Beyond Speech Recognition:
Improving Voice-driven Access to Computers

Susumu Harada, Jacob O. Wobbrock, James A. Landay

I. Introduction

For individuals with limited mobility and motor control, having access to a computer may be one of the few options available to them for achieving greater independence, obtaining or retaining employment, staying connected with people and information around them, and expressing themselves creatively.

Speech recognition technology holds great potential for such users with motor impairments due to the hands-free interaction it affords without significant investment in specialized hardware. Its accuracy and performance have been steadily improving, leading to accurate recognition engines such as Dragon Naturally Speaking from Nuance as well as the Windows Speech Recognizer in the Vista operating systems.

One major deficiency in today's speech-based input technology is the absence of the analogue to direct manipulation that has made the mouse such a successful input device. While speech recognition systems excel at spoken text entry and command-and-control-style interaction, they lack the ability to perform continuous and fluid controls that are possible using the mouse.

Such an ability to perform direct manipulation tasks using voice is essential for users with limited use of their hands, especially for those who depend on such capability to successfully operate computer applications for their employment and daily well-being. Such tasks may arise when using diagramming or drawing tools, selecting unnamed items or regions in a user interface, performing continuous browsing tasks such as panning, scrolling, and zooming, or even when interacting with various games and social applications that require fluid input, such as Second Life. However, speech-based control of computers has not yet reached a point where it can provide the same level of access to application functionality afforded by the keyboard and mouse.

To address these limitations, a system called the Vocal Joystick engine [1] has been developed at the University of Washington that can capture the non-speech parameters of human vocalization such as loudness, pitch, and vowel sounds for continuous control of the mouse cursor. As shown in Figure 1, various vowel sounds are assigned to the radial directions, and while the user vocalizes some sound, the mouse pointer continues to move in the corresponding direction, changing direction and speed as the user changes sound and loudness, respectively. The system processes continuous vocal characteristics every 10 milliseconds, resulting in a highly responsive interaction where a change in vocal properties initiated by the user is reflected immediately in the interface.

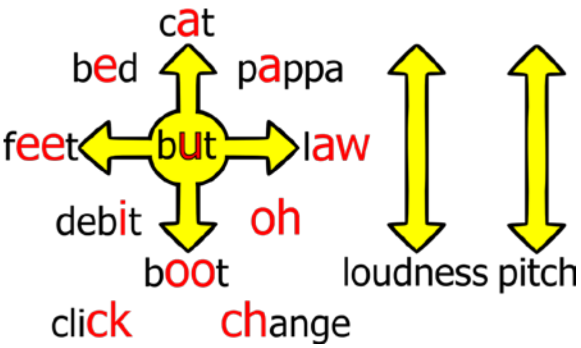

Figure 1: The Vocal Joystick "compass" shows the sounds mapped to each direction. The red vowels in each word approximate the sound corresponding to that direction. The "neutral" vowel (represented by the word "but") can be used to change loudness or pitch without any directional change. The Vocal Joystick also tracks loudness and pitch, as well as discrete non-vowel sounds such as "ck" and "ch."

Over the past years, we have been developing a number of applications around the Vocal Joystick engine in order to explore the potential of using non-speech vocalization as a new input modality, as well as conducting evaluations to determine how well people can learn to use such a modality. We present the representative works to this date, as well as our ongoing work in integrating the non-speech vocal input with existing speech-based input, as well as designing voice user interface (VUI) widgets.

II. Methodology and Results

A. Performance Evaluations

We initially conducted a series of preliminary user studies to understand the usability and capability of the Vocal Joystick pointer control method, both for novice users with minimal training as well as with a smaller set of users with greater experience [3]. Our findings indicated that the Vocal Joystick pointer control follows Fitts' law, enabling us to compare its performance to other well-studied pointing devices, and that its performance by experienced users was roughly 70% of the performance of hand-operated joysticks.

Susumu Harada is a Ph.D. candidate in the Computer Science and Engineering department at the University of Washington.

Jacob O. Wobbrock is an Assistant Professor in the Information School at the University of Washington.

James A. Landay is an Associate Professor in the Computer Science and Engineering department at the University of Washington.

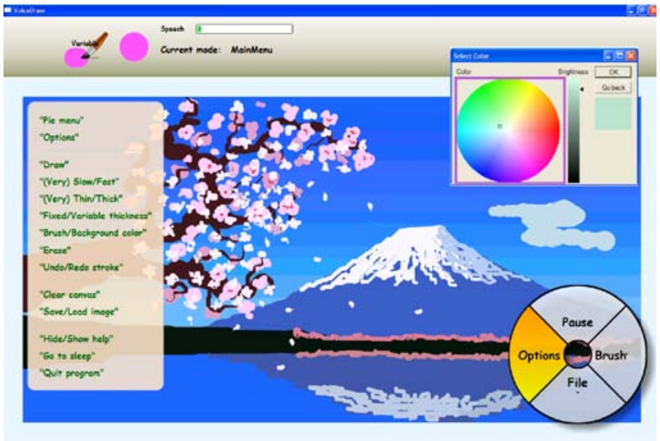

Figure 2: The VoiceDraw application created specifically to leverage the continuous input capability of the Vocal Joystick engine. The first author created the painting in the background using only his voice in about 2.5 hours.

Following our preliminary evaluations, we completed a 2.5 week longitudinal study involving five participants with and four without motor impairments to investigate the learning curve of the Vocal Joystick system 0. All participants exhibited significant improvements ranging from 20% to 50% over the study period, and attained performance levels comparable to those of previously measured experienced users.

B. Applications

We have also built a number of applications to explore the design space of using non-speech vocalization as an input modality. VoiceDraw [4] is one such example, a custom drawing application built to enable the creation of fluid freehand drawings using the Vocal Joystick vowel mapping to control the brush movement, as well as the loudness to control the brush thickness (Figure 2). We worked with an “electronic voice painter,” who has had a spinal cord injury for over 30 years, in designing and evaluating the program, as well as in developing a number of novel voice widgets and interactions, such as the vocal marking menu and continuous undo (Figure 3).

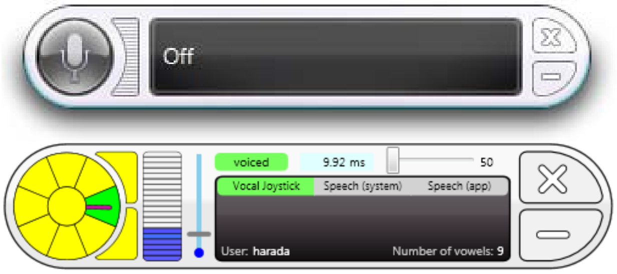

Figure 4: Windows Vista’s Speech Recognition bar (top), and our Vocal Joystick bar (bottom), providing the ability to seamlessly switch between the standard command-and-control/dictation mode and non-speech vocalization mode for fluidly controlling pointers and other voice user interface widgets.

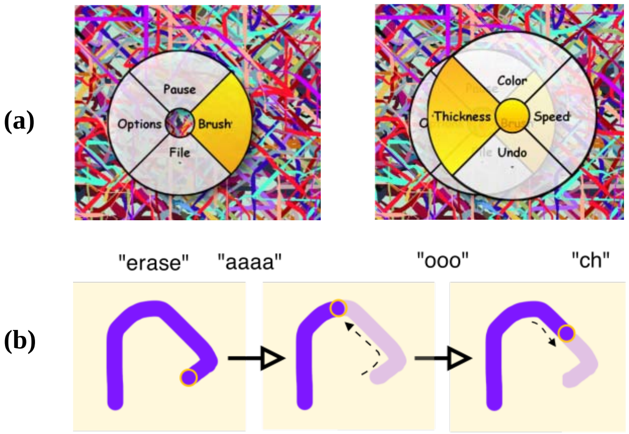

Figure 3: (a) The vocal marking menu invoked by the discrete sound “ck,” followed by a vowel sound corresponding to the direction of the desired menu item, and the discrete sound “ch” to select the item to drill into. (b) Continuous undo enables incremental erasing of a stroke with the sound “aaa,” and incremental restoring with “ooo.”

We are currently working to seamlessly integrate the continuous input capability of the Vocal Joystick engine with existing speech-based command-and-control and dictation interaction methods on the Windows Vista operating system (Figure 4). We hope that this, along with a set of training games we are also developing, will enable users across the world that rely on hands-free computer access to gain greater control over their applications than is currently possible.

III. CONCLUSION

There is a great deal of expressivity in the human voice that is currently not being utilized for controlling computer interfaces hands-free. By augmenting the traditional speech-recognition interfaces with our fluid and continuous Vocal Joystick processing engine, we can greatly expand the expressivity of voice-driven interface manipulations beyond what is currently possible. Our project seeks to make this a reality by developing and evaluating novel voice-driven applications and interaction methods that harness the vocal characteristics beyond what is captured by traditional speech recognition engines.

IV. REFERENCES

[1] Bilmes, J.A., Li, X., Malkin, J., Kilanski, K., Wright, R., Kirchhoff, K., Subramanya, A., Harada, S., Landay, J.A., Dowden, P., and Chizeck, H. The Vocal Joystick: A voice-based human-computer interface for individuals with motor impairments. *HLT/EMNLP 2005* ACL, 995-1002.

[2] Harada, S., Wobbrock, J.O., Malkin, J., Bilmes, J.A., and Landay, J.A. Longitudinal study of people learning to use continuous voice-based cursor control. *CHI ’09* ACM, 347-356.

[3] Harada, S., Landay, J.A., Malkin, J., Li, X., and Bilmes, J.A. The Vocal Joystick: evaluation of voice-based cursor control techniques. *ASSETS ‘06* ACM, 197-204.

[4] Harada, S., Wobbrock, J.O., and Landay, J.A. VoiceDraw: a hands-free voice-driven drawing application for people with motor impairments. *ASSETS ‘07* ACM, 27-34.