Looking for Sound? Selling Perceptual Space in Hierarchically Nested Boxes

Roel Vertegaal
Department of Ergonomics
University of Twente
The Netherlands
roel@acm.org

Barry Eaglestone
Department of Computing
University of Bradford
United Kingdom
B.Eaglestone@scm.brad.ac.uk

ABSTRACT
ISEE3D is a 3D musical sound browser which uses nested boxes to allow browsing through a hierarchy of perceptually organized sound spaces.

Keywords
multimedia, database, musical sound, 3D, dataglove.

INTRODUCTION
The content-based retrieval of musical audio data is a challenging problem of human-computer interaction [7]. In spite of the importance of musical audio in the entertainment industries, this problem has received relatively little attention. Musical audio is often stored as a collection of bytes with little in terms of task-level description attached. To allow sound designers, musicians and composers alike to search their collections of musical sounds they would need:

1) Some form of task-level description and organization of their audio objects.
2) A way of representing such descriptions.

In this paper, we present our approach to the second problem: ISEE3D, a visualization tool for structured browsing of musical sounds based on the Intuitive Sound Editing Environment (ISEE) [5]. In order to reduce the complexity of the first problem, ISEE confines its users to browsing sounds made by real or virtual musical instruments, as experienced by musicians using MIDI synthesizers and samplers. The novelty of ISEE3D lies in the integration of hierarchical organization and sound object description into a single 3D visuo-spatial metaphor. This is made possible by two paradigms:

1) The use of a limited set of perceptual parameters that describe both the sound contents of database objects as well as the relations between them;
2) The use of level of refinement (scale) for organizing the application of these perceptual parameters within the objects.

In ISEE3D, instrument objects are represented by 3D boxes. The space enclosed by these boxes is made up by three parameters that describe the sonic possibilities of the instrument in high-level perceptual terms (see fig. 2). By moving through the perceptual space inside the box, specific sounds can be located and played. Smaller boxes mark the position of other, more refined instrument objects that provide higher definition of that particular area of perceptual space. One can enter these boxes by pointing at them using a dataglove. This way, a perceptually organized hierarchy of nested perceptual spaces can be traversed using 3D positioning and some simple pointing gestures. This metaphor is similar to the *n*-Vision approach [1] for organizing n-dimensional virtual worlds. We will first discuss our perceptual organization scheme and problems with previous work. We will then provide an example of the system at work.

PREVIOUS WORK
When representing instrumental sounds, perhaps the most important issue is how to organize them using the right semantics. We devised a perceptual classification scheme based on expert analysis of instrumental sound [5]. Our first classification parameter is *Envelope*, determining whether the sound sustains or decays. The next parameters are the *Harmonicity* and *Brightness* of the sound. The final classification parameter is the transient behaviour (*Articulation*), describing the synchronicity of overtones at the start of a sound. The last three parameters are similar to the three semantical parameters of timbre discrimination identified by Grey [3]. In ISEE, the four high-level parameters are used to define instrument objects at different levels of refinement , as well as to organize these objects as nodes and subnodes of a hierarchical tree (see [6]). In this tree, the root object contains a broad range of crude instrumental characteristics (such as decaying vs. sustaining sounds). Positions in its parameter space are marked as hyperlinks to more refined instrument objects. These subnodes provide a higher definition of that portion of parameter space (e.g., sustaining instruments only). By continued selection of more and more refined subnodes, searches are structured until a leaf object is reached containing a high-definition description of a specific instrument (e.g., a saxophone).

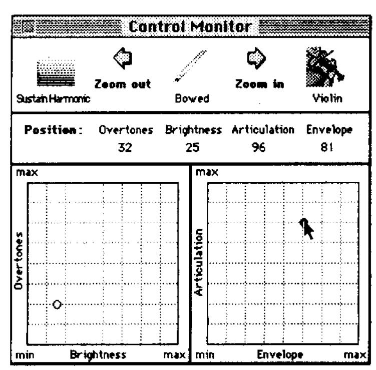
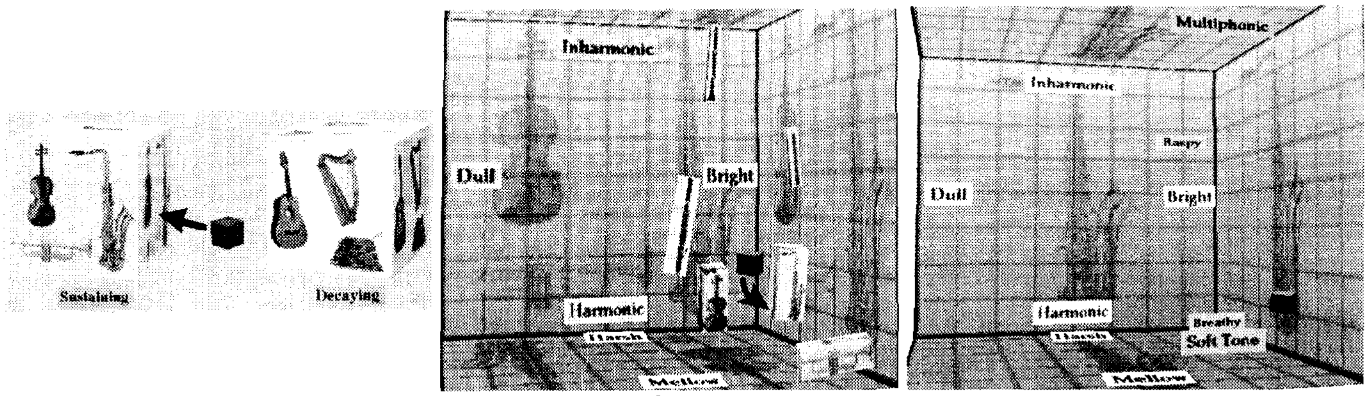


Figure 1. The 2D ISEE Browser.



a b c

Figure 2. ISEE3D Browser: a) selecting Sustaining sounds; b) locating Saxophone object; c) entering Saxophone object.

INTEGRALITY OF REPRESENTATION

In the first 2D version of ISEE, classification and four-dimensional sound specification were separated into two viewers: a hierarchy tree browser for random access of instrument objects, and an object browser in which the four timbre dimensions of an object could be explored by moving within two 2D projections of parameter space (fig. 1). By pressing a zoom button, users could jump from the current object to the closest subnode in the tree. During user trials with different input devices (see [6]), we identified some potential problems with this type of representation:

1) The control structure of the 2D projections corresponds well with that of 2D input devices. However, when a dataglove was used, movement showed high integration in three dimensions. This suggested at least 3 of the 4 timbre parameters to be perceived as integrally related to one another [2]. We wanted to reflect this in the representation.
2) Although subnodes are represented in the hierarchy tree browser, only the closest subnode is represented in the object browser (the Violin icon in fig. 1). Consequently, there are two potentially conflicting conceptual models for navigation.

THE ISEE3D BROWSER

Fig. 2 shows how the above issues were solved in the ISEE3D browser. Instrument objects are represented as 3D boxes (fig 2a). These boxes contain sound, and smaller boxes (fig 2b). A red cube functions as a sound picker. A query starts by dragging the picker with a dataglove towards one of two boxes containing sustaining or decaying instruments (fig 2a). After making a pointing gesture, the box closest to the picker is opened, in this case the Sustaining box (fig 2b). The fabric of space inside this box is made up of instrumental sound. When moving the picker through this space, the sound of a synthesizer is modified according to the three high-level parameters: *Harmonicity* (y-axis), *Brightness* (x-axis) and *Articulation* (z-axis) (*Envelope* need not be represented). Subnodes in the hierarchy are directly represented at positions inside this space. When the picker is moved towards the Saxophone object, the sound will change into that of a bright, relatively harmonic reed instrument. By making a pointing gesture, the Saxophone box is entered (fig 2c). The sound quality improves, and the dimensions of space are now specifically targeted at the Saxophone, allowing bluesy, multiphonic play high up on the y-axis, and soft, breathy play in the far right corner. Auditioning of the current sound automatically takes place whenever grabbing the picker with the glove. When a satisfactory sound is located, a keyboard can be used for further play.

ISEE3D runs on any PowerMac with Quickdraw 3D. Sound synthesis on a YAMAHA TG77 is controlled by ISEE running on the same machine. A Powerglove is used to control picker movement in ISEE3D by MIDI commands. It is connected using a Doepfer MOGLI interface. Hand and finger movements are translated into ISEE MIDI commands using Opcode's Max software [4].

CONCLUSIONS

ISEE3D is a 3D visualization tool for structured browsing of instrumental sound that integrates hierarchical organization and sound object description into a single 3D visuo-spatial metaphor. Instrumental sounds can be located in a perceptually organized hierarchy of nested perceptual spaces using 3D positioning and some simple pointing gestures. Although examples were limited to real instrument simulations, our paradigm can also be used for more artistic sound object definitions. We hope ISEE and ISEE3D are a step towards the much-needed development of direct manipulation interfaces for sound retrieval from MIDI keyboards and computer software such as Apple's QuickTime™ Musical Instruments.

REFERENCES

1. Feiner, S. Worlds within Worlds. In Proceedings of UIST 1990. Snowbird, Utah: ACM, 1990.
2. Garner, W.R. *The Processing of Information and Structure*. Lawrence Erlbaum, Potomac, Md., 1974.
3. Grey, J.M. *An Exploration of Musical Timbre*. Ph.D. Dissertation. Dept. of Psychology, Stanford Uni., 1975.
4. Puckette, M. and Zicarelli, D. MAX. *Opcode Systems, Menlo Park, CA*, 1990.
5. Vertegaal, R. and Bonis, E. ISEE: An Intuitive Sound Editing Environment. *Computer Music Journal* 18(2), 1994, pp. 21-29.
6. Vertegaal, R. and Eaglestone, B. A Comparison of Input Devices in an ISEE Direct Timbre Manipulation Task. *Interacting with Computers* 8(1), 1996.
7. Wold, E., Blum, T., Keislar, D., and Wheaton, J. Content-based Classification, Search and Retrieval of Audio. *IEEE Multimedia* 3(3), 1996, pp. 27-36.