. The therapist can then review the individual recordings and the automated scores through a web interface, provide feedback to the child, and adapt the training program as needed. We validated the system through a pilot study with children diagnosed with apraxia of speech, and their parents and speech therapists. Here we describe the overall client-server architecture, middleware tools used to build the system, the speech-analysis tools for automatic scoring of recorded utterances, and results from the pilot study. Our results support the feasibility of the system as a complement to traditional face-to-face therapy through the use of mobile tools and automated speech analysis algorithms.

Categories and Subject Descriptors

K.3.1 [**Computers and Education**]: Computer Uses in Education– *Computer-assisted instruction (CAI) Computer-managed instruction (CMI).*

K.4.2 [**Computers and Society**]: Social Issues – *Assistive technologies for persons with disabilities.*

I.2.7 [**Artificial Intelligence**]: Natural Language Processing– *Speech recognition and synthesis.*

General Terms

Experimentation, Human Factors.

Keywords

Childhood apraxia of speech, speech therapy, automated speech analysis

1. INTRODUCTION

Childhood apraxia of speech (CAS) is a neurological pediatric speech sound disorder (SSD) that debilitates oro-motor planning, and execution. CAS can delay acquisition of skills including the control of tone, breathing, intensity, and vocalization [3]. It also impairs the child's ability to correctly pronounce sounds, syllables, and words. CAS can thus render the child unable to start articulating the first sounds and words and can lead to a serious communicative disability. CAS can be difficult to diagnose and monitor due to a high co-morbidity with other speech and language disorders and a lack of specific tools [22]. It is known that, by working intensely with a trained speech therapist, those with CAS can overcome their motor planning and motor programming difficulties (articulation capabilities) [3].

However, the ratio of children with CAS to the number of qualified clinicians is growing at a high rate. According to the literature [14], current estimates of children with CAS fall between 5-6%. Due to the increasing number of children needing intervention and the shortage of trained therapists, there is an ever-increasing gap between the quality and duration of needed therapeutic interventions and what is available (because of time constraints and expenses) [3]. Thus, there is a need for practical and cost-effective technological interventions to complement traditional face-to-face therapy sessions. CAS therapy usually comprises of verbal, auditory, and visual interaction between a therapist and the child using game-like activities [31], which makes it a good candidate for technology-based alternative solutions, as these can provide not only remote and automatic monitoring but also interactive training.

As a step towards this goal, we present a multi-tier system that enhances the administration of traditional CAS therapy for use in in-home settings. We adopt the Nuffield Dyspraxia Program (NDP3), an assessment and intervention package for children with severe speech sound disorders including CAS [5, 31]. NDP3 follows a bottom-up approach and provides speech therapy that can be adapted to the child's individual needs and progress; this makes NDP3 ideal for our purposes. As illustrated in Figure 1, our system consists of three major components: (1) a mobile app

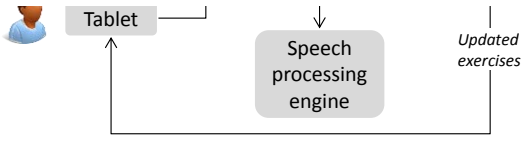

Figure 1. General overview of the CAS therapy system showing the server, mobile clients, and the remote therapy management system.

ASSETS '13, October 21 - 23 2013, Bellevue, WA, USA

running on a tablet, which allows the patient to practice speech exercises at their homes, (2) a therapy management interface, which allows clinicians to assign exercises and monitor progress, and (3) a speech processing engine, which performs automated diagnostics on the patient's recordings. This enables remote therapy sessions where the therapist and the patients need not be in the same geographical location, and helps overcome barriers of access to speech therapy due to distances, lack of specialists and equipment, etc. This architecture differentiates our system from existing mobile speech therapy tools [1, 2, 6, 7], which are standalone applications with no automated speech assessment or remote monitoring capabilities.

We have conducted an initial pilot study with children with CAS/SSD to test the complete system during speech therapy. During the study, CAS patients performed NDP3 exercises on the tablet under the supervision of a therapist. Then, we conducted a semi-structured interview with patients, parents, and therapists to explore issues related to the usability of the system (tablet and clinician interface), level of engagement of the children during therapy, preferences between tablet and paper-based therapy, and areas of improvement for the system.

The rest of the paper is organized as follows. Section 2 provides a brief overview of child apraxia of speech, the Nuffield dyspraxia program, and summarizes past work on computerized speech therapy. Section 3 presents our system architecture, including the mobile client, the server, clinician web interface, and speech processing engine. Section 4 describes the experiments used to validate the system and summarizes comments and recommendations from the participants. The paper concludes with a discussion and directions for future work in Section 5.

2. BACKGROUND AND RELATED WORK

2.1 Childhood apraxia of speech

The American Speech-Language-Hearing Association (ASHA) defines childhood apraxia of speech (CAS) as a "*neurological childhood (pediatric) speech sound disorder in which the precision and consistency of movements for underlying speech are impaired in the absence of neuromuscular deficits (e.g., abnormal reflexes, abnormal tone)*" [3]. Although children with CAS usually have no damage to muscles or nerves, the area of the brain sending signals to the muscles is damaged or not fully developed; thus, CAS patients have marked difficulty in motor programming, motor planning, and correctly producing sounds, syllables, and words [28]. In addition, such children often have an oro-motor dyspraxia, a difficulty in coordinating precise and consistent movements of the articulators (tongue, lips, jaw, and palate) required to produce speech (and to achieve an acceptable pronunciation of a given word) [9]. The speech of children with CAS is usually unintelligible to unfamiliar listeners due to phonemic speech errors and articulatory abnormalities. In the absence of treatment, this neuromuscular developmental disability can delay the acquisition of speech skills and phonological abilities, thus causing severe communicative disability [12]. Hence accurate and timely intervention is critical for children with CAS. Intervention involves repeated speech therapy sessions between the therapist and child, which can continue for several years.

2.2 The Nuffield Dyspraxia Programme

Our proposed system is based on the Nuffield Dyspraxia Programme (NDP3), an intervention program for children in the developmental age range of 3-7 years with severe speech sound disorders including CAS [5, 31]. NDP3 has been designed to address the effects of CAS such as articulation of individual consonants and vowels, sequencing sounds together, and maintaining prosodic accuracy [5, 31]. Therapy for childhood apraxia comprises of two components: assessment and intervention. In the NDP3 protocol, assessment and treatment for children with CAS is performed using a bottom-up approach; see Figure 2(a). An initial assessment provides a measure of the child's current speech skills based on which the therapy is designed, starting from isolated speech sounds and progressing to complex syllable structures, then to sentences, connected speeches, and the full range of speech [5, 31]. These tests also indicate the extent to which an intervention may be effective in helping a child regain a dimension of oro-motor functionality.

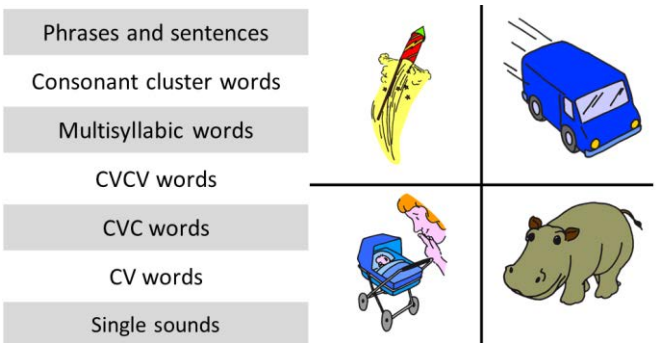

Figure 2. (a) NDP3 "brick wall" showing the bottom-up therapy approach. (b) Sample NDP3 exercise with four stimulus images [5].

The NDP3 protocol requires regular therapy sessions under the supervision of a speech therapist. The intervention approach comprises of a therapy manual, 1,800 picture cards involving 750 different images and 550 line-drawn worksheets. This includes a set of picture cues to represent single consonants, vowels, diphthongs, and words at each of the phonotactic levels; see Figure 2(b). Word creation by joining of sounds or syllables is facilitated by transition worksheets, while sequencing worksheets provide repetitive practice. Guided by the instruction manual, the picture cards are presented to the child as stimuli through tabletop games to elicit the target utterance. The child is asked to produce specific sounds, syllables, or words compliant with their therapy level in these activities. On correctly completing an event, the therapist usually presents the child with simple rewards. On an incorrect response, the therapist follows the instructions to assist the child so as to elicit the correct response. This approach works from the child's strengths and builds skills in incremental steps in a cumulative way [5, 31]. The NDP3 assessment follows a multi-layer approach where the bottom layer consists of single speech sounds, the next layer is CV words, followed by CVC, CVCV and multisyllabic words, consonants clusters, and finally connected speech in the form of phrases and sentences; see Figure 2(a). It relies on the production of 1) all the single consonants, vowels, and diphthongs, 2) a set of 20 single words at each phonotactic structure (CV/VC, CVCV, CVC, CCV and multi-syllabics) through picture naming [C: consonants, V: vowels], and 3) phrases and sentences through imitation with pictures.

2.3 Previous work on computerized therapy

Automated therapy is a subcategory of technological approaches to health care known as "tele-medicine" or "tele-practice". Traditional CAS therapy requires a child to undergo extended

therapy sessions with a trained speech therapist in a clinic. This can be both logistically and financially prohibitive, thus paving the way for remote and automated therapy tools. Studies have shown that children with SSD/CAS have higher levels of engagement and reduced error response with computer-based intervention as compared to traditional therapy [16]. Waite et al. [29] investigated the feasibility of remote assessment of childhood SSDs by therapists and compared it with face-to-face interaction. They found a high level of agreement between the two methods (single-word articulation: 92%; speech intelligibility: 100%; and oro-motor tasks: 91%). Further, researchers have also shown that internet-based tele-rehabilitation sessions can be as effective as clinic-based sessions [11]. These previous studies have emphasized the impact that interactive and remote speech therapy can have.

A number of tools have been developed to facilitate general speech therapy. Speech recognition software such as Dragon Dictate and other speech processing techniques have been used for the assessment of pathological disorders where the acoustic characteristics of the voice produced are affected due to laryngeal and vocal tract disorders [17, 18, 21]. However in speech sound disorders such as CAS, the child's voicing is unaffected; they instead struggle with articulation errors. General tools to facilitate speech therapy include STAR (Speech Training, Assessment, and Remediation), which assists therapists in treating children with articulation problems) [10], and Ortho-Logo-Paedia (OLP), which is intended for use in in-home settings [23]. However, these systems do not cater to the specific articulation problems of children with CAS and other SSDs.

In the specific context of CAS and SSD, a few software programs and tele-rehabilitation tools have also become available [15], such as Phoneme Factory Sound Sorter (PFSS) [32], Sound Contrasts in Phonology (SCIP) [30], and Speech Assessment and Interactive Learning Systems (SAILS) [25]. These tools assist the child in developing phonological patterns and phonemic contrasts. The main drawback of these systems, however, is the absence of automatic feedback, which makes it hard to adapt the therapy regimen on-the-fly based on the specific needs of each child.

Mobile technology provides opportunities to gain richer data and improve the experience of patients undergoing clinical interventions. Touch-based devices such as tablets and smartphones are intuitive and engaging as compared to desktop and paper-based alternatives, and are also very cost effective for in-home therapy sessions. This has led to the development of generic speech therapy applications for mobile platforms, such as PocketSLP [6], ArtikPix [2], and Speech with Milo [7], which usually focus on articulation problems. Of particular interest in our case is Apraxiaville [1] a mobile tool developed for children with apraxia. It includes features such as voice recording, self-scoring, and animated stimuli. However, this app only supports three levels of therapy (single sound, CV-VC-CVC, and multisyllabic words), whereas the NDP3 protocol in addition also supports cluster word and phrase and sentence formation. Furthermore, most of the current apps, including Apraxiaville, are standalone tools with no remote and automated speech assessment or feedback capabilities. In contrast, our system includes automated speech-processing capabilities to provide timely feedback to both the therapist and the child, and also the ability to manage multiple patients remotely.

3. SYSTEM ARCHITECTURE

To remotely administer NDP3 therapy at the home, the system should be able to: (1) prompt the child with the appropriate stimuli on the mobile platform, (2) record the child's speech response and stream it back to the server, (3) identify the individual consonants and vowels produced, and the errors made, through speech analysis algorithms, (4) provide feedback to the child, (5) provide reports to the therapist detailing the child's progress, and (6) facilitate the creation or modification of exercises by the therapist based on performance results. This section explains the various modules in the system we have designed for this purpose and the software technologies that supported the development.

3.1 Mobile client

The mobile client provides an alternative to (and is modeled after) the face-to-face session between the therapist and the child. The application provides visual stimuli to the child, records his/her speech response, and provides feedback. Following recommendations in the literature [8, 19, 24], we designed a user interface specially tailored for children. Children have lower manual dexterity and generally less experience with tablets than adults. This results in unique behaviors when compared to those

(a) Home screen

(b) Exercise screen

Figure 3. Screenshots of the mobile app. (a) Home screen shows the different courses that a child is enrolled in, and the exercises in one of those courses (VC words). (b) Images in an exercise as seen by the child during speech practice; the images correspond to words *up, ice,* and *off* respectively. Also shown are the buttons- home, record, and help.

observed with adults, such as *holdovers* (hitting a button a few times for a single action), unintentional swipes, and problem in target acquisition (e.g., difficulty in performing the precise motion for selecting a target and lifting without slipping [13]). Thus, we designed the UI to minimize these types of errors; as an example, to avoid holdovers the interface has distinct buttons to start, stop, and replay the recording of the speech response. Once the *record* button is pressed, the *stop* button shows up on the screen and the *record* button becomes inactive. Thus even if the child presses the *record* button multiple times it will not lead to multiple recordings. After stopping, the *play and record* buttons are reactivated to play the recorded speech or to record a new utterance.

Figure 3 shows the user interface for the child. The app provides a home screen that allows the child to browse through the various courses remotely assigned by the therapist, each course comprising of multiple exercises. The child can choose any of the available courses for practice. Once the child selects a particular exercise in a course, he/she is presented with a set of stimulus images; see Figure 3(b). Tapping on the red circle will start the recording of an utterance. Once finished, the child can playback the recorded sound, practice the stimuli again or move to next image by tapping on the corresponding image. Speech utterances are time-stamped and uploaded to the server asynchronously in the background. Tapping on the help button (question mark icon in Figure 3(b)) provides assistance to the child by showing the word corresponding to the stimulus image.

The mobile app was developed using the PhoneGap and jQuery framework; see Figure 4. PhoneGap is an open source framework for creating native applications with web technologies. It provides APIs to access hardware features of the device, such as network, microphone, storage, etc. It also allows the usage of jQuery Mobile (a specialized JavaScript framework) which supports asynchronous activity, e.g. uploading audio recording on the background. To further minimize latency, the app stores the entire collection of images in the NDP3 protocol on the tablet. In this way, on assigning a new exercise, only the metadata and not the complete images need to be downloaded from the server.

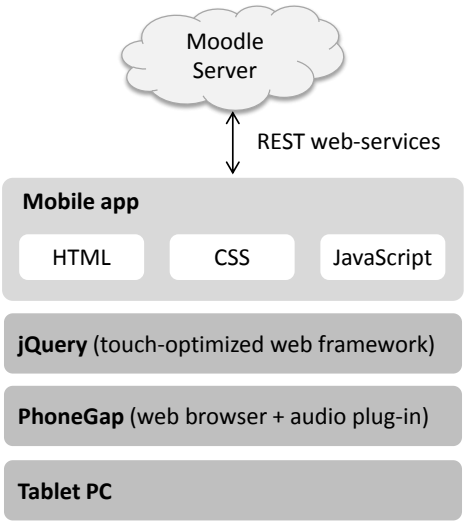

Figure 4. Mobile client architecture showing the software tools used for developing the mobile app user interface and protocol used for communication with the server.

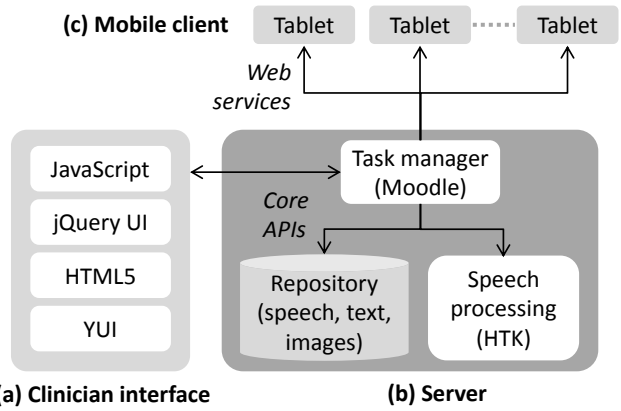

Figure 5. Software components used to develop (a) the clinician interface and (b) server architecture. (c) Mobile clients.

3.2 Server

The server provides logic control to manage therapy for multiple patients, create and store therapy exercises, and store incoming speech recordings from patients; it also hosts the speech analysis engine. Upon receipt of each recording, the task manager invokes the speech-analysis routines and stores the returned results on a centralized relational database. The database also provides storage for the profile for each patient, including their history of speech recordings, results of speech processing, clinician's assessments and annotations.

The server runs Moodle [4], an open-source learning management software, which acts as the task manager. Moodle's modular and object-oriented architecture facilitates information sharing, and integration with other software. It also provides course management functionalities such as user profiles, course pages, secure access by the user, and scheduling of events. Moodle is an example of a LAMP stack (LAMP: Linux, Apache, MySQL, and Perl) [20]. It comes with a web server (Apache), a database (MySQL) and a scripting interpreter (PHP). In our current version, the task manager and other applications run on a Linux machine with Ubuntu Server OS 12.04.

Our framework makes extensive use of Moodle's web services and core APIs; see Figure 5. *Web services* allow us to create fully parameterized generic methods and provide seamless communication between the server, the mobile client, and external applications. For example, the mobile client uses web services to download therapy exercises from the server, and to upload the recorded speech files in the audio repository on the server for speech analysis. Along with this, we have also developed a web service to facilitate real-time feedback from the speech analysis engine; this will allow us to provide rewards to the child based on progress during a therapy session.

The *Moodle core APIs* provide a number of tools for data manipulation, enrollment, secure access management, reporting, etc. Our implementation uses the data manipulation APIs to manage therapy courses, enrolments APIs to manage enrolment of patients to these courses, access APIs to provide secure access to the therapists, and file APIs to store NDP3 images and related files into the server. We have also developed a reporting API which provides data for the reporting UI in the clinician interface.

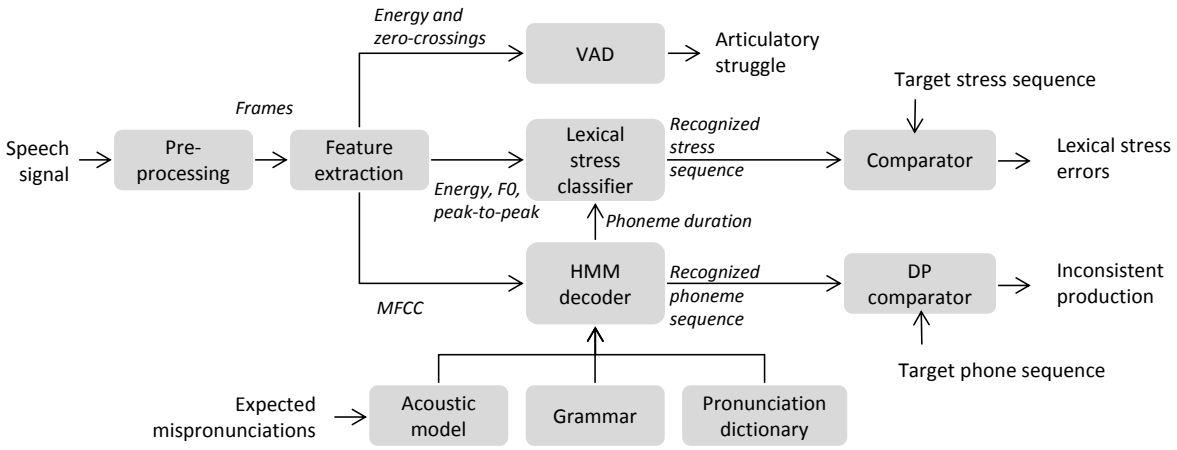

Figure 6. Block diagram of the speech-analysis engine

To facilitate communication between the Moodle database and the speech-processing module, we have created a background process which serves as an interface between the two. This process avoids software dependency and concurrency issues between the Apache server and the speech-processing module, and allows asynchronous communication. Namely, the background process periodically queries the audio repository for recent uploads of audio files, invokes the speech-analysis scripts, and uploads the results to the Moodle database.

3.3 Clinician interface

The clinician UI provides three basic functionalities: managing patients, creating new exercises, and monitoring progress, all done remotely over the web. The *patient management* functionality allows the clinician to add/remove patients and view their profiles. Each profile stores the name, picture, and date of birth of the child, parent's name and contact, and time/date of last access. It also stores all the courses to which the patient has been enrolled and the currently available (unenrolled) courses. The interface also provides the means for the therapist to *create (or edit) exercises* based on the NDP3 protocol. The exercise-creation functionality consists of a canvas, which allows the therapist to create exercises using a drag-and-drop selection of stimuli (i.e., NDP3 images), and a textbox to provide comments and instructions for each exercise. The third functionality provides the clinician with the ability to *view results and reports*. With this, the clinician can analyze the performance of each child, play recordings of individual utterances, and produce comprehensive reports.

As illustrated in Figure 5, the clinician's UI is implemented using JavaScript, PHP, and HTML5. JavaScript is an object-oriented, client-side scripting language and it allows for the development of enhanced UI. The drag-drop feature has been built using HTML5 and the jQuery UI library, while the functionality for data handling and communication with the server is provided by PHP.

3.4 Speech analysis

The speech-analysis module identifies errors made in the child's utterance based upon criteria recommended by ASHA for the analysis of CAS [3, 12], and quantifies them for presentation to the therapist. The three segmental and suprasegmental features of CAS validated by ASHA are (1) inconsistent errors on consonants and vowels in repeated productions of syllables or words and differential use of a certain phoneme or sound class in different word positions (i.e. inconsistency and variability) [3], (2) lengthened and disrupted coarticulatory transitions or struggle between sounds and syllables (i.e. articulatory struggle) [12], and (3) inappropriate prosody, especially in the realization of lexical stress [9, 28]. Figure 6 shows a block diagram of the speech processing steps used to identify errors in the child's utterance. These steps include:

- **Preprocessing**: This stage removes the DC offset, applies a pre-emphasis filter, and segments the speech signal into 25 ms frames (15 ms overlap).
- **Feature extraction**: Several kinds of features are extracted for each frame. The average energy and zero-crossing rate of each frame is calculated and passed to a voice activity detector. The maximum and average pitch and the peak-to-peak amplitude are also calculated and fed to a lexical stress classifier along with the average energy. Finally, Mel Frequency Cepstral Coefficients (MFCCs) are extracted and fed as inputs to a speech decoder.
- **Voice activity detector (VAD)**: A VAD is used to discriminate between speech and non-speech (silence) segments on a frame-by-frame basis based on average energy and zero-crossing rate. Frames identified by the VAD as containing non-speech will indicate the presence of articulatory struggle, specifically '*groping errors*' made by the child, i.e. frames where the child has struggled to produce the requirements.
- **Lexical stress classifier**: This module is used to classify the child speech into strong-weak (SW) and weak-strong (WS) stress patterns. The classifier is based on a multilayer perceptron with the input feature vector consisting of the following acoustic measures: mean and maximum energy over nucleus (the syllable vowel), mean and maximum pitch over nucleus, peak-to-peak amplitude over nucleus, and durations of the syllable and nucleus. Duration information is obtained from the decoder, which provides the recognized phoneme sequence with time boundaries. A pairwise variability index (PVI) is calculated for each acoustic measure to determine the degree of asymmetry across pairs of neighboring syllables and to make the features speaker independent [9]. PVI for any acoustic feature x_i is given by:

$$PVI_i = \frac{x_i^{(1)} - x_i^{(2)}}{(x_i^{(1)} + x_i^{(2)})/2}$$

where $x_i^{(1)}, x_i^{(2)}$ are the acoustic features of the first and second syllables consecutively. The classified stress patterns are then compared with the correct (target) stress patterns to identify the lexical stress errors made by the child. Preliminary results (not included here) show that the classifier has an overall accuracy of 88.7% (SW: 91.1%; WS: 85.9%) [26].

- **HMM decoding**: Mispronunciations in the child's utterance are detected by means of an HMM (hidden Markov model) decoder. For this purpose, we use a grammar lattice consisting of the correct phoneme sequence (known from the exercise given to the child) and expected mispronunciations of each phoneme (recorded by a therapist after assessment of 20 children with CAS); this lattice is used by an HMM decoder along with acoustic models to generate a sequence of phonemes from the child's utterance. The HMM acoustic models consist of tied-state triphones trained using 40 hours of child speech corpus from Oregon Graduate Institute of Science and Technology (OGI) [27]. The recognized phoneme sequence is then compared to the target phoneme sequence through a dynamic-programming string alignment procedure using HResults tool in the HTK toolkit [33]. Three kinds of mispronunciations are identified: insertion, deletion and substitution mispronunciations. These mispronunciations are used to identify inconsistent and variable speech produced by the child.

4. VALIDATION STUDIES

We designed a technology feasibility (pilot) study to determine if our proposed system would be a reliable alternative to traditional therapy. We wanted to observe the kinds of interaction between the child and the tablet app in the context of a speech therapy session, and explored the advantages (and shortcomings) of tablet-based therapy as compared to traditional paper-based exercises. In what follows we describe our experimental protocol, participant demographics, and results from the study.

4.1 Participants

Four children (three male and one female, ages 3-7) and four speech therapists participated in our pilot study. All children were clinically diagnosed with apraxia of speech (3 with medium and 1 with severe apraxia) and were already undergoing speech therapy. Hence, the therapists were aware of the speech sound skills and disabilities of each child. Likewise, the children had prior experience with the traditional picture-card based NDP3 therapy procedure but no experience with the tablet based therapy. We received approval from the Institutional Review Board prior to the study. During the study, parents were given the option to accompany the children for the session or leave the clinic; one of the parents opted to leave the child with the therapist for the session. No compensation was provided to the participants.

4.2 Experiments

During the experiments each child underwent an NDP3 session with the therapist using a tablet (Samsung Galaxy Tab 2 10.1, Android 4.1) running the CAS mobile app (see section 3.1). Prior to the start of the session, the therapist briefed the participants about the goal of the study, and then asked the parent(s) to sign a consent form. The therapy session progressed with the therapist assigning the exercises to the children from the NDP3 protocol. The experimental session was similar to previous sessions with the therapist, with the exception that paper-based exercise sheets were replaced by the tablet and exercises were assigned using the clinician's interface on the Moodle server. At the start of each session, the child would sit with the therapist with the tablet placed on a desk in front of them. The therapist would explain the current activity (single sound, transition etc.) to the child, following which the speech practice would start. Each session lasted for 10 minutes and each child practiced approximately 10 words (unless they wanted to practice more). Assistance was provided if children had difficulty with the tablet at any point during the exercise. After each session, we conducted a semi-structured exit interview using the questionnaires in Table 1, which were designed to capture the participant's preference between the tablet and the paper-based exercises, as well as their likes and dislikes about the system.

Table 1. Exit questionnaires for children, clinician and parents

Child questionnaire
What did you think about doing exercises on the tablet?
What did you like about the exercises on the tablet?
What did you not like about the exercises on the tablet?
What can we do to make the exercises more fun?
Would you like to play with exercises on tablet again?
Would you prefer to do these exercises on the tablet or using paper cards and worksheets?

Clinician and parent questionnaire
Please describe the child's behavior/responses while completing the activity on the tablet.
Was the child able to maintain attention on exercises?
What did you like about the exercise on the tablet?
Did the child need any help completing the exercises?
What did you not like about the exercises on the tablet?
What could we do to make the exercises more engaging for the child and the overall application more usable?
If this application was available to you, how often would you want to use it in clinic/home?
Would you prefer that the children use tablet or paper to complete their at home exercises?

4.3 Discussion of results

Overall, we found that all participants (children, parents and clinicians) enjoyed working with the tablet. When asked *what did you think about doing the exercises on the tablet*? The most common response from children was "being able [to] record and listen to my own voice;" parents and therapists also indicated that the option for children to record/playback their speech was very appealing. Children also commented that the app was "fun", "good", or "I like working with tablets;" in reference to one of the images in the exercises, one of the children (C4) responded "I do like the cow, I like the cow. The cow that walked." When asked to describe the behavior/response of the child, therapists and parents found children very engaged during the therapy session. One parent (P3) commented that his "child always enjoys working with

iPad and tablets". However, one parent (P4) said "[his child] tried initially, but then lost interest – not fun enough."

Therapists also evaluated the clinician UI on the server. In reference to our adoption of the NDP3 protocol, one therapist (T3) commented "A familiar product so comforting that you know it (similar to Speech Builder marketed by Nuffield);" she also stated that "I like the idea of the server and the clinician remotely controlling the exercises that are released and monitoring progress and updating the exercises remotely". Another therapist (T4) pointed out the convenience of being able to listen to recordings immediately, and that our system compared favorably with other software products: "The server and ability to listen to recordings and change exercises immediately and monitor how much practice the child is doing (and how successfully) is great! I think most SLPs would be very interested to buy this package. There are a lot of apps available for speech pathology now but none of them are any good."

When asked if *children needed any help completing the exercise*? Parents P1 and P3 stated "Initially. He was able to work independently after a few trials" and "only needed a little demonstration," respectively. In general, children needed some initial demonstration and help if they got confused with the UI but otherwise were able to perform the tasks independently. This trend was seen across all children except child C4, who required continued help in working with the application. In reference to this child, the therapist (T4) said that "she needed prompts and guidance to scroll down list of exercises, to press stop at the end of recordings," but also suggested this was due to the child's age (C4 was the youngest child in the group); therapist (T4) also indicated that in its current form the system is suitable for children to use in their homes.

When asked *what could we do to make the exercises more engaging?* Many parents and children suggested incorporating rewards, animations, and background music. This sentiment was also echoed by therapists, who suggested we use more colors and animations that pop up as a reward after each stimulus. One therapist (T2) noted that the "current activities on the tablet are not interesting enough to distinguish it from paper-based exercises". He said that providing rewards in the form of badges, stars, cheering sound etc. would lead to positive reinforcement and make the therapy more engaging. Another therapist (T4) suggested progressive goals i.e. "rewards building to certificates, with a goal to get a certain number of stars a week. This goal might be low for young children or when the child is just beginning an activity and then would get higher as the activity got easier or the child older". They also suggested the use of games and puzzles to make therapy more engaging. Parents (P1, P2, and P4) suggested some form of real-time feedback from the system such as "well done" or "try again".

When asked *what could we do to make the overall application more usable*? Parent (P3) mentioned the need for audio prompts associated with each NDP3 image to guide the child in making the correct utterances. Another parent (P1) suggested showing the target word under each image to help the child to make the correct utterance[1]. Therapists also stressed the need for audio prompts accompanying the help text, to help children gradually learn the letter-to-sound relationships. One therapist (T4) suggested the use of an animated visual aid for phoneme production to help the child visualize lip/mouth movements. Therapists also requested a feature for controlling the number of exercises that appear in each course page on the tablet –see Figure 3(a). One therapist (T2) also said - "it was not obvious that there can be more than two-three pictures in an exercise unless you tap the furthest image. In addition, therapists suggested an offline-mode where sessions can be conducted without internet access[2]. This would require that the tablet locally stores all the speech recordings and sync with the server depending on internet/bandwidth availability. Finally, one child (C1) found the system to be slow, while another child (C3) felt the default audio instructions in the mobile app occurred too frequently: "lady speaking too much". Therapists also emphasized the need for real-time speech error detection to reduce mistakes in the utterances made by the children; as an example, one therapist (T3) stated that "I would want an alert to tell the child to stop practice if s/he is off target. Maybe cue the child to stop practice after making 3 or 4 errors on one item". Another therapist (T4) noted that "If you put speech recognition in and some games, and make the reward part more interesting, then that with the speech pathologist's interface will make it very appealing". These comments strongly support our overall design, which includes an automated speech-analysis module –see section 3.4.

Overall, children, parents, and therapists in our pilot study had a positive response towards this new therapy tool when compared to paper-based CAS therapy. All children said that they would like to do the exercises again; interestingly, one child (C1) asked if he could take the tablet home. Parents also commented that if the tablet and the app were available in their home they would use it daily with their children. Finally, therapists liked the idea of remotely controlling the exercises and monitoring the child's progress. One therapist (T2) commented that a tablet-based system would be especially useful for home-based therapy sessions, which strongly aligns with the objective of this work.

5. CONCLUSION

Technology based interventions have tremendous potential in improving the delivery of speech therapy. They can also help overcome geographical barriers and address the issues of shortage of therapists and equipment.

In this paper we have presented an automated speech therapy system for childhood apraxia of speech. The system provides mechanisms to not only deliver therapy exercises but also remotely monitor the child's performance and modify the therapy regimen. We reviewed CAS and existing work on therapy tools for CAS, including the NDP3 protocol. We then described our system architecture, including the software modules used for developing mobile therapy app, speech analysis engine, and clinician's interface. Finally, we conducted a user study to validate the system and discussed results from a semi-structured interview of the participants. Overall, we found that the tablet app was widely liked by the participants. Feedback from the study participants indicates that additional features such as animations, rewards, audio and visual aid on the tablet would make the therapy more engaging and speed-up the learning process. Implementation of these features is

[1]In the current implementation, the child must tap the help button to see the target word –see Figure 3(b).

[2]Our current implementation assumes there is network connectivity (WiFi or 3G/4G) during a therapy session in order to upload the speech recordings to the server.

currently underway for the next release of the software. We also have plans to incorporate games and puzzles as part of the speech therapy; we believe that these improvements to further the appeal of (and compliance with) regular practice.

In contrast to existing mobile tools [1, 2, 6, 7], which are standalone apps, our system includes an automated speech analysis engine that provides quantitative speech assessment results to the therapists. This enables the therapist to remotely monitor progress and adapt the therapy regimen as needed. This is particularly advantageous because each child and his/her speech disability is unique and requires individualized care. Results from our pilot study support the feasibility of our system as complement to traditional face-to-face speech therapy.

6. ACKNOWLEDGMENT

This work was made possible by NPRP grant # [4-638-2-236] from the Qatar National Research Fund (a member of Qatar Foundation). The statements made herein are solely the responsibility of the authors.

7. REFERENCES

[1] *Apraxiaville*. Available: http://smartyearsapps.com/service/apraxia-ville/
[2] *ArtikPix*. Available: http://rinnapps.com/artikpix/
[3] ASHA Ad Hoc Committee on Apraxia of Speech in Children. *American Speech-Language-Hearing Association*, 75,2007.
[4] *Moodle*. Available: http://moodle.org
[5] *Nuffield Dyspraxia Programme*. Available: http://www.ndp3.org/dyspraxia-therapy-overview.html
[6] *Pocket SLP*. Available: http://pocketslp.com/
[7] *Speech With Milo*. Available: http://www.speechwithmilo.com/
[8] Anthony, L., Brown, Q., Nias, J., Tate, B. and Mohan, S. Interaction and recognition challenges in interpreting children's touch and gesture input on mobile devices. *ACM conference on interactive tabletops and surfaces*,pp. 225-234, 2012.
[9] Ballard, K. J., Robin, D. A., McCabe, P. and McDonald, J. A Treatment for Dysprosody in Childhood Apraxia of Speech. *Journal of Speech, Language, and Hearing Research*,vol. 53, 1227-1245,2010.
[10] Bunnell, H. T., Yarrington, D. M. and Polikoff, J. B. STAR: Articulation Training for Young Children. *International Conference on Spoken Language Processing*,pp. 85-88, 2000.
[11] Constantinescu, G. A., Theodoros, D. G., Russell, T. G., Ward, E. C., Wilson, S. J. and Wootton, R. Home-based speech treatment for Parkinson's disease delivered remotely: a case report. *Journal of Telemedicine and Telecare*,vol. 16, 2, 100-104,2010.
[12] Forrest, K. Diagnostic Criteria of Developmental Apraxia of Speech Used by Clinical Speech-Language Pathologists. *American Journal of Speech-Language Pathology*,vol. 12, 3, 376-380,2003.
[13] Froehlich, J., Wobbrock, J. and Kane, S. Barrier pointing: using physical edges to assist target acquisition on mobile device touch screens. *ACM SIGACCESS conference on Computers and accessibility*,pp. 19-26, 2007.
[14] Gaines, R., Missiuna, C., Egan, M. and McLean, J. Educational outreach and collaborative care enhances physician's perceived knowledge about Developmental Coordination Disorder. *BMC Health Serv Res*,vol. 8, 1, 1-9,2008.
[15] Georgeadis, A., Brennan, D. M., Barker, L. N. and Baron, C. R. Telerehabilitation and its effect on story retelling by adults with neurogenic communication disorders. *Clinical Aphasiology Conference*,pp. 639-652, 2003.
[16] Jamieson, D. G., Kranjc, G., Yu, K. and Hodgetts, W. E. Speech intelligibility of young school-aged children in the presence of real-life classroom noise. *Journal of the American Academy of Audiology*, 508-517,2004.
[17] Kolles, H. and Feiden, W. Computer-assisted speech recognition in diagnostic pathology. Development of the DragonDictate. *Der Pathologe*, 439-442,1995.
[18] Maier, A., Haderlein, T., Stelzle, F., Nöth, E., Nkenke, E., Rosanowski, F., Schützenberger, A. and Schuster, M. Automatic speech recognition systems for the evaluation of voice and speech disorders in head and neck cancer. *EURASIP Journal on Audio, Speech, and Music Processing*,2010.
[19] Mich, O. Evaluation of software tools with deaf children. *international ACM SIGACCESS conference on Computers and accessibility*,pp. 235-236, 2009.
[20] Moore, J. and Churchward, M. *Moodle 1.9 Extension Development*. Packt Publishing, 2010.
[21] Moran, R. J., Reilly, R. B., de Chazal, P. and Lacy, P. D. Telephony-based voice pathology assessment using automated speech analysis. *IEEE Transactions on Biomedical Engineering*,vol. 53, 3, 468-477,2006.
[22] Newbury, D. and Monaco, A. Genetic Advances in the Study of Speech and Language Disorders. *Neuron*,vol. 68, 309-320,2010.
[23] Oster, A. M., House, D., Protopapas, A. and Hatzis., A. Presentation of a new EU project for speech therapy: Ortho-Logo-Paedia. *Proceedings TMH-QPSR, Fonetik* pp., 2002.
[24] Rick, J., Harris, A., Marshall, P., Fleck, R., Yuill, N. and Rogers, Y. Children designing together on a multi-touch tabletop: an analysis of spatial orientation and user interactions. *Conference on Interaction Design and Children*,pp. 106-114, 2009.
[25] Rvachew, S., and F. Brosseau-Lapre *Speech perception intervention*. Interventions for Speech Sound Disorders in Children, Brookes Pub, 2006.
[26] Shahin, M. A., Ahmed, B. and Ballard, K. J. Automatic classification of unequal lexical stress patterns using machine learning algorithms. *IEEE Spoken Language Technology Workshop (SLT)*,pp. 388-391, 2012.
[27] Shobaki, K., Hosom, J. P. and Cole, R. A. The OGI kids' speech corpus and recognizers. *International Conference on Spoken Language Processing*,pp., 2000.
[28] Shriberg, L. D., Campbell, T. F., Karlsson, H. B., Brown, R. L., Mcsweeny, J. L. and Nadler, C. J. A diagnostic marker for childhood apraxia of speech: the lexical stress ratio. *Clinical Linguistics & Phonetics*,vol. 17, 7, 549-574,2003.
[29] Waite, M., Cahill, L., Theodoros, D., Busuttin, S. and Russell, T. A pilot study of online assessment of childhood speech disorders. *Journal of Telemedicine and Telecare*, 92-94,2006.
[30] Williams, A. *Multiple oppositions intervention*. Interventions for speech sound disorders in children, Brookes Pub, 2006.
[31] Williams, P. and Stephens, H. *Nuffield Centre Dyspraxia Programme*. Interventions for Speech Sound Disorders in Children. Brookes Pub., 2010.
[32] Wren, Y., S. Roulstone, and A.L. Williams *Computer-Based Interventions*. Interventions for Speech Sound Disorders in Children. Brookes Pub, 2006.
[33] Young, S. J., Evermann, G., Gales, M. J. F., Hain, T., Kershaw, D., Moore, G. and Odell, J. e. a. *The HTK Book, version 3.4*. Cambridge University, 2006.