Univ Access Inf Soc (2006) 4: 237–245
DOI 10.1007/s10209-005-0010-z

SHORT PAPER

Adam J. Sporka · Sri H. Kurniawan · Pavel Slavík

Acoustic control of mouse pointer

Published online: 29 September 2005
© Springer-Verlag 2005

Abstract This paper describes the design and implementation of a system for controlling mouse pointer using non-verbal sounds such as whistling and humming. Two control modes have been implemented—an orthogonal mode (where the pointer moves with variable speed either horizontally or vertically at any one time) and a melodic mode (where the pointer moves with fixed speed in any direction). A preliminary user study with four users indicates that the orthogonal control was easier to operate and that the humming was less tiring for the users than whistling. The developed system may contribute as an inexpensive, alternative pointing device for people with motor disabilities.

Keywords Pointing devices · Motor disabilities · Acoustic input · Assistive technologies · Melodic interaction

1 Introduction

The research and development of assistive technologies is currently an important part of the field of human–computer interaction. Much effort has been made to create some forms of assistance for computer users with disabilities that reduce their capabilities to use conventional devices. Over the years, numerous alternative user interfaces for computer users with motor disabilities have been invented and reported. Typical solutions include devices that utilize speech recognition techniques (e.g., IBM ViaVoice), eye-trackers and various breath controllers (e.g., sip-and-puff controller [11]). Speech recognition software is known to be particularly useful for textual input [3], while additional devices are usually employed as pointing devices, allowing the control of the mouse pointer [16]. These special hardware devices are usually less affordable than traditional devices such as mice and keyboards.

An alternative to these methods may be found in the use of non-verbal acoustic sounds, such as whistling or humming. It has been demonstrated by Igarashi and Hughes [10] that the non-verbal acoustic sound produced by the users may be used to control various parameters of a system.

This paper presents an innovative pointing device that may be installed on a standard home computer. The system, called *Whistling User Interface*, allows the users to control the on-screen mouse pointer through non-verbal sounds like whistling, humming or hissing.

As opposed to the sip-and-puff controllers, the method proposed in this paper does not demand users to maintain physical contact with the input device. The system can be implemented on a standard PC or PDA device for mobile applications and does not require any specific hardware other than a microphone and a sound card that is able to digitize the audio input signal.

1.1 Related work

The use of sound modality has been frequently addressed in recent HCI research. There are many kinds, implementations and applications of user interfaces where sound is used to mediate or enhance the information presentation and navigation.

Information sonification techniques allow multidimensional data to be presented to the users by synthesizing sound signals. The parameters of these such as

A. J. Sporka (✉) · P. Slavík
Department of Computer Science and Engineering,
Faculty of Electrical Engineering,
Czech Technical University in Prague,
Karlovo náměstí 13, Praha 2, 12135 Czech Republic
E-mail: sporkaa@fel.cvut.cz
Tel.: +420-224-357470
E-mail: slavik@fel.cvut.cz
Tel.: +420-224-357617

S. H. Kurniawan
School of Informatics, University of Manchester,
PO Box 88, M60 1QD Manchester, UK
E-mail: s.kurniawan@co.umist.ac.uk
Tel.: +44-161-2008929

pitch, volume, timbre, etc. are modulated over time to reflect the development of the data along a progress of an independent variable. This enables the users to detect the underlying emerging patterns. To give an example, Barra et al. [2] demonstrates the use of such a scenario on the sonification of web server. The pitch of *tremolo* tones played by a synthesized string-ensemble indicates the current workload of the hardware; while separate *staccato* tones represent individual requests for any particular document [1] that uses sonification to present the activity of the human brain. Different EEG values were mapped on different parameters of the sound signal. These parameters could be used in real time to synthesize a sound that could reflect the processes of the brain. The concept of audio progress bar with spatial effect has been discussed [18]. Franklin and Roberets [6] demonstrate the possibility to use sonification to display pie charts to the visually impaired.

Sound is also often being used to enable or support navigation within the user interface. Special acoustic patterns—often referred to as auditory icons or earcons [4, 8] and further investigated [12–14]—are employed to indicate the current context of the user interface or attract the users' attention on various system events. As an example, users may be continuously informed how close their task is to completion, or from whom they are receiving an instant message.

The methods of user data input and application control by means of sound produced by the user have been investigated well, especially regarding the design and applications of the speech recognition. These have been reported to be successful for both text input [3] and GUI navigation [5]. Rabiner and Juang [15] provide a good introduction to the field.

The use of non-speech audio input has been demonstrated, [9] where a game controlled entirely by singing is described. As already mentioned, Igarashi and Hughes [10] describe a hybrid input method based on combination of speech recognition and singing: a spoken command is followed by a tone that may specify the command parameters depending on its pitch and length.

The method proposed in this paper extends the existing range of input techniques in which the non-verbal audio is used. This paper provides an overview of the interaction method and the results of a preliminary user study.

2 The input method

The overview of the system used to implement the input method is shown in Fig. 1. For the purposes of the method, the melody has been defined as the development of the pitch of the tone of whistling over time.

The sound of user's whistling is received by the microphone and digitized by the sound card. The digitized sound is processed in frames that are 1,024 samples long. The melody of whistling is tracked using fast Fourier transform (FFT), employing the FFTW library [7] to perform the FFT computation.

The frequency at which point the energy transfer is the highest is considered the pitch of the tone. The volume level of a frame is determined as a simple sum of the absolute values of the samples of the frame.

A tone is recognized if the sound exceeds the volume level of the user-defined threshold and lasts as long as the volume level is maintained. As no noise cancelling filters have been implemented, the described method is sufficient only for environments with low background noise.

Two different assignments (control modes) of the tonal primitives to the actual movements of the mouse pointer have been defined; namely, the orthogonal and the melodic control mode. Both modes make use of the long tones (used to control the cursor movement) and the short tones (used to emulate the mouse click).

2.1 Orthogonal control mode

In this mode, the mouse pointer may be moved either horizontally or vertically, which is determined by the pitch of the tone at its beginning (the initial pitch). If a tone is started below a specified threshold f_t, the mouse pointer is to move only to the left or to the right. Similarly, if a tone is started above f_t, the pointer will move only up or down. The actual direction and speed of motion at any given time is determined by the difference in the current pitch and the initial pitch. A positive difference (the initial pitch is lower than the current pitch) makes the cursor move up (or to the right). If the difference is negative (the initial pitch is above the current pitch), the cursor moves down (or to the left).

The speed of the cursor at any time is directly dependent on the magnitude of this difference. This allows the user to precisely control the speed according to requirement (e.g., slow down once the cursor is close to the target, etc.) or completely reverse its motion.

As previously noted, the click of the left button is emulated when the user produces only a short tone. Some control tones are shown in Fig. 2.

Figure 3a depicts the state diagram of the orthogonal control mode. The cursor may only be controlled by

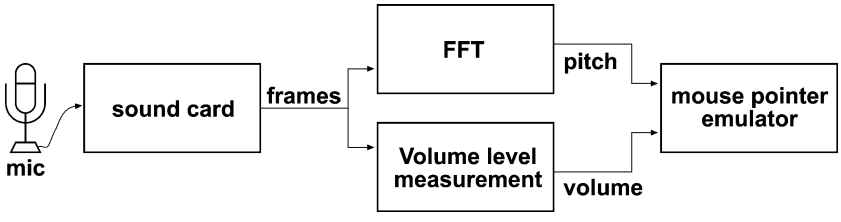

Fig. 1 Block diagram of the system

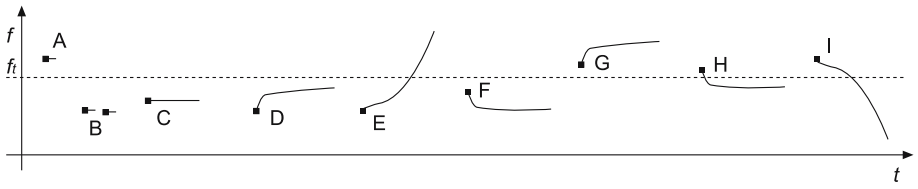

Fig. 2 Orthogonal mode—examples of control tones: *t* time, *f* pitch, f_t threshold pitch; *A* click, *B* double click, *C* no motion, *D* motion to the right, *E* fast motion to the right, *F* motion to the left, *G* motion up, *H* motion down, *I* fast motion down

tones that last more than a threshold length t_t (typically 0.2 s). If the tone does not exceed this t_t, a mouse click is emulated at the position that the cursor exists at that particular point of time.

If the initial pitch f_s of tone is greater than the threshold t_f, the system is locked for the vertical cursor motion, otherwise only the horizontal motion is enabled (see transitions 2–3 and 2–4). As the tone changes its pitch f_c, the cursor velocity is updated appropriately and the cursor is moved accordingly. When the tone is stopped, the system returns to the initial state, waiting for another tone. An example of use of this control mode is shown in Fig. 4.

2.2 Melodic control mode

This mode allows the cursor to move in any direction (not restricting the motion along the directions of the *x*- and *y*-axes only). However, the speed of motion is fixed to a user-defined value: the cursor either moves or it is idle.

The left mouse button click is emulated in the same way as in the orthogonal control. If a longer tone than the threshold t_t is detected (Fig. 3b, transition 2–3), the cursor motion is commenced. The direction of motion is directly dependent on the current pitch of the tone. The pitch of the tone may be any level from the control octave—the interval $<base\ tone,\ base\ tone + 12\ semitones>$. The base tone pitch is selected by the users to reflect their actual range of whistling.

Figure 5 shows an example of the assignment of directions for the C_3 note (approx. 1,050 Hz) chosen the base tone pitch. In this assignment, when the C_3 note is being produced, the cursor moves up, E_3 yields a motion to the right; $G\#_3$ makes the cursor move diagonally down left, and so on. An example of the use of this control mode is shown in Fig. 6.

3 The implementation

The prototype of the system has been realized as a win32 application and requires a Microsoft Windows 98 or newer operating system to run. The system was written in Microsoft Visual C++ (version 6.0) making use of the FFTW library. The application is available for download [17]. A snapshot of the application is provided in Fig. 7.

4 The preliminary user study

To investigate the usability of the system, two separate evaluation sessions were run.

4.1 Method

Four regular computer users, which in this study are defined as people who use computers at least 5h a week,

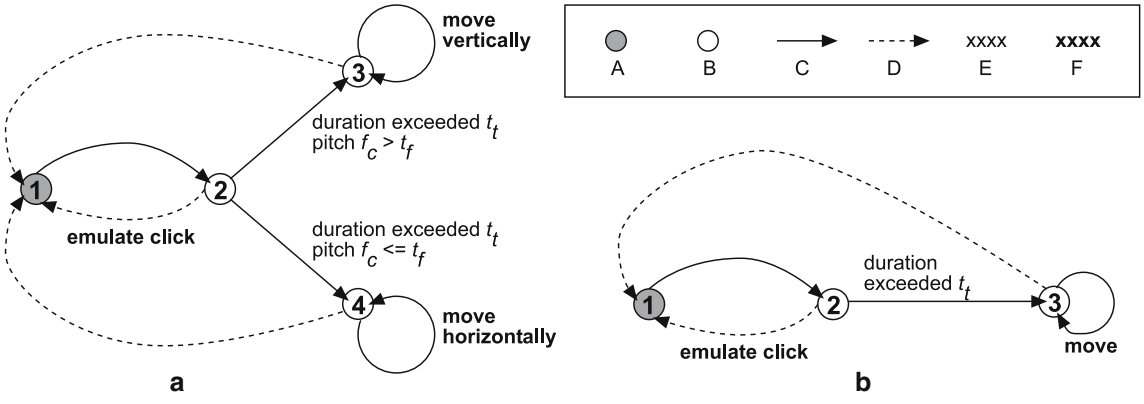

Fig. 3 Function described by means of the state diagrams. a Orthogonal mode. b Melodic mode. Key to the legend: *A* initial state, *B* other state, *C* transition with no sound on input, *D* immediate transition when no sound is being received, *E* additional condition of a transition, *F* action initiated upon a transition

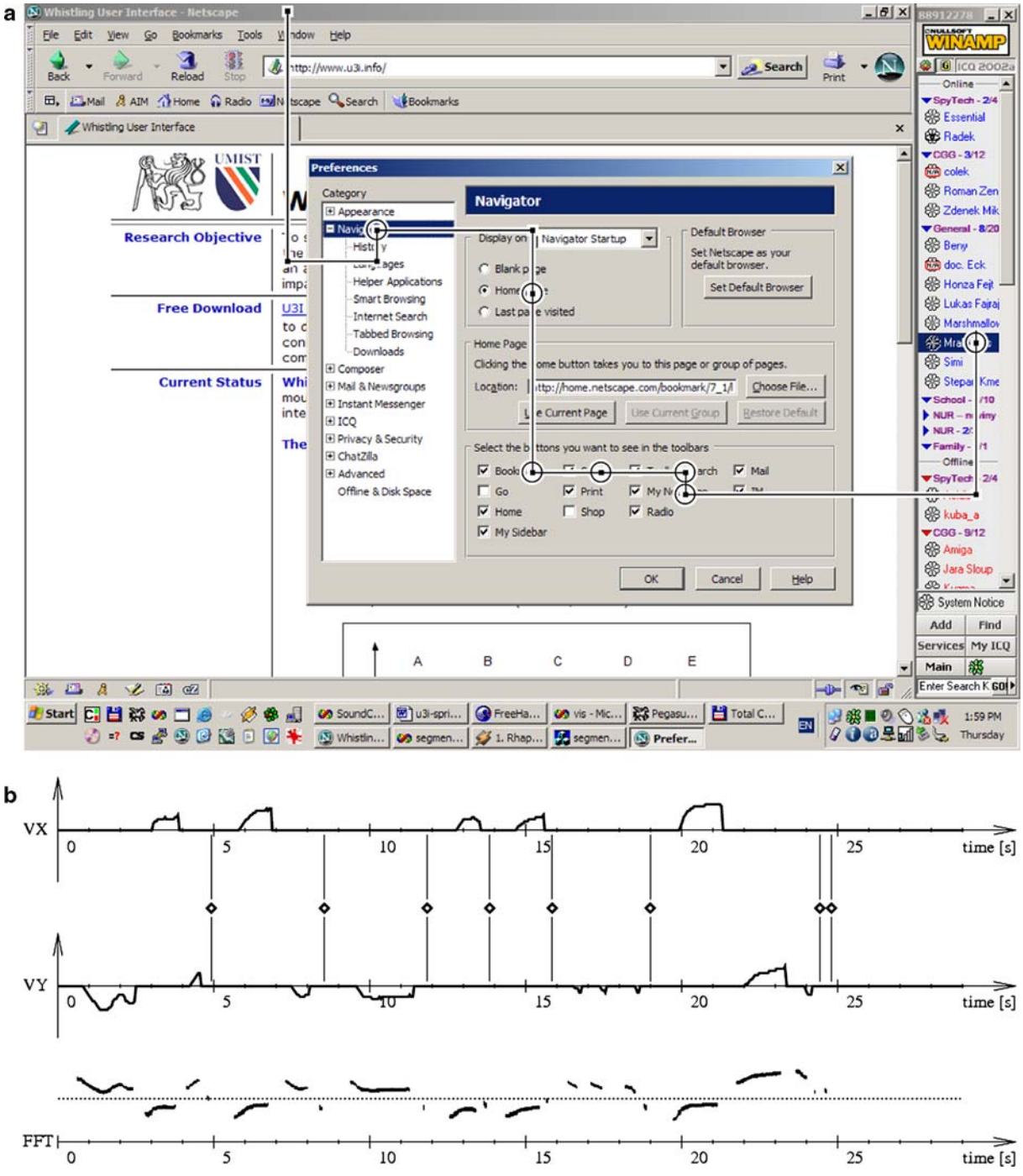

Fig. 4 Example of use of the orthogonal mode. a Trajectory of the mouse pointer. The traces of the individual movements are delimited with *small squares*. The mouse clicks are marked with *circles*. The cursor moved from the left to the right. b *VX*, *VY* relative horizontal and vertical velocity of the cursor, respectively. The clicks are marked with *rhombs*. *FFT* the frequency analysis of the input signal

participated in the study. The demographics data of these users are listed in Table 1.

In the first session, the participants were asked to control the mouse through whistling, while in the second session they were asked to use hissing/humming to control the mouse. All participants either had no visual impairment or wore corrective lenses at the time of the experiment. For both sessions, a Pentium 4 PC (1.7 GHz) running Windows XP with a 17 in. monitor with a resolution of 1,024×768 pixels and a standard blank (blue) background was used. Each participant tested the system in a quiet room to minimize noise interference from the surrounding environment, accompanied only by the experimenter. The participants also wore headsets

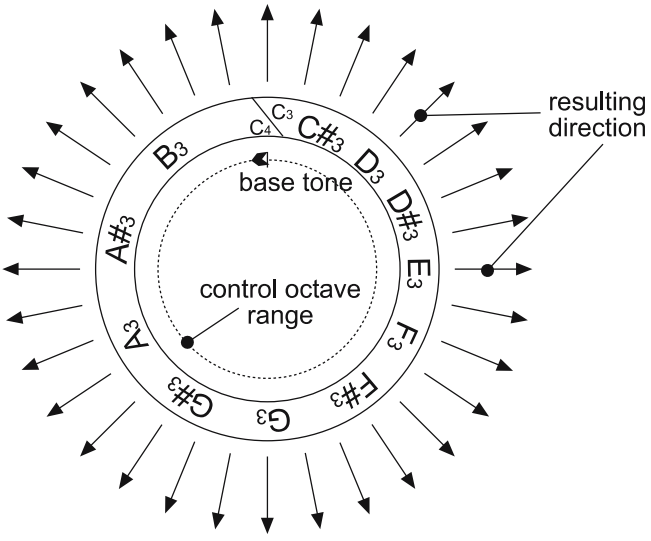

Fig. 5 Melodic mode—the assignment of directions of cursor's motion to different pitches within the control octave

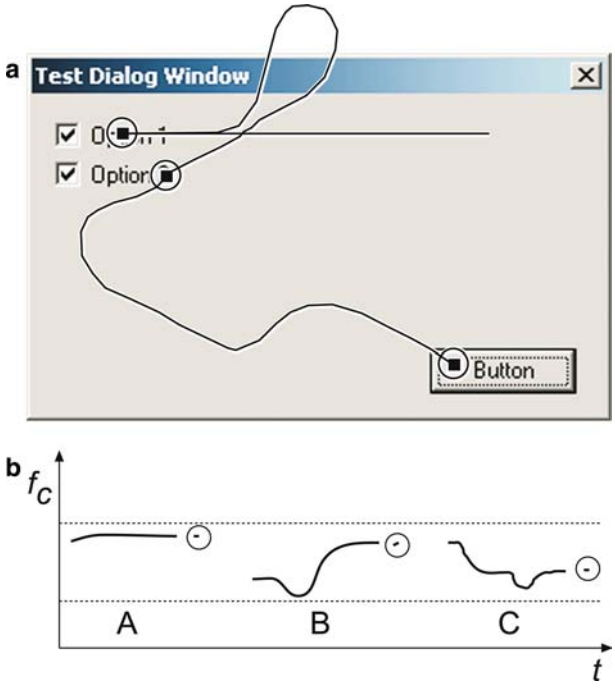

Fig. 6 Example of use of the melodic mode. a Trajectory of the pointer and mouse clicks. b Appropriate control tones. Individual movements are labeled with *letters* and start with *small squares*. Mouse clicks are marked with *circles*. The pitches in the control octave are located between the *dotted lines*

throughout the sessions to further minimize the extraneous noise recorded by the computer. Participants' mouse movements were recorded using Camtasia Studio 2 screen capture software.

4.1.1 The first session

The session started with the filling in of a demographic questionnaire by the participants, either by themselves or with assistance from the experimenter. This was followed by computer-based tasks, and ended with a post-session interview.

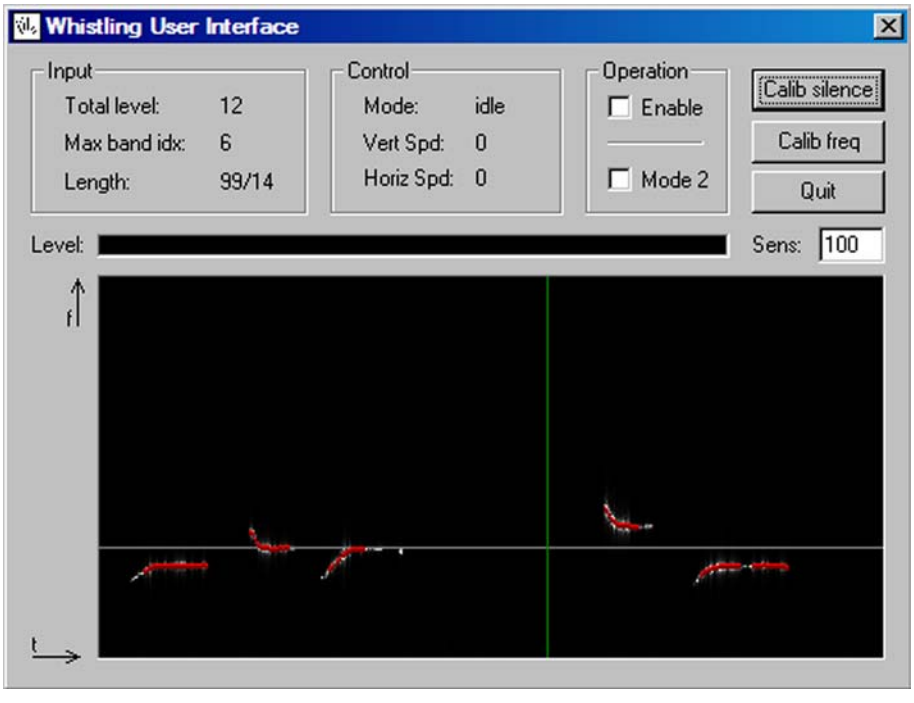

Fig. 7 A snapshot of the U^{3}I prototype application

Table 1 Participants' data

Subjects	S1	S2	S3	S4
Gender	M	M	F	F
Motor disability	Missing fingers (accident)	None	Arthritis (cannot move her finger joints well)	None
Age	40	23	67	19
Average weekly computer use (h)	>20	10–20	5–10	>20

At the beginning of this session the experimenter informed the participants that the purpose of the study was to investigate how easy it was for them to control the mouse pointer through whistling rather than to measure their performance in using the system. The experimenter then demonstrated the two control modes to the participant. The participants were reminded that any noise they made might affect the mouse pointer movement. They were then given 5 min to try the system out before the actual experiment started.

The participants were then asked to move the pointer to various objects on the screen and to click five icons. The participants were allowed to take breaks of any time length between tasks to prevent fatigue from affecting their performance. The tasks varied in the directions and angles of pointer's movement. However, the distances from the current pointer to the next target were kept fairly constant. The icons were standard MS Windows icons displaying folders numbered 1–5 (the number allows the participants to see the sequence of targets to click). Once an icon was clicked, it disappeared, so that only the icons to be clicked remained displayed. The screenshots of the stimuli before and after the first click made are shown in Fig. 8. A picture of a user taking part in the usability study is in Fig. 9.

Two participants (S1 and S4) tested the orthogonal control first and the other two (S2 and S3) tested the melodic control first. This experimental design was aimed at balancing the control mode, gender, age, or disability.

4.1.2 The second session

The same four participants were recruited for the second session 1 month later. Testing the system with the same group of participants allowed a comparison of the ease of use of different types of sound input (whistling vs hissing vs humming). The same setup and equipment was used. However, the stimuli (the locations of the icons) were changed to minimize familiarity, although the 1 month gap between the first and second sessions

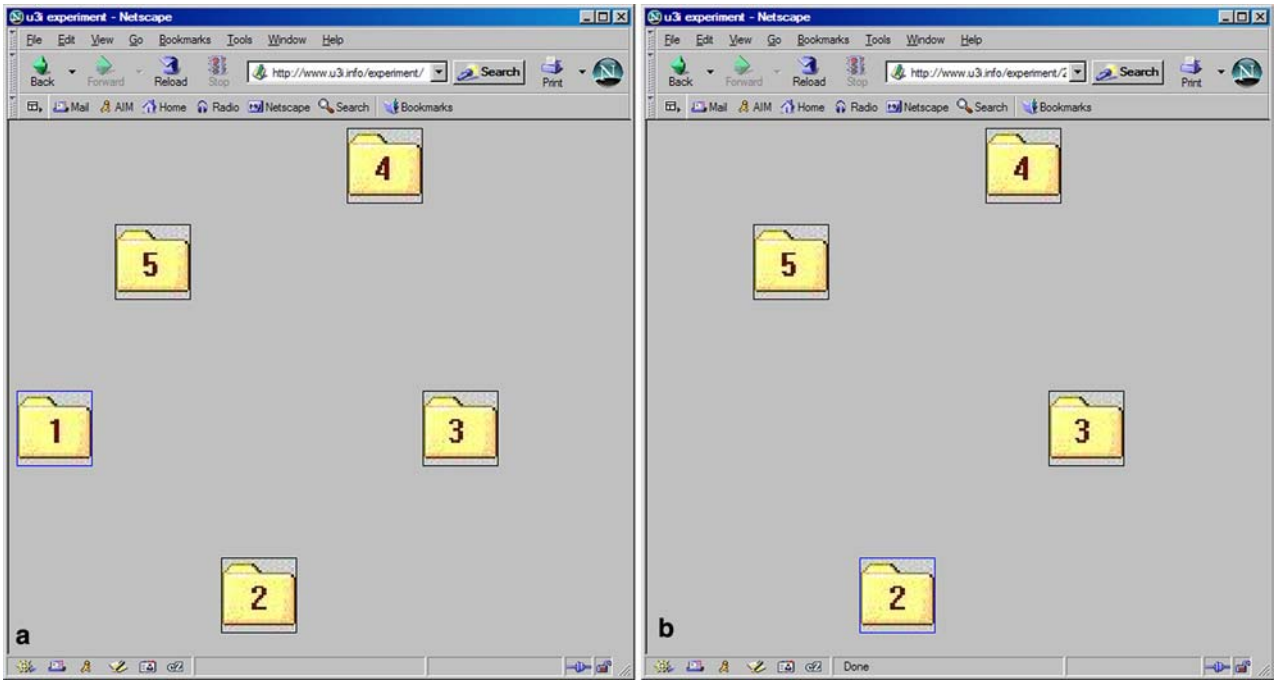

Fig. 8 The stimuli for the user study, the first two of the web pages that the users were asked to sequentially browse using the U^3I

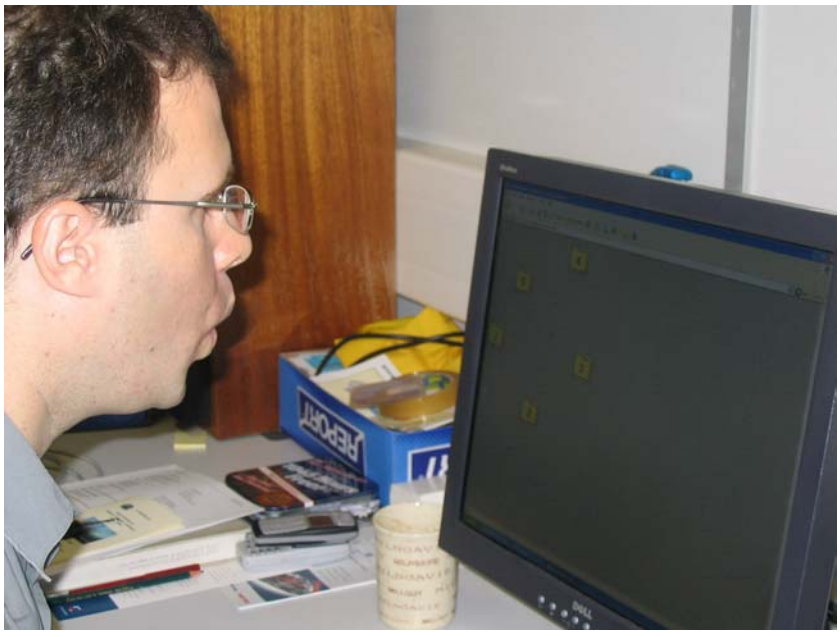

Fig. 9 The user study in progress

might ameliorate this familiarity problem. Because in the first session S1 and S4 tested the orthogonal control first, in this session, S1 and S4 tested the melodic control first, while S2 and S3 tested the orthogonal control for the second session, to fully balance the experimental design.

4.2 Results

Because there were only four participants, a proper statistical analysis could not be performed in this preliminary study.

4.2.1 The first session

The time measures of the first session indicate that, on an average, the participants took twice as long to arrive at a target icon and click it when using the melodic control (an average of 2.6 s, the standard deviation is not reported because there are only four participants) than when using the orthogonal control (1.4 s). When the screen capture was analysed, it showed that all participants overshot the target when using the melodic control.

In the post-session interviews, the participants stated that they felt they could control the pointer much better using the orthogonal control than using the melodic control, in line with the objective performance results (i.e., the time taken to finish the tasks). When asked how they thought these control modes would be useful for them, all participants answered that the orthogonal mode would be useful as an alternative way to control the mouse. Three answered that the melodic mode would be "a fun way to move the mouse on the screen" or "may be good for drawing". One said that she could not think how this mode would be useful for her. All said that they felt comfortable using the system, even though this was the first time they used it and were certain that they would master both control modes, if they were given enough time to learn it properly.

4.2.2 The second session

There were some major problems with operating the system through hissing in the second session. Two participants were unable to hiss properly. The other two could finish the tasks in the orthogonal mode (albeit with a lot of difficulty) through hissing. However, these two were unable to even home in on the first target in the melodic mode.

They were then instructed to repeat the tasks through humming or singing the tones. They were successful in finishing the tasks in both modes. However, they took slightly longer time compared to the time taken to finish the tasks through whistling (1.8 s for the orthogonal mode and 3 s for the melodic mode). The analysed screen capture indicated that, similar to the whistling operation, all participants overshot the target when using the melodic control. Examining the application screen, it was apparent that humming and singing produced signals that were less pure than whistling. Therefore, the movement control was not as refined as the one performed through whistling. The participants still thought that the orthogonal mode was easier to control and operate than the melodic mode. It can be concluded that, in general, the orthogonal mode was

easier to operate than the melodic mode in the context of cursor movement tasks.

In the post-session interview, three participants indicated that they preferred to control the mouse through humming or singing rather than whistling, because whistling was more tiring than humming. The last participant said that he preferred whistling because he felt that this enabled him a better control over the mouse. These data reflect a trade-off between fatigue and better control, i.e., whistling, which provides better control, is more strenuous than humming.

4.3 Discussion

The results of the user study indicated that the orthogonal control was easier to perform than the melodic control in both sessions. This result might be biased by the nature of the task, which is a point-and-click task. It is possible that the melodic mode would have been considered easier if the task involved curvature-drawing or trajectory-based task.

In both control modes, the users were able to complete the tasks. The users reported that they felt comfortable using the system. The participants indicated that humming or singing was less tiring than whistling.. However, from a technical point of view, whistling produces purer sound, and therefore is more precise, especially in melodic mode. The preliminary user study also indicates that the system is not appropriately operable through hissing.

5 Conclusion and future work

This paper reports on the design and implementation of a whistle-operated pointing device. The key benefits of this system include: low computation power needed (especially suitable for mobile devices), short learning curve (as indicated from the user study), easy installation, and no special device required.

The preliminary user study indicated that the system is usable in both melodic and orthogonal modes and that humming was the preferred mode of input. However, the users favoured the orthogonal mode, as they found it more intuitive and comfortable.

Currently, the system is a working prototype. However, since no noise detection and filtering routines were implemented, the system is very sensitive to acoustic interferences. In order to be able to use the system in everyday situations, it should be built in a more robust manner.

Immediate goals are to extend the interaction methods described above so that they enable the emulation of both mouse buttons and the drag-and-drop operations, and to investigate the possibilities of integrating the non-verbal sound input and speech recognition techniques. In such a setup, speech recognition may be used to issue discrete commands or type text while our method may be used to control the mouse pointer.

The user study only involved four participants. A larger scale study would allow statistical analysis of the results. A study that asked more thorough questions on user acceptance would also be helpful to understand the factors involved in transforming a system from being useful and usable into a system that is acceptable.

References

1. Baier G, Herman T (2004) The sonification of rhythms in human electro-encephalorgam. In: Barrass S, Vickers P (eds) Proceedings the 10th International Conference on Auditory Display, Sydney (ICAD), 2004, pp 1–5
2. Barra M, Cillo T, De Santis A, Umberto FP, Negro A, Scarano V, Matlock T, Maglio PP (2001) Personal webmelody: customized sonification of web servers. In: Hiipakka J, Zacharov N, Takala T (eds) Proceedings of the 7th International Conference on Auditory Display. Laboratory of Acoustics and Audio Signal Processing and the Telecommunications Software and Multimedia Laboratory, Helsinki University of Technology, Espoo, pp 1–9
3. Basson S (2002) Speech recognition and accessible education. Speech Technol Mag 7(4) [on-line]. http://www.speechtechmag.com/issues/7_4/avios/
4. Blattner MM, Sumikawa DA, Greenberg RM (1989) Earcons and icons: their structure and common design principles. Hum–Comput Interact 4:11–44 (Lawrence Erlbaum, Hillsdale, NJ)
5. van Buskirk R, LaLomia M (1995) The just noticeable difference of speech recognition accuracy. Proceedings of ACM CHI'95, Conference on Human Factors in Computing Systems, vol 2. ACM Press, New York, p 96
6. Franklin KM, Roberts JC (2003) Pie chart sonification. Proceedings of the Seventh International Conference on Information Visualization, IEEE, London, pp 4–9
7. Frigo M, Johnson SG (2005) The design and implementation of FFTW3. Proceedings of the IEEE. Special Issue on Program Generation, Optimization, and Platform Adaptation, vol 93, pp 216–231
8. Gaver WW (1993) Sythesizing auditory icons. ACM INTERCHI'93 Conference on Human Factors in Computing Systems. ACM Press, New York, pp 228–235
9. Hämäläinen P, Mäki T, Pulkki V, Airas M (2004) Musical computer games played by singing. In: Evangelista G, Testa I (eds) Proceedings of the Seventh International Conference on Digital Audio Effects, Naples
10. Igarashi T, Hughes JF (2001) Voice as Sound: using non-verbal voice input for interactive control. In: Proceedings of UIST 2001. ACM Press, Orlando, FL, pp 155–156
11. Kitto KL (1993) Development of a low-cost sip and puff mouse. In: Proceedings of 16th Annual Conference of RESNA. RESNA Press, Las Vegas, pp 452–454
12. Nicol C, Brewster SA, Gray PD (2004) A system for manipulating auditory interfaces using timbre spaces. In: Jacob R, Limbourg Q, Vanderdonckt J (eds) Proceedings of CADUI. ACM Press, Madeira, pp 366–379
13. Nicol C, Brewster S, Gray P (2004) Designing sound Towards a system for designing audio interfaces using timbre spaces. In: Barrass S, Vickers P (eds) Proceedings of the 10th International Conference on Auditory Display, Sydney 2004, pp 1–5
14. Pirhonen A, Brewster S, Holguin C (2002) Gestural and audio metaphors as a means of control for mobile devices. Proceedings of the CHI 2002 Conference on Human Factors in Computing Systems. ACM Press, New York, pp 291–298
15. Rabiner L, Juang BH (1993) Fundamentals of speech recognition. Prentice Hall, Englewood Cliffs, NJ (ISBN 0130151572)

16. Sibert LE, Jacob RJK (2000) Evaluation of eye gaze interaction. Proceedings of CHI 2000 Conference on Human Factors in Computing Systems. ACM Press, The Hague, pp 281–288
17. U3I Project homepage (2005) [On-line] http://www.u3i.info
18. Walker A, Brewster SA (2000) Spatial audio in small display screen devices. Pers Technol 4:144–154