

CHANTI: Predictive Text Entry Using Non-verbal Vocal Input

Adam J. Sporka

Department of Computer
Graphics and Interaction
Faculty of Electrical
Engineering
Czech Technical University
in Prague
Prague, Czech Republic
sporkaa@fel.cvut.cz

Ondřej Poláček

Department of Computer
Graphics and Interaction
Faculty of Electrical
Engineering
Czech Technical University
in Prague
Prague, Czech Republic
polacond@fel.cvut.cz

Torsten Felzer

Institute for Mechatronic
Systems
Technische Universität
Darmstadt
Darmstadt, Germany
felzer@ims.tu-darmstadt.de

Paul Haiduk

Department of Psychology
Technische Universität
Darmstadt
Darmstadt, Germany
haiduk@psychologie.tu-
darmstadt.de

Sri H. Kurniawan

Department of Computer
Engineering
University of California
Santa Cruz
Santa Cruz, CA, U.S.A.
srikur@soe.ucsc.edu

I. Scott MacKenzie

Department of Computer
Science and Engineering
York University
Toronto, ON, Canada
mack@cse.yorku.ca

ABSTRACT

This paper introduces a text entry application for users with physical disabilities who cannot utilize a manual keyboard. The system allows the user to enter text hands-free, with the help of “Non-verbal Vocal Input” (e.g., humming or whistling). To keep the number of input sounds small, an ambiguous keyboard is used. As the user makes a sequence of sounds, each representing a subset of the alphabet, the program searches for matches in a dictionary. As a model for the system, the scanning-based application QANTI was redesigned and adapted to accept the alternative input signals. The usability of the software was investigated in an international longitudinal study done at locations in the Czech Republic, Germany, and the United States. Eight test users were recruited from the target community. The users differed in the level of speech impairment. Three users did not complete the study due to the severity of their impairment. By the end of the experiment, the users were able to enter text at rates between 10 and 15 characters per minute.

Author Keywords

Human-computer interaction, assistive technology, scanning, ambiguous keyboard, predictive text entry, non-verbal voice input, user study

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

CHI 2011, May 7–12, 2011, Vancouver, BC, Canada.

Copyright 2011 ACM 978-1-4503-0267-8/11/05...\$10.00.

ACM Classification Keywords

H.5.2 Information Interfaces and Presentation: User Interfaces—*Input devices and strategies (e.g., mouse, touchscreen)*

General Terms

Design, Experimentation, Human Factors, Measurement, Performance.

INTRODUCTION

Interacting with a computer often requires entering text in one form or another, especially when the computer is a communication aid; for example, in the context of Internet chat or email. The standard input device for text entry is the manual keyboard. It seems perfect for this task, since it can generate a large number of input signals – typically over 100 – accommodating all of the characters of the Latin alphabet, Arabic numerals, punctuation and special characters, as well as function keys and modifier keys – at the same time, on a single device. Some experienced typists are even able to enter text with a keyboard faster than they can utter text aloud. However, on second thought, it must be acknowledged that the manual keyboard is *not* perfect – if it were, it would be usable by everybody.

Unfortunately, persons with physical disabilities are often unable to operate a standard keyboard. Therefore, to fully utilize modern communication technology, alternative input methods are needed to enter text. Depending on the type and severity of the disability, the number of different input signals may be very limited. Sometimes everything has to be conveyed with a single touch of a button – the actuation of a single switch.

The alternatives in the minimalist case are based on *scanning* or *indirect* text input. Instead of directly entering a character, the user accepts “suggestions” highlighted by the software, which cyclically moves the highlight through a set of options – persisting on each for a *scan delay* τ (between 0.5 s and 2.0 s), before moving on [20, 2].

The shortcoming of scanning systems is that they are comparatively slow – hardly a true alternative. The scanning-based solution QANTI [6] – the predecessor of our current approach, described below – is a step in the right direction, yet it still does not compete with other input methods.

For users who can speak, *speech recognition* might be a faster alternative for text entry. However, the voice of many persons with physical disabilities is subject to conditions such as dysarthria, so producing the same vocal output (with tolerable variations) for the same word is often impossible. As a consequence, no computer really “understands” such users, applying ordinary speech recognition. “Non-verbal Voice Interaction” (NVVI), involving humming or whistling [24], is a possible answer.

This paper presents the software system CHANTI, for “vo-Cally enHanced Ambiguous Non-standard Text Input”. The tool attempts to accelerate the QANTI “Scanning Ambiguous Keyboard” [15] by using NVVI, thus allowing users to select (normally scanned) items *directly*.

The remainder of the paper is organized as follows. After reviewing NVVI, the CHANTI system is introduced, detailing the program structure and ideas behind the software. The usefulness of the system was investigated in an evaluation involving eight participants from three countries and with different levels of speech impairment – all belonging to the target population. The study organization, individual sessions, and overall results are presented and discussed. The paper concludes with a brief summary and a quick look to the future.

RELATED WORK

Non-verbal Voice Interaction

Non-verbal voice interaction (NVVI) is based on interpretation of non-verbal sounds produced by the user, such as humming or whistling. Various acoustic parameters of the sound signal (pitch, volume, timbre, etc.) are measured over time and the data stream is interpreted as an input channel. A formal description of NVVI is given, for example, by Poláček et al. [18].

Watts and Robinson [24] propose a system where the sound of whistling triggers the commands for OS UNIX. Igarashi and Hughes [12] used non-speech sounds to extend the interaction using automated speech recognition, pointing out that non-speech sounds are useful to specify analog (numeric) parameters. For example, the user could produce an utterance such as “volume up, *aaah*”. In a response, the system would increase the volume of the TV for as long as the user held the sound of “*aaah*”.

An emulation of the mouse device and its use as an assistive tool is described by Sporka et al. [22] and Bilmes et al. [4]. NVVI has been successfully employed as an input technique for gaming, as described by Sporka et al. [21] (Tetris) and Hämäläinen et al. [10] (platform arcade games for children).

NVVI has also been used as a means of artistic expression. A live performance designed and described by Levin and Lieberman [14] involved real-time visualization of voice, both speech and non-speech sounds. Depending on the sound the performing artists were producing, the system rendered non-figural patterns on a screen located behind the artists. Samaa Al-Hashimi created Blowtter [1], a NVVI-controlled plotter. Another system where non-verbal sounds are used to control a process of drawing is described by Harada et al. [11].

NVVI allows realtime control, unlike automatic speech recognition where the system waits for completion of an utterance. The language independence inherent in NVVI may simplify cross-cultural deployment. NVVI may thus enhance the class of applications controllable via the input acoustic modality.

Keyboard Input Methods

In an *ambiguous keyboard*, a single key is associated with more than one character of the alphabet, while a *dictionary* is checked for candidates matching a sequence of keys [16, 3].

A well-known example is the phone keypad, where most keys are associated with 3 or 4 characters. When used with dictionary-based disambiguation, the intended word appears in about 95 % of all cases for English [19]. If there is more than one matching candidate, the user scrolls through a list to select the desired word. The list typically not only contains *exact* matches, but also words extending (or *completing*) the entered key sequence; i.e., the software tries to *predict* which word the user intends to enter.

In a *scanning keyboard*, the keys are cyclically highlighted. The user selects the currently highlighted key by actuating the dedicated switch. Typically, a small number of options is presented at a time. If the space of options is larger, it must be hierarchically subdivided and the user needs to make a number of selections to reach the desired option.

A *scanning ambiguous keyboard* (SAK) is a combination of the two methods. It brings together the advantages of scanning (i.e., text entry using a single input key) and ambiguous keyboards (ideally demanding only one keystroke per character, although using less keys than characters).

QANTI, the predecessor of CHANTI

QANTI [6] is an implementation of a SAK, specifically targeted for persons with physical disabilities who are unable to utilize a standard keyboard. It was developed as a fully implemented system, rather than as a proof-of-concept prototype.

An exhaustive parameter search by MacKenzie and Felzer [15] identified a layout with four virtual keys as the most favorable (demanding the smallest number of scan steps per character for a given dictionary). Three virtual keys, each covering about one third of the alphabet, are used to produce a *code sequence*, while a fourth virtual key moves the focus to a frequency-ordered list of candidate words.

SAKs work only with dictionary words, so a separate mechanism is required to add new entries to the dictionary (see below).

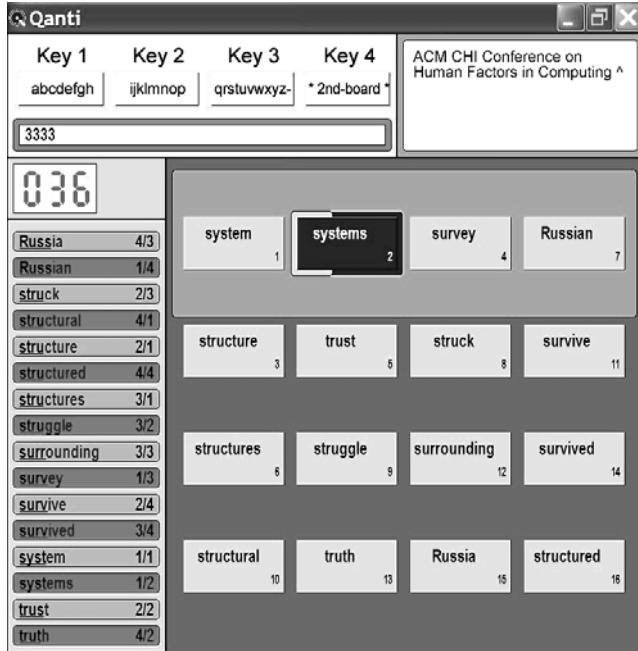


Figure 1. QANTI in candidate selection mode.

This is the basic concept of QANTI (see fig. 1 for a screenshot). To enter a word, the user first produces a code sequence with the help of the four linearly scanned virtual keys in the *sequence selection area* (top left of the screen). While the sequence is entered, a list of candidate words is constantly updated with the 16 most frequent candidates displayed both in the bottom left area of the screen (in alphabetical order) and on the buttons of a large 4×4 board on the bottom right (in frequency order). QANTI supports English and German as default languages. The dictionary may be user-extended.

Once the desired word appears, the user changes to *candidate selection mode*, where the 16 buttons are scanned in a *row-column* fashion. Having selected the candidate, the user chooses among 16 finalization options shown on the buttons of the row-column scanning board. The options determine the way the selected candidate is rendered into the entered text (top right area of the screen), for example, turning the first character into a capital letter, or appending a space, comma, or period at the end.

For initiating selections, QANTI supports *intentional muscle contractions* [8] as an input signal. This feature emphasizes

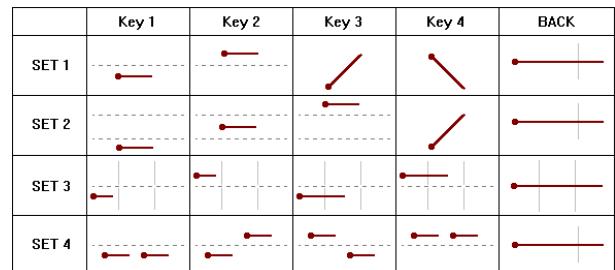


Figure 2. NVVI gestures used with CHANTI. Dark thick lines are pitch profiles over time. Horizontal dashed lines are user-specified pitch thresholds. Vertical solid lines are duration thresholds.

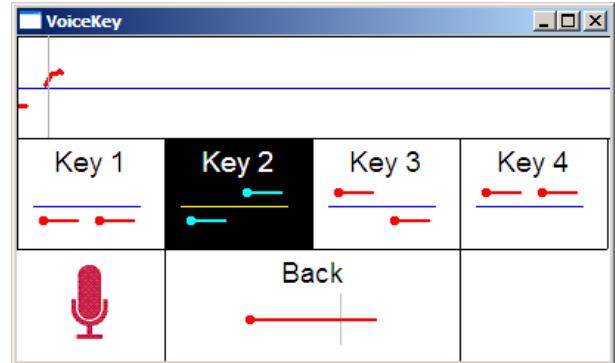


Figure 3. VoiceKey. The user has just produced a gesture G2. Upper part is the detected pitch profile. Lower part shows available gestures to produce.

the target group, since it suffices to merely issue tiny contractions of a single muscle of choice, and thus requires a minimum of physical effort. As a consequence, even someone with a very severe disability can enter text reasonably fast, provided that one muscle (e.g., the brow muscle) can be reliably controlled (also [7]).

QANTI becomes a ready-to-use system through its menu mode. When the user applies a special mechanism involving the fourth virtual “sequence key”, the buttons of the 4×4 board are re-labeled, giving access to several higher-order menu functions. In this mode, the user has the choice to correct errors (either in the entered text or in the current code sequence), to enter line breaks, to configure the scan delay, or to copy the entered text to the clipboard or to disk.

One menu option invokes an ordinary (non-ambiguous) on-screen scanning keyboard, offering a total of 64 virtual keys. This “Full Keyboard” allows the user to enter arbitrary character sequences – in particular non-dictionary words (e.g., proper names) – by adhering to a three-step scanning scheme [9] (with the 64 keys arranged in four groups with 16 buttons each). The menu also includes an add-to-dictionary feature for new words.

THE CHANTI SYSTEM

Design of CHANTI

CHANTI combines the philosophy of QANTI and NVVI. CHANTI is built on a scanning-based tool (maintaining the

hierarchy in the space of options presented to the user), but accepts NVVI gestures for directly selecting items. The structure of the user interface is close to that of QANTI in that text is entered word-by-word. Words are selected from a vocabulary. Each word is ambiguously entered as a code sequence. After a code sequence is specified, the user disambiguates the selection by choosing the word from a list of candidates corresponding to the entered code sequence. Various functions, such as simple editing commands and character-based virtual keyboards for entry of out-of-vocabulary words are available through a menu.

As opposed to QANTI, where the interaction is performed by a single switch activated in specific time slots to make a selection, CHANTI is controlled exclusively by NVVI gestures which are directly assigned to the choices. This provides faster access to the individual choices, compared to the scanning approach in QANTI.

An NVVI gesture is a tone or a sequence of tones with defined characteristics, such as pitch, pitch inflection, or duration, produced by the user. There were four different sets of gestures, equivalent in function, from which the user may choose. There are four NVVI gestures in each set, *Key 1* through *Key 4* (used either to enter a code sequence or advance in the menu) and the *BACK* gesture to reverse the effect of the last gesture produced. This gesture can be used multiple times (multi-level undo). The gestures *Key 1* through *Key 4* correspond to the four keys of QANTI. All gestures are shown in Figure 2.

NVVI has been implemented by VoiceKey (Figure 3), a standalone program that recognizes gestures produced by the users and communicates them to the main application of CHANTI. Other features of the acoustic signal (volume, timbre) were not used in this setup.

The structure of the user interface of CHANTI is shown in Figure 5. Initially, CHANTI awaits either one of the gestures *Key 1*, *Key 2* or *Key 3* to commence entering the code sequence (Fig. 4a) or *Key 4* to enter the menu. When a code sequence is entered, the gesture *Key 4* initiates candidate word selection mode (Fig. 4b). The user then needs to produce two gestures: One selects a row, the other selects a column. Subsequently, the user chooses a finalization option (Fig. 4c). The user may also choose to clear the code sequence and start over by selecting the “Delete Sequence” command in the candidate word selection or finalization option modes.

The menu allows the execution of simple commands (insert a space, remove the last character, remove the last word, insert a new-line character) and access to special modes. These are character-based virtual keyboards. The available characters on these keyboards are organized in a matrix. The user selects a desired character by specifying coordinates using the gestures (similar to selecting candidate words).

Items “Start Test” and “Stop Test” were included for the experiment to start and stop data collection.

THE STUDY

The Study Organization

The purpose of the study was to gauge the first impression of CHANTI, how users would adapt to CHANTI over time, and whether they would be willing to accept CHANTI as their typing tool (and for which type of text).

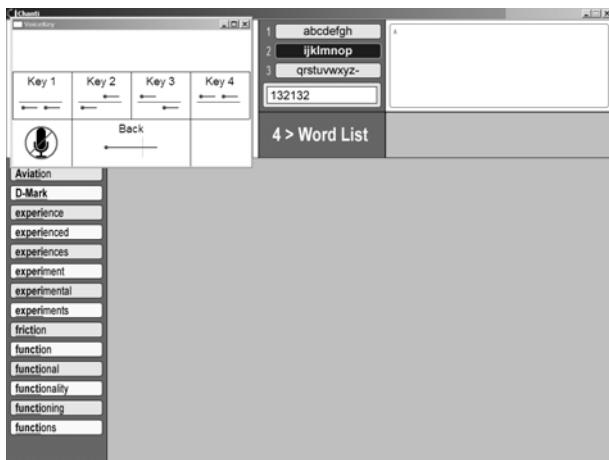
Since we aimed at studying users’ insights, we organized the study as longitudinal and qualitative. Generally, participants were asked to use CHANTI for minimum of 30 minutes each day, over the course of 7 days.

For participation in the study we invited eight participants from three countries: Germany, the Czech Republic, and USA. This allowed us to test CHANTI in three different language contexts. The participants covered a range from no speech impairment to severe dysarthria. All participants were screened for being able to produce NVVI gestures during the Day 1. Three participants were excluded from the study because of the severity of their dysarthria which prevented them from producing the NVVI in required resolution or extreme fatigue even after very short exposure to the system. In Germany the users were recruited via interviews with clients of several local healthcare institutions. In the Czech Republic the users were recruited in cooperation with a local association of paraplegic people. In USA the user was recruited through personal contact.

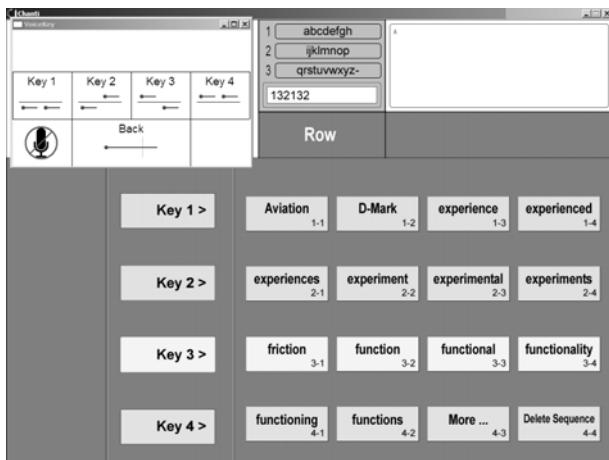
We used the same dictionaries for English and German as used in QANTI. The Czech dictionary was based on a frequency dictionary compiled at the Charles University in Prague [23].

The sessions were outlined as follows:

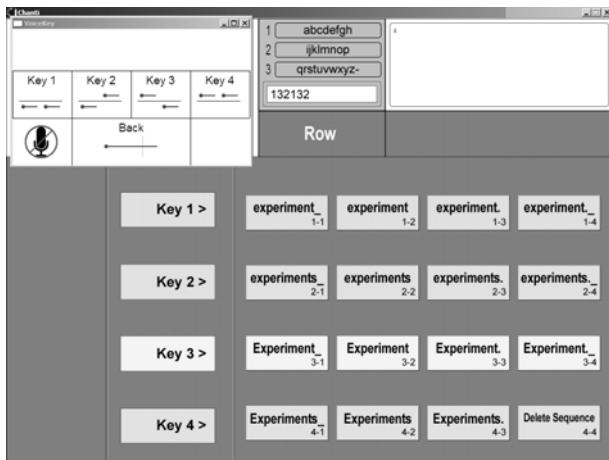
- Day 1 — System set-up and pre-test interview: The participants were asked how they use the information and communication technology (ICT) as well as about their specific disability-induced problems relating to text input. The participants’ assistive technology was discussed. The participants were screened for their sufficient ability to control their voice and trained in using NVVI (see below) and then determined which set of gestures was best for them.
- Day 2 — The participants’ capacity to use NVVI was checked. The session continued with the first exposure to CHANTI and first-impression interview.
- Days 3 through 6 — Continued exposure to CHANTI: Participants were asked to begin writing using CHANTI. Their performance was measured on example phrases, between 5 and 15 words per phrase. These phrases were randomly selected from a text file. Separate collections of phrases were assembled for English, German, and Czech.
- Day 7 — Last day of exposure to CHANTI. A post-test interview took place, in which the participants were asked about their overall experience using CHANTI, and what they considered the strong and weak points of the interaction.



a) part of the code sequence entered



b) selecting a word from the candidate list



c) selecting how the word should be finalized

Figure 4. Illustration of CHANTI in various stages of operation.

NVVI Training

It is known that training is needed for proper use of the NVVI modality [21]. For this purpose we developed a sim-

ple training module which generated a random sequence of gestures and then prompted the users to produce the gestures one-by-one. The users were considered ready for the study if they were able to produce 15 gestures out of 16 without a mistake.

Participant 1

Miloš¹ is 30 years old. He is quadriplegic since birth. He is an IT specialist in a small company based in Prague. On his request, his participation was remote. The interviews were conducted over the telephone and e-mail.

He is able to use a desktop computer while sitting at his desk by a mouth-held stick through which he can type on the keyboard as well as move the mouse. When laying in bed, his laptop is suspended on a platform above him, allowing him to use the stick to type on the keyboard as well as operate a small tablet that emulates mouse control. Apart from the Sticky Keys utility available in Microsoft Windows, he uses no other assistive technology. He reports typing as fast as 100 CPM. He spends 10 hours per day using a computer. He frequently uses shortcut keys. He likes exploring new technologies and interaction techniques.

Miloš is able to use the system tools of the machines he administers through a remote desktop facility. He frequently uses Microsoft Word, composes the HTML and PHP code, edits music, etc. He does not use social networks such as Facebook. He uses the ICQ instant messaging network for quick exchanges of short messages, rather than for extensive chatting. Regarding the use of computers, he feels no disadvantage against other users.

However, he relies on the help of others in hardware-related problems, including switching the machine on. He reports problems using the mouth-held stick on capacitive-sensing devices, such as touchpads or touch switches. He is willing to invest time in training new assistive devices and solutions. He reported having spent three months training for tablet use.

Interaction with CHANTI

Miloš was using CHANTI in Czech. Miloš was eager to participate in the test. He spent 12 hours using CHANTI over 7 days, which was well over the limit set by the design of the experiment. He spent about two hours testing all the sets and deciding which set of gestures to use. ("I was trying hard to find out which set would be best for me.") Finally he opted for gesture set #1 (see Fig. 2). He felt that these gestures were the easiest for him to produce and yielded the lowest error rate. He reported that gesture set #2 was the most difficult to use. ("It was not easy to precisely hit those three separate tones.")

When using CHANTI to write unconstrained text, he noticed that the dictionary was not complete (the Czech dictionary contained only about 30,000 word forms) and tried composing words letter-by-letter, which he found very slow.

¹The participants are represented by fictional names to protect their privacy.

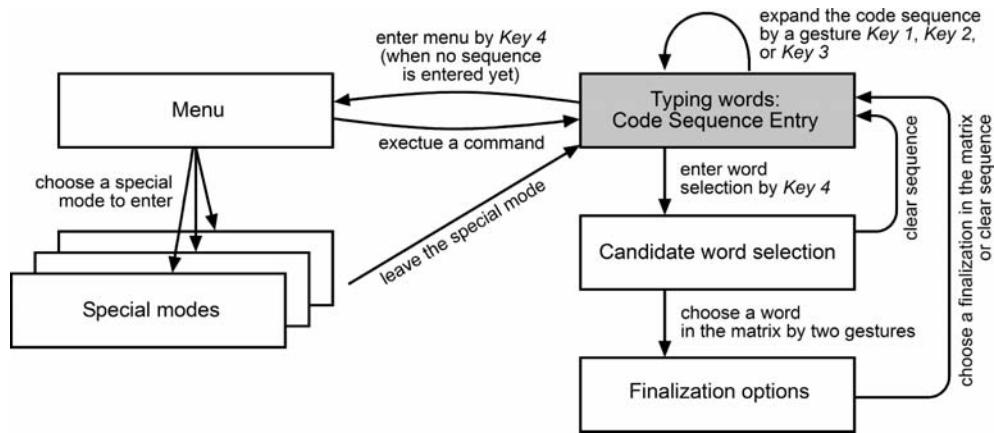


Figure 5. CHANTI user interface state diagram. The initial state is grey.

Initially, his median performance using CHANTI was about 5 CPM. The median performance over his later trials, measured using the internal test facility, was about 15 CPM (with in-vocabulary words only).

He found the structure of CHANTI simple, yet versatile. He liked the fact that it can be controlled by only five distinct voice gestures. He found CHANTI a useful tool for somebody who can not use other means of control. However, he found typing using CHANTI too slow for his needs.

Participant 2

Petr is 19 years old. He is quadriplegic since an accident two years ago. He is a senior-year high-school student who spends typically 2 to 4 hours at a computer daily. He uses the computer to access study materials, to communicate with friends over e-mail, to access the telephone and watch movies. He relies on other family members to assist him daily with the hardware setup.

He uses head motion tracker SmartNav4 by NaturalPoint which emulates a mouse and Click-N-Type software that emulates a keyboard. The device is based on tracking a reflective dot placed on the user's forehead. He is able to type at about 30 CPM using this setup. His performance is generally better when using a head rest or laying in bed. However, he reported that he can not use glasses, as the reflection distorts the output from the device.

He reported limited experience with eye tracking technology, but had tried the system introduced by Fejtová et al. [5]. He was vaguely aware of breath controllers but he was not aware of other methods of the text input, including the use of the mouth-held stick.

Interaction with CHANTI

Petr was using CHANTI in Czech. Petr reported that he found the system easy to learn and that he was able to learn to use the system rather quickly. "I only need to remember the ranges of letters for individual gestures, so that when producing gestures I would be too far off the range."

Throughout the experiment, Petr used gesture set #4. He acquired a "steady rhythm" producing gestures, roughly at the rate of 50 to 60 gestures per minute which he would interrupt when making a decision, such as selecting a word from the list. In such moments he would vocalize his thoughts in a soft voice, so as not exceed the volume threshold.

He used CHANTI in various locations with different acoustic qualities (kitchen, living room, office). He was visibly frustrated when acoustic interferences resulted in an undesired behavior of CHANTI. These interferences included background noises (he was not using a noise-cancelling headset) or a long reverberation of the room, causing the system to register longer tones than actually produced. However, he was aware of the need for calibration and requested it when he felt that it would improve the responses of the system.

He finds that the system would be best used for writing short text messages. ("I would not want to write a whole novel using this.") However, he pointed out that since the Czech language uses an extensive system of declension and conjugation, a word often needed to be specified until the very last letter, since only the ending of the word would determine the desired word form. This somewhat decreased the utility of the word prediction method used by CHANTI as opposed to English and its much simpler morphology.

His median typing rate was about 6 CPM on the first day of the data collection and about 10 CPM in the last day. He reported that in general he liked the input method but he felt frustrated when the word he was attempting to write was not in the dictionary and instead he had to type the word letter-by-letter using the virtual keyboard.

He expressed his wish to use the system again. He reported that he would use the system as an alternative to his current assistive technology when he gets tired moving his head. He and his caretaker were interested in a production-level implementation of the system.

Participant 3

Gabriele is 45 years old. Since her thirteenth birthday she has Friedreich Ataxia (FA). She used to work as an IT professional in university administration. She frequently wants to use a computer for education, entertainment, or gaming.

Due to the progressing symptoms of FA, she has significant problems using a keyboard. This keeps her from using a computer for communicating via chat or email. Typing one sentence can take up to fifteen minutes. Medium speech problems still make it impossible for her to use regular voice recognition. She immediately liked the idea of software that makes text entry more practicable for her.

Interaction with CHANTI

Gabriele was using the German version of CHANTI. Gabriele decided that gesture set #1 was the easiest and most efficient for her. The symptoms of Friedreich Ataxia also heavily affect breathing and therefore speech and humming. She needed three hours before she was proficient enough with the training tool. The large amount of unintended input was frustrating for her: Not being able to switch the microphone off, she constantly produced input, for example, by coughing. It took her four days of intense practice with the experimenter before starting the first session.

Furthermore, she could hardly create humming sounds short and strong enough. The difficulty for Gabriele was to control her sound and breathing. While trying to make a short hum (*Key 1, Key 2* in gesture set #1), she could not control the timing. VoiceKey is designed to allow the program to interpret a humming sound as either *Key 1* or *Key 2*. The hum has to be shorter than half a second, otherwise the program performs a *BACK* operation.

On the second day of the experiment, we prolonged the time threshold to 700ms. This meant that Gabriele was able to get *Key 1* and *Key 2* correct more often, but a longer time threshold also meant that she had to hum longer to perform a *BACK* key. A longer hum is more exhausting and also bears the danger of unintended inputs because of the difficulty to hold a tone for longer time (eventually the program would then rate a long hum as either *Key 3* or *Key 4*).

On average, she made 154 corrections for a sentence with 60 characters during the first day of the experiment and did not exceed 2.5 CPM. She felt she would constantly improve. At the end of the experiment, the number of corrections dropped to 47 per 60 characters.

She only reached 5 CPM after the last session. This was mainly due to the number of corrections she was forced to make. In addition, her low typing rate was caused by a symptomatic eyes dysfunction, which complicates perception of information displayed on different parts of the screen. She would have preferred a bigger screen instead of the 14.1" notebook display used in the experiment. She indicated a willingness to continue practicing and working with the software on her own.

Participant 4

Rolf is 39 years old and has been diagnosed with Friedreich Ataxia at the age of 15. He uses a wheelchair since 1988, and he has considerable motor problems, which also affect his voice. Rolf is with a small software company and mostly works from home. His work requires him to use a computer for up to 8 hours every day, and despite the fact that his disease has progressed quite far, he is still able to use a standard keyboard and mouse (albeit at a modest typing rate of 15–30 CPM, depending on the time of day).

Due to his vocal difficulties, he is unable to use ordinary voice recognition software. He is able to communicate verbally, but the variations in his speech are too large for a computer program – he already tried several possibilities (mostly causing frustrating experiences). In addition, impaired fine motor control makes it difficult for him to use head/eye trackers.

Rolf is very motivated to find an assistive tool allowing him to interact with a computer at a rate comparable to an able-bodied person. When he was asked about being a participant in the CHANTI evaluation, he immediately accepted the invitation, gladly saying: “This could help me a lot”.

Interaction with CHANTI

Rolf was using CHANTI in German. Rolf also decided for gesture set #1, even though producing ascending or descending gestures was not easy for him. Nevertheless, he indicated that this profile worked best for him, for example, as far as timing is concerned. When using CHANTI for the first time, the participant needed almost 1000 seconds for a sentence with just 64 characters (entry rate: 4 CPM). The main reason for this was the heavy need for error correction (e.g., the *BACK* gesture) to take back erroneous selections.

However, the participant’s results gradually improved. During the test week, Rolf spent more than 2 hours per day practicing with the program, and he was finally able to reach peak rates of 12 CPM. He reported that he liked the look of the program, and the colors. Besides, he commented: “At the beginning, I spent a lot of time looking for the intended word – later, I started to remember the position of the candidates, at least for frequent words; I’m sure I can beat my ‘manual lower bound’ of 15 CPM with longer practice”.

Participant 5

Sarah is a 32 years old graduate student. She is paraplegic due to a sporting accident that happened when she was 18. She was also diagnosed with thyroid problem, which made her suffer from a symptom similar to carpal tunnel syndrome last year.

Before last year, Sarah did not use any assistive system. However, because of her thyroid problem, since last year she had been using on and off speech recognition software and screen reader (to read out the information so that she can lay down while working without looking at the screen). She spends 6–10 hours a day with her computer on weekdays, although she usually does not use her computer on week-

ends and holidays. She uses various computer applications, including social networking (Skype occasionally), email (almost all the time), and Facebook (occasionally). She develops C programs as a part of her graduate work.

Interaction with CHANTI

Sarah was using the English version of CHANTI. Sarah is a quick learner. She went from a median speed of 3.9 CPM on day one with a median correction of 16 out of 32 characters (50%) in her first session to a median of 12.3 CPM on the fourth day with a median of 5 corrections out of 27 characters (19%). It should be noted that on her fifth day, she was not feeling well, and while in the first trial on Day 5 she managed 15.15 CPM with only 3 corrections out of 28, her performance deteriorated quickly within minutes, possibly due to her thyroid problem, to 8.5 CPM and then down to 7.6 CPM with 16 corrections out of 24 characters.

We interviewed her at the end of her first day, and the day after her fifth session (she could not communicate effectively at the end of her fifth session). Her first impressions of the system were quite positive. She stated, “After a while I was able to get more of the hang of it. I find it easiest, more convenient to use the Full Keyboard in CHANTI.” She did, however, complain about some of the key arrangements, stating that it was not intuitive for her that the shift key was in the bottom right quadrant of the full keyboard. She was using gesture set #1.

In her final debriefing, she made several remarks, which are summarized below:

- “Sometimes when using the word completion method of typing, the word I was looking for would never show up in the window. It was frustrating to have to delete each letter and then go to the full keyboard to type it up.”
- “After a short while I felt that I was starting to memorize certain common inputs such as the space and I also felt I was becoming more proficient in writing, i.e., I was coming up with quicker ways of typing and looking for more efficient ways.”
- “I also found it a little difficult with punctuation. I was hoping when using the word completion interface that when typing ‘I’m’ that I could type ‘Im’ and when choosing the word that I meant that one of the options would be ‘I’m’.”

When we asked her very specific questions, such as whether the system helped her to do her work, she provided many constructive criticisms that we can take home and use for further refinement of the system. The remarks are summarized below:

- On whether the system helped her to do her work, she said, “I found myself editing a lot of my typing and believe it could be made a bit more intuitive or perhaps have more functionality to help with short cuts.”
- When asked about the on-screen real estate, she stated, “I would like to resize some of the keyboard windows so that

I can see some screen space where I might be doing other work.”

- On whether the system provided sufficient contextual help, she answered, “I found myself guessing as to what action to take to get the end result I wanted. Or sometimes I would choose some menu options believing that option would be available and when it wasn’t I would go back.”
- On menu arrangement, she thought it took a lot of steps to return to the main menu at times and sometimes she had to click on *DONE* or *BACK* to go back to the main menu which was not intuitive for her.

She did have several positive comments about the system. She thought that for most part the design was consistent, intuitive, and quite easy to learn, and she had fun playing with it. She said she would definitely use the system again on her own if she could get a copy of the system.

DISCUSSION

Most users were satisfied with CHANTI and would use the method on their own if given the possibility or their condition worsened so that they were not able to use their present technology. All participants who were able to use the NVVI notably improved their performance using CHANTI over the course of the experiment (see Table 1). The peak performance of the users was typically between 10 and 15 CPM. This is comparable to QANTI [6]. In fact, Rolf also participated in an earlier study evaluating QANTI – he reached around 12 CPM there as well.

Effects of the speech impairment. Though pitch-driven NVVI, as used by CHANTI, does not require the articulation of facial muscles as in speech interaction, the users still need to promptly control their breathing and vocal folds. In our study we identified a “threshold of applicability” of NVVI: 3 participants out of 8 were not able to complete the first session (and therefore did not participate further). They had severe problems producing NVVI sounds due to their speech impairment induced by ataxia. The other two ataxic participants, Gabriele and Rolf, were able to produce NVVI sounds. While Rolf reached the performance similar to participants with no speech impairment, Gabriele frequently needed to correct malformed NVVI gestures, and this limited her performance at 5 CPM. Clearly, her speech impairment was more severe than that of Rolf. Gabriele was on the borderline of the target group. In particular, her speech problems caused a lot of frustration at the beginning (before she was able to proceed with the first test session), which almost made her decline participation in the experiment.

The NVVI threshold of applicability is lower than that of automatic speech recognition (ASR), as evident by Rolf’s participation. Rolf reported that he could not use any system of speech recognition for interaction with the computer due to a different quality of his speech that was not compatible with the current ASR engines. Rolf represented “an ideal target group” of NVVI; i.e. the people who are able to speak but who cannot use the ASR as NVVI is more robust to speech

Table 1. Overview of the Results.

Participant (country and language)	Disability, condition	Dysarthria	Typical type rate (without CHANTI)	Mean type rate on first day*	Mean type rate on last day*	Gesture set	Would use CHANTI again?
Miloš (CZ)	Congenital malformation	No	100 CPM	14 CPM	21 CPM	#1	no
Petr (CZ)	Quadriplegia (accident)	No	30 CPM	6 CPM	12 CPM	#4	yes
Gabriele (DE)	Friedreich ataxia	Yes	very low	very low	5 CPM	#1	yes
Rolf (DE)	Friedreich ataxia	Yes	15–30 CPM**	4 CPM	11 CPM	#1	yes
Sarah (US)	Paraplegia (accident), thyroid problem	No	0–150 CPM	5 CPM	11 CPM***	#1***	yes

* Including the time spent on any corrections

** Depending on current condition

*** Sarah was using the *Full Keyboard* mode. Her peak performance was 16 CPM but during her last session the condition deteriorated and reached 8 CPM after an attack.

impairments than the ASR. NVVI thus expands the range of applications of the vocal modality in assistive technologies by an important margin.

The capacity to speak in a person can change notably over a short period of time and thus the performance of NVVI can vary, as documented by Sarah's participation on the last day: Sarah started out with typing rate of 16 CPM but during the session her condition worsened rapidly and her performance dropped to 8 CPM.

Language model. While NVVI is intrinsically language independent, CHANTI is by design intended to be used in the context of a specific language. When exploring the potential of CHANTI, Sarah and Miloš found the dictionary limited. Both users tried the Full Keyboard mode whereby any string could be typed. Miloš found the method very slow and would use it only to write a specific out-of-vocabulary word while Sarah switched to this method entirely. Her choice could be compared to switching off the predictive text entry method T9 available on mobile telephones. In Kurniawan's study [13] the participants were not using T9 as they found it distracting especially because of the incorrect predictions and subsequent recovery.

To fully accommodate the Czech language and its complex morphology, the method of the word selection should be changed. For example, the nouns could be selected in their basic form and the desired case could be chosen only in the next step. This would effectively reduce the size of the dictionary and thus the need for extensive browsing of the list of candidates.

CONCLUSION

When mentioning the acoustic modality for text entry, very often the automatic speech recognition (ASR) is mentioned as an example of the assistive technique suitable for this task. However, not all motor impaired people can use ASR due to the speech impairments that accompany their motor disability. This study has shown that the NVVI is a viable acoustic modality for text entry even for some of those who are not capable of speech intelligible by the ASR due to conditions such as ataxia.

In this paper we presented CHANTI, a method for text entry based on a combination of QANTI, a one-switch predictive text entry method, and non-verbal vocal input (NVVI). The

main goal of the presented study was to see how users learn using CHANTI during their initial exposure to the system.

CHANTI did not outperform some assistive text input techniques in terms of the type rate: By the end of the experiment, the users were able to enter text at typical rates between 10 and 15 characters per minute.

However, CHANTI uses standard off-the-shelf hardware with no modifications needed for the system to run and therefore is inexpensive to deploy. As suggested by one participant, CHANTI can be used as an alternative system for specific conditions, such as when taking rest from the current assistive tool.

Future work

For languages with complex morphology (which includes the Czech language), the prediction mechanism should be more informed by the grammar of the target language so that the user may more optimally perform the disambiguation of the candidates. Also, the finalization options should be language dependent. The function of adding custom words to the dictionary should be enabled in CHANTI.

In the design of the study we were following the methodology described by Mahmud et al. [17] who reported that the performance of the use of NVVI reached a plateau by day 5. We did not detect a similar pattern in our data. A continuous study mapping the learning curve should be therefore carried out to determine the typing rate of this method in experienced users.

The focus of this study was on testing of the main principle of CHANTI. For this reason, some functions (such as adding new words to dictionary) were omitted for reasons of simplicity and should be implemented before CHANTI is made available for practical use.

ACKNOWLEDGMENTS

This research has been partially supported by the Ministry of Education, Youth and Sports of the Czech Republic (research program MSM 6840770014), and an EU/FP7 project VitalMind (IST-215387).

We are grateful to all our participants and their caretakers for their involvement in the study.

REFERENCES

1. S. Al-Hashimi. Blowtter: A voice-controlled plotter. In *Proceedings of HCI 2006 Engage, The 20th BCS HCI Group conference in co-operation with ACM, vol. 2, London, England*, September 2006.
2. M. Baljko and A. Tam. Motor input assistance: Indirect text entry using one or two keys. In *Proc. ASSETS 2006*, pages 18–25. ACM Press, 2006.
3. M. Belatar and F. Poirier. Text entry for mobile devices and users with severe motor impairments: handiglyph, a primitive shapes based onscreen keyboard. In *Proc. ASSETS 2008*, pages 209–216. ACM Press, 2008.
4. J. A. Bilmes, X. Li, J. Malkin, K. Kilanski, R. Wright, K. Kirchhoff, A. Subramanya, S. Harada, J. A. Landay, P. Dowden, and H. Chizeck. The vocal joystick: A voice-based human-computer interface for individuals with motor impairments. Technical Report UWEETR-2005-0007, University of Washington, 2005.
5. M. Fejtová, J. Fejt, and L. Lhotská. Controlling a PC by Eye Movements: The MEMREC Project. In K. Miesenberger, J. Klaus, W. Zagler, and D. Burger, editors, *Computers Helping People with Special Needs*, volume 3118 of *Lecture Notes in Computer Science*, pages 623–623. Springer Berlin / Heidelberg, 2004. 10.1007/978-3-540-27817-7_114.
6. T. Felzer, I. S. MacKenzie, P. Beckerle, and S. Rinderknecht. Qanti: A Software Tool for Quick Ambiguous Non-standard Text Input. In *Proc. ICCHP 2010*, pages 128–135, Berlin, Heidelberg, 2010. Springer-Verlag.
7. T. Felzer and R. Nordmann. Speeding up hands-free text entry. In *Proc. CWUAAT'06*, pages 27–36. Cambridge University Press, 2006.
8. T. Felzer, R. Nordmann, and S. Rinderknecht. Scanning-based human-computer interaction using intentional muscle contractions. In *Proc. HCI International 2009*, pages 509–518, Berlin, Heidelberg, 2009. Springer-Verlag.
9. T. Felzer and S. Rinderknecht. 3dscan: An environment control system supporting persons with severe motor impairments [Poster Proposal]. In *Proc. ASSETS 2009*, pages 213–214. ACM Press, 2009.
10. P. Hämäläinen, T. Mäki-Patola, V. Pulkki, and M. Airas. Musical computer games played by singing. In G. Evangelista and I. Testa, editors, *Proceedings of 7th International Conference on Digital Audio Effects, Naples, Italy*, pages 367–371, 2004.
11. S. Harada, T. S. Saponas, and J. A. Landay. Voicepen: augmenting pen input with simultaneous non-linguisitic vocalization. In *Proceedings of the 9th international conference on Multimodal interfaces, ICMI '07*, pages 178–185, New York, NY, USA, 2007. ACM.
12. T. Igarashi and J. F. Hughes. Voice as sound: using non-verbal voice input for interactive control. In *UIST '01: Proc 14th Annual ACM Symp on User Interface Software and Technology*, pages 155–156, New York, NY, USA, 2001. ACM Press.
13. S. Kurniawan. Older people and mobile phones: A multi-method investigation. *Int. J. Hum.-Comput. Stud.*, 66(12):889 – 901, 2008.
14. G. Levin and Z. Lieberman. In-situ speech visualization in real-time interactive installation and performance. In *Proc. 3rd Intl. Symp. on Non-Photorealistic Animation and Rendering*. ACM, 2004.
15. I. S. MacKenzie and T. Felzer. SAK: Scanning Ambiguous Keyboard for Efficient One-Key Text Entry. *ACM Transactions on Computer-Human Interaction (TOCHI)*, 17(3):11:1–11:39, 2010.
16. I. S. MacKenzie and K. Tanaka-Ishii. Text entry with a small number of buttons. In I. S. MacKenzie and K. Tanaka-Ishii, editors, *Text entry systems: Mobility, accessibility, universality*, pages 105–121. Morgan Kaufmann, San Francisco, 2007.
17. M. Mahmud, A. J. Sporka, S. H. Kurniawan, and P. Slavík. A comparative longitudinal study of non-verbal mouse pointer. In *Proceedings of INTERACT 2007, Rio de Janeiro, Brazil; Lecture Notes in Computer Science (LNCS) vol. 4663, Part II.*, pages 489–502. Springer-Verlag Berlin Heidelberg, 2007.
18. O. Poláček, Z. Míkovec, A. J. Sporka, and P. Slavík. New way of vocal interface design: Formal description of non-verbal vocal gestures. In *Proceedings of CWUAAT 2010. Get page range and stuff.*, 2010.
19. M. Silfverberg, I. S. MacKenzie, and P. Korhonen. Predicting text entry speed on mobile phones. In *Proc. CHI'06*, pages 9–16. ACM Press, 2000.
20. R. C. Simpson and H. H. Koester. Adaptive one-switch row-column scanning. *IEEE Transactions on Rehabilitation Engineering*, 7(4):464–473, 1999.
21. A. J. Sporka, S. H. Kurniawan, M. Mahmud, and P. Slavík. Non-speech input and speech recognition for real-time control of computer games. In *ASSETS '06: Proc. 8th Intl. ACM SIGACCESS Conf. on Computers and Accessibility*, pages 213–220, New York, NY, USA, 2006.
22. A. J. Sporka, S. H. Kurniawan, and P. Slavík. Whistling user interface (u3i). In *8th ERCIM International Workshop "User Interfaces For All", LCNS 3196, Vienna*, pages 472–478. Springer-Verlag Berlin Heidelberg, June 2004.
23. Ústav Českého národního korpusu FF UK. Český národní korpus: Srovnávací frekvenční seznamy. <http://ucnk.ff.cuni.cz/srovnani10.php> Retrieved 6 January 2011, 2010.
24. R. Watts and P. Robinson. Controlling computers by whistling. In *Proceedings of Eurographics UK*, 1999.