Intuitive Human Interfaces for an Audio-database

Barry Eaglestone
Department of Computing
University of Bradford
Bradford, BD7 1DP, UK
B.Eaglestone@comp.bradford.ac.uk

Roel Vertegaal
Department of Ergonomics
Faculty of Philosophy and Social Sciences
University of Twente
P.O. Box 217, 7500 AE Enschede, The Netherlands
R.Vertegaal@bsk.utwente.nl

Abstract

Database technology can now host multimedia applications through the representation of sounds and images, but such new applications also require extensions to HCI technology. This paper examines the problems of querying and manipulating audio information. We argue that no single "style" of user interface can provide a complete solution, and propose two novel types of interface to complement conventional database languages. The first is gestural, and allows users literally to reach into spaces of sounds and to "grab" the required objects. The second involves retrieval by mimicry. The main part of this paper describes our research into the viability of the gestural interface. We have experimented using the ISEE (Intuitive Sound Editing Environment) interface, a four-dimensional perceptually-based space of sounds. Our experiments have involved a user population and a range of multidimensional input devices, and have provided strong evidence that the approach is viable, but that the choice of input devices has a significant impact on the usability of the system. The second proposed interface, which we are currently researching, involves the use of neural networks within the data model to derive perceptually-based attributes. The neural networks can be trained on expertly created sound spaces, together with vocal imitations of the sounds, and subsequently used to retrieve on the basis of vocal imitations of the required sounds.

1. Introduction

Audio information is now an important dimension in multimedia systems. In particular, there are a number of specialist applications in the arts and in the sound industries concerned with music, television, cine and video, where there is a requirement for large databases of sounds [EA91a,JA90]. However, current database solutions to this problem are inadequate. Though physical storage of sounds is well researched (e.g. through the ESPRIT multimedia projects), this is less true of sound retrieval and manipulation interfaces. This weakness is particularly acute for artistic design applications in which design objects must be retrieved, manipulated and evaluated on the basis of their perceptual features [EA93a, EA93b].

This paper proposes two novel "styles" of intuitive audio-database interface which complement conventional database languages. The first is a gestural interface [VE94], which allows users to select sounds from a database by literally reaching into spaces of sounds and "grabbing" the required objects. The second involves querying the database by vocally mimicking the required sound - query-by-imitation (QBI) [DE91]. Both techniques are being researched as part of the work of the Sound Information Technology (SIT) project [EA91a]. SIT is researching a technology to support audio-related design projects, and involves specialists in music, digital signal processing, and computer science. The research is focused and coordinated within the framework of an extended IPSE architecture [EA91b], and has concentrated mainly upon the repository/database component [EA91b,EA93a,EA93b]. Ideas are being implemented and tested through the construction of a music composition demonstrator.

The structure of the paper is as follows. Section 2 analyses the problems of providing intuitive interfaces to audio-information, and includes the review of related research. The proposed gestural interface (that of the ISEE system) is described and discussed in section 3. Section 4 describes our experimentation with the ISEE interface and various input devices. Experimental results and conclusions are in section 5. Query-by-imitation is briefly discussed in section 6, and finally, sections 7 and 8 gives details of conclusions and acknowledgements.

2. Intuitive Querying of Audio-Information

Though artistic designers may be constrained by functional requirements, their main criteria for the acceptability of designs are usually non-functional, since they are to do with subjective aesthetic judgement [EA93a]. Where the product includes an audio component (as in multimedia authoring, musical composition and instrument design, soundtrack creation, etc.), it is therefore necessary that the design support system should support the representation of audio signals and the manipulation of them on the basis of their perceptual properties. Standards already exist for representation of audio signals. These are typically stored as sample values (perhaps using some destructive compression techniques) (e.g. AES 3-1985, MADI, 8-bit A-law, 16-bit linear, 24-bit linear, ISO/MPEG-Audio-standard), but can also be represented in an analysed form (e.g. group additive synthesis representation [EA90]), or as parameters for some sound synthesis algorithm. However, we know of no general database languages for perceptually-based definition and manipulation of audio information. There are two gaps in our knowledge which hinder the creation of this type of facility:

* lack of sophisticated and widely accepted perceptual models of sound;

* lack of knowledge about the mapping between audio signals and their perceptual properties.

These respectively deny us a standard vocabulary for characterising perceptual sound, and a basis for automatic derivation of perceptual properties from audio signals.

These two problems have been researched, mainly in the field of computer music and psycho-acoustics, but we believe the results provide a basis for more general applications, for example in multimedia design systems. In particular, progress has been made through the study of timbre space representations - a timbre space is a multidimensional space of sound, where each dimension models the variability of sounds with respect to some perceived characteristic. The viability of timbre spaces as perceptual interfaces to objects in an audio-database depends on achieving adequate characterisations of sounds within a manageable number of dimensions. However there are a number of results which indicate that these objectives are achievable. [WE74, GR75, PL76], for example, have established that it is possible to explain differences in timbre with far fewer degrees of freedom than are needed by most sound synthesis algorithms. [WE85] suggested using multidimensional scaling techniques [SH74] to map sounds into a timbre space. He derived a timbre space from a matrix of timbre dissimilarity judgements made by humans comparing all pairs of a set of timbres. In such a space timbres that are close sound similar, and timbres that are far apart sound different. However, this work highlights two problematic areas. Firstly, the manual construction of dissimilarity matrices is clearly not viable for indexing large audio-databases. Secondly, multidimensional scaling can lead to a proliferation of dimensions. [PL76] established that when using this technique, the number of timbre space dimensions increases with the variance in the assessed timbres.

The problem of implementing a mapping between sounds and their perceptual features is an example "soft confusion" problem [ZA93], since the nature of such a mapping is sensitive to many contextual criteria, including the identity and health of the listener, the acoustics, and juxtaposition with other sounds, and the evaluation of the correctness of solutions is subjective. Candidate technologies for this type of problem include connectionist theory, fuzzy logic and fuzzy set theory. Advancements have been made towards automatic creation of timbre spaces using connectionist technology. For example, Feiten and Ungvary [FE91] are making progress in training neural networks to automate the organization of sounds in a timbre space based upon the timbre characteristics identified in [CO84]. There is also evidence that for sound retrieval purposes (rather than manipulation) from classes of 'similar' sounds, a few parameters may suffice. Grey [GR75] for example experimented with similarity comparisons of 16 closely related re-synthesized instrument stimuli with similar envelope behaviour (varying from wind instruments to strings). He concluded that one dimension could express instrument family partitioning, another could relate to spectral energy distribution, and a third could relate to the temporal pattern of (inharmonic) transient phenomena.

A query language can use a timbre space interface to an audio-database by allowing users to specify coordinates in the space. The sounds at or neighbouring specified locations may then be retrieved, or alternatively sound synthesis parameters may be generated (by interpolation between neighbouring timbres). The latter is necessary if the language also allows modifications to stored sounds e.g. for sound synthesis control in music applications, as in the ISEE system [VE94]. A crude approach to implementing a perceptual timbre space interface is as an aggregation or set (a relational view for example) comprising for each sound in the space, its object identifier and timbre space coordinate values. Conventional retrieval techniques can then be used to access and re-synthesise the specified sound or neighbouring sounds from their physical representations. However, this approach imposes problems if automated interpolation is also required, for example to synthesize the sound associated with an unoccupied location in timbre space. An alternative or complementary approach is to represent the timbre space in terms of the functional mappings from perceptual coordinate to sound synthesis algorithm parameter values. In this latter approach the timbre space is characterised by its behaviour, rather than by describing each of its inhabitants, and can therefore be implemented in an object-oriented database through encapsulation of the mapping functions within the timbre space object. The work of Lee and Wessel [LE91, LE92] provides a basis for this second implementation. They have successfully trained a neural network to generate parameters for several synthesis models with timbre space coordinates as input, thus providing timbral interpolation automatically. The neural network code therefore provides the mapping function (method) encapsulated with the timbre space object. This approach however involves substantial computational power in order to train the neural network.

Though specific to music, the above research addresses some general problems of providing perceptually-based facilities for querying design objects. The timbre space solution described is an example of a perceptually-based object space in which each dimension models some variable perceived property of the objects. There are a number of advantages to using such a model as an interface. Objects can be retrieved by describing some of their perceptual properties, and then projecting along the axes corresponding to the other uninstantiated perceptual properties. The fuzzy nature of this type of querying means it can be viewed as an extension of the class 2 search in [BU73, SH90]. It is therefore appropriate to return objects which are in some sense "closest" to the query key. In fact the object space metaphor provides an implementation of a notion of "distance" between objects. A further advantage is the ability to "browse" or "explore" object spaces. This can be provided if the interface supports some "pointing" input device, such as a mouse or joystick, with which the user may "point to" locations within the space. If, in addition, the system provides instant retrieval (playback) of the objects as they are pointed to, we believe users can learn to navigate their way around the spaces, and locate objects with some required property without first having to formulate a query.

The approach does have limitations, in that it does not provide direct definition or manipulation of all design object attributes, or completeness with respect to the underlying data or object model. A consequence is that users are restricted to a pre-specified subset of objects, taken from a possibly infinite domain. In this respect it must be considered as a presentation model which is complementary to other more conventional querying facilities.

The following three sections describe the ISEE timbre space interface, and our experimental research, using ISEE, into the viability and effectiveness of the above type of "point and play" interface to an audio-database.

3. The ISEE Timbre Space Interface

The practicality of using timbre space as a basis for a sound design system is demonstrated by the ISEE (Intuitive Sound Editing Environment) system [VE94]. ISEE is a synthesizer and synthesis model independent user interface designed for musical sound design applications in both composition and performance. However, we also view ISEE as a demonstrator system with which the general concept of a perceptually-based multidimensional object space interface can be evaluated. Accordingly, the following description of ISEE omits technical details concerning signal processing and musical applications - those interested should refer to [VE92,VE94].

The ISEE interface is a four-dimensional timbre space. The dimensions were identified through qualitative observation of the working methods employed by expert designers of synthesized sounds. Because of their high level of abstraction, these parameters have important orthogonal properties which make them suitable as a basis for the high level ISEE sound synthesis model. The actual implementation of the abstract parameters depends on the required refinement of synthesis control. ISEE refers to a scaled implementation of the four 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. The four parameters are presented to the user as a pair of two-dimensional spaces called the Control Monitor (see figure 1). The first two of the abstract timbre 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 [GR75]. The Violin instrument space (see figure 2) 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.

The problem of defining the functions which map from timbre space dimensions to synthesis parameters is simplified by decomposition of the timbre space into a hierarchy of instrument spaces, each of which defines sub-classes of "similar" sounds. Separate mapping functions are then defined for each sub-class. 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 instruments spaces from grosser higher level spaces. 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.

Fig. 1. The Control Monitor application is used to control and monitor the position in the hierarchy (depicted by the middle icon) and the position in the current instrument space (indicated by the two dots). Two buttons are used to zoom out to the parent space (Harmonic) or zoom in to the child space (Violin) closest to the 4D position indicated by the dots.

Fig. 2. An example partial taxonomy of instrument spaces.

ISEE embodies the essential features of a perceptually-based multidimensional object space interface: each dimension models variability of a particular perceptual feature, dimensions are orthogonal, and the interface behaviour is characterised by mapping functions which relate locations in the space to specific objects. Two factors which determine the usability of this type of interface are 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.

4. Experiments with a hardware audio database interface

We believe that the choice of appropriate hardware input devices can increase the efficiency of an audio-database query by increasing the sense of direct manipulation of the audio-objects [FI67,SH87]. This is particularly true in music applications where many users will have a strong bond with direct manipulation of sound using musical instruments of some sort. We tested this assertion through experimentation using three low-cost hardware input devices (two general, low-dimensional devices and one specialized multidimensional device) to assess their usability in this relatively new application of information technology. The ISEE general sound specification model (described in the previous section) was used to provide the generalised low to high-dimensional mapping needed for gestural control of the search process.

4.1 Input Devices

Many studies have tested and compared usability of input devices for manipulation of on-screen graphic objects. We have therefore used this literature to assess which input devices were most suitable to test as timbre-parameter controllers for querying an audio-database.

According to [CH88] a multidimensional input device does not necessarily perform better in a multidimensional task than a low-dimensional input device. [JA92] states that performance in a multidimensional task with simultaneous control of all dimensions depends on the perceptual composition of the task's dimensions. We therefore selected both multi- and low-dimensional devices to control the ISEE interface. ISEE control consists of four medium resolution timbre parameters and one discrete scale parameter. The experiments focused on the usability of the input devices controlling the four medium resolution parameters, of which the control integration is unknown.

Another issue is whether absolute or relative devices are easier to use for audio-database query tasks. With an absolute device, the position of the device directly corresponds to the position of the controlled parameter. A relative device controls the direction in which the parameter changes. The nulling problem (i.e. the inconsistency between the state of the device and the state of the parameter that occurs when changing the parameter an input device operates on) that will occur when switching between the two 2-dimensional coordinate systems (see figure 1) using an absolute low-dimensional device might, according to [BU86], easily be solved by using a relative device instead. However, when using an absolute device, the position within an instrument space matches that of the controlling limb, which, according to [KE68, KE73, SH87], reduces cognitive processing load and corresponds more closely to control of most musical instruments. We therefore experimented with both direct and relative interfaces.

An excellent taxonomy of current input devices is given by [MA90]. It indicates, that multidimensional devices are mainly absolute devices. Of the relative devices, the mouse is the most commonly used. Joysticks can be changed from absolute to relative, which makes them ideal for comparing the usability of an absolute vs. a relative device. [PA91] has studied the use of two multidimensional input devices, the Nintendo Power Glove and the Polhemus 3Space Isotrak, as virtual reality controllers. However, since the magnetic field sensors of the Polhemus cannot be mass-produced, the Polhemus is too expensive to consider for regular database query applications. The low-cost Nintendo Power Glove, however, uses ultrasonics to sense three spatial coordinates and variable resistor material for sensing finger bend. Finger bend can be used to provide information about the status of the glove and to control the scale parameter. The three spatial coordinates can be used to control the first three ISEE parameters. The low resolution of the roll information of the Power

Glove (only 12 positions) however, would make specification of the fourth ISEE parameter rather crude. The roll information would only be useful for our purpose when used as a relative controller. This makes it impossible, however, to assess the control integration of all four ISEE parameters with a Power Glove.

4.2 Methods

We selected the following input devices: Apple Standard Mouse (a relative input device); Gravis Advanced MouseStick II - an optical high-resolution joystick (absolute or relative); 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 [CO90] was used with a group of 15 paid subjects who were asked to search an audio-database for objects matching target sounds using the different input devices. The experiments were conducted using an audio-database containing a broad selection of musical instrument sounds. These were generated by simple FM synthesis [CH73] (technical details of the parameters of the instrument space can be found in the appendix). An Apple Macintosh SE was used to filter the erratic Power Glove information and record the experiments. An Apple Macintosh LC was used as the database query platform running the ISEE system. A YAMAHA SY99 synthesizer was used to generate sounds according to the synthesis specification generated by the database objects. All systems were interconnected by MIDI, a general synthesizer LAN.

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 interfaces each provided feedback in the form of tones corresponding to the current position in the ISEE search space.

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 an order effect due to training, 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 sound, and then locate it in the audio-database, using one of the four interfaces. The location of the target sound could be seen in the Control Monitor while the target sound was being played (five times), and in a separate window throughout the rest of the experiment. After the initial sounding of the target sound, the indicators in Control Monitor centred, with the sound changing accordingly, giving the subjects an audio-visual cue to start manipulating the indicators with the input device. The tones were repeated throughout the experiment to give the subjects sufficient auditive feedback on their position. When the match was considered good enough, the subjects released the input device and recording was stopped.

Recording the experiments in this way 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.

5. Experimental Results and Conclusions

The efficacy of each device was established by measuring the time needed to reach the appropriate 4-dimensional position within a certain accuracy (where accuracy is overall Euclidian distance to target in 4-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 MIDI 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.

Fig. 3. The mean time (in msec) for each input device used in the experiments.

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. Figure 3 shows the mean time for each device. 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 ($p < 0.0001$).

It is clear from these findings that the Power Glove was not very effective in this 4-dimensional task. The subjects found it physically tiring, and very hard to control. However, the bad performance of the glove can also be partly attributed to the lag that occurred because of the filtering, the insufficient resolution of the device beyond 3 degrees of freedom and the fact that this device had not been used before by any of the subjects. During regular audio-database queries a mouse will suffice. When a keyboard is used for additional pitch and loudness specification, the absolute joystick is the most likely option, since it can easily be placed on top of the control panel of the keyboard. Also, in dark studio circumstances, absolute control can be very useful. The relative joystick only seems useful in a musical context where subtle changes in timbre need to be made, and no sudden jumps may occur. The ISEE user interface was judged by most subjects as being "pleasant to work with" and "intuitive".

6. Future work towards a "Query-by-imitation" interface

This penultimate section briefly describes our current research towards automating the creation of timbre space, and extending the intuitive search facilities.

ISEE timbre spaces used in HCI experiments (see sections 4 and 5) were manually created by a sound design expert. The functions which map from timbre space dimensions to synthesis parameters were "hand coded". The problems of perceptual object space definition is analogous to the more general problem of indexing within information retrieval systems. Both are skilled labour-intensive tasks which involve induction of object semantics so as to facilitate object retrieval with high precision and recall. However, the work of [FE91,LE92] indicates that automatic creation of timbre spaces may be possible using neural networks. Neural networks can be trained, using expertly created spaces as training sets, and can then be used to re-apply the induced design expertise in order to create other object spaces.

A "query-by-imitation" (QBI) search facility is a potential by-product of the above strategy. The idea for this type of interface was researched by Vertegaal, De Koning and Oates [DE91]. They researched an interface where-by users query an audio-database by vocally or textually imitating the required sound. The idea is based upon observed use of this vague form of query specification by film soundtrack engineers [EA91a]. However, it was also observed that the lack of an onomatopoeia indexing system could then necessitate a lengthy search for the required audio material, even when there was a common understanding between the requester and the searcher of what the required object sounded like. The search mechanism of the system proposed in [DE91] is based upon the conversion of the sound imitations into phonetic keys, which are then matched with stored sounds using conventional search techniques. The phonemes in fact provide a canonical form for both queries and stored sounds. Their approach builds upon techniques developed for speech recognition.

We believe our proposed neural network solution for automatic creation of timbre spaces can be extended to include vocal imitations of the characteristic component sounds, and thus provide a mechanism for associating sound imitations with locations in perceptual space. The advantage over the phoneme-based searching is its greater generality. The system does not rely upon a notional phoneme-based granularity of sounds, and can be adapted to the individual user, since users can generate their own training set of vocal imitations. We also believe that the QBI approach has some generality. There is an obvious analogy with image-databases whereby the user queries the database by sketching the require object. We are currently developing and experimenting with a prototype implementation of the above QBI system.

7. Conclusions

The gestural and QBI interfaces to audio-databases described are complementary to conventional database languages. For example, we envisage a situation in which QBI provides a mechanism by which a user specifies an approximate location within an object space. The user then refines the query, perhaps through a gestural search. Finally, the required multimedia information is retrieved through the use of some database model-complete language.

We believe that the techniques have some generality within artists design systems, since they provide a general strategy for perceptually-based manipulation of physically represented non-textual design objects.

The main contribution of this paper is the experimental evidence concerning the choice of input device for a gestural interface. The experimental results strongly indicate that the mouse, already used in most applications, is the best input device for audio-database queries with a timbre space approach. The absolute joystick is best used when a keyboard is applied to specify pitch and loudness. Multi-dimensional input devices need not necessarily perform better in a multidimensional query task.

8. Acknowledgements

We would like to thank Apple Computer Inc., and in particular, S. Joy Mountford of the Apple Human Interface Group, for supporting the above research, and Dr Michael Clarke of the University of Huddersfield Music Department for facilitating the experimental work. We would also like to thank the other members of SITRG, Geoff Davies, Tamas Ungvary, Bernhard Feiten and Iain Millns, for their contributions to our research.

References

[BU73] W.A. Burkhard. Some Approaches to Best-Match File Searching, pp 230-236, Comms ACM, 16(4), 1973.

[BU86] W. Buxton. There's More to Interaction than Meets the Eye: Some Issues in Manual Input, in User Centered System Design: New Perspectives on HCI, D.A. Norman and S.W. Draper, Editor. 1986, Lawrence Erlbaum Associates: Hillsdale, N.J. p. 319-337.

[CH73] J. Chowning. The synthesis of complex audio spectra by means of frequency modulation. Journal of the Audio Engineering Society, 1973. 21(7): p. 526-534.

[CH88] M. Chen, S.J. Mountford, and A. Sellen. A Study in Interactive 3-D Rotation Using 2-D Control Devices. Computer Graphics, 1988. 22(4): p. 121-129.

[CO84] R. Cogan. New images of Musical Sound. Havard U P, 1984.

[CO90] H. Coolican. Research Methods and Statistics in Psychology. 1990, London: Hodder & Stoughton.

[DE91] K De Koning and S. Oates. Sound Base: Phonetic Searching in Sound Archives. International Conference on Computer Music, Montreal, 1991, pp 433-436.

[EA90] B.Eaglestone and S. Oates. Analytical tools for Group Additive Synthesis. International Conference on Computer Music, Glasgow, 1990.

[EA91a] B. Eaglestone and A. Verschoor. Dichtslaande deuren en mens-machine interfaces, Kennissystemen jrg 5 nr 5 mei 1991, pp 17-21.

[EA91b] B. Eaglestone and A. Verschoor. An Intelligent Music Repository International Conference on Computer Music, Montreal, 1991, pp 437-440.

[EA93a] B. Eaglestone, G.L.Davies and T. Ungvary. An Extended Version Model for Artistic Design. 5th International Conference on Computing and Information, IEEE, Sudbury, Canada, 1993.

[EA93b] B. Eaglestone, G.L. Davies, M. Ridley and Hulley N. Implementation of an Artists Version Model using Extended Relational Database Technology. Advances in Databases, BNCOD-11, Keele, UK, July 1993, Lecture Notes in Computer Science 696, Springer Verlag, 1993, pp 258-276.

[FE91] B. Feiten and T. Ungvary. Organisation of Sounds with Neural Nets. Proceedings of the 1991 ICMC, Montreal, International Computer Music Association, 1991.

[FI67] P. Fitts and M. Posner. Human Performance. London, Prentice-Hall, Inc. 1967.

[GR75] J. Grey. An Exploration of Musical Timbre. Ph.D. Dissertation, Dept. of Psychology, Stanford University. CCRMA Report STAN-M-2, 1975.

[JA90] M. Jaslowitz, T. D'Silva and E. Zwaneveld. Sound Genie - An Automated Digital Sound Effects Library System, SMTE Journal, May 1990, pp 386-391.

[JA92] R.J.K. Jacob and L.E. Sibert. The Perceptual Structure of Multidimensional Input Device Selection, in Proceedings of ACM CHI'92 Conference on Human Factors in Computing Systems. 1992, p. 211-218.

[KE68] S. Keele. 1968. Movement Control in Skilled Motor Performance. Psychological Bulletin 70, 1968, pp 387-402.

[KE73] S. Keele Attention and Human Performance. Pacific Pallisades, Goodyear Publishing Company, 1973.

[LE91] M. Lee, A. Freed, D. Wessel. Real-Time Neural Network Processing of Gestural and Acoustical Signals, Proceedings of the 1991 ICMC (International Computer Music Association, Montreal, 1991), pp. 277-280.

[LE92] M. Lee and D. Wessel. Connectionist Models for Real-Time Control of Synthesis and Compositional Algorithms. Proceedings of the 1992 ICMC, San Jose, International Computer Music Association, 1992.

[MA90] J.D. Mackinlay, S.K. Card, and G.G. Robertson. A Semantic Analysis of the Design Space of Input Devices. Human-Computer Interaction, 1990. 5: p. 145-190.

[MO85] A. Monk. Statistical Evaluation of Behavioural Data, in Fundamentals of Human-Computer Interaction, A. Monk, Editor. 1985, Academic Press: London. p. 81-87.

[PA91] R. Pausch. Virtual Reality on Five Dollars a Day, in Proceedings of ACM CHI'91 Conference on Human Factors in Computing Systems. 1991, p. 265-270.

[PL76] R. Plomp. Aspects of Tone Sensation. London, Academic Press, 1976.

[SH74] R. Shepard. Representations of Structure in Similar Data: Problems and Prospects. Psychometrica 39, 1974, 373-421.

[SH87] B. Shneiderman. Designing the User-Interface: Strategies for Effective Human-Computer Interaction. Reading, MA, Addison Wesley. 1987.

[SH90] D.Shasha and T.-L. Wang. New Techniques for Best-Match Retrieval. pp 140-158, ACM TOIS 8(2), 1990.

[TR77] B. Truax. Organizational Techniques for c:m Ratios in Frequency Modulation. Computer Music Journal, 1977. 1(4): p. 39-45.

[VE92] R. Vertegaal. ISEE: ontwerp en implementatie. Music Technology Dissertation, Utrecht School of the Arts, The Netherlands, 1992.

[VE94] R. Vertegaal and E. Bonis. ISEE: An Intuitive Sound Editing Environment. Computer Music Journal, to be published 1994.

[WE74] D. Wessel. Report to C.M.E. University of California, San Diego, 1974.

[WE85] D. Wessel. Timbre Space as a Musical Control Structure. In C. Roads and J.

Strawn, ed. Foundations of Computer Music. Cambridge, MA, MIT Press, 1985.

[ZA93] L.A. Zadeh. Soft Confusion and Fuzzy Logic 5th International Conference on Computing and Information, IEEE, Sudbury, Canada, 1993.

Appendix - Experimental Instument Space Definition

The four parameters of the instrument space were defined as follows. 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 [TR77] for a more detailed explanation)). The Brightness parameter was used to control the cutoff 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 perceptual mapping as possible with simple FM.