Design and Evaluation of a Vocalization activated Assistive Technology for a child with Dysarthric speech

by

Nayanashri Thalanki Anantha

A thesis submitted in conformity with the requirements for the degree of Masters of Health Science in Clinical Engineering
Institute of Biomaterials and Biomedical Engineering
University of Toronto

© Copyright 2013 by Nayanashri Thalanki Anantha
Design, Training and Evaluation of a Vocalization activated Assistive technology for a participant with dysarthric speech

Nayanashri Thalanki Anantha
Masters of Health Science in Clinical Engineering
Institute of Biomaterials and Biomedical Engineering
University of Toronto
2013

Abstract

Communication disorders affect one in ten Canadians and the incidence is particularly high among those with Cerebral Palsy. A vocalization-activated switch is often explored as an alternative means to communication. However, most commercial speech recognition tools to date have limited capability to accommodate dysarthric speech and thus are often prematurely abandoned. We developed and evaluated a novel vocalization-based access technology as a writing tool for a pediatric participant with cerebral palsy. It consists of a high quality condenser headmic, a custom classifier based on Gaussian Mixture Modeling (GMM) and Mel-frequency Cepstral Coefficients (MFCC) as features. The system was designed to discriminate among five vowel sounds while interfaced to an on-screen keyboard. We used response efficiency theory to assess this technology in terms of goal attainment and satisfaction. The participant’s primary goal to reduce switch activation time was achieved with increased satisfaction and lower physical effort when compared to her previous pathway.
# Table of Contents

Table of Contents ........................................................................................................ iii

1 Introduction .................................................................................................................. 1

1.1 Assistive Technology .............................................................................................. 1

1.2 Access Technologies ............................................................................................... 2

1.3 Problem .................................................................................................................... 2

1.4 Roadmap .................................................................................................................. 3

2 Chapter 2 .................................................................................................................... 4

Background ....................................................................................................................... 4

2.1 Cerebral Palsy (CP) and GMFCS ............................................................................ 4

2.2 Dysarthria .................................................................................................................. 4

2.3 Automatic Speech Recognition (ASR) as an access technology ......................... 6

2.4 ASR optimization for dysarthria ............................................................................. 7

2.5 Case description ...................................................................................................... 7

2.5.1 Health status ...................................................................................................... 8

2.5.2 Physical Capabilities and limitations .................................................................. 8

2.5.3 Access technology history ................................................................................ 8

3 Chapter 3 .................................................................................................................... 10

Research Objectives ...................................................................................................... 10

4 Chapter 4 .................................................................................................................... 11

Methods ......................................................................................................................... 11

4.1 Rationale for selecting input commands ............................................................... 11

4.2 Data acquisition ...................................................................................................... 12

4.3 Speech recognition software ................................................................................ 14
4.3.1 Pre-processing ........................................................................................................ 14
4.3.2 Feature extraction .................................................................................................. 15
4.3.3 Classification ......................................................................................................... 18
4.3.4 Software link to WIVIK interface ........................................................................ 19
4.4 Measurements ........................................................................................................ 20
  4.4.1 Video Recording .................................................................................................... 20
  4.4.2 Response Efficiency ............................................................................................ 21
ISO 9241-9 .................................................................................................................... 22
  4.4.3 Appropriateness ................................................................................................... 25
  4.4.4 Impact .................................................................................................................. 28
4.5 Experimental Protocol ............................................................................................. 29
4.6 Data Analysis ........................................................................................................... 35
5 Chapter 5 ..................................................................................................................... 37
  Results .......................................................................................................................... 37
   5.1 Response efficiency ............................................................................................... 37
      5.1.1 Response effort ............................................................................................... 37
      5.1.2 Rate and Immediacy of reinforcement: ............................................................ 38
   5.2 Appropriateness ...................................................................................................... 39
      5.2.1 Switch efficacy ................................................................................................ 39
      5.2.2 Satisfaction ...................................................................................................... 41
   5.3 Outcomes ............................................................................................................... 42
      5.3.1 Goal attainment .............................................................................................. 42
      5.3.2 Use of technology ......................................................................................... 42
6 Chapter 6 ..................................................................................................................... 44
  Discussions ................................................................................................................... 44
7 Conclusion .................................................................................................................. 46
7.1 Contributions...............................................................................................................................................46
7.2 Concluding remarks ....................................................................................................................................46
7.3 Future work..................................................................................................................................................47
8 References......................................................................................................................................................48

Table of Figures

Figure 1: Components of an access solution within the user’s environment [3]................................. 2
Figure 2: Gross Motor Function Classification system [7]........................................................................... 5
Figure 3: The spectrograms in the first row are for the same vowel and the second row are for different phonemes ....................................................................................................................... 12
Figure 4: Sennheiser HSP4 head worn microphone ..................................................................................... 13
Figure 5: Formant frequency range for fundamental vowels [29].............................................................. 15
Figure 6: Calculation of MFCC .................................................................................................................. 16
Figure 7: Fourier transform of a windowed signal .................................................................................... 16
Figure 8: Mel frequency filters ................................................................................................................. 17
Figure 9: Quadrant layout of the WIVIK onscreen keyboard ................................................................. 20
Figure 10: Properties to evaluate and measures used ........................................................................... 21
Figure 11 Pictorial Children’s Effort Representation Table [15]............................................................... 24
Figure 12 Modified PCERT with Wong-Baker images [28]..................................................................... 24
Figure 13: Experimental protocol ........................................................................................................... 30
Figure 14: PCERT results

Figure 15: Speed-accuracy curve for game timing

Figure 16: Switch efficacy results

Figure 17: Satisfaction (QUEST 2.0) results

Figure 18: GAS scores
1 Introduction

1.1 Assistive Technology

In the United States (US) ‘Technology-Related Assistance of Individuals with Disabilities Act of 1988’ (P.L.100-407), AT is defined as “any item, piece of equipment, or product system, whether acquired commercially off the shelf, modified, or customized, that is used to increase, maintain, or improve functional capabilities of individuals with disabilities”. Assistive technology includes different aids and tools to compensate for functional and/or sensory impairments by providing a means to hear (hearing aids), read (screen readers), move (wheelchairs), speak (augmentative communication systems) and manage self-care (environment control systems). These technologies can greatly influence the quality of life for users by enhancing their participation in daily activities and increasing their level of independence [1]. However, improved functionality alone does not necessarily ensure successful adoption of an AT. Depending on the ease of use and comfort to the user, AT abandonment rates can range between 8% and 75% in the first 3 months of use [2]. Thus, consideration of the user’s contentment with an AT is essential in order to avoid abandonment of the device and loss of resources.
1.2 Access Technologies

An ‘access technology’ is a form of AT that translates a user’s intentions into a functional activity [3] (Figure 1). This technology consists of two components, namely an access pathway and a signal processing unit. The access pathway measures a physiological signal or a physical movement and produces a corresponding electrical signal. This signal is then analyzed by the signal processing unit and a control signal is generated to drive an external device, such as a communication aid or an environmental control unit.

![Diagram](image)

**Figure 1: Components of an access solution within the user's environment [3]**

1.3 Problem

This thesis aimed to address an elusive challenge faced by clinicians at Holland Bloorview, specifically the search for an access pathway for a very bright pediatric client with cerebral palsy who did not have sufficient motor control to effectively use mechanical switches. She however had some vocalizations of highly variable quality. While she could string words together in partial sentences, intelligibility of these phrases was extremely low, even for a familiar communication partner. Furthermore, attempts to articulate words was physically effortful and she tired easily. Clinicians at Holland Bloorview had
trialed, without much success, numerous mechanical as well as speech recognition solutions. The latter initiatives included industry-standard speech recognition systems and a custom word-recognition system. In this study, we focused on the development and evaluation of a novel access technology for this client with message composition on the computer as the target functional activity. The proposed access technology consists of a microphone (access pathway) and a custom vowel recognition algorithm (signal processing unit). The access technology generates commands relevant to an on-screen keyboard for writing.

1.4  Roadmap

The remainder of the thesis is organized as follows. Chapter 2 provides background on extrapyramidal CP and speech recognition, closing with a case description of the client of interest. Following a statement of objectives (Chapter 3), we outline the methods pursued (Chapter 4). The results are then presented (Chapter 5) and discussed (Chapter 6).
2 Chapter 2

Background

2.1 Cerebral Palsy (CP) and GMFCS

CP is a neurological disorder caused by permanent brain injury that occurs before, during or shortly after birth [4]. It often results in constrained physical activity due to restricted body movement, limited muscle coordination and poor posture [4]. These motor disorders are often accompanied by impairments in communication, sensation, perception, and cognition, as well as seizure disorders [4]. Depending on the location of neurological injury, CP is classified as pyramidal (spastic) or extrapyramidal (non-spastic) [5]. Occurring in 70% of individuals with CP, pyramidal CP is characterized by increased muscle tone, resulting in heightened muscle contractions, which vary little with emotion or sleep. On the contrary, extrapyramidal CP causes movement disorders that are dependent on activity and emotion and are therefore less intense during sleep or relaxation. However, abnormal involuntary movements increase during emotional stress or intense activities. Cognitive impairments and seizures are uncommon with this type of CP [5].

The Gross Motor Function Classification System (GMFCS) is a 5 level classification system that describes the gross motor function of children and youth with CP [6] (Figure 2). This age-dependent system rates individuals on the basis of their self-initiated movement with particular emphasis on sitting, walking, and wheeled mobility [6]. Scores on the GMFCS scale indicate the degree to which an individual relies on mobility assistance, ranging from level I (individuals that do not require assistance) to level V (non-ambulatory individuals).

2.2 Dysarthria

Dysarthria is a collection of speech disorders due to neurological injury that results in poor
function of the muscles controlling speech. Consequently, dysarthria is characterized by poor articulation, pitch inconsistencies, and reduced volume of speech [8]. Often caused by diseases affecting nerves and muscles, this condition is prevalent in individuals with extrapyramidal CP [8].

Figure 2: Gross Motor Function Classification system [7]
2.3 Automatic Speech Recognition (ASR) as an access technology

Automatic speech recognition (ASR) is the computer-driven transcription of spoken language into readable text in real-time. Upon capturing spoken words into a microphone or telephone, this technology allows a computer to identify the words that have been spoken and subsequently generate corresponding written text [9]. Based on the algorithm used, it can be designed to perform speaker dependent or speaker-independent recognition.

ASR systems for computer access allow a user to enter text and commands using their voice. This type of system has the potential to greatly enhance the comfort and productivity of computer-based tasks, particularly for individuals with physical disabilities who find the use of mechanical technologies too cumbersome, painful or slow. A survey of ASR users revealed that usability depends not only on speed and accuracy but also on user satisfaction [10]. For users who retain some vocalizations, ASR systems have been found to be superior to other technologies for computer access, not only in terms of speed and accuracy, but also with respect to user satisfaction [10]. This was shown for a population of individuals with disabilities who exhibited non-spastic movements [10].

Speech recognition can be classified into continuous speech recognition and isolated word recognition. Speech uttered continuously makes the recognition process harder as it is difficult to automatically segment words (i.e., define the start and end points of words). The start and end of each word is affected by the preceding and following word. Another factor affecting speech recognition is ‘co-articulation’. The production of each phoneme is affected by the production of surrounding phonemes. The rate of speech also affects recognition; in continuous speech, the words are connected together without defined pauses, making faster speech harder to parse and recognise [13].
On the other hand, isolated word recognition operates on single words at a time, requiring a pause between each word. Hence, in this case the end points of words are easier to find and are not affected by the pronunciation of other words. Since word segmentation is simplified and more accurate, the words should in theory be easier to recognise algorithmically. Though the vocabulary will be limited in this case, isolated word recognition may be suitable for controlling an access technology interface where a full-blown speech-to-text system is not required.

2.4 ASR optimization for dysarthria

For individuals with physical limitations but typical speech capabilities, commercially available ASRs, such as DragonNS or SpeakQ, have demonstrably enhanced the writing experience [10]. However, there are a significant number of people with both physical and speech disabilities for whom such a tool might not work efficiently. Studies investigating the use of commercial ASRs with dysarthric populations have realized recognition rates of approximately 80% for small vocabularies [11]. However, it was found that such a system was not accurate enough to provide reliable access for individuals with poor articulation who require support while writing [11] [12]. This limitation was due to the fact that ASRs are usually optimized for typical speech patterns. Therefore, to meet the needs of individuals with dysarthria, ASR systems must be designed to contend with high levels of variability in pronunciation and volume.

2.5 Case description

The participant (6 years, 10 months old; female) had a history of extrapyramidal type CP, GMFCS level II-III. At the time of the study, she was a student of the Integrated Educational Therapy (IET) program in the Bloorview School Authority, and a client of the Communication
and Writing Aids Service (CWAS) at the Holland Bloorview Kids Rehabilitation Hospital. She was referred to the Pediatric Rehabilitation Intelligent Systems Multidisciplinary (PRISM) Laboratory at the Bloorview Research Institute. Her primary caregiver is her mother.

2.5.1 Health status

The participant’s hearing and vision are intact. According to a documented psychological assessment, her verbal and non-verbal skills are extremely good and she is cognitively bright. A neurological examination demonstrated evidence of hyperkinetic movements which she is able to consciously control. In addition, she has dystonic posturing, particularly in the lower extremities.

2.5.2 Physical Capabilities and limitations

The participant can walk independently using a walker and hinged ankle-foot orthotics (AFO), but cannot climb stairs or cruise. She has limited ability in all motor areas. Her highest level of motor function is in her legs, followed by her oral muscles, arms, and then hands. When sitting, the participant tends to extend and push back in her chair if unsupported. However, putting weights or straps on her legs for support provides improved stability. Opening her hand or moving her arm while maintaining a relaxed state is challenging. However, her motor control in her upper limbs improves in low-stress situations. The participant’s speech volume is low and the pitch is variable. She exhibits difficulty in breathing in stressful situations leading to heavy breathing and has difficulty verbalizing full sentences and words. Because her speech can be unpredictable, she often requires voice amplifiers and clarification strategies.

2.5.3 Access technology history
For written communication, the participant uses five buddy button switches, which are positioned horizontally on a table with a computer in front of her. She activates these switches with her fists. Each button corresponds to a function on a WIVIK onscreen, virtual, quadrant keyboard. When using this access technology, her communication rate is very slow and her constant need for assistance often leads to frustration. Overall, her use of this particular communication pathway has diminished over time. Previously, the participant tried and abandoned various other technologies including a head pointer, a head mouse, two head switches, an ASR (both SpeakQ and DragonNS) and a VoiceGP module which translated five trigger words into switch closures for operating the WIVIK Quad keyboard.
3 Chapter 3

Research Objectives

In light of the case history presented in the previous chapter, the objectives of this thesis were to:

1. Design a speech/vocalization recognition system for the participant such that a minimum of five code words/vocalizations can be discriminated for use of an onscreen WIVIK quad keyboard with an accuracy greater than or equal to 90%.

2. Train the participant with this new AT and systematically evaluate the designed system against her existing AT using Response Efficiency Theory.
4 Chapter 4

Methods

4.1 Rationale for selecting input commands

After an evaluation of the participant’s continuous speech by a certified Speech-Language Pathologist (SLP), five isolated words of interest to the participant were selected as verbal commands to operate the on-screen keyboard. A set of training and evaluation sessions were performed with a recognition system developed for our participant by CWAS (Communication and Writing Aids, Bloorview). Only low levels of accuracy (below 40%) could be achieved and it was difficult to train the participant to produce consistent sounding words. We therefore decided to use isolated phonemes (isolated vowel sounds) for the recognition process. Each vowel possesses a distinct harmonic spectrum. The features which spectrally distinguish vowels are called formants. Formants are vocal tract resonances and are positioned at different frequencies for different vowels. We collected some pilot data from our participant to confirm this formant separation and to confirm vowel sounds as a reliable input for our participant. As exemplified in Figure 3, our pilot data indicated a clear formant separation among different vowels and a strong level of consistency for the same vowel. Thus five isolated fundamental vowel phonemes (a, e, i, o, u) were chosen to replace the five buddy button switches mentioned
earlier.

Figure 3: The spectrograms in the first row are for the same vowel and the second row are for different phonemes

4.2 Data acquisition

We selected a high quality Sennheiser HSP4 microphone for data acquisition (Figure 4). The microphone was connected to our participant’s PC running a Windows 7 (32-bit) operating system via a standard PCI express sound card.

The HSP4 is a high quality pre-polarised condenser head mike for ‘hands-free’ professional applications. Its adjustable neckband is unobtrusive and comfortable to wear. It also has a twist-proof microphone boom. It can be attached to the left or right ear as convenient. The chosen
microphone ensured that the participant would be comfortable over an extended period of wear (up to an hour in a single sitting).

**Figure 4: Sennheiser HSP4 head worn microphone**

In terms of technical specifications, the Mike has a cardioid pick-up pattern with a frequency response of 40 – 20,000 Hz ± 3 dB. It has a sensitivity of 4 mV/Pa. It features excellent capture of speech and very good suppression of ambient noise. It has superior feedback rejection and high maximum sound pressure level. As our participant would be using this system in a classroom environment, these features were important considerations in microphone selection.

The signal from the microphone was imported to the speech recognition software which was installed as an application on the participant’s computer. The speech processing and recognition algorithm will be explained in the next section.

Initial training data were acquired over a period of two weeks (10 days). We collected three sessions a day and three trials per session. Each trial consisted of vocalizing the 5 phonemes once. Two and four months following initial training data collection, we collected an additional nine sessions over a period of five days where each session consisted of a single trial. Eliminating inconsistent data from all the above sessions, we used 75 training data sets for each
vowel sound. This training data were collected over extended periods of time to account for any variations in consistency. Nonetheless, we did train our participant to utter these five sounds consistently with the help of an SLP during the first two weeks and at the beginning of the two and four month follow up data collection sessions.

This training data was collected in both the control environment – PRISM lab while training with the SLP as well as the natural environment. The natural environment for our participant meant the participant’s home and her classroom. These data collections were equally distributed between the two natural environments to account for the noise interference while designing the system. However the first two weeks were conducted in the control environment to allow the participant to get trained with uttering the vowels consistently.

4.3 Speech recognition software

The speech recognition algorithm consisted of the following stages.

4.3.1 Pre-processing

Since the microphone has very good noise rejection, filtering was relatively straightforward. The signals were low-pass filtered with a cut-off of 10 kHz. Frequencies in excess of 10 kHz are typically used for speaker identification but were not of interest here. The lower frequencies were required for phoneme differentiation. The typical range of formant frequencies for vowels for an adult speaker is well below 10 KHz (Figure 5). However, given the dysarthria of our participant and that the formant frequency range varies for child speakers, we admitted frequencies up to 10 kHz.
### Feature extraction

Following pre-processing, data reduction was performed to extract features and form speech templates from the training data. These templates were then used to classify the incoming speech signal during speech recognition.

We selected the Mel-Frequency Cepstral Co-efficients (MFCC). This feature vector is based on the ‘Mel-scale’ frequency which comes from the observation that humans can differentiate low-frequency sounds better than high-frequency sounds. The advantage of MFCC is its ability to reduce the amount of information in a Fourier transform of a frame of speech. MFCC is a more efficient means of data reduction (as compared to Linear Predictive Coding (LPC) employing formants). It reduces the amount of information in a Fourier transform of a frame of speech to a small set of parameters representing the nonlinear perception of sound in the human ear. The cepstral domain is obtained from the discrete cosine transform of the logarithm of the spectral power at each of the Mel frequencies. This representation unveils time delays, harmonics and positions of fundamental frequencies. MFCCs are thus a result of log-log warped frequency spectrum and can be plotted over time similar to a spectrogram. These co-efficients are sufficient to differentiate among the fundamental vowel sounds. Although using Linear Predictive Coding

<table>
<thead>
<tr>
<th>Vowel</th>
<th>Main formant region</th>
</tr>
</thead>
<tbody>
<tr>
<td>u</td>
<td>200–400 Hz</td>
</tr>
<tr>
<td>o</td>
<td>400–500 Hz</td>
</tr>
<tr>
<td>a</td>
<td>800–1200 Hz</td>
</tr>
<tr>
<td>e</td>
<td>400–600 and 2200–2600 Hz</td>
</tr>
<tr>
<td>i</td>
<td>200–400 and 3000–3500 Hz</td>
</tr>
</tbody>
</table>

**Figure 5: Formant frequency range for fundamental vowels [29]**
(LPC) classification with formant features would be less computationally intensive and have higher classification accuracy, formant tracking for five classes of vowels for dysarthric speech has low accuracy. We did extract these formant features but obtained accuracies of less than 60%. For this reason, we chose MFCC features which had better tracking accuracy of over 90%.

The steps for calculating MFCC are summarized by the block diagram in Figure 6.

![Figure 6: Calculation of MFCC](image)

The following steps were used to derive MFCCs

1. We took the Fourier transform of a windowed excerpt of a signal

![Figure 7: Fourier transform of a windowed signal](image)
2. We mapped the powers of the spectrum obtained above onto the Mel scale, \( m \), using triangular overlapping windows. Triangular filter windows are used because they are more closely spaced for the lower frequencies which are what we aim to look at in order to extract the Mel scale features.

\[
m = 2595 \log_{10} \left( 1 + \frac{f}{700} \right)
\]

3. We took the logs of the powers at each of the Mel frequencies, \( m \).

4. We applied the Discrete Cosine Transform (DCT) to the Mel log powers.

5. The amplitudes of the resulting spectrum were the MFCC coefficients.

Figure 8: Mel frequency filters

For our study, we used 45 MFCC co-efficients with 25 filter bank windows and an overlap size of 0.5.
4.3.3 Classification

For classification of the incoming signal into one of the five vowel classes, we employed Gaussian mixture modeling. A Gaussian Mixture Model (GMM) is a parametric probability density function represented as a weighted sum of Gaussian component densities. The density for a given vowel class is given as

\[ p(x) = \sum_{i=1}^{M} w_i g(x \mid \mu_i, \Sigma_i) \]

where \( x \) is a \( D \)-dimensional continuous-valued data vector of MFCC coefficients, \( w_i, i = 1 \ldots M \), are the mixture weights, and \( g(x \mid \mu_i, \Sigma_i), i = 1 \ldots M \), is the \( i^{th} \) component Gaussian density with \( D \times 1 \) mean vector, \( \mu_i \) and \( D \times D \) covariance matrix, \( \Sigma_i \). The GMM parameters, namely, mixture weights and the mean and covariance of each component density, are estimated from the training data using the iterative Expectation-Maximization (EM) algorithm.

Each component density is a multivariate Gaussian function of the form,

\[ g(x \mid \mu_i, \Sigma_i) = \frac{1}{(2\pi)^{D/2} |\Sigma_i|^{1/2}} e^{-\frac{1}{2} (x - \mu_i)' \Sigma_i^{-1} (x - \mu_i)} \]

where \(|\cdot|\) denotes the determinant and the prime denotes the transpose. The complete Gaussian mixture model is parameterized by the mean vectors, covariance matrices and mixture weights from all component densities.

We used the above method of training the system with MFCC features. One GMM was trained per vowel sound. Once the GMMs were trained, we compared the Mahalanobis distance in MFCC feature space from the incoming signal to each trained class. The incoming signal was assigned the label of the closest trained class.

To account for any unintended vocalizations, breath sounds and other disturbances, we needed a ‘no output’ result where the system does not activate. For this purpose, we designed three tests. In the first test, we checked for the presence of sound using a maximum energy calculation. If
this test was positive, then the signal proceeded to the second test where we checked if the incoming signal was periodic. If yes, then MFCC extraction and classification were invoked. The third test checked the shortest Mahalanobis distance against a threshold value. This threshold value was empirically calculated from the training data set as the distance beyond which the occurrence of a vowel sound was less than 5% probable. If the distance exceeded this threshold then the classifier defaulted to the “no output” class. This prevented any activations when the participant vocalized unintentionally. Likewise, for the first and second test, violation of either test condition resulted in “no output”.

4.3.4 Software link to WIVIK interface

The output from the recognition software triggered one of five relays, emulating the selective activation of 5 mechanical switches. The output hardware was a 5 volt low-cost USB-based 8-Channel Data Acquisition Module (DLP-IO8-G; DLP Design. This module was custom-wired to a series of relays that communicated with a Prentke Romich Company USB switch interface box. This switch interface was used with the WIVIK on-screen keyboard. WiVik is a virtual on-screen keyboard that allows people with physical disabilities to access any application within Microsoft Windows (like a word processor) via switch access. WiViK also includes user-customizable word prediction that facilitates typing. We took advantage of this feature, and customized the WiViK user profile to include words that our participant commonly uses.

We used the hierarchical quadrant layout of the keyboard as shown in Figure 9. Once a quadrant was selected, the keyboard reconfigured, dividing the content of the selected quadrant among four new quadrants. At any level of the selection hierarchy, each of the first four vocalization switches corresponded to one quadrant while the fifth switch reverted back to the original quadrant layout. Recall that each “switch” was activated by the production of a vowel sound. During our trials, the selected letters or words (from word prediction) were displayed in a word document.
4.4 Measurements

4.4.1 Video Recording

A SONY HandyCamDCR-SR88 was used to capture a visual record of some participant data acquisition sessions. The perspective of the camera was optimized to capture the access pathway being used and/or the potential access site for the new pathway. The videos served several purposes: 1) to provide visual confirmation of the activations recorded by the software programs, 2) to provide visual confirmation of the unsuccessful activation attempts manually noted by the researchers, 3) to facilitate analysis of the time required for switch activation and explain delays in feedback/software response, and 4) to record the time required for a user to have a request acknowledged and fulfilled.

Figure 10 groups the various measurement tools into three domains: response efficiency, appropriateness and impact.
4.4.2 Response Efficiency

Response efficiency theory states that when individuals have the opportunity to choose between two or more functionally equivalent alternatives, they will select the option that they perceive as most efficient [14].

The factors which affect efficiency are listed below [14]:

- Response Effort
- Rate and Immediacy of Reinforcement
- Quality of Reinforcement
- Switch Efficacy
- Satisfaction
- Goal Attainment
- Usage
- PCERT
- Direct Observation
- Interview
- Game
- QUEST 2.0
- GAS
- Automated usage log
1) Rate of reinforcement – The frequency at which the individual receives the expected result.
2) Quality of reinforcement – The match between the individual’s expectations and the actual result.
3) Response effort – The amount of physical and cognitive effort required of the individual.
4) Immediacy of reinforcement - The delay between the technology response and the user’s indication of his/her intent.

These factors contribute to the overall perception of efficiency but this can vary depending on the individual and their environment [14].

These measures are designed to assess the overall value, in terms of costs and gains, of a given access pathway. They relate to the efficiency concept defined by ISO 9241-9.

---

**ISO 9241-9**

The ISO 9241-9 is the International Standard describing the ergonomic requirements for non-keyboard input devices for office work with visual display terminals (ISO, 2000). To maintain a standardized approach of subjectively measuring the user-technology match, the following key definitions are adopted.

**Appropriateness:** An appropriate input device is effective, efficient and satisfactory for the tasks being performed and the intended work environment.

**Effectiveness:** Accuracy and completeness with which users achieve specified goals.
**EFFICIENCY:** Resources expended in relation to the accuracy and completeness with which users achieve goals.

**SATISFACTION:** Freedom from discomfort and positive attitudes of the users towards the use of the product.

*Response Effort*

*Modified Pictorial Children’s Effort Rating Table (PCERT)*

The PCERT is a pictorial effort rating scale based on the original Children’s Effort Rating Table (CERT) developed in 1994 by Williams and colleagues [15]. The scale combines verbal descriptors specific to the child population with pictorial representations (Figure 11). It has been tested for reliability and validity with children aged 10 to 15 years [16]. For the purposes of this study, our participant was asked to use a modified PCERT to gauge the amount of effort required to use her access pathway. There is no record in the literature of the PCERT being used with children with multiple and severe disabilities. However there is a lack of evidence in general for effort rating scales for this population and therefore, a modified version of the PCERT was adopted based on its success with other pediatric populations. In the modified version, the descriptors from the original PCERT have been combined with images from the Wong-Baker pain scale [16]. This version was validated in an earlier study with a similar protocol to deliver access technologies for children in need of a communication access [28]. These images were chosen as being more meaningful to children in wheelchairs than the original images that depicted a child climbing a set of stairs. Additionally, as many children in this population may require partner assisted auditory scanning to provide their ratings, the scale was reduced from 10 items to 6 items to facilitate self-reporting (Figure 11).
Additionally, the caregiver of the participant was asked to assess the amount of effort associated with using the access pathway across the following categories:

- perceived physical effort required for the participant to activate the switch
- perceived cognitive effort required for the participant to activate the switch
- perceived physical effort required for the caregiver to set up the switch/interact with the switch
- perceived cognitive effort required for the caregiver to set up/interact with the switch
The same Modified PCERT (Figure 12) was used for caregiver assessments in order to maintain consistency.

**Rate and Immediacy of Reinforcement:**

Reinforcement for switch activation was provided in the form of auditory and/or visual feedback. The following measurements were made while the participant performed a computer activity (writing) that requires the use of her switch.

- time required to activate the switch
- time between activation and software response/feedback

**Quality of Reinforcement:**

The following questions relate to the quality of reinforcement and were answered through direct observation and interview with the caregiver and/or participant.

- Is the switch activation feedback clear and discernible by the participant?
- How clearly can the desires of the user be interpreted by the caregiver?
- Can the access pathway be used in a variety of different contexts, and interfaced with a variety of programs? If not, how is it limited?

**4.4.3 Appropriateness**

Since achieving a good contextual fit is essential to successful technology adoption, it is important to determine whether or not a response efficient switch is also appropriate for the user. ISO 9241-9 defines an appropriate device as one that is effective, efficient and satisfactory for
the tasks being performed in the intended work environment (ISO, 2000). Thus, in addition to response efficiency, efficacy and satisfaction were also measured for each access pathway.

**Switch Efficacy**

This measure is designed to assess the performance of a switch and consists of the activities below. It also provides the measure of effectiveness as defined by ISO 9241-9. Switch efficacy was measured using a game that is of interest to the participant. In addition to gathering information about participant likes and dislikes from caregivers, the participant was allowed to try out several different games in the initial data collection sessions until one was found that she appeared to enjoy. Criteria for the game were as follows: must require low cognitive effort, must have clear correct and incorrect responses, must have clear feedback and must have an adjustable pace.

The participant was given up to 3 training sessions with each game/test in order to familiarize herself with the task. A 1-minute break between training sessions and a 3-minute break after the last training session was given to ensure full recovery before the trial.

The sensitivity and specificity of the switch were calculated based on data manually recorded during the game. Three trials were completed with the participant using an activation window (amount of time the stimulus will remain present for her to activate her switch) with which she was comfortable (3-5 seconds). There was a one-minute break between trials. Subsequent trials involve changing the activation window time so that the speed-accuracy curve could be mapped. The trials began with a short activation window, which was gradually increased until the participant’s accuracy reached a plateau.

Data for sensitivity and specificity measures were collected over a 2-week period during 3 sessions. This was collected for the existing technology during assessment and for the new technology at the 4 and 8-week stages. The data from all 3 sessions were pooled and calculations were based on total counts. During each session, the participant was asked to play either ‘Splat the clown’ or ‘Load the truck’. These games are available on helpkidzlearn.com. Both these
games met our criteria of requiring low cognitive effort, captivating the interest of the participant, and offering clear feedback and an adjustable pace. In the first game (Splat the clown), an object would appear on the screen for a fixed amount of time (set to 10 seconds for her existing technology and 5 seconds for the new technology- determined after trials with different timing and fixing the one where the participant’s accuracy reached a plateau). Initially the timing was set to 1 second, 3 seconds and then to 5 seconds, the duration at which the participant was most consistent. This observation is shown in the Results (5) section. She was required to activate her switch to release the object to hit the clown. In the second game (Load the truck), for each game she was required to hit two switches. One was to move the object that appears to the truck and the other switch to load the object into the truck.

For her existing technology, all 5 buttons were placed in their original positions and one switch was selected for each game. Sequentially all buttons were used to play the same game to determine switch efficacy. The same arrangement was applied to the new technology where phonemes where chosen sequentially to play each game.

Satisfaction

The Quebec User Evaluation of Satisfaction with assistive Technology (QUEST 2.0) is an outcome measure that focuses on consumer satisfaction with assistive technology [17]. The QUEST ascertains a person’s positive or negative valuation of the dimensions of an assistive device as influenced by their expectations, perceptions, attitudes and personal standards [18]. The psychometric properties of the QUEST 2.0 have been verified for individuals with disabilities and it can be used for both adults and adolescents [17] [19]. The QUEST is a questionnaire that can be self-administered or interview-based and requires respondents to rate their satisfaction with each of 12 variables on a five point scale, with respect to their AT, and subsequently, subjectively identify the 3 most important variables. The measured variables relate to the environment, the user and the AT. The QUEST 2.0 was administered to determine user and/or caregiver satisfaction with the technology. This was administered with the existing technology at Phase 1-2 during assessment. This was prior to the development of the new
technology, so she did not have any experience with the new technology. This was done to ensure that her ratings were not biased based on the new technology. This measure represents the satisfaction parameter defined by ISO 9241-9.

### 4.4.4 Impact

**Goal Attainment**

Goal Attainment Scaling (GAS) is an individualized, participant-centered (or clinician-centered) outcome measure designed to capture and measure goals of intervention from a participant or clinician perspective [20]. It is internationally recognized as a tool that helps children and families set realistic goals and focus their attention on a target [20] [21]. The use of GAS, its psychometric qualities and clinical utility in pediatrics are well documented [20] [21] [22] [23] [24]. Because the GAS is flexible in nature, it allows individuals with separate and unique goals to be compared in terms of their success in attaining their respective goals (Cusick, McIntyre, Novak, Lannin, & Lowe, 2006). The GAS is also more sensitive to change than norm-referenced measures [22].

To measure goal attainment, GAS uses a five-point scale ranging from -2 to +2 [21]. The levels are usually represented as follows: -2 = ability at the time the goal was set, -1 = slight improvement, 0 = expected level of improvement, +1 = improvement that slightly exceeds expectations and +2 = improvement that greatly exceeds expectations [21] [23]. In our study, we also added a sixth level of -3 and +3 to indicate deterioration and very high levels of improvement, respectively. This has been done in previous studies in order to address the issue of possible floor and ceiling effects in the GAS method, and has been shown to be more sensitive to changes than the traditional scale [20] [24]. Several studies have recommended that individuals administering the GAS have specific goal-setting training in order to minimize the likelihood of proposing goals that are too easily achieved or having a scale in which the increments do not represent equal levels of difficulty. The GAS was administered by the
researcher trained in the method and goals were set collaboratively with the family. Our intention in invoking GAS was to assess whether or not the access pathway contributed to overall goal attainment.

**Use of the technology**

- For the existing access pathway (five mechanical switches), an estimate of daily and/or weekly usage of the pathway, as determined by the caregiver, was recorded.
- After the training period, an estimate of daily/weekly usage as determined by the primary caregiver, was recorded

### 4.5 Experimental Protocol

The study was approved by the Research Ethics Board (REB) at Holland Bloorview Kids Rehabilitation Hospital. The study was organized into six phases (Figure 13).

**Phase 1:**

This phase consisted of the initial visit with the participant to determine her existing access pathway, possibilities for a new access pathway (determined through consultation with parents, therapists and teachers and observation of the participant), reasons for seeking a new pathway, and expectations for switch use. Additional information regarding the participant’s cognitive, physical and health attributes was collected using the modified Participant Access Questionnaire [25]. Video of the participant’s switch use was acquired during this time for use in developing the new access pathway. The initial visit was under 2 hours.
Phase 2:

The new access pathway for the participant was developed. This phase involved 3 follow-up sessions with the participant in order to test different versions of the new access pathway and record more data where necessary. These sessions varied in location (school, home or at Holland Bloorview), depending on the participant and family’s preference. These iterations in the design process helped to ensure that the design adhered to the desired criteria for response efficiency and satisfaction.

Phase 3:

Evaluation of the efficiency of the existing access pathway. Switch efficacy (where applicable), response efficiency, satisfaction and use of the technology (where applicable), and

Figure 13: Experimental protocol
outcomes were measured as per the methods outlined in section 4.5. For switch efficacy and response effort, three data collection sessions were held over a 2-week period at the convenience of the user. This required up to 1.5 hours of participant time, and between 0.5 and 0.75 hours of parent time.

Phase 4:

Phase 4 entailed the delivery of the new access pathway and training related to setup and function. The new access pathway was delivered 4 months after the initial visit. GAS was used to establish goals and appropriate achievement scales in relation to the new access pathway. An individualized switch training schedule was developed as per the participant’s availability.

Phase 5:

Individualized switch training was performed for 8 weeks following switch delivery. Training was performed two to three times per week in one-on-one sessions, for 0.5 hrs. Training was focused on skill acquisition with a Most-to-Least prompt fading hierarchy [26].
<table>
<thead>
<tr>
<th><strong>Table 1 : Hierarchy of prompts [26]</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Independent</strong></td>
</tr>
<tr>
<td><strong>Visual</strong></td>
</tr>
<tr>
<td><strong>Indirect (Verbal or Nonverbal)</strong></td>
</tr>
<tr>
<td><strong>Direct Verbal</strong></td>
</tr>
<tr>
<td><strong>Gesture</strong></td>
</tr>
<tr>
<td><strong>Positional</strong></td>
</tr>
<tr>
<td><strong>Modeling</strong></td>
</tr>
<tr>
<td><strong>Partial Physical Assistance</strong></td>
</tr>
<tr>
<td><strong>Full Physical Assistance</strong></td>
</tr>
</tbody>
</table>

The overall Switch Training Paradigm [26] is described below. In order to motivate the participant and encourage meaningful participation in switch training, individualized games/activities were used. These were of interest to the student, age appropriate, and relevant to the participant’s curriculum. The participant progressed from one training stage to the next when...
she reached a level of “independent” use (as per the prompting system above) at least 80% of the time.

**Table 2: Overall switch training paradigm [26] [27]**

<table>
<thead>
<tr>
<th>STAGE</th>
<th>TRAINING TECHNIQUE</th>
<th>ACTIVITY DESCRIPTION</th>
</tr>
</thead>
<tbody>
<tr>
<td>Introducing AT</td>
<td>N/A - Exploratory &amp; experiential learning</td>
<td>Participant tolerates the positioning of AT equipment in relation to her body</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Participant responds to AT generated experiences</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Participant attends to &amp; shows interest/pleasure in on-screen sounds, images or movement</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Participant independently explores the switch and its method of activation</td>
</tr>
<tr>
<td>Motor Movement (cause&amp; effect)</td>
<td>Graduated Guidance with Immediate Reinforcement - Most-to-Least prompt fading hierarchy</td>
<td>Press &amp; Hold: presses and holds the switch to achieve a desired effect</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Press &amp; Let Go: presses and releases a switch to achieve a desired effect</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Press It Again: activates a switch a number of times to keep an activity playing</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Turn On &amp; Off: activates a switch to start and stop an activity</td>
</tr>
<tr>
<td>Skill Acquisition</td>
<td>Most to Least Prompts - Errorless learning</td>
<td><strong>One Switch Training:</strong></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Timing: presses a switch in response to an on-screen cue</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Positional: tracks an object as it moves across the screen, pressing a switch when the object is in a target area</td>
</tr>
<tr>
<td>STAGE</td>
<td>TRAINING TECHNIQUE</td>
<td>ACTIVITY DESCRIPTION</td>
</tr>
<tr>
<td>-------</td>
<td>-------------------</td>
<td>---------------------</td>
</tr>
<tr>
<td></td>
<td>- Decrease probability of developing prompt dependency</td>
<td><strong>Two Switch Training:</strong></td>
</tr>
<tr>
<td></td>
<td>- Most rapid skill acquisition method</td>
<td>This or That: differentiates the actions of two different switches</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Start &amp; Stop: uses one switch to start and another to stop an activity</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Move &amp; Choose: uses two switches to complete simple “move &amp; choose” activities</td>
</tr>
<tr>
<td></td>
<td></td>
<td><strong>Formal Scanning Training:</strong></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Always Right: chooses one item from three on-screen options</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Specific Target (empty cell): chooses a specific target from three on-screen options that include two empty cells</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Completing Sequences: completes simple sequences by choosing the correct target from three options</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Specific Target (three options): selects a specific target from three on-screen options in response to a question or request</td>
</tr>
<tr>
<td></td>
<td></td>
<td><strong>Training on Meaningful &amp; Functional Use of Switch</strong></td>
</tr>
</tbody>
</table>

Phase 6:
Phase 6 consisted of two follow-up evaluations. The first occurred 4 weeks after the new access pathway was delivered and the second, another 4 weeks later (i.e., 8 weeks after the new access
pathway had been delivered). Switch efficacy, response efficiency, satisfaction with the technology, use of the technology and goal attainment were measured as per the methods outlined in section 4.5. For switch efficacy and response effort, three data collection sessions were held over a 2-week period at the convenience of the family. This required up to 1.5 hours of student time, and between 0.5 and 0.75 hours of parent time.

4.6 Data Analysis

Switch Efficacy

Specificity and sensitivity of the input device was calculated based on the cumulative data collected at each experimental phase (with the old access pathway, 4 weeks post-introduction of the new pathway, and 8 weeks post-introduction of the new pathway). The speed-accuracy curves were plotted based on the cumulative data for each experimental phase.

\[
\text{Specificity} = \frac{\text{True Negatives}}{\text{True Negatives} + \text{False Positives}}
\]

\[
\text{Sensitivity} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Negatives}}
\]

Response Efficiency

Response efficiency was reflected in the cumulative data for each phase. Qualitative comparisons were made between phases.

Goal Attainment
GAS scores were analyzed using a comparison between scores at the 4- and 8-week follow-ups. Goal attainment was characterized by an improvement of at least 2 scale levels on the individualized GAS scales, indicating that the expected outcome was achieved [20] [24].

Satisfaction, Usage and Contextual Impact

Results from these measures were assessed qualitatively or quantitatively as appropriate, and comparisons were made between the information collected in relation to the old access pathway and that collected in relation to the new access pathway and training programs.
5 Chapter 5

Results

5.1 Response efficiency

5.1.1 Response effort

Results from the PCERT scale for both existing and new technologies are presented in Figure 14. In terms of physical effort, the results clearly indicate that the new technology requires far less effort as compared to her existing technology (‘hard’ to ‘easy’). Her main goal to reduce physical effort (reported under GAS in later sections) was clearly achieved. By the end of training, she rated the new technology as requiring even less effort (‘very easy’). In terms of cognitive effort, the new technology required more effort than the existing technology during the earlier weeks of training (‘starting to get hard’ to ‘easy’). Though by the end of 8 weeks of training, she rated the new technology as necessitating the same amount of cognitive effort as that of the existing technology (‘easy’).

Her caregiver rated more physical effort for the existing technology compared to the new technology (‘starting to get hard’ to ‘easy’). Cognitive effort required by the caregiver was similar for both technologies (‘easy’).
5.1.2 Rate and Immediacy of reinforcement:

*Time required to activate a switch*

With her existing technology, depending on the time of the day and level of exhaustion, the participant required between 5-20 seconds to physically hit a switch. Also the switches to the extreme left and extreme right took an additional 5 seconds to reach.

With her new technology, it took her between 1-5 seconds to vocalize the desired phoneme. This was after the 8-weeks of individualized switch training. At the 4-week mark, it took her between 4-8 seconds to vocalize and activate the switch. Clearly, with training her activation time decreased. The time required is usually for cognitive processing rather than physical vocalization. Once she decided her option, vocalization was immediate.

![PCERT results](image.png)

Figure 14: PCERT results
**Time between activation and feedback/software response**

For both the existing and the new technologies, there was no lapse time between activation and feedback/software response. The response was immediate following activation. With the existing technology, the feedback was the same as the software response wherein a quadrant of the WIVIK keyboard was selected immediately after hitting the corresponding switch. With the new technology, the feedback entailed a visual display of the phoneme vocalized. This feedback was presented simultaneous to the software response which was the WIVIK quadrant selection. These system behaviors were consistent throughout and did not change with participant training.

**Quality of Reinforcement:**

Observations were made and an interview was conducted with the participant and caregiver to answer the questions relating to quality of reinforcement as outlined in section 4.4.2.

On direct observation with the participant, it was clear that with both her existing, and new technology, activation feedback and software response were clearly understood and discernible by the participant. The participant is cognitively very bright and her vowel vocalizations are clearly understood by her caregiver. So her desires are always clearly interpreted by her caregiver with respect to use of her technology. Currently different contexts for switch use are not being explored for this participant.

### 5.2 Appropriateness

#### 5.2.1 Switch efficacy

Results for measures of efficacy and satisfaction are presented in Figure 15 and Figure 16 respectively.
In each of these games, the participant required 5 hits at the correct time to win the game. During this game, false positives, false negatives, true positives and true negatives were recorded. For these recordings, errors due to classification errors from the new technology were also taken into account. The participant understood the difference between errors due to her selection and those due to the software. She was not bothered by this and understood that the new technology would not be 100% accurate (overall software accuracy was 94%).

The speed-accuracy curve for the selection of the object timing in the games is shown in Figure 15. Ten trials were performed for each of the speed levels or until a plateau was reached with a comfortable timing.

For the existing technology the sensitivity and specificity were 0.85 and 0.95, respectively. For the new technology, at the 4-week period these metrics were 0.8 and 0.85 and at the 8-week period they improved slightly to 0.85 and 0.85, for sensitivity and specificity, respectively. There was a decrease in specificity from the existing technology to the new technology accounting for the classification errors in the new technology which did not exist in the existing technology.
5.2.2 Satisfaction

*QUEST 2.0:*

The results for QUEST 2.0 are presented in Figure 16. It was filled out by the caregiver with discussions with the participant where applicable. The three items that were considered most important were ‘Dimensions’, ‘Easy to use’ and ‘Effectiveness’. The QUEST score for the existing technology was 3.875 out of 5. For the new technology, it was 4.0 at the 4-week stage and 4.5 at the 8-week stage.
5.3 Outcomes

5.3.1 Goal attainment

In the beginning of the study, two goals were set with the caregiver and the participant using the Goal Attainment Scale (GAS). The first goal was to effectively reduce the amount of time required for switch activation which was stated as very important for the participant for school activities (i.e. writing). The second goal was to develop her competency with a new technology that demanded less effort for switch activation. At the end of the 4-week stage, for goal 1 she was at the ‘less than expected’ (-1) outcome. At the end of 8-week training she was at the ‘more than expected’ (+1) outcome.

For goal 2, at the 4-week stage she was at the ‘Expected’ (0) level. At the end of the 8-week training, she was at the ‘more than expected’ (+1) level. An improvement of 2 points or greater from the starting score is considered clinically significant with GAS. The results for GAS scores are shown in Figure 17.

5.3.2 Use of technology

For the existing technology, the participant used her switch 1-2 times per week to perform writing activities (homework). She typically used it for a maximum of 30 minutes at a time. This technology had been used for a year at the beginning of the study. The caregiver reported that its use had gradually diminished over time due to the enormous physical effort required.
Her new technology was used 2-3 times a week during the training period of 8 weeks to play games and perform writing. As this technology required less effort and the participant was very interested in the new access technology, she was able to use for about an hour at a time. This could also be due to decreased time for switch activation. She was regularly using it after training 3-4 times a week for 30-40 minutes each time. Her usage has decreased over the summer though.

Figure 18: GAS scores
Chapter 6
Discussions

The response efficiency results indicate that there was a decrease in physical effort with the new technology as compared to the existing technology. The amount of cognitive effort required was however higher. This was because the new technology required her to associate the corresponding phoneme with the quadrant on the WIVIK keyboard. In the initial stages this association confused her at times. For e.g., the character she intends to choose is ‘e’ which is in the first quadrant; but the phoneme corresponding to the first quadrant is ‘a’. This took her between 3-5 seconds to decide. As a result the time for switch activation was higher. However, by the end of training, this concept had become easier for her to grasp. She then rated the cognitive effort as lower than before. Combining PCERT results with quality and rate of reinforcement data, the new voice-controlled technology appears to be much more enabling than her existing mechanical switch arrangement. Also the existing technology had already been modified a few times prior to the development of the new technology in terms of arrangements of the mechanical switches and enabling best available access to the switches. This was done by her Occupational Therapist (OT) at school with whom our participant would use the technology regularly.

When it comes to switch efficacy, her existing technology initially had better results than her new technology. By the end of 8 weeks of training however, the new technology matched but did not exceed the old in terms of efficacy. Note that from the technology standpoint, the existing access solution is errorless when activated as it consists of mechanical buttons directly connected to the computer. The buttons are assigned to specific computer keys. There is no classification involved and thus no errors occur. In contrast, the new technology may have errors attributable to algorithmic classification of vocalizations and to participant error (e.g., making the incorrect vocalization). Participant error however significantly reduced with training. It is important to note that although participant could hit the right button almost every time with her existing
technology, it required much physical effort and she would become frustrated. As a result she would end the task. Thus the total number of activations with her existing technology would be lower than with the new technology.

The QUEST scores indicated that the participant and her caregiver were very satisfied with the new technology and this increased with training. The sections which were rated higher than the existing technology were ‘comfort’, ‘easy to use’, ‘dimensions’, ‘effectiveness’ and ‘weight’. Ease of use was rated higher after the 8-week training period. These scores are important to ensure that the technology is continued to use. Continued satisfaction and ease of use will mitigate abandonment of the technology.

From the GAS scores, it is shown that the goals set by the family and participant were achieved partly after the 4-week training period and exceeded by the end of the 8-week training stage. At the end of the 4-week period, goal 1 was not achieved as the new solution still required more cognitive effort and thus the activation time required was higher. This is expected given the new association that the participant had to learn between phoneme and keyboard quadrant.

The participant’s usage of the new access technology was higher than that of her existing technology. Frequency of usage increased as did her total time per use. This was a result of decreased effort required and faster activation time. Her usage decreased over the summer because there were not enough writing activities to perform. It is expected that her usage will increase with school starting again in the fall. The family does ensure usage 1-2 times per week to keep the participant familiar with the technology. Additional training will be provided if required when school re-opens for the participant.
7 Conclusion

7.1 Contributions

This thesis developed and evaluated a vocalization recognition system as an access pathway for a participant with extrapyramidal CP and dysarthria. The major contributions of this thesis are as follows:

1. Developed a five-code, voice-activated phoneme recognition system to assist a child with dysarthria in the task of computer-based writing. To our knowledge, there is no equivalent vocalization-based access pathway available commercially.

2. Implemented a classification algorithm (Gaussian Mixture Models with MFCC) for the system which is computationally less expensive than methodologies (Hidden Markov Models (HMM) with MFCC and Linear Predictive Coding (LPC) with formants) for required phoneme recognition. While HMM might offer higher accuracy, it is computationally intensive and requires larger training data sets. LPC with formants would be less intensive but formant tracking accuracy is generally low. Our current algorithm struck a balance between computational efficiency and accuracy, operating in real-time while achieving greater than 90% accuracy as demanded by the participant.

3. Strengthened the evidence that the “access technology delivery protocol” can improve satisfaction, usage and participation via an individualized access solution. While this protocol has achieved positive results for children with disabilities using other access pathways [28], this thesis adds to the validation of the protocol for a phoneme recognition access pathway.

7.2 Concluding remarks

This thesis developed and evaluated a 5-vowel voice-activated system to enable writing activity in a case study context. The majority of the results showed that this technology is more response
efficient than the participant’s existing access solution. The individualized goals were achieved and thus encouraged continued use of the technology.

The access technology proposed in this thesis could be customized to other participants with extant vocalizations. While the overall analytical methodology would remain similar, the classifier and features would be individualized as required to accommodate for the unique vocalizations of another participant. Also depending upon the target functional task, the number of classes could be augmented or diminished as needed. In general, the proposed system could serve as a prototypical access solution for participants capable of producing only a finite set of vocalizations.

7.3 Future work

This project involved training and testing in the control environment and in one of the natural environments-home. Future work would involve training the participant in the classroom environment and testing the system with the same measures as in the control environment. This would ensure that the access technology is sustainable in the classroom environment which was one of the goals set by the parent and the teacher. This process has been scheduled to be continued at the PRISM lab starting in fall 2013 and the participant has been enrolled in a research study for the same purpose.

Also the next steps in terms of improving the technology would be to modify the data collection in relation to the type of vocalization. More intuitive data like words relating to the quadrants of the on-screen keyboard or colors would make the process easier and reduce the cognitive load on the participant. If such a method is implemented, the feature extraction would involve extracting the segments relating to the vowels in the words and using only that information to classify the different words instead of using information from the entire word. This way the feature extraction and classification remains less computationally intensive despite of choosing more complex data. This would have to change on a case by case basis depending on the capabilities of the participant, the requirements for the level of access and the interface chosen.
8 References


