YORK UNIVERSITÉ UNIVERSITY

redefine THE POSSIBLE.

Digital Tangible Games for Speech Intervention

Foad Hamidi

Technical Report CSE-2012-02

March 2012

Department of Computer Science and Engineering
4700 Keele Street, Toronto, Ontario M3J 1P3 Canada

Digital Tangible Games for Speech Intervention

Foad Hamidi

This document has been prepared in support of the qualifying examination.
Department of Computer Science and Engineering

Supervisory Committee Members:

Melanie Baljko

Richard P. Wildes

Michael Longford

Table of Contents

1. Introduction 7

2. Clinical Speech Intervention 11

2.1. Disorder vs. Disability 11

2.2. Speech Language Disorders 12

2.3. Clinical Intervention and Speech Sound Disorders 13

2.4. A Categorization of Computer Speech Language Systems 17

3. The Use of Digital Media for Speech Intervention 21

4. The Use of Automatic Speech Recognition (ASR) 25

4.1. ASR and Speech Language Intervention 28

4.1.1. Box of Tricks 30

4.1.1. ARTUR 32

4.1.2. Speech Viewer II 34

4.1.3. Summary 37

4.2. ASR and Second Language Training 38

4.2.1. A Counter Intuitive Success Story 40

4.3. ASR and Speech Rehabilitation 42

4.4. ASR and Speech Input 43

4.5. Conclusion 47

5. Serious Game Design 51

5.1. Games Supporting Learning 52

5.2. Games Supporting Rehabilitation 54

5.3. Games Supporting Language Intervention 55

5.4. Uncertainty, Challenge and Game Design 56

5.4.1. Uncertainty 56

5.4.2. Dynamic Difficulty Adjustment 59

5.5. Games and Gender 59
6. Tangible User Interfaces (TUIs) 61
6.1. Overview 61
6.2. Situatedness and Shared Space 63
6.3. Flexibility in Design 65
6.4. Engagement of Movement 65
6.4.1. Movement and Learning 67
6.4.2. Movement, Speech and Language 69
6.4.3. TUI-based Speech Intervention 70
6.4.4. Movement and Engagement 74
6.5. Design Approaches Relevant to TUIs 76
7. Assessment 81
7.1. Challenges 82
7.2. Assessing Computer Language Intervention: *Fast ForWord* 85
7.2.1. Assessing *Fast ForWord* Using Randomized Controlled Trial Studies 89
7.2.2. Assessing *Fast ForWord* Using Case Studies 92
7.2.3. Summary 94
8. A Brief Description of the Tangible Toy 95
9. Conclusion 97
10. References 103

1. Introduction

In this document, I summarize and synthesize extant research that is relevant to the planned topic of my Ph.D. dissertation: the design and evaluation of digital interactive objects that are used as a support for *articulation drills,* a common mode of speech therapy for children and adults. My dissertation will focus on the support of speech therapy for children, which is presently deployed via “playful interaction”, albeit typically using “low-tech” toys and props. That the introduction of a *digital* plaything will be of clinical value is a key conjecture of the planned dissertation work. In my proposed system, I plan to combine *tangible user interfaces* (TUIs) and *automatic speech recognition* (ASR) to create the plaything that will be used in the context of a game that supports speech intervention.

This review aims to answer the following questions:

- What are the design characteristics of computational media that have already been developed for speech and language intervention?
    - To what extent has game design been used?
    - To what extent has *automatic speech recognition* (ASR) been used?
    - To what extent have *tangible user interfaces* (TUIs) been utilized?
    - To what extent have these three techniques been combined?
- What is the evidence, if any, that supports the combination of TUIs and ASR for this application domain?
- What kind of evaluation methodologies have been previously used to assess computer-based approaches to speech intervention?

In section 2, I first characterize and identify the speech disorders and user population that benefit from *repetitive articulation drills* (section 2.1 and section 2.2). This is followed by a brief discussion of current non-digital clinical methods for speech intervention (section 2.3). Next, I will categorize computer speech language systems and situate the proposed system with respect to other extant applications (section 2.4).

Section 3 presents a brief outline of the proposed system. The system provides a novel approach to speech intervention by combining *automatic speech recognition* and *tangible interface.* This approach has not been taken before and is a planned key contribution of the dissertation.

Section 4 focuses on the prior use of ASR in domains for non-standard speech: for clinical speech intervention (4.1), for second language training (section 4.2), for speech rehabilitation (section 4.3), and for assistive technology, enabling the use of speech as an alternative input technique (section 4.4).

Section 5 focuses on prior research on *serious games* in the domains of learning (5.1), non-speech rehabilitation (5.2) and non-speech clinical intervention (5.3). In section 5.4, two game characteristics that contribute to the engagement of games, *uncertainty* and *dynamic difficulty adjustment* are discussed. The section ends with a brief discussion of gender and game design (section 5.5).

In section 6, the relevance of *tangible user interfaces* (TUIs) and their role in engaging movement and spatial object manipulation is described. First, TUIs are introduced and their various aspects and characteristics are discussed (section 6.1 and section 6.2). TUIs will be used in the form of digital toys that elicit movement in the proposed system. Evidence from cognitive science and from developmental theory is provided that shows movement and speech are deeply linked. Furthermore, it is shown how object manipulation and physical movement provide a deeper sense of presence and serve to facilitate learning. The section ends with a brief discussion of design approaches relevant to TUIs (section 6.3)

Section 7 focuses on the important question of assessment (i.e., what should be measured and how) as it pertains to the proposed system. While this question is considered throughout the review and relevant assessment methodology are examined and presented in various sections of the report, in section 7 specific approaches towards assessing the proposed system are discussed. In human-computer interaction research, experimental design with dependant and independent variables is often viewed as the most effective approach for evaluating systems. Given the logistics of evaluating an intervention system for people with disabilities, there are alternative, more appropriate methods that can be used. Using studies on an existing computer language intervention system, the pros and cons of these assessment methods are discussed.

Section 8 provides a brief description of the tangible toy in the light of the reviewed research. Design decisions informed by research findings will be presented that address issues encountered in previous work.

2. Clinical Speech Intervention

2.1. Disorder vs. Disability

First, I carefully distinguish between *disorder* and *disability*, in order to delineate what is within the scope of a clinical intervention (disorder) from what is part of the fabric of an individual's subjective experience (disability). Conventional theory of disability holds that the subjective experience of disability arises from an impairment in functioning that may arise from an inability (the individual's or society's) to circumvent physiological disorder (Goggin & Newell, 2003).

Disability affects a large and varied population. Worldwide, at least one person out of ten experiences disability due to physical, mental, or sensory impairments (i.e., more than five-hundred million people), while at least one out of four is adversely affected by disability (United Nations, 2004). In recent years, there has been a major shift in understanding disability and its social and personal implications (Rosenbaum & Stewart, 2004). This shift is most apparent in the work of the *World Health Organization* (WHO) that has challenged the traditional *medical model* of disability that emphasizes the negative consequences of disease as defining factors in disability. As an alternative, WHO has provided a broader *social model* of disability (World Health Organization, 2001). In this model, disability is not viewed as a condition attributed to an individual due to disease or impairment. Rather, it is viewed as a social construct representing the dynamic interaction of the individual and society (Rosenbaum & Stewart, 2004). The social model of disability recognizes *contextual factors,* such as the availability and quality of intervention programs, as important factors in understanding and addressing disability (World Health Organization, 2001). Assistive technology holds great potential for helping this population with tools to improve quality of social and personal life. It also has potential to improve clinical interventions. In this work, I focus on the review of prior research that is relevant to the design of systems to supplement clinical speech intervention targeting speech errors.

Speech is an important social tool, the use of which greatly impacts many aspects of social life such as communication, development, education, social and psychological well-being and employment, among many others. In Canada, according to the *2001 Participation and Activity Limitation Survey* (PALS), 155,000 children aged five to fourteen (or about 4% of all children in that age group) have some form of disability (Statistics Canada, 2003). Of this population, 38% (approximately 58,500 children) reported receiving special education services, half of which services for speech or language difficulties (Statistics Canada, 2008). In the United States, also, roughly 5% of all children are affected by various speech disorders by first grade (NIDCD Health Information, 2010). According to the United States Department of Education, speech, language, and hearing impairments account for 24% of all Special Education students in the United States. This amounts to almost a million and a half individuals (U.S. Department of Education, 2005). Thus, there is a large population of children who would benefit from technological systems that make clinical speech intervention more effective.

2.2. Speech Language Disorders

Speech language disorders is a broad umbrella term that encompasses both lower-level *speech sound disorders,* which affect sound production and articulation, and higher-level *language disorders,* which involve linguistic processing and thus speech. These disorders can be *congenital* (existing before or at birth), or *acquired* (existing due to an external factor such as a disease or accident). *Developmental disorders* are disorders that appear during childhood and slow or impede the natural development of a child. *Language disorders* consist of *expressive* and *receptive disorders,* affecting active language creation and expression and passive language reception and comprehension, respectively. In addition to verbal language, these disorders also affect reading and writing. *Speech sound disorders* (also known as *phonological disorders*) are lower-level disorders that generally occur in children and may be due to physiological, neurological or developmental disorders. These disorders have historically affected the largest subgroup of children with speech or language disorders (Weiner, 1981). Speech errors caused by speech sound disorders are of four types: they either involve the omission of a sound, the addition of extra sounds, the distortion of a sound, or the substitution of one sound with another (Bauman-Waengler, 2004). These disorders are further categorized into *articulation* and *phonemic disorders* based on the underlying cause of the speech errors. *Articulation* or *phonetic disorders* are caused by difficulty in the physical production of sound. There are various causes for this disorder, ranging from inadequate learning of motor sequences to impaired motor planning (e.g., *apraxia of speech)*, or disorders involving muscle innervation (e.g., *dysarthria*). *Phonemic* or *phonological disorders* are caused by difficulty with the sound system of the language and may include difficulty distinguishing between similar yet different sounds (Bauman-Waengler, 2004).

2.3. Clinical Intervention and Speech Sound Disorders

Various comprehensive programs of intervention have been developed and are already in use that do not make use of computational supports (software and/or devices); these programs, such as the *Lindamood Phoneme Sequencing Program* (LiPS) (Lindamood & Lindamood, 1998), provide a comprehensive umbrella for reading, writing and speech intervention.

There are a number of computer-based speech language intervention systems available commercially. However, none target speech sound disorders specifically. Examples of these systems include the *Earobics* software (Cognitive Concepts, 2000), a reading intervention suite, the *Laureate Learning* software (Semel, 2000; Wilson & Fox, 1997), a language intervention program mainly targeting listening comprehension and vocabulary development and *Fast ForWord* (Tallal et al., 1996), a listening comprehension intervention system (Of these *Fast ForWord* will be further discussed in section 7.2). The design of many speech and language intervention systems (both computational and non-computational) is based on the idea of sustained exercise supported by feedback. Motor learning theory in speech development postulates that the two elements of repeated practice and accurate feedback are necessary to establish automaticity and skill transfer in untrained situations (Wiepert & Mercer, 2002).

Interventions that target speech sound disorders in particular are typically not computer-based. They are *speech language pathologist* (SLP) directed and involve the elicitation and repetition, often through play, of spoken utterances chosen to include the problematic sounds (Secord, 1981). Elicitation and repetition is used to engage the child in the purposeful physical production of the problematic speech sounds, in effect, allowing them to practice and to improve their ability to produce them correctly. In the case of phonemic disorders, this method involves the use of *minimal pairs,* words that differ in only one or two sounds, and provides the child with practice discerning and producing problematic sounds. Repetitive articulation drills are the basis of many clinical interventions for speech sound disorders.

The use of games and exercises during intervention sessions is prevalent. These games and exercises are mostly based on the use of "low-tech" props, such as dolls and pictures. These toys are used by the SLP to motivate and to elicit speech from the client. Following the production of speech by the client, the SLP provides constructive feedback and instructions addressing the problems causing the specific errors present. To be effective, these sessions are designed to be followed by hours of practice and exercise, in the form of homework activities, at home or school. The amount of practice and exercise varies according to the client, but can amount to additional hours outside of the time spent in direct contact with the SLP.

During direct contact with the SLP, various techniques are used to provide corrective feedback. These include using slower rate (Weismer, 1997), emphatic stress on target forms (Weismer, 1997), growth-relevant recasts, defined as the correct repetition of a mispronounced utterance without any further explanation (Camarata et al., 1994; Nelson et al., 1996), focused stimulation (Cleave & Fey, 1997; Fey et al., 1993), incidental teaching (Kaiser et al., 1992), scaffolding (Schneider & Watkins, 1996), and mediation (Miller et al., 2001). In some methods*, modified speech* is used, for instance, in the *Fast ForWord* method, specific sound transitions are slowed down (discussed in detail in section 7.2).

In these face-to-face clinical sessions, the use of recasts without detailed feedback is sometimes enough to correct pronunciation in the short term (Lyster, 1998). This method, referred to as *imitation* or *auditory stimulation,* involves the repetition of a target sound by the client after the SLP has presented a number of examples; Auditory stimulation, is widely used and has been suggested by a number of clinicians as the first method that should be used for speech elicitation (Bernthal & Bankson, 2004; Bauman-Waengler, 2004; Secord, 2007). Other clinicians hypothesized that detailed feedback is not necessary for proficient learners and a redirection to the correct alternative, when a mistake is made, is often enough for correction because of the user's familiarity with the correct sounds in the linguistic inventory (Nicholas et al., 2001).

Two approaches toward speech training can be distinguished (Vicsi et al., 2000). In the *process-oriented* approach, used by SLPs, children attempt to correct their speech on the basis of instructions on how to use speech organs while forming sounds (Povel, 1991). In the *product-oriented* approach, used in natural language acquisition, children learn how to control their speech organs through references to acoustic speech signals (Daniloff et al., 1980).

Several assessment tools have been used clinically in the past for the assessment of speech articulation and receptive and expressive language. These include the *Comprehensive Assessment of Spoken Language* (CASL) (Carrow-Woolfolk, 1985), the *Test of Language Development-Primary* (TOLD-P) (Newcomer & Hammill, 1982), the *Token Test for Children* (DiSimoni, 1978), and the *Goldman-Fristoe Test of Articulation* (GFTA) (Goldman & Fristoe, 1986). These tools have also been used to compare the effects of an intervention using pre- and post-intervention assessments (Loeb et al., 2001; Gillam et al., 2008).

The prevalent clinical methods for intervention are *behavior-based*, meaning they involve intensive speech therapy, typically administered by a SLP in one-on-one sessions. These sessions involve repetition of problematic words and sounds, with corrections by the SLP through the techniques described earlier (e.g., recasts). Behavior-based approaches are generally effective for the alleviation or even elimination of speech disorders (Johnson & Jacobson, 1998).

Although behavior-based approaches are effective for many clients, they are resource intensive. They require frequent one-on-one sessions with a professional clinician. As well, behavior-based approaches also often require compliance with a regime of homework exercises and practice without direct supervision of the SLP. This has created the need for supports and frameworks to help clients (and their caregivers and supports), since the regime may require sustained efforts and persistence.

Computer-based systems such as *Earobics* (Cognitive Concepts, 2000), *Laureate Learning Software* (Semel, 2000; Wilson & Fox, 1997), and *FastForward* (Tallal et al., 1996), allow the SLP to design exercises to be completed in his or her absence. Using appealing visual and audio features as well as basic gaming elements such as rewards and scorekeeping, the exercises motivate children to interact with the computer and use their speech. The systems also create activity logs and reports automatically, which allow the SLP to assess system use by the client. These systems generally do not take advantage of ASR technology (systems that do use ASR are covered in section 4) and their interfaces are limited to the traditional desktop configuration and do not use TUIs. Finally, game characteristics of *uncertainty* and *dynamic difficulty adjustment* are not employed.

2.4. A Categorization of Computer Speech Language Systems

The computational intervention systems described above (e.g., *Earobics*) are primarily designed for use by children in order to correct or improve speech errors due to congenital or acquired speech impairments. In section 4 below, a number of other computational systems will be discussed. Some of these are designed to support speech and language intervention, but not all.

Collectively, I will refer to these as *computer speech language systems*. I categorize the systems as follows:

- Therapeutic systems
    - Intervention systems: used mostly for children to correct or improve speech errors due to congenital or acquired speech impairments. Systems that, if adopted into clinical use would have been described in section 2.3.
    - Rehabilitative systems: used mostly for adults with acquired rather than congenital speech impairments (e.g. due to disease or accident for example a stroke)
- Training systems: used by both adults and children to learn or to improve skills in a new language or dialect or to reduce accent

Rehabilitation systems and intervention systems typically serve different clients and are used in different contexts. Rehabilitation can be geared towards a variety of client populations, but most typically computational rehabilitation systems are geared to older adults. Intervention systems are generally geared towards children. It is conventionally believed that children are more receptive to intervention than adults; it is known that the earlier an intervention is applied the higher the chances of success are. Research in developmental psychology has shown that the rate of human learning and development is most rapid during the early childhood years (Caplan & Caplan, 1993). Rehabilitation, defined as "a dynamic process of planned adaptive change in lifestyle in response to unplanned change imposed on the individual by disease or traumatic incident" (Gunasekera & Bendall, 2005), typically involves older adults who have lost some ability due to accident or disease.

Previous research has identified needs and characteristics, such as decreased sensory acuity and slower responses specific to the users of rehabilitative systems that make them different from children with disabilities that are typically the user group of intervention systems (Flores et al., 2008).

The contexts of use differ as well. Intervention systems are usually used in specialized classrooms, private therapy sessions, and speech clinics, whereas rehabilitative systems are usually used in hospitals and rehabilitation centers.

Computational training systems typically target either adults or children and are intended to support learning or the improvement of skills in a new language or dialect or to reduce accent. These computational systems differ from computational intervention systems in key ways: they target users without impairments and are used either in a regular classroom setting or at home. It is usually assumed that the target users do not have speech impairments, already know a mother language well, and are using the program to learn or develop proficiency with another unfamiliar language. The emphasis is on carrying over knowledge from one language to another. Furthermore, the target population is usually adult. This means when ASR is incorporated in the application, the feedback is given mostly in the form of scores or diagrams which are abstractions that are usually understandable to an adult. These forms of feedback are not suitable for children. As will be discussed in detail in section 4, complex forms of feedback can be disorienting and confusing for children.

3. The Use of Digital Media for Speech Intervention

Given the current state of computer speech language systems used in a clinical setting, as discussed in section 2, several opportunities are identified:

- Incorporating ASR as part of the intervention system
- Using interactive toys as the primary interface to the system
- Using game design techniques to motivate higher amounts of interaction

I plan to implement a digital media approach to the support of speech therapy. A digital interactive toy will be designed to engage both speech and movement through spatial object manipulation; it is intended both to support sessions with an SLP as well as to support the homework phase of speech therapy. Previous research has identified these goals as important for success in speech therapy (Bunnell et al., 2000).To this end, *automatic speech recognition* (ASR) is employed to engage speech and *tangible user interface* (TUIs) are employed to engage movement; these two technologies are combined in the context of an engaging therapeutic game.

In my planned approach, the system's responsive interaction can be viewed as an *extension* of the SLP's role via virtual asynchronous presence. While SLPs, through their nuanced feedback can support the process-oriented aspect of the intervention, the proposed system would support a product-oriented process in support of the SLPs work, by engaging the children in speech. The system is going to support the continuation of speech exercises in the absence of the SLP.

By interaction with speech, the proposed system aims to achieve two results: firstly, to motivate the child to repeat words and phrases to practice speech; secondly, to record the child's speech during practice for later analysis by the SLP. The aim of the ASR module in the proposed systems is to engage speech not to provide detailed feedback.

The therapeutic system is envisioned as consisting of both hardware and software modules:

- A customizable digital plaything used for therapeutic game play at home or school (D-Toy)
- A *data collection module* (DCM) that gathers information during game play for later analysis
- A suite of software tools for use by the SLP that allows for interaction with the two previously described parts of the system (SLP-UI)

The system is envisioned to be used in the following manner: The SLP customizes the game for each client and specifies settings such as what words or phrases are to be practiced, how closely should the pronounced words resemble a target pronunciation, how much exercise in terms of hours spent with the system are required until the next session and so on. Once these parameters are set a "play period" is started during one-on-one therapy sessions with the SLP. The client will then have time to use the game at home or in other settings when the SLP is not present. The "play period" ends when the SLP reviews the client's interaction with the system by accessing the data gathered by DCM and viewing summary reports provided by the system in the form of user-friendly statistics and charts. The evaluation of the system will entail evaluation of the usability of *D-Toy,* the functionality of the DCM and the usability of SLP-UI.

The system is intended to target intervention that addresses speech sound disorders. However, the use of elicitation and repetition of certain words and phrases is also useful for therapeutic methods that address other disorders, such as expressive language disorders. Thus, the system ultimately may be potentially useful in other therapeutic contexts (e.g., rehabilitation or second language learning).

Many aspects of the design of *D-Toy* are informed by the literature review that follows. These aspects will be described in section 8, after the previous research review is presented.

4. The Use of Automatic Speech Recognition (ASR)

Automatic speech recognition (ASR) refers to a series of techniques combining signal processing, statistical modeling, and machine learning to analyze and to interpret human speech typically by deciphering input acoustic signals into phones or other linguistic elements such as syllables, words or phrases (Cohen et al., 2004). ASR often makes use of *Hidden Markov Models* (HMMs) which are probabilistic predictive models that can be defined at different linguistic levels, ranging from phonemic to phrasal.

ASR is used in many applications in the disability community, such as for speech-to-text communication technologies and for command interpretation systems for hands-free computer use. Numerous automatic speech-recognition software modules have been developed in the last decade both for commercial and research purposes. Some of the more well-known systems include *Dragon Naturally Speaking* (Dragon Systems, 1999), *CMU Sphinx* (Walker et al., 2004), IBM's *ViaVoice* (IBM, 1998) and *Microsoft Speech* (Microsoft Corporation, 2009).

Traditionally, the goal of ASR has been to identify words, even in the face of deviation from the training pronunciation. Extant ASR modules aim to *recognize*, rather than *evaluate* or *analyze,* input speech. In traditional ASR, the input speech is not known and the goal is to recognize it.

When ASR has been deployed in *computer speech language systems* for users who do not have standard speech, it has been in support of the tasks to *assess* input speech in the context of remediation. The input speech is already "known" in the sense that the user has been provided with a target utterance. Researchers who use ASR in therapeutic or training systems that involve non-standard speech input should be aware of these discrepancies between the traditional ASR and ASR in intervention-based systems. For the latter, the goal in using ASR techniques is to assess how closely the input resembles an expected result, as opposed to recognizing unknown speech.

The potential utility of ASR for enhancing education and therapy for people with disabilities is recognized in previous research (Zhao, 2007). However relatively little research has been conducted in this area (Okolo & Bouck, 2007). This is mainly due to challenges non-standard speech poses for ASR, which will be described below.

ASR modules are generally developed for speakers with clear speech. These modules use databases of phones or diphones that are derived from human speech samples. When the input speech differs from the modeled speech, due to reasons such as when the input speech is produced by a speaker with an accent or a speech impairment, the performance of the ASR module degrades. The performance further degrades when speech is affected by environmental noise, distortion and sound quality change (Huang et al., 2001).

An *error,* for the purposes of this review, can be understood as either a *recognition error,* where input is “correct” but the system fails to recognize it, or a *speech error,* where input speech significantly deviates from a standard model. Recognition errors can be further categorized into *false positives* and *false negatives.* A *recognition result* is the output of the ASR in the form of a lexicographical interpretation for some particular input acoustic signal. If the lexicographic candidate matches the input speech, then the ASR module has produced a true positive. If the ASR module produces a lexicographic candidate that does not match the input speech, then this is a *false positive* (an erroneous recognition result when it does not exist). If the ASR module fails to produce a lexicographic candidate for a speech signal that correspond to a correct utterance of the candidate, then a *false negative* (i.e., erroneously not recognizing a result that does exist) recognition error is produced. Despite rapid improvements in ASR technology, some researchers believe that because sound and specifically speech is a noisy input channel, errors are an inevitable part of ASR technologies (Danis & Karat, 1995). In the presence of non-standard speech, the ASR module produces low confidence scores for predicted candidates, reflecting the high possibility of recognition errors.

When using ASR for speech intervention, additional challenges are present. In addition to recognition correctness, the ASR module needs additionally to distinguish among degrees of correctness for a given lexicographic target. If the ASR module is unable to produce a lexicographic candidate, then the system must decide whether this error is an instance of speech error due to poor speech or a recognition error due to poor system performance. In this situation, the system may exploit additional information in the form of information about the expected input.

The remainder of this section is ordered as follows: In section 4.1, speech intervention systems that are designed for children are reviewed. In section 4.2, results from second language training systems are presented. In section 4.3, ASR-based speech rehabilitative systems are reviewed (there are many systems aimed for speech rehabilitation, but the focus here is on the ones that use ASR). In section 4.4, the use of ASR for interpreting speech input for assistive technology applications is examined. In section 4.5, the use of ASR for making speech an alternative input mode is discussed and two main techniques, *input restriction* and *acoustic training,* that are employed for ASR with non-standard speech are discussed. These sections will identify and address pitfalls and challenges of ASR techniques in this domain, as well as the gains that have been made.

4.1. ASR and Speech Language Intervention

In this section, ASR-based speech language intervention systems are reviewed. Other types of computational intervention systems will be reviewed in section 5.3.

A number of computer intervention systems that use ASR have been designed to engage children in repetitive articulation drills, usually in the context of games, in the absence of SLPs. These systems also assist in record keeping and reporting. In this section, these systems will be reviewed and important questions pertaining to the role of ASR in their design will be addressed. These questions are: to what degree do errors recognized by ASR correspond to errors detected by human listeners? In the face of the discrepancy discussed in the previous section, can ASR provide automatic constructive feedback based on the quality of the input speech? And can ASR be used effectively with children's speech?

Previous research has shown that, in principle, it is possible to distinguish between "good" and "poor" pronunciations of a target word spoken by a child. For instance, Series (1993) conducted "off-line" experiments in which comparisons were made between the probability of a speech pattern that was conditioned on standard speech *Hidden Markov Models* (HMMs) and a pattern that was conditioned on word-level HMMs. Russell et al. (1996) developed a system for assessing the quality of children's pronunciation, based on making a comparison of the child's input utterance of a known word with two different models of the word: a word-based model formed from a continuous HMM of the word and a general speech model created by pooling phone HMMs. They developed the following heuristic: if the word-specific model was found to provide a better fit to the utterance than the general speech model, then it was assumed that the utterance was a good example of the intended word; otherwise, it was assumed to be a poor model of the word. The system demonstrated an *equal error rate* (ERR) of 30% in distinguishing between correct and minimally contrasting incorrect words.

Bunnell et al. (2000) describe a computer system for children, *the Speech Training, Assessment, and Remediation* (STAR) system that was designed to distinguish between the segments /r/ and /w/. Intervention was achieved through playing a role-playing game in which “aliens” are to understand spoken speech. They conducted an experiment in which likelihood ratios, as calculated automatically by the ASR module, were compared with perceptual quality ratings, as provided by human judges. The results showed high correlation between the two measures for substitution errors. In other words, the system worked well when /r/ and /w/ were misarticulated. However, the ASR module produced many false positives (the results correlated poorly for correctly articulated examples).

Kewley-Port et al. (1991) presented another system in which recorded templates of the child’s best production were elicited and then used as standards against which to measure the acceptability of new utterances. The degree of success of the system was not clear. The researchers conjectured that recognition error rates as high as 20%, a rate within the capabilities of a small vocabulary speaker-dependent system, would be acceptable for articulation training. However, adoption of this approach has been limited by its shortcomings: that training is required for each individual, and that target words and phrases that consist of segments not producible by the child are not allowed for. The overall efficacy of these approaches is unclear.

4.1.1. Box of Tricks

Vicsi et al. (2000) developed a speech intervention system, *Box of Tricks,* for children with hearing impairment. *Box of Tricks* uses ASR to detect and to provide feedback about speech mistakes and was originally devised to support Hungarian, but has subsequently been expanded to also support English, Swedish and Slovenian. *Box of Tricks* is used to train for vowels and also correct misarticulated fricative sounds.

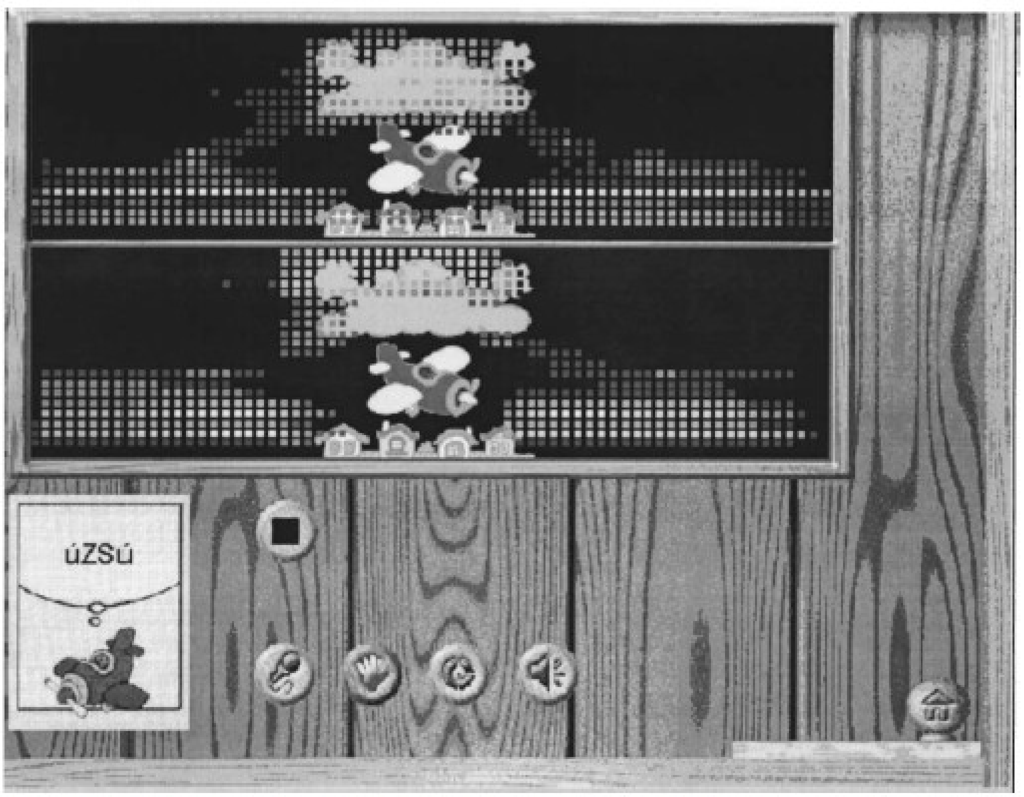

Figure 1. Visual representation of a reference and an input pronunciation spectrograms (Vicsi et al., 2000)

The researchers have adopted a *product-oriented* approach in which children use references to acoustic speech signals to correct their speech. The goal of *Box of Tricks* is to teach children to modify their speech on the basis of visualizations of their speech signals. Picture-like images of energy, changing in time, fundamental frequency, voiced or unvoiced detection, intonation, spectrum, spectrogram (cochleogram) and spectrogram differences were used for the visualization. A sample screen shot of the program is shown in Figure 1.

For the visualizations, a filter was developed and applied that produces a representation based on inner ear processing rather than FFT spectra. The researchers hypothesized that the visualization generated by this filter would be a more intuitive representation of speech as interpreted by humans. The representation of input pronunciation was shown in alignment with a representation of a target pronunciation. Parts of the representation reflected more important features of the speech; to draw the children's attention to these parts, they were highlighted by amusing background pictures.

Box of Tricks did not provide instruction to the children about how to correct their speech, however. Users can see that the feedback indicates that there is something different in the input speech from the desired speech, but it is not clear how this difference can be decreased without corrective instructions.

As a component of this work, Vicsi et al. (2000) proposed the *average spectrum distance* (ASD) as a metric that indicates how close an input pronunciation is to a target pronunciation and corresponds to the human evaluation of speech. This metric is calculated as the distance of the spectrum components of an input pronunciation from the averaged target examples. In order to derive the parameters for this metric, an experiment was conducted that involved collecting human speech samples, followed by human judge evaluations. The scores were then correlated with automatically generated scores in order to find out whether input pronunciations could be reliability categorized.

The researchers claim that this approach provides meaningful feedback to children and allows them to use the system by themselves. However, no studies were conducted to measure how useful the representations are for children.

4.1.1. ARTUR

Bälter et al. (2005) developed a prototype of a computer system for speech intervention for children with hearing impairments to be used in the absence of SLPs. The system aims to identify problematic pronunciations and provide corrective feedback. An animated head with exposed internal parts of the face and mouth, referred to as the *ARticulation TUtoR* (ARTUR), was constructed. ARTUR was utilized to provide feedback based on the input. The researchers hypothesized that for children with hearing problems, the visualization of the movement of vocal organs is more useful than acoustic signal visualization. A database of possible errors and responses, in which each response was mapped to one or more errors, was constructed. The feedback, in the form of spoken commands and corresponding animation was chosen from this database. The researchers claim that showing the hidden parts of human head involved in speech production is key to the correction of speech impediments. In the final implementation of the system, audio input is to be supplemented with video footage of the user for more accurate categorization of pronunciation error.

The program was tested with two groups of children in a Wizard-of-Oz study. The children in the first group were six years old and the ones in the second group were between nine and eleven years old. In addition to children, an adult with English as second language also used the program and provided feedback.

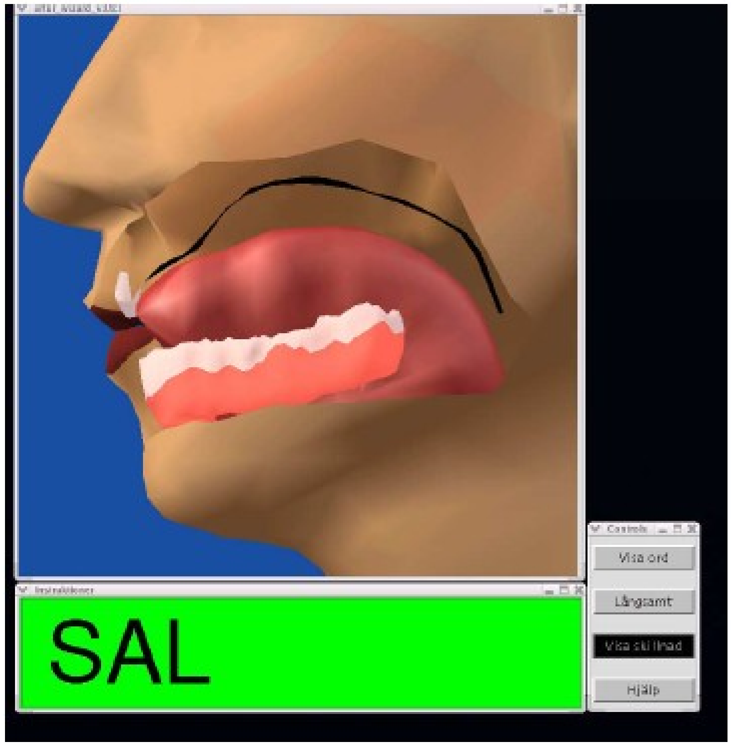

Figure 2. User interface of articulatory feedback (Bälter et al., 2005)

The empirical qualitative data demonstrated that the children, especially the older group, liked the idea of playing with a computer and being given explicit feedback. However, while they (and especially the older group) liked the program in general, they found the visual feedback confusing and unhelpful. This was found of both the image representation of speech organs and the accompanying animation. The children suggested that adding more game-like features, such as goals and rewards, to make it more engaging. Also, they found the user interface of the program, as well as the anatomy of the vocal tract (e.g. the hard palate), unclear. When compared to interaction with the SLP, older children described the interaction as more relaxed.

The adult user suggested that it should be possible to make a few practice pronunciations before being evaluated. The wizard who simulated automatic feedback observed that the choices available for feedback were not well-grained and precise. Many situations were not covered and, as a fall-back strategy, encouragements were used after repeated errors. They suggested that repeating the same feedback after same error is not a good strategy, and also that general ambiguous feedback should be used for times when the error is hard to classify.

4.1.2. Speech Viewer II

A commercial speech therapy system, *Speech Viewer II,* was developed to help adults with speech impairments improve their speech (Öster, 1995). This system visualized speech signals and waveforms. Figure 3 shows speech visualization as produced by this system.

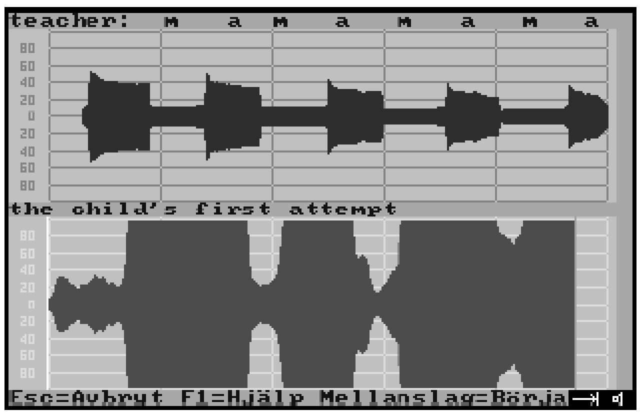

Figure 3. Speech Viewer II displays loudness as produced by the therapist (upper wave) and the client (lower wave) (Öster, 1995)

While this system was originally developed for adults, several studies examined its potential for children with hearing impairments and encountered mixed results. Two studies have shown that it does not work well for use by children with hearing impairments. The first study showed that the program did not have any advantages over traditional speech therapy for vowel training for children with profound hearing impairments (Ryalls et al., 1994). The second study tested a vowel accuracy feedback with children with hearing impairments and showed that the system produced modest gains but exhibited inaccuracies and inconsistencies in feedback (Pratt et al., 1993).

The use of *Speech Viewer II's* visualizations have been unsuccessful for improving pronunciation in children, specifically correct sound production. However, it has been promising when used to improve prosodic features of speech for children with hearing impairments. Öster (1989) conducted a study with two deaf children who were trained using the program for ten minutes twice weekly over an eight week period. For each child a different skill was targeted. The first child, a fifteen-year-old boy had difficulty with producing durational contrast between phonologically long and short vowels. The second child, a thirteen-year-old girl, had difficulties producing voicing contrasts between voiced and voiceless velar stops. Both children were reported to have improvements in the areas targeted. Öster (1995) also conducted a study with a five-year-old deaf boy who had difficulty controlling the loudness and pitch of his speech. While detailed information about the amount of training, methodology and the results of the intervention is not provided, the researcher claimed that using the program and specifically its graphical interface allowed the SLP to communicate better with the child resulting in improved loudness and pitch.

Öster et al. (2003) have conducted initial experiments with a system that produces visual maps for training Swedish *sibilant fricatives*, fricatives with higher-frequency and acoustic energy than non-sibilant fricatives, to hearing-impaired users. The system is designed to supplement speech intervention. The user is provided with a visual representation of his or her speech that is shown in relation to a visualization of a target pronunciation. The researchers hypothesize that having this feedback will help increase the frequency of correct pronunciations.

Initial maps were created for three sibilant fricatives, /s/, /C/ and /S/, using the speech of a normal hearing girl. The subject was instructed to pronounce CV combinations, where C was a sibilant fricative /s/, /C/ and /S/ and V was either /i/ or /u/. The data files were manually labeled and subjected to cepstral analysis before being input to a multilayer perceptron that outputs the positions of the points to be shown in the feedback map in the form of points each corresponding to a fricative and placed with respect to their frequency range. Figure 4 shows an instance of the map.

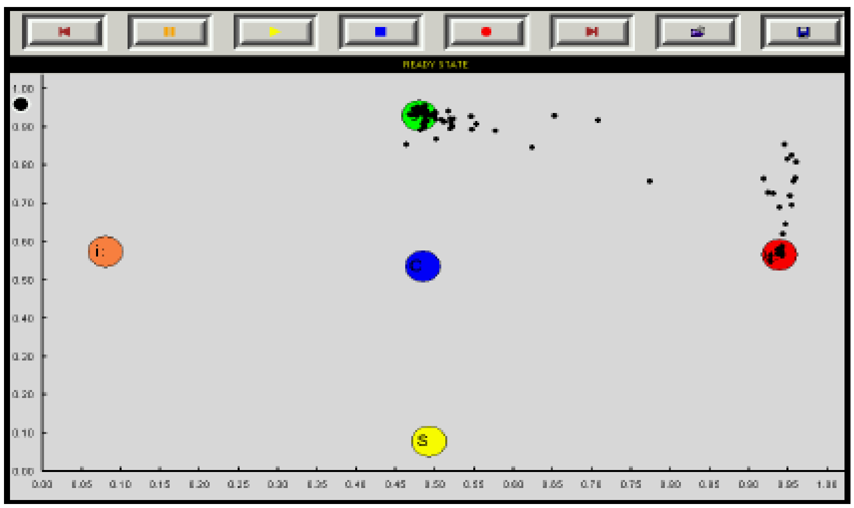

Figure 4. Map interface where input points (black dots) are placed relative to trained data (fixed color circles) for Swedish sibilant fricatives (Öster et al., 2003)

The speech of three severely hearing-impaired children when pronouncing the fricatives, using CVC combinations where C was one of the sibilant fricatives mentioned before and V was either /i/ or /u/, was recorded and mapped against the created maps and it was found that the visualizations corresponded well with the visualization produced.

While this system shows it is possible to create visualizations that correspond with non-standard input speech, it does not discuss the usefulness of this approach for children. The input data is severely restricted, to CVC combinations rather than words, and the visualizations are shown in terms of time and frequency, an unintuitive approach for children. The project is in its initial phase and no user studies are conducted.

Neri et al. (2003) have identified a major problem with providing comparable waveforms, a popular form of feedback (e.g., Speech Viewer II), to the user. While showing a target and input waveform motivates the user to try to emulate the target waveform by modifying their pronunciation, it does not necessarily lead to correcting pronunciation and might be misleading. Two pronunciations might be correct and contain the same phonetic content but still have waveforms that are very different from each other. Neri et al. (2003) assert that even a trained phonetician cannot extract information needed to correct pronunciation from this feedback, let alone a user that does not have any training in interpreting this form of feedback.

4.1.3. Summary

In the reviewed systems, the provision of meaningful and consistent feedback to children was shown to be the key challenge. Previous research has identified the main cause of this problem as the fact that ASR technology is not mature enough to provide corrective analytic feedback and can only provide evaluative feedback (Hincks, 2002; Menzel et al., 2000). Another possible cause for this problem is the discrepancy mentioned in the previous section: ASR modules being originally developed for *recognizing* speech rather than *evaluating* or *analyzing* speech.

The importance of using ASR to facilitate an interactive dialogue between the user and computer is identified previously (Cohen et al., 2004). An important requirement for this dialogue to be successful is that the feedback provided by the computer be motivating and understandable. In the face of the tension between this requirement and the challenge outlined above, the proposed system will be designed such that the ASR module is used to provide interactivity rather than detailed feedback.

ASR has been successfully used in language interventions other than speech. For example, Adams (2006) developed a computer application, *The Reading Assistant* that uses ASR to promote literacy. The system provides feedback based on mispronounced words. Each time a mispronunciation is detected, the words are highlighted and the user can request their meaning or listen to their correct pronunciation. After each session, the program creates reports for both students and teachers that include statistics on usage patterns and words per minute accuracy. The user's voice is also recorded allowing for subsequent review. While a preliminary experiment showed that using the software can improve fluency, the study did not include a control group (Adams, 2006).

Studies by Higgins and Raskind (Higgins & Raskind 1995; Raskind & Higgins 1999) and Wetzel (1996) showed that students with learning disabilities showed significant improvements over a control group with respect to word recognition, spelling, reading and writing after using a computer program with an ASR module. Relevant intervention systems that target areas other than speech will be further reviewed in section 5.3.

4.2. ASR and Second Language Training

Second language training systems that use ASR face many of the same issues as speech intervention systems: they deal with non-standard speech and they need to *analyze* or *evaluate* the speech rather than *recognize* it. In addition to these issues, non-native speakers produce different formant structures from native speakers (Arslan & Hansen, 1997). This causes recognition difficulties for ASR systems that attempt to match and recognize formant structures. As with intervention systems, doubts have been expressed about whether errors recognized by the system correspond to ones recognized by native listeners (Derwing et al., 2000). Furthermore, the problem of providing meaningful and accurate automatic feedback is also present here.

Despite these issues, in recent years, a number of applications in this area, most notably *The Rosetta Stone,* have incorporated ASR to help users test their language skills. Furthermore, ASR modules have been successfully incorporated into *computer assisted language learning* (CALL) and *computer assisted pronunciation training* (CAPT) courseware (ISLE, 1999; Franco et al., 2000; Witt & Young, 2000). Some researchers view usage of ASR as a natural choice for second language training application (Bemstein, 1994).

Do the feedback and accuracy scores provided by second language training systems correspond to the accuracy of speech as perceived in the real-world? Wildner (2002) observed that native speakers sometimes received lower ASR-based scores than non-native speakers. Also, Mackey and Choi (1998) found that the scoring algorithm in another ASR-based program was inaccurate and resulted in many false positive errors (for example, words whose last syllable were omitted were often accepted as accurate). In a Japanese language learning program, it was found that the scores given by the system differed from a professional teacher's ratings for pronunciation exercises provided (Miura, 2004). Reesner (2002) criticized the French version of the program and found that there was little basis for the presented scores.

Second language learning systems also face the same problem as speech intervention systems in producing useful visualizations. In such systems, the feedback in the form of sound waves and pitch curves was found to be problematic. To compare their input speech with the stored model, the users had to speak with the same speed with which the model phrase was spoken.

There have been a number of studies that demonstrate that off-the-shelf ASR systems cannot recognize non-native speech well. Coniam (1999) tested the accuracy of *Dragon Naturally Speaking*, a state-of-the art-system at the time of the study, with ten Cantonese speakers to see how accents affect performance. The study demonstrated, perhaps unsurprisingly, that the system was considerably less effective for Cantonese-accented English. It has been observed that ASR performs poorly (around 11% accuracy) for non-native speakers (Glass & Hazen, 1998; Teixeira et al., 1997; Zavaliagkos et al., 1995). Previous research has shown that recognition rates of ASR modules increase for non-native speakers when acoustics (Diakoloukas et al., 1997; Nguyen et al., 1999; Zavaliagkos et al., 1995) and pronunciation (Humphries et al., 1996; Liu & Fung, 1999) are taken into account and incorporated.

4.2.1. A Counter Intuitive Success Story

One intriguing study found that very rudimentary feedback was helpful for accent reduction (Mitra et al., 2003). In this study, sixteen children between the ages of twelve and sixteen were chosen from an Indian English median school, where English was the primary medium of teaching. The curriculum of the school was taught in English and the children could read and write English at different levels of competency, but their mother tongue was not English and they had English poor pronunciation. The children were grouped into four groups of four and were given access to a computer for three hours a week. They were provided with "Ellis", an English language learning program with no ASR support, four classic English-language films that the children could choose to watch during their time at the computer and the previously mentioned *Dragon Naturally Speaking* program. The children were given the objective of making their speech understood by the *Dragon Naturally Speaking* program. The program either accepted or rejected input speech and did not provide corrective feedback. Pre-experiment pilot trials were conducted with various configurations, and the version trained on a female voice was used as this was found to offer the highest recognition rates for the speech of children of both sexes.

The children were given a demonstration of the resources (i.e. programs and films) available on the computer but no other instruction or intervention was provided during the experiment period. The researchers tested the following hypotheses in this study: could *Dragon Naturally Speaking* be used to help learners modify their speech to reduce accent? Can children collaborate to use computers to achieve improved intelligibility of speech with minimal intervention?

Rather surprisingly, the approach was effective. To measure improvements in speech and whether they carry to real-life situations, four human judges were provided with video clips of children speaking at different evaluation points. A measure of the percentage of words correctly recognized was calculated. Significant improvements over a five month period were observed. Furthermore, the word recognition rates by the ASR module were found to be closely correlated with the human judges' assessments of pronunciation accuracy (e.g., an improvement of 117% was observed, as assessed by the human judges and of 79%, as assessed to the ASR module).

A key point in this study is that ASR was used only to provide a binary decision on whether a word was pronounced properly or not. This approach avoids the problem of complicated and inconsistent feedback by focusing on *recognition* rather than *evaluation*. This approach does not avoid the problem of recognition errors in the sense that the system might accept words that are mispronounced (i.e., *false positive*) and reject ones that are properly pronounced (i.e., *false positive*). The computer and microphone facilitated an interactive context in which children were motivated to try multiple repetitions of words and phrases. Thus, even in the absence of detailed corrective feedback the children were able to figure out ways to improve their speech and to imitate correct pronunciations without further guidance.

4.3. ASR and Speech Rehabilitation

ASR has been used in speech rehabilitation systems. Piper et al. (2010) developed an augmented pen-and-paper interface in which the pen, augmented with a camera and microphone, can detect patterns on a paper, playback a recording of it and record the user's voice.

The design of the system was based on a field research of SLPs providing speech rehabilitation for elderly adults with aphasia caused by stroke. Using this ethnographic approach, they found that rehabilitation under the supervision of an SLP often takes place during ongoing sessions that are customized to the client. The sessions are often face-to-face, collaborative, and multimodal. They further observed that, for the elderly population, a lot of paper-based solutions are used.

Using these observations, they designed and created a prototype of the augmented pen. The prototype was shown to 15 SLPs who observed that the approach can be successful because using pen and paper is a familiar medium for their clients and the system affords independent practice for clients who live alone and also help with communication barriers between clients and SLPs. They did preliminary testing of the interface with two elderly users who were successful completing basic tasks such as selecting words and repeating pronunciation using the interface. The researchers also aim to use the system as an assistive communication device where the pen is used to select desired phrases on paper to be vocalized.

4.4. ASR and Speech Input

ASR has been used to provide speech as a mode of input for assistive technology applications ranging from environmental controls to computer access. In these systems, speech provides an alternative to the keyboard and mouse or other input devices that require fine motor skills to operate. Speech can also be used in conjunction with other input modes such as keyboard or mouse to increase efficiency and/or reliability (Oviatt, 1997). Due to the diversity of their disabilities, the users of these applications have a wide range of speech quality ranging from clear to severely dysarthric.

ASR has been used for computer input and environmental control for more than three decades. Research has shown that ASR in combination with button presses increases the rate of text entry for individuals with physical disabilities (Treviranus et al., 1991). For example, Manasse et al. (2000) found that an individual with severe traumatic brain injury was able to learn to use voice-input technologies to create greater amounts of text when using it than working with a keyboard.

The use of speech as a mode of input poses special challenges for technology designers and the disability community (Hawley, 2002; Rosen & Yampolsky, 2000). Dysarthria is a speech disorder common in people with various physical disabilities. It has multiple causes including neurological damage brought about by various conditions such as strokes, head or neck injuries, cerebral palsy and degenerative diseases such as *amyotrophic lateral sclerosis* (ALS) (Bauman-Waengler, 2004). Patients with dysarthria show a wide range of speech errors. Due to this high degree of *intra-speaker variability*, a system that works well for one person with dysarthria might not work well for other people with dysarthria (Hasegawa-Johnson et al., 2006). Another important characteristic of dysarthric speech is that the same person can show different speech patterns at different times, causing a high degree of intra-speaker variability, making the speech especially challenging for ASR systems to analyze (Netsell & Abbs, 1979). Despite these challenges ASR has been used for more than three decades to provide environmental and computer control for people with dysarthria (Clark & Roemer, 1977; Cohen & Graupe, 1980). Various studies have shown that in some cases this technology is faster than using the keyboard for people with motor disabilities who have difficulty using the keyboard (Chang, 1993; Thomas-Stonell et al., 1998; Hux et al., 2000).

Hawley et al. (2003) designed a speech interface for adults with dysarthria (shown in figure 5). The interface was designed to enable users to both access practical computer applications, specifically an augmented communication application and an environmental control system, and practice their speech. The goal in speech practice with this software was to change the speech in such a way that the computer can better recognize it (as opposed to making the speech more intelligible for humans listeners). To this end, an interface was developed that used a small vocabulary recognizer with models stored for each word that were then compared to input from the user. A best fit was chosen for each word, which corresponded to the input the system could recognize more easily and not necessarily to an intelligible pronunciation. Users practiced their speech by comparing two similarity measures shown as bars when the word was pronounced. One measure corresponded to a closeness of fit of the current utterance and another to a log probability of the model generating the word on the Viterbi path. The users had the option of listening to the current best fit example and attempt to modify their pronunciations to raise the bars. This system stored the input utterances so that a database was created, making recognition more robust over time.

Figure 5. User interface showing two bars representing similarity measures (Hawley et al., 2003)

The system was tested with five individuals with dysarthria. Improvements in speech recognition were observed for three of the users over a five year period. These improvements might be due to the expansion of the database rather than actual speech improvement. It was observed that while training improved system performance, the physical and health conditions of users made extensive training difficult for them.

Over the years, two main strategies are developed, which can be used in combination, for improving ASR usability in the presence of low confidence scores occurring with dysarthric and other non-standard forms of speech: *input restriction* and *acoustic training*.

The first technique, *input restriction,* simplifies the recognition task by restricting the input to a limited number of isolated words, rather than continuous speech (Rosen & Yampolsky, 2000). ASR modules that are designed to accept this kind of input are sometimes referred to as *small vocabulary* systems. Rosengren et al. (1995) showed that adapting the vocabulary for each user improved accuracy rates from 28% to 62%. Hamidi et al. (2010) developed a speech module for use by people with severe dysarthria which was designed to accept 30 to 40 spoken keywords, each corresponding to an alphanumeric input or a system command (the mapping from spoken keyword to system command was tailored to the application domain). For each user, a list of keywords that he or she could say clearly was compiled during one-on-one sessions. Once the keyword list was defined, it was used to compose text and to navigate the World Wide Web (Spalteholz et al., 2008).

The other technique, *acoustic training,* entails to the process of recording speech samples from a user to create or to augment an acoustic database. Systems that use this technique are sometimes referred to as *speaker dependent.* A related technique, *adaptation,* is the process of refining the acoustic templates during use. Systems that use this technique are sometimes referred to as *speaker adaptive.* Persons with mild dysarthria have been able to achieve ASR accuracy levels similar to those of people without dysarthria after adaptation or acoustic training (Ferrier et al., 1995; Chen & Kostov, 1997; Hawley, 2002). Recent empirical evidence has shown that some speech errors perceived by human judges do not interfere with the recognition accuracy of speaker adaptive systems. Thus, after adaptation or training, these systems might actually perform better than humans at recognizing a user's dysarthric speech (Thomas-Stonell et al., 1998). For example, Fried-Oken (1985) found that, despite having unintelligible speech, people with limited use of hands due to brain and spinal cord injuries were able to use ASR for software interaction and preferred speech to other input modes and that these users were able to produce consistent speech that was recognizable by the computer after acoustic training. A related approach (Patel & Roy, 1998; Patel, 1999) has identified the potential of prosodic features of dysarthric speech, such as pitch, duration and intensity, as additional information useful for automatically recognizing speech. An ASR system trained on the prosodic information of a particular individual can then act as a "familiar facilitator" between them and an unfamiliar communication partner.

Acoustic training is useful for communication and control systems, such as speech-to-text modules for people with motor disabilities, where the goal of the system is to recognize as many input utterances as possible even if they are pronounced incorrectly or with a strong accent. However, this method is clearly unsuitable for therapeutic or training systems that aim to improve the user's speech over time. In these contexts, using this technique defeats the purpose of improving speech by awarding the users for repeating incorrect pronunciations.

4.5. Conclusion

There are a number of key observations that can be drawn from the previous section. The first observation is the strength of evidence that shows that children responded positively to computer language intervention and training programs (Bälter et al., 2005; Adams, 2006; Mitra et al., 2003). The qualitative feedback from the children was positive, providing evidence to support the hypothesis that computer-based games are appealing to children and will provide a useful tool to support and supplement learning and rehabilitation.

A second observation is that intervention-oriented systems have not matured. Most systems are prototypes that are not fully developed or evaluated. All the reviewed systems are less than a decade old. These systems have accomplished very little penetration to the intended user population (i.e., uptake by SLPs and/or clients). This is due to several reasons. Firstly, clinical uptake must be preceded by convincing demonstrations of efficacy, but evaluation of the systems is difficult and presents challenges (see section 7.1). Secondly, the multidisciplinary nature of the work means that researchers from multiple fields (e.g., computer science and speech language pathology) must collaborate, which in of itself is challenging and time consuming. The systems reviewed here, given their early stage of development, may yet be still in design and prototyping stages. Guidelines for the development and design of these systems have not yet been developed, and investigators often derive their own standards and design decisions at each step anew. In section 7, possible approaches for assessment of such systems will be examined.

A third observation is that all projects, especially *Box of Tricks* and ARTUR*,* faced the main challenge of designing effective feedback. This is directly related to the idea that there is a discrepancy between the original design goal of ASR modules (i.e., *recognition* of speech) and their perceived role in computer speech language systems (i.e., *analysis* and *assessmen*t of speech). By relying on ASR modules to provide detailed corrective feedback, the designers have risked providing the user with confusing and inconsistent feedback. Furthermore, abstract representations such as waveforms and closeness scores are unintuitive for children and have not been helpful in correcting speech.

Looking forward, one technique would be to use ASR to *engage* rather than *evaluate* speech, given the goal of facilitating sustained practice through the elicitation of multiple repetitions of target words and phrases. As demonstrated by Mitra (2003), it can be effective to subordinate the accuracy of ASR to its use as a facilitator and “encourager” of interaction. The most promising results were observed when the emphasis was on providing a context for multiple trials and opportunities to practice speech as opposed to providing complex corrective feedback.

5. Serious Game Design

Serious games are characterized as "games that do not have entertainment, enjoyment or fun as their primary purpose" (Chen & Michael, 2006). Zyda (2005) defines a serious game as: "a mental contest, played with a computer in accordance with specific rules, which uses entertainment to further government or corporate training, education, health, public policy, and strategic communication objectives." Thus, in addition to having characteristics most associated with games, such as being entertaining, fun and enjoyable, serious games also have other objectives such as facilitating learning and training, rehabilitation or artistic expression. Chen and Michael (2006) advocate the view that, through this shift in emphasis, games should become more responsible, relevant and important. The potential of games for training and learning has led to the relatively new area of research known as "game-based learning" (Prensky, 2000).

Thus, games that support speech language therapy, and therapeutic games in general, can be considered serious games. As mentioned in section 3 and 4, the use of games is prevalent in clinical speech intervention, as well as, computer intervention systems developed for children. An important point is that the essential gaming elements of playfulness and fun should be preserved and supported in the design of speech language intervention systems since these are what make them effective in the first place (Malone, 1982). This is the main motivator for incorporating elements of game play in therapeutic applications. In the following sections, I review previous research that demonstrates the effective use of games to support learning and rehabilitation. Next, two useful game characteristics, *uncertainty* and *a dynamic difficulty adjustment,* relevant to the proposed design are reviewed. Finally, the issue of gender and game design is briefly discussed.

5.1. Games Supporting Learning

There has been a great interest in the potential of using computer games for learning. For instance the main reason that a third of the American public high school students do not graduate was described as “boredom” rather than trouble with coursework (Bridgeland et al., 2006) and games are frequently identified as a means to increase interest in the classroom (Bergin, 1999). Some schools in the United States are experimenting with new innovative curricula that are based entirely on games (Corbett, 2010).

Löwgren and Stolterman (2004) recognize a “crucial difference between games and most work-oriented productivity programs”, in that, for the most part, users are motivated to play games not for extrinsic factors such as to perform or produce something but for intrinsic motives such as reaching a higher score, competing with fellow game players or solving a puzzle. Game interaction is a process of enticement, relationship, and fulfillment.

Games are also seen as a way to present problems in a way that decreases the fear and stigma around failure. Students behave differently when problem solving in game settings than traditional academic setting. Evidence points to the possibility that games weaken the stigma around failure. When problems are presented in an academic test, students generally prefer low levels of uncertainty and choose problems well below moderate challenge (Clifford, 1988; Harter, 1978). This choice, however, changes when problems are presented as games, in which case students will take greater risks than in other contexts such as the classroom or social interactions (Clifford & Chou, 1991). Thus, games can remove the stigma due to the implications of failure for social status and esteem that is hypothesized to deter individuals from tackling academic tasks with higher levels of uncertainty (Howard-Jones & Demetriou, 2009). In fact, recently, in a more radical view, game developers have recognized that "making failure fun" can have a beneficial effect on learning and remove the stigma and vulnerability of engaging in a learning situation (Corbett, 2010). There are many studies that examine the potential of gaming for learning and the impact of various factors such as gender and genre effects. For example, research has shown that playing action games that involve survival instincts, can improve visual-spatial thinking, which is an important skill for engineering tasks. This effect was observed only after ten hours of playing and was more apparent in women than men, with the effects almost eliminating the initial difference between them (Feng et al., 2007).

The aforementioned interest in computer gaming, combined with the increasing involvement of children with digital technology, point to computer games specifically as a promising tool to promote learning. According to the results of a national survey published by the Rideout et al. (2010), in the United States 60% of children between the ages of eight to eighteen use computer games on hand-held or console devices for an average of two hours on a typical day. In the last ten years, the percentage of children playing digital games has increased by more than 50%, and the amount of time they spend playing games has almost doubled. These numbers support the innovative schools' approach of capitalizing on this great interest of children in digital games and using it for learning.

A concern that is commonly-raised is that games may have detrimental effects on the players that are not offset by their positive effects. Griffiths (2005) has shown that many fears about links between computer games and the development of pathologies or serious health risks are unfounded. Given that the incidence of computer game addiction is itself very low (Griffiths &

Davies, 2005), it seems unlikely that even excessive use of games could lead to problematic gambling behavior in later life.

5.2. Games Supporting Rehabilitation

Rehabilitation is similar to intervention in that it often involves the performance of repetitive tasks that target specific areas for improvement. A large amount of practice over a long period of time is most often required to achieve results in the therapy.

Research has shown that rehabilitation results improve when patients are motivated (Maclean et al., 2002). Demotivation that can lead to resignation is often experienced during rehabilitation due to lack of immediate increase in the user's capabilities (Burke et al., 2009). It has been shown that during rehabilitation sessions, games increased motivation which, lead to patients finishing more exercises and achieving better results (Rego et al., 2010).

By focusing their attention on the game, patients may forget they are involved in rehabilitation (Flores et al., 2008). Further, games distract the patient's attention and can aid in pain management (Burke et al., 2009).

In the light of these possibilities and potentials many rehabilitative game have been developed in recent years (Alankus et al., 2010; Flores et al., 2008). The fact that games create motivation in the context of rehabilitation confirms the idea that they will be motivating in the context of intervention.

5.3. Games Supporting Language Intervention

Computer games have been used as the basis of interventions targeting the areas of communication, social pragmatics and learning, in addition to intervention targeting speech production (see section 4.1). In this section, two projects are briefly presented that demonstrate the application of games to other intervention target areas: autism and learning disability. These show the versatility of computer games and their applicability to interventions that target areas ranging from communication disabilities to various cognitive impairments.

Previous work has identified the potential for computer games to be used to supplement clinical interventions for children with varying degrees of autism, in particular to develop communication skills (Burns, 2003). More concretely, Cockburn et al. (2008) found that the use of games, even for short amounts of time, can help children with autism develop facial expression and face recognition skills. Tanaka et al. (2010) implemented a game that motivated children with autism to practice facial expressions, such as smiling and frowning that used the technique of computerized facial recognition. The children were motivated to practice facial expression, since these were needed in order to progress in the game.

The Arrowsmith Program is designed to alleviate learning disabilities through exercises and games (Arrowsmith, 2007). The conjecture behind the approach is that problems in high level functions, such as writing and copying, have their underlying root in impairments in specific elemental cognitive abilities, such as symbol recognition and sequencing (Young & Burrill, 1997), and that games and exercises that target and strengthen these underlying skills will also improve high level functions and learning more generally. The first step of the program involves a detailed evaluation and assessment process that identifies elemental cognitive impairments for each student. This is followed by an individualized task-oriented program that challenges and alleviates the identified deficit through playing customized games.

A number of reports examined the program's effect in select schools in the *Toronto District School Board* (TDSB) where it was introduced. They have documented its positive impact on the student's learning abilities and have also observed that students using the program needed fewer resources (Arrowsmith, 1998; Arrowsmith, 2007). But the main conjecture of the approach was not supported with empirical data.

5.4. Uncertainty, Challenge and Game Design

What are the characteristics of games that make them engaging and motivating? Previous research has identified numerous elements. For example, Malone (1982) hypothesized the reason games are engaging is because they involve elements such as fantasy, curiosity and challenge. Alankus et al. (2010) also found aesthetic appeal important for motivation. Many of these characteristics can be implicit in the design of a game. This section will focus on two game characteristics that contribute to engagement: *uncertainty* and a *dynamic difficulty adjustment*.

5.4.1. Uncertainty

The first game characteristic is uncertainty. During game play, the presence of uncertainty increases motivation through the building of anticipation. Uncertainty is known to provoke emotional responses, such as surprise, into game play (Owen, 2005). Historically, various means such as the use of die throws have been used to simulate randomness and thus introduce uncertainty into game play. For example, the game of *Backgammon* makes use of the dice, which allows the players to combine their skills with an element of chance.

Howard-Jones and Demetriou (2009) conducted three studies examining the effect of uncertainty on games and learning. The first study tested whether the attractiveness of uncertainty survives the pressure imposed by evaluation in a learning game. In the study, fifty pupils (age eleven years old) were asked to play a computer mathematical quiz. Each individual played the game by him- or herself. Before seeing a question, the subject had the choice between having one of two characters, Mr. Certain and Mr. Uncertain, evaluate their answer. The evaluations were the same if they answered the question incorrectly, in which case, they would be given a 0 score. If they answered correctly, Mr. Certain would award them a score of 1 which Mr. Uncertain would award them a score of either 0 or 2. In effect, this approach allowed students to choose the presence of uncertainty in the game. The subjects showed a significant preference for uncertainty.

The second experiment examined the perception of fairness and just reward when uncertainty is present in games. Twenty participants (16 years old) were divided into pairs, with each pair competing in a quiz game against the computer. At each step of the game, two animated dice were rolled and a question displayed. The combined value of the dices could be won if the question was answered correctly. If the question was answered correctly, the players had a choice. They could roll the dice again or pass it to the virtual opponent. If they rolled again and both dice showed ones, then all the accumulated points for the round would be lost. This choice clearly posed a risk. The team that reached 100 first was the winner. The conversation of the teams during the game was recorded and analyzed later. The analysis showed that the players were highly engaged and the uncertainty did not deter them. They used words such as “annoying” and “fun” a lot. Howard-Jones and Demetriou (2009) showed that in a gaming environment, traditional notions of fairness can be subverted in a way that engages children emotionally and might be better for learning. The issue of reward inconsistency was shown not to be a problem when introduced in the context of games.

Howard-Jones and Demetriou's (2009) third experiment was designed to examine the relationship between emotional response and uncertainty in the context of a learning game. The study used *electrodermal activity* (EDA) as a measure of emotional engagement. EDA measures the electrical conductance of skin, which varies in accordance with the activity of sweat glands that are in turned controlled by the sympathetic nervous system and respond to emotional arousal (Martini & Batholomew, 2003). Sixteen adults played a computer game and had their EDA measured during play. The two conditions both involved playing a quiz game. In one condition, an element of uncertainty was introduced via the possibility of making a random choice (i.e., a die) whereas in the second condition the random choice option was not available. An increase in emotional response was observed in the condition with uncertainty when compared to the condition without uncertainty. The increased emotional response to the element of uncertainty was observed even once educational material was introduced to the conditions. The researchers conclude that the presence of uncertainty can transform the emotional experience of learning and make it more engaging. Further, they noted that uncertainty increases engagement with the whole experience including learning, rather than just the game. This is in accord with other research that has shown declarative memory formation and memory for items encoded in a positive emotional context is improved, compared to neutral or negative contexts (Erk et al. 2003; Brierley et al., 2007).

The benefits of uncertainty dovetail nicely with the shortcomings of ASR, as discussed in sections 4. The use of ASR entails errors in recognition, but these errors might be incorporated and disguised as elements of uncertainty in the proposed system. Uncertainty can both add a motivating and engaging element to an interactive system and also provide a solution to the problem of recognition errors.

5.4.2. Dynamic Difficulty Adjustment

A second game characteristic is *dynamic difficulty adjustment*, a method by which the challenge level of the game is changed dynamically during game play, in order to avoid both user boredom (e.g., if the game is too easy) and frustration (i.e., if the game is too hard) (Hunicke, 2005). Research has shown that in non-human primates, the maximum dopamine is released when the likelihood of receiving reward for success is about 50% likely, in other words when the probability of success is half way between totally unexpected and completely predictable (Fiorillo et al., 2003)[1]. Several studies on rehabilitative games have shown that an appropriate level of challenge, maintained by automatic difficulty adjustments, is an important element in keeping the user motivated (Burke et al., 2009; Alankus et al., 2010).

5.5. Games and Gender

Research has demonstrated that some differences in user experience are correlated to gender. It has been noted that, among children, boys play in larger groups than girls and that girls often are (slightly) ahead of boys in their ability to initiate fantasy play without realistic props (Hughes, 1991). Further gender differences were also observed in collaborative interaction — boys take, and girls relinquish control of the mouse (Inkpen et al., 1995).

---

[1] Motivation to pursue a variety of pleasures, including sex, food, gambling and computer gaming is linked to dopamine levels in the human brain (Elliot et al., 2000; Koepp et al., 1998).

Fails et al. (2005) conducted a study in which two versions of a collaborative content-infused story game, *the Hazard Room,* were tested with separate groups of young girls and boys (4-6 years old). One version of the game used a desktop interface and the other one used a tangible interface. While the small-scale comparison revealed no statistical differences with respect to learning between the two versions, interesting differences were observed between the interaction of boys and girls with the system. The researchers observed that boys tended to point and touch more than girls in both environments. When questioned about the stories after the interaction, the girls were more creative in narrative processing and used rewording rather than rote recitation of stories. Both groups preferred the tangible interface. However, girls seemed more interested in interacting with the game in general and exhibited more patience. The contrast in being patient and following the narrative was more evident with the desktop version of the game (Fails et al., 2005).

A designer bias has been observed towards male users and a corresponding interest imbalance on the part of game designers towards designing games for males (Cassell & Jenkins, 1998). This imbalance has been addressed by games specifically designed for girls. The question of gender is important and should be taken into account in the design of the proposed system.

6. Tangible User Interfaces (TUIs)

6.1. Overview

Tangible user interfaces (TUIs), also known as *physical manipulatives*, *3D manipulatives* or *digital manipulatives* (Manches et al., 2009; Zuckerman et al., 2005), are a relatively new set of computer interfaces that transcend the traditional desktop and screen-based system. TUIs are characterized by embedded electronic microcontrollers, sensors and actuators in either existing or newly-designed physical objects (Ishii, 2008). The diversity of TUIs and the range of possibilities they provide means they have great potential for future research and development (Ishii & Ullmer, 1997; Jacob et al., 2008). These interfaces are becoming more versatile as the novel and improved computational materials are developed, such as digitally-enhanced paper and fabrics (Buechley et al., 2008; Buechley et al., 2009). For example, Buechley et al. (2008) have developed sewable and washable electronic components that can be sewn into fabric.

There are two main categories for TUIs. In the first are the TUIs that *extend or augment the traditional desktop computer configuration* (i.e., by supplementing or replacing its existing interface). In the second category are *the stand-alone computational devices*, which embed the computation so that the computer becomes an invisible platform background. Weiser (1999) foresaw this latter category as a natural direction for computing and believed that computers "will weave themselves into the fabric of everyday life until they are indistinguishable from it". For instance, the "thrifty faucet" (Togler et al., 2009) monitors the usage patterns of a shower faucet and provides statements pertaining to water use to household members.

The study of the relationship between tangibility and design has a long history. For example, Baskinger and Gross (2010) pointed out that “the hand axe reminds us that design has always been about interaction and interaction has always been tangible.” With the advent of new embedded electronics, computation continues to migrate more than ever into physical objects, especially those that are not traditionally computational; this serves to delineate a novel class of computational objects altogether, thus giving weight to the idea that “tangible interaction is the physical embodiment of computation” (Baskinger & Gross, 2010). As a practical application, TUIs can be designed so that appearance expresses function (Redström, 2008); this approach underlies *reality-based interaction* design and will be discussed in section 6.2.3. In this approach, the fitting of the interaction to the user’s perceptual and motor capabilities is seen as the main means to improve usability. The very recent emergence of these interfaces means that research into the potential uses is still in its infancy. It is a great opportunity for TUI designers to set thoughtful examples for the future generations of these interfaces.

This section addresses the question of the essential characteristics of TUIs that support their use for therapeutic and learning applications. Specifically, what are the characteristics that support their use in the proposed system? The three key characteristics are:

- Support for collaboration through shared space and situatedness
- Affordance for the engagement of movement
- Flexibility in the design of the physical interface of the system

In the following subsections, these characteristics of TUIs and their applications in relevant areas will be identified and reviewed.

6.2. Situatedness and Shared Space

TUIs allow electronics to be embedded in one or more physical objects, which can then interact with the user and each other TUIs. In the latter case, TUIs can be considered in terms of an assemblage that is manipulated by human agents or as an assemblage that reacts to environmental or other input. This leads to an important characteristic: *situatedness*. TUIs can be moved and placed in relation to each other and thus are suitable for geographically distributed and collaborative interactions. Their physical presence can provide real-time feedback about the state of the system, creating a *sensor network user interface* (SNUI) (Merrill, 2007). Since the physicality of TUIs allows changes in their location and configuration to be immediately visible to the user, they can be used to increase legibility and seamlessness between people and information (Ishii, 1997; Ishii et al., 2008). This allows for the interpretation of their order and relative position with respect to each other and the physical environment, thus making them even more articulate than centralized interfaces such as desktop systems (Blackwell & Edge, 2009).

Siftables (Merrill, 2007) is an example of a system that utilizes this characteristic of TUIs to implement a novel form of interaction. *Siftables* consist of a number of small tangible units that each encapsulates an LCD display, embedded sensors, a microcontroller and a wireless transceiver. Each unit senses the other units in its proximity and can communicate with them or a central computer unit wirelessly. The collection of units implements a sensor network and is designed to allow the users to apply their physical manipulation skills (i.e., sorting and sifting) to digital objects. While, the researchers have developed this system as a general interface for application in various contexts, they have provided an example application in the form of a photo sorting interface where pictures are represented by images on siftable units and can be grouped or sorted by grouping or sorting the units.

TUIs support collaboration and interaction in a shared space, both when users utilize the interface for the same system and when they each have their own TUI (Fernaeus & Tholander, 2005; Suzuki & Kato, 1995). TUIs allow users to monitor each other's gaze to achieve interaction more easily than when interacting with a graphical representation on a display (Suzuki & Kato, 1995). They can also increase the visibility of other members' activity, better communicating the current state of their work to their peers (Stanton et al., 2001; Suzuki & Kato, 1995; Fernaeus & Tholander, 2006). TUIs often allow concurrent interaction, sharing control between collaborating learners (Zuckerman et al., 2005). Previous research has shown that computational technology can support new forms of collaboration and encourage creativity (Hamidi & Baljko, 2010).

In the context of collaborative activity, TUIs provide the system designer with novel physical conditions to control the interaction pace and the effort required from the user. Some factors that can be used to increase the effort needed to finish a task include increasing the size of TUIs and the complexity of the manipulations in order to increase the effort needed. The pace of the interaction can be slowed down by using props and obstacles. Using these factors effectively allows for more social engagement opportunities during the interaction and more time for reflection and communication (Stanton et al., 2006).

6.3. Flexibility in Design

TUIs are also characterized by accessibility. They are more accessible than traditional desktop systems to young children, people with learning disabilities or novices (Zuckerman et al., 2005). TUIs have a lower threshold of participation than that typically required by a traditional desktop system (Hornecker and Buur, 2006). It is possible to design TUIs such that interaction is more natural or familiar than with other types of interfaces (Jacob et al., 2002; Dourish, 2001). TUIs have been used to develop engaging interfaces for very young children (Zuckerman et al., 2005). For example, Martinussen et al. (2007) developed a TUI for small children to access and play media. In this system, dolls were embedded with *radio-frequency identification* (RFID) tags to represent animation clips that were then played when the child placed them in a bowl equipped with an RFID tag reader. In this project, the researchers explored the effect of lowering the threshold of access and control over digital media and observed that providing such playful control affords more physical activity around media usage as well as more dynamic and active interaction with the media.

Barriers to interaction were also addressed by Bhat (2010), who developed a tangible musical instrument for children with cerebral palsy who had poor bimanual coordination. This instrument was designed to help develop musical ability, bimanual coordination and increase social participation for these children.

6.4. Engagement of Movement

TUIs afford the engagement of movement and physical interaction with objects, a key characteristic that will be utilized in the proposed system. A defining characteristic of TUIs is their *tangibility*. TUIs incorporate sensors, which allow them to recognize pressure, heat, light, speech, proximity to other objects among other phenomena. TUIs also incorporate actuators, which allow them to create movement, sound and visual feedback that can be used to communicate system information to the user. In addition, other actuators that can create other phenomena such as heat and moisture are emerging. Thus, TUIs have responsive physicality.

The *physicality* of TUIs allow for designers to incorporate feedback and cues in the interaction that engage haptic and proprioceptive senses. These are especially useful for promoting the users' understanding of three-dimensional and physical forms and hypothesized to provide an advantage over visual representations alone (Marshall, 2007). This is especially relevant for learning applications in domains such as molecular biology, chemistry and dynamic systems where the concepts are inherently spatial, either literally or metaphorically (Marshall, 2007). TUIs afford concrete physical manipulation, which supports situated learning in which objects in the everyday environment are augmented and used for teaching (Marshall, 2007; Klemmer et al., 2006).

TUI design at present has mostly focused on using tangibles in expressing spatiality and metaphorical spatial structures (e.g. narrative) (Marshall, 2007). The potential of other dimensions of use such as the mass, texture, temperature or malleability, has yet to be explored and utilized. The characteristics that have been explored show the great potential of TUIs for different applications.

The *tactile-kinesthetic* mode of learning is a form of learning that "comes through touching and physical sensation and is anchored by movement and touch" (Silverman, 2006). TUIs support for movement and tactile interactions, provides them with great potential to support this form of learning (Revelle et al., 2005). This is important in the light of the wide recognition of the diversity of intelligences and their effect on learning that has prompted research into novel teaching methods (e.g., the *tactile-kinesthetic* mode) and technology that can support them (Snyder, 2000).

6.4.1. Movement and Learning

Movement and learning have a deep connection and interdependence. In this section, the following questions are discussed: what is the nature of the evidence that supports this point and how do TUIs, given their tangibility and physicality, support this connection?

Automaticity, which is defined as a virtuosity in action that is reached by high degrees of practice (Freeman & Nunez, 2000), is central to gross bodily movements (i.e., coordinated movement needed for tasks such as walking or jumping), fine motor tasks (i.e., coordinated movement needed for tasks such as writing) and speech. In the philosophy of psychology, *embodiment theory* holds that automaticity is necessarily anchored in motor learning (Freeman & Nunez, 2000). This theory holds that bodily experiences play an integral role in the formation of human cognition and that cognition forms through the ways the body moves and interacts with its environment.

Previous research has shown that learning memory is improved when movements are aligned with what is to be remembered. Actors have been shown to better recall, after five months monologues during which they moved on stage rather than ones where they were standing on stage (Noice et al., 2000).

The importance of interaction with physical toys or objects that involves the use of touch and movement has long been recognized in education and development theory. A comprehensive discussion of research in these areas is out of the scope of this document, but at a minimum it is noted that Jean Piaget, a central figure in the area of child development research, held the view that development and movement are closely related and believed in the notion that in infancy intelligence is action (Flavell, 1963). Additionally, a core axiom of the Montessori Method, a popular and influential educational method, is that children build a collection of impressions about the physical world through movement, exploration, and manipulation (Lillard, 2007). In this view, the child's hands and developing brain are closely related and essential learning is achieved through deep concentration which is more available through working with hands and physical material (Lillard, 2007).

As mentioned previously, in recent years, these observations have led to the recognition of the *tactile-kinesthetic* mode of learning that engages movement to facilitate learning (Silverman, 2006). TUIs are seen as natural tools to support this mode of learning (Zuckerman et al., 2005).

TUIs engage children in playful learning (Price et al., 2003). For example, Rogers et al. (2002) conducted a study in which children mixed colors either using digital tools, physical tools or a combination of both digital and physical tools. They found mixing physical action and digital effects leads to increased engagement and reflection. In another project, Brotto Furtado et al. (2008) used the combination of a vibrating pen and a physical talking agent to allow children work on an activities notebook that focuses on learning material.

6.4.2. Movement, Speech and Language

There is evidence that movement and speech are intimately connected. What is the nature of the evidence that supports this view? What evidence, if any, is there that this enhances the speech intervention process? In this section, answers to these questions are discussed.

Iverson and Thelen (1999), in their synthesis of research from cognitive science and gesture research, inventoried neurological evidence that demonstrates that speech and movement have a deep connection and interdependence. First, they identified the evidence that shows that some speech and motor functions share underlying mechanisms in brain regions. For example, speech production and motor sequencing both use precise timing (Ojemann, 1984). Also, there is strong evidence that the hands and arms and the vocal tract are represented in neighboring sites in brain regions (Fried et al., 1991). Second, brain regions typically associated with motor functions (e.g., motor cortex, premotor area, cerebellum) are activated when speech is produced. When motor tasks are performed, brain regions that are recruited for language and speech (e.g., Broca's area) are activated (Erhard et al., 1996). Finally, patterns of breakdown and recovery in certain speech and motor functions are closely linked. For instance, impairment in non-speech motor sequencing may be affected by speech disorders. Studies of children with *Specific Language Impairment* (SLI) showed that these children's motor movements were also negatively impacted and that these children fell under the standard and matched *Developmental Coordination Disorder* (DCD) levels (Hill, 1998). SLI is a speech impairment of unknown causes. DCD is a childhood disorder characterized by poor coordination and clumsiness.

Results from gesture research also point to a close relationship between speech and movement (Iverson & Thelen, 1999). Gesture production can improve speech in individuals with aphasia, an effect that might be due to the simulation of relevant parts of the brain (Hanlon et al., 1990). Research has shown that teaching gestures to infants accelerates their language learning abilities: Goodwyn et al. (2000) conducted a study in which a group of infants were taught a small set of communicative gestures and encouraged to use them in daily interactions. Their longitudinal study showed that these infants attained milestones, such as the first word or symbol acquisition and the five-symbol, approximately one month earlier than other children. These results are confirmed in infants who learn *American Sign Language* (ASL) (Folven & Bonvillian, 1991). Thal and Tobias (1992) found that a single test of communicative gesture production is a reliable predictive test that can distinguish between truly delayed late talkers (i.e., ones who stayed delayed after a year) and late bloomers (i.e., ones who started talking within a year). In this experiment, it was found that late talkers who used more gestures were more likely to start talking after one year.

6.4.3. TUI-based Speech Intervention

As of yet, very few systems employ TUIs as the basis for language and speech intervention. For instance, Hengeveld et al. (2009) developed *LinguaBytes*, a set of digitally-augmented dolls, pictogram cards, and puzzle pieces. These TUIs were designed for use by toddlers (between 1 and 4 years old) with multiple disabilities, such as cognitive and perceptual-motor disabilities and were intended to aid with therapist-supported language and communication skills learning.

Figure 6. ***E-scope's*** **interface (Hummels et al., 2006)**

The goal was to devise a set of activities that would help children expand their vocabulary and to elicit a greater amount of communication between the child and others. The TUI tool engaged children in interactive storytelling and puzzle-solving games. Children interacted with the system by placing digitally augmented tangible items in the form of enhanced dolls, cards labeled with pictograms and puzzle pieces on an augmented play surface. The interaction involved the narration of a story. The story was paused periodically after the mention of an object or character, prompting the children to place corresponding items on the surface in order for the story to continue. These items were each embedded with *radio-frequency identification* (RFID) tags which were recognized by RFID tag-readers embedded in the play surface. With any given configuration of items, the system would narrate a corresponding story through a screen and speakers. The tangible items were created to correspond to keywords in the stories. Parents and therapists were provided with additional tags, so that they could create new embedded objects, such as the child's own plush toy. In evaluations of the TUI, Hengeveld et al. (2009) observed that the children had longer attention spans and that the TUIs themselves slowed down interaction, which provided more control to the children over the timing of the material. The researchers concluded that the TUI afforded a more natural and accessible interaction for the children.

Figure 7. *LinguaBytes*' interface (Hengeveld et al., 2009)

Another tangible interface was developed for interactive storytelling for the same user population (Hummels et al., 2006). *Explorascope (E-scope)* consists of a wooden toy in the shape of two rings, one non-moving ring on the bottom and a rotatable one on top, that is wirelessly connected to a computer with a screen. The toy was designed to implement an interactive story telling activity for use in the context of communication and language therapy. It could be used in different ways. One method was to move it over tagged images placed on the floor. Motions would activate a story line corresponding to the chosen images. Another method was to rotate its upper ring that moved over images placed under the cover of the toy. To choose a specific image or symbol, the user had to rotate the upper ring until the image was visible through an opening in the cover. A prototype of the system was developed and a preliminary experiment was conducted with three children with multiple disabilities. Observations from the experiment were overall positive and showed that two of the three children were motivated to interact with the toy. The therapists who were present at the experiment believed the approach was promising because of its adaptability, playful approach and integration of multimedia material.

The flexibility of TUIs has prompted their use as accessible interfaces for children with disabilities. One of the design issues for assistive technology is the diversity of user needs. The standard desktop input and output devices, such as the mouse, keyboard and monitor, pose many challenges for users with disabilities, such as the need for fine motor skills and good vision. It is easy to see the appeal and potential of developing TUIs that are easier to use by this population. In fact, many technological alternative input methods such as single and puff switches share the aim of transcending the conventional desktop paradigm with TUIs.

For assistive applications, the flexibility of TUIs not only allows for a more diverse user population but also allows for the development of general customizable frameworks that can be customized based on the scenario of use. Individualized education programs for people with disabilities are long recognized as good examples in which an awareness of the ecology of the particular user with respect to a team solution is used to address the complexity of the individual requirements (Chang & West, 2006; Rainforth & York-Barr, 1993). This approach is previously recommended for language intervention systems (Eriksson et al., 2005) where it is desirable to provide the SLP with a general customizable framework that he or she can adapt based on the needs of each client.

6.4.4. Movement and Engagement

Next, I examine how the engagement of movement can support an increased sense of *presence,* a sense related to the level of engagement in an experience (Steuer, 1992). Specifically, the support of this sense by TUIs will be examined.

Results from *virtual reality* (VR) research literature have demonstrated the importance of bodily movement in the user's experience of *presence* and *telepresence* (Slater et al., 1998; Slater & Usoh, 1994), where *presence* is defined as the sense of the user's experience of being in an environment through technology (Gibson, 1979), and *telepresence* is defined as presence through a communication medium (Steuer, 1992). Slater et al. (1998) found a significantly positive correlation between body movement and presence using an experimental design with movement and no-movement conditions. The degree of presence was measured using standardized questionnaires. VR research aims to identify factors that contribute to and develop systems that afford presence and telepresence. Chief among these factors are vividness, system response speed, range of input channels and the resemblance of mappings to real world counterparts. In the context of VR, *vividness*, defined as the depth and breadth of sensory perception, is identified as a central factor that affects presence (Steuer, 1992). *Depth* is defined as the quality or resolution of the sensory channels and, *breadth* is defined as the diversity of sensory channels that are engaged.

The subjective experience of presence in virtual worlds may bear some similarity to the feeling of engagement in non-virtual words that include computational objects. Factors that impact the former may also be relevant for the latter. Presence is a subjective experience and so far has been typically measured using self-reporting, via standardized questionnaires, a method open to interpretation and not as precise as objective measurements (Slater et al., 1998).

VR applications aim to utilize aspects of human perceptual systems that are optimized for interactions with the real world and make use of mappings that are adapted and modeled on the human body and its real world environment. This goal is similar to the *reality-based interaction* (RBI) approach as implemented by TUIs, where the fitting of the interaction to the user's perceptual and motor capabilities is seen as the main means to improve usability.

RBI is an approach in which interaction is based on real life approaches and capitalizes on people's extant knowledge of physics, body awareness and skills, as well as environmental and social awareness and skills (Jacob et al., 2008). An important motivation for this is that "basing interaction on pre-existing real world knowledge and skills may reduce the mental effort required to operate a system because users already possess the skills needed" (Zukerman et al., 2005). In other words, this approach aims to create natural and intuitive interactions based on the user's experience in the physical world. This approach has been used for the augmentation of everyday objects, for example "musicBottles" (Ishii et al., 2001), in which different musical sounds are stored in a set of augmented physical bottles and can then be played back by opening the bottles, and "I/O Brush" (Ryokai et al., 2004), in which an augmented brush allows young children to "pick up" colors and patterns from objects and apply them to a digital canvas. These interfaces explore new functionality and artistic expression through computational interfaces.

A conjecture of TUIs is their support of *empathy*, in the sense of engendering an emotional connection with a virtual persona or character. For instance, Johnson et al. (1999) developed a TUI in the form of a plush toy, which had a virtual counterpart on a computer screen. When the children moved the plush toy, its movements, captured by gesture recognition software, were reflected on the virtual counterpart. The researchers hypothesized that touching and moving a physical version of the virtual character helped children relate to the character more closely. The quality of empathy is often discussed in the context of *sympathetic interfaces,* which are characterized as interfaces that provide a correspondence between physical manipulation and digital effects. These interfaces are hypothesized to be helpful by interpreting the user's actions given the context, as well as being inviting and friendly (Johnson et al., 1999). In previous research, the importance of interactive systems that incorporate emotional as well as functional or cognitive elements into the interaction is recognized and encouraged (Picard, 2000).

6.5. Design Approaches Relevant to TUIs

The novel characteristics of TUIs have given rise to introspection into the processes whereby computational media are designed. User-centered and usability design arose in tandem with the increasing prevalence of GUI-based interfaces in traditional desktop computing. Presently, renewed design approaches, such as *reflective design* (Sengers et al., 2005), *thoughtful interaction design* (Löwgren & Stolterman, 2004), and *ludic design* (Gaver et al., 2004) are emerging and explicitly seek to apply critical thinking to the design process and resultant computational media..

Dunne (1999) observes that computer technology has matured to a point where it is possible to apply art concepts such as *provocation*, *ambiguity* and *strangeness* to explore the aesthetic potential of computational artifacts. TUIs have been used in areas such as cinema or visual arts to provide a depth of experience to the user and to invite reflection of one's own subjective experience into encounters with art works. The two aspects to electronic objects stand in tension:

the solidity of their physicality and the fluidity of electronic behavior. Design will thus link these two aspects: the material and immaterial. By designing objects in such a way as to simulate or project human psychological desires and needs, it becomes possible for the user to experience new narrative situations in which the object has an active role. This approach is already realized in projects such as *eyeLook* and *Connectibles.* For example, in the *eyeLook* project, eye contact sensing is used to enable mobile devices to be used as interfaces that act on innate human nonverbal turn taking cues (Dickie et al., 2005). In *Connectibles,* computational objects are used as interactive gifts to establish social connection between members in a social network (Kalanithi, 2007). Thus, TUIs can support social values, such as gift sharing and the use of social conventions such as eye contact for interaction (Jacob et al., 2008).

Dunne (1999) advances the argument that designers should go beyond relentless innovation for its own sake and should apply critical thinking to their designs. Ideas and values embedded in electronic objects are not immutable but man-made. By applying their knowledge and worldview to the design of electronic artifacts, designers, in effect, embed their conscious or unconscious values into their creations. The electronic object is a crystallization of the designer's vision: a vision formed by his or her particular point of view and value system. That TUIs combine design and technology means that social and aesthetic values can be embodied into electronic objects through their integration into design. According to activity theory, consciousness is realized by what we do in practical everyday life in relation to people and artifacts (Kaptelinin & Nardi, 2006). Tools are viewed as mediators between people and the world. Thus, the intentionality hidden in the design of artifacts becomes an important subject of analysis.

Reflective design, another similar approach, emphasizes reflection on the unconscious values that are embedded in technology (Sengers et al., 2005). In this method, the user is engaged such that he or she can bring unconscious values to the surface so that they are available for conscious choice. Sengers et al. (2005) provide several strategies to achieve this goal including giving the users license to participate in the interpretation of the design, using digital scaffolding by gradually moving from the familiar to the unfamiliar, introducing novel parts of interaction slowly, providing dynamic feedback that encourages reflection and inspiring rich feedback from the user.

Löwgren and Stolterman (2004) propose *thoughtful interaction design* as a methodology to enhance interaction design through the application of techniques from disciplines such as architecture and industrial design. Thoughtfulness is needed in interaction design because digital artifacts affect people's lives and unintended side effects might result in unforeseen problems and issues. In this approach, the "good" in the design has to be defined with respect to a situation and continuously reevaluated with respect to that situation. Two relevant desirable qualities proposed for digital artifacts in this approach are *tight coupling*, which refers to a minimized distance between the user's intentions, actions and the effects of actions, and *pliability*, which refers to the quality of the material being responsive to the user's actions to the point of being tactile. Both these qualities are supported by tangible interfaces.

Another relevant design approach is *ludic design* that focuses on designing for people as playful creatures. This approach designs for curiosity and exploration of new values (Gaver et al., 2004). Picard (2000) argues that focus on cognition has inadvertently led to a discounting of emotion in interaction, which should be countered by methods that bring sensitivity to user emotions in interface design. This is especially important for learning as it is shown in previous research that memory for items encoded in a positive emotional context is improved, compared to neutral or negative contexts (Erk et al. 2003). Furthermore, declarative memory formation is found to be enhanced in emotional contexts (Brierley et al., 2007). Thus, learning might be enhanced by incorporating ludic elements into the design.

The design approaches outlined above provide valuable ways for the designers of TUIs to explore the design space and make decisions pertaining to the affordances of the system and quality of interaction.

7. Assessment

An assessment of an intervention system may address several questions: what is the relative efficacy (i.e., in terms of significant therapeutic effects caused by it and how do they compare with other extant interventions) of the intervention? Are there any adverse effects? What are the effects of the longitudinal change on the intervention effects? What data concerns the big picture that captures the complexities of the learning process? When there are several aspects present in the intervention, how to evaluate the effect of each aspect separately and compare it with alternative approaches?

There are two dependent variables of interest in evaluating learning and therapeutic systems. *Knowledge retention* refers to how well learned concepts are retained. In this case, the retention of the learned pronunciations after interaction with the system should be evaluated. *Generalizability* refers to whether learned concepts are applicable to everyday situations outside the context of learning or intervention. In the case of the proposed system, it should be examined whether the usage of the learned pronunciations are limited to the context of therapy or generalized to everyday speech.

With respect to the above questions, there are several challenges in the evaluation of user interfaces and more generally, systems designed to achieve clinical objectives for speech intervention. These challenges involve both methodological issues, which concern the design of experiments suitable for the measurement of specific effects in terms of quality and quantity, as well as, logistical issues, that concern the selection of participants as well as facilitating participation.

In this section, various evaluation methodologies that exist will first be examined. The challenges of this application domain and the shortcomings of common evaluation methodologies will be examined. Alternative methodologies that have been used successfully to evaluate similar therapeutic and training systems are discussed. This is followed by an in-depth case study of the evaluation of *Fast ForWord,* a popular computer language intervention system.

7.1. Challenges

Traditional experimental design is a well-known method for testing hypotheses about properties of user interfaces. This method is oftentimes used to examine hypotheses that concern psychological phenomena or medical interventions, but also could concern the use of software systems. Experimental design usually involves using control and treatment groups that are exposed to conditions that differ from each other only in the factors that are under study. *Randomized Controlled Trials* (RCTs) are a popular form of experimental design in which the subjects are randomly assigned to one or more conditions, whose effects are then compared after the conclusion of the experimental trials (Pelham & Blanton, 2007). There are different variations of the method that, for example, involve using different groups of participants for each condition (known as a *between-subjects* design) or using the same group of participants for every condition (known as a *within-subjects* design). Various measurements are taken under each condition and compared, resulting in the acceptance or rejection of a hypothesis regarding the effects of applying the specific conditions (Pelham & Blanton, 2007). In user interface design, the hypotheses are formed in terms of predictions about various properties, such as ease of use, speed and accuracy that the system may afford during interaction. In the case of speech intervention, the hypotheses concern predictions about various improvements of speech, as measured by standard speech tests, following the intervention and maintained in the period of time following withdrawal of the "treatment".

In the evaluation of the efficacy of complex systems such as one that uses both ASR and a TUI, an important issue is disentangling the impacts of each of these components on the overall intervention. The evaluation of an intervention as a sum of its parts is difficult. This difficulty, for instance, is seen as the cause of a shortage of experimental research comparing the actual benefits of TUIs and equivalent graphical representations (Marshall, 2007). As well, it is known that in the absence of any formal intervention, there is still some skill acquisition due to naturally-occurring cognitive development and maturation. A major advantage of the RCT methodology is the simplification: if variations of the intervention are created such that they only vary in only the aspects under study.

The shortcomings of the RCT methodology for evaluating interfaces developed for people with disabilities have been recognized previously (Lazar et al., 2010). The main reason is that research in this realm is often exploratory and is "a hybrid of quantitative and qualitative research or primarily qualitative" (Lazar et al., 2010). Experimental design requires the control of the experimental conditions, which is difficult to accomplish in a target population of people with disabilities. The diversity of the population is often connected to confounding factors. Many members of this population have multiple forms of physical and mental disabilities that differ from person to person. Also, it is often the case that many of the subjects have had exposure to various other forms of intervention and this may have an effect on their interaction with the system under study. In addition, subjects have changing health and living conditions, any of which may affect the experimental conditions.

An alternative design is *small-n design* in which conditions as applied to a small number of individuals are studied closely. Rather than using one or more groups of subjects, this approach focuses on a small number of participants who are representative of the user population. These studies provide more detailed information about the participants and their interactions. In assistive technology, where research is oftentimes exploratory and using a small number of participants at the early stages of research is a priority, this type of study is common practice (Lazar et al., 2010). In the presence of significant results, the researchers can validate the results of a small-n study for larger groups of users by replicating the experiments for larger groups of participants (Taber, 2000). In the absence of definite results and where mixed results are observed, the research can perform a Power Analysis to decide the sample size that would yield a definitive answer to whether the effects are significant or not (Fails et al., 2005).

In the case of speech intervention systems, there are additional factors that contribute to the complexity of the assessment. An important target user population is children, which means the experiments will involve additional logistical efforts. For instance, these subjects require accompaniment and supervision. Also, informed consent is required, which must be obtained from the child's legal guardian and who might be biased against or towards different types of interventions (possibly impacting the experimenter's ability to obtain consent). There are specific difficulties that arise with assessing therapeutic and learning effects, specifically the determining length of the duration that is needed over which significant results, if produced, may be reasonably observed. There are also challenges in implementing the random assignment of conditions; parental bias towards various forms of intervention or teaching method which might lead parents to interfere with the effects by communicating, sometimes unconsciously, expectations or doubts to the children (Loeb et al., 2001). There is also the difficulty of separating individual teacher or practitioner effects from program effects (Lillard, 2007). A solution to this might involve choosing participants from the clients of practitioners or teachers with similar methods. This may increase the *internal consistency* (i.e., the measurement of effects with accuracy within the participant group) of the study but might compromise the *external consistency* (i.e., generalizablity) of the study's results. Another solution involves pausing any intervention programs other than the one under study for the duration of the study. This solution would deprive participants of positive results that would have occurred due to paused interventions.

7.2. Assessing Computer Language Intervention: *Fast ForWord*

In this section, *Fast ForWord,* is examined in more detail, with a particular focus on the methodologies that were used for its evaluation. *Fast ForWord* provides intervention that aims to improve speech comprehension. It is among the most popular and most studied computer language intervention systems and successfully uses games and computer activities to deliver language intervention.

The design of *Fast ForWord* was based on two conjectures. The first is that many language capacities and not only speech are based on skills that involve the processing of rapidly changing sensory inputs; these temporal skills are recruited not only for the perception and reception of speech sounds, but also its production (albeit indirectly) (Tallal, 1980, 1990). For example, the targeted deficit causes children to have problems discriminating between many speech syllables, such as [ba] and [da], "which are characterized by very rapid frequency changes or formant transitions that occur during the initial few tens of milliseconds" (Tallal et al., 1996). The targeting of speech comprehension serves to improve underlying temporal processing skills, in this case the recognition and sequencing of the spectrotemporal structure of speech. Carry-over effects are hypothesized to other language capacities that entail the temporal processing (Tallal, 1980, 1990). The skills targeted included: In tone discrimination, detection of individual phoneme changes, matching phonemes to a target, identifying matched syllable pairs and discriminating between minimal pair words.

The second conjecture is that these temporal processing abilities can be improved through exercises and that the exercises can make use of *scaffolding*. Scaffolding is a technique where skills are mastered at a slower speed which is then increased and each step is based on improvements gained earlier. *Fast ForWord* implements this scaffolding through a suite of exercises that are delivered to the children in the form of graphical computer games and computer-based activities. The initial exercises make use of specially modified speech in which the speed of change of the critical formant transitions has been slowed down (Gilliam, 2008). This modified speech is easier for children with language comprehension impairments to understand. Thus, the instruction involves modifying the speech signal to a slower rate and gradually reducing the changes (i.e., speeding up the transitions until they are the same as the original speech) as improvements in the comprehension and creation of speech are detected (Merzenich et al., 1996).

In sum, the design of *Fast ForWard* is thus based on these conjectures: first, that deficits in skills involving temporal processing of rapidly changing sensory inputs are a cause of many speech comprehension and production errors, and second, that these deficits can be reduced or eliminated by using computer-based games and activities that involve scaffolding in the form of using speech in which transitions between sounds are slowed down initially and gradually are speeded up.

Fast ForWord was designed to target relevant brain regions that process fast-changing acoustic input. Temple et al. (2000) used the *Functional Magnetic Resonance Imaging* (fMRI) brain imaging technique to confirm changes in brain activity in children with developmental dyslexia after using the program. Two groups of children, one with developmental dyslexia and one without this condition participated in the experiment. fMRI scans of both groups were taken before the intervention. Children with dyslexia then used *FastForword* for 28 days for 100 minutes per day. Three sets of behavioral tests, evaluating early reading skills, were conducted before and after the intervention. The tests were *the Comprehensive Test of Phonological Processing* (CTOPP) (Mitchell, 2001), the *Clinical Evaluation of Language Fundamentals* (CELF–3) (Semel et al., 2000), and the *Woodcock Reading Mastery Test* (WRMT-R) (Woodcock, 1987). After the intervention, it was observed that for children with dyslexia several brain regions, including the *temporo-parietal* region and the *left frontal* region showed increased activity. Both of these regions typically have reduced activity in children with dyslexia compared with children without this condition. Behavioral tests repeated after the intervention showed improvement in the reading skills of the children. The researchers concluded that children with dyslexia have reduced cortical activity in the mentioned regions and that this activity can increase with the use of the *Fast ForWord* intervention.

Different versions of the program have been developed that use the same strategy to target language skills other than speech comprehension. *Fast ForWord-Language* (FFW-L) not only targeted speech reception skills but also reading skills. In addition to skills targeted in the original version of the program, the skills of recalling commands, and comprehension of grammatical morphemes and complex sentence structures, among others, were also addressed (Gilliam, 2008).

The developers of *Fast ForWord* claimed that training with the system for only six weeks would translate into one to one and a half years of gains compared with absence of intervention (Merzenich et al., 1996; Tallal et al., 1996).

Two influential and controversial studies investigated specifically the impact of the *Fast ForWord* intervention and confirmed the two conjectures at its core (Merzenich et al., 1996; Tallal et al., 1996).

In the first study, seven children between the ages of five and nine with language learning disabilities received three hours of language intervention daily totaling sixty hours over four weeks. The intervention consisted of playing two games in the *Fast ForWord* suite, called "Circus Sequence" and "Phoneme Identification", in addition to eight other speech and language exercises presented by trained SLPs. All the exercises used recorded modified speech. The children also listened to one to two hours of stories with modified speech every day. At the end of the training phase, the children improved significantly on measures of speech discrimination and language comprehension. There were no control groups.

A second study was conducted, this time with a larger participant group totaling twenty-two children from five to ten years old with language impairments who were divided into two matched groups, based on age and language skill level. All children received three and a half hours of language intervention and one to two hours of listening to stories at home. However, one group received exercises and stories in which recorded modified adaptive speech was used, and the other groups received exercises and stories that used unmodified natural speech. After training, both groups showed statistically significant improvements. However, a significant between-group difference was observed in the areas of temporal processing, speech discrimination and grammatical comprehension, where the group that received intervention with modified speech showed significantly greater improvements. No significant between-group difference was observed for improvement to language comprehension. The children in the group that received intervention with modified speech were retested six weeks after intervention and it was observed that the gains were retained. The authors concluded the computer training, especially training using modified speech, was effective at improving underlying language skills, especially temporal processing.

A number of independent studies have questioned the results of the first two studies and have suggested that the modified speech approach does not have significant advantages over similar methods (Gillam et al., 2008; Loeb et al., 2001; Pokorni et al., 2004). In the following sections, these studies and the methodology used are briefly reviewed.

7.2.1. Assessing *Fast ForWord* Using Randomized Controlled Trial Studies

As mentioned in section 7.1, *Randomized Controlled Trials* (RCTs) are an established method for comparing the effects of conditions using experimental trials (Pelham & Blanton, 2007). Pokorni et al. (2004) conducted a RCT study of *Fast ForWord* with fifty-four children with language and reading skills that were at least one year below their grade level. The children were randomly assigned to one of three possible interventions: (a) FFW-L, a version of *Fast ForWord* described before, (b) Earobics (Cognitive Concepts, 2000), a computer reading intervention system, and (c) the Lindamood Phoneme Sequencing Program (LiPS) (Lindamood & Lindamood, 1998), a non-computerized intervention program, both mentioned in section 3. The children received three hours of daily intervention for twenty days. The researchers conducted tests on three subsets of the *Clinical Evaluation of Language Fundamentals, Third Edition* (CELF–3) (Semel et al., 2000). The measured subsets were (a) Concepts and Directions, (b) Recalling Sentences and (c) Listening to Paragraphs. The assessments were conducted four to six weeks prior and six to eight weeks after the completion of the intervention. No significant differences among the interventions were found. These results were confirmed by another similar RCT study (Cohen et al., 2005).

Another study used an even larger number of participants. Gillam et al. (2008) divided two-hundred and fourteen children into four groups. Each group used one of four language interventions for five days per week over six weeks. The interventions were (a) FFW-L, (b) *academic enrichment* (AE) which consisted of a suite of computer games that were designed to be educational but not to address any language skills, (c) *computer-assisted language intervention* (CALI), which consisted of seven computerized instructional modules selected from the Earobics software (Cognitive Concepts, 2000) and *Laureate Learning software* (Semel, 2000; Wilson & Fox, 1997), and (d) *individualized language intervention* (ILI), a suite of non-computerized exercises and activities typically employed in SLP therapy. The last intervention was provided by a SLP whereas the other interventions were provided by a computer. The Earobics/Laureate Learning software-based exercises were similar to the ones provided by FFW -L in that they targeted cognitive, processing, and language skills, but without the use of modified speech.

Two measures were used to assess the effects of the interventions. The first measure was *backward masking* (Marler et al., 2001), which measures an individual's ability to distinguish between sounds in time. This measure evaluates the ability to recognize a sound that is followed rapidly by another sound (i.e., *mask*). The second measure was derived from the *Comprehensive Assessment of Spoken Language* (CASL) (Carrow-Woolfolk, 1985), a norm-referenced, standardized test of expressive and receptive language. Five subsets of the suite (Antonyms, Syntax Construction, Paragraph Comprehension, Nonliteral Language, and Pragmatic Judgment) were used and a sum of scores from them was calculated. A high degree of correlation among measurements derived using these five suite subsets suggests that they all measure the same underlying language skill (Gillam et al., 2008).

The children were tested immediately after treatment, three months after treatment and six months after treatment. Immediately after the treatment, all children improved significantly as measured by the backward masking and CASL tests. Also, all conditions lead to improvements in the measures skills and there was no significant difference between the conditions. However, at the six-month follow-up point of measurement, children in the two conditions that involved language-specific computer exercises, FFW-L and CALI, showed higher phonological awareness scores than children in the other two conditions; thus, retention was higher for language-specific computer-based interventions. The results showed that the FFW-L approach, which employed the use of modified speech, was not more effective than the other language-specific computer-based interventions or other computer-based interventions, for example the exercises used in CALI, that target similar language skills.

The researchers stated that, due to the absence of a control group with no intervention, it was difficult to attribute whether the improvements were the result of the interventions or other confounding factors, such as practice effects, maturation or spontaneous recovery. If the improvements were due to the interventions, then computer-delivered training has potential to be as effective as human-delivered training. On the other hand, if the results were due to confounding factors, then all methods are similarly ineffective. The researchers conclude that the results suggest that intervention, in the form of intensive daily experiences in listening and responding to verbal input, had a positive effect on language skills, whether delivered by a computer software program or a human agent.

7.2.2. Assessing *Fast ForWord* Using Case Studies

All of the independent studies reviewed in the previous section, demonstrated the effects of different intervention conditions using experimental designs with relatively large groups of subjects. Another evaluation methodology is the *in-depth case study*. The case study focuses on a small number of individuals. These studies often involve obtaining a large number of measurements over a long period of time (Lazar et al., 2010). These studies are especially useful in the presence of many inter-related factors. In these cases, taking one or two measures might not be enough to capture the “big picture” and provide a complete characterization of the effects of an intervention.

Loeb et al. (2001) used this methodology to examine the effect of using FFW-L in the home setting. They conducted four case studies, each for a child with language impairments. The researchers stated that, rather than examining specific effects of the program (i.e., efficacy), they aim to uncover areas in which large or consistent changes might be observed by surveying a large number of measures. To this end they chose to use a case study design that allows a more in-depth study of a few participants and the use of many measures.

Various sets of measures were used in the study. The *Test of Language Development-Primary* (TOLD-P) (Newcomer & Hammill, 1982), the *Token Test for Children* (DiSimoni, 1978), and *the Goldman-Fristoe Test of Articulation* (GFTA) (Goldman & Fristoe, 1986) were used to assess speech comprehension and articulation. The *Developmental Sentence Scoring* (DSS) (Lee, 1974) was used to assess changes over time in the syntax used by children in their oral narratives. It also assessed the use of grammatical devices in spontaneous speaking tasks. The *Criterion Referenced Inventory of Language-Pragmatics* (CRIL-P) (Wiig, 1990) and the *Pragmatic Protocol* (Prutting & Kirchner, 1987) were used to assess changes in pragmatic skills that affect communication competence. The first assessment tool allows for a structured elicitation of speech acts whereas the second assessment tool allows for the assessment of linguistic and paralinguistic features of a conversation and for the evaluation of the extent to which problem behaviors interfere with the conversation as a whole. Finally, three sets of tests, the *Test of Phonological Awareness* (TOPA) (Torgesen & Bryant, 1994), the *Woodcock Reading Mastery Test – Revised* (WRMT-R) (Woodcock, 1987) and the *Gray Oral Reading Test-3* (GORT-3) (Wiederholt & Bryant, 1992) were used to assess changes in reading.

All the children showed improvements after using FFW-L in their homes. However, the observed improvements were more moderate than those previously reported. Only 32% of the assessed standardized language measures improved significantly, and of these, only 61% were retained after three months. In a closer examination, it was discovered that there were no consistent pattern to the improvements. More importantly, the effects might not have been due to using the specific approach of FFW-L with modified speech. Instead it might have been the game-like characteristics of FFW-L combined with the encouragement of administering parents that provided benefits. The researchers concluded that using speech in the context of games benefitted the children in the case studies. However, in the case of FFW-L, the magnitude, generalizability and retention of the developed skills were not as dramatic as originally reported (Merzenich et al., 1996; Tallal et al., 1996).

7.2.3. Summary

An examination of the methodologies used to assess *Fast ForWord* provides insights that will be applied to the assessment of the proposed system. RCTs and case studies are different methodologies, which provide answers to different questions. Case studies allow for more in depth and comprehensive examination of the effects of a specific intervention. They are especially useful for examining system designed for diverse user populations such as people with disabilities. On the other hand, RCTs can answer precise questions about the relative efficacy of interventions.

8. A Brief Description of the Tangible Toy

This document discussed and synthesized prior work relevant to the development of a computer speech intervention system that involves the combining of ASR and TUIs in the form of a digital toy that will be prototyped within the context of my dissertation. As described in section 3, a central part of the proposed system is the development of *D-Toy,* an interactive tangible toy that that will be used by children. To contextualize further this undertaking in the light of the reviewed research, research results pertaining to the design of *D-Toy* will be briefly described here.

In the planned dissertation, ASR and TUIs will be combined in the design of *D-Toy* such that game play requires the use of speech and movement in the support of speech intervention. No prior intervention system has combined this technology with ASR. TUIs will be used to achieve two key results in the proposed system; they will be used to engage motor movement in addition to speech and they are used to divorce the toy (from the perspective of the child) from the computer and thus increase novelty and immersion in the game.

The proposed system allows a shift in focus from providing accurate feedback to creating an engaging and pleasurable interaction for the child. This approach will simplify the interaction for the child and allows him or her to focus on intuitive language learning skills rather than deciphering complex computer feedback. Inexact results (i.e., those identified by the ASR system with low confident scores) will be disguised as feedback incorporating and supporting uncertainty, a desirable game characteristic. The performance of the child during game play will be monitored by the system and *dynamic difficulty adjustment* will be used to maintain engagement with the game.

In the light of the various factors that can affect the interaction and the novelty and exploratory nature of the project, in depth case studies with a small number of participants over longer periods of times is an appropriate assessment methodology for the evaluation of the system, especially the interaction with *D-Toy.*

9. Conclusion

In this document, I have discussed and synthesized previous research that is relevant to the following questions outlined in section 1: what previous work has been conducted that involves the design of computational media intended for language intervention? To what extent has ASR, TUIs and game design have been used in the development of computer language intervention programs, on their own and in combination with each other? What is the nature of the evidence that supports the design rationale of combining TUIs and ASR in a speech intervention system? What assessment methods have been previously used for the evaluation of similar approaches to speech intervention? What research questions have been raised in this area?

This review has demonstrated that game design and ASR have been previously used for speech intervention for children. In section 4.1, two such systems, ARTUR (Bälter et al., 2005) and Box of Tricks (Vicsi et al., 2000), are reviewed. Additionally, a novel use of *Speech Viewer II,* originally developed for adults, for intervention for children along with a related prototype were also examined (Öster, 1995; Öster et al., 2003).To a much smaller degree, TUIs have been as well: *LinguaBytes* (Hengeveld et al., 2009) and *E-scope* (Hummels et al., 2006) are two interfaces developed to aid with speech intervention for children. However, despite this previous work, important possibilities, such as combining ASR and TUIs, remain unexplored. No system has so far used a *combination* of these techniques and technologies to address speech intervention.

Section 2 addressed the topic of the current non-digital methods for speech intervention that are in use for the population of children with speech sound disorders. Speech sound disorders, lower-level disorders that have various causes and generally occur in children, were described.

Individuals in the target user population need speech intervention that can alleviate or eliminate speech errors, improving communication and social interactions. This population forms the largest subgroup of children with speech or language disorders. Many interventions that address these disorders are especially suitable to be supplemented by digital systems because they involve the elicitation and repetition of spoken utterances chosen to include the problematic sounds. This can be simulated using automatic systems.

This section also discussed research that shows that behavior-based therapies involving intensive face-to-face sessions with a SLP are, in general, effective for the alleviation or even elimination of speech disorders. Furthermore, games that involve "low-tech" props, such as dolls and pictures are often employed during intervention sessions. The use of *imitation* or *auditory stimulation* that involves the correct repetition of a mispronounced utterance without any further explanation followed by trial by the client is widely used and suggested to be the first method that should be used for speech elicitation. Also, it is noted that for proficient learners detailed feedback is not necessary and a redirection to the correct alternative, is often enough for correction. This is relevant to the proposed system because it is intended to supplement rather than replace intervention provided by the SLP. In this sense, detailed corrective feedback is provided by the SLP and the proposed system supports the continuation of the exercises in the absence of the SLP.

Section 2 also set out a broader framework of systems that are intended to intervene and bring about speech modification: *rehabilitative systems* aim to improve speech impaired due to disease or accident, and *training systems* aim to improve skills in a new language or dialect and/or to reduce accent. *Intervention systems* aim to correct or improve speech errors due to congenital or acquired speech impairments.

Section 3 provides an overview of the planned dissertation work, which frames the questions pursued here in this document.

Section 4 provided a review of systems that use ASR for intervention, rehabilitation or training. From this review, the following points can be taken away:

- Extant ASR engines have been developed for use by people with standard speech, for the most part. Deviations in input speech from the training models that are based on standard speech result in recognition errors and the presence of low confidence recognition scores. Systems that have been developed for people with non-standard speech have addressed this issue by using *input restriction* and *acoustic training.* The first strategy, *input restriction,* is relevant to the proposed project and will be used. It entails configuring the ASR engine to expect a restricted set of input words at any point in the interaction, thereby simplifying the recognition task and often decreasing recognition errors.

- ASR technology as of present often provides unreliable and approximate feedback in the presence of ambiguous or erroneous speech. This results in unintuitive and inconsistent feedback, which user studies consistently reveal to be confusing to users and in some cases even worse than no feedback at all.

- ASR, especially when combined with gaming elements, has been successful in motivating children to use their speech when interacting with the systems. Even systems that provide only "binary" feedback (i.e., whether the input speech was recognized or not) have been effective in situation in which motivation was maintained.

A useful observation stemming from section 4 is that ASR is more powerful as a tool to *facilitate and motivate interaction* rather than as a tool to *provide automatic analysis and detailed feedback.* Recognition of these two different and distinct potential roles of ASR is hypothesized as a key to its successful development in my planned dissertation work.

In section 5, serious games, defined as games that have goals other than achieving entertainment as their primary focus, are discussed. Previous research was reviewed that shows that serious games have been successfully used for learning and rehabilitation. In the context of rehabilitation, games were shown to result in increased motivation and led to the patients finishing more exercises and achieving better results. One aspect of gaming, the use of *uncertainty¸* is especially relevant to the proposed project (section 5.4.1). Evidence was reviewed that shows how *uncertainty* often provokes emotional responses and increases engagement in the game. With respect to the current project, a promising strategy is embracing the uncertainty in the outcome of ASR with non-standard speech as if it were an intentional game-like characteristic.

In section 6, the characteristics of TUIs were examined. Key characteristics of TUIs that make them suitable for education and intervention applications were discussed. These include *tangibility, flexibility* and *situatedness.* Evidence from previous research was presented that demonstrates the important of engaging movement for learning and development. TUIs can support the *tactile-kinesthetic* teaching mode that uses movement and touch. Also, they support playful learning, as well as, the development of *sympathetic interfaces* that can engage children's emotions and aid in learning. Also, previous research was discussed that points to a deep connection between movement and speech, both from a neurological perspective and from gesture research (Iverson & Thelen, 1999). Combining TUIs and ASR allows for the engagement of both movement and speech. Research was also presented that supports the use of TUIs for the development of *reality-based interaction* (RBI), where the interaction capitalizes on people's extant knowledge of the physical world. Several TUIs were described and were shown to exhibit the qualities described above.

Finally, in section 7, the question of assessment was addressed. The first issue was noting the various purposes that "evaluation" has served in the research literature, ranging from exploratory research that involve qualitative aspects of interaction to experimental methods that focus on the precise quantitative aspects of systems. In addition, challenges involved in evaluating the system for children with disabilities were reviewed. These challenges include difficulty of controlling for external factors and the diversity of the target population. Randomized controlled trials and case studies of the system *Fast ForWord,* were examined in more depth in section 7.2. Evidence in section 7 shows that in depth case studies with a small number of participants over longer periods of times has merits for the evaluation of novel intervention systems, where there are many aspects to the interaction and a number of factors might be involved. RCTs were shown to be suitable for comparing a small number of factors that differ between variants of an approach and need to be measured with precision.

For my dissertation, I plan to design and develop a speech intervention system for children by combining ASR and TUI technologies in the form of an interactive game using a computational toy. With respect to the proposed project, I have both reviewed relevant research about similar extant systems and approaches, identifying their issues and strengths, and presented background research results that inform my design decisions.

10. References

Adams, M. J. (2006). The promise of automatic speech recognition for fostering literacy growth in children and adults. In *The International Handbook of Literacy and Technology*, Volume 2, McKenna, M.C., Labbo, L. D., Kieffer, R. D. and Reinking, D. (Eds.), Lawrence Erlbaum Associates, Inc., Mahwah, New Jersey, 109-128.

Alankus, G., Lazar, A., May, M., and Kelleher, C. (2010). Towards customizable games for stroke rehabilitation. In *Proceedings of the 28th International Conference on Human factors in Computing Systems* (CHI '10), ACM, New York, NY, 2113-2122.

Arrowsmith. (1998). *A Report on the Effectiveness of the Arrowsmith Program in the Toronto Catholic District School Board.*

Arrowsmith. (2007). *Report on the Arrowsmith Program in the Toronto Catholic District School Board.*

Arslan, L. and Hansen, J. (1997). Frequency characteristics of foreign accented speech. In *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing* (ICASSP’97). Volume 2, IEEE, New York, NY, 1123 – 1126.

Bälter, O., Engwall, O., Öster, A., and Kjellström, H. (2005). Wizard-of-Oz test of ARTUR: A computer-based speech training system with articulation correction. In *Proceedings of the 7th International Conference on Computers and Accessibility* (ASSETS'05), ACM, New York, NY, 36-43.

Baskinger, M. and Gross, M. (2010).Tangible interaction = form + computing. In *Interactions*, ACM, New York, NY, 17(1), 6-11.

Bauman-Waengler, J. (2004). *Articulatory and Phonological Impairments: A clinical focus (2nd ed.)*, Peasron, Boston, MA.

Bernthal, J. A. and Bankson, N. W. (2004). *Articulation and Phonological Disorders*. Allyn and Bacon, Inc., Boston, MA.

Bemstein, J. (1994). Speech recognition in language education. In *Proceedings of CALICO'94*, 3741.

Bergin, D. A. (1999). Influences on classroom interest. In *Educational Psychologist*, 34(2), 87–98.

Bhat, S. (2010). TouchTone: An electronic musical instrument for children with hemiplegic cerebral palsy. In *Proceedings of the 4th International Conference on Tangible, Embedded, and Embodied Interaction* (TEI'10), ACM, New York, NY, 305-306.

Blackwell, A. F. and Edge, D. (2009). Articulating tangible interfaces. In *Proceedings of the 3rd International Conference on Tangible and Embedded Interaction* (TEI'09), ACM, New York, NY , 113-118.

Bridgeland, J. M., Dilulio, J. J., and Morison, K. B. (2006). *The Silent Epidemic: Perspectives of High School Dropouts*. Bill and Melinda Gates Foundation, Seattle, WA.

Brierley, B., Medford, N., Shaw, P., and David, A. S. (2007). Emotional memory for words: Separating content and context. In *Cognition and Emotion*, 21(3), 495–521.

Brotto Furtado, A. W., Falcão, T. P., Gomes, A. S., Eduardo, C., Rodrigues, M., and Sonnino, R. (2008). E-du box: Educational multimedia with tangible-enhanced interaction. In *Proceedings of the 7th ACM Conference on Designing Interactive Systems* (DIS '08). ACM, New York, NY, 139-146.

Buechley, L., Eisenberg, M., Catchen, J., and Crockett, A. (2008). The LilyPad Arduino: Using computational textiles to investigate engagement, aesthetics, and diversity in computer science education. In *Proceeding of the 26th Conference on Human Factors in Computing Systems* (CHI '08). New York: ACM, 423-432.

Buechley, L., Hendrix, S., and Eisenberg, M. (2009). Paints, paper, and programs: First steps toward the computational sketchbook. In *Proceedings of the 3rd International Conference on Tangible and Embedded Interaction* (TEI '09), ACM, New York, NY, 9-12.

Bunnell, H.T. Yarrington, D.M. and Polikoff, J.B. (2000) STAR: Articulation training for young children. In *Proceedings of the International Conference on Spoken Language Processing* (InterSpeech'00), Volume.4, ISCA, 85-88.

Burke, J. W., McNeill, M. D. J., Charles, D. K., Morrow, P. J., Crosbie, J. H., McDonough, S. M. (2009). Optimizing engagement for stroke rehabilitation using serious games. In *The Visual Computer: International Journal of Computer Graphics*, 25 (12), 1085-1099.

Burns, M. (2003). Fast ForWord Products Open a Child's Window to Language. In *Autism Asperger's Digest,* Volume 1, Zysk, V. (Ed.), Future Horizons, Arlington, TX, 89-96.

Camarata, S. M., Nelson, K. E., and Camarata, M. N. (1994). Comparison of conversational-recasting and imitative procedures for training grammatical structures in children with specific language impairment. In *Journal of Speech and Hearing Research,* 37, 1414–1423.

Caplan, T, and Caplan, F. (1993). *The Early Childhood Years: The 2 to 6 Year Old*. Putnam: New York.

Carrow-Woolfolk, E. (1985). *Test for Auditory Comprehension of Language*. DLM Teaching Resources, Allen, TX.

Cassell, J., and Jenkins, H. (1998). *From Barbie to Mortal Kombat: Gender and computer games*. MIT Press, Cambridge, MA.

Chang, H. P. (1993). Speech input for dysarthric users. In *Meeting of the acoustical society of America.* 2aSP7.

Chang, E. and West. M. (2006). Digital ecosystems and comparison to collaboration environment. In *WSEAS Transactions on Environment and development,* 2, 1396-1404.

Chen, F., and Kostov, A. (1997). Optimization of dysarthric speech recognition. In *Proceedings of the 19th Annual International Conference of the Institute of Electrical and Electronics Engineers*, IEEE, New York, NY, 1436–1439.

Chen, S., and Michael, D. (2006). *Serious Games: Games that Educate, Train and Inform*. Thomson Course Technology PTR, Boston, MA.

Clark, J. A. and Roemer, R. B. (1977). Voice controlled wheelchair. In *Archives of Physical Medicine and Rehabilitation*, 58(4), 169-175.

Cleave, P. L., and Fey, M. E. (1997). Two approaches to the facilitation of grammar in children with language impairments: Rationale and description. In *American Journal of Speech-Language Pathology,* 6, 22–32.

Clifford, M. M. (1988). Failure tolerance and academic risk-taking in ten- to twelve-year-old students. In *British Journal of Educational Psychology,* 58, 15–27.

Clifford, M. M., and Chou, F. C. (1991). Effects of payoff and task context on academic risk taking. In *Journal of Educational Psychology,* 83(4), 499–507.

Cockburn, J., Bartlett, M., Tanaka, J. Movellan, J. and Schultz, R. (2008). SmileMaze: A tutoring system in real-time facial expression perception and production in children with Autism Spectrum Disorder. In *Proceedings from the IEEE International Conference on Automatic Face and Gesture Recognition,* IEEE, New York, NY, 978-986.

Cohen, A. and Graupe, D. (1980). Speech recognition and control system for the severely disabled. In *Journal of Biomedical Engineering*, 2(2), 97-107.

Cohen, M. H., Giangola, J. P. and Balogh, J. (2004) *Voice User Interface Design*. Addison-Wesley, Boston, MA.

Cohen, W., Hodson, A., O'Hare, A., Boyle, J., Durrani, T., McCartney, E., Mattey, M., Naftalin, L. and Watson, J. (2005). Effects of computer-based intervention through acoustically modified speech (Fast ForWord) in severe mixed receptive–expressive language impairment: Outcomes from a randomized controlled trial. In *Journal of Speech, Language, and Hearing Research*, 48, 715–729.

Cognitive Concepts. (2000). Earobics [Computer software]. Houghton Mifflin, Evanston, IL.

Coniam, D. (1999). Voice recognition software accuracy with second language speakers of English. In *System*, 27, 49–64.

Corbett, S. (2010). Learning by playing: Video Games in the classroom. In *The New York Times*, September 15th 2010.

Daniloff, R.G.,Wilkox, K., and Stephens, M.I. (1980). An acousticarticulatory description of children's defective (s) productions. In *Journal of Communication Disorders,* 13, 347–363.

Danis, C. and Karat, J. (1995).Technology-driven design of speech information systems. In *Proceedings of the Symposium on Designing Interactive Systems* (DIS'95), ACM, New York, NY, 17-24.

Derwing, T. M., Munro, M. J., and Carbonaro, M. (2000). Does popular speech recognition software work with ESL speech? In *TESOL Quarterly*, 34, 592-603.

Diakoloukas, V., Digalakis, V., Neumeyer, L., and Kaja, J. (1997). Development of dialect-specific speech recognizers using adaptation methods. In *Proceedings of the IEEE Conference on Acoustics, Speech, and Signal Processing* (ICASSP'97), IEEE, New York, NY, 1455-1458.

Dickie, C., Vertegaal, R., Sohn, C. and Cheng, D. (2005) eyeLook: Using attention to facilitate mobile media consumption. In *Proceedings of the 8th Symposium on User Interface Software and* Technology (UIST'05), ACM, New York, NY, 103-106.

DiSimoni, F. (1978). *The Token Test for Children.* Teaching Resources Corporation, Boston, MA.

Dragon Systems. (1999). *Dragon NaturallySpeaking SDK, C++ and SAPI Guide and Reference.* Dragon Systems, Newton, MA.

Dunne, A. (1999). *Hertzian Tales: Electronic Products, Aesthetic Experience and Critical Design*. RCACRD Research Publications, London, UK.

Elliot, R., Friston, K. J., and Dolan, J. R. (2000). Dissociable neural responses in human reward systems. In *Journal of Neuroscience,* 20(16), 6159–6165.

Erhard, P., Kato, T., Strupp, J.P., Andersen, P., Adriany, G., Strick, P.L., and Ugurbill, K. (1996). Functional mapping of motor in and near Broca's area. In *Neuroimage*, 3 (1), S367.

Eriksson E., Bälter O., Engwall O, Öster A-M and Kjellström H. (2005). Design recommendations for a computer-based speech training system based on end-user interviews. In *Proceedings of the 10th International Conference on Speech and Computers (SPECOM '05),* 483-486.

Erk, S., Kiefer, M., Grothe, J., Wunderlich, A. P., Spitzer, M., and Walter, H. (2003). Emotional context modulates subsequent memory effect. In *Neuroimage*, 18, 439–447.

Fails, J. A., Druin, A., Guha, M. L., Chipman, G., Simms, S., and Churaman, W. (2005). Child's play: A comparison of desktop and physical interactive environments. In *Proceedings of the 2005 Conference on Interaction Design and Children* (IDC'05). ACM, New York, NY, 48-55.

Feng, J., Spence, I., and Pratt, J. (2007) Playing an action videogame reduces gender differences in spatial cognition. In *Psychological Science*, 18, 850–855.

Fernaeus, Y. and Tholander, J. (2005). Looking at the computer but doing it on land: Children's interactions in a tangible programming space. In *Proceedings of the 11th International Conference on Human-Computer Interaction* (HCI'05), 3-18.

Fernaeus, Y. and Tholander, J. (2006). Finding design qualities in a tangible programming space. In *Proceeding of the 24th Conference on Human Factors in Computing Systems* (CHI '08), ACM, New York, NY, 447-456.

Ferrier, L., Shane, H., Ballard, H., Carpenter, T., and Benoit, A. (1995). Dysarthric speakers' intelligibility and speech characteristics in relation to computer speech recognition. In *Augmentative and Alternative Communication*, 11, 165–174.

Fey, M. E., Cleave, P. L., Long, S. H., and Hughes, D. L. (1993). Two approaches to the facilitation of grammar in children with language impairment: An experimental evaluation. In *Journal of Speech and Hearing Research*, 36, 141–157.

Fiorillo, C. D., Tobler, P. N., and Schultz, W. (2003). Discrete coding of reward probability and uncertainty by dopamine neurons. In *Science*, 299, 1898–1902.

Flavell, J. H., (1963). *The Developmental Psychology of Jean Piaget*. D. Van Nostrand, New York, NY.

Flores, E., Tobon, G., Cavallaro, E., Cavallaro, F. I., Perry, J. C., and Keller, T. (2008). Improving patient motivation in game development for motor deficit rehabilitation. In *Proceedings of the 2008 International Conference on Advances in Computer Entertainment Technology* (ACE '08). ACM, New York, NY, USA, 381-384.

Folven, R., and Bonvillian, J. D. (1991). The transition from nonreferential to referential language in children acquiring American Sign Language. In *Developmental Psychology*, 27, 806-816.

Franco, H., Abrash, V., Precoda, K., Bratt, H., Rao, R., Butzberger, J., Rossier, R. and Cesari, F. (2000). The SRI EduSpeak system: Recognition and pronunciation scoring for language learning. In *Proceedings of NLP and Speech Technologies in Advanced Language Learning Systems* (InStil'00), 123–128.

Freeman, W. J., and Núnez, R. (1999). *Reclaiming Cognition: The Primacy of Action, Intention and Emotion.* Imprint Academic, Thorverton, UK.

Fried, I., Katz, A., McCarthy, G., Sass, K.J., Williamson, P., Spencer, S.S., and Spencer, D.D. (1991). Functional organization of human supplementary motor cortex studied by electrical stimulation. In *Journal of Neuroscience*, 11, 3656–3666.

Fried-Oken, M. (1985). Voice recognition device as a computer interface for motor and speech impaired people. In *Archives of Physical Medicine and Rehabilitation*, 66, 678–681.

Gaver, W.W., Bowers, J., Boucher, A., Gellerson, H., Pennington, S., Schmidt, A., Steed, A., Villars, N. and Walker, B. (2004). The drift table: Designing for ludic engagement. In *Proceeding of the 20th Conference on Human Factors in Computing Systems* (CHI '04), ACM, New York, NY, 885-900.

Gibson, J. J. (1979). *The Ecological Approach to Visual Perception*. Houghton Mifflin, Boston, MA.

Gillam, R. B., Loeb, D. F., Hoffman, L. M., Bohman, T., Champlin, C. A., Thibodeau, L., Widen, J., Brandel, J., Friel-Patti, S. (2008). The efficacy of Fast ForWord Language intervention in school-age children with language impairment: A randomized controlled trial. In *Journal of Speech Language Hearing Research,* 51, 97–119.

Glass, J. R. and Hazen, T. J. (1998). Telephone-based conversational speech recognition in the JUPITER domain. In *Proceedings of the 4th International Conference on Spoken Language* (ICSLP'98), 61–64.

Goldman, R. and Fristoe, M. (1986). *Goldman-Fristoe Test of Articulation.* American Guidance Service, Circle Pines, MN.

Goodwyn, S. W., Acredolo, L. P., and Brown, C. A. (2000). Impact of symbolic gesturing on early language development. In *Journal of Nonverbal Behavior,* 24, 81–103.

Goggin, G., and Newell, C. (2003). *Digital disability: The social construction of disability in new media*. Rowman & Littlefield, New York, NY.

Griffiths, M. (2005). Video games and health. In *British Medical Journal,* 331, 122–123.

Griffiths, M. D., and Davies, M. N. O. (2005). Videogame addiction: Does it exist? In *Handbook of Computer Game Studies*. Goldstein, J. and Raessens, R. (Eds.), MIT Press, Boston, MA, 359–368.

Gunasekera, W. S. L., and Bendall, J. (2005), Rehabilitation of neurologically injured patients, In *Neurosurgery*. Springer Science, Berlin, 407-421.

Hamidi, F., Baljko, M., Livingston, N., Spalteholz, L. (2010). CanSpeak: A Customizable speech interface for people with dysarthric speech. In *Proceedings of the International Conference on Computers Helping People with Special Needs* (ICCHP10), Springer, Berlin, 605-612.

Hamidi, F., and Baljko, M. (2010). Collaborative Poetry on the Facebook Social Network. In *Proceedings of the International Conference on Supporting Group Work* (GROUP10), ACM: New York, 305-306.

Hanlon, R.E., Brown, J.W., and Gerstman, L.J. (1990). Enhancement of naming in nonfluent aphasia through gesture. In *Brain and Language*, 38, 298–314.

Harter, S. (1978). Pleasure derived from cognitive challenge and mastery. In *Child Development*, 45, 661–669.

Hawley, M. S. (2002). Speech recognition as an input to electronic assistive technology. In *British Journal of Occupational Therapy*, 65(1), 15-20.

Hawley, M., Enderby, P., Green, P., Brownsell, S., Hatzis A., Parker, M., Carmichael, J., Cunningham, M., ONeill, P. and Palmer, R (2003). STARDUST: Speech training and recognition for dysarthric users of assistive technology. In *Assistive Technology – Shaping the Future*. Craddock, G. M. (Ed.), IOS Press, Amsterdam, 959-964.

Hengeveld, B., Hummels, C., Overbeeke, K., Voort, R., van Balkom, H., and de Moor, J. (2009). Tangibles for toddlers learning language. In *Proceedings of the 3rd International Conference on Tangible and Embedded Interaction* (TEI '09), New York: ACM, 161-168.

Higgins, E. I., and Raskind, M. H. (1995). Compensatory effectiveness of speech recognition on the written composition performance of postsecondary students with learning disabilities. In *Learning Disability Quarterly*, 18, 159–174.

Hill, E.L. (1998). A dyspraxic deficit in specific language impairment and developmental coordination disorder? Evidence from hand and arm movements. In *Developmental Medicine and Child Neurology*, 40, 388–95.

Hincks, R. (2002). Speech recognition for language teaching and evaluating: A study of existing commercial products, In *Proceedings of the 7th International Conference on Spoken Language* (ICSLP'02), 733-736.

Hornecker, E. and Buur, J. (2006). Getting a grip on tangible interaction: A framework on physical space and social interaction. In *Proceeding of the 22nd Conference on Human Factors in Computing Systems* (CHI '06), ACM, New York, NY, 437-446.

Howard-Jones, P.A., and Demetriou S. (2009). Uncertainty and engagement with learning games. In *Instructional Science*, 37 (6), 519-536.

Huang, X., Acero, A., and Hon, H. W. (2001). *Spoken Language Processing: A Guide to Theory, Algorithm, and System Development.* Prentice Hall, Upper Saddle River, NJ.

Hughes, F. P. (1991). *Children, Play, and Development.* Allyn and Bacon, Needham Heights, MA.

Hummels, C., Helm, A., Hengeveld, B., Luxen, R., Voort, R., Balkom, H., and Moor, J. (2006). Explorascope: An interactive, adaptive educational toy to stimulate the language and communicative skills of multiple-handicapped children. In *Proceedings of Art Abilitation,* Brooks, A.L. (Ed.), 16-24.

Humphries, J., Woodland, P., and Pearce, D. (1996). Using accent-specific pronunciation modeling for robust speech recognition. In *Proceedings of the International Conference on Spoken Language,* 2324 – 2327.

Hunicke, R. (2005). The case for dynamic difficulty adjustment in games. *In Proceedings of the 2005 ACM SIGCHI International Conference on Advances in computer entertainment technology* (ACE '05). ACM, New York, NY, 429-433.

Hux, K., Rankin-Erickson, J., Manasse, N., and Lauritzen, E. (2000). Accuracy of three speech recognition systems: Case study of dysarthric speech. In *Augmentative and Alternative Communication*, 16 (3), 186–196.

IBM Corporation. (1998). *Speech Manager Application Programming Interface Developers' Guide: IBM ViaVoice SDK for Windows.*

Inkpen, K., Booth, K., Gribble, S., and Klawe, M. (1995). Give and take: Children collaborating on one computer. In *Proceeding of the 11th Conference on Human Factors in Computing Systems* (CHI '95), ACM, New York, 258-259.

Ishii, H. (2008). Tangible Bits: Beyond pixels. In *Proceedings of the 2nd International Conference on Tangible and Embedded Interaction* (TEI'08), xv-xxv.

Ishii, H. and Ullmer, B. (1997). Tangible bits: Towards seamless interfaces between people, bits and atoms. In *Proceeding of the 13th Conference on Human Factors in Computing Systems* (CHI '97), ACM, New York, NY, 234-241.

Ishii, H., Mazalek, A., and Lee, J. (2001). Bottles as a minimal interface to access digital information, In *Proceeding of the 22nd Conference on Human Factors in Computing Systems* (CHI '01), ACM, New York, NY, 187-188.

ISLE D3.3 (1999). Recognition of learner speech, In *ISLE Deliverable,* D3.3.

Iverson, J.M., and Thelen, E. (1999). Hand, mouth, and brain: The dynamic emergence of speech and gesture. In *Journal of Consciousness Studies,* 6, 19–40.

Jacob, R. J. K., Girouard, A., Hirshfield, L. M., Horn, M. S., Shaer, O., Solovey, E. T., and Zigelbaum, J. (2008). Reality-based interaction: A framework for post-WIMP interfaces. In *Proceeding of the 24th Conference on Human Factors in Computing Systems* (CHI '08), ACM, New York, NY, 201-210.

Jacob, R.J.K., Ishii, H., Pangaro, G. and Patten, J. (2002). A tangible interface for organizing information using a grid. In *Proceeding of the 18th Conference on Human Factors in Computing Systems* (CHI '02), ACM, New York, NY, 339-346.

Johnson, M. P., Wilson, A., Blumberg, B., Kline, C., and Bobick, A. (1999). Sympathetic interfaces: Using a plush toy to direct synthetic characters. In *Proceeding of the 15th Conference on Human Factors in Computing Systems* (CHI '99), ACM, New York, NY, 152-158.

Johnson, A. and Jacobson, B. (1998). *Medical Speech-Language Pathology: A Practitioner's Guide.* Thieme, New York, NY, 550–555.

Kaiser, A. P., Yoder, P. J., and Keetz, A. (1992). Evaluating milieu teaching. In *Causes and Effects in Communication and Language Intervention,* Warren, S. F., and Reichle, J. (Eds.), Brookes, Baltimore, 9–48.

Kaptelinin, V. and Nardi, B. A. (2006). *Acting with Technology: Activity Theory and Interaction Design*. MIT Press, Boston. .

Kalanithi, J. (2007). *Connectibles: Tangible Social Networking.* MS Thesis, MIT Media Lab, Cambridge, MA. .

Kewley-Port, D., Watson, C.S., Elbert, M., Maki, D., and Reed, D. (1991). The Indiana speech training aid (ISTRA) II: Training curriculum and selected case studies. In *Clinical Linguistics and Phonetics*, 5(1), 13-38.

Klemmer, S. R., Hartmann, B., and Takayama, L. (2006). How bodies matter: Five themes for interaction design. In *Proceedings of the 6th ACM Conference on Designing Interactive Systems* (DIS '08), ACM, New York, NY, 140-149.

Koepp, M. J., Gunn, R. N., Lawrence, A. D., Cunningham, V. J., Dagher, A., Jones, T., Brooks, D. J., Bench, C. J. and Grasby, P. M. (1998). Evidence for striatal dopamine release during a video game. In *Nature*, 393, 266–268.

Lazar, J., Feng, J. H., Hochheiser, H. (2010). *Research Methods in Human-Computer Interaction*. Wiley, New York, NY.

Lee. L. (1974) *Developmental Sentence Analysis*. Northwestern University Press, Evanston, IL.

Lillard, A. (2007). *Montessori: The Science behind the Genius.* Oxford University Press, Oxford.

Lindamood, P. and Lindamood, P. (l998). *The Lindamood Phoneme Sequencing Program for Reading, Spelling, and Speech.* PRO-ED, Inc., Austin, TX.

Liu, W. K. and Fung, P. (1999). Fast accent identification and accented speech recognition. In *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing* (ICASSP'99), IEEE, New York, NY, 221 - 224.

Loeb, D. F., Stoke, C., Fey, M. E. (2001). Language changes associated with Fast ForWord-Language: Evidence from case studies. In *American Journal of Speech Language Pathology*, 10, 216-230.

Löwgren, J. and Stolterman, E. (2004). *Thoughtful Interaction Design*. MIT Press, Boston.

Lyster, R. (1998). Negotiation of form, recasts, and explicit correction in relation to error types and learner repair in immersion classrooms. In *Language Learning*, 48, 183-218.

Mackey, A., and Choi, J. Y. (1998). Review of TriplePlayPlus! English. In *Language, Learning and Technology,* 2, 19-21.

Maclean, N., Pound, P., Wolfe, C. and Rudd, A. (2002). The concept of patient motivation: a qualitative analysis of stroke professionals' attitudes. In *Stroke*, 33, 444–448.

Malone, T. W. (1982). Heuristics for designing enjoyable user interfaces. In *Proceeding of the 1982 Conference on Human Factors in Computing Systems* (CHI '82), ACM, New York, NY, 63-68.

Manasse, N. J., Hux, K., and Rankin-Erickson, J. L. (2000). Speech recognition training for enhancing written language generation by a traumatic brain injury survivor. In *Brain Injury*, 14, 1015–1034.

Manches, A., O'Malley, C., and Benford, S. (2009). Physical manipulation: Evaluating the potential for tangible designs. In *Proceedings of the 3rd International Conference on Tangible and Embedded Interaction* (TEI'09), ACM, New York, NY, 77-84.

Marler, J. A., Champlin, C. A., and Gillam, R. B. (2001). Backward and simultaneous masking measured in language-learning impaired children who received intervention with Fast ForWord or Laureate Learning systems software. In *American Journal of Speech-Language Pathology*, 10, 258–268.

Marshall, P. (2007). Do tangible interfaces enhance learning? In Proceedings of the 1st International Conference on Tangible and Embedded Interaction (TEI'07), ACM, New York, NY, 163-170.

Martini, F. and Bartholomew, E. (2003). *Essentials of Anatomy and Physiology*. Benjamin Cummings, San Francisco, CA.

Martinussen, E. S., Knutsen, J., and Arnall, T. (2007). Bowl: Token-based media for children. In *Proceedings of the 2007 Conference on Designing for User Experiences* (DUX '07), ACM, New York, NY, 3-16.

Menzel, W., Herron, D., Bonaventura, P. and Morton, R. (2000). Automatic detection and correction of non-native English pronunciations. In *Proceedings of NLP and Speech Technologies in Advanced Language Learning Systems* (InStil'00), 49-56.

Merrill, D., Kalanithi, J., and Maes, P. (2007). Siftables: towards sensor network user interfaces. In *Proceedings of the 1st international conference on Tangible and embedded interaction* (TEI'07). 75-78.

Merzenich, M. M., Jenkins, W. M., Johnson, P., Scheiner, C., Miller, S. L., and Tallal, P. (1996). Temporal processing deficits of language-learning impaired children ameliorated by training. In *Science*, 271, 77–81.

Microsoft Corporation, (2009). *Microsoft Speech SDK 5.1 for Windows Applications.*

Miller, L., Gillam, R., and Peña, E. (2001). Dynamic Assessment and Intervention of Children's Narratives. Pro-Ed, Austin, TX.

Mitchell, J., J. (2001). Comprehensive Test of Phonological Processing. In *Assessment for Effective Intervention*, 26, 3, 57-63.

Mitra, S., Tooley, J., Inamdar, P. and Dixon, P. (2003). Improving English pronunciation - an automated instructional approach. In *Information Technologies and International Development,* 1(1), 75-84, MIT Press, Boston, MA.

Miura, K. (2004). Tell me more Japanese. In *CALICO Journal,* 21(3), 659-670.

Nelson, K. E., Camarata, S. M.,Welsh, J., and Butkovsky, L. (1996). Effects of imitative and conversational recasting treatment on the acquisition of grammar in children with specific language impairment and younger language-normal children. In *Journal of Speech and Hearing Research,* 39, 850–859.

Neri, A., Cucchiarini, C., and Strik, W. (2003). Automatic speech recognition for second language learning: How and why it actually works. In *Proceedings of the 15th International Congress of Phonetic Sciences* (ICPhS'03), 1157–1160.

Netsell, R. and Abbs, J. (1979). Acoustic characteristics of dysarthria associated with cerebellar disease. In *Journal of Speech and Hearing Research*, 22, 627-648.

Newcomer, P. L. and Hammill, D. D. (1982). *Test of Language Development-Primary.* Pro-ED, Austin,TX.

Nguyen, P., Gelin, P. H., Junqua, J. C., and Chien, J, T. (1999). N-best based supervised and unsupervised adaptation for native and non-native speakers. In *Proceedings of the 1999 IEEE International Conference on Acoustics, Speech, and Signal (*ICASSP'99), IEEE Press, New York, 173 – 176. .

Nicholas, H., Lightbown, P.M. and Spada, N. (2001). Recasts as feedback to language learners. In *Language Learning*, 51, 719-758.

NIDCD Health Information, (2010), Accessed December 5th 2011, <http://www.nidcd.nih.gov/health/statistics/vsl/Pages/stats.aspx>.

Noice, H., Noice, T., and Kennedy, C. (2000). Effects of enactment by professional actors at encoding and retrieval. In *Memory*, 8, 353–363.

Ojemann, G.A. (1984). Common cortical and thalamic mechanisms for language and motor functions. In *American Journal of Physiology,* 246, R901–R903.

Okolo, C. M., and Bouck, E. C. (2007). Research about assistive technology: 2000-2006. What have we learned? In *Journal of Special Education Technology*, 22(3), 19-35.

Öster, A-M. House D., Green P. (2003). Testing a new method for training fricatives using visual maps in the Ortho-Logo-Pedia project (OLP), Phonum 9- Fonetik, 89-92.

Öster, A-M. (1995). Teaching speech skills to deaf children by computer-based speech training. In *Proceedings of 18th International Congress on Education of the Deaf,* Tel-Aviv, Israel.

Öster, A-M. (1989). Applications and experiences of computer-based speech training. In *STL-QPSR,* 1, 59-62.

Oviatt, S., DeAngeli, A., and Kuhn, K. (1997). Integration and synchronization of input modes during multimodal human-computer interaction. In *Referring Phenomena in a Multimedia Context and their Computational Treatment (ReferringPhenomena '97).* Association for Computational Linguistics, Stroudsburg, PA, 1-13.

Owen, M. (2005). *An Anatomy of Games*. Available at: <http://www.futurelab.org.uk/resources/publications_reports_articles/discussion_papers/Discussion_Paper259>.

Patel, R. (1999). Prosody conveys information in severely impaired speech. In *Proceedings of the European Speech Community Association Workshop of Dialogue and Prosody*, Veldhoven, The Netherlands, 129-135.

Patel, R. and Roy, D. (1998). Teachable interfaces for individuals with severe physical and speech disabilities. In *American Association for Artificial Intelligence Workshop on Assistive Technology*, Madison, WS. 40-47.

Pelham, B. and Blanton, H. (2007). *Conducting Research in Psychology: Measuring the Weight of Smoke*. Thomson Higher Education, Belmont, CA.

Picard, R.W. (2000). *Affective Computing.* MIT Press, Boston, MA.

Piper, A. M., Weibel, N., Hollan, D. (2010). Introducing Multimodal Paper-digital Interfaces for Speech-Language Therapy. In *Proceedings of the 12th International Conference on Computers and Accessibility* (ASSETS '10). ACM, New York, NY, 203-210.

Pokorni, J. L.,Worthington, C. K., and Jamison, P. J. (2004). Phonological awareness intervention: Comparison of Fast ForWord, Earobics, and LiPS. In *Journal of Educational Research,* 97, 147–157.

Povel, D.J. (1991). The visual speech apparatus: Theoretical and practical aspects. In *Speech Communication,* 10, 59–80.

Pratt, S, Heintzelman, A., and Deming, S. (1993). The efficacy of using the IBM Speech Viewer vowel accuracy module to treat young children with hearing impairment. In *Journal of Speech and Hearing Research*, 36, 1063–1074.

Prensky, M. 2000. *Digital Game-Based Learning.* McGraw-Hill, New York, NY.

Price, S., Rogers, Y., Scaife, M., Stanton, D. and Neale, H. (2003). Using 'tangibles' to promote novel forms of playful learning. In *Interacting with Computers*, 15, 2, 169--185.

Prutting, C. A. and Kirchner, D. M. (1987) A clinical appraisal of the pragmatic aspects of language. In *Journal of Speech and Hearing Disorders*, 52, 105-119.

Rainforth, B. and York-Barr, J. (1997). *Collaborative Teams for Students with Severe Disabilities: Integrating Therapy and Educational Services.* Paul H. Brookes Publishing Co., Baltimore, MD.

Raskind, M. H., and Higgins, E. L. (1999). Speaking to read: The effects of speech recognition technology on the reading and spelling performance of children with learning disabilities. In *Annals of Dyslexia*, 49, 251–281.

Redström, J. (2008). Tangled interaction: On the expressiveness of tangible user interfaces. In ACM Transactions Computer-Human Interaction, 15, 4, 1-17.

Reesner, T. (2002). "Tell Me More French" Software review. In *CALICO Journal,* 19, 419-428.

Rego, P., Moreira, P. M. and Reis, L. P. (2010). Serious games for rehabilitation: A survey and a classification towards a taxonomy. In *Proceedings of the 5th Iberian Conference on Information Systems and Technologies* (CISTI'10), 1-6.

Revelle, G., Zuckerman, O., Druin, A., and Bolas, M. (2005). Tangible user interfaces for children. In *Proceeding of the 21th Conference on Human Factors in Computing Systems* (CHI '05), ACM, New York, NY, 2051-2052.

Rideout, V., Roberts, D. F., and Foehr, U. G. (2005). *Generation M: Media in the Lives of 8-18 Year-olds.* The Henry J. Kaiser Family Foundation, Washington, DC.

Rogers, Y., Scaife, M., Gabrielli, S., Smith, H. and Harris, E. (2002). A conceptual framework for mixed reality environments: Designing novel learning activities for young children. In *Presence: Teleoperators and Virtual Environments,* 11, 6, 677-686.

Rosen, K. and Yampolsky, S. (2000). Automatic speech recognition and a review of its functioning with dysarthric speech. In *Journal of Augmentative and Alternative Communication,* 16, 48-60.

Rosenbaum, P. and Stewart, D. (2004). The World Health Organization international classification of functioning, disability and health: A model to guide clinical thinking, practice, and research in the field of cerebral palsy. In *Seminars in Pediatric Neurology*, 11, 5-10.

Rosengren, E., Raghavendra, P., and Hunnicutt, S. (1995). How does automatic speech recognition handle severely dysarthric speech? In *Proceedings of the 2nd TIDE Congress*, Placencia Porrero, I. and de la Bellacas, P. (Eds.), IOS Press, Amsterdam, 336–339.

Russell, M., Brown, C., Skilling, A., Series, R., Wallace, J., Bonham, B., and Barker, P. (1996) Applications of automatic speech recognition to speech and language development in young children. In *Proceedings of the Fourth International Conference on Spoken Language Processing* (ICSLP'96), IEEE Press, New York, NY, 176-179.

Ryalls, J., Michallet, B, and Le-Dorze, G. (1994). A preliminary evaluation of the clinical effectiveness of vowel training for hearing-impaired children on IBM's Speech Viewer. In *Volta Review*, 96, 19–30.

Ryokai, K., Marti, S., and Ishii, H. (2004) I/O brush: drawing with everyday objects as ink. In *Proceeding of the 20th Conference on Human Factors in Computing Systems* (CHI '04), ACM, New York, NY, 303-310.

Schneider, P., and Watkins, R. V. (1996). Applying Vygotskian developmental theory to language intervention. In *Language, Speech, and Hearing Services in Schools*, 27, 157–170.

Secord, W (2007) *Eliciting Sounds: Techniques for Clinicians, 2nd edition.* Thomson Delmar Learning, Clifton Park, NY.

Semel, E., Wiig, E. H. and Secord, W. A. (2000). Clinical Evaluation of Language Fundamentals. Psychological Corporation, London.

Semel, E. (2000). *Following directions [Computer software].* Laureate Learning Systems, Winooski, VT.

Sengers, P., Boehner, K., David, S., and Kaye, J. (2005). Reflective design. In *Proceedings of the 4th Decennial Conference on Critical Computing* (CC'05). Bertelsen, O.W., Bouvin, N. O., Krogh, P.G., and Kyng, M. (Eds.). ACM, New York, NY, 49-58.

Series, R.W. (1993). A speech training aid. In *Proceedings of the ESCA Workshop on Speech and Language Technology for Disabled Persons* (SLTDP-93), 173-176.

Silverman, F. (2006). Learning styles. In *District Administration*. 42, 9, 70-71.

Slater, M., Steed, A., McCarthy, J., Maringelli, F. (1998). The influence of body movement on subjective presence in virtual environments. In *Human Factors*, 40, 1998.

Slater, M. and Usoh, M. (1994). Body centred interaction in immersive virtual environments. In *Artificial Life and Virtual Reahty,* Magnenat Thalmann, N., and Thalmann, D. (Eds.), John Wiley and Sons, 125-148.

Snyder, R.F. (2000). The relationship between learning styles/multiple intelligences and academic achievement of high school students. In *The High School Journal*, 83, 11-20.

Spalteholz, L., Lin, K. F., Livingston, N. and Hamidi, F. (2008). KeySurf: A character controlled browser for people with physical disabilities. In *Proceedings of the 17th World Wide Web Conference* (WWW'08), 31-39.

Stanton, D., Bayon, V., Neale, H., Ghali, A., Benford, S., Cobb, S., Ingram, R., O'Malley, C., Wilson, J. and Pridmore, T. (2001). Classroom collaboration in the design of tangible interfaces for stroytelling. In *Proceeding of the 17th Conference on Human Factors in Computing Systems* (CHI '05), ACM, New York, NY, 482-489.

Statistics Canada. (2008). *Educational Services and the Disabled Child.* Uppal, S., Kohen, D., and Khan, S. (Eds.), Ottawa, Canada.

Statistics Canada. (2003). *Children with Disabilities and Their Families.* (Catalogue no. 89-585), Ottawa, Canada.

Steuer, J. (1992). Defining Virtual reality: Dimensions determining telepresence. In *Journal of Communication*, 42, 4, 73-93.

Suzuki, H. and Kato, H. (1995). Algoblocks: An open programming language. In *Proceedings of the First International Conference on Computer Support for Collaborative Learning* (CSCL'95), 349-355.

Taber, K. S. (2000). Case studies and generalisability - Grounded theory and research in science education, In *International Journal of Science Education*, 22, 5, 469-487.

Tallal, P. (1980). Auditory temporal perception, phonics, and reading disabilities in children. In *Brain and Language,* 9, 182–198.

Tallal, P. (1990). Fine-grained discrimination deficits in language-learning impaired children are specific neither to the auditory modality nor to speech perception. In *Journal of Speech and Hearing Research,* 33, 616–617.

Tallal, P., Miller, S. L., Bedi, G., Byma, G., Wang, X., Nagarajan, S.S., Schreiner, C., Jenkins, W. M., and Merzenich, M. M. (1996). Language comprehension in language-learning impaired children improved with acoustically modified speech. In *Science*, 271, 81-84.

Tanaka, J. W, Wolf, J. M., Klaiman, C., Koenig, K., Cockburn, J., Herlihy, L., Brown, C., Stahl, S., Kaiser, M.D., and Schultz, R.T. (2010). Using computerized games to teach face recognition skills to children with autism spectrum disorder: The Let's Face It! program. In *Journal of Child Psychology and Psychiatry,* 51, 944-952.

Teixeira, C., Trancoso, I. and Serralheiro, A. (1997). Recognition of non-native accents. In *Proceedings of EUROSPEECH-1997*, 2375-2378.

Temple, E., Poldrack, R.A., Protopapas, A., Nagarajan, S. Salz, T., Tallal, P., Merzenich, M.M., and Gabrieli, J.D.E. (2000) Disruption of the neural response to rapid acoustic stimuli in dyslexia: Evidence from functional MRI. In *Proceedings of the National Academy of Sciences,* 97. 25, 13907-13912.

Thal, D.J. and Tobias, S. (1992). Communicative gestures in children with delayed onset of oral expressive vocabulary. In *Journal of Speech and Hearing Research,* 35, 1281–1289.

Thomas-Stonell, N., Kotler, A. L., Leeper, H. A., and Doyle, P. C. (1998). Computerized speech recognition: Influence of intelligibility and perceptual consistency on recognition. In *Augmentative and Alternative Communication,* 14, 1, 51–56.

Togler, J., Hemmert, F., and Wettach, R. (2009). Living interfaces: The thrifty faucet. In *Proceedings of the 3rd International Conference on Tangible and Embedded Interaction* (TEI'09), 43-44.

Torgesen, J. and Bryant, P. (1994) Test of Phonological Awareness. Pro-Ed, Austin, TX.

Treviranus, J., Shein, F., Haataja, S., Parnes, P., and Milner, M. (1991). Speech recognition to enhance computer access for children and young adults who are functionally non-speaking. In *Proceedings of 14th Annual RESNA Conference,* 308–310.

U.S. Department of Education. (2005). *To Assure the Free Appropriate Public Education of all Americans: Twenty-Seventh Annual Report to Congress on the Implementation of the Individuals with Disabilities Education Act.*

United Nations. (2004). *World Programme of Action Concerning Disabled Persons: Current Situation.* Available on http://www.un.org/esa/socdev/enable/diswpa04.htm.

Vicsi, K., Roach, P., Öster, A. M., Kacic, Z., Barczikay, and Tantoa, A., Csatári F., Bakcsi Zs. and Sfakianaki A. (2000). A multimedia, multilingual teaching and training system for children with speech disorders. In *International Journal of Speech technology*, 3, 289-300.

Walker, W., Lamere, P., Kwok, P., Raj, B., Singh, R., Gouvea, E., Wolf, P., Woelfel, J. (2004). *Sphinx-4: A Flexible Open Source Framework for Speech Recognition.* Sun Microsystems.

Weiner, F. F. (1981). Treatment of phonological disability using the method of meaningful minimal contrast: Two case studies. In *Journal of Speech and Hearing Disorders,* 46, 97-103.

Weiser, M. (1999). The computer for the 21st century. In *Mobile Computing Communication Review*, 3, 3, 3-11.

Weismer, S. E. (1997). The role of stress in language processing and intervention. In *Topics in Language Disorders*, 17, 4, 41–52.

Wetzel, K. (1996). Speech-recognizing computers: A written communication tool for students with learning disabilities? In *Journal of Learning Disabilities,* 29, 371–380.

Wiederholt, J. L. and Bryant, B. R. (1992) *Gray Oral Reading Test: 3rd edition.* Pro-Ed, Austin, TX.

Wiepert, S.L., and Mercer, V.S. (2002). Effects of an increased number of practice trials on Peabody Developmental Gross Motor Scale scores in children of preschool age with typical development. In *Pediatric Physical Therapy*, 14, 22-28.

Wiig, E. H. (1990). *Wiig Criterion Referenced Inventory of Language*. Psychological Corporation, New York, NY.

Wildner, S. (2002). Learn German Now! Version 8, Software review. In *CALICO Journal*, 20, 161-174.

Wilson, M. S., and Fox, B. J. (1997). *Micro-LADS [Computer software].* Laureate Learning Systems, Winooski, VT.

Witt, S. M., and Young, S. J. (2000). Phone-level pronunciation scoring and assessment for interactive language learning, In *Speech Communication*, 30, 95-108.

Woodcock, R. W. (1987). *Woodcock Reading Mastery Tests – Revised*. American Guidance Services, Circle Pines, MN.

World Health Organization, (2001*), International Classification of Functioning, Disability and Health* (ICF). World Health Organization, Geneva, Switzerland.

Young B. A., and Burrill, D. F. (1997). Correlates of a test of motor symbol sequencing performance. Poster presented at the 105th APA Annual Convention.

Zavaliagkos, G., Schwartz, R., and Makhoul, J. (1995). Adaptation algorithms for BBN's phonetically tied mixture system. In *Proceedings of ARPA Spoken Language Systems Technology Workshop,* 82-87.

Zhao, Y. (2007). Speech technology and its potential for special education. In *Journal of Special Education Technology*, 22, 3, 35-41.

Zuckerman, O., Arida, S., and Resnick, M. (2005). Extending tangible interfaces for education: Digital Montessori-inspired manipulatives. In *Proceedings of the 21st Conference on Human Factors in Computing Systems* (CHI '05), 859-868.

Zyda, M. (2005). From visual simulation to virtual reality to games. In *IEEE Computer*, 38, 9, 25-32.