# VoicePen: Augmenting Pen Input with Simultaneous Non-Linguisitic Vocalization

Susumu Harada, T. Scott Saponas, James A. Landay
Computer Science and Engineering
University of Washington
Seattle, WA 98195 USA
{harada, ssaponas, landay}@cs.washington.edu

## ABSTRACT

This paper explores using non-linguistic vocalization as an additional modality to augment digital pen input on a tablet computer. We investigated this through a set of novel interaction techniques and a feasibility study. Typically, digital pen users control one or two parameters using stylus position and sometimes pen pressure. However, in many scenarios the user can benefit from the ability to continuously vary additional parameters. Non-linguistic vocalizations, such as vowel sounds, variation of pitch, or control of loudness have the potential to provide fluid continuous input concurrently with pen interaction. We present a set of interaction techniques that leverage the combination of voice and pen input when performing both creative drawing and object manipulation tasks. Our feasibility evaluation suggests that with little training people can use non-linguistic vocalization to productively augment digital pen interaction.

## Categories and Subject Descriptors

H.5.2 [**Information interfaces and presentation**]: User Interfaces – *Voice I/O.*

## General Terms

Design, Human Factors.

## Keywords

Voice-based interface, pen-based interface, multimodal input.

## 1. INTRODUCTION

Pen-based input on tablet computers is used for a wide variety of applications such as note taking, sketching, architectural design [9], graphic design, website design, and animation. Typically, users manipulate a pen on the tablet's surface by controlling pen position, whether the pen is touching the surface, and the state of the pen's barrel button(s). This allows users to draw brush strokes, click buttons, and operate menus.

However, many tasks can benefit from simultaneous *continuous* manipulation of multiple parameters. For example, in drawing applications, one may want to change brush thickness and opacity at the same time (Figure 1), or change the color of the brush while stroking (see Figure 2 for examples of other brush parameters that may be continuously manipulated). This is in contrast to discrete manipulation such as picking a different brush type or color. When creating two-dimensional animations, the user may wish to specify moving elements that not only change their location but also their orientation or scale simultaneously.

**Figure 1: Example of a drawing created with pen input augmented with non-linguistic vocalization. The pen pressure controlled the brush thickness and non-linguistic vocalization controlled the opacity simultaneously as the strokes were being drawn.**

On a typical tablet, users have at their disposal the pen's pressure on the tablet surface to control one of these extra parameters. However, this only allows the control of one extra parameter, and pen pressure may not be the ideal control for many of these parameters.

Investigations into bimanual interfaces such as Bricks [8] and TouchMouse [13] offer some promise for controlling multiple continuous parameters simultaneously. For example, a track pad or keyboard could be used to control a drawing brush's opacity while the pen's pressure controls thickness. However, Tablet PC's, PDAs, or tabletop displays do not have a readily accessible keyboard much of the time and as such there are limited possibilities for traditional bimanual interaction.

In this paper, we investigate voice as a potential secondary modality due to its low cost and expressiveness. Previous work [6][24] has utilized speech to augment pen in the form of verbal


*ICMI'07*, November 12–15, 2007, Nagoya, Aichi, Japan.

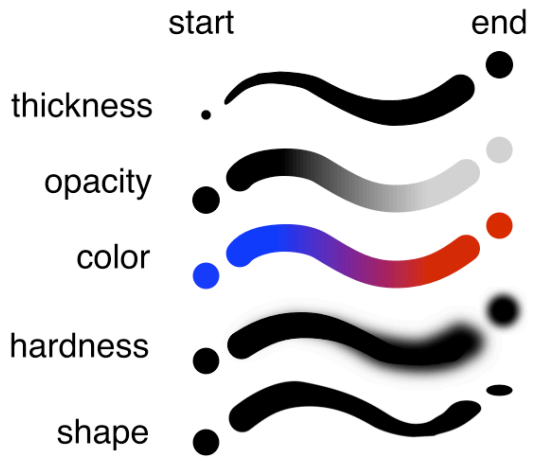



**Figure 2: Various parameters of a brush stroke that may be manipulated while the stroke is being drawn.**

commands, such as saying "show me all the hotspots in this area" while drawing a circle on a map. However, these discrete commands do not allow fluid control of continuous parameters. Instead, we focus our investigation on using non-linguistic vocalizations, or voice sounds that are not words such as the vowel utterance "aaah."

We propose four new voice-augmented pen input techniques primarily related to stroke making. We describe how varying loudness while making continuous vowel sounds can be used to control continuous parameters such as the brush drawing parameters of thickness and opacity. We focus on tasks that require simultaneous input of multiple continuous values.

The rest of the paper is organized as follows: in Section 2, we describe the characteristics of pen and non-linguistic voice input as well as our four interaction techniques. In Section 3, we describe our prototype implementation of these interaction techniques in a system called VoicePen, and discuss the results of a feasibility study we conducted to explore the potential of our techniques. We survey related work in Section 4. Finally, we discuss future work and conclude in Section 5.

# 2. NON-LINGUISTIC VOICE + PEN INPUT

In this section we first describe pen input and non-linguistic voice input separately. We then describe a set of interaction techniques that combine the pen and voice modalities simultaneously.

## 2.1 Digital Pen Input

The digital pen is a good input modality for positioning, direct manipulation, sketching, and performing other single-handed input on a tablet computer. The pen input space usually consists of at least the pen position on the screen surface and whether or not the pen is contacting the surface. In some cases, pens can also indicate a pressure being applied by the user, the state of one or two barrel buttons on the pen, pen angle to the surface, and pen height from the surface (when hovering). Tablet PCs make use of this input space in traditional WIMP interfaces as well as in those specific to digital pens, such as Windows Journal or Denim [23].

Position is usually used for targeting actions, such as clicking a button or menu, or for making brush strokes when the pen is dragged on the surface. Barrel buttons are used for actions such as bringing up context menus. Pen pressure is often used for controlling brush stroke thickness or invoking a context menu.

The primary advantage of pen input is that it allows a person to use their fine motor control for direct manipulation of objects as well as leverage already acquired skills of using analog drawing instruments.

## 2.2 Non-Linguistic Vocalization

Non-linguistic vocalizations are sounds that humans can produce that are not words or phrases in a language. For example, vowel-like sounds such as "aaaiiiuuu", humming, changes in loudness or pitch, or tonguing (e.g., "tatatatata") can be considered non-linguistic vocalizations. There are a number of useful features that can be extracted from such vocalizations for the purpose of providing input to a computer system, such as loudness, pitch, vowel quality (how close a sound is to a particular vowel), etc. Some key characteristics of non-linguistic vocalization that make it a viable option as an input modality include:

1. Continuous and immediate input

   Due to the relatively simple nature of the signal, non-linguistic vocalizations can be recognized almost immediately after a few audio frames have been processed, as opposed to waiting for an entire word or phrase to be uttered, as is the case in standard speech recognition systems. Also, unlike clicking a button or selecting a tool from a palette, the duration of the vocalization and the variations during the vocalization can be used to provide continuous input.

2. Direct manipulation through voice

   The shorter processing time required for non-linguistic vocal input enables tighter coupling between the user's intentions expressed through the production of voice sounds and the system's feedback and response, leading to a greater sense of direct manipulation through one's voice.

3. Robustness

   The simplicity of most features in non-linguistic vocalization can lead to more robustness in the recognition process than with standard speech recognition systems, and can also require little or no user adaptation.

## 2.3 Interaction Techniques

Given these characteristics of pen and non-linguistic vocalization, we sought to explore the potential of combining the two modalities for performing the following three categories of tasks on a Tablet PC.

The first category of tasks we explore is brush stroke creation. Here the user's objective is to draw a certain stroke using the digital stylus as a virtual brush. In existing systems, a user may pre-select a brush color, shape and size, and begin creating the stroke using the stylus, possibly modifying the pressure of the pen against the tablet surface to continuously vary the thickness of the stroke if the tablet is pressure sensitive.

The second category of tasks is object manipulation. Here the stylus is used to "pick up" objects displayed on the screen and move them about the workspace, possibly manipulating their geometric features, e.g., by scaling and rotation.

The third category is workspace navigation. Here the user controls the current view of the workspace they are interacting with by performing actions such as zooming, scrolling and panning. In cases of applications that deal with three-dimensional

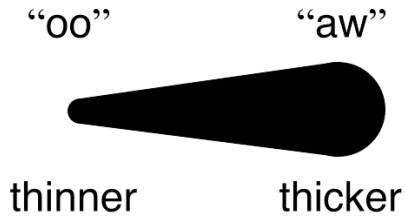



**Figure 3: In thickness control mode, uttering “oo” continuously decreases the brush thickness while “aw” increases it, as long as the user sustains the utterance. The user can vary the loudness of the utterance to control the rate of change in either direction.**

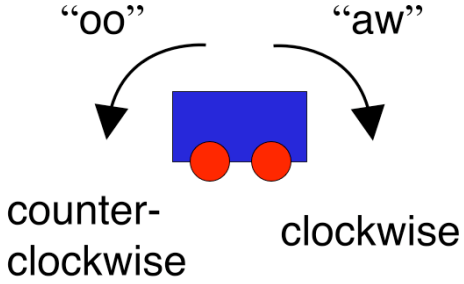
representations, navigation may involve a more complex manipulation of the viewport location and orientation.

We implemented four pen-voice interaction techniques to investigate the feasibility of combining pen and non-linguistic vocalization to support the above categories of tasks.

In all cases, the main principle is to augment pen input with voice, providing continuous and simultaneous input to manipulate an additional parameter beyond those able to be controlled by the pen position and pressure. We focus on vocal manipulation of a single scalar parameter value, and use two vowel sounds (“aw” and “oo”) to either increase or decrease the value. Due to the inherent arbitrariness of mapping vowel sounds to a change in some value, any other of the many possible vowel combinations could have been used. The two vowels we used were chosen based on the relative ease of maintaining the mouth shape while vocalizing the vowels. The loudness of vocalization is mapped to the rate of change in the corresponding direction. The parameter value begins increasing or decreasing as soon as the user begins vocalizing, and stops changing as soon as the user stops vocalizing.

### *2.3.1 Brush Stroke Thickness Control*

In the brush stroke thickness mode, the pen is used to specify the path of the stroke, with pressure sensing turned off so that it does not affect the thickness of the stroke. Instead, the vowel sounds are used to control the thickness. The same technique can be extended to control a variety of other brush properties, such as those listed in Figure 2. This mode was used to determine whether the user is able to manipulate one parameter using voice while at the same time using the pen to control the position of the brush. This also simulates the situation in which the pointing device may not have pressure sensitivity, such as when using a mouse or a PDA.[1] As shown in Figure 3, the user vocalizes the vowel “oo” to make the brush thinner and “aw” to make the brush thicker.




[1] **Figure 5: In rotation control mode, uttering “oo” continuously rotates the object counter-clockwise while “aw” rotates it clockwise, as long as the user sustains the utterance.**


**Figure 4: In opacity control mode, uttering “oo” continuously decreases the opacity while “aw” increases it, as long as the user sustains the utterance.**

### *2.3.2 Brush Stroke Opacity & Thickness Control*

Under the opacity control mode, the pen’s pressure is activated to control the thickness of the stroke, while the user’s voice is used to vary the opacity of the ink. Such a mode may be used to create brush strokes with effects similar to watercolor, where the lightness of the ink varies across the stroke. As with the brush stroke thickness mode, a different brush property may be substituted for opacity. This mode was also designed to investigate whether users are able to manipulate two continuously varying parameters simultaneously using two different modalities. The vowel “oo” is used here to make the brush lighter, while the vowel “aw” is used to make the brush darker (see Figure 4).

### *2.3.3 Object Translation & Rotation*

Object manipulation mode allows the user to move an object around the screen by directly tapping down on the object and dragging the stylus around while holding it down on the tablet surface. We augment this basic manipulation with the ability to control the rotational orientation of the object using voice in parallel to the translation of the object’s location using the pen. Such a feature may be used when creating an animation, in which an object needs to undergo translation and rotation simultaneously. The technique can also be applied to other transformation parameters such as scaling. Figure 5 shows that the vowel sound “oo” is used to rotate counter-clockwise and “aw” is used to rotate clockwise.

### *2.3.4 Workspace Navigation*

We explored controlling one aspect of workspace navigation using voice, namely zooming in and out of a large workspace. In tasks such as map navigation, exploring the space typically involves having to manipulate a scroll bar or a zoom control widget, requiring the user to take the pointer off of the primary point of interest, when bimanual control is not available. [14] and [16] show that users can easily get lost when panning or zooming within large documents without proper feedback and navigation control. We explore the ability to use voice to perform the

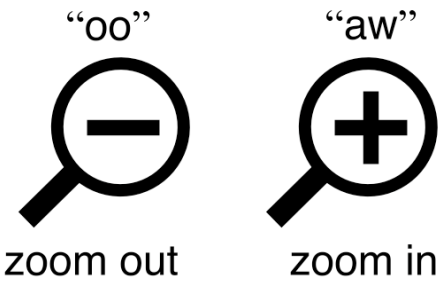



**Figure 6: In zoom control mode, uttering “oo” continuously decreases the current zoom level while “aw” increases the zoom level, as long as the user sustains the utterance.**

zooming while the pen remains on the primary point of interaction. "oo" is used to zoom out and "aw" is used to zoom in (see Figure 6).

# 3. FEASIBILITY STUDY

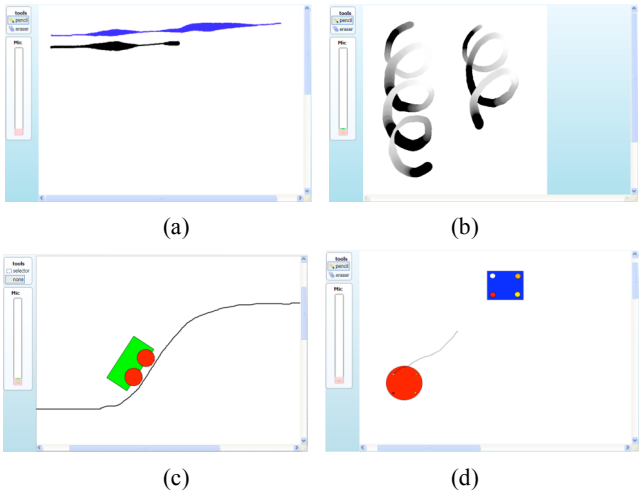
To understand the feasibility of our interaction techniques, we conducted a feasibility study in the lab with a prototype implementation. Here we describe our prototype, VoicePen, and the tasks that participants engaged in using VoicePen. The study results and participant feedback provide insight into the possibilities of simultaneous non-linguistic voice and pen input.

## 3.1 VoicePen Prototype

The VoicePen prototype is written in C# and XAML on the Microsoft .NET 3.0 platform. Vowel recognition and loudness detection is performed using the VocalJoystick engine [3]. The canvas and network components are based on the SketchWizard library [7].

The interface consists of a *drawing canvas* or *workspace* on a Tablet PC (see Figure 1). The workspace can be put in several modes, one for each of our proposed interaction techniques. Depending on the mode, voice and pen input have different effects on the canvas.

In *brush thickness* mode, the user can draw brush strokes where the thickness of the stroke is adjusted through voice interaction. *Stroke opacity & thickness* mode allows similar brush stroking except thickness is controlled through the pen's pressure on the tablet's screen surface and opacity of the stroke is controlled through voice interaction. In both of these modes, the brush preview next to the canvas shows the current thickness or opacity of the brush, which can be changed even when the user is not actively inking. The third mode is *object translation & rotation* mode; where an object such as the car in Figure 5 can be dragged around by the pen while simultaneously being rotated by the user's voice. VoicePen's fourth mode is *canvas zooming* mode. Users can use their voice to zoom in and out of a canvas while simultaneously creating brush strokes. Zooming allows navigation of a much larger canvas than the available screen space as well as the ability to make very detailed drawings.



(a) (b)

(c) (d)

**Figure 7: Screenshots of the four tasks: (a) brush thickness task, (b) stroke opacity & thickness task, (c) rotation task, and (d) zoom task.**

In all of our techniques, the loudness of the user's vocalization reported by the underlying VocalJoystick engine was linearly mapped to the rate of change of a parameter value (e.g., brush size or rotation speed). The Vocal Joystick engine itself uses an exponential mapping between the raw audio energy level and the loudness value it reports [3].

## 3.2 Study Participants and Tasks

Seven people participated in our feasibility study. Three were male, four were female, one was left-handed, six were experienced with tablet computers, six were university students, and three were computer science students. Participants ranged in age from 18 to 45. All studies took place in our laboratory. Participants used a Tablet PC with VoicePen while wearing a headset microphone. Sessions lasted approximately one hour.

Participants' sessions began with a brief demographic questionnaire and training VoicePen to adapt to the user's voice. The adaptation process consists of having the participants make each of the "oo" and "aw" vowel sounds for about two seconds at a comfortable loudness. More details on the adaptation process can be found in [3].

The study session consisted of participants using each of our four interaction techniques to complete a task. Prior to trials in each mode, we explained to the participants the interaction technique including a printed reference for the vowel mapping (see Figures 3 through 6). They were then given a blank canvas to experiment with the VoicePen mode for a few minutes before starting that task's five trials. For each interaction technique, participants completed five trials of the task using VoicePen in the corresponding mode. Figure 7 shows one trial of each of our tasks. The other four trials are different instances of the task. In the cases of *brush thickness* mode and *stroke opacity & thickness* mode, each trial asked the participants to reproduce a target stroke with varying thickness or varying thickness and opacity, respectively. The user was asked to use VoicePen in the current mode to draw a stroke as similar as possible to the target stroke (user strokes were drawn next to target strokes).

In the *object translation & rotation* mode, the tasks consist of using the pen to translate a car image along a path while using voice to rotate the car such that the car's wheels stay parallel to the target path. The participants were asked to try to animate the car with more emphasis on smoothness than precision. Trials of the *canvas zooming* mode consisted of zooming into a small beginning target and starting a stroke in the center of that target and, while continuously stroking, zoom out to find the end target and zooming into the end target to finish the stroke on a very small sub-target point. These targets can be seen in Figure 7d.

## 3.3 Results & Discussion

An emergent theme from our study is that the participants found the first two trials of each technique challenging and the last two easier and often fun. Our participants attributed their initial difficulty to several factors: coordinating the control of multiple simultaneous variables (position & thickness, position & rotation, etc.), learning the mapping of vowels to directions of change, adapting to loudness sensitivity, and making certain vowel sounds. After only one or two trials, many users said they no longer had to think of the vowel mappings. Users also found that with little experimentation they had adequate control of loudness. Multiple users remarked that they were "happy" or "pleased" with how close their drawings were to the target strokes (This was more common with their last two trials of a task even though

those trials were meant to be more difficult). Despite users' lack of familiarity with VoicePen's multimodal input, everyone was able to use VoicePen to complete all the tasks.

Prior to our study we did not know whether using VoicePen for an hour or longer would be straining or tiring. When asked about voice strain only one of our participants said the study was tiring or straining. However, that user also said the continuous vocalizations did not "bother them much" and they would still be happy and able to continue using VoicePen for a while longer.

No single interaction technique was a dominant favorite in terms of difficulty or enjoyment. Some users found the rotation task fun and easy. Other participants really enjoyed and became proficient with the stroke opacity & thickness mode. For example one user remarked while controlling brush thickness with their voice, "this is cool … this is fun … I like this. I think it's pretty easy to control … these are like watercolor or something. They are really pretty." Another user said of drawing using pen pressure for thickness and voice for brush opacity, "This looks so good. This is so much fun. I really like this one." Another participant said, "I see how this could be really useful once you get used to it." We believe the variance in favorite tasks suggests that none of our study tasks are the "killer app" for VoicePen, but that each of our interaction techniques has appeal to some subset of users.

### *3.3.1 Vowel Mappings*

In each of VoicePen's modes, the vowels "oo" and "aw" are mapped to opposite directions of change of a parameter (brush thickness, brush opacity, rotation angle, and zoom level) and loudness is mapped to the magnitude of change. We asked each of our participants about the mappings we chose, other vowel mappings that might make more sense for them, and what other kinds of vocal mappings they would like.

Some users suggested that they might prefer a direct mapping of vowel and loudness while performing continuous vocalization. In this situation, if a user wants to make a medium thick stroke, the user must maintain a medium loudness continuous "aw" sound. However, these users also said they might not want that type of mapping all the time. They speculated the existing method of vocalization, adjusting a parameter up or down, would probably be best for them in most circumstances.

For brush thickness and zooming, some users said the shape of their mouth when making a vowel sound contributed to whether they liked the mapping or whether it felt natural. One user stated "It was nice to have 'aw' for going bigger [in thickness] because your mouth is more open." Users suggested that it made sense and was easy to remember if an open mouth vowel, such as "aw," corresponded to thicker or zoomed in and a narrow mouth vowel, such as "oo," decreased thickness or zoomed out.

Some users said that just the tone of the vowels also played a role, "'oo' and 'ah' for lighter and darker felt really natural" because 'oo' has a *light* sound to it and 'ah' sounds *dark*. All of these comments suggest that there may not be a single mapping that is inherently intuitive to the majority of the users, and that they should be customizable by each user based on their preference

Several participants suggested non-vowel mappings such as pitch and loudness for controlling thickness or opacity. One user suggested making a louder vocalization for a thicker brush and a quieter vocalization for a thinner brush. Another user offered the idea of using pitch to control opacity, with high pitch making the brush lighter and low pitch causing the brush to become darker. These non-vowel mappings suggested by our participants are interesting alternatives to VoicePen's vowel mappings and should be explored in future work.

### *3.3.2 Participants' Techniques & Strategies*

While using VoicePen to complete our study tasks, participants developed their own techniques and strategies. One example of this is some participants' use of short loud bursts to make a large change (such as zooming all the way out or reaching the maximum thickness) followed by a quiet continuous adjustment from that new level. Multiple participants discovered this possibility on their own and found it useful during several trials of both brush tasks as well as zooming tasks. Knowing this user strategy, we could alter the design of our interaction techniques to take advantage of loud vowel bursts.

In the rotation tasks, some users preferred to do rotations in one smooth constant loudness vocalization until the car had rotated to the desired orientation. Other participants found it better to make a series of very short quiet vocalizations until the car was properly rotated. Some users initially tried to rotate the car through one continuous vocalization where they start with their loudness very soft and increase it until the car is rotated properly. However, this leads to users overshooting their desired orientation and engaging in a correction rotation in the opposite direction. These participants eventually adopted one of the previously mentioned approaches of a constant appropriate loudness or multiple short vocalizations.

Participants had various strategies for adjusting brush thickness and opacity. Many users paused their pen movement mid-stroke to adjust the thickness or opacity of the brush with their voice (using the brush preview to the left of the canvas as a guide). However, after a couple of trials, participants began to anticipate brush changes they wanted to make and were able to use loudness and timing to make brush changes without pausing. This approach allowed users to replicate the target strokes by making one continuous fluid stroke of their own. In fact, by the last two trials of each brush tasks, participants needed little or no pausing of the pen for brush adjustment.

Another unanticipated behavior is that multiple users continuously vocalized even when the vocalization had no effect. For example, when the brush thickness reached the maximum value, participants continued to vocalize the vowel despite the absence of a corresponding change. Users said they did this because it felt more "natural" to them to always be vocalizing as long as they were drawing. Users also said they did this because they have a mental model of a direct mapping from vowel + loudness to thickness even though they know that is not how VoicePen works.

### *3.3.3 Fun & Novel Expressiveness in VoicePen*

All of our users remarked multiple times that the VoicePen techniques were fun and enjoyable. When asked about why it was fun, some users said that part of it comes from the novelty of using non-linguistic vocalization as an input mode. Others said that using their voice like this for input was just naturally fun like singing. We think that our participants were discovering that there is a pleasurable nature to this style of input.

Our participants also thought that controlling brush parameters with non-linguistic vocalizations simultaneously with pen input was different enough from traditional Tablet PC drawing that it leads to different drawings. We believe this represents expressiveness unique to VoicePen; while there may be drawings

that would only be made using traditional digital pen input, there are many drawings that would only be made with VoicePen's new input techniques. This opens an interesting future research direction in how new VoicePen interaction techniques can support new and desirable expressiveness in creative tasks such as drawing art or creating graphic designs.

### *3.3.4 VoicePen Limitations*

While VoicePen's interaction techniques offer many benefits, we have identified several limitations of this approach and our prototype implementation.

The primary complaint users had about VoicePen was its loudness sensitivity. Although the maximum rate of change of a parameter value (e.g., brush size or rotation speed) was normalized based on each user's average loudness during adaptation, some users found the mapping to be too sensitive, while others found it to be not responsive enough. Ideally, we would like to allow users to adjust the sensitivity to their preference through a control panel. Further investigation is needed to determine what an ideal mapping function should be between non-verbal vocalizations and control parameters [21]. A few participants also had initial difficulty with pronouncing the right vowel sound. However, this was mitigated by having the participant practice the vowel sound and undergo the adaptation process.

A fundamental issue with VoicePen is that voice is a busy channel. When users make linguistic utterances such as in talking to themselves or someone else they are also uttering the vowels "oo" and "aw" and thus are manipulating VoicePen controls. This means that in scenarios where someone desires to talk while drawing, VoicePen cannot be used. However, some issues with the noisy channel of voice can be minimized by having the system only interpret incoming audio signals when users are actively using the pen. This allows the user to lift the pen from the tablet surface when speaking to themselves or others.

Participants found controlling multiple brush variables, such as thickness and opacity, in addition to pen position, difficult at times. One participant remarked "it's hard to think of thickness and opacity at the same time." This suggests both that there is some learning time associated with discovering how to control multiple variables at once and that it is just difficult to control multiple parameters. However, we observed most participants got a "feel" for it and were able to replicate target strokes.

Certainly there is a social component to our use of voice and using VoicePen in the presence of others can be somewhat embarrassing. In fact, there are numerous situations in which it is probably socially unacceptable to be making non-linguistic vocalizations. However, this is not limited to our techniques; speech input, and sometimes mice and keyboards, can suffer from the same problems. We think there are still many situations where VoicePen may be used without negative social consequences.

Non-linguistic vocalizations such as vowel sounds do not necessarily have obvious, universal, or intuitive mappings to user interface interactions. In our feasibility study, not all participants found all vowel mapping easy to remember or intuitive. For those users on those tasks, they had to spend time thinking about which vowel to use. One user said, "I don't think you'll find a good one. You just got to pick one and get used to it… you don't zoom in and out in real life, so there is no natural [mapping]." This problem can be mitigated by enabling users to quickly experiment with different vowel and loudness mappings. Our participants also gave us some insights into mappings that might be more universal, such as mouth shape to control brush thickness.

# 4. RELATED WORK

We looked to applications of bimanual interaction as a potential source of interaction techniques in which the input parameter controlled by the second hand may be also be controlled by voice. A large body of work has examined the domain of bimanual interaction. We survey only a small sample of work from this research domain that has fostered many novel and interesting interaction techniques. We also discuss previous research that examines the use of pen pressure in tablet applications, novel drawing applications, and uses of spoken input in conjunction with pen input.

## 4.1 Bimanual Interaction

Much research has been directed at investigation of using two hands to interact with computers. From this work a number of interaction techniques have been proposed. Bricks, one of the earlier systems presented by Fitzmaurice et al. [8], demonstrates how simple graspable input devices manipulated by both hands can significantly enhance the input vocabulary. The Toolglass and Magic Lens widgets by Bier et al. [2] enable the use of the non-dominant hand for positioning a see-through tool palette for operations by the input device in a user's dominant hand. Kurtenbach et al. [19] also explored the use of similar widgets in the context of a commercial paint system using a puck and stylus. Hinckley, *et al*. explored the use of a new input device called the TouchMouse along with a standard touchpad for performing map navigation tasks [13].

Researchers have also evaluated the human capability and efficacy of many bimanual input techniques [1][5][12]. Such work describes the benefits of bimanual interaction, including the lowered cognitive load compared to unimodal interaction and time efficiency due to parallelization [20][25]. However, Kabbash, *et al*. suggests bimanual interaction may not yield better performance when not designed appropriately [18]. Many of these research activities were spurred by the initial theoretical framework for bimanual interaction presented by Guiard in [10]. Our exploration of using voice as the secondary modality to augment pen input draws inspiration from these projects exploring bimanual interaction.

## 4.2 Pressure Input

A number of input techniques have been proposed that leverage the pressure sensitivity of digital tablets. Ramos *et al*. present a study exploring the human ability to specify distinct stylus pressures and a taxonomy of "pressure widgets" that leverage this ability [28]. Another technique called Zliding, proposed by Ramos *et al.*, examines stylus pressure for control of the scale of zooming [29]. They compare their zooming method to using discrete keys and an isometric input device. They found their technique to be comparable to the other alternatives, and cite the users' preference for their technique in being able to perform the task with just one hand. Ramos and Balakrishnan have also suggested a technique called pressure marks, where the user varies the pen pressure throughout a stroke to issue commands [27]. They also situate a number of pressure-based interaction techniques within the context of an application for annotating digital video sequences [26]. Our work utilizes vocal parameters in a manner similar to pen pressure, i.e. as a continuous input modality, which suggests its potential applicability in similar interaction methods as those

described above involving pen pressure. Further investigation is necessary to determine how the controllability of vocalization compares to that of pen pressure.

## 4.3 Drawing

Several systems have been proposed to investigate the use of bimanual interaction techniques in the domain of artistic drawing and other creative activities. HabilisDraw [4] allows a user to manipulate various virtual drawing objects projected on a touch sensitive table using both hands. Raisamo presents the use of a trackball and a mouse to manipulate virtual carving sticks to sculpt away at a two dimensional block [30]. However, none of them offer continuous and simultaneous control of multiple drawing parameters in conjunction with a digital pen.

## 4.4 Pen + Voice Input

Many researchers have explored combining pen input with speech [6][17][24]. However, most of the work focuses on the integration of spoken commands with pen gestures, and does not explore the use of continuous non-linguistic vocalizations as a form of input.

There have been a number of voice-based systems that attempt to provide continuous control using non-linguistic vocalizations [11]. Igarashi et al. [15] proposed the use of non-linguistic vocal parameters as a means for achieving direct manipulation via voice. Subsequent systems have used voice for continuous input [3][22][31], however primarily as a replacement for mouse pointer control and not in the context of multimodal augmentation of a pointing device.

# 5. CONCLUSIONS & FUTURE WORK

In this paper, we have explored the potential for non-linguistic vocalization as an additional modality to the digital pen through our VoicePen prototype for the Tablet PC. Our feasibility study suggests our four interaction techniques may be useful and viable in a variety of digital pen tasks. We are also encouraged by our participants' enthusiasm about VoicePen and believe many more interaction techniques and tablet applications could be built using simultaneous non-linguistic vocalization and pen input.

In the future, we seek to better understand human capabilities for generating and controlling non-linguistic vocalizations in conjunction with pen input through empirical studies in the lab. For example, we hope to discover how many levels of loudness a person can comfortably and reliably generate. We also hope to look at how complex of a drawing people can replicate as well as to what extent people are capable of varying two or three parameters simultaneously and the associated learning curve. We would also like to directly compare existing input techniques to VoicePen. In particular, we would like to compare non-linguistic vocalization to pen pressure for controlling brush attributes such as thickness, color, curvature, and opacity, as well as to the use of keyboard or other button-based interfaces that allow the attribute value to be changed in a discrete or fixed increment. In our feasibility study we did not have a direct comparison of existing techniques to VoicePen techniques. However, participants did use vocalizations to control thickness in one task and a standard Tablet PC pressure scheme for controlling thickness in another task. One user commented on these two methods that using vocalizations was "pretty easy to control" and that it is "hard to control thickness of a line with pen pressure. I feel like I'm drawing the same stroke, but they come out different." Further investigations into such techniques and measures of human capability are needed to establish better foundation for designing effective future VoicePen interaction techniques.

Another future direction we would like to pursue is to create a more complete drawing application that utilizes VoicePen interactions as a test bed for experimenting with other ways of combining non-linguistic vocalization with the digital pen input. Such an application could also allow for a better understanding of VoicePen as an expressive tool for doing creative tasks. Based on our feasibility study, it is clear that VoicePen is a unique experience with its own affordances and benefits. In the drawing case in particular we are interested in exploring with artists and graphic designers how these techniques can support existing and new kinds of expressiveness.

Our VoicePen prototype and feasibility study suggest a great potential for simultaneous non-linguistic voice input with digital pen input. We found that participants quickly learned our interaction techniques, enjoyed using VoicePen, and found vocalization to provide new expressiveness. Participants developed their own techniques and strategies that worked well and were pleased with how their drawings matched the trial targets. We believe our interaction techniques are feasible and open the possibilities for a variety of new digital pen based input.

# 6. ACKNOWLEDGEMENT

We would like to thank Jeff Bilmes, Xiao Li, Jonathan Malkin, Richard Davis, Kayur Patel, and our study participants for their time and feedback.

# 7. REFERENCES

[1] Balakrishnan, R., and Hinckley, K. The role of kinesthetic reference frames in two-handed input performance. In *UIST '99: Proc. 12th annual ACM symposium on User interface software and technology* (New York, NY, USA, 1999), ACM Press, pp. 171–178.

[2] Bier, E. A., Stone, M. C., Pier, K., Buxton, W., and DeRose, T. D. Toolglass and magic lenses: the see-through interface. In *SIGGRAPH '93: Proc. 20th annual conf. on Computer graphics and interactive techniques* (New York, NY, USA, 1993), ACM Press, pp. 73–80.

[3] 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. In *HLT '05: Proc. conf. on Human Language Technology and Empirical Methods in Natural Language Processing* (Morristown, NJ, USA, 2005), Association for Computational Linguistics, pp. 995–1002.

[4] Butler, C. G., and Amant, R. S. HabilisDraw DT: a bimanual tool-based direct manipulation drawing environment. In *CHI '04: Proc. SIGCHI conf. on Human factors in computing systems* (New York, NY, USA, 1994), ACM Press, pp. 1301–1304.

[5] Buxton, W., and Myers, B. A study in two-handed input. In *CHI '86: Proc. SIGCHI conf. on Human factors in computing systems* (New York, NY, USA, 1986), ACM Press, pp. 321–326.

[6] Cohen, P. R., Johnston, M., McGee, D., Oviatt, S., Pittman, J., Smith, I., Chen, L., and Clow, J. Quickset: multimodal interaction for distributed applications. In *MULTIMEDIA*

*'97: Proc. fifth ACM international conf. on Multimedia* (New York, NY, USA, 1997), ACM Press, pp. 31–40.

[7] Davis, R. C., Saponas, T. S., Shilman, M., and Landay, J. A. SketchWizard: Wizard of Oz Prototyping of Pen-based User Interfaces. *UIST '07* (to appear).

[8] Fitzmaurice, G. W., Ishii, H., and Buxton, W. A. S. Bricks: laying the foundations for graspable user interfaces. In *CHI '95: Proc. SIGCHI conf. on Human factors in computing systems* (New York, NY, USA, 1995), ACM Press/Addison-Wesley Publishing Co., pp. 442–449.

[9] Gross, M. D., and Do, E. Y.-L. Ambiguous intentions: a paper-like interface for creative design. In *UIST '96: Proc. 9th annual ACM symposium on User interface software and technology* (New York, NY, USA, 1996), ACM Press, pp. 183–192.

[10] Guiard, Y. Asymmetric division of labor in human skilled bimanual action: The kinematic chain as a model. *Journal of Motor Behavior*, 19 (1987), 486–517.

[11] Harada, S., Landay, J. A., Malkin, J., Li, X., and Bilmes, J. A. The Vocal Joystick: evaluation of voice-based cursor control techniques. In *Assets '06: Proc. international ACM SIGACCESS conf. on Computers and accessibility* (New York, NY, USA, 2006), ACM Press, pp. 197–204.

[12] Hinckley, K., Pausch, R., Proffitt, D., Patten, J., and Kassell, N. Cooperative bimanual action. *In CHI '97: Proc. SIGCHI conf. on Human factors in computing systems* (New York, NY, USA, 1997), ACM Press, pp. 27–34.

[13] Hinckley, K., Czerwinski, M., and Sinclair, M. Interaction and modeling techniques for desktop two-handed input. In *UIST '98: Proc. 11th annual ACM symposium on User interface software and technology* (New York, NY, USA, 1998), ACM Press, pp. 49–58.

[14] Igarashi, T., and Hinckley, K. Speed-dependent automatic zooming for browsing large documents. In *UIST '00: Proc. 13th annual ACM symposium on User interface software and technology* (New York, NY, USA, 2000), ACM Press, pp. 139–148.

[15] Igarashi, T., and Hughes, J. F. Voice as sound: using non-verbal voice input for interactive control. In *UIST '01: Proc. 14th annual ACM symposium on User interface software and technology* (New York, NY, USA, 2001), ACM Press, pp. 155–156.

[16] Jul, S., and Furnas, G. W. Critical zones in desert fog: aids to multiscale navigation. In *UIST '98: Proc. 11th annual ACM symposium on User interface software and technology* (New York, NY, USA, 1998), ACM Press, pp. 97–106.

[17] Julia, L., and Faure, C. Pattern recognition and beautification for a pen based interface. In *ICDAR '95: Proc. Third International Conf. on Document Analysis and Recognition (Volume 1)* (Washington, DC, USA, 1995), IEEE Computer Society, p. 58.

[18] Kabbash, P., Buxton, W., and Sellen, A. Two-handed input in a compound task. In *CHI '94: Proc. SIGCHI conf. on Human factors in computing systems* (New York, NY, USA, 1994), ACM Press, pp. 417–423.

[19] Kurtenbach, G., Fitzmaurice, G., Baudel, T., and Buxton, B. The design of a GUI paradigm based on tablets, two-hands, and transparency. In *CHI '97: Proc. SIGCHI conf. on Human factors in computing systems* (New York, NY, USA, 1997), ACM Press, pp. 35–42.

[20] Leganchuk, A., Zhai, S., and Buxton, W. Manual and cognitive benefits of two-handed input: an experimental study. *ACM Transactions on Computer-Human Interaction (TOCHI)* 5, 4 (1998), 326–359.

[21] Malkin, J., Li, X., and Bilmes, J. Energy and loudness for speed control in the Vocal Joystick. In *IEEE Automatic Speech Recognition and Understanding Workshop* (November 2005).

[22] Mihara, Y., Shibayama, E., and Takahashi, S. The migratory cursor: accurate speech-based cursor movement by moving multiple ghost cursors using non-verbal vocalizations. In *Assets '05: Proc. 7th intl. ACM SIGACCESS conf on Computers and accessibility* (New York, NY, USA, 2005), ACM Press, pp. 76–83.

[23] Newman, M. W., Lin, J., Hong, J. I., and Landay, J. A. Denim: An informal web site design tool inspired by observations of practice. *Human-Computer Interaction* 18, 3 (2003), 259–324.

[24] Oviatt, S., Cohen, P., Wu, L., Vergo, J., Duncan, L., Suhm, B., Bers, J., Holzman, T., Winograd, T., Landay, J. A., Larson, J., and Ferro, D. Designing the user interface for multimodal speech and pen-based gesture applications: State-of-the-art systems and future research directions. *Human-Computer Interaction 15* (2000), 263–322.

[25] Owen, R., Kurtenbach, G., Fitzmaurice, G., Baudel, T., and Buxton, B. When it gets more difficult, use both hands: exploring bimanual curve manipulation. In *GI '05: Proc. 2005 conf. on Graphics interface* (Univ. of Waterloo, Waterloo, Ontario, Canada, 2005), Canadian Human-Computer Communications Society, pp. 17–24.

[26] Ramos, G., and Balakrishnan, R. Fluid interaction techniques for the control and annotation of digital video. In *UIST '03: Proc. 16th annual ACM symposium on User interface software and technology* (New York, NY, USA, 2003), ACM Press, pp. 105–114.

[27] Ramos, G. A., and Balakrishnan, R. Pressure marks. In *CHI '07: Proc. SIGCHI conf. on Human factors in computing systems* (New York, NY, USA, 2007), ACM Press, pp. 1375–1384.

[28] Ramos, G., Boulos, M., and Balakrishnan, R. Pressure widgets. In *CHI '04: Proc. SIGCHI conf. on Human factors in computing systems* (New York, NY, USA, 2004), ACM Press, pp. 487–494.

[29] Ramos, G., and Balakrishnan, R. Zliding: fluid zooming and sliding for high precision parameter manipulation. In *UIST '05: Proc. 18th annual ACM symposium on User interface software and technology* (New York, NY, USA, 2005), ACM Press, pp. 143–152.

[30] Raisamo, R. An alternative way of drawing. In *CHI '99: Proc. SIGCHI conf. on Human factors in computing systems* (New York, NY, USA, 1999), ACM Press, pp. 175–182.

[31] Sporka, A. J., Kurniawan, S. H., and Slavík, P. Whistling user interface (U3I). In *User Interfaces for All* (2004), pp. 472–478.