HANDS-FREE VOICE-DRIVEN DRAWING AND DIAGRAM CREATION METHOD FOR PEOPLE WITH MOTOR IMPAIRMENTS

Susumu Harada, Jacob O. Wobbrock and James A. Landay
University of Washington

ABSTRACT

Many of today's computer applications rely heavily on the use of a mouse, especially those that require fluid manipulations such as drawing or diagramming applications. This reliance on manual input method presents a great challenge to those with motor impairments, especially those attempting to enter the workforce, where the use of such applications is becoming increasingly essential. Although speech recognition technology has become more reliable for dictation and hands-free command-and-control of computers, they still remain unsuitable as an alternative to the mouse. To address this issue, we have developed an application called VoiceDraw that enables fluid creation of drawings using voice only. In this paper, we present the features of the VoiceDraw application and our user-centered design sessions with a "voice painter," and discuss the potential of our method as a low-cost solution for enabling hands-free access to modern computing applications.

BACKGROUND

One of the key requirements for effective use of drawing and diagramming programs is fluid mouse pointer control. However, many users with motor impairments have difficulty moving a mouse cursor or performing a variety of other continuous control tasks (1). To address this issue, eye-tracking (2) and head-tracking (3) have been investigated as methods for continuous control, but these have notable drawbacks, including the need for high-end equipment and high cost, configuration, and maintenance. Prior work shows that these factors can be significant barriers to adoption and retention of assistive devices (4).

Speech input is an attractive low-cost alternative for hands-free input, yet many of the existing attempts at controlling the mouse pointer using voice fall short of the flexibility demanded by drawing and diagramming applications. In the domain of commercially available products, Dragon Naturally Speaking and Windows Vista speech recognition engines both provide methods in which the user can either click a point on a screen by subdividing it into nine grid cells and recursively calling out the grid number containing the desired target, or move the pointer at a fixed speed by issuing commands such as "move mouse left". The first method does not allow specification of smooth paths, and the second method suffers from slow speed.

In academia, a number of systems have been proposed that attempt to use non-speech vocal parameters (such as volume, pitch, and vowel sounds) to control the mouse pointer (5,6,7,8), but none offers the fluid control of both the mouse's direction and speed. For a comparative discussion of various voice-based pointer control methods, the reader is directed elsewhere (9).

STATEMENT OF THE PROBLEM

Our goal was to bring the productive and creative activities on the computer, such as drawing and diagramming, within the realm of possibility of individuals with motor impairments. Having identified the lack of an effective hands-free mouse alternative as one of the major impediments, we endeavored to make voice-

Figure 1. Philip, a self-described "electronic voice painter" with spinal cord injury, using the microphone to create his artwork on the desktop PC running our *VoiceDraw* software.

based mouse pointer control more effective. We also sought to integrate such fluid pointer control with the power of speech-based discrete command input method to create a system that can be controlled entirely via voice. To this end, we envisioned a fully functional drawing program whose functionalities can be easily accessed using speech commands, and in which the brush can be controlled fluidly using the voice without hampering the creative process.

RATIONALE

The use of computers for work or creative production is becoming increasingly integral in today's society. Such an "advancement" can come as both a blessing and a curse for those with motor impairments. While computers hold the potential for facilitating a wide variety of complex tasks, their current dependence on manual input methods only serve to deepen the gap between those who are able to use them and those who are not. Abilities ranging from creating artistic works and graphical designs to making diagrams and figures for documents are valuable skills, especially for working professionals, but they remain largely inaccessible to people who cannot use the mouse. Thus, we seek to bridge this digital divide by developing a hands-free alternative input method to make such tasks possible without the use of hands.

DESIGN AND DEVELOPMENT

We investigated the use of non-speech vocalizations as a method for enabling hands-free fluid pointer control essential for expressive interaction. In contrast to speech recognition, non-speech vocalizations refer to vocal sounds that do not correspond to any words or phrases in a language. For example, pitch, volume, and vowel quality are all continuous features of voice that may be manipulated by users in control tasks. To evaluate the effectiveness of such a method, we developed VoiceDraw, a drawing application that allows users with motor impairments to fluidly create artwork by using the non-speech properties of their voice. Our process included interviews, field investigations, and user-centered design sessions over a period of two weeks with a self-described "electronic voice painter" who formerly used Dragon Dictate and Microsoft Paint to produce his own artwork. Based on our sessions with this artist, we made numerous iterative design refinements.

Field investigations with the "Voice Painter"

Despite the absence of existing tools that enable drawing on a computer using voice, we were able to find an artist named Philip Chavez[1]. Mr. Chavez (Figure 1) has been creating art on his computer using his voice for over 15 years. He has had a spinal
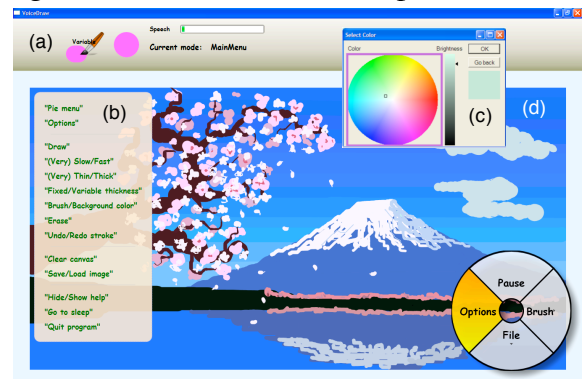

Figure 2. A screenshot of the VoiceDraw application showing (a) the status bar, (b) help overlay, (c) color picker, and (d) canvas area. The first author created this painting using only his voice in about 2.5 hours.
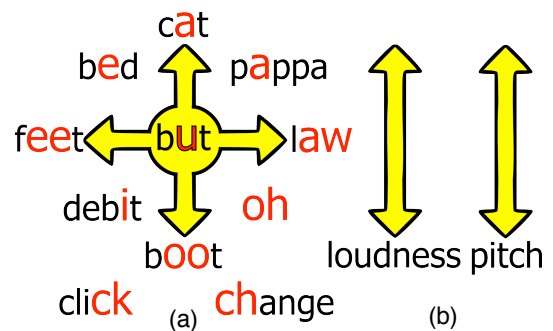

Figure 3. (a) Directional mapping of vowel sounds (in bold red) used in VoiceDraw. The middle vowel is a neutral sound not mapped to any direction. The two non-vowel sounds on the bottom are the discrete sounds "ck" and "ch". (b) The loudness and pitch are also tracked and can be used to control brush speed or thickness.

---

[1] The artist has specifically requested that we use his real name and that we credit his artwork, which can be found at http://www.evoiceart.com/.
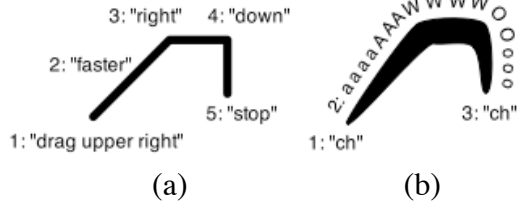
Figure 4. Sample interaction of drawing a stroke using (a) speech-based cursor control, and (b) VoiceDraw. With speech-based control, changes are discrete and incremental. With VoiceDraw, changes are continuous and fluid. Loudness of the utterance can be mapped to stroke thickness or brush speed.
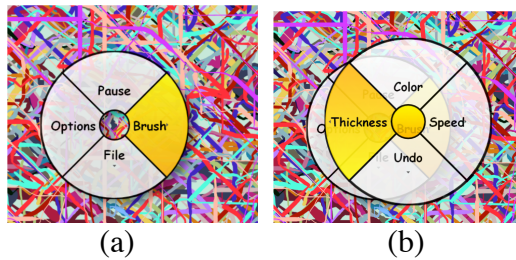

Figure 5. The vocal marking menu supports menu navigation using only non-speech vocalizations. The menu is invoked by issuing the discrete sound "ck". (a) The user finishes uttering "aww" (right), and is about to open the submenu by issuing the discrete sound "ch"; (b) the user finishes uttering "eee" (left) within the submenu and is about to execute the command by issuing the discrete sound "ch".
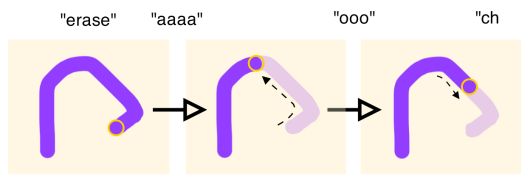

Figure 6. Continuous undo and redo are initiated by issuing the "erase" command, followed by continuous utterance of the "up" vowel ("aaa" as in "cat") to incrementally erase from the end of the stroke, and continuous utterance of the "down" vowel ("ooo" as in "boot") to incrementally restore the portion of the stroke. The change is committed with the discrete sound "ch".

cord injury at the C4-C5 level for about 30 years, and has limited movement in his shoulder. As a result, he has no dexterity in his elbow or wrist, and no sensation in his hands.

Under his previous method, the voice painter controls the mouse pointer using the speech cursor feature of Dragon Dictate, where he utters commands such as "drag mouse upper right," upon which the pointer begins moving at a slow fixed speed until another command is issued. The time spent on each art piece ranged from a few hours to 40 hours, with some taking up to 100 hours. Although he has tried a number of mouse alternatives, including an eye tracker, head tracker, and a mouth-operated joystick, but has settled on speech due to its relative reliability and comfort.

The feedback we obtained through our design sessions lead to the VoiceDraw application features described next. The details of the content of the feedback and the respective design guidelines can be found in (10).

VoiceDraw System Overview

VoiceDraw builds upon the Vocal Joystick engine (11) for the 2-D control of the paintbrush and discrete sound recognition used for mode switching. The system can also recognize spoken phrases for executing commands such as "erase". The Vocal Joystick engine continuously processes a user's vocalizations and recognizes the corresponding vowel sounds. The pointer moves in the direction of vowel sounds based on the mapping shown in Figure 3. This occurs for as long as the user continues vocalizing, at a speed proportional to the loudness of the vocalization. The system also recognizes discrete sounds, such as "ck" and "ch", which can be mapped to input events such as mouse-click and toggle.

Under the default setting, the loudness of the vocalization is mapped to the thickness of the stroke, allowing for a more realistic manipulation of brush strokes. In contrast to the sequence of discrete speech commands that must be issued using Dragon Dictate, VoiceDraw enables the creation of smooth strokes using continuously varying vowel vocalization, as shown in Figure 4.

The vocal marking menu (Figure 5) allows the user to invoke menu commands without leaving VoiceDraw's non-speech vocalization modality. The

voice-optimized color picker (Figure 2c) also leverages the 2-D vowel mapping by having the user choose the color by moving a selector within a color wheel, rather than having to click on a small color palette. The continuous nature of the non-speech vocalization also made such features as the continuous undo (Figure 6) possible, in which the user can incrementally erase or unerase parts of the stroke after a mistake without having to redo the entire stroke.

EVALUATION

In addition to the numerous feedbacks we gathered from the voice painter throughout the iterative design process, we had him create a painting of his choosing and observed the time spent and the reported satisfaction and fatigue levels. Figure 7 shows the resulting artwork created using VoiceDraw compared to the corresponding artwork created using the prior method. The VoiceDraw artwork took a third of the time it took to create a similar painting with his previous method, despite the fact that the VoiceDraw piece was also three times larger. The voice painter did not report any significant vocal fatigue issue other than the desire to take occasional sips of water, and instead commented on the immersive nature of VoiceDraw as it allowed him to focus on the creative task rather than the operation of the tool. The detailed views of the paintings in Figure 10c and Figure 10d indicate the richer vocabulary of strokes supported by VoiceDraw over the voice painter's previous method. VoiceDraw's strokes exhibit unconstrained smooth paths and variable stroke thicknesses. The voice painter expressed great satisfaction in having been able to produce a piece that much more closely mimics his original inspiration, Jackson Pollock.

DISCUSSION AND CONCLUSIONS

We have presented VoiceDraw, an application that investigates the use of voice as a viable modality for creating freeform drawings on a computer without manual interaction. Through repeated field investigations involving a self-described electronic voice painter, we identified a number of key design goals and solutions for upholding them. The results from this work suggest that VoiceDraw can significantly enhance the process of voice-based drawing, and that it opens up a new realm of creative expression that was previously not available to people with motor
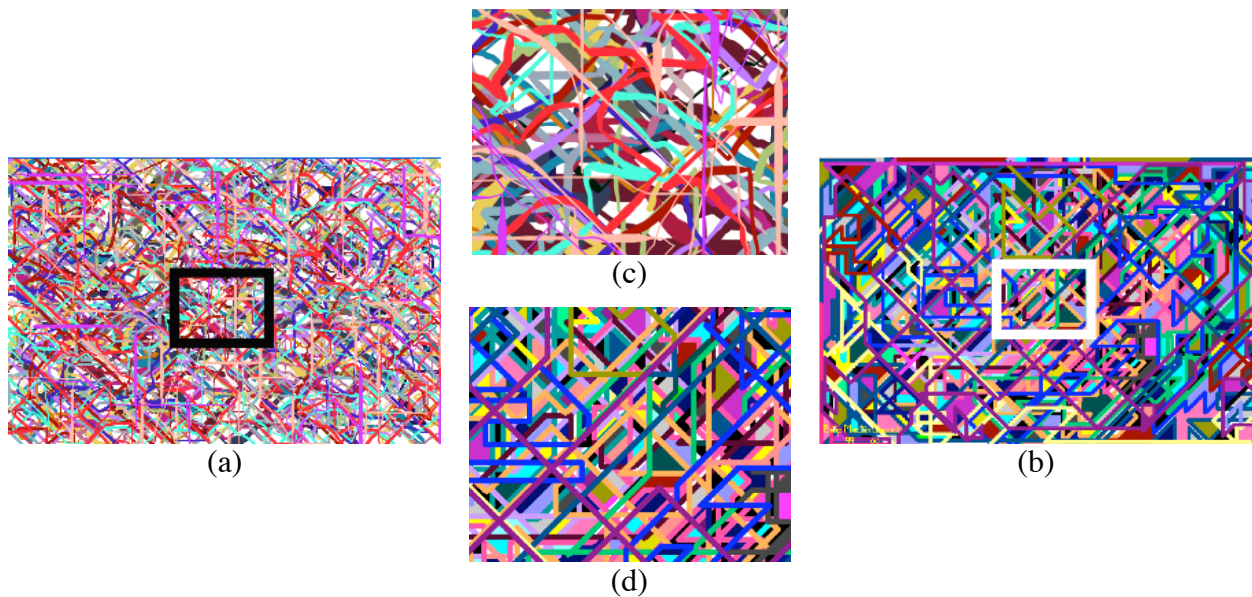

Figure 7. Results from the painting task by the voice painter. (a) The result using VoiceDraw. (b) An earlier piece that the voice painter had created using Microsoft Paint and Dragon Dictate which he recalled from memory to inspire (a). The detailed view of a region of each painting ((c) for (a), (d) for (b)) reveals the difference in the stroke attributes with each tool. The VoiceDraw painting (a) took 3 hours; Microsoft Paint version (b) was estimated to have taken 9 hours.

impairments. The experience gained through building the VoiceDraw application can inform future design of applications optimized for voice control. We are also developing a modified version of the application for supporting the creation of diagrams and figures beyond artistic paintings, and a generalized toolkit with which developers can make other applications accessible via the fluid vocal input methods we presented. With continued development of these tools, the digital divide created by the heavy reliance on mouse and keyboard input can be bridged.

REFERENCES

1. Keates, S., Hwang, F., Langdon, P., Clarkson, P. J., and Robinson, P. (2002). Cursor measures for motion-impaired computer users. *Proceedings of the ACM SIGCAPH Conference on Assistive Technologies (ASSETS '02).* Edinburgh, Scotland. New York: ACM Press, 135–142.

2. Hornof, A. J., and Cavender, A. (2005). EyeDraw: Enabling children with severe motor impairments to draw with their eyes. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '05).* Portland, OR. New York: ACM Press, 161–170.

3. Kjeldsen, R. (2006). Improvements in vision-based pointer control. *Proceedings of the ACM SIGACCESS Conference on Computers and Accessibility (ASSETS '06).* Portland, OR. New York: ACM Press, 189–196.

4. Phillips, B., and Zhao, H. (1993). Predictors of assistive technology abandonment. *Assistive Technology,* 5, 36–45.

5. Igarashi, T. and Hughes, J.F. (2001). Voice as sound: Using non-verbal voice input for interactive control. *Proceedings of the ACM Symposium on User Interface Software and Technology (UIST '01).* New York: ACM Press, 155-156.

6. Dai, L., Goldman, R., Sears, A. and Lozier, J. (2004). Speech-based cursor control: a study of grid-based solutions. *Proceedings of the ACM SIGACCESS Conference on Computers and Accessibility (ASSETS '04)*, Atlanta, GA. New York: ACM Press, 94-101.

7. de Mauro, C., Gori, M., Maggini, M. and Martinelli, E. (2001). Easy access to graphical interfaces by voice mouse. *Technical report*, *Università di Siena*. Available from the author at: maggini@dii.unisi.it.

8. Sporka, A.J., Kurniawan, S.H. and Slavík, P. (2006). Acoustic control of mouse pointer. *Universal Access in the Information Society*, 4(3), 237-245.

9. Harada, S., Landay, J.A., Malkin, J., Li, X. and Bilmes, J.A. (2006). The Vocal Joystick: Evaluation of voice-based cursor control techniques. *Proceedings of the ACM SIGACCESS Conference on Computers and Accessibility (ASSETS '06).* Portland, OR. New York: ACM Press, 197-204.

10. Harada, S., Wobbrock, J.O. and Landay, J.A. (2007). VoiceDraw: A hands-free voice-driven drawing program for people with motor impairments. *Proceedings of the ACM SIGACCESS Conference on Computers and Accessibility (ASSETS '07).* Tempe, AR. New York: ACM Press, 27-34.

11. 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. (2005). The Vocal Joystick: A voice-based human-computer interface for individuals with motor impairments. *Proceedings of the conference on Human Language Technology and Empirical Methods in Natural Language Processing.* Morristown, NJ. Association for Computational Linguistics, 995–1002.

Susumu Harada, Paul G. Allen Center for Computer Science and Engineering Box 352350, University of Washington, Seattle, WA 98195-2350 USA

Alternative text description for figures

Figure 1: Figure 1 shows Philip, a self-described "electronic voice painter" with spinal cord injury, using the microphone to create his artwork on the desktop PC running our VoiceDraw software.

Figure 2: Figure 2 shows a screenshot of the VoiceDraw application showing the status bar, help overlay, color picker, and canvas area.

Figure 3: Figure 3 shows the directional mapping of vowel sounds used in VoiceDraw, represented as a directional compass with eight directions, each with a word representing the vowel sound corresponding to that direction. From the "north" direction going clockwise, the sounds are "a" as in "cat", "a" as in "pappa", "aw" as in "law", "oh" as in "oh", "oo" as in "boot", "i" as in "debit", "ee" as in "feet", and "e" as in "bed". There is also a "neutral vowel" sound in the middle of the compass, which is "u" as in "but"., which is not mapped to any direction. There are also two non-vowel sounds on the bottom called discrete sounds, which are "ck" as in "click" and "ch" as in "change". The figure also shows two vertical arrows representing the fact that loudness and pitch are also tracked continuously and can be used to control brush speed or thickness.

Figure 4: Figure 4 shows a sample interaction of drawing a stroke using the speech-based cursor control and also VoiceDraw. With speech-based control, changes are discrete and incremental, such as issuing the commands "drag upper right. faster. right. down. stop." With VoiceDraw, changes are continuous and fluid, such that a similar stroke can be made by the utterance "ch. aaaaaaAAAWWWOOooo. ch," where the initial and final "ch" puts the brush down or lifts it off the canvas, and the loudness of the utterance can be mapped to stroke thickness or brush speed.

Figure 5: Figure 5 shows the vocal marking menu, which is a pie-shaped hierarchical menu that supports menu navigation using only non-speech vocalizations. The menu is invoked by issuing the discrete sound "ck", upon which the user can utter a directional vowel sound such as "aww" (right) to select the menu item at the corresponding direction, and issue the discrete sound "ch" to either open the submenu (if the menu item has nested menu items) or to issue the corresponding command.

Figure 6: Figure 6 shows the continuous undo and redo feature, which is initiated by issuing the "erase" command, followed by continuous utterance of the "up" vowel ("aaa" as in "cat") to incrementally erase from the end of the stroke, and continuous utterance of the "down" vowel ("ooo" as in "boot") to incrementally restore the portion of the stroke. The change is committed with the discrete sound "ch".

Figure 7: Figure 7 shows the results from the painting task by the voice painter. There are two paintings shown, one created using VoiceDraw, and another created using Microsoft Paint and Dragon Dictate some time ago which the voice painter recalled from memory to inspire the first piece. There are also detailed views of a region of each painting that reveals the difference in the stroke attributes with each tool, in which VoiceDraw's strokes exhibit unconstrained smooth paths and variable stroke thicknesses while the other piece exhibit rigit, fixed-width strokes constrained to 45 degree increments.