Blowtter: A Voice-Controlled Plotter

Sama'a Al Hashimi
Lansdown Centre for Electronic Arts, Middlesex University
Hertfordshire, United Kingdom
Samaa.alhashimi@gmail.com

The use of microphones as an input device has predominantly represented a linguistic rather than a paralinguistic gateway to computer-based technologies. However, speech is not the only channel of communication that the voice may employ. Moreover, the output of many voice-controlled applications tends to be confined to on-screen displays. By contrast, this paper investigates the combination of non-verbal vocal input with physical outputs. The paper reviews existing work which exploits non-verbal vocalisation, and discusses the relative merits of the verbal and non-verbal as inputs. It presents the *mic-board,* an attempt to find a new application for the microphone. The *mic-board* is used to control *Blowtter*; a voice-controlled plotter. The main aim is to introduce the concept of a *voice-physical* installation and to explore the use of the non-verbal channel of voice as a complementary input mechanism to the verbal channel.

Keywords: voice, paralanguage, plotter, voice-physical, speech-recognition, non-speech recognition, mic-board

1. INTRODUCTION

The availability of a wide variety of input mechanisms today reflects the high potential of the human body as a rich source of input. Recently, voices, fingers, hands, and eyes have all been explored and employed in different multimedia applications as an input. This has, to a certain extent, transformed part of the computer monitor into a mirror that reflects the user sitting in front of it while also displaying the visuals processed behind it. A lot of these techniques, however, are not enough to break the barrier between the user and the computer. The widespread reliance on the visuals, and hence the desktop screen holds many users back from the perception of an amorphous computer that any device may incorporate or embody. Only recently has more attention been driven towards what O'Sullivan and Igoe refer to as "Physical Computing" or using computers to affect the physical world:

> When asked to draw a computer, most people will draw the same elements: screen, keyboard, and mouse. When we think "computer," this is the image that comes to mind. In order to fully explore the possibilities of computing, you have to get away from that stereotype of computers. You have to think about computing rather than computers. Computers should take whatever physical form suits our needs for computing. [8]

With the increasing ubiquity of computer vision and video tracking, physical movement has almost always been associated with and expected from the user. When expected from the computer, physical movement is often linked with android (humanoid robot) behaviour either in reaction to a remote control or to spoken commands. In this paper, however, I aim to reflect my interest in programming computers to *complement* users' deficiencies rather than merely to imitate their abilities. *Voice-visual* applications, for instance, may counterbalance our inability to measure voice characteristics and produce visuals in reaction to them. Igarashi and Hughes refer to the use of non-verbal voice to control interactive works as "voice as sound" [4]. On the other hand, our inability to naturally use voice to physically control and affect real inanimate objects calls attention to the development of *voice-physical* applications. It is important that developers adequately exploit the computer's capacity to respond multimodally to voice not only by visuals but also by movements, odours, or any kind of physical output. According to Levin, a computer may cause "uniquely ephemeral dynamic media to blossom from the expressive 'voice' of a human user" [5]. *Voice-physical* installations are likely to prove a fruitful expansion of the possibilities already shown to be inherent in voice-visual and other existing forms of voice-controlled installations. Such an expansion seems to have a natural synergy with the move towards "The Invisible Computer" described by Norman [7].

In light of this, I was prompted to investigate the possibility of programming voice to control inanimate objects in what could be referred to as *vocal telekinesis*. An initial development in this direction was *sssSnake*; a *voice-physical* version of the classic 'Snake' game. In this two-player game, one player makes an 'ssss' sound to control a virtual snake projected on the surface of an installation table. The other player makes an 'ahhhh' sound to move a real coin placed on top of the table. The 'sss' and 'ahh' sounds are distinguished through frequency-range differentiation rather than speech recognition. The game is based on utilizing the difference in frequency between the high-pitched 'ssss' and the relatively low-pitched 'ahhh'. The position of the players round the table, which is fitted with a microphone on each of its four sides, determines the direction of the coin or snake. A plotter is hidden below the table. Its head, to which a magnet is attached, is programmed to move away from the source of the 'ahhh' sound and pull the coin with it. The development and usability testing of *sssSnake* has inspired and paved the way for the development of *Blowtter* which will be discussed thoroughly in the rest of this paper.

2. BLOWTTER

Ali Abbas, an Iraqi orphan whose arms were amputated following an attack in Baghdad, was later fitted with prosthetic limbs. Later on, he started drawing palm trees and temples with his feet [2]. I remembered him when the plotter that I was testing for *sssSnake* started moving in reaction to voice. This led to the development of *Blowtter*.

2.1 Concept

Blowtter is a voice-controlled plotter that allows a disabled user to blow into a *mic-board* in order to draw. The *mic-board* is a small square board that consists of four microphones, one on each of its four sides. Blowing into one of the microphones moves the head of the plotter in the opposite direction as if pushing it. The notion of *blowing* fits neatly with the action of *pushing* the head. Blowing is used as an input in order to facilitate the directness and continuity required for drawing which is not easily achieved by using a spoken command repetitively such as saying "move…move…move" or "move, 10". Igarashi and Hughes further explain the advantage of this technique:

> […] one can say “Volume up, ahhhhhh”, and the volume of a TV set continues to increases while the “ahhh” continues. The advantage of this technique compared with traditional approach of saying “Volume up twenty” or something is that the user can continuously observe the immediate feedback during the interaction. One can also use voiceless, breathed sound. [4]

Where appropriate, speech commands are also used in *Blowtter*. For example, "start plotter" and "end plotter". Saying "up" raises the head and allows for moving it without drawing. Saying "down" moves the pen into the paper. Saying the number of the pen causes it to be selected.

The advantage of using blowing lies in the possibility of targeting it with one microphone without interference through nearby microphones. Blowing is perceptually voiceless despite the fact that the microphones may detect it and the computer may measure its voice characteristics. Unlike voiced sounds which involve the vibration of the vocal folds while being generated, blowing and other unvoiced sounds only involve air passing through the larynx without causing the vocal folds to vibrate. Furthermore, the range of amplitude values that voiced sounds can produce is wider and louder than the range of amplitude values generated by blowing. This fact is exploited in specifying a higher threshold for speech than for blowing. This will be explained in detail in the following section.

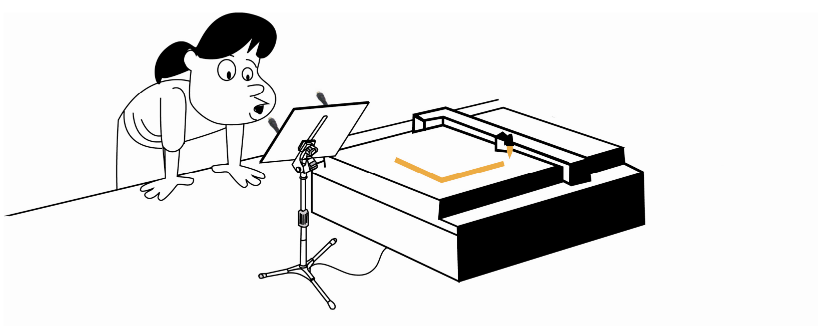

FIGURE 1: *Blowtter*: a voice-controlled plotter that allows a perhaps disabled user to blow into a *mic-board* in order to draw.

2.2 Technical Details

The main hardware requirements for the development of *Blowtter* include a DXY-880 Roland plotter, four USB audio adapters, four microphones, and a fast computer. The application has been programmed in Macromedia Director/Lingo. Three Xtras (external software modules) are used: asFFT, Chant, and Direct Communication.

2.2.1 Using asFFT Xtra to detect voice characteristics

The vocal signal is analysed using the Fast Fourier Transform (FFT) algorithm which the asFFT Xtra employs. The Xtra transforms the voice signal into frequency-based data. Through programming it to measure and compare the different volume levels at each microphone, the software recognizes which microphone of the four the user is blowing into. The microphone at which the maximum amplitude is detected is the only one that the software uses to detect voice and to determine the direction for moving the pen. In other words, even when the user blows hard enough for nearby microphones to pick up the sound, the software is programmed to compare the amplitudes and to only react to the microphone through which the loudest amplitude is detected. To avoid the head reacting to any voice input other than blowing, an amplitude threshold is specified in order not to allow noises above it to move the head. When a volume above the threshold is detected, the software assumes that the user is speaking rather than blowing and the speech recognition Xtra, Chant, is activated. Furthermore, to avoid the head reacting to soft ambient noises, another lower threshold is specified. Any noises below it are completely ignored. The Hewlett-Packard Graphics Language (HPGL) was used to program the plotter to move its head in response to blowing.

2.2.2 Using Chant Xtra to recognize speech

Chant Speech Kit, a speech-recognition Xtra for Director, is used to recognise the commands spoken by the user. The Xtra supports a number of speech recognition applications, in this case Microsoft SAPI. It consists of a command, grammar, and dictation vocabulary list. For *Blowtter*, however, only the command list is enabled and customized to consist only of the commands that control the plotter. Thus, detection errors are minimized.

2.2.3 Using Direct Communication Xtra to establish a parallel or serial connection with the plotter

Direct Communication Xtra allows Director to communicate with an external device either through the serial or, as in this case, the parallel port. Care must be taken to match the settings at the computer and at the plotter end.

3. THE MICBOARD AS AN ALTERNATIVE INPUT MECHANISM

The investigation of paralinguistic vocal control of interactive media does not only involve singing, humming, or making voiced sounds as ways to control a computer. Blowing has also been utilized innovatively in very few interactive works. The *Yacht* game by Nintendo DS, for instance, allows the player to blow into the microphone to propel a boat and steer it in the sea. The weight of breath determines the speed of the yacht. Another Nintendo game, *Candles*, also allows the player to blow into the microphone to blow out candles; the larger the size of the candle, the harder the player must blow. In both cases, users probably believe that the velocity of their breath is the determining factor, whereas it is really the volume of the sound created by blowing on the microphone.

Blowing "Windows" by Matsumura at The Royal College of Art, London, is another application that employs blowing [6]. It allows the user to blow into one end of a hose of which the other end is directed towards a computer's screen in order to control, move, and rearrange desktop icons. The duct contains a wireless microphone that detects the blow and measures its intensity. This intensity as well as the size of the file represented by the icon determines the speed at which the icon moves. The hose also contains tilt switches that detect the angle at which the hose is held and determine which side of the desktop to rearrange.

Another remarkable application of blowing is *Kirifuki* which allows for interaction with visual desktop objects by inhalation and exhalation. The system consists of a breath microphone switch containing a Polhemus sensor that detects the orientation of the user's head, a projector that projects the desktop on a desk, and a magnetic gyroscope [3]. The microphone differentiates between inhaling and exhaling by means of comparing between their acoustic signals. Iga and Higuchi claim that the randomness of the exhale signals is less than that of the inhale signal, and by measuring and comparing the local peak numbers the system recognizes the sound input as either an inhale or an exhale [3]. The implementation of this technique allows for a variety of interaction mechanisms. The technique can be applied to make a blow diffuse the icons around the mouse pointer, while an inhale may assemble these icons again. Another interesting application is programming the blow to cut a virtual object, while an inhale pastes it again. If the object being manipulated is three-dimensional, an inhale deflates it while an exhale inflates it. The system may also be used for drawing where an exhale sprays a shape while an inhale erases it.

One last noteworthy installation is *Blow Up*. Developed by Scott Snibbe, this installation allows the sender to blow into a group of twelve small impellers that control an array of twelve large fans which replicate and magnify the speed and movement of the small impellers on the receiver's end [9]. The array of small impellers is placed in one side of the gallery, while the fans they control are placed in another side. Through the development of this installation, Snibbe seems to emphasize that the visual and physical impact of breath on nearby objects can be more obvious than the audible impact. He believes that the physical interaction of our voices and breaths with surrounding media has a significant influence on our inference about the existence of these body activities [9].

In addition to the exploitation of blowing as an input source, *Blowtter* involves a utilization of the un-voiced-ness of blowing to propose a new use for the microphone. The fact that blowing is un-voiced makes it possible to place the four microphones used to control *Blowtter* in a very close position to each other on the *mic-board*. The size of the board is around 12x12 cm and the distance between one microphone and the other is around 11 cm. This minimizes the need to move from a microphone to another where only the user's face is expected to move slightly to direct the air stream into one of the microphones. For such an application designed especially for disabled users, minimizing movement is very necessary. The *mic-board* may be used by disabled users as an alternative to a joystick in some modified versions of existing games. Further implementations may involve increasing the numbers of microphones. Hence, several microphones may be placed on a circular rather than a square board or even forming a matrix of microphones. Blowing into a microphone on the right side of the board and then into a microphone on the left side would draw a straight line from right to left. Moving the head slightly in a certain pattern while blowing sequentially into one microphone after another would draw the pattern or shape that the user's head composes while moving. This would increase the opportunity of allowing smoothly graduated control.

4. CONCLUSION

The development of *Blowtter* is still in progress and user-testing it will certainly reveal many unexpected results. One of the main aspects that I aim to investigate while testing it is the integration of certain paralinguistic components as a complementary input mode to certain speech input applications in order to create a synergistic combination that might let the strengths of each mode overcome the weaknesses of the other. One possible outcome could be a *voice-physical* avatar or robot that is pre-programmed to recognize paralinguistic and linguistic utterances and physically display their equivalent in sign language. This robot may be portable or may also be connected to or placed next to a television for the hard of hearing. Comparing this interactive work with other voice-controlled applications that only involve non-speech input will make it possible to research further paralinguistic voice-related dimensions. One facet that requires further research, for instance, is the use of non-speech voice as an alternative input in comparison to the use of it as a complementary input mechanism to speech recognition.

On the basis of an empirical evaluation of previous work [1], it seems more appropriate to employ paralinguistic input alone to control a game or an entertaining work than to control a practical and serviceable application. Interaction with a game allows for less accuracy than for a functional application. A game usually offers a chance to try again after losing. Error and loss are meaningful components of the action in most games but surely not part of any practical application. For this reason, a game is especially likely to benefit from exclusive use of paralinguistic input, whereas in the case of a practical everyday application, this kind of input is likely to be of greatest benefit when used with a complementary input mechanism.

One advantage of using blowing in *Blowtter* rather than using voiced sounds as in *sssSnake* is that it ensures accuracy by minimizing microphone interference. The underlying technique is shared in both applications, but the extent to which a technical inaccuracy is perceptible and tolerable by the user is not. Because of a limitation of asFFT when used with multiple microphones, two microphones could wrongly detect the same amplitude level even when the player was much closer to one microphone than the other while making loud sounds in *sssSnake*. Occasionally, this complicated the amplitude comparison operation and caused the coin to momentarily deviate. That was, nonetheless, hardly noticeable by the players who were, anyway, running and laughing. For a disabled user sitting still and using *Blowtter* to draw accurately, however, such deviation is intolerable and unquestionably perceptible. Therefore, the un-voiced-ness of blowing and the inherent notion of having to bring the mouth very close to the 'blown' object, which in this case is one of the microphones, make it possible to obviate this problem.

Another important aspect to keep in mind while developing a voice-controlled work is its context. Since fun and laughter are part of an entertaining system, making non-speech sounds to control a game is reasonable and even attention-grabbing to passers-by. On the other hand, a user interacting with a practical voice-controlled interface will most probably not want an audience watching and laughing. As *Blowtter* is not designed as a game, this is another reason behind choosing blowing rather than vocalizing to control it.

The use of *Blowtter* as a drawing tool for physically impaired users could have significant potential. Non-speech voices can provide this user segment with a new dimension for exploring their artistic talents and communicating their thoughts. Exploring the untapped dimensions of voice-control might empower them in other ways too.

REFERENCES.

[1] Al Hashimi, S. (2005) Beyond Using Voice as Voice. *In Proceedings of the 16th International Conference for Advanced Studies in Systems Research, Informatics and Cybernetics*. Baden-Baden, Germany

[2] Gordon, Y. (2006) Ali has the World of Art at his Feet. Kingston Guardian (http://www.kingstonguardian.co.uk/display.var.704393.0.0.php).

[3] Iga, S. and Higuchi, F. (2002) Kirifuki: Inhaling and Exhaling Interaction for Entertainment Systems. *Transactions of the Virtual Reality Society of Japan TVRSJ.7* No.4, 445-452.

[4] Igarashi, T. and Hughes, F. (2001) Voice as Sound: Using Non-verbal Voice Input for Interactive Control. *Proceedings of the 14th Annual Symposium on User Interface Software and Technology, ACM UIST'01*, 11-14 November, pp.155-156. Orlando, Florida.

[5] Levin, G. (1999) Interface Metaphors and Signal Representation for Audiovisual Performance Systems (http://acg.media.mit.edu/people/golan/thesis/proposal/).

[6] Matsumura, E. (2005) Blowing 'Windows'. (http://www.interaction.rca.ac.uk/alumni/04-06/Eriko/html/01_puppet.htm).

[7] Norman, D. (1999) *The Invisible Computer: Why good products can fail, the personal computer is so complex and information appliances are the solution*. MIT Press, Cambridge.

[8] O'Sullivan, D. and Igoe, T. (2004) *Physical Computing: Sensing and Controlling the Physical World with Computers*. Thomson Course Technology PTR, Boston.

[9] Snibbe, S. (2005) Blow up. (http://www.snibbe.com/scott/breath/blowup/index.html).