

Voice art: Investigating paralinguistic voice as a mode of interaction to create visual art

Dharani Perera
School of Engineering and IT
Deakin University, Melbourne,
Australia
++44 20841 14850
deehansika@gmail.com

R. T. Jim Eales
School of Computing Science
Middlesex University, London,
NW4 4BT, UK
++44 20841 14396
j.eales@mdx.ac.uk

Kathy Blashki
School of Engineering and IT
Deakin University, Melbourne,
Australia
+61 438 328 266
Kathy.blashki@gmail.com

ABSTRACT

In this paper, we report on our investigation into people's ability to use the volume of their voice to control cursor movement to create drawings. Early analysis of the results show changing the volume of the voice as an interaction method is a concept that is easily understood by users. People find changing the volume of their voice comfortable, natural and intuitive. With motivation, training and practice use of volume to control drawing tasks shows great promise. This is especially hopeful for artists with upper limb disabilities who show remarkable endurance, patience and determination to create art with whatever means available to them. We have also identified several design recommendations that may improve the control and performance of such a system. We believe that volume control has wider implications beyond assisting artists with upper limb disabilities. Some possible implications may be: as an alternative mode of interaction for disabled people to perform tasks other than creating visual art or for hands busy environments and as a voice training system for people with speech impairments.

Categories and Subject Descriptors

H.5 Information Interfaces and Presentation (e.g., HCI), H.5.2 User Interfaces, *Voice I/O*

General Terms

Human factors, Design, Experimentation

Keywords

Voice, volume, visual art, interaction, people with upper limb disabilities

1. INTRODUCTION

Visual art is an important activity for many people with physical disabilities. It provides a vehicle for creative expression, communication and sometimes financial independence. People with upper limb disabilities demonstrate remarkable endurance, patience and determination to adapt their remaining capabilities such as facial gestures, voice or foot movements to create various forms of art. Creating art using

existing tools can be a tedious and time-consuming process. One significant advantage of digital tools for disabled artists is that they require little or no help from a human assistant when compared to artists who use non-digital tools [8,9]. Research into alternative forms of digital input would suggest that there are significant opportunities to design and develop new ways in which artists with upper limb disabilities can be supported by digital technologies [3, 6, 7].

Voice recognition technologies have proven to be a particularly promising mode of interaction. However, speech commands lack sophistication when it comes to creating visual art. Artistic expression generally requires an interaction style that can support a certain freedom of movement.

"With dragon dictate you have to remember very precise commands, and it's very slow and restrictive in what you can say...I don't have the patience for it. And when I am planning my art work I don't want to be thinking about every word I say." Tom Yendell (personal communication).

Vocal paralanguage are the general sounds people make that do not have speech elements, for example laughter, whistling and hissing. It can be characterised by numerous properties such as volume, pitch, phase and timbre. The user may intentionally change these properties that may be mapped to different actions. Most people with upper limb disabilities have vocal paralanguage abilities; unlike speech recognition it is not governed by a need for a language model and is universal for all languages.

Lately, there has been significant research interest in audio to visualisation mapping. Various strategies have been employed to establish novel mapping techniques between visual properties and voice characteristics. Al Hashimi [1], Igarashi and Hughes [4] and Sporka *et al.* [11] have successfully investigated paralinguistic voice interfaces for control of cursor movement, scrolling, clicking on icons and control of output devices such as plotters.

This paper presents an innovative design for a drawing system that may be controlled by varying the volume of different sounds made by people. The aim of this research was to investigate people's ability to control the volume of their voice to create drawings on the screen for artistic expression.

2. Design Concept

Our design concept is based on people's ability to change the volume of the sounds they make. People often change the volume of their voice in day to day communication and are able to do it intentionally [2, 5, 10]. Volume is also a continuous vocal characteristic that goes beyond the capabilities of discreet speech commands. Therefore it may be better suited for artistic expression which generally requires a certain freedom of

movement. This study investigated if people are able to change the volume of their voice to control the direction of motion of the cursor to create drawings on a screen.

2.1 Prototype Design

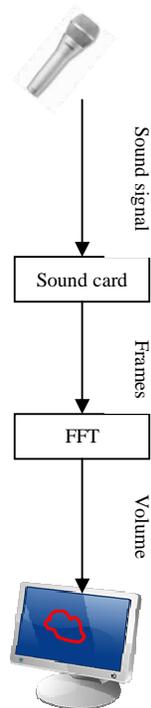


Figure 1. Overview of the system.

Figure 1 shows the overview of the system. The users make sounds into the microphone. The voice signal captured by the microphone is analysed using the Fast Fourier Transformation (FFT) algorithm. The algorithm transforms the voice signal into frequency-based data. The prototype is programmed to compare the volume levels (energy transfer) at each frequency band. The highest volume level is mapped to the direction of motion. The mapping is based on a circular gauge, such as a speedometer. The mapping allows the user to control the full range of 360 degrees. The lower volumes are mapped to lower angles and higher volumes to higher angles. The volume range is set to the users' comfort range (described further in section 2.1.2.1). The range is divided into equal steps which will be mapped to equally divided angles. A voice sample is averaged over 2000 counts, as this proved to be the best reaction time during development. As no noise cancellation is built in, the described prototype is effective only for environments with low background noise.

2.1.1 Experimental setup

The prototype was installed on a machine with an Intel® Pentium® M 740 processor, 1.73 GHz and 512 MB of system RAM. An Electret® condenser microphone was used to capture the voice. Figure 2 shows the interface.

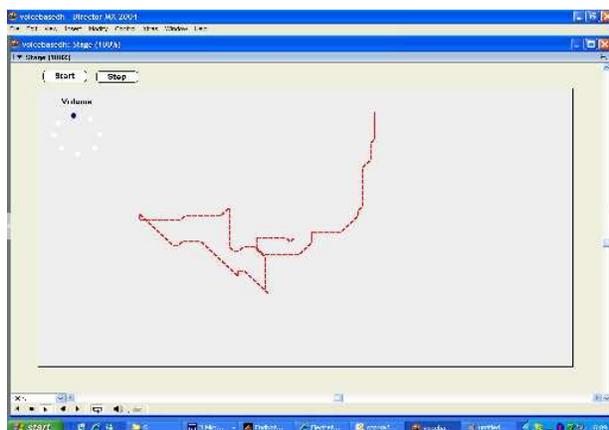


Figure 2. Screen capture of the interface.

In the top left corner of the screen the participants were provided with a circular gauge that showed the direction their voice was currently controlling. The cursor started to move from the middle of the screen. The "Start" and the "Stop" buttons started and stopped the voice capture. If at any

point the participant stopped the voice capture in the middle of the drawing, the cursor remained in the current position and continued to move from there when the voice capture was started again. A video camera was set up to capture the screen and the sounds made by the participants and the recordings were later used for analysis.

2.1.2 Procedure

The experimental process consisted of three parts. The participants first filled out a preliminary questionnaire that was used to gather some basic personal information such as gender, age and any voice impairments. Then they completed the experiment (section 2.1.1.1). After the experiment the investigator asked a few open ended questions about thoughts and feelings on the system. These interviews were tape-recorded. The recordings were transcribed later for analysis. Four males and two females, aged 20 to 65 years, were used to test the system. None had upper limb disabilities nor speech disorders or vocal training. Only one participant had experience with a speech recognition based system.

2.1.2.1 Experiment

The investigator explained the purpose and procedure of the experiment. Then, the participants were given 10 to 15 minutes practice time with the system. During the practice time they were exposed to several previously set volume ranges and were asked to choose the range that best suited their voice. Once they were comfortable with the experimental setup the testing started. The participants were exposed to four experimental conditions (Table 1). For example the 8 directions allowed the participants to move in steps of 45 degree angles, while the 20 directions allowed movement in steps of 18 degree angles. The participants completed four tasks under the four experimental conditions, 16 tasks in all. Table 1 shows the experimental conditions and the tasks.

Table 1. Experimental conditions and tasks

Experimental condition	Tasks for each condition
8 directions	Draw square
12 directions	Draw circle
16 directions	Draw triangle
20 directions	Draw cloud

The participants were told that any sound they made would be captured by the system and result in a modification of the drawing. The participants completed all the tasks for one experimental condition and then moved on to the next experimental condition in the order shown in Table 1. The participants were allowed to take breaks anytime during the experiment and ask questions from the investigator. There was no time limit set to complete the tasks. On average participants took approximately 30 minutes to complete the experiment.

3. Results and Discussion

3.1 Input strategies

The participants used two input strategies to control the drawing task.

Strategy One: Control via changing the volume of a single sound. Only one participant used this method. The "Ahhhhhhhh"-sound was used. During the practice session the participant learned how to lower and increase the volume of this sound to control the direction.

Strategy Two: Use of multiple sounds to control direction. During the practice sessions the participants experimented with several sounds and identified three to four sets of sounds that would control the four main directions (up, down, left, right). The participants would then change the volume of the identified sounds to acquire neighbouring directions.

Both strategies worked with equal merit. However, two participants who used the second strategy got confused during the testing as the identified sounds wouldn't always result in the desired direction. This could be due to the fact that the same sound can be produced at different volumes and control of the volume takes more practice.

Generally the participants were quick to understand the concept behind the design. They identified ways to change the volume to control the directions and found varying the volume of their voice comfortable and intuitive.

However, volume control using both strategies suffered the following drawbacks.

- It is very difficult to produce the correct volume level to obtain the desired direction at the initial burst of the sound. In most cases the participants made a sound that they thought would result in the desired direction but then had to correct the volume. Since the drawing continued during this correction time it resulted in undesired curves.
- It is difficult to hold the volume of the voice within a constant volume range. The volume oscillated between the ranges resulting in incorrect drawings on the screen. The oscillations were more frequent in higher angles as higher angles resulted in narrower range values.
- The participants found it difficult to hold the desired volume for a long period of time. The volume level dropped over time, especially for higher volumes. In many instances to the human ear it seemed like the participants were holding the same volume level, but the volume level captured by the system recorded a drop. This may be due to the fact that the microphone captures the varying pressure waves in the air and converts them into different electric signals. Over time the pressure caused by the voice may drop, the system was more sensitive to these drops than the human ear. This led to a lot of confusion among the participants.

3.2 Visual Feedback

The system provided two types of visual feedback: the drawings on the screen and the circular gauge in the top left corner of the screen (Figure 2). Participants were divided on the use and the usefulness of the visual cues. Three out of the six participants only used the drawing on the screen for visual guidance. Out of the 3 participants who used the circular visual gauge, one used it extensively during the experiment, the second one mainly during practice session as a tool to identify the sounds that would control a desired direction and the last took it as an added aid to monitor oscillations and drops in volume. All participants, while performing the task, produced volumes out of the selected range. The system provided no feedback for this situation which confused the participants.

3.3 Shapes

The square shape was best drawn with the eight direction experimental condition. While all participants found it difficult

to control the volume to draw a perfect square, many were able to draw a shape that closely resembled it.

It was very difficult to obtain a smooth circle shape. Most participants drew a shape that closely resembled a "puddle of water". The circle shape was best drawn with the 20 direction experimental condition.

The triangle proved to be the most difficult to draw. None of the experimental conditions seemed to be of any particular help. The participants found it especially difficult to control diagonal directions. They had more control over the four main directions. This may be due to the fact that during the practice sessions they spent time mainly learning to control these directions.

Participants were most comfortable with drawing the cloud shape. The oscillations and the dropping of the volume complimented the task. Most participants deliberately oscillated the volume of their voice in a playful manner to obtain the cloud shape. The shape was best drawn in the 20 direction condition.

3.4 Number of Directions

The eight directions experimental condition was best to draw square shapes.

As the number of angles increased, it was easier to draw circular and irregular shapes. The undesired "jaggies" caused due to oscillations and dropping of the volume were smoother for higher number of angles.

However, a higher number of angles resulted in frequent oscillations as the bandwidth of the volume ranges was narrower.

3.5 Fatigue

All participants reported that performing the tasks was not tiring and the number of tasks was not overwhelming. They agreed that the opportunity to rest during the task helped considerably. However, the video recordings showed that most participants got tired trying to hold the high volumes.

Heavy breathing and sighing when the microphone was close to the mouth was picked up by the system and modified the drawings. Coughing, sneezing didn't affect the system as most participants used the "Stop" button in these situations.

3.6 Guidance, learning curve and practice

Initially, the investigator didn't give any guidance to the participants. However, if the participant sought assistance to perform a certain task the investigator suggested several strategies as a help.

During the 10 to 15 minutes practice time they were able to identify ways to control their voice to move in the four main directions. However, it is evident that a lot more training and practice is needed to obtain a level of control that is needed to create accurate drawings of the given shapes. Most participants agreed that they would improve with time.

4. Implications for Technology

Changing the volume of the voice as an input method is a concept that is easily understood by users. People find changing the volume of their voice, comfortable, natural and intuitive. With motivation, practice and training, use of volume to control drawing tasks shows great promise. This is especially helpful for artists with upper limb disabilities who show remarkable endurance, patience and determination to create art with whatever means available to them [9].

When developing voice based systems better averaging functions should be built-in to control the undesired oscillations of voice. The system should also have improved functions to cancel out unwanted noise such as heavy breathing and sighing.

The visual feedback of the drawing is sufficient for direction control. However it would be useful to provide a visual gauge that shows the dropping of volume, oscillations in volume and instances where volumes outside of the set volume ranges are made.

A microphone that captures the sound signals in a different mechanism to the vibration of air; such as a throat microphone may be less prone to capturing volume drops and may provide more intuitive control. Therefore it will be useful to test the system with different types of microphones.

It is very difficult to produce the correct volume level to obtain the desired direction at the initial burst of the sound. In most cases the participants made a sound that they thought would result in the desired direction but then had to correct the volume. Since the drawing continued during this correction time it resulted in undesired curves. A possible design solution could be to let the user first obtain the desired direction and then move the cursor to create the drawing. In the event the voice drops below the desired volume range the drawing will stop and the user will get chance to set the direction. This method may be most useful for drawing straight lines and may hinder the drawing of curvy lines and the deliberate rise and drop of volume to create more abstract shapes like a “cloud”.

It is important to give users a choice to select different modes that may suit different situations. A few of such choices could be:

- Fixing the direction to draw straight lines or no fixing for more curvy and abstract shapes.
- Number of directions 8, 12, 16, 20, ... N.
- Preferred volume range.

The use of two microphones, one to control the angles in the right and one for the left may provide more intuitive control of direction. This would also reduce the need to use very high ranges of volume that most participants had trouble with.

5. Conclusions and future directions

We have investigated people’s ability to use the change of volume to control the direction of motion to create drawings in a digital medium. The study captured the instant usability of the system. However on reflection it seems that with motivation, training and practice use of volume to control drawing tasks shows great promise. We believe this is more relevant for artists with upper limb disabilities who have much more time to develop their skills. In the future we plan to build the suggested design recommendations into the system and test it for long term learnability. Volume is a continuous vocal characteristic that goes beyond the capabilities of discreet speech commands. Therefore it is better suited for artistic expression that generally requires a certain freedom of movement. We also believe that volume control has wider implications beyond assisting artists

with upper limb disabilities. Such possible implications may be: as an alternative mode of interaction for disabled people to perform tasks other than creating visual art or for hands busy environments and as a voice training system for people with speech impairments.

6. REFERENCES

- [1] Al Hashimi, S, Blowtter: a Voice-Controlled Plotter, in *Proceedings the 20th BCS HCI Group Conference in Cooperation with ACM*, London, UK, 2006.
- [2] Banse, R. and Scherer, K. R. 1996, Acoustic profiles in vocal emotion expression, *J Pers Soc Psychol*, 70, 3, (1996), 614-36.
- [3] Gorodnichy, D. C., Malika, S., and Roth, G. Nouse: Use your nose as a mouse -a new technology for hands-free games and interfaces, in *Proceedings of the International conference on vision interface (VI2002)*, Calgary, Canada, 2002, 354-361.
- [4] Igarashi, T. and Hughes, J. F. 'Voice as Sound: Using Non-verbal voice input for interactive control', in *Proceedings of the 14th Annual Symposium on User Interface Software and Technology UIST'01*, ACM Press, Orlando, Florida, USA, 2001, 155-156.
- [5] Nass, C. and Brave, S. *Wired for speech: how voice activates and advances the human-computer relationships*. MIT Press, Cambridge, UK, 2005.
- [6] Oviatt, S., and Cohen, P. Multimodal interfaces that process what comes naturally. *Communication of the ACM* 42, 3(2000), 45-53.
- [7] Oviatt, S., DeAngeli, A., and Kuhn, K. Integration and synchronization of input modes during multimodal human-computer interaction, in *Proceedings of the SIGCHI conference on Human factors in computing systems*, (Atlanta, Georgia, United States, 1997), ACM Press, 415-422.
- [8] Perera, D., Blashki, K., and Li, G. Visual art for the disabled: A user requirement study on the needs of visual artists with upper limb disabilities, in *Proceedings of the 2006 IRMA International Conference*, (Washington, DC, USA, 2006).
- [9] Perera, D., Eales, R. T. J., and Blashki, K. The Drive to create: An investigation of tools to support disabled artists, to appear in the *Proceedings of the Creativity and Cognition 2007*, (Washington, DC, USA, 2007), ACM Press.
- [10] Scherer, K. R. 1986, Vocal Affect Expression: A Review and a Model for Future Research, *Psychological Bulletin*, 99, 2, (1986), 143-165
- [11] Sporka, A. J., Kurniawan, S. H. and Slavik, P. Acoustic control of mouse pointer, *Universal Access in the Information Society*, 4, 3,(2005), 237-245.