

# An Evaluation of Input Devices for Use in the ISEE Human-Synthesizer Interface

Roel Vertegaal  
Department of Ergonomics  
University of Twente  
P.O. Box 217  
7500 AE, Enschede  
The Netherlands

Barry Eaglestone  
Department of Computing  
University of Bradford  
Bradford, BD7 1DP  
United Kingdom

Michael Clarke  
Department of Music  
University of Huddersfield  
Huddersfield, HD1 3DH  
United Kingdom

R.Vertegaal@bsk.utwente.nl

B.Eaglestone@comp.brad.ac.uk

AL0008P@pl.hud.ac.uk

## Abstract

The research described investigates the impact of input devices on human performance in four-dimensional timbre space navigation. The experiments used ISEE, a high-level synthesizer-independent user interface. Four different device types (mouse, relative joystick, absolute joystick and dataglove) were used to reach target positions in perceptual space using audio-visual feedback conditions. Data was analysed for speed and accuracy of the devices. Results indicate a highly significant effect of the choice of input device on the efficacy of timbre manipulation, and therefore have significant implications for the design of intuitive interfaces for direct manipulation of sounds in composition and performance.

## 1 Introduction

Musician-synthesizer interaction is problematic since user interfaces must resolve conflicting requirements: simple, direct and intuitive real-time control of sounds, and constructive control of inherently complex synthesis technology. Our research explores whether this conflict can be resolved using a user interface based on the perceptual parameters of timbre space, in conjunction with appropriate input devices. Timbre space provides a general abstraction of musical instrument control by modelling the perceptual degrees of freedom for sound modification. In this way a timbre space user interface can provide independence from synthesizer characteristics, thus enabling users to manipulate sound purely within the perceptual domain. The practicality of this approach is demonstrated by ISEE (described in § 2.2), a four-dimensional synthesis user interface based on the timbre space paradigm. In ISEE, timbre space dimensions are based upon expert sound design practice, and the complexity of sound synthesis is managed through hierarchical decomposition of the timbre space. However, musical validity of this form of interface depends on a means for direct, intuitive, real-time manipulation of sounds. We have therefore evaluated the use of generic low- and multidimensional input devices for ISEE. In our experiments, four different input devices (mouse, relative joystick, absolute joystick, dataglove) were used to reach target positions in perceptual space using audio-visual feedback. Data was analysed for speed and accuracy. The results indicate that the choice of input device has a significant impact on system usability. The mouse was the fastest and most accurate device, followed by the absolute joystick, relative joystick and Power Glove.

## 2 Background

### 2.1 The Control Problem

With the advent of graphical user interfaces in sound synthesis systems, one would expect the notion of direct manipulation of timbre to have gained ground. According

to Nelson [1980], direct manipulation is a user interface technique where objects and actions are represented by a model of reality. Physical action is used to manipulate the objects of interest, which in turn give feedback about the effect of the manipulation. An important aspect of direct manipulation is the principle of transparency, where attention shifts from issuing commands to observing results conveyed by feedback. This requires feedback to be consistent with the user's expectations of the task's results. Shneiderman [1987] argues that with direct manipulation systems, there may be substantial task-related semantic knowledge, but users need to acquire only a modest amount of computer-related semantic and syntactic knowledge. Task-related semantics should dominate the users' concerns, reducing the distraction of dealing with the computer semantics and syntax. Current synthesis user interfaces are based on the direct use of synthesis model parameters, which need not necessarily behave in a perceptually linear or consistent fashion. For example, to change the brightness of an FM synthesized tone, one could change the output level of a modulator. Though most of the time this seems to affect the brightness of the sound, this method could result in noise due to aliasing when modulator feedback is active, resulting in a loss of correspondence between the task-related semantics and synthesizer-related semantics. A more direct mapping between task-related semantics (I want to make a sound brighter) and synthesizer-related semantics (then I need to change the output level of the modulator or the feedback level or both) could easily be achieved if control would operate at a higher level of abstraction. Achieving true direct manipulation of timbre is a step to be taken before we can test generic input devices, since it helps operating those devices in a more meaningful way, possibly improving their performance. A consistent task-related, low- to high-dimensional mapping between control information and synthesis information, ideally based on human perception and cognition of timbres is a first step in this direction.

Studies into timbre control systems have traditionally focused on performance rather than generic sound

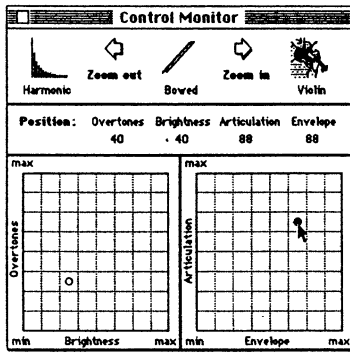


Fig. 1: The ISEE controller tool

synthesis. They are often based on the use of innovative hardware controllers [Bauer and Foss, 1992; Cadoz et al., 1984; Gibet and Marteau, 1990; Waisvisz, 1985] (see [Vertegaal, 1994] for a review). However, the use of these as generic controllers is limited, because researchers fail to develop accompanying formalisms for the low- to high-dimensional mapping of control data to synthesis data. Also, most of these systems are intended to be idiosyncratic for artistic reasons and their usability is hardly ever empirically evaluated. Fortunately, some research into generic control formalisms has been done, and can form the basis for further evaluation and development of synthesizer control mechanisms and techniques.

In [Buxton et al., 1982], the *Objed* system is described as a part of the SSSP, a computer composition environment that was one of the first to introduce direct manipulation principals in digital sound synthesis. Subsequent graphical MIDI editors were all based on the same principle: that of manipulating sliders to control on-screen synthesis model parameters. However, early on the authors recognized that approach to be no more than a substitute, and that timbre should ideally be controlled according to perceptual rather than acoustical attributes. They also emphasized the importance of minimizing non-musical problems of the sound synthesis task and permitting the composer to understand the perceptual consequences of their actions.

In [Lee et al., 1991; Lee and Wessel, 1992] it is demonstrated how a Mattel Power Glove was used in combination with a neural network to produce real-time control of timbre during performances. As a control mapping a timbre space was used in which a limited number of sounds were organised in a geometrical model according to their perceived timbre differences [Wessel, 1985]. This approach elegantly features all constraints for achieving a direct manipulation of timbre, including a well-based formalism for the real-time mapping of low-dimensional perceptual parameters to high-dimensional synthesis model parameters. However, Plomp [1976] indicates that when constructing timbre spaces, the number of timbre space dimensions increases with the variance in the assessed timbres. This makes it difficult to derive a generalized synthesis model from this strategy. When trying to reduce the number of dimensions artificially by using several less varied timbre spaces the dimensions of the different timbre spaces might not correlate, which could cause usability problems if used as synthesis parameters. Generic use of timbre space is also inhibited by the need to use existing sound examples judged by a human panel. How could a musician construct his own timbre spaces? What if he wants to generate totally new sounds?

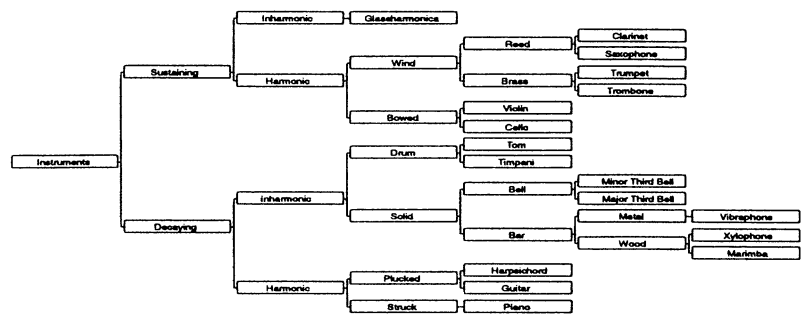


Fig. 2: A partial taxonomy of instrument spaces

## 2.2 The ISEE Timbre Space Model

The practicality of using timbre space as a basis for a sound design system is demonstrated by the Intuitive Sound Editing Environment (ISEE) [Vertegaal and Bonis, 1994]. ISEE is a synthesizer and synthesis model independent user interface designed for sound synthesis applications in both composition and performance. The following description of ISEE omits technical details—those interested should refer to [Vertegaal and Bonis, 1994].

ISEE attempts to generalize the timbre space paradigm for generic user interface purposes by concentrating on the defining dimensions of timbre space. Assuming these parameters have orthogonal properties, every point in space can be defined by combining synthesis data associated with the projections of its coordinates. ISEE features a carefully selected set of perceptual timbre parameters, identified through qualitative observation of working methods employed during expert sound synthesis practice [Vertegaal, 1992]. Four timbre parameters are presented to the user by two two-dimensional projections of the four-dimensional timbre space they constitute (see figure 1). The first two of these parameters relate to the spectral envelope and the last two to the temporal envelope: the *Overtones* parameter controls the basic harmonic content; the *Brightness* parameter controls the spectral energy distribution; the *Articulation* parameter controls the spectral transient behaviour as well as the persistent noise behaviour; and the *Envelope* parameter controls temporal envelope speed. The first three parameters are similar to those identified by Grey [1975]. The high level of abstraction of the ISEE timbre parameters can be utilized to avoid the proliferation of timbre space dimensions [Plomp, 1976], consistently keeping the number parameters presented to the user to a minimum. This problem is one of defining functions which map from timbre space dimensions to synthesis parameters, and is simplified by decomposition of the timbre space into a hierarchy of spaces, each of which defines sub-classes of related timbres (see figure 2). Separate mapping functions are then defined for each sub-class. The actual implementation of the ISEE timbre parameters thus depends on the required refinement of synthesis control. ISEE refers to a scaled implementation of the four timbre parameters as an *instrument space*, because as well as allowing control of the timbre, it also defines the range and type of pitch and loudness behaviour of the instrument(s) it encloses.

Generally, each component instrument space is organised using the following heuristics: from low to high, from harmonic to inharmonic, from mellow to harsh, and from fast to slow. The instrument space

hierarchy (see figure 2) is based upon a categorisation scheme derived from expert analysis of existing instruments using think-aloud protocols, card sorting and interview techniques. Using the hierarchy, the user can structure a search by *zooming in* on specific instrument spaces from grosser higher level spaces, thus selecting constraints for each of the four timbre parameters. Alternatively, when interested in a broader perspective of instruments, the user can jump to a broader instrument space by *zooming out*. More expert users can also make use of a traditional hierarchy browser, for example, when constructing new instrument spaces. Instrument spaces in the hierarchy can also be selected using program changes following the General MIDI specification. The Violin instrument space is a good example of a refined application of ISEE timbre parameters. In this space, the Overtones parameter describes the relation of the bow to the bridge, from *flautando* to *sul ponticello*. The Brightness parameter relates to the bow pressure on the string, the Articulation parameter controls the harshness of the inharmonic transient components (the force with which the bow is dropped on the strings) and the Envelope parameter controls the duration of the attack.

There are two main factors which affect the usability of the ISEE user interface: the visual representation of the space, and the means by which users specify locations within it. The following section describes our experimental research toward establishing the significance of the second of these factors.

### 3 Materials and Methods

We selected three input devices to empirically establish their impact on performance in a four degree of freedom (DOF) instrument space navigation task: the Apple Standard Mouse (a relative input device); the Gravis Advanced MouseStick II—an optical joystick (absolute or relative); the Nintendo Power Glove (absolute and relative). Our sample population consisted of music students from the Department of Music of the University of Huddersfield, England, with experience in the use of electronic instruments and synthesized sounds, but with little experience in sound synthesis. A repeated measures design [Coolican, 1990] was used with a group of 15 paid subjects who were asked to reach for target positions in the Sustaining instrument space using the various device types. The Sustaining space contained a broad selection of sustaining musical instruments generated in real-time by simple FM synthesis on a YAMAHA SY99. The Overtones parameter was used to control the harmonicity of the spectrum using FM frequency ratios ( $c:m = 1:1, 2:1, 3:1, 4:1, 5:1, 1:2, 1:4, 1:3, 1:5, 4:5, 6:5, 1:9, 1:11, 1:14, 2:3, 3:4, 2:5, 2:7, 2:9$ ; see [Truax, 1977] for a more detailed explanation). The Brightness parameter was used to control the cut-off frequency of the low-pass filter. The Articulation parameter controlled the ratio of the higher partials' attack rate to the lower partials' attack rate. The Envelope parameter controlled the duration of the attack. These mappings were designed by an expert to approach as consistent a mapping as possible with simple FM. An Apple Macintosh SE was used to filter the erratic Power Glove information and record all movement during the experiments. An Apple Macintosh LC was used to run the ISEE system. All systems were interconnected by MIDI.

Four interfaces were constructed. In the first, the mouse was used to change the coordinate indicators in the Control Monitor (see figure 1) by clicking and dragging

the indicator dots. The joystick was used in the second and third interfaces. In the second the joystick provided absolute control—the position of the stick corresponded directly to the position of the indicators in the Control Monitor. In the third, the joystick provided relative control—the position of the stick controlled the speed and direction of the Control Monitor indicators. In both, the two buttons on the top of the stick were used to select the coordinate system to be controlled with the stick. The fourth interface used the Power Glove for four-dimensional positioning in the Control Monitor. Motion on the Y-axis controlled Overtones, the X-axis controlled Brightness, the Z-axis controlled Articulation, and roll information was used to control the Envelope parameter in a relative fashion. Holding the wrist level would produce no change, rolling the wrist anti-clockwise would decrease the Envelope parameter and rolling clockwise would increase the Envelope parameter. The glove was engaged by clutching and inactive when not clutching.

The subjects were given five minutes to get used to each device, except for the Power Glove, with which they were allowed to practice for 15 minutes because of the special technique involved. Each subject was given 10 test blocks of four experiments, one for each of the four device types. To prevent order effects, the order of the 4 types of input devices in each test block as well as the order of the test blocks was randomised. A questionnaire was answered by each subject after the experiments.

In each experiment the subject was required to listen to a timbre and locate it in the instrument space, using one of the four interfaces. At the start of each experiment, the location of the target could be seen in the Control Monitor window while the target sound was played five times. After this, the Control Monitor indicators would centre, with the sound changing accordingly, giving the subjects an audio-visual cue to start manipulating the indicators with the input device. The target position remained visible throughout the experiment in a separate window similar to that of Control Monitor. The stimulus tones (with a 1.5 sec. duration and a C3 pitch) were repeated throughout the experiment to give the subjects sufficient auditory feedback on the position of the Control Monitor indicators. When the match was considered good enough, the subjects released the input device. All movement during the experiments was recorded. This enabled us to simulate retroactively an experiment where the subject would have been required to reach a certain accuracy criterion, which would then automatically terminate the trial.

The efficacy of each device was established by measuring the movement time needed to reach the four-dimensional target position within a certain accuracy (where accuracy is overall Euclidean distance to target in four-dimensional space). This combines speed and accuracy into a single measure and removes the effect of individual subjects' subjective accuracy criteria for terminating trials. Since a subject might briefly, inadvertently pass through a point that lies within the required accuracy, retroactive analysis allows us to correct this by measuring the time until the subject passed the criterion for the last time during the trial. The data was recorded at millisecond accuracy using a sequencer. The accuracy criterion was set to 1.13 cm in 4-D Euclidean distance to target, which was the 75th percentile of the final accuracies achieved over all trials in this experiment by the least accurate device, the Power Glove. The choice of the 75th percentile is not critical; analysis with other criteria gave similar results.

## 4 Results

Analysis of variance showed that the choice of input device had a highly significant effect on performance ( $F(3, 483) = 68.99, p < 0.001$ ). This indicated that differences in performance were related to the choice of input device and not just due to differences between subjects. Table 1 shows the mean movement time and mean overall accuracy for each device.

Device:	Mouse	Abs. Joystick	Rel. Joystick	Power Glove
MT:	4,917	7,139	10,308	24,950
Accuracy:	1.35	1.61	2.04	8.97

Table 1: Mean movement time (msec) and accuracy (mm)

All differences were highly significant. The mouse was 1.5 times faster than the absolute joystick (paired two-tailed t-test;  $p < 0.0001$ ), 2.1 times faster than the relative joystick ( $p < 0.0001$ ) and 5.1 times faster than the Power Glove ( $p < 0.0001$ ). The absolute joystick was 1.4 times faster than the relative joystick ( $p < 0.0001$ ) and 3.5 times faster than the Power Glove ( $p < 0.0001$ ). The relative joystick was 2.4 times faster than the Power Glove ( $p < 0.0001$ ).

The mean accuracy is the smallest 4-D Euclidean distance to target reached during the experiment (Table 1). The difference between the mouse and the absolute joystick was not significant ( $p > 0.15$ ). The mouse was 1.5 times more accurate than the relative joystick ( $p < 0.005$ ) and 6.7 times more accurate than the Power Glove ( $p < 0.0001$ ). The absolute joystick was 1.3 times more accurate than the relative joystick ( $p < 0.05$ ) and 5.6 times more accurate than the Power Glove ( $p < 0.0001$ ). The relative joystick was 4.4 times more accurate than the Power Glove ( $p < 0.0001$ ). Movement pattern analysis indicated *Overtones* and *Brightness* to be perceptually closer related than any other combination of the 4 timbre parameters (see [Vertegaal, 1994] for a detailed review).

## 5 Conclusions

It is clear from these findings that the Power Glove was not very effective in this four-dimensional task. The subjects found it physically tiring and very hard to control. The bad performance of the glove can be attributed to the lag that occurred because of the filtering and the insufficient resolution of the device beyond 3 degrees of freedom. During regular sound synthesis applications a mouse will suffice. When using a keyboard, the absolute joystick is the most likely option, since it can easily be placed on top of the control panel of the keyboard. The relative joystick is useful when rapidly switching between the two Control Monitor parameters sets and where subtle changes in timbre need to be made.

The *Overtones* and *Brightness* parameters were considered to be the most intuitive and useful parameters. Subjects were able to integrate the use of these two parameters effectively. The 2 x 2-D visual representation in Control Monitor corresponds nicely with the control structure of the best performing devices. Further evaluation of the presentation interface is however necessary. The subjects considered ISEE a useful tool which liberated them from technicalities without restricting their artistic freedom.

## Acknowledgements

We would like to thank Apple Computer Inc., and in particular, S. Joy Mountford of the ATG Design Center for supporting the above research. We would also like to thank Kurt Schmucker, Tamas Ungvary, Ernst Bonis and Tom Wesley for their valuable support.

## References

- [Bauer and Foss, 1992] Will Bauer and Bruce Foss. GAMS: An Integrated Media Controller System. *Computer Music Journal*. 16 (1) pp. 19-24, 1992.
- [Buxton et al., 1982] W. Buxton, S. Patel, W. Reeves, et al. Objed and the Design of Timbral Resources. *Computer Music Journal*. 6 (2), 1982.
- [Cadoz et al., 1984] C. Cadoz, A. Luciani and J. Florence. Responsive Input Devices and Sound Synthesis by Simulation of Instrumental Mechanisms: The Cordis System. *Computer Music Journal*. 8 (3), 1984.
- [Coolican, 1990] Hugh Coolican. *Research Methods and Statistics in Psychology*. Hodder & Stoughton, 1990.
- [Gibet and Marteau, 1990] Sylvie Gibet and Pierre-Francois Marteau. Gestural Control of Sound Synthesis. *Proceedings of the 1990 ICMC*, Glasgow, International Computer Music Association. pp. 387-391, 1990.
- [Grey, 1975] John M. Grey. *An Exploration of Musical Timbre*. Ph.D. Dissertation. Dept. of Psychology, Stanford University, 1975.
- [Lee et al., 1991] Michael Lee, Adrian Freed and David Wessel. Real-Time Neural Network Processing of Gestural and Acoustical Signals. *Proceedings of the 1991 ICMC*, Montreal, International Computer Music Association. pp. 277-280, 1991.
- [Lee and Wessel, 1992] Michael Lee and David Wessel. Connectionist Models for Real-Time Control of Synthesis and Compositional Algorithms. *Proceedings of the 1992 ICMC*, San Jose, International Computer Music Association, 1992.
- [Nelson, 1980] Ted Nelson. Interactive systems and the design of virtuality. *Creative Computing*. 6 (11-12), 1980.
- [Plomp, 1976] Reinier Plomp. *Aspects of Tone Sensation*. Academic Press, 1976.
- [Shneiderman, 1987] Ben Shneiderman. *Designing the User-Interface: Strategies for Effective Human-Computer Interaction*. Addison Wesley, 1987.
- [Truax, 1977] Barry Truax. Organizational Techniques for c:m Ratios in Frequency Modulation. *Computer Music Journal*. 1 (4) pp. 39-45, 1977.
- [Vertegaal, 1992] Roel Vertegaal. *Music Technology Dissertation*. Utrecht School of the Arts, The Netherlands, 1992.
- [Vertegaal, 1994] Roel Vertegaal. *An Evaluation of Input Devices for Timbre Space Navigation*. MPhil Dissertation. University of Bradford, UK, 1994.
- [Vertegaal and Bonis, 1994] Roel Vertegaal and Ernst Bonis. ISEE: An Intuitive Sound Editing Environment. *Computer Music Journal (in press)*, 1994.
- [Waisvisz, 1985] M. Waisvisz. THE HANDS: A Set of Remote MIDI-controllers. *Proceedings of the 1985 ICMC*, Burnaby, Canada, International Computer Music Association, 1985.
- [Wessel, 1985] David Wessel. Timbre Space as a Musical Control Structure. In: Roads, C. and J. Strawn (Ed.), *Foundations of Computer Music*. MIT Press, pp. 640-657, 1985.