Virtual Humans that Conduct Speech Therapy

There are compelling reasons to explore the potential of virtual humans to conduct speech, language and other behavioral therapies. The reality is that proven clinical treatments developed in the laboratory are difficult to scale up to the real world. Reasons for this situation are varied and complex, but include the following issues: (a) knowledge about how to conduct the therapy is vested in relatively few individuals, and these individuals are typically researchers who focus on developing and demonstrating the effectiveness of their treatments. Efforts are made to disseminate the information however training a sufficient number of clinicians to conduct new treatments while assuring both fidelity and quality is a major effort. Moreover, there are few programs in place that support transfer of behavioral treatments from laboratories to clinics or hospitals; (b) effective treatments for individuals with neurological disorders who are acquiring and/ or recovering speech and language skills are typically both intensive and extensive, requiring many sessions, lasting an hour or longer, conducted over weeks or months. In the current healthcare climate, reimbursed face to face treatment is usually insufficient in terms both of its intensity and length; (c) there are significant barriers to introducing behavioral treatments for neurological diseases and disorders, including informing and convincing physicians about the value of the treatment, and developing reimbursement models that assure access to treatment by patients; (d) even if all of these problems could be solved, and every physician miraculously recommended an efficacious, reimbursable speech therapy to those patients with a neurological disorder likely to benefit from the treatment, a residual problem would remain of training a sufficient number of therapists to treat all of the patients in need of treatment.

One solution to this problem is to develop computer-based *virtual therapy systems* that provide accessible, inexpensive, engaging and effective treatment to individuals on standalone or networked computers. Initially, these systems would most likely be available in hospitals and clinics, with certified clinicians conducting assessment and orientation sessions, and monitoring patients during or following use of the virtual therapy system. Within a few years, we believe that systems will be accessible via the internet for home use, with a substantial number of clinician functions, including ongoing assessment and treatment capable of being conducted by the virtual therapist. Human clinicians will still play an essential role by monitoring sessions, addressing potential problems (most of which will be identified automatically by the system) and interacting with patients during and following sessions. In this scenario, patients and clinicians are working together to optimize treatment outcomes and financial gain (for the clinician, through economies of scale), and for society through less expensive and more accessible treatments and increased function and quality of life for millions of communicatively disabled individuals.

Beyond the major potential benefits to patients and society, virtual therapy systems provide new and unprecedented opportunities for experimental studies that can provide deep insights into the nature of the therapeutic process; that is, to manipulate and study the patient and client behaviors that lead to successful treatment outcomes. For example, we might hypothesize that therapeutic outcomes are enhanced not only by the fidelity of treatment (the extent to which the clinician faithfully adheres to the principles and practice of the treatment) but also by the degree of rapport and trust that is established during treatment sessions. Variables of interest that could be manipulated experimentally using a virtual therapist and virtual therapy system to establish rapport might include frequency of feedback, response rate, eye gaze, facial expressions and communicative gestures (Saxon and Schneider, 1995; Schmidt and Lee, 1999; Borod et al., 1985; McFarland, 2001; McNeil,1992; Duncan et al., 1995). Rapport could be measured by independent observations and analysis of patient-therapist gestures as well as self reports. The use of a virtual therapist, programmed to interact with the client in specific and well defined ways, provides the opportunity to manipulate and model the underlying clinical process. This is difficult and perhaps impossible to do with human experimenters for some variables, as it is well known that gestures and other nonverbal behaviors are typically unconscious. It is therefore complex (and costly) to conduct experiments in which human experimenters attempt to manipulate posture, head movements, facial expressions, gaze and hand and arm gestures in different experimental conductions. Virtual humans, however, can be programmed to produce specific behaviors in response to patient behaviors. Thus, virtual therapists provide unprecedented opportunities to conduct controlled experiments on patient- clinician interaction, while measuring patient responses (behavioral, physiological) and attitudes toward the therapist and therapy, and effects of these interactions on treatment outcomes.

An important first step toward this new paradigm for investigating changes in social dynamics during therapy is to develop virtual human therapy system and demonstrate its effectiveness in clinical trials relative to therapy conducted by human clinicians. In the remainder of this article, we describe a system under development at CSLR, in collaboration with researchers at the National Center for Voice and Speech that is designed to conduct speech therapy with individuals with Parkinson disease.

A Virtual Speech Therapist for Individuals with Parkinson Disease

In the United States alone, over 20 million adults have a neurological disorder that may impair their ability to communicate orally (Yorkston et al., 1988). The prevalence of disordered communication is particularly high in the over six million individuals worldwide diagnosed with Idiopathic Parkinson disease (IPD). At least 89% of these individuals have disordered speech and voice (Logemann et al., 1978); however, only 3-4% of these individuals receive speech treatment (Hartelius et al., 1994). With IPD predicted to increase significantly as the general population ages, there is a growing need for an accessible, inexpensive, effective treatment for disordered communication for these individuals.

Research by Professor Lorraine Ramig and her colleagues at the National Center for Voice and Speech (NCVS) at the Denver Center for Performing Arts (DCPA) has a long history of published efficacy data (see Fox et al., 2002 for a review) and has generated the first short- and long-term experimental efficacy data for a speech treatment for individuals with IPD, including the first Type 1 evidence for efficacious speech treatment for Parkinson disease (Ramig et al., 2001; Goetz, 2003).

The speech of individuals with PD is characterized by a raspy or hoarse voice, by monotonic and unemotional (non-expressive) speech, and is produced with insufficient volume. Changes in voice are often one of the first symptoms of Parkinson disease noticed by others, and progressive loss of communication skills by individuals with IPD can be one of the most devastating effects of the disease; as one patient said, *"When you don't talk loud, other people stop listening."* After

LSVT, patients make comments such as: *“My voice is alive again!” “I can talk to my grandchildren on the phone for the first time in years!”*

In addition to documented long-term improvements in the speech intelligibility and communication abilities of these patients, and their post-treatment reports of increased self esteem and self confidence in social situations, research has documented that individuals who have undergone LSVT treatment benefit from additional acoustic, aerodynamic, and physiologic changes associated with treatment, including positive changes in phonation, articulation, swallowing, and facial expression (Ramig et al., 1995; Ramig & Dromey, 1996; Smith et al., 1995, Ramig et al., 1996; Ramig et al., 2001; Dromey et al., 1995; El Sharkawi et al., 2002; Spielman, et al., 2001). Preliminary imaging results with Positron Emission Tomography have identified post-speech treatment changes consistent with improved neural functioning (Liotti et al., 1999; 2003; Narayana et al., 2005).

<u>The Need for Virtual LSVT Therapy:</u> Despite the major progress described above, the “real world” treatment of speech and voice remains an unmet need for the vast majority of individuals with IPD. The current efficacious dosage of LSVT is 16 individual treatment sessions in one month with an LSVT-certified speech clinician. Despite impressive efforts to train certified LSVT clinicians by Dr. Ramig and her colleagues at the LSVT Foundation, there are today only about 2500 certified clinicians in Australia, Europe, Latin America and the United States, thus most individuals with IPD worldwide do not have access to certified LSVT clinicians. For many individuals with IPD, access to treatment is not feasible due to physical disability, geographic barriers or financial reasons; in the U.S. and elsewhere, lack of consistent or comprehensive insurance reimbursement for speech therapy is an obstacle to treatment. Computer-based speech therapy presents a potential solution to this problem by providing access to LSVT through a virtual LSVT speech therapist that interacts with patients, initially in clinics and eventually in their homes, to provide an engaging and efficacious treatment.

The community of LSVT practitioners recognizes the need for the proposed LSVT-VT system, and clinicians who have been informed about or seen videos or demonstrations of the initial LSVT-VT laboratory prototype have been uniformly positive. Indeed, one-hundred percent of the clinicians surveyed at the 2004 American Speech Language Hearing Association convention expressed interest in employing the LSVT-VT in their clinical practice. Comments such as this one from an LSVT clinician were representative: “I have treated over 100 individuals with LSVT and have had remarkable results. However, because the treatment is so intensive, I can only treat five patients per month. The LSVT-VT is a hopeful and I believe realistic solution that will allow me to treat many more patients. I am happy to commit to testing your system during development and eager to integrate the system into my daily practice when it becomes available.”

In May 2004, with support from grants from the U.S. NIH NIDCD and the Coleman Institute of Cognitive Disabilities, we began development of an LSVT Virtual Therapist (LSVT VT). At the time of this writing, we have spent one year designing, developing and testing the LSVT VT system, and have just completed LSVT sessions with a 74 year old female with Parkinson disease, who used the LSVT VT as part of her therapy. In the remainder of this section, we describe the process of developing the LSVT VT system, the challenges encountered and insights gained, and initial results from our first patient.

<u>The Practice of LSVT</u>**:** LSVT treatment focuses on teaching individuals to think and talk LOUD. The rationale for the mode and frequency of exercises and delivery of treatment are based upon the underlying physiology for the disordered sensory motor function and the neuropsychological profile of these individuals. These features together with principles of motor learning and skill acquisition form the rationale for the LSVT exercises, their dosage and mode of administration. This rationale has been detailed in Fox, Morrison and Ramig (2002) and the details of the treatment have been summarized in Brin, Vilecovic, Ramig, & Fox (2004); Ramig, Countryman, Fox & Sapir (2001) and Ramig, Fox, and Sapir (2004).

The 16 LSVT treatment sessions are administered to an individual with IPD by a certified LSVT clinician in four individual sessions a week for four weeks. The first half (approximately 30 minutes) of every treatment session includes three daily tasks—maximum sustained vowel duration, maximum fundamental frequency range, and maximum functional phrases. In task 1, *maximum sustained vowel duration*, the patient is taught to produce and sustain the vowel "ah" with good vocal loudness. The LOUD "ah" exercises provide the foundation for correct production of loud speech, and serve as a stepping stone for bringing increased loudness into everyday speech production. The clinician stimulates the patient to produce a sustained "ah" sound, and then follows a systematic process of shaping the patient's "ah" into a loud good quality healthy sounding voice. A unique aspect of LSVT is that shaping is achieved by minimizing cognitive load, thus, the clinician gives simple instructions: "Do what I do" and then models the target behavior. In task 2, *maximum fundamental frequency range*, the patient learns to generalize the feeling of increased loudness learned in the LOUD "ah" task to using increased loudness while changing pitch either higher or lower. Subjects are instructed to start with an "ah" and perform a minimum of 15 repetitions stretching high in pitch and similarly, to start with an "ah" and perform a minimum of 15 repetitions stretching low in pitch. Once again, the therapist will stimulate, shape and stabilize loudness and quality of voice, while the individual is performing both high and low pitch changing exercises. Task 3, *maximum functional phrases*, is designed to bring the increased amplitude of movement achieved with loudness in sustained "ahs" and high/low "ahs" into speech. Each patient generates a list of 10 functional phrases that he or she says every day (e.g., "What's for dinner" "I love you."). During treatment, the patient reads these phrases using his or her LOUD "ah" voice. These 10 phrases are read five times every treatment session. The repetitions are used to shape a good quality LOUD voice, to increase ease of using a LOUD voice, and to over learn loudness in these everyday functional phrases.

The second half of each session (approximately 20-30 minutes) includes a progressive hierarchy of speech tasks. In addition, embedded throughout every session are activities designed to train sensory calibration, in which the individual learns to calibrate their perceived vocal effort with the LOUD voice they are supposed to produce. The goal of these tasks is for the patient to learn to apply their LOUD voice to increasingly complex speech tasks that require more cognitive effort and greater degrees of endurance to maintain loudness for longer and longer periods of time. Week 1 of treatment focuses on words and phrases; Week 2 of treatment focuses on sentences; Week 3 of treatment focuses on continuous reading, and Week 4 of treatment is focused on conversational speech, and when possible, occurs outside of the therapy room. The materials used for the hierarchy change from day-to-day and gradually build in complexity and duration.

Sensory Calibration: One of the greatest challenges of LSVT therapy is sensory calibration, or generalization of loud speech from the clinic to the real world. Individuals with PD believe they are speaking much louder than they are. This problem has been associated with both motor and sensory disorders (Fox et al., 2002). When stimulated to speak louder, patients often complain "I can't talk like this – it feels too loud" or "People will think I am crazy if I talk like this." The goal of sensory calibration is to enable patients to learn to overcome the mismatch between their perceived vocal effort and loudness of their speech and the actual loudness of their speech. Calibration exercises aim to adjust the patient's sensorimotor perception of vocal loudness and effort so that what initially felt and sounded too loud is now perceived to be within normal limits. As treatment progresses, the increased effort and loudness becomes more automatic. Calibration assessment and exercises are embedded into every treatment session. Calibration assessments are made during informal conversations with the therapist during sessions. Activities include: helping patients recognize the need for louder speech (i.e., their speech is too soft); convincing patients that their loud speech is within normal limits; and making them comfortable with using louder speech in daily living. To achieve sensory calibration, the LSVT VT system uses a number of exercises that require continuous speech production, with patients receiving feedback and encouragement for speech produced at targeted loudness levels. Exercises include "faux" spoken dialogue interactions with the Virtual Therapist, during which the VT asks open-ended questions in engaging, individualized topics of interest to the patient, such as how they prepare favorite meals, or what they like about certain books, etc. The system uses speech recognition to track the person's speech, compute loudness, and provide feedback and encouragement. An important part of calibration is generalization of the new voice outside of the treatment room to the patient's real world environment. Day one through day sixteen of treatment, patients are given assignments to practice their voice exercises at home and to use their loud voice in functional situations (i.e. when ordering meat at the deli counter, when talking to their children on the phone).

Designing the LSVT VT system

The main tasks leading to development of the LSVT VT system prototype included (a) specifying and implementing the "rules of engagement" for controlling the verbal and nonverbal behaviors of the virtual therapist in response to vocalizations produced by the patient; (b) designing the multimedia human computer interface for each of the five LSVT exercises; (c) testing and refining the system through iterative design-and-test cycles; and (d) incorporating the LSVT VT within the treatment sessions for an initial patient with Parkinson disease. Development of the initial LSVT VT system (tasks a-c) occurred over an eleven month period, and included collaboration between researchers at CSLR and NCVS and LSVT "Vets"; individuals with Parkinson disease who had previously completed LSVT, and worked with the research team to test the system and provide detailed feedback about their experiences. The final task—integration and evaluation of the LSVT VT prototype into LSVT therapy with an actual patient—was conducted at NCVS by an expert LSVT clinician who was not involved in the system development process.

Developing Rules for Governing Interaction between the Patient and Virtual Therapist: To develop the LSVT VT, it was necessary to (a) specify and codify the rules that clinicians use to interpret patient's verbal behaviors during LSVT tasks, and (b) to understand and emulate, via the Virtual Therapist, the verbal and nonverbal behaviors that clinicians produce in response to patients' vocalizations.

The first step in the development process was to specify and codify the rules that govern clinician responses to patient vocalizations in each LSVT task. For example, in the sustained "ah" task, the clinician's response to the patient depends on the loudness and duration of the patient's vocalization relative to prior vocalizations (if any). Fortunately, these rules were already developed in the context of a previous project that led to the design of an "LSVT Companion," a hand held device that measured and stored data regarding the loudness, duration and pitch of a patient's utterance as it guided them through LSVT exercises. The LSVT Companion also provided auditory as well as visual feedback to the patient using computer graphics presented on the screen of the device (Halpern et al. 2004). We modified these rules as needed for the LSVT VT.

Our efforts focused on providing patients with immersive and effective experiences for each of the five LSVT exercises. We attempted to do this by providing highly informative real time feedback to patients during the exercises and excellent modeling behavior by the virtual therapist.

Once the rules had been implemented, significant effort was devoted to (a) designing computer displays that provide accurate and intuitive feedback to the patient about his or her vocalizations, and (b) designing the verbal and nonverbal behaviors of the Virtual Therapist. The Virtual Therapist's behaviors were divided into those that occurred during the patient's speech productions, and those that occurred after the patient's speech productions. The design of animated feedback displays was guided by principles of multimedia design in learning tasks (Mayer, 2001) and prior work with the LSVT Companion. In addition, we investigated a variety of novel dynamic displays for providing real time feedback to patients about their vocalizations and also motivating them to achieve specific goals. These displays, and the verbal and nonverbal behaviors of the Virtual Therapist, were then tested and refined through several design-and-test cycles by members of project team and LSVT Vets. This process resulted in a complete LSVT VT system that was tested by a 74 year old woman with Parkinson's disease during May 2005. Our goal was to enable the patient to independently complete all five LSVT exercises during a one hour session, and provide the patient with an engaging and satisfactory experience. The system achieved this goal in the final session, which included frequent breaks controlled by the Virtual Therapist during which the patient would drink water and converse informally with the virtual therapist (measurements of loudness of the patient's speech were made during these dialogs).

Figures 1a-e show the system interface, respectively, for training sustained loud /ah/ phonation and for training /ah/ phonation with rising and falling pitch patterns, spoken functional phrases, reading out loud, and spoken dialog interaction. A video showing system use by Ms. Halpern can be viewed via the CSLR Web Site (LSVT-VT-DEMO 2004). For example, in sustained "ah" phonation, the patient's voice causes a car to travel forward if the patient's vocalization meets or exceeds the loudness goal set by the clinician, indicated by a horizontal line on the display. The car continues to move forward as long as the person maintains this loudness level. The patient attempts to achieve the goal of moving the car a certain distance past the distance achieved on the last trial, indicated by the vertical lines in the figure.

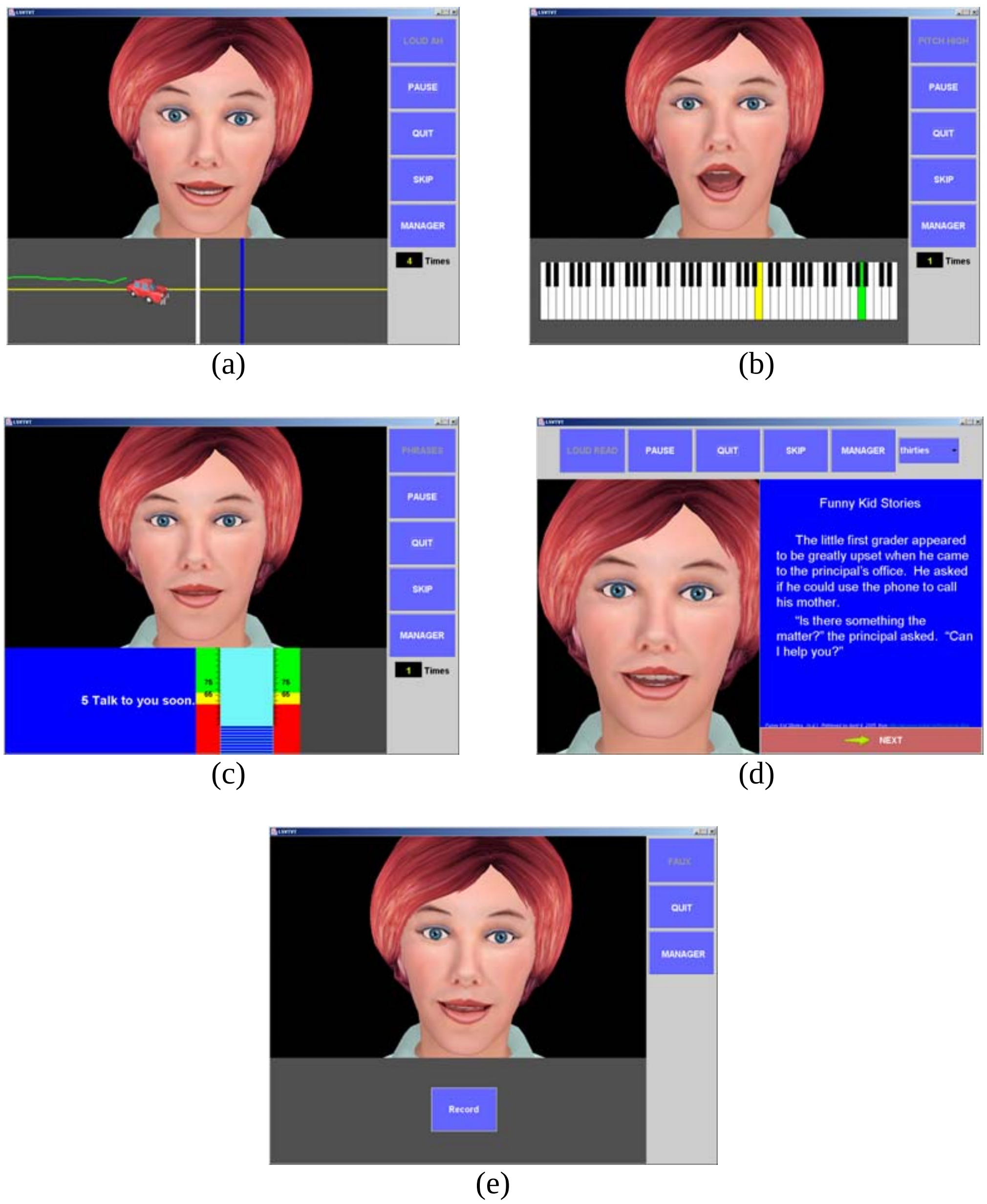

Figure 1. LSVT-VT system interface for (a) training sustained loud /ah/ phonation and for (b) training /ah/ phonation with rising and falling pitch patterns (c) spoken functional phrases, (d) reading out loud (e) spoken dialog interaction.

While the patient vocalizes, Marni provides encouragement by smiling and nodding her head and saying phrases such as "go go go". Following the trial, Marni provides feedback to the patient based on their performance; e.g., Marni congratulates the patient for achieving their goal, or encourages the patient if they didn't, and then models the desired behavior. The verbal behaviors, head movements and facial expressions produced by Marni were designed by the research team based on analysis of video tapes of LSVT clinicians during LSVT sessions and based on feedback of the LSVT Vets.

Feedback from LSVT Vets: Five LSVT Vets—individuals with IPD who completed LSVT—served as consultants on our project. Tester 1 is 61 years old, tester 2 is 71 years old, tester 3 is 56 years old, tester 4 is 58 years old, and tester 5 is 83 years old. The LSVT Vets played a key role in participatory design of the LSVT VT system through their many important suggestions and insights that led to significant changes in the system. They were perceptive, forthcoming and articulate in expressing their opinions, criticisms and suggestions for improving the system. Patients provided specific suggestions on the feedback displays for specific tasks and Marni's verbal and nonverbal behaviors. Some examples of suggestions in August 2004 included: (a) On Marni: Tester 01 felt that the agents were looking above or below her, but not at her. She also felt that the eyes were not "real" as though she was looking at a glass eye. Tester 02 thought the agent should be female and younger in their 20's or 30's, he thought this type of agent would seem more upbeat." (b) On sustained "ah": Tester 01 liked the idea of the car (moving forward), but would prefer to have a photo of a real object rather than an animated object. Tester 02 liked having the blue line appear from the previous phonation, so that he had a goal of what to beat for the next phonation. Tester 02 and 03 liked to have the car stop if the person falls below the target yellow SPL line. Tester 01 and 04 wanted to have the car continue to move if it fell below the target SPL line. (c) On High and Low- Pitch change: Tester 01 did not like the car with headlights, she thought it was too confusing. She liked a piano, but thought we could make some improvements. Tester 02 liked the car better than the piano. Testers 03 and 04 liked the keyboard best. They did not like the car. They liked the keyboard because it is intuitive.

A second testing session was conducted in November 2004 using an LSVT system that incorporated many of the LSVT vets' previous suggestions. In addition to recording and implementing many new suggestions from this testing session, we administered a survey to 4 of the LSVT Vets after testing the system. The results of the survey, summarized in Table 1, reveal that the 4 testers were uniformly positive in their ratings of the system.

A few additional comments made by the LSVT vets about improvements to the Virtual Therapist from the August to November session include:

> *Like the addition of the light in her eyes, she looks more alive. Like the more animation of face. Eyebrows are better, eyes are good. More body movement than before.*

And comments made regarding their opinion of the importance of this project include:

> *Very important from a patient standpoint, as a motivator, as a tool to measure progress.*
> *Multiplying- can help 100 people at a time, no limit.*
> *If had this at home would practice more.*

Table 1. Results of User Testing of Initial LSVT-VT system.

		Rating					
		1	**2**	**3**	**4**	**5**	
Question	**1 =**	**# of testers rating**					**5 =**
How easy was this to use?	Not easy			1	1	2	Easy
How fun was this to use?	Not fun				2	2	Very Fun
Please rate the appeal of the program	unappealing				1	3	Very appealing
Please rate how engaging you found the virtual therapist to be.	Not engaging		1	2	1		Very engaging
Please rate how much you liked the feedback provided by the virtual therapist.	Not at all			2	1	1	Very Much
Please rate how much the virtual therapist helped you.	She was not helpful				3	1	She was very helpful
Did the virtual therapist make you feel frustrated?	Very frustrated				1	3	Not at all frustrated
Does the computer make it difficult for you to focus on your speech exercises?	Not able to focus				1	3	Very able to focus
How much did you have to think about what the program wanted you to do?	Required constant thought				2	2	Required no thought at all
Please rate how much you liked or disliked the program.	Disliked greatly				1	3	Liked very much
Does the computer help you to focus on your speech exercises?	Does not help				2	2	Is very helpful
How easy is it to interpret the visual display at the bottom of the screen?	Not easy				1	3	Very easy
Is the verbal feedback useful? (e.g. the virtual therapist saying, “good job”).	Not useful				4		Very useful
Is the visual feedback at the bottom of the screen useful?	Not useful				2	2	Very useful
Please rate how motivated the program’s verbal and visual feedback made you feel.	Very unmotivated				3	1	Very motivated
Were you embarrassed to use this program?	Very embarrassed				1	3	Not at all

First Tests of the LSVT VT System in a Clinical Setting

During May 2005, we incorporated the LSVT system into LSVT treatment sessions conducted by an LSVT clinician at NCVS with a 74 year old female patient. This patient was 8 years post diagnosis and presented with mild IPD, characterized by dyskinesias and tremor in her right dominant hand. Fig X shows a picture of the patient (VT P1 using the system. VT P1 did not own a computer. Her previous computer experience was limited to two classes that she had taken at a library to learn to use the library computer system.

System Performance and Improvements During Testing: Our first test of the LSVT VT was designed to provide an initial evaluation of the system, to identify and fix problems, and to gauge system usability and the patient's experiences. The patient was informed that we were introducing a new computer system as part of the therapy and that we expected to discover some problems with the system. The patient understood the situation perfectly, and was tolerant and helpful at all times. The LSVT VT system was first introduced during the fourth session at the end of the first week of treatment and the patient did a few exercises utilizing the system. The entire fifth session was conducted using the LSVT VT system with the therapist sitting next to the patient to provide help as needed. Unfortunately, problems were encountered in each exercise; the researcher who observed the session presented a list of 16 program changes. Session 6 was conducted with the therapist only. In session 7, the patient used the system on her own for the whole session with minimal help from the therapist. The number of suggested changes was 5. Sessions 8 and 9 were conducted by the therapist. In session 10, the client worked independently; the therapist intervened to help the patient click on icons to move to new exercises (the patient had difficulty controlling the mouse due to lack of previous computer experience and tremor) and to shape the patient's speech when poor quality utterances were detected. After this session, 9 changes to the system were suggested, but these reflected more subtle improvements, such as "The clinician commented that the agent needs "more non verbal smiles". Sessions 11 and 12 were conducted with the LSVT VT system. The patient used the system successfully during these sessions, although several new problems were discovered, and a total of 8 specific suggestions for program changes were generated. Session 13 was conducted with the therapist only. Session 14 was conducted with the LSVT VT system with minimal clinician intervention and one suggested change. Session 15 was conducted with the LSVT VT system. The system performed well and the patient worked independently with the LSVT VT the entire session without the therapist in the room. Session 16 was conducted with the therapist only. Throughout all sessions conducted with the LSVT VT the therapist spoke with the patient at the beginning and end of each session to review her homework exercises and assign new homework for the next day. Immediately after session 15 the patient was interviewed.

Patient Interview: A ten minute interview was conducted by LSVT researcher Angela Halpern immediately after the 15th session. A video of the interview can be obtained from the CSLR Web site (LSVT-VT-DEMO, 2004), as well as video clips showing the speech of the patient before and after LSVT. One of the most salient features of the interview is the patient's voice—it is loud and clear at all times while answering Angela's questions, without any cues from Angela to maintain loudness. The patient's responses were informative and very positive, especially considering the many problems she encountered while using the system; we believe her responses bode well for the future of LSVT VT, since the program will improve steadily in the near future. Informative sections of the interview are presented here:

Angela: *In general, after doing the session with Marni how do you feel, how does Marni make you feel as you're going through the session?*
VT P1: *It's somewhat different. You know, it's nice to have a live person, but I feel that I really interact with Marni. Somehow I have a sense of a personality.*
Angela: *Oh good, good. So when you said you said you feel a sense of a personality with Marni can you elaborate a little bit?*
VT P1: *I think the prompting that she gives me and the encouragement that she gives me and the way I get aggravated with her (laughing) when she keeps wanting to know what steps I take to get ready for bed and to get ready to meet the day in the morning.*

Angela: *Do you feel like Marni is encouraging as you're going though the therapy?*
VT P1: *Yes. Of course, not to the same degree that I feel it with Jill*.*
Angela: *Okay. And what would be the difference in the encouragement you get from Jill versus the encouragement you get from Marni?*
VT P1: *Other than Jill is a live person with a lot of personality, and I think that there's no way Marni could sparkle that way.*
Angela: *Alright, good. What elements do you like most about Marni's personality?*
VT P1: *Personality. Well, this isn't personality, but I like the way she looks. Uh, personality does seem to come through in her eyes especially. I'd have to think about personality some more.*

Angela: *But you do get a sense of a personality with her?*
VT P1: *Yes, I do because I feel the urge to talk back.*
Angela: *Okay, alright, what elements do you like least about her personality?*
VT P1: *I don't think of anything that about her personality that rubs me the wrong way or feels irritating.*

Angela: *So how do you feel in general, after you've finished a session with Marni versus after you finish a session with Jill?*
VT P1: *In both instances I feel as though I've made progress. And, I almost want to say elated…*
Angela: *Good.*
VT P1: *…Because of a several reasons. I sense such a difference in my voice and can hear it as well. And the um, sense of having done something that will continue to be of great benefit.*
Angela: *Good. Do you feel the same sense of effort when you're working with Marni versus with Jill?*
VT P1: *Yes. She keeps me on my toes.*
Angela: *Do you feel more effort with one versus the other, or is it pretty much the same?*
VT P1: *I would say the same.*
Angela: *Okay, okay.*
VT P1: *Yes, I feel motivated to do my best with either.*

Lessons Learned and Future Challenges

At first glance, development of LSVT using a virtual therapist appears straightforward—LSVT is based on well established principles and well articulated procedures, and these procedures have been validated by over a decade of efficacy data, and by the ability of LSVT experts to train and certify clinicians who are then able to conduct LSVT therapy and report successful outcomes with their patients. The clinician measures the loudness (and/or pitch and/or duration) of each vocalization and provides feedback to the patient, either by reinforcing the patient when the desired behavior is achieved, or, when the patient's speech production is deficient in some way, by modeling (producing) the behavior that the patient is expected to produce. Moreover, patients' vocalizations can be measured for loudness and duration with sufficient accuracy using commercially available instruments (a SPL meter and stop watch), and rules have been specified to guide clinicians' responses to patients' vocalizations in terms of these measurements.

Despite the elegance and apparent simplicity of LSVT, and some initial success with our first patient, we learned that inventing an LSVT Virtual Therapist is not as straightforward as it first appeared to be, and that the eventual system we would like to build requires advances in some technologies. One of the most interesting findings of the first year of our research effort was the discovery, through analysis of video taped sessions of LSVT and in depth discussions with the LSVT researcher clinicians, of the subtle yet important perceptual judgments made routinely during the course of treatment, and how these judgments affect clinician responses to patient behaviors. For example, during initial sessions of LSVT, the clinician "shapes" the patient's vocalizations, especially in the context of sustained "ah" exercises, thus producing consistent loud, high quality patient vocalizations. During initial sessions and throughout the course of treatment, clinicians *continuously* judge the quality of every vocalization produced by the patient. Good vocalizations are loud and strong and steady in terms of both loudness and fundamental frequency. Bad vocalizations are characterized by several features, including a "too sudden" onset or bad attack, a strained or hoarse voice or irregular pitch, which can be manifest in variety of ways. LSVT clinicians are highly sensitive to voice quality during loud LSVT training, and also pay close attention to the *visual* behaviors produced by the patient, including relaxed and upright posture, seeing the patient take a deep breath before vocalizing, having an upright head position and a wide open mouth (during sustained "ah" production). Of course, these visual behaviors will correlate with acoustic measurements that can be estimated automatically using signal processing and pattern classification methods, but it is clear that computer vision has a role to play in future LSVT VT systems. We have observed, and the LSVT researchers have reported, that the models they produce to demonstrate desired visual and acoustic behaviors to patients are highly responsive to the behaviors the patients have just produced. Effective LSVT clinicians will exaggerate or emphasize "correct" behaviors they have observed on the previous speech production as a way of pointing out the behaviors they wish patients to continue to produce. Likewise, when the patient produces a sufficiently bad vocalization, the clinician will stop the patient from vocalizing and immediately model the desired behavior.

Emulating the complex perceptual judgments and social interactions of LSVT clinicians and their patients is beyond the capabilities of the technologies deployed in our current LSVT VT system. At present, the speech signal processing technologies in our LSVT VT system are limited to detecting loudness and fundamental frequency changes. Moreover, we are not using computer vision to detect important visual behaviors that clinicians commonly use to detect good and bad speech production behaviors. While we are confident that vocalization quality can be classified accurately from the acoustic signal once sufficient data have been collected from patients and annotated to reflect the different perceptual judgments made by clinicians, in the absence of these data we are currently unable to emulate the important perceptual judgments that clinicians use to categorize utterances as good or bad. Thus, future work must focus on accurately measuring the quality of patient's vocalizations using both speech recognition and computer vision, and developing effective interactive animation sequences for modeling and shaping high quality loud speech. Because these problems remain unsolved at present, LSVT VT can be used to provide an effective component of the LSVT regimen, but it can not yet provide a complete solution. An alternative solution that we are investigating is to have suitably qualified patients use the LSVT VT system after an initial LSVT session, and deal with voice quality issues by having the clinician monitor and respond to suspicious utterances, including those that are identified as problematical by the system.

References

Borod, J., Koff, E., Lorch, M., & Nicholas, M. (1985). Channels of emotional communication in patients with unilateral brain damage. Archives of Neurology, 42: 345-348.

Brin, M., Vilecovic, M., Ramig, L., & Fox, C. (2004) Dysphonia due to Parkinson disease: Pharmacological, Surgical and Behavioral Management Perspectives. In: Casper, J., and Sapienza, C., (Eds.) Vocal Rehabilitation in Medical Speech-Language Pathology. (pp 209-269) Austin: Pro-Ed.

Dromey, C., Ramig, L., & Johnson, A. (1995). Phonatory and articulatory changes associated with increased vocal intensity in Parkinson disease: A case study. Journal of Speech and Hearing Research, 38: 751-63.

Duncan, S., McNeill, D.M., & McCullough, K-E. (1995). "How to transcribe the invisible --- and what we see," KODIKAS/CODE Ars Semiotica, vol. 18 nos. 1-3, 75-94.

El-Sharkawi, A., Ramig, L., Logemann, J., Pauloski, B., Rademaker, A., Smith, C., Pawlas, A., Baum, S., & Werner, C. (2002). Swallowing and voice effects of Lee Silverman Voice Treatment: A pilot study. Journal of Neurology, Neuropsychiatry, and Psychiatry, 72(1): 31-36.

Fox, C., Morrison, C., Ramig, L, & Sapir, S. (2002). Current perspectives on the Lee Silverman Voice Treatment (LSVT®) American Journal of Speech Language Pathology, 11: 111-123.

Goetz, C. (2003). Personal communication.

Halpern, A., Matos, C., Ramig, L., Petska, J., Spielman, J., Bennett, J., Will, L. LSVTC- A PDA Supported Speech Treatment for Parkinson's Disease. Presented at the International Association of Logopedics and Phoniatrics, Brisbane, Australia, 2004.

Hartelius, L., & Svensson, P. (1994). Speech and swallowing symptoms associated with Parkinson's disease and multiple sclerosis: A survey. Folia Phoniatrica et Logopedica, 46(1): 9-17.

Liotti, M., Vogel, D., New, P., Ramig, L., Mayberg, H., Cook, C., & Fox, P. (1999). A PET study of functional reorganization of premotor regions in Parkinson's disease following intensive voice treatment (LSVT). Neurology, 52(6), Suppl .2, A348-9 (Abstract).

Liotti, M., Ramig, L., Vogel, D., New, P., Cook, CI., Ingham, R.J., Ingham, J.C., & Fox, P.T. (2003). Hypophonia in Parkinson's disease: neural correlates of voice treatment revealed by PET. Neurology, 60(3):432-40.

Logemann, J., Fisher, H., Boshes, B., & Blonsky, E. (1978). Frequency and co-occurrence of vocal tract dysfunctions in the speech of a large sample of Parkinson individuals. Journal of Speech and Hearing Disorders, 42: 47-57.

LSVT-VT-DEMO, 2004. The Web site: http://cslr.colorado.edu/beginweb/animated_speech_therapist/anim_therapist.html provides a demo of the LSVT-VT during treatment.

Mayer, R. (2001) Multimedia Learning. Cambridge, UK: Cambridge University Press.

McFarland, D.H. (2001). Respiratory markers of conversational interaction. Journal of Speech Language and Hearing Research, 44: 128-143.

McNeill (1992): Hand and mind: what gestures reveal about thought. Chicago, IL: The University of Chicago Press.

Narayana, S., Vogel, D., Brown, S., Franklin, C., Zhang, W., Lancaster, J., Fox, P. (2005). Mechanism of Action of Voice Therapy in Parkinson's Hypophonia - A PET study, Human Brain Mapping meeting in Toronto, June 12-16.

Ramig, L., Countryman, S., Thompson, L., & Horii, Y. (1995). A comparison of two forms of intensive speech treatment for Parkinson disease. Journal of Speech and Hearing Research, 38: 1232-1251.

Ramig, L., & Dromey, C. (1996). Aerodynamic mechanisms underlying treatment-related changes in vocal intensity in patients with Parkinson disease. Journal of Speech and Hearing Research, 39: 798-807.

Ramig, L., Countryman, S., O'Brien, C., Hoehn, M., & Thompson, L. (1996). Intensive speech treatment for individuals with Parkinson's disease: short- and long-term comparison of two techniques. American Academy of Neurology, 47: 1496-1504.

Ramig, L., Countryman, S., Fox, C., Sapir, S., (2001). Speech, voice and swallowing disorders. In: Factor, S., and Weiner, F., (Eds). Parkinson Disease: Diagnosis and Clinical Management. Demos Publishing: New York (NY). pp. 75-87.

Ramig, L., Sapir, S., Fox, C., Countryman, S. (2001). Changes in vocal intensity following intensive voice treatment (LSVT) in individuals with Parkinson disease: a comparison with untreated patients and noraml age-matched controls. Movement Disorders, 16, 79-83.

Ramig, L., Sapir, S., Countryman, S., Pawlas, A., O'Brien, C., Hoehn, M., & Thompson, L. (2001). Intensive voice treatment (LSVT®) for individuals with Parkinson's disease: A two year follow-up. Journal of Neurology, Neurosurgery & Psychiatry. 71, 493-498.

Ramig, L., Fox, C. and Sapir, S. (2004). Parkinson's disease: Speech and Voice Disorders and Their Treatment with the Lee Silverman Voice Treatment (LSVT). Seminars in Speech and Language. 25, 2, pp.169-180.

Saxon, K., & Schneider, C. (1995). Vocal exercise physiology. San Diego, CA: Singular Publishing Group.

Schmidt, R. and Lee, T. (1999) Motor control and learning: a behavioral emphasis. 3rd Edition. Champaign, IL: Human Kinetics.

Smith, M., Ramig, L., Dromey, C., Perez, K., & Samandari, R. (1995). Intensive voice treatment in Parkinson's disease: Laryngostroboscopic findings. Journal of Voice, 9: 453-459.

Spielman, J., Ramig, L., & Borod, J. (2001). Preliminary effects of voice therapy on facial expression in Parkinson's disease. Journal of the International Neuropsychological Association, 7(2): 244.

Yorkston, K., Beukelman, D., & Bell, K. (1988). Clinical management of dysarthric speakers. Boston: Little, Brown & Company.