Robust Speech Messaging on the Power Wheelchair Assistant System

Developers: Han Joo Chae, Jun Woo Park, Yunchan Paik, Ramakrishna Battala
Advisors: Asim Smailogic, Daniel P. Siewiorek

1. Introduction

1.1 Virtual Coach Simulator

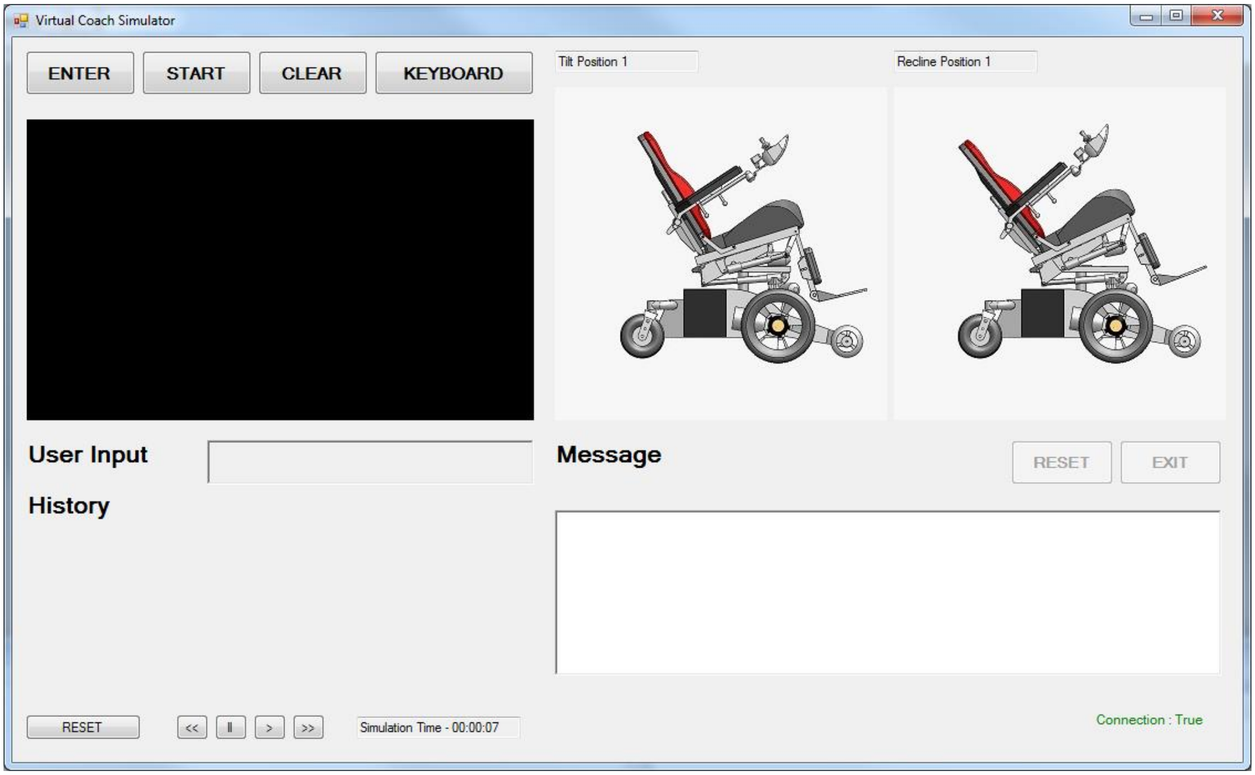
The Virtual Coach simulator is mainly developed to overcome the limitations of the existing Virtual Coach system of Carnegie Mellon University (CMU). The original Virtual Coach system was implemented on a powered wheelchair with a number of required hardware. It is obvious that the system lacks a few crucial aspects: portability and efficiency. Since the Virtual Coach device is not very portable, we would have to ship the device to the location beforehand every time we need to have a demonstration. In addition, the hardware components of Virtual Coach system are very expensive. It would cost much if the device gets damaged during the shipping process or during the demonstration. Therefore, the simulator software we have developed could liberate Virtual Coach from these constraints of portability and the dependency on the hardware including the cost from possible hardware malfunctions.

1.2 VoicePredict

The project includes not only porting the Virtual Coach system to a simulator software but also enhancement of the original system. The older version of Virtual Coach uses a joystick as a communication method between the user and the wheelchair. However, a joystick is very inefficient and requires a huge learning curve when the user wants to type a message. Other existing input interfaces for mobile devices can be considered such as 9-digit keypad, miniature keyboards and more recently touch-based inputs. These interfaces are being used by mobile users to enter text into applications like email, messaging, or internet browsing. It is widely acknowledged that these low-level input methods are "clumsy" and lack the speed, accuracy, and user-friendliness of a full-size keyboard. To address these problems, this project will incorporate a "predictive speech-to-text" prototype, as a new user interface paradigm for interacting with CMU's Virtual Coach. This technology has been incorporated into first iteration of a product, called "VoicePredict" by TravellingWave, which in turn claims to be one of the first multimodal mobile user experiences. Broadly voice prediction technology has the potential to become a ubiquitous interface for a variety of computing platforms including the personal computer, the embedded technology industry, and assistive technologies. The research project involves extending speech-to-text technology with voice prediction for integration with CMU Virtual Coach that involves assistive technologies. It is anticipated that the end product will result in a powered wheel chair with an inherent multimodal user interface that will significantly benefit the lives of the elderly and people with disability.

2. Design and Implementation Detail

The simulator has a great advantage in the sense that it is very portable and can be demonstrated in any circumstances. The simulator user interface (UI) basically consists of five major parts: the Virtual Coach section, the current state section, the user command section, the message section and the nurse interface as described in Figure 1 and 2. To integrate C# based Virtual Coach together with C++ based VoicePredict, Named Pipe has been adopted as an IPC (Inter-process Communication) solution.



Tilt Position 1

Recline Position 1

Simulation Time - 00:00:07

Connection : True

Figure 1. Overview of the Virtual Coach Simulator UI

2.1 Virtual Coach Section

The Virtual Coach Section performs as same as the original virtual coach system. It gives the user suggestions to move the wheelchair in a timely manner depending on the current state of the chair. The timing information for these actions can be set at when the simulator starts for the first time using the nurse interface which will be discussed in detail in the later section. The avatar gives suggestions both in text and audio. In addition, animations are provided so that the user can easily understand the recommended action. The user would then have the choice of accepting, snoozing or dismissing the suggestion given by the avatar.

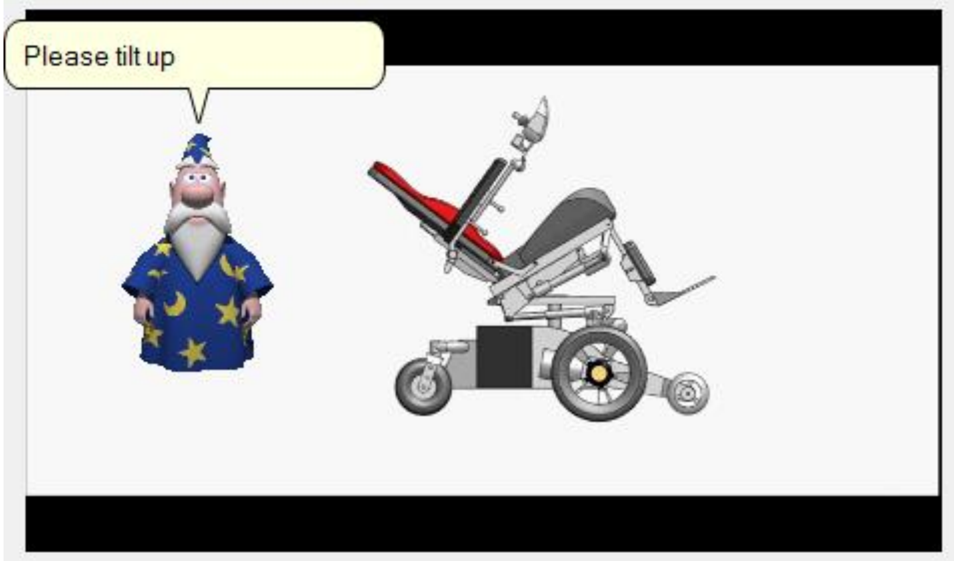


Figure 2. Avatar and Animation

Communication between a user and the simulator is done through voice commands using Push-to-talk functionality. The user has eight different commands to give when the avatar makes a suggestion: tilt up/down, recline up/down, message, mute, snooze and dismiss. Once the user gives any one of those commands, the system would move on to the next state corresponding to the given command. For example, the system will be in state #1 initially as shown in the state diagram below in figure 3, and when the user says “tilt,” the system will move on to the state #2 and then wait for the user commands, up, down or done. The system will stay in state #2 so that the user can freely tilt the wheelchair unless the user gives the command, “done.” When the user says “done,” it will go back to state #1 and wait for another user command. Once the user satisfies the suggested condition or gives a dismiss command, the system becomes state #4. In this state, only the tilt and recline commands are enabled for the user to freely move the chair. Detailed state transitions are described in Figure 3.

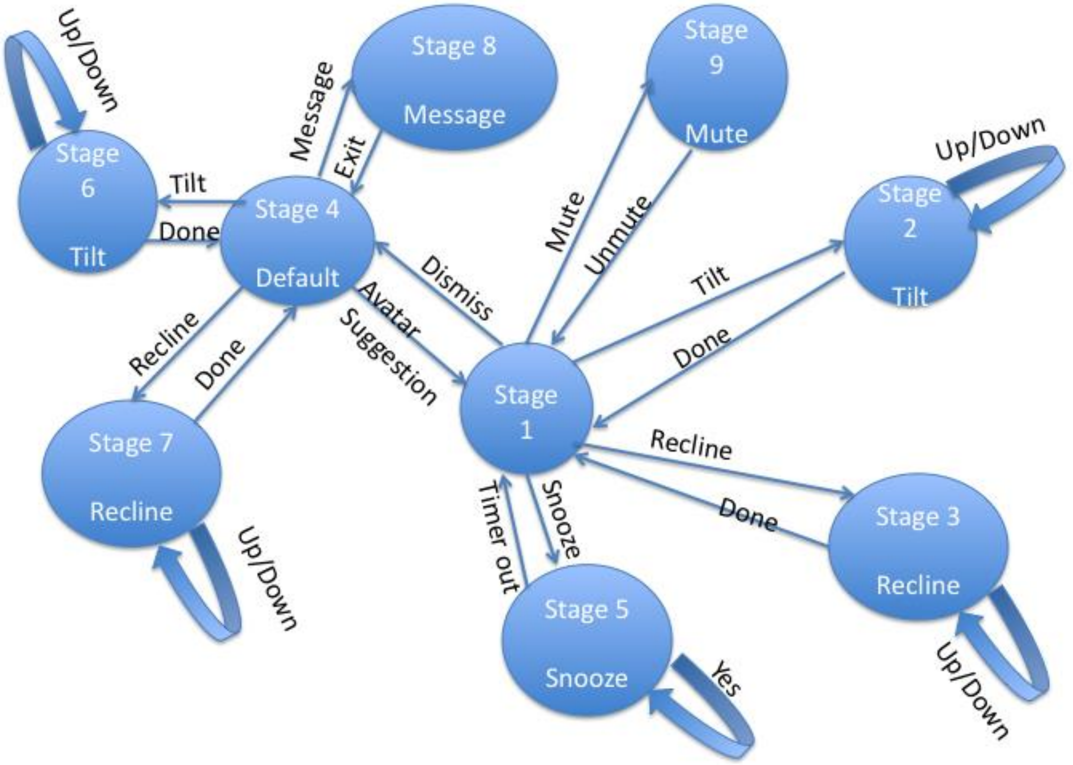


Figure 3. State Diagram of Push-to-talk Commands

2.2 Current State Section

The current state section is supposed to give a user a visual feedback of the tilt and recline motions that one has made. This section is divided into two parts: tilt and recline.

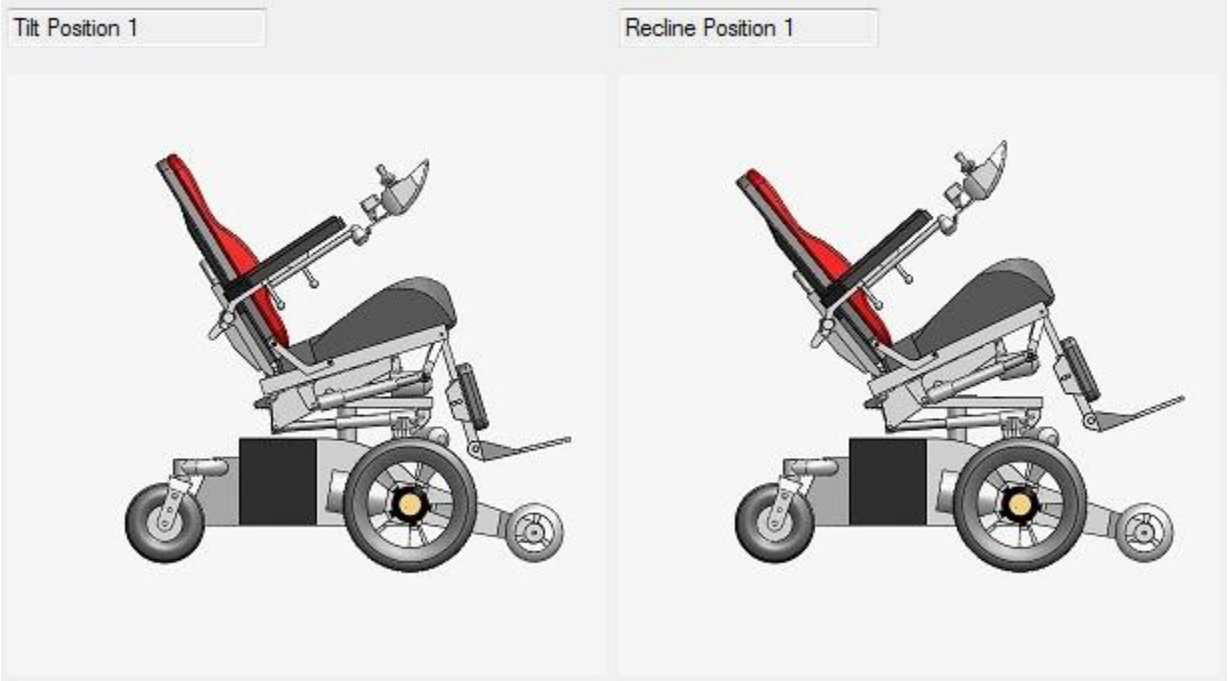
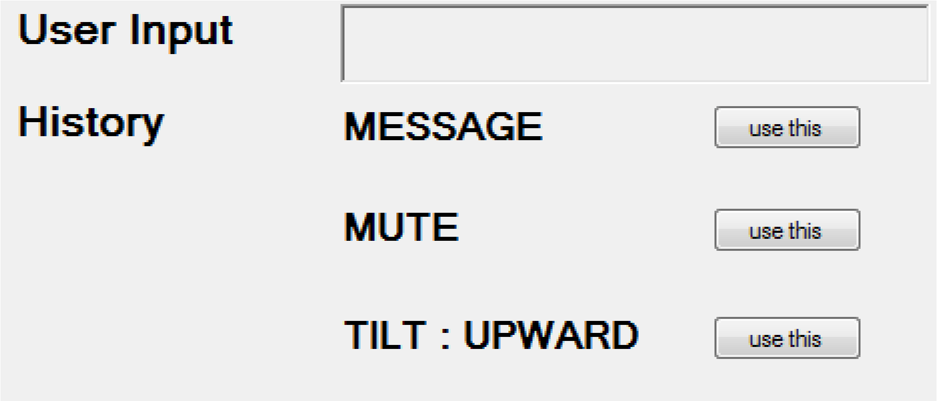


Figure 4. Current State Section

2.3 User Command Section

The User Command section displays the last user command recognized by the simulator. Since it displays how the system has recognized and interpreted what a user just said, it could enhance the usability of the system. Below the last user command, it displays the input history of the three last commands made by the users so that they can see and reuse some of the recent commands that have been used.



MESSAGE

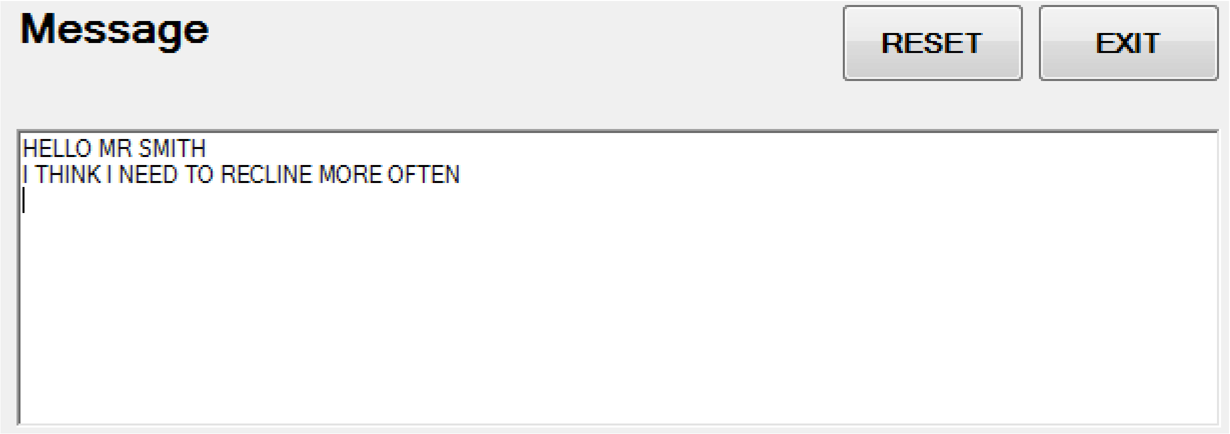
MUTE

TILT : UPWARD

Figure 5. User Command Section

2.4 Message Section

The message section is for the messaging capabilities from the user to the clinician. This allows the users to type a message to their clinician using VoicePredict technology. The message session is initiated by giving the message command. Once the system goes into the message state it fully enables VoicePredict functionalities that are included in Traveling Wave's SDK. The user would speak out a word and then start typing the word one just said using the virtual keyboard. While the user types in each character of the word, the possible word options detected by the VoicePredict technology will appear above the virtual keyboard. The user can then simply choose the word if it is on the list. Otherwise, the user can type in the next letter to see the next option. The exit button on the Message section, as shown in Figure 6, is used to end the messaging activity.

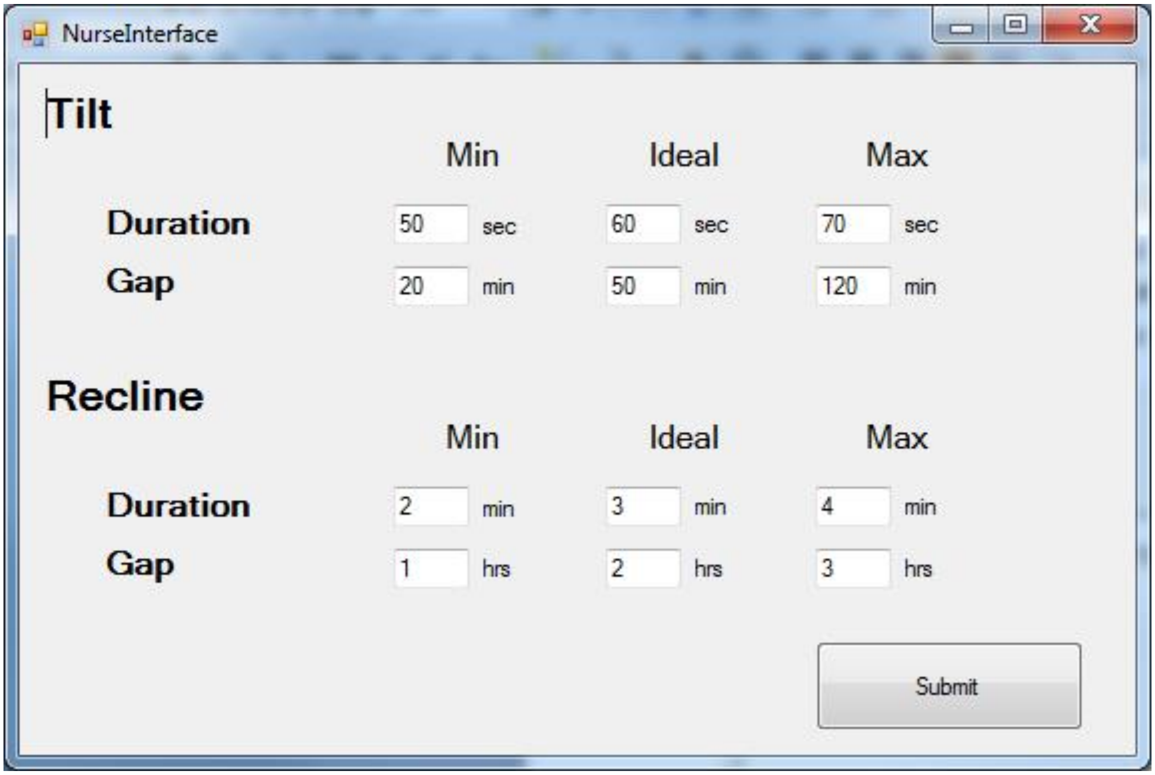


HELLO MR SMITH
I THINK I NEED TO RECLINE MORE OFTEN

Figure 6. Message Section

2.5 Nurse Interface

The Nurse Interface is to set the timing information for the Virtual Coach to give users timely suggestions. The parameters consist of duration for which the action has to be performed and the degree to which the action has to be performed such as recline angle, tilt angle, leg elevation duration, etc. All these information is stored locally and the avatar is used to trigger the timely suggestions to the user.



Tilt

	Min		Ideal		Max	
Duration	50	sec	60	sec	70	sec
Gap	20	min	50	min	120	min

Recline

	Min		Ideal		Max	
Duration	2	min	3	min	4	min
Gap	1	hrs	2	hrs	3	hrs

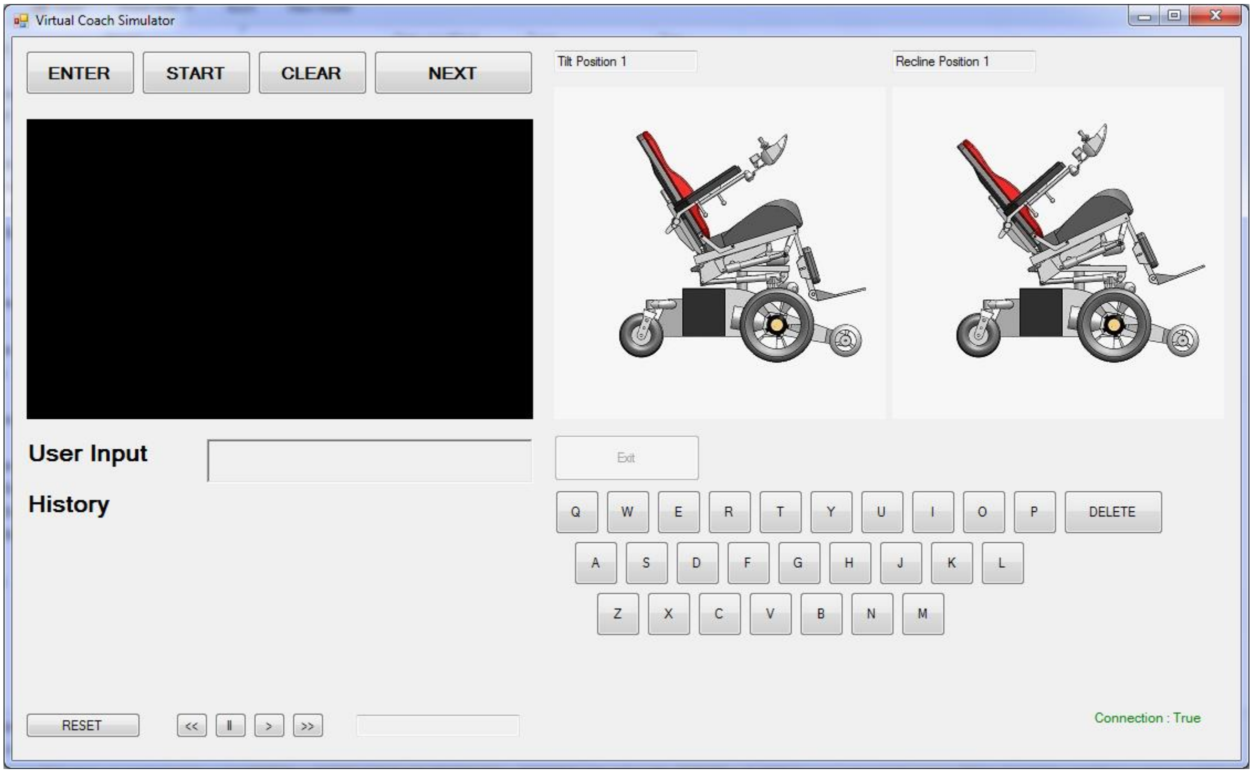
Figure 7. Nurse Interface

3. User study

The user study is designed to capture and analyze the advantages and pitfalls of “VoicePredict” compared to other method of interaction such as virtual keyboard and 9-digit keypad, which are most common in mobile environment. The study has been conducted in two parts to have more analytical tests. For detailed instruction on the user study please refer to Appendix A.

3.1 User Study Part 1

In the first part we compared virtual keyboard and voice recognition functionality called “Push-To-Talk” in the context of Virtual Coach Simulator. The Simulator through Avatar and animations will give the test subjects several suggestions. Then, they were required to respond to the suggestion in a preset way through either virtual keyboard command input or push-to-talk voice commands. For each method of input, running time, and number of button clicks were measured. Running time is measured as sum of {time between start of first button pressed} for each of five possible operations. To eliminate fatigue bias, the order of operation was randomized. Also, to eliminate the learning curve effect, the subjects was given a demo by the experts beforehand, and there would be sufficient time to practice the interface before the data is collected.



Tilt Position 1

Recline Position 1

Connection : True

Figure 8. Interface of the user study application used for part 1 of the study

3.2 User Study Part 2

The second part of the study compared the strength of VoicePredict over regular virtual keyboard and 9-digit keypad in writing out some phrases. In this part, the test subjects would be given two different sentences to type out using three different input methodologies. Similar to part 1, the test subjects were given sufficient amount of instruction and practice time to get familiar with the interface of the application / simulator. Again, the running time, and number of button clicks will be measured and analyzed. In this part total running time will be measured as time between Form load and the correct phrase is completed.

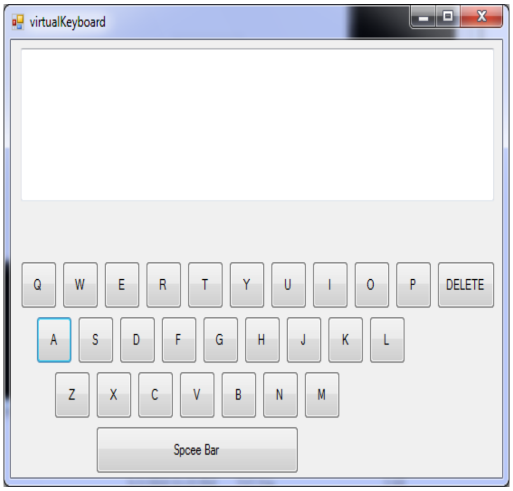


Figure 9. Virtual Keyboard user study app

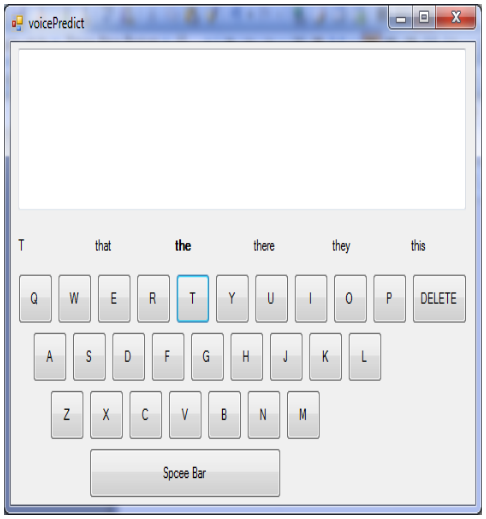


Figure 10. Voice Predict user study app

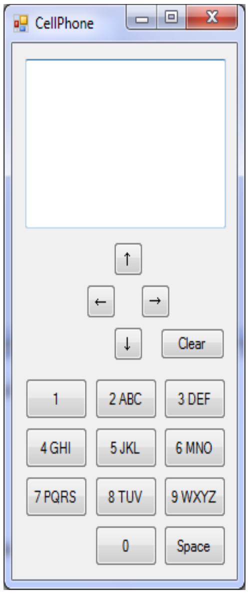


Figure 11. Cell phone interface simulator

4. Results and Analysis

4.1 Observed Result – Part 1

Amount of time to complete an operation

	Recline Up		Recline Down		Tilt Upward		Tilt Downward		Message	
User #	Key	Voice	Key	Voice	Key	Voice	Key	Voice	Key	Voice
1	9.39	1.85	7.63	1.63	6.67	1.46	5.69	1.51	3.31	0.79
2	5.46	1.89	5.11	1.65	6.36	1.93	4.9	1.68	2.75	1.15
3	8.39	1.99	6.13	2.02	5.74	1.93	5.04	1.7	4.39	0.79

4*	5.23	1.6	5.22	1.63	4.94	1.8	4.22	1.86	3.37	0.81
5*	5.38	1.5	4.35	1.41	4.38	1.45	3.63	1.49	2.34	0.69

Table 1. Individual data on the amount of time to complete an operation. The time is measured in seconds. The data labeled as "Key" refers to the time to give command in Virtual Keyboard. The data labeled as "Voice" refer to the time to give command using Push-to-Talk.
* Subject #4 and #5 are two of our researchers.

Average and Standard Deviation for "Amount of time to complete an operation"

Operation	Recline Up		Recline Down		Tilt Upward		Tilt Downward		Message	
	Key	Voice	Key	Voice	Key	Voice	Key	Voice	Key	Voice
Average	6.7700	1.7660	5.6880	1.6680	5.6180	1.7140	4.6960	1.6480	3.2320	0.8460
Standard Deviation	1.9690	0.2067	1.2560	0.2200	0.9571	0.2423	0.7923	0.1522	0.7736	0.1763

Table 2. Average and Standard Deviation for "Amount of time to complete an operation". Units are in seconds. Numbers are calculated using Average() and STDEV() functions of EXCEL.

Number of Keystrokes to complete an operation

Recline Up		Recline Down		Tilt Upward		Tilt Downward		Message	
Key	Voice	Key	Voice	Key	Voice	Key	Voice	Key	Voice
15	4	13	4	12	4	10	4	8	2

Table 3. Number of keystrokes to complete an operation

4.2 Data Analysis – Part 1

The results where as expected, the voice input was much better in terms of the amount of time and of the number of key strokes to give commands. On average, two word commands using voice took about between 1.648 seconds to 1.766 seconds for the users and same commands using virtual keyboard input which ranged between from 4.696 seconds up to 6.77 seconds. The only one word command "MESSAGE" using voice took about 0.846 seconds and same command using virtual keyboard input took about 0.846 seconds. As table 3 suggests, the number of key strokes was another advantage of voice input.

An interesting point about the data is that the time to give command for the voice input seems to depend on number of words in the command, and the keyboard input seems to depend on number of characters in the command. This may imply that the time to give command is mostly depending on the number of key strokes (since # of key strokes for voice increase linearly by the number of words as well), however, more data is needed to confirm this.

Finally, the time to give command in virtual keyboard input was distributed more widely compared to voice input. In other words, the response time for voice input did not vary significantly by the user, but the response time for keyboard input varied significantly by the user.

4.3 Obeserved Results – Part 2

Amount of time to complete sentence 1

User #	Virtual Keyboard	VoicePredict	9-digit Keypad
1	17.86	37.31	33.56
2	17.66	26.91	27.78
3	22.3	24.21	32.3
4	20.35	32.33	54.12
5*	16.67	15.45	29.73
6*	16.7	18.54	27.45

Table 4. Individual data on the amount of time to complete sentence 1. The time is measured in seconds.
* Subject #4 and #5 are two of our researchers.

Amount of time to complete sentence 2

User #	Virtual Keyboard	VoicePredict	9-digit Keypad
1	27.64	46.97	50.21
2	21.51	29.76	35.88
3	22.83	35.16	40.52
4	25.21	36.64	66.18
5*	23.41	28.07	52.48
6*	19.56	22.65	35.11

Table 5. Individual data on the amount of time to complete sentence 2. The time is measured in seconds.
* Subject #4 and #5 are two of our researchers.

Average and Standard Deviation of the amount of time to type in a message

	Virutal Keyboard		VoicePredict		9-digit Keypad	
Sentence #	1	2	1	2	1	2
Average	18.59	23.36	25.7917	33.2083	34.1567	46.73
Standard Deviation	2.2591	2.8244	8.2306	8.4231	10.0755	11.9565

Table 6. Average and Standard Deviation for the amount of time to type in a message. Units are in seconds. Numbers are calculated using AVERAGE() and STDEV() functions of EXCEL.

Number of key strokes to complete sentence 1

User #	Virtual Keyboard	VoicePredict	9 Digit Keypad
1	43	33	81
2	43	31	77
3	43	34	97
4	43	28	82
5*	43	30	77

6*	43	31	77

Table 7. Individual data on the number of key strokes to complete sentence 1. The time is measured in seconds.
* Subject #4 and #5 are part of this research group

Number of key strokes to complete sentence 2

User #	Virtual Keyboard	VoicePredict	9 Digit Keypad
1	57	44	107
2	57	28	95
3	57	33	105
4	57	38	110
5*	57	35	105
6*	57	36	95

Table 8. Individual data on the number of key strokes to complete sentence 2. The time is measured in seconds.
* Subject #4 and #5 are part of this research group

Average and Standard Deviation of the number of key strokes to type in a message

	Virutal Keyboard		VoicePredict		9-digit Keypad	
Sentence #	1	2	1	2	1	2
Average	43	57	31.1667	35.6667	81.8333	102.8333
Standard Deviation	0	0	2.1370	5.3166	7.7567	6.3377

Table 9. Average and Standard Deviation for the number of key strokes to type in a message. Units are in seconds. Numbers are calculated using AVERAGE() and STDEV() functions of EXCEL.

4.4 Data Analysis – Part 2

The results from this part of the study did not agree to the study conducted at TravellingWave and the common beliefs of the researchers in this research group. For most of the users virtual keyboard input were the fastest, the voice predict was second, the 9-digit keypad was the slowest. For sentence 2, all users performed best with the virtual keyboard method, followed by VoicePredict method, and 9-digit keypad in that order.

For sentence 1, there were some disagreements to the order of the performance. User 1 performed better using 9 digit keypad compared to VoicePredict and User 5 performed better using voice predict compared to virtual keyboard. A questionnaire revealed that user 1 was sending more text messages compared to rest of the users. This outlier was probably due to different level of skills each user possesses. This also explains extra-ordinary low performance of user 4 in 9-digit keypad input; because that user responded that he/she almost never send text messages.

Another interesting point is the performance of the researchers compared to other users. For VoicePredict, the researchers showed significantly better results compared to the other users in terms of the task completion time. This result suggests that the experimental design was not able to eliminate the learning curve completely, and the researchers who have been exposed to VoicePredict since the beginning of the semester performed better compared to others who only had few minutes of VoicePredict experience. This also implies that long term research may be required to better assess the performance of VoicePredict.

However, despite these outliers, the result generally shows that Virtual Keyboard was more efficient compared to VoicePredict, and VoicePredict more efficient compared to 9-digit keypad. Study conducted at TravellingWave and common beliefs of the researchers in this group expected VoicePredict to be more efficient compared to Virtual Keyboard. The behavior of the users during the study procedure might partially explain the cause of this result. When using Virtual Keyboard and 9-digit keypad, the users with sufficient practice time had constant focus on the buttons. However, when they were using VoicePredict, every time after pressing a key, the users skim through the wordlist from left to right to search the word they desire, and then make a decision whether to press another key or select a word from the list. What VoicePredict is different from other methods is the skimming action which takes small amount of time, and the fact that they need to make a decision before they take an action (In other methods, the users have to think about which button to press while VoicePredict requires users to choose between the wordlist and the keyboard, even before the users choose which button or word to press). However, the data collected in this study is not sufficient to accept or deny any possibilities, and as suggested by TravellingWave, the layout of the interface might cause some impact on the result as well. More research including CogTool modeling would be useful to find the definite cause of the situation.

Also the data on the number of key strokes between virtual keyboard and VoicePredict did not agree to the conclusion derived by the internal study of TravellingWave. VoicePredict in this study performed much less compared to that of TravellingWave's study in reducing number of key strokes. This could be resulted due to many different reasons, one of them being the choice of the message. The longest words in the phrase used in study were seven letter words, but most of the words were 3-5 letters long. VoicePredict would not be able to reduce number of key strokes significantly for shorter words since it requires one or two keyboard presses plus a click to choose a word.

Also the accuracy of the prediction engine caused some effect on the number of keystrokes when using VoicePredict. Some words in the phrases used in this study was never recognized and showed up in the suggestion list during duration of the study. The word "JUMPS" was noted that it does not exist in the dictionary, however some words in the dictionary such as “FOX” and “DOG” was never showed up in the suggested word list before spelling out the word. Also words such as “LAZY”, “BROWN”, and “EXPERT” never showed up as suggestion until three or more characters were typed in. Since the execution time increases somewhat proportional to the number of keystrokes, this also may have been the cause of the low performance of VoicePredict compared to Virtual Keyboard.

5. Conclusion

Overall, this study was not designed to make concrete decision on the strength of VoicePredict over other methodologies of input. Rather, this study aims to act as a starting point of the research on VoicePredict. The researchers concluded that further in-depth research is required to make more decisive action. Despite the efforts of the researchers, there were some issues in the experimental design that might have caused unintended impact on the result of the study. First, the experiment environment was not under complete control. The users 2, 3, and 4 were in the same location during the study, however, user 1, 5, 6 were in different environments during the study process. Since the performance of the voice recognition software is heavily dependent on the environment generally, it is not impossible to think that this caused some adverse impact on the result. Also, the number of the users in the study was too small to make any significant conclusions. Although, users were diversified in terms of the major, it cannot be denied that they were all university students in highly technical university. Much larger and diverse pool of users would improve the plausibility of the result. Finally, as suggested by TravellingWave, the user study did not fully utilize the full potential of VoicePredict software. The dictionary was remained static during the study, and VoicePredict were not able to gather any usage information to improve the prediction in the long run. This might have helped in terms of controlling the performance of each user in this study, however in the long-run research, this feature might cause huge discrepancy between actual performance of VoicePredict and the result of a future study.

Appendix A: Manual for User Tests

Part 1

1. Give the user instruction on how to give commands to the simulator. Followed by a demo done by the expert.
2. Give the user time to practice with the interface and commands.
3. The user will be given following tasks in random order. User will be required to perform the task in either using voice or virtual keyboard
    a) Tilt up operation
    User will be given a recommendation to tilt the chair up. Time will be measured since beginning of the tilt operation to when tilt up command is recognized by the simulator

    b) Tilt down operation
    User will be given a recommendation to tilt the chair down. Time will be measured since beginning of the tilt operation to when tilt down command is recognized by the simulator

    c) Recline up/down operation
    Similar to Tilt up/down operation, User will be required to perform recline up and down functionality.

    d) Message
    User will be given a task of entering message mode and exit using the exit button

4. Repeat #2 with different method of input.

Part 2

1. Give the user instructions, on how to give commands to the simulator. Followed by a demo done by the expert.
2. Give the user time to practice with the interface and commands.
3. The user will type two sentences (In random order). Also each user will be given a different sequence of input methods.

    The sentences are:

    Sentence 1: The quick brown fox jumps over the lazy dog

    Sentence 2: All question asked by five watched expert amaze the judge