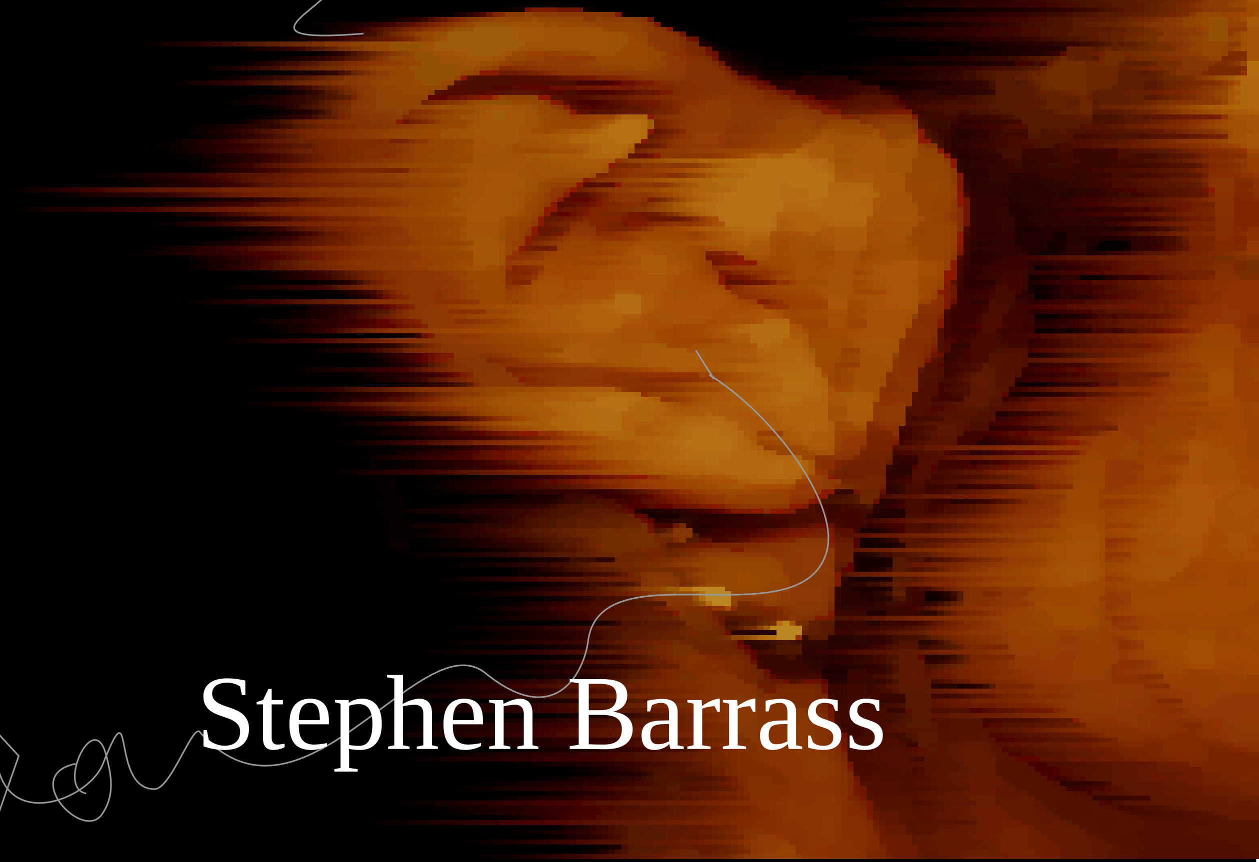
Auditory Information Design



Stephen Barrass

A thesis submitted for the degree of
Doctor of Philosophy of
The Australian National University

July 31, 1997

Declaration

Some material in this thesis has previously been made public. Parts of Chapters 3 and 10 appeared in a paper for the International Conference for Auditory Display 1996. Parts of Chapters 4, 5 and 10 appeared in a paper for the Asia-Pacific Conference on Human-Computer Interaction 1996. Chapter 6 is in a book on the Csound audio processing language to be published by MIT Press in 1997. Parts of Chapter 7 were presented at the Australian Computer Music Conference 1994. Other parts of Chapter 7 appeared in a paper, co-authored with Phil Robertson, in the proceedings of SPIE ‘95. Parts of Chapters 6, 7 and 8 appear in the proceedings of the International Conference on Auditory Display 1994. Parts of Chapters 7 and 8 are in an article for the Organised Sound journal 1,2 published by Cambridge University Press in 1996. Parts of Chapter 9 were presented at the Australian Computer Music Conference 1995.

Some of the tools and demonstrations described in this thesis utilise tools developed by other people, and I am pleased to declare their contributions. Don Bone wrote the tools for modelling colour output devices which were used to construct the sound space in Chapter 7. Matthew Hutchins wrote graphic interfaces and colour mapping algorithms that were extended to become the Personify tool in Chapter 9. The demonstrations in Chapter 10 build on the work of Simon Kravis in the RiverAndRain visualisation, Rochelle O’hagan and Danielle Landy in the GeoViewer, and Chris Gunn in the MicroSeismic Animator.

Except where otherwise stated, I declare that this thesis is my own original work

Stephen Barrass
July 31, 1997

Abstract

The prospect of computer applications making “noises” is disconcerting to some. Yet the soundscape of the real world does not usually bother us. Perhaps we only notice a nuisance? This thesis is an approach for designing sounds that are useful information rather than distracting “noise”. The approach is called TaDa because the sounds are designed to be useful in a Task and true to the Data.

Previous researchers in auditory display have identified issues that need to be addressed for the field to progress. The TaDa approach is an integrated approach that addresses an array of these issues through a multifaceted system of methods drawn from HCI, visualisation, graphic design and sound design. A task-analysis addresses the issue of usefulness. A data characterisation addresses perceptual faithfulness. A case-based method provides semantic linkage to the application domain. A rule-based method addresses psychoacoustic control. A perceptually linearised sound space allows transportable auditory specifications. Most of these methods have not been used to design auditory displays before, and each has been specially adapted for this design domain.

The TaDa methods have been built into computer-aided design tools that can assist the design of a more effective display, and may allow less than experienced designers to make effective use of sounds. The case-based method is supported by a database of examples that can be searched by an information analysis of the design scenario. The rule-based method is supported by a direct manipulation interface which shows the available sound gamut of an audio device as a 3D coloured object that can be sliced and picked with the mouse. These computer-aided tools are the first of their kind to be developed in auditory display.

The approach, methods and tools are demonstrated in scenarios from the domains of mining exploration, resource monitoring and climatology. These practical applications show that sounds can be useful in a wide variety of information processing activities which have not been explored before. The sounds provide information that is difficult to obtain visually, and improve the directness of interactions by providing additional affordances.

Keywords:
Auditory display, sonification, information design, interaction design, sound design, task analysis, computer-aided design, human-computer interaction, user interface, multimedia interfaces, visualisation, data representation.

Acknowledgements

The work in this thesis grew out of a project on colour applications and modelling (CAM) for scientific visualisation at the Commonwealth Scientific and Industrial Research Organisation (CSIRO), and I would like to thank Don Bone, Phil Robertson, Matthew Hutchins, Duncan Stevenson, Chris Gunn and Dionne Smith for what I learnt from them in that work. Ideas of extending the colour framework to sounds arose from conversations with my brother Tim, who has always been a source of imagination and inspiration for me. Thanks to Kim Marr, Joni Bryan and Anne-Marie Crowe who read my thesis proposal and said “go for it” when I wasn’t too sure. Thanks to Phil Robertson for his positive response and support for the proposal.

My research was funded by a scholarship from CSIRO, and a professional development grant from the Advanced Computational Systems Cooperative Research Centre (ACSys CRC). I would like to thank my principal supervisor, Phil Robertson of Canon Informations Systems Research Australia (CISRA), and the rest of my thesis committee, David Worrall of the Australian Centre for Arts and Technology (ACAT), and Paul Thistlewaite and Jeffrey Yu from the Australian National University (ANU) for their academic support throughout the process.

There could be no better place than CSIRO to work and study, and I am very grateful for the environment and the people there. My colleagues helped me and taught me by reviewing and discussing my work - thanks to Matthew Hutchins, Neale Fulton, Simon Kravis, Phil Robertson, David Worrall, Don Bone, Ken Tsui, Danielle Landy, Tao Lin, and John Lilleyman. The demonstrations of my ideas depended on the willingness of people to let me to work on one of their projects - thanks to Simon Kravis for RiverAndRain, Rochelle Ohagan and Danielle Landy for GeoViewer, and Chris Gunn for MicroSeismic Animator. Thanks to Don Bone for helping me to use his colour modelling tools, and to Matthew Hutchins for his work on the graphical interfaces that are the basis for the Personify tool. There are many other people at CSIRO to thank as well - some lent me their ears by taking part in the Gold and Dirt experiment, others lent me their voices for the cOcktail demonstration, and others lent me stories for the EarBenders database. My enthusiasm for my topic was kept alive by the interest of those around me, and I particularly remember Neale Fulton’s drop-in visits as sparks that helped keep my research spirit alive. Thanks to the other postgrads for sharing the spirit and talking the thesis talk - Matthew Hutchins, Ron Li, Kevin Moore, Andrew Lui, Mark Cameron - good luck.

One of the most important and enjoyable events in the course of this project was the opportunity to meet other researchers at the International Conference on Auditory Display (ICAD). The forum provided by ICAD helped focus my research and was a vital part of my experience. Thanks to the organisers of ICAD - Greg Kramer, Steve Frysinger and Stuart Smith, and to CSIRO and ACSys for funding my attendance.

Finally there are the people who kept me sane during this single-minded endeavour. Lunch-hours were regularly stretched in the company of Stephen Pratt, Ninh Hyne, and Sarah May, and I thank them for the cappuccinos, bad-jokes, good advice, and hot gossip. Eternal thanks Mum and Dad for my education, for listening and talking with me about what I was doing no matter what it might be, and for a place that always feels like home. Thanks to Linda, Robert, Lauren and Joel for fun and games, more dinners and another place that feels like home. A woof to Ziggy for all the walks up Mt. Ainslie, and for teaching me to both throw and fetch the frisbee. Finally, my heartfelt thanks to Bettina for being with me through all the ups and downs on a long ride, but at last the end of the tunnel is in sight and its sunny outside so let me pick you some flowers.

Contents

Figures

Tables

1 • Introduction

Sounds are distinguished from noises by being useful, informative, answering a question, and supporting a task or activity. In the kitchen the sounds of food frying, liquids boiling, timers going off, and pots clanging help the chef prepare a meal. On a construction site the workers use the hammering of nails, clattering of tiles and revving of engines to coordinate their tasks in a common project. On a farm a stockman may listen for cattle hidden in the bush, or diagnose a faulty pump that is making an erratic sound. Sounds are a natural consequence of activity in the physical environment, and although we are seldom aware of it, we are always listening and using sounds in many ways.

1.1 Motivation

The computer-based workplace is unnaturally quiet...and disquietingly unnatural...

Contrast the informative soundscape of the everyday world with the silence of the computer-based workplace. More and more people are spending more and more time carrying out their workday activities in this environment, and there is a suggestion that just as sounds help us in the everyday world they may also be helpful in computer-based activities too. This suggestion is sometimes met with the reservation that, although it might be fun, a “noisy” computer could become annoying. This concern highlights the need to design useful sounds, rather than amusing novelties or distracting noise. Just like the soundscape of the real world, a well-designed soundscape for the computer-based workplace will seldom be noticed, except through the marked effects of its absence.

Sound design is a well-developed practice in cinema. The first movies were silent, but the silent era did not last long. Within 30 years a soundtrack with church bells, train whistles and accompanying music, had become a part of the movie experience. Today the moviegoer is immersed in a multi-layered 3D soundscape of voices, sounds, music and special effects that is often essential to the understanding of the film. The simple bells and whistles of multimedia computer interfaces are reminiscent of the first movie soundtracks, and they too signal the end of a silent era. Although there is much to learn from existing practices, the design of sounds to support information-processing activities is a new challenge because of the types of information involved, and the need to communicate that information clearly and unambiguously. Some major obstacles that need to be addressed in auditory display were described by Smith in a panel on Human Perception and Visualisation [Smith S. (1990)]...

The first obstacle is the prevailing sonification model, which is simply to map data to sound parameters arbitrarily. The resulting sound is typically unpleasant and lacking any natural connection to the data represented (one intuitively feels that medical images, for example, ought to somehow sound different from demographic data or satellite imagery). Models of sonification more sensitive to the kinds of data presented must be developed.

The second major obstacle is the lack of suitable sound generation hardware. Sonification requires a general purpose real-time sound synthesis capability, preferably at an affordable price.

Finally, the third major obstacle is the nearly total absence of the kinds of models that allow design of computer graphics software systems that can run successfully on hardware made by many different manufacturers. The principle reasons for this situation are the *lack of a satisfying comprehensive theory of timbre perception and the lack of an agreed upon theory of timbre generation.*

These translate directly into the situation we observe today: multiple incompatible sound-generation devices, each accompanied by its own suite of non-standard application packages.

The need for a linkage between the characteristics of the data and the auditory perception of the data is reiterated in other surveys of the state-of-the-art. Kendall makes some suggestions about how this linkage might be made in an article titled Dream Machines [Kendall G.S. (1991)] ...

Some classifications of sound events tend to be categorical. Excitation functions are typically of discrete types such as hitting, scraping, blowing, vocal glottis etc. Some classifications of sounding objects are similarly categorical - metal, wood, hollow, solid, vocal tract etc. These simple categorical distinctions can potentially be exploited in auditory presentations to communicate important distinctions in the data.

Beyond these categorical distinctions, the essential goal is that perceptually continuous auditory attributes are scaled and mapped to data attributes in a way that is meaningful to the observer. Relevant changes in data should insure a change in what is perceived. Changes in what is perceived should signify meaningful changes in the data. The appropriate scaling functions will probably not exist a priori. Psychophysical scaling experiments may be needed in order to create perceptual scaling functions through which collections of auditory stimuli are mapped. This is made feasible only by utilizing a limited number of auditory tokens with well-understood perceptual properties. This suggests that sets of tokens be developed and scaled in advance.

Issues of perceptual representation are also highlighted by Frysinger [Frysinger S.P. (1990)]...

Some serious questions which must be addressed if we are to generate reliable and meaningful auditory displays. We must discover the set of truly useful auditory parameters and understand their perceptual transfer functions so that displays can be designed to take advantage of them. Likewise we need to understand which data analysis tasks can most benefit from Auditory Data Representation, and what types of displays apply to them.

Kramer found that the psychoacoustic interaction between acoustic parameters in his multidimensional sonifications had a significant influence on the comprehension of the display, and concluded that the realisation of a balanced auditory display with independently observable auditory dimensions may not be possible in practice. Like Frysinger he also notes the importance of a task-oriented approach [Kramer G. (ed) (1994b)] ...

Sonification system design should be heavily task dependent. Techniques that are applicable to one task, e.g. real time monitoring, may not be as effective on other tasks, e.g. exploration of a high-dimensional data set. Practical use of the techniques (and of sonification in general) will prove or disprove their utility.

The need for demonstrations of practice, and tools to support practice, is raised by Scaletti in her list of open questions in sonification from ICAD'92 [Scaletti C. (1994)]...

Broad categories for further study include:

Applications: further and more sophisticated examples of sonification applied to specific problems

Sonification science: studies in the perception, cognition and neurophysiology of data-driven sound

Systems: Needs for future hardware and software include: integrated sonification/visualisation languages, tools for getting from an imagined sound to a realised sound, the integration of sonification tools into mass market software like spreadsheets or statistical analysis packages, and more and better tools for exploratory sonification.

This thesis is motivated by the need address obstacles identified by previous researchers as important for progress in the field of auditory display, which I have listed as

- usefulness of the sounds in an activity
- faithful representation of data relations
- semantic linkage to the application domain
- psychoacoustic control
- device-independent display specification
- computer aided tools for auditory information design
- demonstrations of practical application

1.2 Thesis

My thesis is an approach for designing sounds to support information processing activities. The approach focuses design on an auditory representation to meet the information requirements of the activity. This focus is at the core of a system of methods that address the array of issues raised by previous researchers. A task-analysis addresses the issue of usefulness. A data characterisation addresses the issue of faithful representation. A case-based design method addresses semantic linkage with the application domain. A rule-based method addresses psychoacoustic control. A perceptually linearised sound space addresses device-independent specification. This multifaceted system is flexible to cope with the obstacles encountered in design practice. Most of these methods have never been applied in auditory display before, and each has been adapted specially for this design domain.

1.3 Layout and Overview

The layout of the thesis chapters is organised to reflect the TaDa design process. An introduction to each Chapter is given here to give you an overview of what is to come.

Chapter 2: Previous approaches

Approaches to auditory display have been classified into the semiotic types of lexical, syntactic and semantic [Blattner M.M. Papp A.L. and Glinert E.P. (1994)] which place different emphasis on learnability, organisation, and discrimination. However there are other approaches to the design of information displays that raise other issues about usefulness, usability, social value and the realisation of the design on a device. This chapter begins with an introduction to semiotic terms, followed by descriptions of approaches to design that emphasise syntactic, semantic, pragmatic, perceptual, task, connotation and device issues. The collection of a broad range of issues may help us to develop a broad and flexible approach to auditory information design

Chapter 3: Designing useful sounds

This chapter proposes an approach for designing useful sounds. The approach builds on Scaletti's working definition of sonification, which is analysed to have two parts - one part has to do with information requirements and the other with information representations. The requirements part addresses issues of usefulness in a task and the selection of useful data relations to display. The representation part addresses the need to ensure that people can hear the required information in the display. These parts are shaped into a framework that focuses on the design of an information representation to meet the information requirements of a task. The phrase *auditory information design* indicates this focus on *useful information* which is at the core of this approach.

Chapter 4: TaDa: information requirements analysis

The purpose of the TaDa approach is to design sounds that carry useful information. The core of the approach is a meeting of information requirements with an information representation. This chapter describes the methods used to elicit the information requirements of a design problem. The first section introduces scenario analysis as a technique for capturing key features of the problem. The next sections describe the particular flavour of task analysis and data characterisation used to decompose the problem, and give a detailed account of the parts of analysis.

Chapter 5: EarBenders: case-based design from stories

Designers often base a new design on a previous version that has proven successful in similar problems. A previous solution can be a way to quickly come to grips with a design problem, and provides a starting point for a top-down process of iteration. This chapter introduces the case-based method of design by example, and describes how it has been adapted for auditory information design. The case-based method relies on a rich source of examples to be effective, but as yet there are not many examples of auditory display design to draw upon. An alternative resource was developed by collecting stories about everyday listening experiences into a database, which I call EarBenders. The information requirements of a design problem can be used to search this case-base for everyday examples which share a similar task, data and information structure with the problem. The ability to do this search required each story to be analysed with the TaDa Information Requirements Analysis developed in the previous chapter. In addition an auditory characterisation was developed to describe the sounds in each story, and provide a footing for auditory design. The sound characterisation also provides an opportunity to extract principles of sound design from regularities between auditory structure and information structure in the example cases. The case-based design of auditory information is demonstrated on a problem in a geological visualisation interface, called the GeoViewer.

Chapter 6: Hearsay: principles for auditory design

This chapter proposes the Hearsay principles to help a designer to meet the information requirements specified by the TaDa analysis. Hearsay integrates principles for information design with observations about auditory perception. Each Hearsay principle was investigated by generating a simple auditory demonstration to confirm that characteristic properties can be heard. The demonstrations show that the required information can be represented by auditory relations, and that the Hearsay principles are applicable in practice. The principles were tried out in a design of an auditory display for Bly's 'dirt and gold' scenario. The display enables a listener to quickly answer the question "is there gold in this pile of dirt?". The effectiveness of this display indicates that the principles are helpful in practice.

Chapter 7: Information-Sound Space: a cognitive artifact

The previous chapter introduced the Hearsay principles for auditory information design, that summarise some knowledge that can help in the design process. Although they are helpful, principles and guidelines can be unwieldy in practice because of the need to keep referring back to them. Principles cannot be simply applied by rote, they have to be learnt and understood. This chapter describes an alternative representation of the Hearsay rules in the form of an Information-Sound Space (ISS). The ISS is a three dimensional spatial organisation of auditory relations that embodies the Hearsay principles. This tool bridges the gap from theory to practice by changing the way the designer can think about and manipulate relations between sounds. Rather than having to follow written principles the designer is able to think in terms of simple spatial structures that represent information relations. The following sections describe the development of the ISS, which is based on the HSL colour space that has been applied in many areas of design including scientific visualisation of data sets. The feasibility of implementing an ISS is investigated in several experiments that draw upon psychoacoustic observations made by Von Bismarck, Slawson, Grey, and Bregman.

Chapter 8: GreyMUMS: an Information-Sound Space

This chapter describes the realisation of an Information-Sound Space (ISS). The raw material for the construction is the McGill University Master Samples (MUMS) reference palette of musical samples that is specifically intended for research into musical timbre. The ISS was constructed in 4 stages - the pedestal, the frame, the grating and the plasticine. The pedestal is 8 equally spaced timbre steps organised in a circle by similarity. The frame is an arrangement of brightness profiles which define the limits of dynamic range in pitch and brightness for each timbre. The grating consists of grids of equal differences in brightness and pitch for each timbre. The grating grids are joined on a central axis and radiate outward like segments of a mandarin. The plasticine is a continuous medium moulded to the grating and frame to model the behaviour of the overall space. The resulting sculpture has the property that there is a relationship between distance and the strength of perceptual grouping between points in the space. A vertical line is a pitch scale, and may be used to represent continuous data. A radial line is a scale of equal brightness increments for a timbre, and may also be used for continuous data. A circle of constant radius is a contour of constant brightness across the range of timbres which can be used to represent categorical data. These properties are a rich area for further experiment with data mappings.

Chapter 9: Personify: computer-aided design tool

Multi-modal interfaces are becoming increasingly important, and designing sounds for the human-computer interface is something more people are going to want to do. This chapter describes the Personify tool that can assist in the design of useful and effective sounds. The tool integrates a principled design approach with a direct manipulation interface. The guidance provided by the tool makes it quick and easy to use, and improves the likelihood of producing an effective display. The chapter begins with an overview of interfaces for handling audio material found in musical tools. This is followed by an overview of tools that are specialised for auditory display design, with attention to the way these tools allow you to handle sounds. The Personify tool is then described in two parts - the Requirements part and the Representation part. The meshing of the parts as a system is demonstrated in a design scenario that involves resource monitoring by satellite data.

Chapter 10: TaDa! demonstrations of auditory design

This chapter describes the design of 4 auditory displays for information processing scenarios drawn from mining exploration, resource planning and climatology applications.

The RiverAndRain scenario is about the siting of a new sewerage treatment works to minimise the environmental impact on a river system. The PopRock scenario involves the assessment of risk in digging a mineshaft. The cOcktail scenario is about modelling climate change from measurements of oxygen isotopes in sea-bed drill-core sites. The LostInSpace scenario is about navigating back to an interesting place through irregular structures in a 3D visualisation of geology. Experiences with the multimedia interfaces that were implemented shows that the sounds can provide information that is difficult to obtain visually, and can improve the usefulness of the display. Besides showing ways that sounds can be useful, the demonstrations also show how the TaDa approach to design works in practice.

2 • Previous approaches

Approaches to auditory display have been classified into the semiotic types of lexical, syntactic and semantic [Blattner M.M. Papp A.L. and Glinert E.P. (1994)] which place different emphasis on learnability, organisation, and discrimination. However there are other approaches to the design of information displays that raise other issues about usefulness, usability, social value and the realisation of the design on a device. This chapter begins with an introduction to semiotic terms, followed by descriptions of approaches to design that emphasise syntactic, semantic, pragmatic, perceptual, task, connotation and device issues. The collection of a broad range of issues may help us to develop a broad and flexible approach to auditory information design.

2.1 Semiotics

Semiotics is a theory of signs and their meanings that has been used to analyse communication media. Some key terms and concepts in semiotics are introduced here (based on Chandler's tutorial [Chandler D.C. (1997)]).

A "sign" is anything from which a meaning may be generated - words, sounds, photographs, clothing etc. A sign has two parts - a "signifier" which is the form that the sign takes, and the "signified" which is what it represents to the person who perceives it. Semiotic principles are commonly divided into 3 kinds - syntactic, pragmatic, and semantic.

- Syntactic principles bear on the way signs are organised to produce meanings. Signs may be organised in parallel (paradigm) or serial (syntagym). Rules for organising signs are called grammars.
- Pragmatic principles bear on the material form of the signifier in the sign. It is usually important for different signs to be perceptually distinct, recognisable and memorable. A set of signifiers that have a special purpose, such as the letters of the alphabet, are called a lexicon, or sometimes a palette.
- Semantic principles bear on what is signified by the sign. The signified is a concept in the mind of an observer, not a material "thing". The signified concept of a ball contains many different objects of different materials, sizes, colours and shapes. Concepts can be learnt from other people, or from experiences. The association between a signifier and the concept that is signified is commonly classified into 3 kinds
    - Symbolic the signifier does not resemble the signified e.g. the spoken word "ball" doesn't sound like a ball, and the written word "ball" doesn't look like a ball.
    - Indexical the signified is causally connected to the signifier e.g. hitting a tennis ball makes a characteristic whacking sound, bouncing a basketball makes a bouncing sound.
    - Iconic the signifier resembles the signified e.g the sampled sound of a cricket chirping in the soundtrack of a movie signifies "real" crickets chirping, and

the photographic image of a person signifies a “real” person.

Signs can signify more than one thing at the same time. The concept that the sign stands for is called the “denotation” and additional signifieds are called “connotations”. Cultural associations generate connotations by metonym and metaphor. A metonym invokes an idea or object by some detail or part of the whole - a picture of a horseshoe may be a metonym for a horse. A metaphor expresses the unfamiliar in terms of the familiar - a picture of a tree may be a metaphor for a genealogy. A metonym is considered more “natural” than a metaphor because it does not require such a leap of imagination (transposition). Connotations may transform the meaning of a sign through emotional overtones, subjective interpretations, sociocultural values and ideological assumptions.

Auditory display techniques were classified as syntactic, semantic and lexical by Blattner, Papp and Glinert [Blattner M.M. Papp A.L. and Glinert E.P. (1994)]. An example of a syntactic approach is the earcon, which is a short musical motif with structure modelled on pictographic writing [Blattner M. Sumikawa D. and Greenberg R. (1989)]. An example of a semantic approach is the auditory icon, which is modelled on the everyday sounds caused by interactions between material objects [Gaver W.W. (1986)]. Lexical approaches map data variations to acoustic variations, and examples include Bly’s multivariate mappings [Bly S. (1994)], Kramer’s “parameter nesting” [Kramer G. (1994a)] and the granular texture technique described by Smith, Pickett and Williams [Smith S. Picket R.M. and Williams M.G. (1994)].

2.2 Syntactic approach

The emphasis in the syntactic approach is on the organisation between auditory signs. Morse code is a syntactic approach where variation in duration and rhythm of a non-speech sound can communicate coded text messages. The earcon is a syntactic method for designing non-speech sounds to represent information in human-computer interfaces. An earcon is built from components that may vary in rhythm, pitch, timbre, register, and dynamics. Each earcon has a unique meaning that must be learnt - for example a tone X with pitch 440 Hz may mean “file”, and tone Y with pitch 600 Hz may mean “deleted”. These earcons can be combined to communicate more complex messages - for example playing X and Y in series produces a rising XY earcon that means “file deleted” [Blattner M. Sumikawa D. and Greenberg R. (1989)]. The syntactic structure of an earcon can be organised by transformations, combinations, inheritance, and polyphony. Blattner et al. suggest that earcons have the advantages of

- *Ease of production: earcons can be easily constructed and produced on almost any computer with tools that already exist for music and audio manipulation.*
- *Abstract representation: earcon sounds do not have to correspond to the objects they represent, so objects that either make no sound or an unpleasant sound can still be represented.*

Earcons were added to a map of a research facility to provide extra information that could be heard by pointing or selecting a region with the mouse [Blattner M.M. Papp A.L. and Glinert E.P. (1994)]. Access privileges of a building were heard by a knocking pattern of a tom-tom drum. Higher restrictions are heard by a faster knock and higher pitch. The presence of computers in a building was heard as a four note flute tune. The presence of an administrative unit in a building was heard by a 3 note saxophone tune. This demonstration of earcons raises an important issue - symbols are categorical and the categorical representation of ordered information requires a decoding phase. The need for more direct representations of ordered values is recognised in the speeding up of rate to represent the higher restriction levels in the access privileges earcon. However this type of representation is not an explicit part of the earcon method.

A major problem with earcons is learnability. Novices are able to learn 4-6 symbolic sounds within minutes, but further learning of up to 10 signals can take hours. Beyond 10, the process is prolonged and some listeners may never learn the catalogue completely [Patterson R.D. (1982)]. There is no standard syntax or lexicon of earcons, and the investment of time and effort in learning a new set may be too great for many applications.

2.3 Semantic approach

The emphasis in the semantic approach is on what is signified by a sound. The semantic method for sounds in user interfaces is called the auditory icon. The auditory icon method is to map what is signified by a familiar everyday sound to objects and events in the user interface [Gaver W.W. (1986)]. Gaver suggests that sounds modelled on real world acoustics are likely to be learnable and easy to understand because humans are adapted to hear information in these kinds of sounds. This suggestion is based on a theory of perception that says that physical variations in acoustic energy generated by interactions between material objects are innately perceived and intuitively understood [Gibson J.J. (1966)].

The design of an auditory icon starts with an analysis of interactions between objects in the interface which would cause sounds in the physical world. For example, moving a file in the Apple desktop GUI involves dragging it between windows. An auditory icon for this event is based on the sound caused by dragging a real file across a real desktop. Gaver demonstrated auditory icons in the SonicFinder, which augmented the Apple desktop GUI with auditory icons for selecting, dragging and copying files, opening and closing folders, selecting scrolling and resizing windows, and dropping files into and emptying the trash can [Gaver W.W. (1994)]. He comments that although many people found the auditory cues useful, others found them irritating or merely entertaining - providing an indication of the challenges facing the designer in real-world practice. In another demonstration, called the ARKola experiment, the auditory icons were designed to provide information that was not visible in the interface [Gaver W.W. Smith R.B. and O’Shea T. (1991)]. Participants collaborated in pairs to control a simulated cola bottling factory made up of 9 interconnected machines, such as a heater and a bottler, with on/off and rate controls. Each participant could only see half of the factory, but could hear a characteristic sound for each machine, which varied in a predictable manner with the rate of operation. Each sound was designed to have a semantic association with the machine it represented, for example the heater made a whooshing sound like a blowtorch, and the bottle dispenser sounded like bottles clattering along a conveyer belt. The participants were able to quickly learn and remember the meaning of the auditory icons in the process of using them. The sounds helped them to track ongoing processes, and to monitor individual machines as well as the overall condition of the factory. The participants could refer to the sounds to discuss unseen elements, and the sounds increased the enjoyment of the activity. Some problems in the discrimination of overlapping sounds were observed, which Gaver et al. suggest could be addressed by a more systematic approach to shaping the psychoacoustics of the sounds.

The auditory icons in the ARKola experiment were able to signify amounts, such as rate or quantity, directly and without the need to refer to a legend. This is an important demonstration that sounds can convey quantitative information, and support decision making based on quantitative data. The fact that the participants learnt the meaning of the sounds very quickly seems to indicate that the semantic design method is intuitive to understand, as Gaver suggested. However the theory that auditory icons are more learnable than other sonic signs has not been supported by empirical investigations. Lucas found no significant difference between auditory icons and earcons in the amount of time taken to learn associated meanings, in the number of errors made in the process of learning, or in the improvement in these factors over two trials. It took significantly less time to learn speech associations, and these were consistently error free. The factor that most influenced the accuracy of recognition of the auditory icons and earcons was an explanation of the rationale behind the sound design [Lucas P.A (1994)].

This result does not accord with the ecological theory, but rather indicates that intermediate mental structures do have a role in listening. This observation is supported by Ballass' experiment which found that the speed and accuracy of identification of the source of a sound depends on the listeners expectations, context and experience. The experiment tested the ambiguity of the relation between a sound and its perceived source by embedding subjectively similar sounds, such as a fuse burning and leaves crackling, in contextual sequences which were biased toward one or the other interpretation. An inappropriate context had a clear negative effect, although an appropriate context did not seem to have a symmetrical positive effect [Ballass J.A. (1994)].

Truax comments that if an audio signal is perceived or measured to have "fidelity" to the original source it is thought to have been successfully reproduced - but this assumption causes a schizophrenic fracture between the original source and its later out-of-context usage [Truax B. (1992a)]. Beck describes how most of the sounds in movies, radio and theatre are not sampled directly from the real world but are deliberately constructed to foster an impression in the mind of the listener, through expectations and cultural experience, as well as physical experience [Beck M. (1996)]. The subjective nature of auditory experience was highlighted by the inconsistency that Mynatt found in her experiment on associations between sampled sounds and interface concepts. [Mynatt E.D. (1994)]. The experiment tested associations between 28 sounds, such as a zipper, and 15 interface concepts, such as a pull-down menu, from a sound to a concept and from a concept to a sound. The same sound was chosen for many different concepts, and there was rarely any convergence on a single sound for a single concept. Mynatt concluded that the design of sounds for the human-computer interface is difficult and painstaking process which depends on the skill of gifted individuals. She summarised her findings as guidelines that may assist other designers

- *Choose short sounds that have a wide bandwidth, and where length, intensity and sound quality are roughly equal.*
- *Evaluate the identifiability of the auditory cues using free form answers.*
- *Evaluate the learnability of the auditory cues that are not readily identified.*
- *After deciding on a collection of sounds that are easily identified or learned, evaluate possible uses of the auditory cues in the interface being designed.*
- *Evaluate possible sets of auditory icons for potential problems with masking, discriminability and conflicting mappings.*
- *Finally the complete interface should undergo usability evaluations with subjects or*

users conducting specified tasks, as well as long-term evaluations that monitor the performance over a long period of use.

The studies of learnability and identification with auditory icons suggest that what is signified depends on listening experience and context. Although the method is based on a theory of innate perception that should ensure consistent interpretation the experimental results show significant individual differences.

2.4 Pragmatic approach

The pragmatic approach emphasises the form of the signifier. The set of signifiers in a lexicon need to be discriminably different to represent different signifieds, and to prevent ambiguous combinations.

Early work on the pragmatics of sounds was prompted by the need to ensure that a pilot could hear warnings and alarms against a background of engine noise and radio conversations. A study of sounds for aircraft display systems was carried out by Patterson. Although his guidelines were intended for sounds in cockpits, they may be helpful in computer interfaces too [Patterson R.D. (1982)].

- *Overall level - The lowest intensity level of a warning sound should be 15dB above the threshold imposed by background noise. The upper limit is 25dB above threshold.*
- *Temporal - Component pulses of a warning sound should have onsets and offsets 20-30 ms in duration. These avoid a startle response in the listener. Pulses should be 100-150 ms in duration with a 150 ms inter-pulse gap for urgent sounds and a 300 ms gap for non-urgent sounds. Distinctive rhythms of 5 or more pulses should be used.*
- *Spectral - The fundamental frequency of a warning should be within the range 150-1000 Hz. There should be four or more component harmonics to help avoid maskings. The overall spectral range of warnings should be 500-5000 Hz.*
- *Ergonomics - Manual volume control should be avoided and automatic control restricted to a range of 10-15 dB variation. There should be no more than six immediate action warnings.*
- *Voice warnings - These should be brief and use a keyword format. They should not be repeated in a background version of the warning. Voice warnings used as immediate awareness warnings should use a full-phrase format and be repeated after a short pause.*

In another study of auditory displays for aircraft, Deatherage not only investigated the design of sounds as alarms but also as aids for spatial information and orientation [Deatherage B.H. (1972)].

Alarm and warning signals:

- *At a minimum, use sounds having frequencies between 200 and 5000 Hz, and if possible, between 500 and 3000 Hz, because the human ear is most sensitive to this middle range.*
- *Use sounds having frequencies below 1000 Hz when signals must travel long dis-*

tances (over 1000 ft.) because high frequencies are absorbed in passage and hence cannot travel as far.

- *Use frequencies below 500 Hz when signals must bend around obstacles or pass through partitions.*
- *In noise, signal frequencies different from those most intense frequencies of the noise are best in order to reduce masking of the signal.*
- *Use a modulated signal to demand attention. Intermittent beeps repeated at rates of one to eight beeps per second, or warbling sounds that rise and fall in pitch are seldom encountered, and are therefore different enough to get immediate attention. If speech is necessary during an alarm, use an intermittent, pure-tone signal of relatively high frequency.*
- *Use complex tones rather than pure sinusoidal waves, because few pure tones can be positively identified but each complex sound is noticeably different from other sounds.*

Spatial information:

- *Use auditory displays to relieve the eyes. Although the eye is better for spatial discrimination, it can look in only one direction at a time. In general, auditory spatial displays are recommended only when the eyes are fully engaged and additional spatial information is needed.*
- *Use auditory displays (other than speech) to present restricted information, such as the following:*
  (a) "Yes-no" information and indications of amount or degree. Auditory displays can represent error or deviation from a course, speed, attitude, or other "normal" condition.
  (b) Continuous information. For example, radio-range signals present practically continuous information about one kind of event - the course the aircraft is flying.
  (c) Automatic information - recorded word signals as from an automatic annunciator.
- *Use auditory displays of tonal or noise signals when speech channels are already fully employed. Most of the auditory displays that utilize tonal signals can be heard through speech, and, conversely, speech can be understood while hearing the tonal signals over the same receiving system.*

Spatial orientation:

- *Confine representation to a single dimension; multidimensional displays are less effective than their visual counterparts.*
- *Provide a standard stimulus to represent the "normal" then make abrupt changes to indicate departures from the normal. Human listeners are sensitive to frequency or intensity changes but poor at identifying a unique signal.*
- *Provide changes in intensity rather than frequency as a spatial cue. Because everyone with normal hearing can detect changes in intensity, it is easier to control these changes.*
- *Use intermittent to repeated changes in a signal rather than a single change followed by a continuous signal. The ear is much more likely to detect changes in a signal*

occurring every second or two than at longer intervals.

- *If absolute identification is required, limit the number of signal categories to four because listeners cannot identify correctly more than a few different intensities, pitches, or interruption rates.*
- *The following "natural" relationships between auditory signals and the dimensions they represent are quickly learned or are perceived with little training:*
  (a) Binaural intensity differences serve to localise (in bearing) the direction of a sound.
  (b) Pitch differences naturally represent up and down (high and low pitch). To indicate climb or "upward pointing" raise the pitch. Combined with binaural changes in pitch from one ear to the other, "left wing high" for instance, can be represented.
  (c) A slow interruption rate is a natural indication of speed - an increase or decrease in interruption rate is immediately perceived as a change in speed (or rate) of interruption.

These recommendations were brought together with other observations, by McCormick and Sanders, to provide some general guidelines that extend to other types of auditory displays outside the cockpit [McCormick E.J. and Sanders M.S. (1983)]. They begin with a list of circumstances where an auditory display is preferable to a visual display:

- *When the origin of the signal is itself a sound.*
- *When the message is simple and short.*
- *When the message will not be referred to later.*
- *When the message refers to events in time.*
- *When sending warnings or when the message calls for immediate action.*
- *When presenting continuously changing information of some type, such as aircraft, radio range, or flight path information.*
- *When the visual system is overburdened.*
- *When speech channels are fully employed (in which case auditory signals such as tones should be clearly detectable from the speech).*
- *When illumination limits vision.*
- *When the receiver moves from one place to another.*

General principles

- *Compatibility - where feasible, the selection of signal dimensions and their encoding should exploit learned or natural relationships of the users, such as high frequencies associated with up or high, and wailing signals with emergency.*
- *Approximation - two-stage signals should be considered when complex information is to be presented, these stages consist of 1) Attention-demanding signal: to attract attention and identify a general category of information 2) Designation signal: to follow the attention demanding signal and designate the precise information within the general class indicated above.*
- *Dissociability - auditory signals should be easily discernible from any ongoing audio*

input (be it meaningful input or noise). For example if a person is to listen to two or more channels the frequencies of the channels should be different if it is possible to make them so.

- *Parsimony - Input signals to the operator should not provide more information than is necessary.*
- *Invariance - The same signal should designate the same information at all times.*

Principles of presentation

- *Avoid extremes of auditory dimensions - high intensity signals, for example, can cause a startle response and actually disrupt performance.*
- *Establish intensity relative to the ambient noise level - this is simply saying that the intensity level should be set so that it is not masked by the ambient noise level.*
- *Use interrupted or variable signals - where feasible, avoid steady-state signals and, rather, use interrupted or variable signals. This will tend to minimise perceptual adaption.*
- *Don't overload the auditory channel - only a few displays should be used in any given situation. Too many displays can be confusing and will overload the operator. (For example, during the Three Mile Island nuclear crisis, over 60 different auditory warning displays were activated).*

Principles of installation of auditory displays

- *Test signals to be used - such tests should be made with a representative sample of the potential user population, to be sure the signals can be detected by them.*
- *Avoid conflict with previously used signals - any newly installed signals should not be contradictory in meaning to any somewhat similar signals used in existing or earlier systems.*
- *Facilitate change-over from previous display - where auditory signals replace some other mode of presentation (e.g. visual), preferably continue both modes for a while, to help people become accustomed to the new auditory signals.*

The pragmatic design of sounds to support human-computer interaction was investigated by Brewster, Wright and Edwards who sought to answer the question “what sounds should be used at the user interface?” Their empirical observations are summarised as guidelines for designing earcons, shown in Table 2-1 [Brewster S.A. Wright P.C. and Edwards A.D.N. (1994)].

Timbre	*Use musical instrument timbres. Where possible use timbres with multiple harmonics. This helps perception and avoids masking. Timbres should be used that are subjectively easy to tell apart e.g. use "brass" and "organ" rather than "brass1" and "brass2".*
Pitch	*Do not use pitch on its own unless there are very big differences between those used (see register below). Complex intra-earcon pitch structures are effective in differentiating earcons if used along with rhythm. Some suggested ranges for pitch are: maximum 5 kHz (four octaves above middle C) and minimum 125 Hz to 150 Hz (an octave below middle C).*
Register	*If this alone is to be used to differentiate earcons which are otherwise the same, then large differences should be used. Two or three octaves difference give good rates of recognition.*
Rhythm	*Make them as different as possible. Putting different numbers of notes in each rhythm was very effective. Patterson says that sounds are likely to be confused if rhythms are similar even if there are large spectral differences. Small note lengths might not be noticed so do not use notes less than eighth notes or quavers. In the experiments described here these lasted 0.125 seconds.*
Intensity	*Although intensity was not examined in this test, some suggested ranges from Patterson are: maximum 20dB above threshold and minimum 10 dB above threshold. Care should be taken in the use of intensity. The overall sound level will be under the control of the user of the system. Earcons should all be kept within a close range so that if the user changes the volume of the system, no sound will be lost.*
Combinations	*When playing earcons one after another, use a gap between them so that users can tell where one finishes and the other starts. A delay of 0.1 seconds is adequate. If the above guidelines are followed for each of the earcons to be combined, then the recognition rates should be sufficient.*

Table 2-1: Guidelines for the pragmatic design of earcons

Auditory representation techniques have also been evaluated from a pragmatic perspective. A system for psychometric evaluation of auditory displays was implemented by Smith, Levkowitz, Pickett, and Torpey [Smith S, Levkowitz H., Pickett R.M. and Torpey M. (1994)]. They classify perceptual discrimination into 3 types

- Detection - the discernment of a shift (or just noticeable difference)
- Recognition - the discernment of a shift and discrimination between different kinds of shifts
- Scaling - the discernment, discrimination and characterisation of a shift

The threshold of shift detection for an auditory parameter is measured by a three-alternative, forced-choice, up/down technique. Evaluations are made against a statistically parameterised data set, in which the mean, standard deviation and other features can be systematically controlled. The system was demonstrated in an investigation of an asynchronous granular synthesis (AGS) algorithm with ten parameters. The results show complex interactions between the parameters, perceptual non-linearities in response to linear changes in the data, and wide variation in the dynamic range of various perceptual effects. Smith et al. concluded that their AGS algorithm is less than ideal for auditory data representation. Some of the issues they raise are that sounds that are interesting and lively in musical and artistic compositions can become irritating when presented repeatedly in an experiment. Sound synthesis algorithms have idiosyncratic parameters that are not aligned with perceptual variations, and which cannot be generalised. They comment that the development of a device-independent and easy to use way to specify timbres is critical before auditory data representation can make a further leap forward.

The types of discriminations measured by Smith et al. are called “capabilities” by Watson and Kidd. They propose a principle of “proclivity” to explain their findings that performance on very familiar sets of stimuli may not be predictable from performance on the discrimination of unfamiliar sounds [Watson C.S. and Kidd G.R. (1994)]. Proclivities reflect responses that cannot be scored as right or wrong, such as judgements of sound quality, of the value of a sound on a perceptual scale, and of the similarity of a sound to another sound. Psychophysical measures of acuity are not sufficient to explain the large individual differences in performance, and central “cognitive” factors of learning, memory and attention may have to be considered if we are to predict the human operator’s performance with arrays of auditory signals. Their observations are summarised by some guidelines

- *Auditory capabilities determined under minimal uncertainty are both valid descriptions and misleading guides.*
- *Auditory capabilities determined under high levels of stimulus uncertainty or under “ecologically valid” conditions can be both invalid descriptions and useful guides.*
- *Performance on auditory identification tasks can be greatly improved through intense training, but pilots, cardiologists, and nuclear power station operators aren't likely to devote the required hours in learning an arbitrary auditory catalogue of more than 7 or 8 discriminable sounds.*
- *One way to convey more information is to use the auditory signal only to indicate (a) that a message has arrived and (b) the priority of that message. The actual message then can be conveyed by voice or by alphanumeric display.*
- *Establishing larger catalogues of auditory signals may be possible, but may benefit from:*
  (a) exploiting pre-existing sound-event associations,
  (b) designing sounds with maximal pairwise discriminability, based on psychoacoustic principles,
  (c) establishing a grammar, or other form of sound organisation.

The perception of overlapping auditory elements has been explained by Bregman in his theory of auditory scene analysis [Bregman A.S. (1990)]. This theory proposes that acoustic elements are grouped into “streams” according to heuristics such as similarity, proximity, closure and familiarity. The heuristics help to predict the outcomes of a collaboration and competition between two auditory processes:

- Primitive grouping - fast, innate grouping by acoustic properties of the proximal stimuli.
- Schema segregation - slower, conscious selection of elements from groups that have been formed by the primitive process, and active restructuring of groups by effort of will.

The primitive level explains auditory illusions and why sounds can be hidden or camouflaged due to the acoustic interactions. The schema level explains why the motivation and experience of the listener can make such a difference to what they hear. Some properties of streams that may be relevant in display design are [Bregman A.S. (1990)]

- Only material grouped into the same stream can obscure or camouflage a target.
- Simple tasks such as counting the number of tones are more accurate if the tones are in the same stream.
- Temporal relations are difficult to make across streams - for example it is very difficult to judge the order of elements in separate streams, or compare the rates of cyclic sequences which have segregated.
- An element in a stream may be captured by another stream with elements that are similar
- A rhythm tends to be defined by sounds that fall in the same stream. The allocation of sounds to different streams affects what rhythms may be heard

Williams says that a knowledge of the potential auditory streams that may arise from a particular acoustic signal is essential in order to predict the possible interpretations of that signal, and lists gestalt heuristics as principles for auditory display design as shown in Table 2-2 [Williams S.M. (1994)]. This approach may help in the design of concurrent and overlapping sounds. However it doesn't tell us how to map data relations to auditory relations.

Similarity	*components which share attributes are perceived as related.*
Proximity	*the closer two components are, the more likely that they belong together.*
Good continuation	*components that display smooth transitions from one state to another are perceived as related.*
Familiarity	*recognition of well-known configurations among possible subcomponents leads to these subcomponents being grouped together.*
Belonginess	*a component can only form part of one object at a time and its percept is relative to the rest of the figure-ground organisation to which it belongs.*
Common fate	*components which experience the same kinds of changes at the same time are perceived as related.*
Closure	*incomplete figures tend to be completed.*
Stability	*having achieved an interpretation that interpretation will remain fixed throughout slowly changing parameters until no longer appropriate.*
Articulation	*the separation of a figure from ground requires energy.*

Table 2-2: Gestalt heuristics

Psychoacoustic observations describe the relation between acoustic variations and what is heard by a listener. This pragmatic method works well for simple displays by ensuring the discrimination of categories and linearity of scales, suggesting that psychoacoustic measurements and theories can assist in the design of an auditory display. However complex multidimensional displays are not easy to measure, and the predictions made from these measurements are not always valid outside the laboratory situation. Perhaps the biggest problem with this approach is that psychoacoustics does not provide guidance about the relation between perception and information.

2.5 Perceptual approach

Graphs show information in sets of abstract numbers. Graphic information is not contained in individual signs, but in the perceptual relations between signifiers. Bertin proposed that graphic relations were a new form of semiology that is distinctly different from other sign systems in the way the signified is perceived [Bertin J. (1981)].

- *To perceive a pictograph, a road sign for example, requires a single stage of perception: what does the sign signify? Stop! All the useful information is perceived. The aim of pictography is to define a set or concept.*
- *To perceive a graphic requires two stages of perception: 1st: What are the elements in question? 2nd: What are the relationships among those elements?*

The signifieds in a graphic are resemblance, order and proportion, transcribed by visual variables that have signifying characteristics shown in Table 2-3.

Perceptual Variable	*Quantitative*	*Ordered*	*Differential*	*Visibility*
X,Y spatial dimensions	*quantitative*	*ordered*	*selective*	*constant (associative)*
size	*quantitative*	*ordered*	*selective*	*variable (dissociative)*
value (lightness)		*ordered*	*selective*	*variable (dissociative)*
texture		*ordered to some extent*	*selective*	*constant (associative)*
colour			*selective*	*constant (associative)*
orientation			*selective*	*constant (associative)*
shape				*constant (associative)*

Table 2-3: Bertin’s signifying properties of the visual variables

A representation is designed by mapping data relations onto visual relations with similar characteristics. Quantitative visual relations signify quantitative data relations, qualitative visual relations signify qualitative data relations. The point at which a visual variable disappears signifies the zero in a data variable. The effectiveness of Bertin’s scheme for visually representing quantitative data was empirically validated by Cleveland [Cleveland W.S. (1985)]. This perceptual scheme is the basis for MacKinlay’s rule-based tool which can automatically generate a graph from a characterisation of the data to be represented [MacKinlay J. (1986)].

At the ICAD’92 conference Bly pointed out that progress in auditory display is impeded by the lack of a principled approach similar to that which has been developed for graphic representation [Bly S. (1994)]. She showed that different auditory representations are not equally effective by challenging researchers to design a display to support the classification of a 6 dimensional multivariate data. The results for the 3 displays that were evaluated ranged from chance to significant. The variation in the results led Bly to conclude that a knowledge of the data structure is crucial for the design of an effective auditory display. This observation is echoed by Scaletti's suggestions of appropriate synthesis techniques based on data characteristics as shown in Table 2-4. These suggestions capture experience and expert knowledge in the design of displays to represent these types of data, and are helpful for other designers faced with similar data.

Data Characteristics	*Suggested synthesis techniques*
Oscillating between states	*timbral interpolation or morphing*
Axes and grids	*resonators, fixed tones*
Comparison	*sums, products, differences, correlation*
Textures and tendencies	*granular synthesis, FM, waveshaping, sonic histogram*
Periodicity detection	*data as samples, autocorrelation*
Virtual objects in VR space	*physical models, sampled sounds*
Data with an attitude	*instrumental sounds, musical scales, sampled sounds*

Table 2-4: Scaletti's suggested synthesis techniques

Hayward similarly characterises data in the seismic domain, as shown in Table 2-5 and Table 2-6. As well as characterising the data and the interpretation task, Hayward also provides a rationale for applying the auralisation technique to seismic data.

Sample rates	*250 to 4000 Hz*
Bandwidth	*usually less than 3 octaves* *10-100 Hz common* *20-500 Hz high resolution* *20-2000 Hz special studies*
Dataset size	*12,000 to several billion samples* *0.2 to 5 seconds/record* *usually < 4000 samples/record* *multiple shots per reflection point*
Interpretation	*Every peak is significant and waveform details (width or phase) are often also interpreted*
Dynamic range	*8 - 24 bits commonly used*

Table 2-5: Characteristics of exploration seismic signals

Sample rates	*1 to 120 Hz* *multiple sample rate streams are common*
Bandwidth	*1 hour period to 40 Hz (17 octaves)*
Dataset size	*Continuous recordings 3-100 Megabytes/day* *Triggered recordings < 1 Megabytes/day*
Interpretation	*Precise timing and identification of phase (P,S etc.) is critical. Much interpretation is based on the envelope and spectra of the event*
Dynamic range	*24 bit recording is common* *Most signals easily represented by 16 bits*

Table 2-6: Characteristics of planetary seismic signals

Hayward says that the direct mapping (auralisation) of seismic data works well because the data is physically constrained by the elastic wave equation which is common to both acoustic and seismic vibration. The resemblance between the acoustic signifier and seismic signified is a basis for everyday interpretations - for example a series of echoes followed by an explosion is recognised as physically ridiculous. Loudness is related to energy, pitch implies harmonic structure, and intervals between echoes infer the size of an echo chamber. He comments that auralisation of arbitrary time-series data, such as stock prices, is often unsuccessful because the data relations are physically unconstrained and generate acoustics that are not matched to the perceptual capabilities and expectations of the listener. This is a similar argument to that given by Gaver for the intuitiveness of auditory icons.

The systematic acoustic relation between the signifier and the signified is clearly a useful way to design sounds, but it doesn't help us to signify symbolic relations in abstract data, like stock prices. The effectiveness of a simple pitch variation over time was evaluated as a means to represent abstract data relations by Flowers, Buhman and Turnage. They found that normal users could quickly and easily understand the basic properties of simple functions, distribution properties of data samples, and patterns of covariation between two variables, and that the auditory display was as effective as a visual graph [Flowers J.H. Buhman D.C. and Turnage K.D. (1996)]. This is an important starting point for a method of signifying properties of auditory elements modelled on Bertin's signifying properties of visual marks. Kendall has suggested such a method when he observed that auditory representations of categorical data should sound categorical, that continuous data should sound continuous, and that uniform steps along the continuum should sound uniform [Kendall G.S. (1991)]. This method would focus on a faithful mapping between auditory relations and data relations. A step in this direction has been taken by Madhyastha and Reed, who proposed a matching of the importance of a data variable to the perceptual weighting of an acoustic variable heard in the display. This method was designed to handle the problem that extreme values of some acoustic parameters can cause others to lose their effect - for example it can be difficult to hear the pitch of a short sound, or the timbre of a high pitched sound [Madhyastha T.M and Reed D.A. (1992)]. The perceptual ranking is informal, i.e. "pitch and rhythm are the most distinguishing characteristics of a melody, and thus, can be considered more significant than, say, volume", but it is an example of an approach to auditory design based on perceptual signification of abstract relations.

The matching of perceptual properties of sounds with the properties of the information to be conveyed may improve the directness and correctness of a display. However the display is a nuisance if the information is not useful. The viewer can simply not look at a useless visual display, but auditory displays are not so easy to ignore and prone to become annoying when the information in them is not useful. The auditory display may be perceived as an annoying noise if the information in it is not useful.

2.6 Task-oriented approach

Semiotic methods have been criticised because they are static in time, and do not include motivation as a factor in either the construction or reading of signs. Wittgenstein observed that meaning of a sign is not the object it signifies, but the way it is used [Wittgenstein L. (1953)]. Signs can be used in many ways to generate families of meanings. Signs can be useful without having a material signified at all - for example the greeting "hello".

Usefulness is one of the criteria often used in HCI to evaluate an interface design. Casner says that different presentations of the same information best support different tasks, and that the usefulness of any information presentation is a function of the task it is being used to support. He built an automatic tool to deduce a graph for a specific task and data [Casner S. (1992)]. Kramer suggests that two broad types of tasks are important in auditory display [Kramer G. (ed) (1994b)]

- *Analysis - tasks where the user cannot anticipate what will be heard and is listening for "pop-out" effects, patterns, similarities and anomalies which indicate structural features and interesting relationships in the data.*
- *Monitoring - a "listening search" for familiar patterns in a limited and unambiguous set of sounds.*

The need to design sounds that provide information relevant to a task was recognised by Frysinger in his proposal of a taxonomy of tasks and data types as a foundation for choosing auditory representations [Frysinger S.P. (1990)]. He suggests that the effectiveness of different auditory display techniques could be evaluated against standard tasks and sets of data. However the definition of a corpus of tasks and parameterised data sets is not trivial. Frysinger points out that the identification of task types is difficult because often the analyst is not able to describe what they are doing very precisely, and a task may consist of a combination or compound of simpler tasks. Sometimes a task is considered to be a small closed action, like pushing a button, other times it can be something bigger, like filling in a form, or something more complex like analysing trends in a data set. Storytelling has been suggested as a way to describe a task by Erickson, who uses a collection of stories to come to grips with the user requirements of an interface [Erickson T. (1996)]. The users are asked to tell stories about their activities which are not expected to be formal or complete or even particularly accurate. A collection of stories contain information about what the users like, and dislike, what works well, what users who are expert in the activity think will work well, and concrete explanations of real problems. Stories are informal, unconstrained and easy to remember. The stories of different users may overlap and provide snapshots of an activity from different perspectives and in different situations. The collection is a rich source of contextual, experiential and concrete knowledge about the problem that can provide a basis for more formal analyses.

A semi-formal method for graphical user interface (GUI) design was extended by Brewster to design useful sounds for the human-computer interface. The method is called the Event, Status, Mode method (ESM). An event marks something that happens at a discrete point in time, due to some action by the user (e.g. a mouse click) or the system (mail arriving). Events depend on context, for example clicking the mouse in one window may select an icon, whilst in another it may position a cursor. Events can be hidden because the system does not display them, or because the user does not perceive them. Status information is similar to an event, except that it is persistent, for example the location of a mouse cursor. Status information can become hidden even if it is visually present, due to visual fixation on other parts of the screen, and because it is static and may fade from attention. A mode is a system context that alters the interpretation placed on events. In one mode typed characters may appear on the screen, whilst in another they may be interpreted as commands. Mode errors occur when the status information is hidden. Information about the events, status and modes in an interaction is characterised by 4 dimensions of feedback.

Information	*Description*	*Sound*
Action dependent/ independent	*Does the feedback depend on a user or system action? Events are action dependent, status and modes are action independent.*	*A keypress activated beep is an action dependent sound.* *A constant tone indicating mode is action independent sound.*
Transient/ sustained	*Is the feedback sustained throughout a particular mode? Events are transient, status is sustained, and modes may be either.*	*A short beep to indicate an error is a transient sound* *Sustained sounds can be habituated, and will be perceived only when it changes in some way, or by conscious attention.*
Demanding/ avoidable	*Can the user avoid perceiving the feedback? Events and modes should be demanding.*	*Sound is attention grabbing so is good for demanding feedback, whereas graphic displays are often missed. Care must be taken in designing avoidable sounds so that they are not demanding by mistake.*
Static/ dynamic	*Does the feedback change whilst it is presented or is it constant? Events are static, status can be static or dynamic.*	*A constant tone is static, music is dynamic*

Table 2-7: Brewster's ESM method

The method begins with the identification of all the modes that are present in an interaction. The events and status information required in that mode are listed and compared with the event and status information that is available in the interface. Discrepancies identify the missing or hidden information that may be causing errors and reducing task performance. The characteristics of the required feedback are then represented by sounds with appropriate characteristics. Brewster demonstrates the method by analysing 7 interface widgets, including a caps-lock key, dialogue boxes, buttons, menus, scrollbars and windows. In every case the source of errors was identified as the avoidability of status information about modes. From this analysis we can surmise that the most useful way that sounds can provide information in an interface is by alerting the user to transient events that may be missed due to visual attention being elsewhere at the time. Brewster tested the method by designing some earcons for the scrollbars, buttons and windows, and evaluating user performance with and without the sounds switched on. He found that the earcons decreased the amount of time to carry out a task, and the amount of time taken to recover from errors. The users preferred the sound enhanced interfaces over visual-only interfaces, and Brewster comments that this was because the sounds provided information that was needed, they were not gimmicks.

Alty lists reasons why multimedia interfaces may be advantageous for control applications [Alty J. (1995)]

- *Telepresence - Multimedia options allow us to regain the natural link between the operator and observables. Avoids events by proxy. Preserves implicit cues.*
- *Measurable Media Differences - Match the medium carrying capabilities with the knowledge output requirements. Differences must be discernable.*
- *Goal - Main factor in choice of a medium is how the presented information will be used. Need a taxonomy for characterising knowledgeNeed a characterisation of Media and their knowledge carrying capabilities. Need to match the two.*
- *Complexity - The medium matters in more complex situations.*
- *Redundancy - For humans more is better. Humans prefer redundancy of information. Multimedia interfaces exploit the whole mind. Useful when information quality deteriorates.*
- *Operator Choice - Operators have hidden goals and implicit knowledge. Often they do not know what information they use. Give them an element of choice. Feedback gives useful pointers to media usage.*
- *Intrusion - Some media such as sound are intrusive. Such media are often useful as context switchers. Intrusive media can obliterate lower level goals, so the system should remember "interrupted" goals and remind operators when the intrusion is over. Media switches can also aid problem solving.*
- *Metaphor - Presenting information in a different medium can be illuminating i.e. music is a visual medium (height and distance metaphors for notes and time).*
- *Synchronisation - Media when used together and synchronised are very powerful. Minimising overload on short term memory.*

Although these seem like very good reasons to use multimedia, Alty comments that in fact there is little empirical evidence to support the view that multimedia interfaces are any better. He tested the validity of these principles in an experiment which compared the performance of subjects using different display media. The task was to control Crossman's Waterbath which involves balancing in-flow, out-flow and heater temperature to prevent error conditions. Information about flowrates, temperature, and water level were shown by combinations of graphics, text, speech, and sound. The results show that differences in the display media do influence task performance, though it is hard to know why or how. Alty observes that the different media have different syntactic, pragmatic and semantic properties, and some are better suited to represent different types of knowledge. He concludes that is it difficult to produce good interfaces by chance or ad-hoc techniques, and that a formal method of multimedia design is necessary.

The approach he proposes has the premise that the goals of the user should determine what information is required and how it should be rendered, and is summed up by a series of questions

- *what is the goal?*
- *what task is needed to achieve it?*
- *what knowledge is required?*
- *how is the knowledge characterised?*

The goal of the user is described in a process flow diagram that allows flexible connections and iterations between tasks. There are 4 types of tasks, shown in Table 2-8.

Task type	*Description*
monitoring	*checking on critical variables to spot deviations as soon as possible, in order to maintain optimal running conditions and plant safety. Activities include Identify, Search, Browse, Instantiate, Scan, Check.*
diagnosing	*identifying causes of deviations so that the plant conditions can be stabilised. Identify, Compare, Derive, Guess, Reason.*
predicting	*identifying potential consequences of plant deviations in order to prevent them. Model, Simulate, Run.*
controlling	*direct impact on the operations of the processing system. Record, Create, Delete, Edit, Alter, Enter, Move, Load, Save.*

Table 2-8: Task types

The knowledge required by each task is characterised in terms of 3 kinds of variables - primitive, derived, or complex. A derived variable is a new variable calculated from a primitive variable(s). Complex variables relate a variable with an organisation in space or time. For example a set of Temperature Variables in Time (i.e. Temperature history) is described as {(Temperature, Time)}. Most variables are complex. The description of a variable is shown in Table 2-9.

Variable	*Description*	*Example*
Name	*Conventional name for the variable*	*Temperature*
Type	*Nominal, Ordinal or Quantitative*	*Quantitative*
Cardinality	*Single-Values, Fixed-Multiple Valued or Variable-Multiple-Valued*	*Single-Valued*
Accuracy	*The accuracy of representation*	*0.01*
Range	*The range of possible values*	*-10 to 300*
Ordering	*Does the variable have an ordering*	*Ascending*
Units	*The units of measurement*	*Celsius*

Table 2-9: Knowledge characterisation

Variable	*Description*	*Example*
Stability	*Static or Dynamic*	*Dynamic/every 5 secs*
Continuity	*Continuous or Discrete*	*Continuous*
Directionality	*Scalar or Vector*	*Scalar*
Derived from	*Variables list*	*[]*
Derivation	*How derived*	*Primitive*

Table 2-9: Knowledge characterisation

The Knowledge characterisation is matched against a characterisation of a display medium, shown in Table 2-10. This characterisation is similar to Bertin's visual variables, but the concrete characterisation for auditory media is not described.

Medium	*Description*	*Values*
Name	*Conventional name for the variable*	
Number	*the variables involved*	
For Each Variable		
	Type	*Nominal, Ordinal, Quantitative*
	Ordering	*Does the variable have an ordering*
	Continuity	*Continuous or Discrete*
Processing	*Possible processing options*	
Carrier Details		
Resource Needs	*Audio or Visual +formula for Resource requirement*	

Table 2-10: Media characterisation

The method was applied to design a multimedia control system for a nuclear power plant. The operators were asked to try out the new multimedia display for a trial period, and were asked to evaluate the display. They generally agreed that the sounds made handling alarms quicker and easier, and helped avoid mistakes related to the analysis of alarms. However the sound was disliked as there were too many sounds in the plant. Alty comments that the impact of the multimedia interface was not as great as it could have been. All operators found the system complex and difficult to use, and needed much more training to make full use of the system. One found it fun to experiment with the new opportunities. Another found the new system too tedious to work with. A number of operators drew attention to the poor quality of the voice output.

Task-oriented design methods help to focus the design on a display that is useful. This is particularly important in auditory display because the sounds cannot be easily ignored. However the usefulness of the sound only becomes apparent with use. The first impression that a listener has of an auditory display is aesthetic, and the usefulness of the display can only become evident if it is used.

2.7 Connotation approach

Most people are used to hearing high quality sounds in music CD’s, movie soundtracks and computer games. Composers and sound designers are concerned with aesthetic, stylistic and affective connotations of the sounds. These connotations reflect the value that society places on both the signifier and the signified in a sign. The connotations of the sounds in an auditory display are likely to influence how well it is received by users, and may be especially important in commercial applications. Positive connotations may encourage users to experiment and learn the display. An example of how sounds can influence credibility is provided by Tkaczevski’s sound design to imply an air of fraud and humour in a computer game [Tkaczevski A. (1996)]. The humorous connotations that arise from the juxtaposition of high tech graphic objects with low tech sounds is surely a lesson to auditory display designers about the importance of auditory connotations! The need to satisfy expectations of audio quality is described by Dougherty in his account of the sound design for the Taligent Operating System [Dougherty T. (1996)]. Engineers on the project had experienced the 8 bit sound quality of the SonicFinder demonstration (which was constrained by the 8 bit sound card that was available at the time), and were concerned that the sounds should be of compact disc quality, and should be very low volume. Dougherty shaped high quality sampled sounds so they could be discriminated at low volumes. He found that people reacted more favourably to the interface when the sounds were switched on, and that they were often influenced by the sounds even when they said they hadn’t noticed them.

2.8 Device approach

A major problem in audio applications is the variation in the characteristics of audio devices. These characteristics include the parameters for adjusting overall volume, equalisation, stereo position, reverberation etc., and the ranges of these parameters. Frequency response, speaker locations, ambient noise and many other factors can influence the display design. Tkaczevski describes an approach for balancing the sounds in computer games across various FM and Wave Table synthesisers on different devices and operating systems. He had to manually edit and evaluate the global controls on each target system, tweaking them to reach a compromise that worked across the board. Although a compromise is a practical solution, it can only work when the designer knows what sounds will be heard and when. It also compromises the capabilities of the better quality devices. These problems have been addressed by the Porsonify toolkit developed by Madhyastha and Reed to support auditory displays [Madhyastha T.M and Reed D.A. (1992)]. Porsonify provides a uniform network interface to sound devices through table driven servers that encapsulate device specific parameters without hiding unique hardware functions. A sonification client can query a device daemon to obtain a description of available parameters and their ranges. This description can be used to automatically constrain the mapping to sounds that can be realised on that device. As part of this work, Madhyastha and Reed raise the need to introduce a data characterisation, and rules for mapping data into sounds in a device-sensitive manner. However this approach does not address the problem that different devices with the same interfaces produce perceptually different sounds, or Smith’s comments that the need for a device-independent method for specifying timbre is a major impediment to progress in auditory display [Smith S. (1990)].

Transportability has also been an issue in satellite imagery and visualisation of data. Different devices produce colours in different ways, and an image that looks good on one device can look terrible on another, even similar, device. The need for a transportable way to specify colours for these applications was addressed by Robertson using the device-independent, perceptually uniform CIE colour system [Robertson P.K. and O'Callaghan J.F. (1986)]. The behaviour of a display device was modelled from perceptual measurements at regular points in the device parameter space and could be used to optimise colour sequences for the dynamic range of a device. This approach was developed in an interactive tool for designing colour images to represent data-sets. The designer is able to tailor the colour selection by manipulating graphic widgets overlaid on slices through the colour space that show the limits and behaviour of the display device [Robertson P.K. Hutchins M. Stevenson D. Barrass S. Gunn C. and Smith D. (1994)].

The approach taken in the design of device independent colour displays may also help in the design of transportable auditory displays. Plainly the characteristics of the output device are a major factor to consider in display design.

2.9 Summary

This chapter described previous approaches to the design of sounds to support information processing activities. The approaches were divided into types - syntactic, semantic, pragmatic, perceptual, task-oriented, connotative, and device-oriented. The syntactic approach focuses on the organisation of auditory elements into more complex messages. The semantic method focuses on the metaphorical meaning of the sound. The pragmatic method focuses on the psychoacoustic discrimination of the sounds. The perceptual method focuses on the significance of the relations between the sounds. The task-oriented method designs the sounds for a particular purpose. The connotative method is concerned with the cultural and aesthetic implications of the sound. The device-oriented method focuses on the transportability of the sounds between different devices, and the optimisation of the sounds for a specific device. Clearly there are many ways to go about designing sounds, and many factors to consider. A comprehensive method will need to address issues that include the psychoacoustic properties of the signifier, the perceptual relations between signifiers, the organisation of signifiers into structures, the learnability of the signified, signification through use, connotations and social values, the need for transportability and reproduction.

3 • Designing useful sounds

Ask not the meaning, but the use. [Wittgenstein L. (1953)]

Shirley Robertson: In terms of utilising hearing in the world, when somebody's going through the [blindness] training program, what we really try and look at is focusing on sounds that are useful and filtering out the sounds that are not useful. So sounds that are useful are sounds that can help you maintain a straight line, and help you to orientate so that you know where you are. Sounds that are not useful in that situation, are sounds that are moveable for example. So sounds of people talking, they're useful in terms of knowing that there's somebody in your way, but they're not useful in terms of an orientation clue. [Swan N. (1996)]

This chapter proposes an approach for designing useful sounds. The approach builds on Scaletti's working definition of sonification, which is analysed to have two parts - one part has to do with information requirements and the other with information representations. The requirements part addresses issues of usefulness in a task and the selection of useful data relations to display. The representation part addresses the need to ensure that people can hear the required information in the display. These parts are shaped into a framework that focuses on the design of an information representation to meet the information requirements of a task. The phrase "auditory information design" indicates the focus on useful information which is at the core of this approach.

3.1 Scaletti's definition of sonification

Scaletti proposed a working definition of sonification as ...
a mapping of numerically represented relations in some domain under study to relations in an acoustic domain for the purpose of interpreting, understanding, or communicating relations in the domain under study [Scaletti C. (1994)].

There are three different relations in this definition - numerical relations, acoustic relations, and domain relations. The observation that information is contained in relations between elements, rather than in the elements of themselves, has been made many times. Gibson says that people and animals obtain information about the environment from the perceptions of energy proportions and ratios [Gibson J.J. (1966)]. Bertin says that graphic information is contained in the relations between visual marks that signify resemblance, order and proportions [Bertin J. (1981)]. Sebba found that consistent pairings of paintings and music were influenced by perceptions of contrast, order and ratios [Sebba R. (1991)].

These observations support a substitution of the word "information" for the phrase *relations in the domain under study* in Scaletti's definition. This substitution allows us to separate the definition into 2 parts that are connected by **information** - the requirements part, and the representation part. The requirements part is about specifying information that is useful in some purpose. The representation part is about displaying information as sounds.

3.1.1 The requirements part of the definition

The requirements part of Scaletti's definition is the phrase *for the purpose of interpreting,* *understanding, or communicating relations in the domain under study*. The purpose has been classified into 3 types of information processing activities - interpreting, understanding and communicating. This focus on particular activities can greatly simplify the design problem, particularly when the domain is large or complex. Rather than designing a generic display to represent everything about a domain, the display can be focused on a subset of information that is useful and relevant to the task at hand. From this perspective the requirements part of the definition could be rephrased *to meet the information requirements of an information processing activity*.

3.1.2 The representation part of the definition

The representation part of Scaletti's definition is the phrase *a mapping of numerically represented relations in some domain under study to relations in an acoustic domain*. This is a mapping from one representation to another. Numerically represented relations are contained in measurements of some phenomenon, be it a physical phenomenon such as sea surface temperatures, or an abstract phenomenon like share trading prices. However it may be difficult for a person to understand much about the phenomenon by reading a list of numbers. A perceptual representation, such as a graph or a sonification, is another way to represent information that can allow a person to more quickly obtain a better understanding of what is being represented.

Although the definition outlines the basic process of representation there are issues of faithful perception and faithful reproduction that are not addressed. Numeric relations in data sets are usually classified into 4 types - nominal, ordinal, interval and ratio. A faithful perception of this information requires that the listener be able to hear these types of relations in the auditory display. If the information in the display is to be understood correctly then the mapping is first a mapping to auditory perception, followed by a mapping to acoustic variations produced by a display device. In this view sounds exist only in the mind of the listener, whilst acoustic vibrations exist even when a listener is not present. From these observations we may rephrase the representation part of Scaletti's definition to *a mapping of* ***information*** *to perceptual relations in the acoustic domain*.

3.2 A definition of auditory information design

If we put the modified parts of Scaletti's working definition of sonification back together we obtain ... *a mapping of information to perceptual relations in the acoustic domain to meet the information requirements of an information processing activity*.

This phrase can be made more succinct by substituting "sounds" for the phrase *perceptual relations in the acoustic domain*. The definition for auditory information design becomes

the design of **sounds** *to support an* **information** *processing activity*.

3.3 An approach to auditory information design

The approach to auditory information design proposed here builds on Scaletti's definition of sonification. The approach focuses on the design of sounds to support an information processing activity. The approach has two parts that hinge on **information**

1. Requirements: analysis of the information requirements of an activity
2. Representation: design of an auditory representation of the information requirements

3.3.1 Requirements

There are a variety of methods for analysing information. Task analysis is a method developed in Human-Computer Interaction (HCI) design to analyse information required to manipulate events, modes, objects and other aspects of user interfaces [Kaplan B. and Goodsen J. (1995)]. This form of analysis is particularly concerned with actions that occur in sequence and parallel, and the feedback of the current state of the interface. Data characterisation is a method developed in scientific visualisation to describe the relations contained in data sets [Robertson P.K. (1991)]. This analysis addresses concerns about the validity and faithfulness of a representation that is to be re-represented in some other form. A combined task analysis and data characterisation can define a mapping from data relations to information that is useful in a task. This mapping may involve transforming or selecting parts of the data set, for example highlighting a region. The combination of task analysis and data characterisation has been demonstrated in a system for designing colour representations called PRAVDA [Bergman L.D. Rogowitz B.E. and Treinish L.A. (1995)], and in an automatic graphing tool called BOZ [Casner S. (1992)]. These tools operate on the task and data descriptors with a rule base that selects a representation scheme. However there is a problem that the addition of a new task to the system requires new rules to be formulated for every type of data. As the number of tasks and data types increases there will be a combinatorial explosion of rules to cope with each special case. This problem can be addressed by introducing an explicit description of information requirements that separates the task and data, but is a function of these influences. When a new task type is added to the system it is only necessary to add a new rule to map that task type to the information requirements. The information requirements are the pivotal point of contact between the analysis and the realisation of the design. I call this the TaDa approach because the design is focused on information that is useful in a task and true to the data.

3.3.2 Representation

Once the information requirements of the activity have been analysed we are in a position to design a representation of that information. This is the stage where the designer maps the required information into sounds in which a person can hear that information. Listening is a complex process which has been described in terms of "innate primitive" and "learnt schema" levels of perception in Bregman's theory of auditory scene analysis. The TaDa approach addresses these two levels with different design methods - a case-based method for schema design, and a rule-based method for primitive design. The case-based method is a source of familiar auditory scenes that have an information structure that can represent the required information. For example the sounds of a kettle coming to the boil may be a familiar schema for an auditory display of temperature levels in a turbine boiler.

The rule-based method aligns auditory structure with information structure in accordance to principles from graphic design and models from psychoacoustics. Examples of these rules are that if the information is ordered then the sounds should be ordered, if the information is categorical then the sounds should be categorical. Equal differences along a continuum should sound equal to the listener, and zero values should sound like they are zero.

The final stage in the design process is to produce specified sounds on an auditory display device. It is called an auditory display device because the sound specifications are perceptual (auditory), rather than device specific (acoustic), so the display can be transported to other devices. The display device may be a compound of hardware, software, synthesis algorithms, samples and audio components such as amplifiers, speakers or headphones. The reproduction of perceptual specifications on a device requires a measurement of the mapping from perceptual coordinates to control parameters, device capabilities and audio ranges. There is no point specifying sounds that cannot be produced by the display so a knowledge of the device characteristics is vital in the design process.

3.3.3 TaDa tiles

The TaDa approach is summarised by an arrangement of six tiles organised into two trapezoidal shapes shown in Figure 3-1. The upper trapezoid is the requirements part of the approach. The lower trapezoid is the representation part of the approach. The component tiles are facets of the design process, and common edges are connections between these facets.

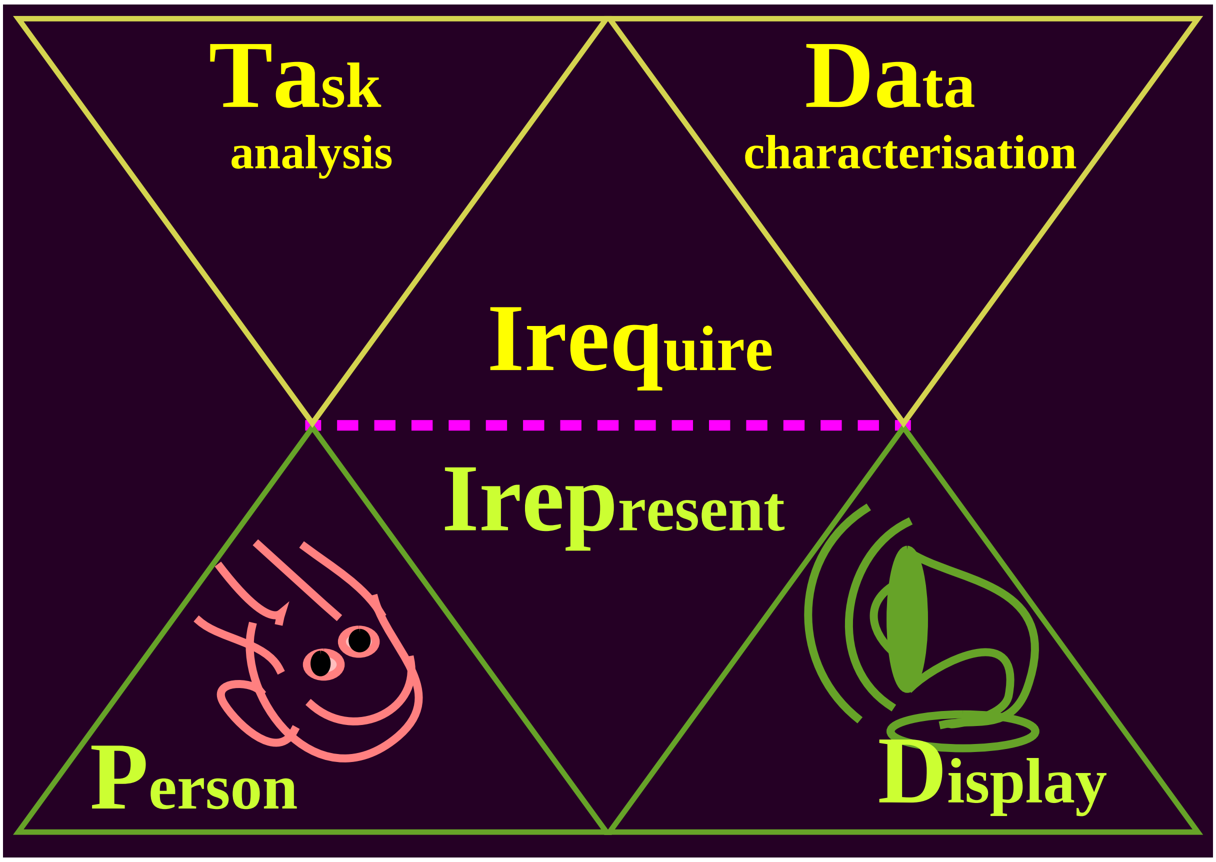


Figure 3-1: Facets of the TaDa approach to auditory information design

The requirements trapezoid consists of the Task analysis (Ta) tile, Information requirements (Ireq) tile, and the Data characterisation (Da) tile. The Ta and Da tile sit upon the Ireq tile to indicate that both of these facets are necessary for the analysis of information requirements. The representation trapezoid consists of the Person (P) tile, Information representation (Irep) tile, and Display device (D) tiles. The Irep tile sits upon the P and D tiles to show that the design depends critically on the types of information relations that can be heard by a human listener, and the types of sounds that can be produced by a particular auditory display device. The core of the TaDa approach is the central diamond where the trapezoids connect. This is where the information requirements (Ireq) are met by the information representation (Irep). The TaDa approach moves through the phases of requirements analysis, design, and representation as shown in Figure 3-2. The phases generally move from left to right, top to bottom through the TaDa tile arrangement. However it is expected that certain facets and connections will be revisited and improved during the process, which is not intended to be strictly linear.

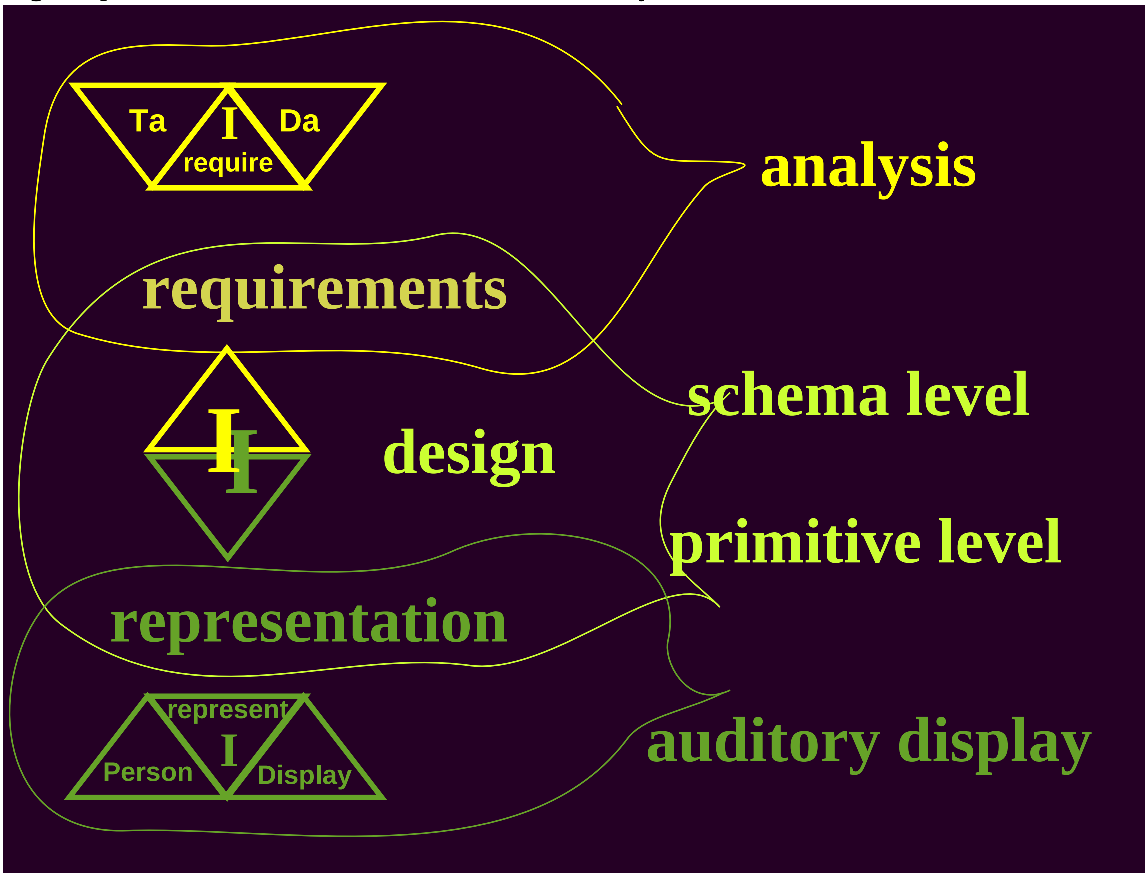


Figure 3-2: TaDa process

3.4 Summary

This chapter proposed the TaDa approach to auditory information design. The approach builds on Scaletti's working definition of sonification, which has two parts - a representation part and a requirements part. Some aspects of the definition were modified based on issues raised in previous approaches to auditory display (see Chapter 2). These issues included the need to analyse useful information using techniques such as task analysis and data characterisation, an explicit differentiation between auditory and acoustic domains, and the need to consider the characteristics of the display device in the design process. A modified version of the definition was proposed as a definition for auditory information design... *the design of* **sounds** *to support* **information** *processing activities.* This definition is the core of the TaDa design approach, which integrates task analysis, data characterisation, perceptual factors, and device characterisation into a multifaceted framework of methods. The framework is described by triangular tiles organised into a requirements trapezoid sitting on a representation trapezoid. The design process is intended to be iterative and the hybrid process of methods is intended to allow the designer scope to be flexible in the process. A multifaceted approach may be particularly beneficial in real world design practice where problems are messy and involve many complex issues.

4 • TaDa: task and data analysis of information requirements

Information is a difference that makes a difference. [Bateson G. (1972)]

Useful information is the answer to a question. [Bertin J. (1981)]

The purpose of the TaDa approach is to design sounds that carry useful information. The core of the approach is a meeting of information requirements with an information representation. This chapter describes the methods used to elicit the information requirements of a design problem. The first section introduces scenario analysis as a technique for capturing key features of the problem. The next sections describe the particular flavour of task analysis and data characterisation used to decompose the problem, and give a detailed account of the parts of analysis. The TaDa analysis is supported by a form interface to an Access™ database.

4.1 Describing the problem with a story

The design of useful information display must involve a description of how the display will be used. This may best be obtained from someone who is involved in that activity. User-centred design has become a prominent topic in Human-Computer Interaction. This method involves the user in the design process, and communication between the designer and the user. Techniques, such as storyboards, scenarios, interviews, and case studies, have developed as a means of information exchange between technical and non-technical groups.

Storyboards are effective for discussing and presenting design situations within collaborative design groups, and between designers and their clients [Lewis C. and Rieman J. (1994)]. Scenario analysis incorporates the knowledge and experience of the user into the design process through the recounting of instances of interaction with the system, which vary from real episodes to more or less constructed stories [Klausen T. and Aboulafia A. (1995)]. Interviews are often used to evaluate an existing system, as in the analysis of a decision support system for the St. John Ambulance service [Wong, Sallis and O'Hare (1995)]. Use-case analysis is based on descriptions of interactions which are written by the designer from the perspective of the system [Kaplan B. and Goodsen J. (1995)]. These techniques force the designer to be specific about the features of the real system and user, rather than addressing abstract issues which may have little real impact on the situation of use [Lewis C. and Rieman J. (1994)].

The method of problem description adopted in this thesis is a text description of an activity written or spoken by the user involved in that activity. The intent is to obtain short story-like descriptions in very general terms at any level of the problem. The word activity is used, rather than task, to indicate a loose association between elements, rather than a structured account. Below, and shown in Figure 4-1, is an example of a problem scenario transcribed from a verbal description and demonstration of the problem by Chris G., a software engineer working on a geophysical visualisation interface at CSIRO.

The GeoViewer is a 3D interactive view of rock strata for mine planning and other geological applications. You can move about anywhere in the space, and see the rock layers. The rock- type of the layers is shown by colour and texture, and a mouse click can popup a text description. You can see more by turning on the transparent view, and speed up the interaction with wireframe views. One problem is that it can be hard to see what type of rock it is when it is transparent, or wireframe. Also the popup text can get in the way of what you are looking at.

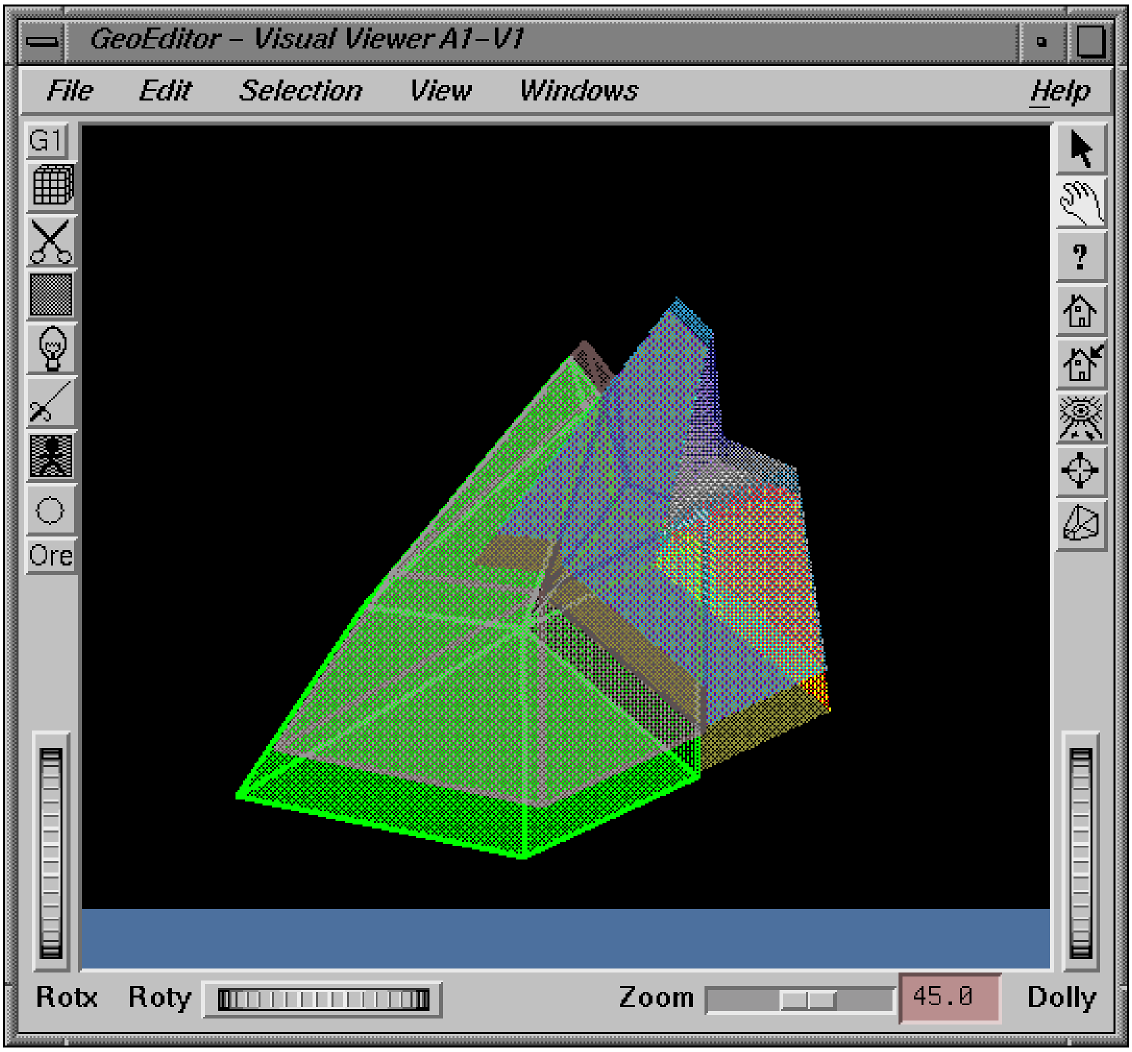


Figure 4-1: The GeoViewer

4.2 A bridge from story to requirements

Once the problem has been described it can be analysed to find key features and requirements of a solution. The analysis identifies key features of the problem description that are relevant in the design domain, and different analyses may be carried out with the same problem description.

The observation that “useful information is the reply to a question” [Bertin J. (1981)] provides a way to identify features that are relevant in the design of an information display. The problem captured by the scenario story is recast as a question, and the range of possible answers is identified. The questions extract the information required to carry out the activity, the relationships between the answers further specify the information type, and the subject of the question is the phenomenon of interest in this activity. The key features (Question, Answers, Subject) are a bridge from the story to the Requirements Analysis. They are an important first step toward the representation of the problem in a standard format that provides a framework for systematic design.

The Scenario Description for the GeoViewer, is shown in Figure 4-2. The story was recast as a Question “what type of rock is this?”. Chris agreed that this question adequately summarised the problem. However he pointed out that there is no definitive answer to this question - there are as many answers as there are types of rocks in geological data sets which can be visualised with this software. The GeoViewer handles this by allowing the user to arbitrarily associate rock types with colours in an editing panel. A similar panel could be made for sound allocations from a palette. Rather than consider all possible answers, a representative subset was chosen to focus the design on a useful implementation. The Answers {coal, sandstone, granite and marble} were selected as typical. The Subject of the Question in this example is the rock-type.

Scenario

	Title	GeoViewer
	Storyteller	C.G.
STORY		The GeoViewer is a 3D interactive view of rock strata for mine planning and other geological applications. The rock- type of the layers is shown by colour and texture, and a mouse click can pop-up a text description. You can see more by turning on the transparent view, and speed up the interaction with wireframe views. A problem is that it can be hard to tell the rock-type when it is transparent, or wireframe. Also the popup text can get in the way
KEYS	*Question*	what rock-type is it?
	Answers	coal, sandstone, granite, marble
	Subject	rock-type
	Sounds	???

Figure 4-2: Scenario Description for the GeoViewer

4.3 Requirements

The TaDa requirements analysis draws on task analysis and data characterisation methods that have been established in HCI and visualisation. TaDa amalgamates variants of these methods to produce an Information Requirements specification that drives the design of an Information Representation. There are three areas of analysis - Task, Information and Data. Each area analyses a different key feature from the Scenario Description. The Task section analyses the Question key, the Information section analyses the Answers key, and the Data section analyses the Subject key.

4.3.1 Task analysis

Task analysis is an established technique in HCI and visualisation design, and the fields proposed here are an amalgam of components borrowed from several different analyses [Wurman R.S. (1989)], [Norman D.A. (1991)], [Robertson P.K. (1991)], [Kaplan B. and MacCuish J. (1993)], [Rogowitz B.E. and Treinish L.A. (1993a).], [Lewis C. and Rieman J. (1994)], [Alty J. (1995)].

These components have been selected for their relevance to designing information in a temporal medium like sound. The task analysis is rooted in the Question key from the Scenario Description.

Generic question

Although there are an unlimited number of questions which may be asked, Bertin proposed that they may all be classified in terms of three levels of information. The subject of the question can be used to make this classification. A question which requires local information is about a single element. A question which requires intermediate information is about a subset of elements. The global question is about all the elements as a whole. By replacing the subject of a question with a generic tag, such as "it", or "they", a range of generic questions is proposed as a classification. New questions can be added to the classification scheme, shown in Table 4-1, as they are encountered.

Local Questions Subject {it}	Intermediate Questions Subject {they,which,what}	Global Questions Subject {everything, anything}
who is it ? what is it ? where is it ? is it ready ? is it time ? is it ok ? how good/bad is it ? how much is it ? what is wrong with it ? is it organised ? what was that ? where did it go ? what does it remind me of ?	where are they ? are they the same ? are they similar ? which is more ? which are the same ? which are similar ? which are different ? what is over there ? where am I ?	is anything here ? what is happening ? is everything ok ? has anything changed ? where am I ?

Table 4-1: Generic questions by information level/subject

Purpose

The purpose, or goal, or aim of a task is identified in most task analyses. A set of Purposes was obtained by interpreting the purposes of the generic questions, as shown in Table 4-2. The 10 Purposes proposed here are an amalgam of the search and compute types that have been defined in graphic display [Casner S. (1992)], and interactive confirmation, navigation, and alert types found in interface design methods [Brewster S.A. (1994)]. Several extra Purposes were added after analysing the uses made of sounds in everyday listening experiences (see Chapter 5). *Relax* is a purpose that captures the way people sometimes use low-level background noise or music to block out unwanted auditory disturbances which interfere with sleep, or mental attention in some activity. *Remember* has been included because sounds can be used to remember things - for example I recently noticed that I had mis-dialled my parents phone number by a change in the familiar tone-dialling tune. Sounds attract attention and are an important part of the engagement of interest in entertainments of all kinds. *Engagement* has been included because it is an important role that sounds have in movies and computer games, and there is potential to use them this way to improve workplace activities too.

Question	Purpose	Description
are they the same ? are they similar ? which are similar ? which is different ? which are different ? what is over there ? what is here ? what is happening ? has anything changed ?	analysis	observe relationships, groupings, trends, outliers, patterns
is it ok ? is it ready ? is it time ? are they the same ? is everything ok ?	confirm	absolute boolean confirmation
who is it ? what is it ? what is wrong ? what state is it ? what is over there ?	identify	absolute identification from a familiar set
how good/bad ? how much is it ? is it organised ? is everything ok ?	judge	absolute classification from a familiar ordered set
which is more ? which are same ? which are similar ?	compare	relative comparison of ordered properties
where are they ? what is here ? where am I ?	navigate	interactive movement through an ordered space
where did it go ? where are they ?	track	track an object through an ordered space
what was that ? has anything changed ?	alert	highlight or draw attention to an element or subset
	relax	mask unwanted noise, de-emphasise highlights
what does it remind me of? has anything changed ?	remember	remember places, times, people, information
what is over there ? has anything changed ?	engage	attract, entertain, maintain interest

Table 4-2: Generic question by purpose

Mode

The distribution of attention between overlapping tasks may help capture the ebb and flow between what is information at one moment and noise the next. An interactive task requires full attention, a monitoring task requires focused attention, and a background task can continue while something else is the focus of attention. The task mode field is directly borrowed from Kaplan and Goodsen, and is shown in Table 4-3.

interactive	manipulation with feedback e.g. tuning a radio
focus	conscious attention to an element e.g. conversation in a noisy room
background	attention focus is elsewhere e.g. watching television while babysitting

Table 4-3: Task attention mode

Type

Sounds can overlap, form patterns and cycles, be ongoing or very short. The analysis of overlapping tasks are a part of the HCI task analysis proposed by Kaplan and Goodsen [Kaplan B. and Goodsen J. (1995)]. Discrete/Procedural tasks are initiated by a single event, linear, short and seldom overlapping. Continuous/Tracking tasks are interactive, overlapping and have undefined closure. Branching/Decision tasks affect the subsequent course of action. The task event type field is borrowed from Kaplan and Goodsen, and is shown in Table 4-4.

discrete/procedural	initiated by a single event, linear, defined closure, seldom overlapping
continuous/tracking	require constant monitoring, constant feedback is used to iteratively make refinements, often overlapping other tasks, fuzzy closure.
branching/decision	affects the subsequent course of action

Table 4-4: Task event type

Style

Researchers in both sonification and visualisation have recognised two quite different styles of information processing tasks. The first is the exploration of data sets for interesting and as yet unknown features, which requires a faithful, veridical or isomorphic representation that preserves structural relationships. The second is the presentation of known features which may involve an intentional transformation of structure to draw attention or highlight or exaggerate. The task style field, shown in Table 4-5, is similar Kramer's exploration and presentation tasks [Kramer G. (ed) (1994b)], Cleveland's analysis and communication tasks [Cleveland W.S. (1985)] and the type I and type II tasks defined by Bergman et al. [Bergman L.D. Rogowitz B.E. and Treinish L.A. (1995)]

exploration	veridical representation that preserves information relations
presentation	intentionally transform the structure to draw attention to particular features, exaggerate details, or segment into regions.

Table 4-5: Task style

4.3.2 Information

The Answers to the Question key contain the information needed to carry out the activity. A characterisation of these answers can specify the information requirements of a display to support that activity.

Level

The level of information describes whether it concerns a single element (local), a group of elements (intermediate) or all of the elements as a whole (global). The level of information in the design scenario is obtained from a classification of the Generic Questions in these terms (see the previous section), for example “what is it?” is a local question.

local	information related to a single element
intermediate	information related to a subset of elements
global	information about all of the elements as a whole

Table 4-6: Information level

Reading

A direct representation can be understood with little training, can be understood almost immediately, and allows judgements which are not readily swayed by the opinions of others [Ware C. (1993)]. Some examples of direct representations are scatterplots, satellite images, and geiger counters. Conventional symbols, on the other hand, depend on learning or a legend to be understood. However they have the advantage that they may carry complex concepts built on layers of reference. Some examples of conventional representations are traffic signs, morse code, and hand gestures.

conventional	learnt, cultural, varies between individuals
direct	little training, immediate, resists bias, cross-cultural

Table 4-7: Information directness

Type

The set of Answers to some questions are qualitative, and others may be quantitative. For example a set of Answers, such as “coal” or “sandstone” that identify materials as categorical types is qualitative. Answers such as “twice as much coal as sandstone”, or “no shale at all” provide information that involves some form of measurement. In data analysis the data relations are characterised as 4 main types - nominal, ordinal, interval and ratio. This characterisation may also be used for the information relations contained in the Answers. Some specialisation of the types occurred when this scheme was used to design the demonstration scenarios described in later chapters. The additional types are boolean, ordinal-with-zero, and ordinal-bilateral.

none	no information is involved
boolean	2 different categories (e.g. yes, no)
nominal	difference without order (banana, apple, orange)
ordinal	difference and order (e.g. low, med, high)
ordinal-with-zero	difference, order and a natural zero (e.g. none, some, lots)

Table 4-8: Information type

ordinal-bilateral	difference, order, central zero (e.g. less, same, more)
interval	difference, order and metric (e.g. Celsius temperature scale)
ratio	difference, order, metric, and natural zero (e.g. mm of rainfall)
unknown	information type is unknown

Table 4-8: Information type

Range

The number of Answers can have a significant effect on the design of a display. Large catalogues may need to be organised and represented differently from small sets of 3 or 4. In the case of continuous information the range of variation is recorded in this field. In exploratory tasks or unknown domains where the range of the answers is unknown the Range is tagged as *unknown*.

Organisation

The information in a display is contained in the interplay of relations between the elements. These relations can be organised in a variety of ways. A characterisation of information organisations developed for graphic design has been borrowed directly from Wurman [Wurman R.S. (1989)]. This scheme may be used to characterise the organisation of the Answers, for example {coal, sandstone, shale, limestone} are a *category* organisation

category	Pertains to organisation of goods or types. Category can mean different models, types or questions. This mode lends itself to organising items of similar importance. Category is well reinforced by colour as opposed to numbers which have inherent value.
time	Works best for events that happen over fixed durations. It is an easily understandable framework from which changes can be observed and comparisons made
location	The natural choice for examining and comparing information from diverse sources or locales. If you were examining an industry you may want to know how it is distributed around the world. Location doesn't always have to refer to a geographical site. Doctors use locations in the body as groupings to study medicine.
alphabet	Lends itself to organising extraordinarily large bodies of information, such as words in a dictionary or names in a telephone directory. As most of us know the alphabet this organisation works when another form such as category or location may not.
continuum	Organises items by magnitude from small to large, least expensive to most expensive, order of importance etc. It assigns value of weight. Which department has the highest rate of absenteeism? What is the smallest company engaged in a certain business? Unlike category, magnitude can be illustrated with numbers or units

Table 4-9: Information organisation

4.3.3 Data

The Subjects key of the Scenario Description identifies the phenomenon of interest. The characterisation of this phenomenon can help the designer select a representational mapping that provides useful information about the relevant aspects of the phenomenon.

Type

In data analysis the general data types (nominal, ordinal, interval, ratio) have been developed to represent different phenomena. For example the Subject "type of rock" is a nominal phenomenon, whereas "percentage of a rock type in a rock sample" is a ratio phenomenon which can be measured and has a zero.

none	
nominal	difference without order (banana, apple, orange)
ordinal	difference and order (green, crisp, ripe)
interval	difference, order and metric (temperature)
ratio	difference, order, metric, and natural zero (rainfall)

Table 4-10: Phenomenal type

Range

The number of categories in a nominal or ordinal phenomenon can have a great bearing on the representational mapping of that phenomenon in a display. The display may need to show some or all of those categories, depending on the information required. In a continuous phenomenon the range of variation is recorded in this field as a basis for scaling transformations which may be needed in the mapping to the display representation.

Organisation

Physical phenomena such as temperature are organised in the physical continua of space, time and energy. Abstract phenomena such as stock prices may be organised by alphabet or category.

category	organisation by difference
time	organisation by time
location	organisation by spatial position
mnemonic	organisation by mnemonic, e.g. alphabet
continuum	organisation by continuous order

Table 4-11: Phenomenal organisation

4.4 Requirements of the GeoViewer

This section demonstrates TaDa Information Requirements Analysis by analysing the GeoViewer geological exploration interface.

Task analysis of GeoViewer

The Question Key is the bridge to the Task Analysis. The Question {what type of rock is it?} from the GeoViewer is transformed to the Generic {what is it?} by removing the subject, and referring to Table 4-1. The Purpose {identify} is looked up from Table 4-2 with the Generic question. The Mode is {interactive} because the question is made through a mouse click which is a manual operation. The Type is {discrete} because the question only occurs at discrete times during the activity. The Style is {exploration} because the user does not know the rock structure before using the GeoViewer.

Information analysis

The Answers Key is the bridge to the Information Analysis. The Answers for the GeoViewer are {coal, sandstone, shale, limestone}. The Reading is {direct} because the diversion of visual attention by popup text is a problem in the visual display design. The Type is {nominal} because the Answers are different but have no ordering. The Level is {local} because each Answer is about a particular rock strata. The Organisation is {category} because the relationships between Answers are not ordered in space or time, but are simply different. The Range is {4} which is indicative of the number of rock types that might be expected.

TaDa Analysis

Generic	what is it?	
Purpose	identify	TASK
Mode	interactive	
Type	discrete	
Style	exploration	
Reading	direct	
Type	nominal	INFO
Level	local	
Organisation	category	
Range	4	
Type	nominal	DATA
Range	4	
Organisation	category, space	

Figure 4-3: Example of TaDa requirements analysis

Data analysis

The Subjects Key is the bridge to the Data Analysis. The Subject in the GeoViewer is {type of rock}. The Type is {nominal} because types of rocks have difference but no order. The Range is {4} because there are only 4 types of rocks in the region of interest. The Organisation is {category, space} because the rock structure is both spatial and material.

4.5 Computer aided support

The TaDa Information Requirements Analysis has been implemented as a graphical form-like interface in an Access™ database. This tool makes it easy to store and recall design problems. The options for each field can be selected from a popup menu, which aids the transcription and analysis of a new problem.

4.6 Summary

This chapter presented a scenario-based method for analysing the information requirements of a design problem. The method recasts the problem as a Question and some Answers about the Subject of the Scenario. These Keys are the bridge between the informal story that describes the problem, and the formal analysis. The analysis is an amalgam of visualisation and HCI task analyses and data characterisation methods specifically selected to support the design of auditory information. There are three parts to the analysis - the Task analysis of the Question, the Information analysis of the Answers, and a Data characterisation of the Subject. The rationale for choosing each analysis field was described, and examples were used to illustrate each field. The method was demonstrated in an analysis of the GeoViewer tool for interactive exploration of rock structures. A form-based interface was implemented in an Access™ database to support the method.

5•EarBenders: case-based design from stories about listening

The first obstacle is the prevailing sonification model, which is simply to map data to sound parameters arbitrarily. The resulting sound is typically unpleasant and lacking any natural connection to the data represented (one intuitively feels that medical images, for example, ought to somehow sound different from demographic data or satellite imagery). Models of sonification more sensitive to the kinds of data presented must be developed.[Smith S. (1990)]

Designers often base a new design on a previous version that has proven successful in similar problems. A previous solution can be a way to quickly come to grips with a design problem, and provides a starting point for a top-down process of iteration. This chapter introduces the case-based method of design by example, and describes how it has been adapted for auditory information design. The case-based method relies on a rich source of examples to be effective, but as yet there are not many examples of auditory display design to draw upon. An alternative resource was developed by collecting stories about everyday listening experiences into a database, which I call EarBenders. The information requirements of a design problem can be used to search this case-base for everyday examples which share a similar task, data and information structure with the problem. The ability to do this search required each story to be analysed with the TaDa Information Requirements Analysis developed in the previous chapter. In addition an auditory characterisation was developed to describe the sounds in each story, and provide a footing for auditory design. The sound characterisation also provides an opportunity to extract principles of sound design from regularities between auditory structure and information structure in the example cases. The case-based design of auditory information is demonstrated on a problem in a geological visualisation interface, called the GeoViewer.

There are two appendices for this chapter. Appendix 5-1 lists the EarBenders case-base. Appendix 5-2 lists the software programs that retrieve relevant cases from the database, and synthesise a sound design from them according to the method.

5.1 Advantages of case-based design

The stories in many professional design journals and magazines are case-studies which convey information about a good design in a way that can be easily understood and assimilated. The popularity of these magazines attest to the value of case-studies in design practice. The designer can use good examples as a starting point, or borrow elements from them. In HCI the intelligent borrowing of good ideas from existing interfaces is promoted as an effective design technique which is less risky, quicker and easier to implement than a totally original design, and has the advantage that users will already be familiar with the borrowed elements [Lewis C. and Rieman J. (1994)]. In a similar vein, comparability analysis uses the precedent system, and lessons learnt from it, as the starting point for the design of the replacement system [Carlow International Inc. (1992)]. The synthesis of a design from examples is suited to design domains where there is little theory and few principles that can provide guidance. A critical survey and comparison of designs can identify new principles of design from examples, as demonstrated by Tufte in his collection and critique of graphs [Tufte E.R. (1983)]. Some advantages of the case-based method are listed below.

Top down

Cases are top down - example cases address the key features of the problem, and provide an account of those features. Most real problems are large and complex, and the reuse of a previously successful design can more quickly lead to a better quality solution. The recall of several previous solutions provides an opportunity to integrate features to synthesise a new design.

Open ended

A case-base is an open-ended store of knowledge. New cases can always be added to improve the depth and breadth of examples. Many different solutions to the same problem can be included, providing a store of variation which can illuminate the features of the design space. The database can change over time to reflect a particular perspective or style as cases are added.

A source of principles

A systematic analysis of cases may lead to the discovery of general design principles. These principles capture regularities in the relations between features of the problem space and features of the solution.

Computer tools

Computer tools can assist a case-based method by making it easy to store and retrieve design cases from a database. Case-based tools have been implemented for designing buildings, computer interfaces, fire engines, and CAD drawings [Maher M.L. Balachandran M.B. and Zhang D.M. (1995)]. The method depends on the retrieval of relevant cases. The retrieval of relevant cases depends critically on a representation that captures features of the cases that are important in a solution.

5.2 Everyday stories as examples

The case-based method depends on a rich source of examples. The small number of examples of auditory display limits the potential for case-based design in this field. However the observation that people use sounds in everyday activities opens the door to an alternative source of examples. I investigated the idea that everyday uses of sounds might serve as design examples in auditory display by collecting stories about listening experiences into a database. The collection was started by an email message to all staff at CSIRO Mathematics and Information Sciences that included some examples of everyday uses of sound, and a request for similar stories.

During the next few weeks if you notice yourself using your hearing to help you, say, find a lost pin, or search for hollow spaces in a wall, or notice something wrong with *your car, or tell whether the dog is hungry, anything at all, then please email me with a brief story about the occasion.*

Nearly 200 stories have since been collected. The initial round of stories were received by email, but many more have come from conversations, newspapers, books, television, movies, and ongoing general observation. Each story is written down and stored in a computer database, along with the name of the storyteller and a title. The database is called EarBenders because “to bend your ear” is an Australian colloquialism to do with telling a story or having a conversation. The stories are about sports, cooking, camping, domestic chores, games, car maintenance and many other activities. Some example stories are shown in Figure 5-1

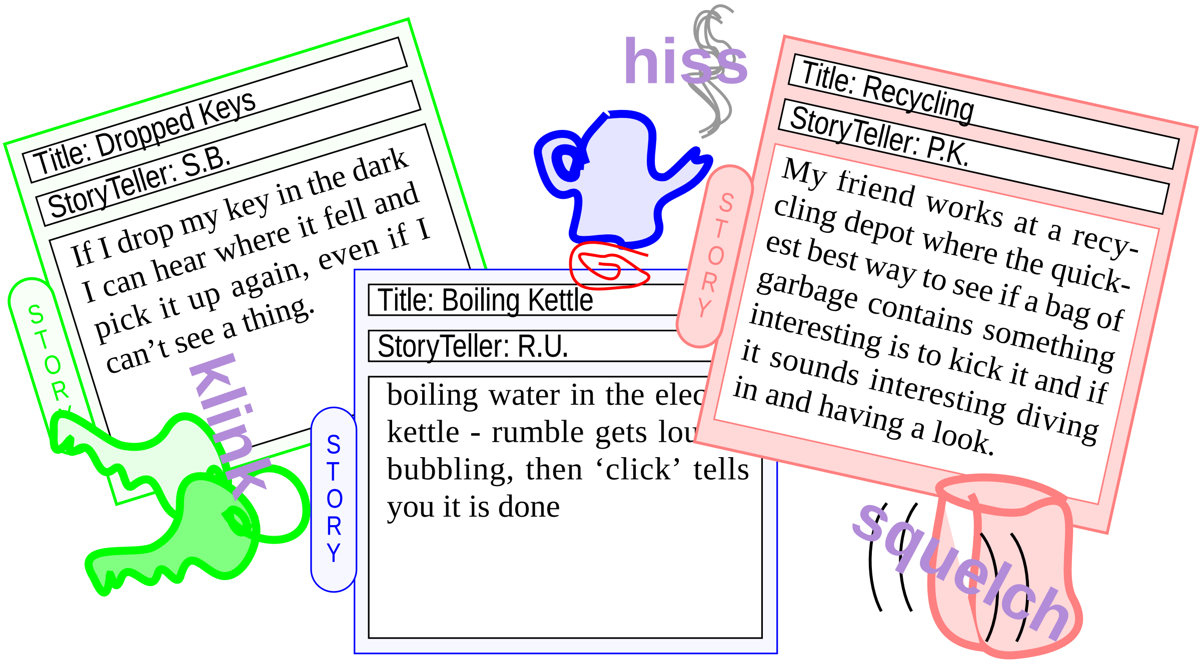


Figure 5-1: EarBenders stories

The stories capture the way sounds are useful, the ways that people hear information in sounds, and the organisation of the sounds. The use of the sounds can be found by recasting each story as a question. In the examples the questions are “where did the key drop?”, “has the kettle boiled?” and “what is in this rubbish bag?”. The answers to these questions are sounds described in the stories. Where the sounds are not explicitly described the objects and events that cause them are. The organisation of the answers helps to identify the syntax of events in the story. If the {rumble, bubbling, click} sequence in the Boiling Kettle story did not occur in the right order then you would know something was wrong. These observations indicate that the stories contain semantic, syntactic and pragmatic elements that are important for designing symbols.

Implementation details

The EarBenders database is currently implemented in Microsoft Access™ 6.0. This tool lets you design a form as the interface to the database records. Keyword searches can be carried out on individual fields, or on all the fields at once. Partial matches can be specified with wildcard characters. An example of the interface for entering an EarBenders story is shown in Figure 5-2.

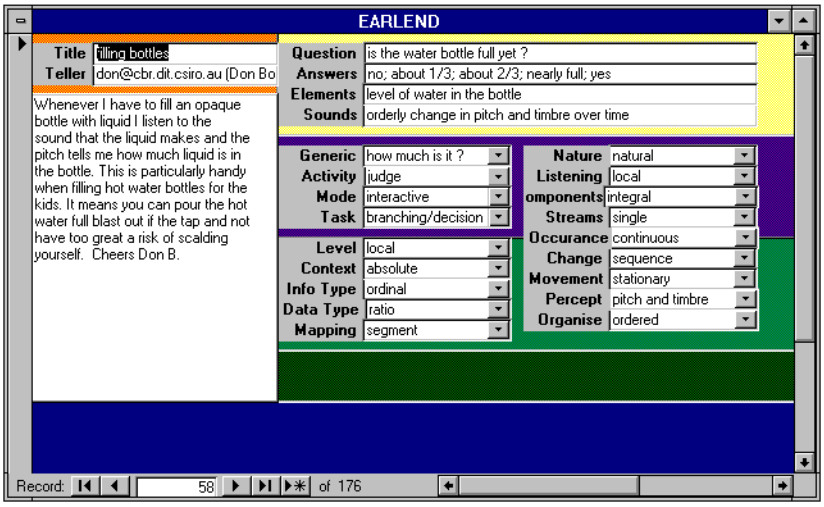


Figure 5-2: EarBenders database form

5.3 Three case-based methods

The EarBenders stories describe situations where sounds have been useful. Collecting and reading the stories was an educational experience which has changed how I think about the uses of sounds, and made me more aware of the many ways that people use them all the time. Entering the stories into a database provides the extra capability to search and retrieve examples that transforms EarBenders from a collection of stories into a resource for case-based design. There is more than one way that EarBenders may assist in auditory display design. Three methods that I have explored are called the metonymic method, the metaphoric method, and the pattern method, each of which is described in the following subsections.

5.3.1 Metonymic method

A metonym connotes the whole by a part of that whole - for example the sound “woof” may be a metonym for a dog. The metonymic method uses sounds that are a part of the design scenario. The sounds in a medical scenario might be heartbeats, breathing, or instrument beeps. Any sounds mentioned in a design scenario may be the beginnings of a metonymic palette. If there aren’t any sounds explicitly mentioned, then objects and events that are described may be a basis for sound design. This method is modelled on Gaver’s auditory icon method where he maps events and actions in the interface to events and actions that cause sounds in the everyday world [Gaver W.W. (1994)].

Keyword searches of the EarBenders database can assist a metonymic design. Keywords from the scenario domain can retrieve other stories about that domain which may provide a broader design perspective. For example a search for {medicine} may retrieve other medical stories that describe uses of medical sounds. If sounds are explicitly described in a scenario then you may search for other stories which have the same sounds in them. For example you might retrieve all stories with the word “crunchy” in them. Keyword searches for objects and events in a scenario may retrieve stories that also include those objects and events. For example, the scenario might involve opening a file, so you might search for “open” or “file” to find stories that have sounds that maybe related to these keywords. Two stories retrieved by the keyword {open} are shown in Figure 5-3. Door-muffle describes how you can hear the sounds from a room become louder when someone opens the door. Window-open describes the whistling that happens when you are driving along with a car window partly open.

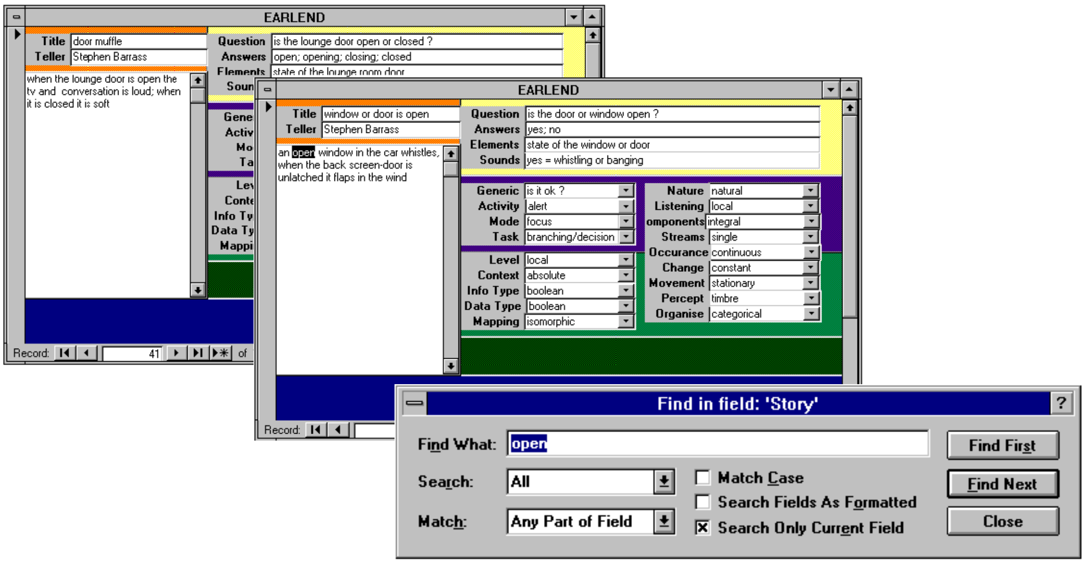


Figure 5-3: Retrieving stories by keyword search

5.3.2 Metaphoric method

A metaphor expresses the unfamiliar in terms of the familiar, for example a tree may be a metaphor for a filing system. A metaphoric design can help when sounds are not a natural part of the design scenario, which is often the case in computer-based applications. Stock prices and internet traffic don't normally make sounds, so are likely candidates for a metaphorical design. But what metaphor to choose? The metaphor needs to be easy to understand and relate to the phenomenon it represents. The information in the metaphor must be information about the phenomenon, or else it is useless. A relevant metaphor can be found by searching EarBenders for a story that can support the information requirements of the design scenario. These requirements are described by a TaDa analysis. The facility to search on TaDa fields was added to EarBenders by appending a TaDa analysis to each story. You can retrieve stories by entering the TaDa requirements of a design scenario into a query form. The cases that are retrieved are a source of useful metaphors. After some trials I found that it was quite rare to find cases that match the query in every way. The solution was to change the retrieval process from an exact match to a search by similarity. The measure of similarity is the number of fields that match. The query returns a ranked list where the top-most cases are most similar in TaDa structure to the design scenario.

TaDa analysis for an EarBenders story

The TaDa analysis of the EarBenders stories is the key to the metaphoric method that allows the retrieval of relevant metaphors. The process of analysis is briefly outlined in this subsection, and is summarised in Figure 5-4. The story we will analyse is the Recycling story introduced earlier. The Question is {what is inside this bag?} The Answers to the Question are {household garbage, bottles, crockery, nappies}. Probably there are many more answers than were given, but a comprehensive list may not be necessary to analyse the information relationships. The Subject of the Question is {inside this bag}. The Sounds that give each Answer are {plasticy, clink, clatter, squelch}. The identification of the Scenario Keys leads to the next stage of TaDa Analysis. First the Question {what is inside this bag?} is analysed in the Task section. The Generic question {what is it?} is obtained by removing the subject of the question, and referring to Table 4-1. The Purpose of the Generic question is {identify} which is found in Table 4-2. The Mode is {interactive} because the question is asked by kicking a bag. The Type is {discrete} because the question has an immediate response. The Style is {exploration} because the answer to the question is unknown. The Answers Key {household garbage, bottles, crockery, nappies} is analysed in the Information section. The Reading is {direct} because the identification can be made without reference or comparison. A new employee will probably need to look inside the bag but quickly learns to make a direct auditory identification to avoid an unpleasant experience. The Type is {nominal} because the Answers are different but have no order. The Range of Answers is {4} in this example. The Level is {local} because the Answer only pertains to the bag that is being kicked. The Organisation is {time}, because the answers are separated from each other in time. The Subjects Key {inside this bag} is analysed in the Data section of the TaDa analysis. The Type is {nominal} because the phenomena are mixtures of garbage objects that have characteristic differences in content but no natural physical order. The Range is {unlimited} to indicate that each garbage bag has a unique content. The Organisation is {location} because the garbage bags are separated by their locations.

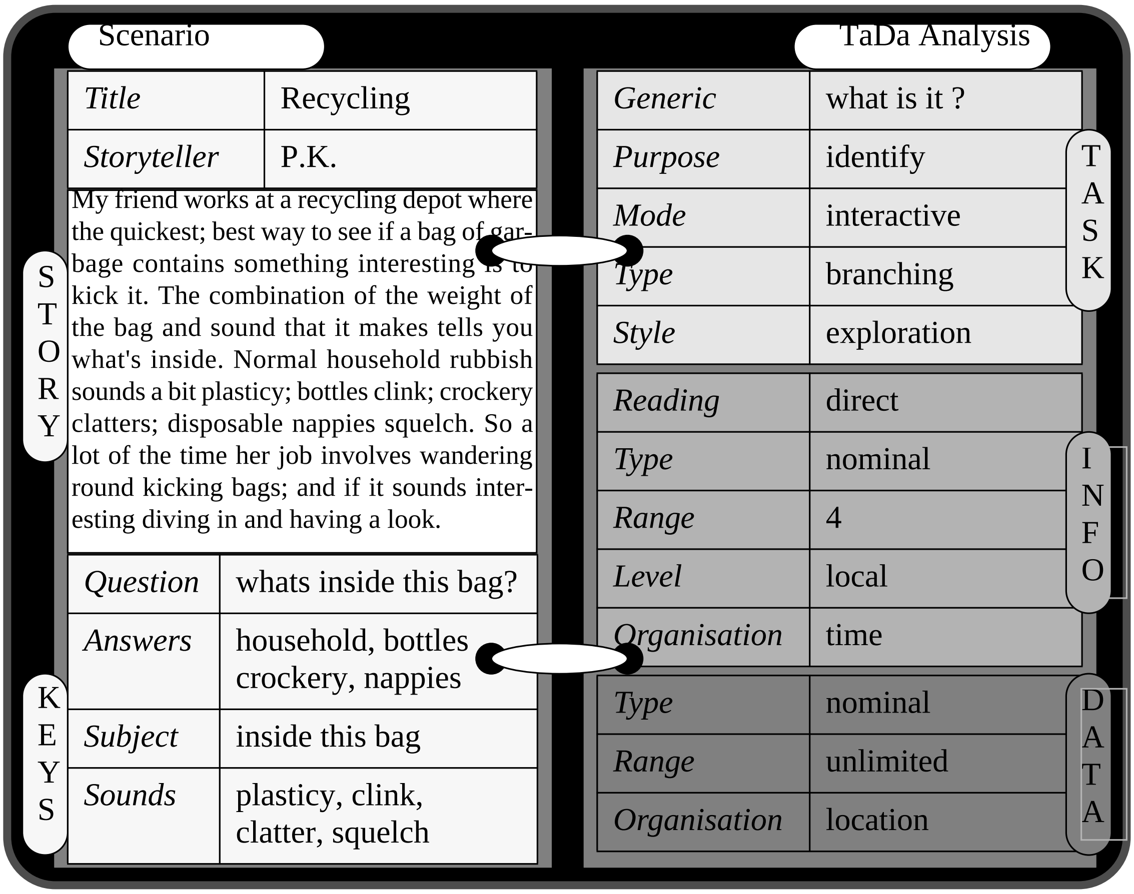


Figure 5-4: TaDa analysis for the Recycling story

Implementation details

The Access™ database can only retrieve records by exact match on all the query fields. This matching procedure does not allow the retrieval of cases by similarity that is necessary in the metaphoric method. The lookup of cases ranked by similarity to a TaDa query was implemented in a Perl program [Wall L. and Schwartz R. (1991)] called lookup.prl (see Appendix 5-2). The lookup process involves saving the Access™ database to a file as ascii text with a special “|” character as a field delimiter. The lookup.prl program accepts the text file as input, and generates a reorganised version ranked by similarity with the query. Similarity is measured by the number of fields that match. The best matches can be taken from the bottom of the generated file, or the file may be re-imported to Access™ and browsed with the forms interface. This process is shown in Figure 5-5.

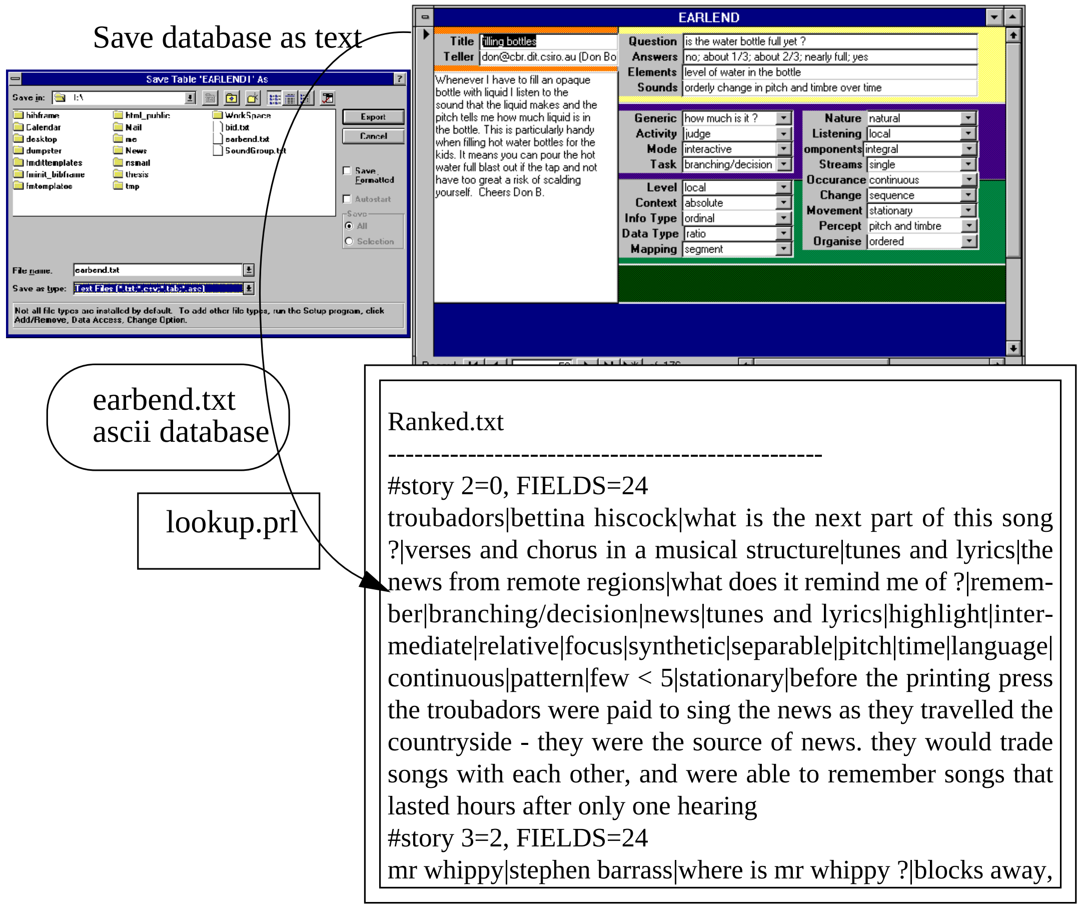


Figure 5-5: Looking up stories by similarity to a TaDa query

5.3.3 Pattern method

A pattern is a regularity in the mapping between the problem and solution domains. The pattern method begins by identifying features shared by different solutions to the same problem that may capture patterns in the mapping between domains. The pattern method was enabled by appending a description of auditory characteristics to each EarBenders case. When a TaDa query retrieves cases with similar information structure it also retrieves the auditory characterisation of each case. The characterisations are scanned for regularities identified by the majority trend in each auditory characteristic. The weighting of each trend is calculated from the number of cases which follow the trend. The result is a synthesis of the auditory characteristics of the retrieved cases. These cases have an information structure which is similar to the information requirements of the design scenario. Regularities in the auditory characteristics of the retrieved cases may capture patterns in the mapping between information and sounds. Hence the design of an auditory display from these regularities may support the information requirements of the design scenario.

The auditory characterisation

The auditory characterisation is both a description of sounds in the cases, and a specification of sounds for the design scenario. The characterisation is headed by the Sounds Key which has been added to the EarBenders Scenario Analysis to capture the auditory features of the story. This Key is a high level description of the sounds that give each Answer in the scenario. However a verbal description does not specify how to produce the soundon an output device. The Sounds Key is a bridge to a more detailed characterisation of the sounds which may assist in the pragmatic and syntactic aspects of the sound design. There is no all-encompassing way to characterise sound, so several different perspectives from music, psychoacoustics, perceptual psychology, and auditory display have been taken up. The characteristics are not necessarily orthogonal or independent, and only through practice will it be possible to determine which, if any, describe important features for auditory design. The current set of characteristics are {nature, level, streams, occurrence, pattern, movement, type, compound, descriptors}. Each is described in the following subsections.

Nature

Everyday sounds are acoustic events generated by physical interactions between objects rubbing and colliding in the environment [Gibson J.J. (1966)]. Everyday sounds also describe artificially generated sounds that have an environmental basis, for example digital recordings of acoustic events, or auditory icon algorithms [Gaver W.W. (1986)]. Musical sounds are generated by specially engineered instruments that shape the acoustics to engage a human listener. These instruments allow control over specific perceptual aspects, such as pitch, rhythm, timbre and loudness. Synthetic sounds are artificially generated sounds that do not have an acoustic basis. The sounds can be generated by electronic circuits or computer algorithms. Vocal sounds are moans, croaks, or other expressive but non-verbal sounds made by people and animals, including clicks and hums that insects make by means other than a vocal tract. Verbal sounds are recognisable words. Auditory Display researchers sometimes distinguish their field as the use of non-speech sound, but there are many ways that words can be used that are not the same as speech.

everyday	acoustic events in the physical environment e.g. knocks, scrapes, rumbles, crashes
musical	musical sounds generated by instruments specialised to shape acoustics to engage musical perceptions e.g. pitch, rhythm, loudness, timbre, etc.
synthetic	synthetic sounds with no acoustic basis e.g. car alarm, keycard beep, computer error quack
vocal	animal communications formants, phonemes, moans, grunts and sighs humming, whistling
verbal	recognisable words, singing

Table 5-1: Nature of the sound

Level

Analytic and holistic listening are musical terms for different listening styles. Picking the flute part in a symphony is an example of analytic listening. This attention to a single element is called local information in Bertin’s graphical method. Holistic listening to the overall sound of the orchestra is analogous to the attention to global information. A good display allows attention to information at more than one level [Bertin J. (1981)].

local	analytic listening to a single element e.g. violin in orchestra
global	holistic listening to many elements e.g. whole orchestra

Table 5-2: Listening level

Streams

The ability to segregate the flute or clarinet from the rest of the orchestra is explained in terms of perceptual streams in Bregman's theory of auditory scene analysis [Bregman A.S. (1990)]. Streams are perceptual groups that form when sounds occur simultaneously and in sequences, as is usual in everyday listening. Bregman's theory explains some of the factors that influence the grouping of acoustic events into perceptually cohesive sounds in the mind of the listener. This grouping enables the listener to consciously select between different sound streams as they become interesting or useful - an ability sometimes called the cocktail party effect. The number of streams may be estimated from the number of sound sources that can be consciously identified in an auditory scene.

single	only a single stream is involved e.g. a voice talking
pair	a pair of streams are involved e.g. bass line and melody
few < 5	a few streams are involved e.g. shaking a box of muesli
some <10	5-9 streams are involved e.g. car sounds while driving
many 10	more than 10 simultaneous streams e.g. the aural scene during a picnic lunch in the park

Table 5-3: Streams

Occurrence

Sounds can be one-off or ongoing, just as tasks may be discrete or continuous [Kaplan B. and Goodsen J. (1995)]. It takes at least 4 seconds for the primitive stream grouping process to stabilise, and once an interpretation of the number of sources has occurred it doesn't matter if one or other of them briefly disappears for a second or two [Bregman A.S. (1990)]. This 4 second hysteresis is found in other perceptions, and provides a basis for classifying the occurrence of sounds. A continuous sound doesn't have a definite beginning or end, and interruptions last for periods less than the 4 second hysteresis of perceptual continuity. A regular sound repeats at predictable intervals that are greater than 4 seconds apart. A sporadic sound repeats at frequent but unpredictable intervals of more than 4 seconds. An isolated sound does not occur often at all.

continuous	an ongoing sound in which breaks are < 4 seconds e.g. a waterfall
regular	a sound that repeats at intervals > 4 seconds e.g. a dripping tap
sporadic	unpredictable repetition at intervals > 4 seconds e.g. wind-chimes in light breeze
isolated	a one-off sound e.g. a dropped key

Pattern

Sounds can vary in many ways over time. A discrete sound may have a typical duration of only about a second or two, and so may be too short to hear a pattern in. A constant sound does not change over its duration. An unpredictable sound changes in unpredictable ways but maintains its essential identity. A cycle is a cyclic variation than can be predicted after it has been heard. A sequence is a sound that has a predictable directed variation.

discrete	a short sound with definite start and end e.g. a hand clap
constant	does not change much e.g. air-conditioner hum
unpredictable	unpredictable variation e.g. wind-chimes, pop-corn popping
cycle	predictable cyclic variation e.g. a squeaky wheel
sequence	predictable directed variation e.g. water bottle filling

Table 5-4: Pattern

Movement

The movement of a sound is relative to the listener. The stationary sound stays in roughly the same location. A change in distance occurs in movements when other cues for spatial location are not available. Jumping sounds shift location in a discontinuous fashion. The location of a smoothly moving sound can be tracked and predicted. A texture sound does not have an identifiable location or movement.

stationary	the sound is fairly stationary e.g. rattling mudguard as you ride a bike
distance	the distance of the sound is changing e.g. walking to a surf beach
jumping	the location of the sound jumps about in space e.g. flying grasshopper wing clicks
smooth	the sound moves smoothly through space e.g. a plane flying overhead
texture	the sound has no identifiable location or movement e.g. the rain

Table 5-5: Movement

Type

The type is a characterisation of the perceptual relations between the sounds. Perceptual psychologists typically classify perceptions as categorical or continuous. A categorical perception has difference but no order. A continuous perception has a unidimensional organisation. Continuous perceptions were further divided by Stevens into metathetic and prothetic types [Stevens S.S. (1966)]. A metathetic perception is not additive, for example the simultaneous occurrence of two sounds of the same pitch is not heard as a sound with an increased pitch. On the other hand a prothetic perception is additive, so for example when two sounds of the same loudness are heard together there is an increase in overall loudness. All of these types of relations have been scaled to create organisations of equal perceptual difference.

categorical	difference e.g. piano note, engine rev, dog bark
metathetic	difference and order e.g. pitch, brightness
prothetic	difference, order, and natural zero e.g. loudness, duration

Table 5-6: Type

Compound

Sometimes it is possible to separate aspects of the sound from its overall identity, for example you may listen specifically to the rate of vibrato, or the depth of tremolo. Other sounds are more integral and it can be hard to hear them as anything but a whole - for example the crunching of gravel, the popping of a cork or a hand clap.

separable	aspects of the sound can be readily heard e.g. vibrato, tremolo, pitch
integral	it is difficult to separate aspects of the sound from the sound itself e.g. crunching gravel, popping cork, hand clap

Table 5-7: Compound

Descriptors

The descriptors are a list of the words used to describe the sounds. Verbal rating scales have been used to measure the primary dimensions of perceived variation between sounds in multidimensional scaling experiments [Von Bismarck G. (1974a)]. Shared descriptors may indicate a consistent variation or similarity relations. Multi-word descriptions may indicate the separability and attention to particular aspects in the sounds.

Auditory characterisation of an EarBenders story

The Sounds Key for the Recycling case is {plasticy, clink, clatter, squelch}. The Sounds Characterisation, shown in Figure 5-6, is a more detailed description of the sounds that may provide a basis for organising the perceptual and psychoacoustic aspects of an auditory display. The Nature is {everyday} because these sounds are generated by everyday physical interactions between materials rubbing and colliding. The Level is {local} because the sounds emanate from the particular bag that is being kicked. The Streams are {few} because, for the most part, each kick forms a cohesive auditory entity. However it may be possible to also hear separate objects, perhaps a plate breaking or a can crushing, as highlights in the overall mass. The Occurrence is {isolated} because the sound only occurs when the bag is kicked. The Pattern is {discrete} because it only occurs once per kick. The Movement is {stationary} because

Character

Sounds	plasticy, clink, clatter, squelch
Nature	everyday
Level	local
Streams	few
Occurrence	isolated
Pattern	discrete
Movement	stationary
Type	category
Compound	integral
Descriptors	plasticy, clink, clatter, squelch

SOUND

Figure 5-6: Auditory characterisation of the Recycling story the sound doesn't move (unless you really send the bag flying!). The Type is {category} because the relations between the sounds are unordered, although some sounds may be more similar to each other than others. The Compound is {integral} because the sound is an overall or holistic identity. The Descriptors are a list of adjectives that were used in the description of the sounds. Shared descriptors may indicate principal dimensions of perceptual variation. In this case of integral sounds the descriptors are purely categorical and are copies of the Sounds themselves.

Implementation details

The pattern method is supported by a Perl program, called synthesis.prl (see Appendix 5-3), which automatically generates an auditory specification from trends in the auditory characteristics of EarBenders cases. This program accepts a file of cases in ascii text format, as output by lookup.prl. The best cases at the bottom of the file are scanned for trends. The most common value for each field is recorded, along with the count of cases that had that value. An auditory design is generated from the majority value of each characteristic, and printed to the standard output. The complete process, shown in Figure 5-7, is tied together by a csh script called casedesign.csh (see Appendix 5-4).

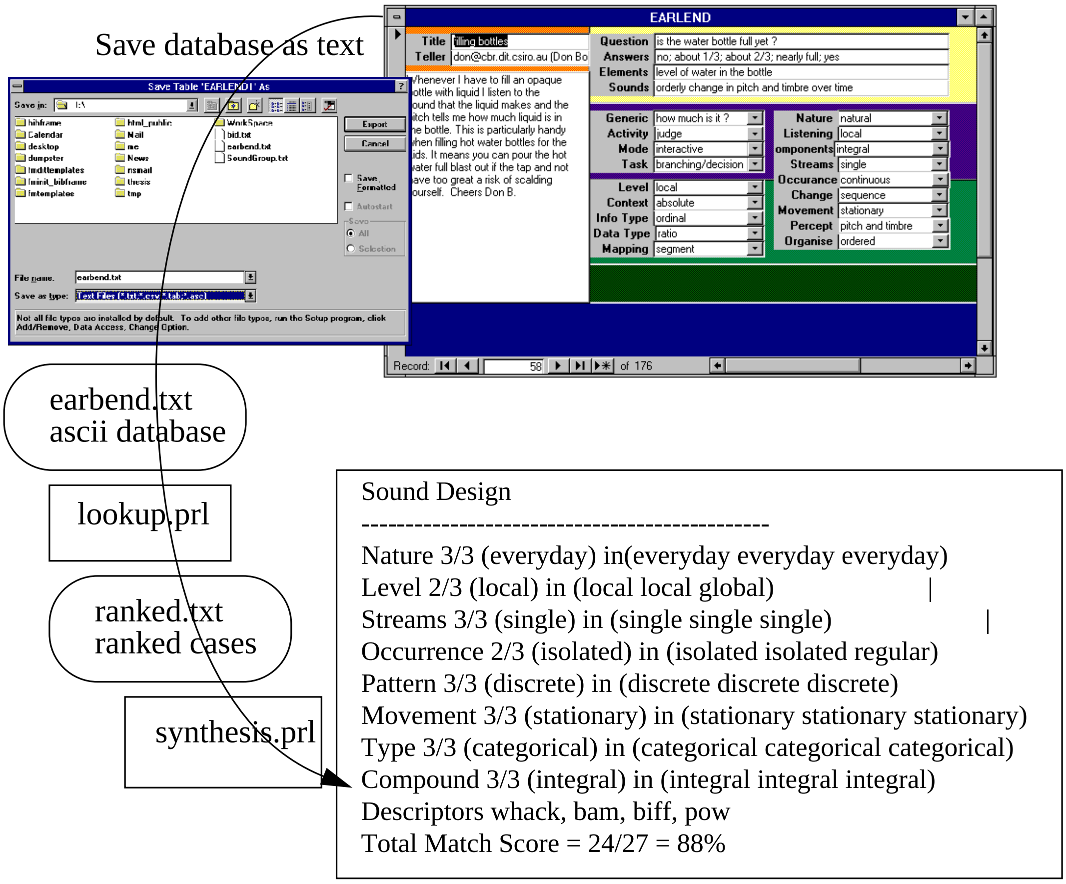


Figure 5-7: From TaDa query to sound design

5.4 Demonstration of EarBenders design

This section demonstrates EarBenders by designing sounds for the GeoViewer scenario that was introduced in the previous chapter. The metonymic, metaphoric and pattern methods are each demonstrated in the following sections. The GeoViewer scenario and analysis are summarised in Figure 5-8

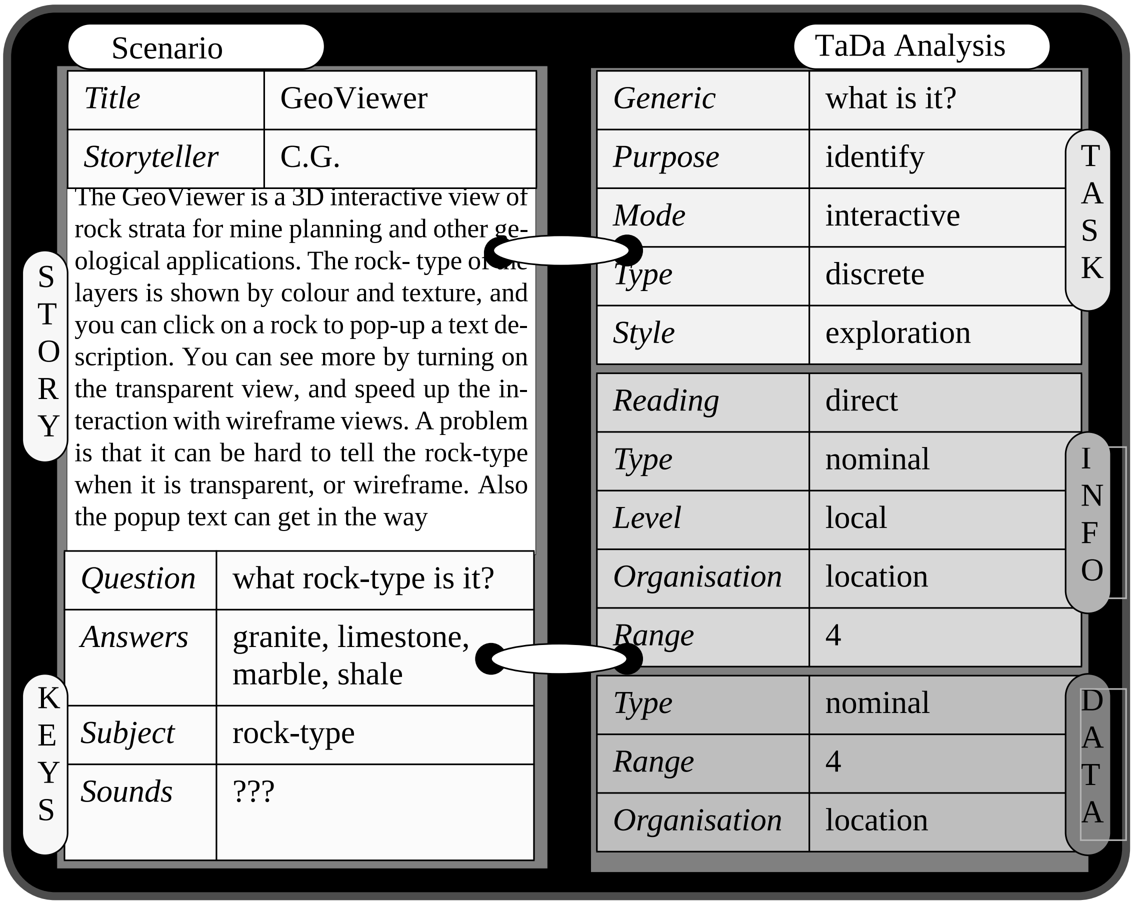


Figure 5-8: TaDa Analysis for the GeoViewer

5.4.1 Metonymic method

In the metonymic method we are looking for sounds that are a part of the design scenario. The GeoViewer is from the domain of mining exploration, and some keywords that may find other examples from similar scenarios are {mining, exploration, geophysics}. I tried searching EarBenders with these fields but did not turn up any cases. The story in the GeoViewer scenario, shown in Figure 5-8, does not mention any sounds so there are no leads there. The subject of the scenario is the rock-type, which is the object of interest. The actions that involve the rock are - clicking on it to find out what type it is, and moving it to see what shape it is. We can search EarBenders for object keywords like {rock, strata, layer}. This returned one story about dropping a rock down a hole to find out how deep it is, which did not seem relevant. The event keywords are {click, tap, knock, impact}. Click turned up a lot of stories which had clicking sounds in them but were otherwise irrelevant. Some possibilities were turned up by {tap} and {knock}which retrieved 4 stories - ripe fruit, gas cylinder, hammering a nail, and finding studs, shown in Figure 5-9 and Figure 5-10.

Scenario

Title	ripe fruit
Storyteller	D.B.

STORY

You can tell whether a water melon is ripe by tapping it and listening - this also works for apples - it must be something to do with the damping of the sound when the fruit isn't ripe, because it sort of rings when it is ripe.

KEYS

Question	is this fruit ripe?
Answers	yes, no
Subject	the fruit
Sounds	yes=ringing no = damped, dull

Scenario

Title	gas cylinder
Storyteller	S.B.

STORY

which gas cylinder has more in it? you tap them and the fuller one has a low dull sound while the emptier one is higher and has a brighter hollower tone

KEYS

Question	which cylinder has more gas in it?
Answers	this one, that one
Subject	the cylinders
Sounds	more gas = lower, duller tone

Figure 5-9: Stories retrieved by the keyword {tap}

Scenario

Title	hammering a nail
Storyteller	R.U.

STORY

A solid sounding knock, then a softer slightly clangy sound means that the nail is bent.
A solid sounding knock, then overly solid (higher pitch) means that the nail has hit something like a pipe or concrete.

KEYS

Question	is the nail ok?
Answers	yes, no
Subject	the nail
Sounds	yes=solid knock no = softer slightly clangy, overly solid high pitch

Scenario

Title	finding studs
Storyteller	S.B.

STORY

you can find wall studs by knocking on the walls and listening - where the stud is sounds solid and dull compared to the hollow sound of the rest of the wall.

KEYS

Question	where is the wall stud ?
Answers	here, not here
Subject	the wall stud
Sounds	here = dull solid not here = hollow

Figure 5-10: Stories retrieved by the keyword {knock}

The knocks and taps in these stories produce sounds that provide quite different information. The sounds in the stories do not seem connected to mining or rocks. However the idea of knocking on something to hear information does suggest tapping on a rock with a hammer. This idea may be the starting point for a design. My first investigation of this design was to go outside and tap some rocks with a hammer. I tapped on marble, granite, sandstone, and a conglomerate. They didn’t sound very different. I closed my eyes and tried to learn the sound each rock made. I could hear how hard I was tapping, and that it was rock that I was tapping, but could not identify what type of rock it was. I couldn’t hear any correspondence between the size of chunks of granite, and the tapping sound. The variations in the physical properties of the rocks don’t seem to have perceptible affects on the sounds generated by tapping them. Tapping metal and wood did produce discriminably different sounds. I could correctly identify rock, wood and metal by tapping them. My experience is that the identification of materials by tapping on them requires familiarity, and the discrimination between materials requires that they have quite different material properties. These observations raised the possibility of a metonymic palette comprised of clearly discriminable and identifiable impact sounds. The impact-like nature of the sound could suggest tapping and connote the scenario.

A metonymic design for the GeoViewer was investigated by assigning an impact sound to each rock type as follows

granite = tennis serve
limestone = door slam
marble = glass break
shale = bell strike

The resulting sounds were clearly different and it was easy to answer the question “is this the same rock type as that?” straight away. After using the interface for a couple of minutes I was able to consistently answer the question “what rock type is it?” from the sounds. The sounds demonstrably do provide the information required by the design scenario.

The sounds also allowed answers to other questions that weren’t in the design scenario. The intersections between rock surfaces can be visually ambiguous in a 3D view, and answering the question “which surface is in front?” involves a change in viewpoint which can be distracting and slow. The sounds let you answer this question by tapping at the intersection and listening for the fore-most surface. A development could provide information about the number and material of overlapping hidden layers. The prospect of unanticipated affordances of the sounds is very encouraging.

The investigation also turned up some problems. Firstly, there is an inconsistency in the degree of the impacts that does not correspond with the consistent nature of tapping with the mouse button. Secondly, the initial impression is that the breaking glass and the tennis serve are out of context because they clearly connote their origins. Nevertheless, after using the interface for a while I found that the sounds began to become a part of the tool, and their use in this context overtook other associations. All the same the initial impression of the tool depends critically on the connotations of the sounds in the context of the other symbols. The final problem was the need to learn the sounds, which is counter to the requirement for a direct and immediate interface. The metonymic method aims at a direct display by taking advantage of sounds that are part of the design scenario. Although the sounds in this design could immediately answer the question “is this the same rock type as that?” it took several minutes of use to learn to answer the question “what rock type is this?”. It is difficult to know how to design a more direct display for categorical information. Auditory icons have a premise of directness, but it is not clear how this can be realised with this scenario involving rock-types. You can specify different materials such as wood, metal and glass in auditory icon algorithms, but these materials are no more a part of the problem domain than the other impact sounds that were tried, and so entail the same need for learning. Ballass found that up to 35 different sources may be attributed to the same everyday sound. He also found that the identification of discriminable but ambiguous everyday sounds depends critically on the expectancy and context of the listener [Ballass J.A. (1994)].

The problem that the answer associated with each sound takes a few minutes to learn can be overcome with verbal sounds. An experiment that compared the learnability of auditory icons, earcons and spoken words found that the time taken to associate interface actions with auditory icons and earcons was similar, but that speech feedback was significantly faster to learn and more reliable [Lucas P.A (1994)]. A verbal design was implemented by sampling a person saying each of the 4 rock-types, and storing the words as audio files.

granite = “granite”
limestone = “limestone”
marble = “marble”
shale = “shale”

Tapping the rocks in the interface produces the spoken response. The “what is it?” question is answered immediately, without any need for learning at all. It also answers the other questions just as well as the non-verbal design. New rock types can be added without introducing ambiguity because words are inherently discriminable. A large catalogue of rock-types could be represented this way, without the need for a long training period that is necessary for systems of more than 6 or 7 symbols [Patterson R.D. (1982)]. Although the verbal design is direct, it also has its problems. The verbal design is probably best suited to local information, since it may be difficult to understand more than a couple of words at the same time. Spoken words may interfere with other speech in the environment, where non-verbal sounds may not. There is some incongruity in that a talking rock requires a leap of imagination (transposition) that is expected in metaphors, but there are no other symbols in the interface to support this metaphor. Metonyms can seem more natural than metaphors because they are a part of the evoked object (contiguous). The words are semantically a part of the design scenario, but not the interaction in the interface. The words typically take about a second to say, and can slow the interaction with the interface down a lot. However speech can be time compressed by a factor of about 2 and still maintain intelligibility [Arons B. (1994)]. Time compression techniques could be used to equalise the durations of the words, and to set the durations to < 0.5s which is comparable with a short impact sound.

5.4.2 Metaphoric method

The metaphoric method is a way to represent the unfamiliar in terms of the familiar. The method works by looking up some EarBenders cases that are similar to the design scenarios in their information structure. The TaDa analysis for the GeoViewer, shown in Figure 5-8, was used to retrieve 3 EarBenders cases - Recycling, Walking and Kitchen, shown in Figure 5-11, Figure 5-12, Figure 5-13. Although the subject matter of each story is very different, they share a similar TaDa structure with the design scenario. The information required in the GeoViewer is the type of a rock strata that is being explored. The Recycling case describes how workers at a recycling depot classify garbage bags by listening to the sounds they make when kicked. The Walking case describes different sounds that footfalls make on different walking surfaces. The Kitchen story describes the identification of different cereals in opaque containers by shaking them.

Scenario

Title	Recycling
Storyteller	P.K.

STORY

My friend works at a recycling depot where the quickest; best way to see if a bag of garbage contains something interesting is to kick it. The combination of the weight of the bag and sound that it makes tells you what's inside. Normal household rubbish sounds a bit plasticy; bottles clink; crockery clatters; disposable nappies squelch. So a lot of the time her job involves wandering round kicking bags; and if it sounds interesting diving in and having a look.

KEYS

Question	whats inside this ?
Answers	household, bottles crockery, nappies
Subject	inside this
Sounds	plasticy, clink, clatter, squelch

TaDa Analysis

TASK	*Generic*	what is it ?
	Purpose	identify
	Mode	interactive
	Type	branching
	Style	exploration
INFO	*Reading*	direct
	Type	nominal
	Level	local
	Range	4
	Organisation	time
DATA	*Type*	nominal
	Range	unlimited
	Organisation	location

Figure 5-11: Recycling case

Scenario

Title	Walking
Storyteller	S.B.

STORY

walking home in the dark
footsteps make different sounds
as I walk over different surfaces
leaves are crackly
concrete is quiet thud
grass is muffled drag
gravel is crunchy

KEYS

Question	what surface am I walking on ?
Answers	leaves, concrete, grass, gravel
Subject	surface
Sounds	rustle, quiet thud, muffled drag, crunchy

TaDa Analysis

TASK	*Generic*	what is it ?
	Purpose	identify
	Mode	focus
	Type	discrete
	Style	exploration
INFO	*Reading*	direct
	Type	nominal
	Level	local
	Range	7+-2
	Organisation	time
DATA	*Type*	nominal
	Range	unlimited
	Organisation	location

Figure 5-12: Walking case

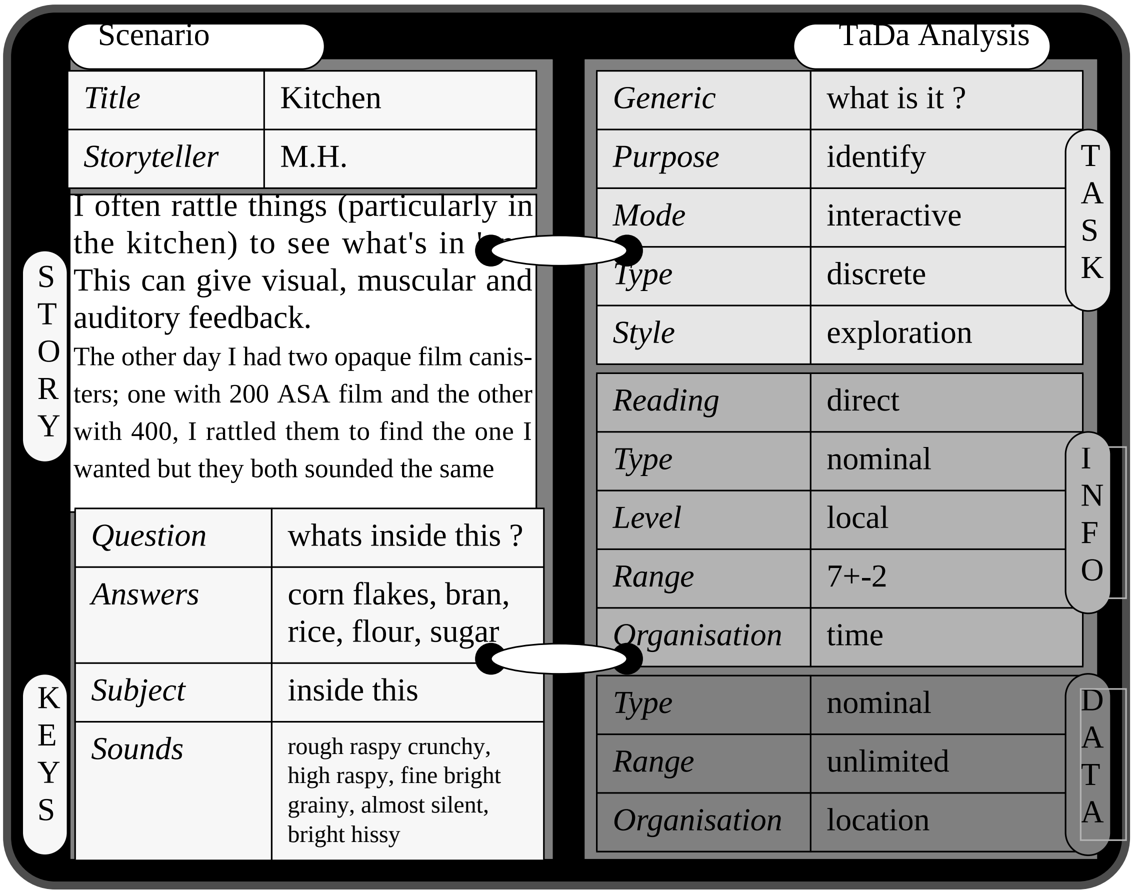


Figure 5-13: Kitchen case

Although the stories provide some interesting metaphors there is a problem. The metaphoric method works best when the suggestion of sounds that are part of the design scenario is weak or non-existent, allowing a leap in imagination. In the GeoViewer scenario the tapping of rocks is an easy sound to imagine, and it is hard to replace this with a metaphorical sound that has other connotations. The semantics of the metaphorical design need to be carefully considered in the context of the design scenario - for example garbage sounds might evoke pollution where none exists. Shaking cereals suggests a movement of the rocks which are stationary. Perhaps the footfalls on different surfaces has some potential to suggest walking over the rocks that could be a basis for a metaphorical design.

5.4.3 Pattern method

The pattern method is a search for trends in a set of solutions to similar problems that may capture consistencies in the mapping from the problem domain to the solution domain. The TaDa query from the GeoViewer scenario retrieved the cases shown in the previous section. The pattern method continues by scanning the auditory characterisations of these relevant cases for trends. The cases shared 6 of the 8 characteristics, shown by a score of 3 in the fields in Figure 5-14 - {everyday, local, single stream, discrete, stationary, category}. The other 2 fields had a 2/3 majority trend - {isolated and integral}. The auditory characterisations are very similar. However the actual sounds and descriptions of the sounds are quite different in each case. The Descriptors field contains 18 words that describe the sounds in the 3 cases. Only “crunchy” appears more than once - to describe both walking on gravel and shaking a container of cornflakes in the kitchen. The variety of descriptions reinforces the notion that it is the relations between sounds rather than the sounds themselves that are important in conveying information relations. The auditory design proceeds by selecting a set of sounds according to the characterisation synthesised from the trends. The actual choice of sounds has not been filled in Figure 5-14. This choice might be made with the metonymic or the metaphoric methods described in the previous sections. These methods consider the connotations of the auditory symbols with respect to the design problem. An alternative to these semantic approaches might be a structured palette of “abstract” sounds that do not have an everyday connotation that will interact with the connotations of the rest of the interface.

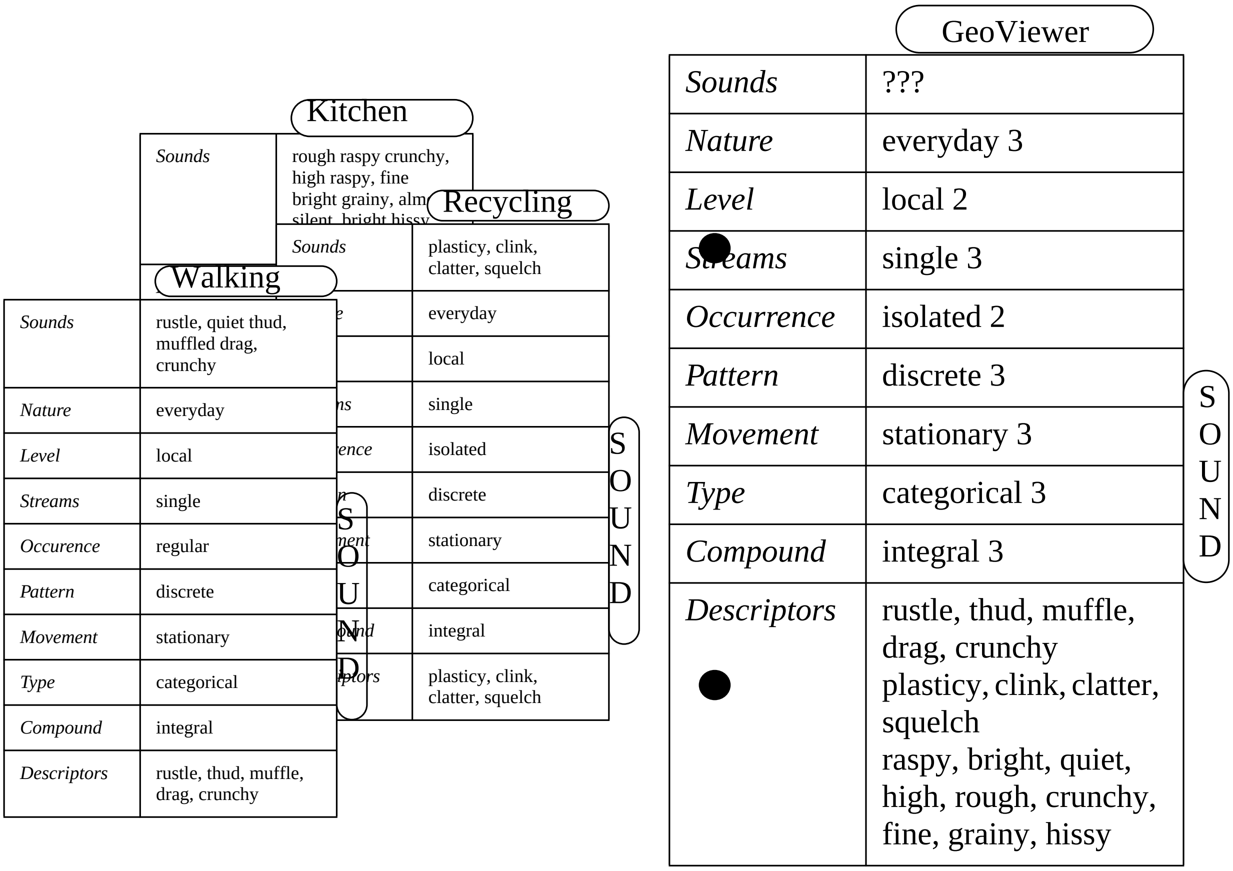


Figure 5-14: Pattern method for the GeoViewer

An example of an abstract palette might be the musical sounds from classical instruments. Although these obviously do have connotations, they are not usually associated with everyday tasks, and so may be more transportable between application scenarios. A palette of musical instrument samples was chosen in accordance with the auditory characterisation from the GeoViewer. However the Nature of the sound was changed from {everyday} to {musical}. The timbres of musical instruments are usually considered to be categorical, so 4 instruments with discriminable timbres were chosen. The pitch and duration of the sounds are the same, to prevent a misleading impression of order between the categories.

granite = tenor sax, pitch A3, duration 0.5s
limestone = cello, pitch A3, duration 0.5s
marble = English horn, pitch A3, duration 0.5s
shale = trombone, pitch A3, duration 0.5s

As with the metonymic designs, the answer to the question “is this the same rock type as that?” was immediate. However learning to answer the question “what rock-type is it?” took longer than it did with the metonymic palette of familiar impact sounds. Perhaps this was because the musical instrument timbres were much more similar to each other than the highly recognisable impacts, and I had to learn the timbral features of each instrument as well as the association with an answer. The musical palette has the advantage that the sounds can be much shorter than either the impact or the word designs. Just how short depends on the instrument. The instrumental sounds in Grey's study of timbre were 350 ms in duration. Hammered and plucked instruments can be identified from durations of less than 100 ms [Grey J.M. (1975)]. Many everyday sounds, such as a tap drip or clock tick, can be identified from durations between 50 and 100 ms [Ballass J.A. (1994)]. These sounds are an order of magnitude shorter than words.

5.5 Summary

HCI designers, bridge engineers and home renovators all use previous examples as a starting point for new designs. Case-based design is a top down method that can help in complex problems and areas of practice where few predictive principles exist. Auditory display is a developing field where case-based design can be helpful. However there are not enough examples of auditory display to draw upon as a case-base. In this chapter it was proposed that everyday uses of sounds could provide a pool of examples that could make case-based design practical in auditory display. A collection of nearly 200 stories about everyday listening experiences was gathered by email and conversation, and entered into a database called EarBenders. The stories were found to contain semiotic elements that could help in the design of an auditory display and 3 methods of case-based design were developed to take advantage of this resource - the metonymic method, the metaphoric method, and the pattern method.

The metonymic method searches the case-base for keywords that are a part of the design scenario. The aim is to design a palette of sounds that are a familiar part of the application domain, and may evoke that context for a listener. The metaphoric method represents the unfamiliar with the familiar. This is suited to scenarios where sounds do not naturally occur, which is often the case in computer-based information processing activities. EarBenders stories can be retrieved by similarity to the information structure of the design scenario, as described by a TaDa analysis. The retrieved stories are a source of metaphors that can support the information requirements of the design scenario. The pattern method captures regularities in the mapping from the design domain to the solution domain. EarBenders stories have an auditory characterisation that can describe the characteristics of an auditory display. The trends in the characterisations of stories retrieved with the TaDa characteristics of the design scenario were used to automatically synthesise an auditory specification. This specification describes the relations between sounds, but not the sounds themselves. The actual choice of sounds to meet the specification might use a metonymic or metaphoric palette, or perhaps an abstract palette of musical sounds that do not have everyday connotations.

Each case-based method of auditory information design was demonstrated in a mining visualisation scenario. The GeoViewer is a 3D interactive view of rock strata for mine planning and other geological applications. A problem is that it can be hard to tell what type of rock you are looking at, so an auditory display was designed to provide information about the rock-type as well. A rock strata that is difficult to visually identify can be heard, without having to divert visual attention to a text. An unexpected advantage became apparent when the interface was used. The sounds allow the front most surface at an intersection of strata to be disambiguated by tapping there, saving on a distracting and computationally expensive change of viewpoint operation. A development could provide information about the number and material of overlapping hidden layers. The demonstration showed that the case-based method is functional in practice, and can provide a good starting point for the design of a useful auditory display.

5.6 Further work

There is great potential to use the Internet to gather many more stories for the EarBenders case-base. Examples from different cultures could allow the possibility of culturally specific metaphors. For example the sounds of cooking in a kitchen in Australia may be quite different from the sounds in a kitchen in Japan.

6 • Hearsay: principles for auditory information design

Perception is the obtaining of information about the environment. [Gibson J.J. (1966)]

Information is a difference that makes a difference. [McAdams M. (1995)]

The EarBenders stories in the previous chapter suggest that natural sounds can convey many different types of information. We can hear information about the environment, but can we hear information in digital data? Evidence that we can comes from a formal investigation carried out by Sarah Bly in the early 1980's. Her investigations also found that the ability to correctly hear the information in the display depends critically upon the design of the mapping of the data into sound. Bly drew attention to the lack of a systematic approach for designing the data-to-sound mapping as a "gaping hole" impeding progress in this field of design practice [Bly S. (1994)].

This chapter proposes the Hearsay principles for auditory information design which may help bridge this gap. These principles can help a designer to meet the information requirements specified by the TaDa analysis. Hearsay integrates principles for information design with observations about auditory perception. Each Hearsay principle was investigated by generating a simple auditory demonstration to confirm that characteristic properties can be heard. The demonstrations show that the required information can be represented by auditory relations, and that the Hearsay principles are applicable in practice. The principles were tested in a design of an auditory display for Bly's 'dirt and gold' scenario. The display that was developed allows a listener to quickly answer the question "is there gold in this pile of dirt?". The effectiveness of this display indicates that the principles can be helpful in practice. This chapter is written in a tutorial-style and includes in-line code fragments for generating each demonstration with the freely available Csound audio synthesis software [Vercoe B. (1991)].

6.1 Advantages of design principles

Design principles can make specialist expertise easier to learn and apply in practice. Some advantages of design principles are listed below.

Making expert knowledge accessible

Principles, guidelines and rules are a way to capture and formalise knowledge. This knowledge can then be used to produce effective designs, without the need for every designer to have in-depth expertise in every aspect of the design problem, or have to design from first principles every time.

Generality

Principles are general observations that have many uses. For example principles of aerodynamics influence the design of bicycles, windmills, bridges and space shuttles.

Constrained guidance

Principles can help you to home-in on a good solution more quickly, by constraining or pruning the design space to a manageable size. A rule system can detect contradictions and exceptions during the design process. However there is the problem that formal methods may impair innovation by forcing all designs into the same mould.

Computer assistance

Rules can be programmed into a computer to deduce the outcome of a design decision, and support interactive simulations and feedback about a design. Computer aided (CAD) tools can allow the designer to work at the level of the problem, to focus on anomalies, innovation and creative synthesis of new solutions. A computer can support the design process by handling repetitive and difficult calculations.

6.2 Principles for information design

This chapter develops some principles for designing useful sounds. The TaDa analysis characterises the information required from the display. The information characteristics are the starting point for design. The information is characterised by five fields - reading, type, level, organisation and range (see Chapter 4). These fields can serve as anchor points for principles that couple the requirements to the representation.

In the quest for a principled approach to auditory display we can look to methods of graphic display that involve similar issues of representation. There has been a great deal of effort put into understanding how graphs can best show different types of information. This effort has resulted in the development of principles for graphic information design that have been broadly applied and found to be effective in practice. This approach to design has progressed to the point where rule-based computer tools can automatically construct a display from descriptions of the task and the data [MacKinlay J. (1986)], [Casner S. (1992)].

The principles of information display developed for graphic design may also be helpful in auditory information design. Some principles that have been consistently identified, and broadly applied, are linked with each of the TaDa information characteristics in Table 6-1.

Reading	*The most direct representation is the one with the shortest psychological description [Norman D.A. (1991)]*
Type	*An appropriate representation provides the information required by the task: neither more nor less [Norman D.A. (1991)]*

Table 6-1: Some principles for information design

Level	*The power of a graphical display is that it allows us to summarise general behaviour and at the same time to examine details [Cleveland W.S. (1985)]*
Organi-sation	*Useful information involves regrouping. The interactive reorganisation of relations between elements can uncover information in the interplay of the data [Bertin J. (1981)]*
Range	*Any undetectable element is useless. Utilise the entire range of variation [Bertin J. (1981)]*

Table 6-1: Some principles for information design

6.3 Principles for auditory design

The combination of principles for information design with observations about auditory perception may produce the systematic approach to auditory display called for by Bly. In the following sections each of the principles is related to auditory perception. A small auditory example of each principle is provided to confirm that the key characteristics can be heard, and to help get a feel for how it might apply in practice. The principles are then tested on a design of an auditory display for Bly's "dirt and gold" scenario. The examples and final display can be generated with Csound.

6.4 Principle of directness

The most direct representation is the one with the shortest psychological description. A less direct scheme exacts a penalty in mental workload that will become apparent by poorer performance under stress [Norman D.A. (1991)].

A direct representation can be understood with little training, can be understood almost immediately, and allows judgements which are not readily swayed by the opinions of others [Ware C. (1993)]. Some examples of direct representations are scatterplots, satellite images, and geiger counters. Conventional symbols, on the other hand, depend on learning or a legend to be understood. However they have the advantage that they may carry complex concepts built on layers of reference. Some examples of conventional representations are traffic signs, morse code, and hand gestures. These symbols are slow to read (several per second), and people can only keep about seven discrete items in short term memory, which may limit the operations that can be performed.

Different degrees of directness are demonstrated by the displays that may be generated from Table 6-2 and Table 6-3. The scenario is a mine rescue in which you might imagine that you have an instrument that measures the level of a dangerous gas as a reading between 0 and 9. The more direct display is a geiger counter-like granular synthesis where the density of grains increases with the level of gas. The other display is a morse code signal that taps out the level as a coded series of long and short tones - for example level 6 is "long, short, short, short, short". A short walk along the mineshaft is simulated by a set of instrument readings. Generate and listen to each walk-through. You can tell immediately when there is more gas with the geiger counter. The morse code is exact, but has to be looked up or learnt. It is suited to robust communications of more complex messages.

geiger.orc	geiger.sco
sr = 8000 ; kr = 800 ; ksmps = 10 ; instr 1 idens tablei p4*12.7, 3 ar grain idens*400, 600, idens, 1000, 100, 0.01, 2, 2, 0.03 out ar endin	f2 0 128 7 1 128 0 ; ramp 1 to 0 f3 0 128 -5 1 128 100 ; exponential 1 to 100 ; walk = 0 2 4 5 3 7 8 ;instr start dur reading 0-9 i1 0 1 0 i1 + . 2 i1 + . 4 i1 + . 5 i1 + . 3 i1 + . 7 i1 + . 8

Table 6-2: Example of direct representation

morse.orc	morse.sco
sr = 8000 kr = 800 ksmps = 10 instr 1 ion = (p4 > 0 ? 0.9*p3 : 0.1*p3) ioff = (p4 > 0 ? 0.1*p3 : 0.9*p3) kamp linseg 0, 0.01, 10000, ion, 10000, 0.01, 0, ioff-0.02, 0 aout buzz kamp, 440, 5, 1 out aout endin	f1 0 8193 10 1 ; sin t 0 300 ; tempo ; 1 .---- 2 ..--- 3 ...-- 4 ....- 5 ..... ; 6 -.... 7 --... 8 ---.. 9 ----. 0 ----- ; walk = 0 2 4 5 3 7 8 ;instr start dur bit ; 0 = ---- f0 6 ; each number is 5 beats + 1 gap i1 0 1 1 ; - i1 + . 1 ; - i1 + . 1 ; - i1 + . 1 ; - i1 + . 1 ; - s ; 2 = ..---, etc.

Table 6-3: Example of conventional representation

6.5 Principle of type

A representation should provide the information required by the task: neither more nor less [Norman D.A. (1991)].

If the task requires qualitative information then use a qualitative representation. If the task requires quantitative information then use a quantitative representation. For example the task of finding a country on a globe can be appropriately supported by colouring the countries in qualitatively different hues. If the task is to find the country with the highest rainfall then hues would make this difficult because large differences in hue do not look ordered and can't be compared.

The design of an appropriate representation requires a description of the information to be represented. The information types identified in the TaDa analysis are shown in Table 6-4.

boolean	2 qualitatively different categories
nominal	qualitative difference without order
ordinal	qualitative difference with order
ordinal-with-zero	qualitative difference, with order and a zero
ordinal-bilateral	qualitative difference, with order and a central zero
interval	quantitative difference, with order and metric
ratio	quantitative difference, with order, metric, and a zero

Table 6-4: TaDa Information types

The information types are further characterised by elementary relations of difference, order, metric and zero that are the building blocks of more complex information structures. Order is a directed difference, which might be expressed as more or less, or low and high. Metric is an equal unit of difference that is consistent, for example a 1 degree rise in temperature is the same no matter what the current temperature is. Zero is a point of correspondence between all scales independent of unit, so for example zero rainfall is the same whether your rain gauge is in mm or inches. The elementary information relations are described in Table 6-5.

difference	qualitative or quantitative
order	directed difference
metric	equal units of difference throughout the range of variation
zero	a point of correspondence between all scales independent of unit

Table 6-5: Elementary information relations

The information building blocks can be aligned with perceptual building blocks that have similar properties. Gibson describes perception as the obtaining of information about the environment from higher order invariants such as stimulus energy, ratios and proportions [Gibson J.J. (1966)]. Sebba found that subjects made consistent judgements of similarity between music and colour structure due to correspondences in perceived order, contrast and ratios [Sebba R. (1991)]. A characterisation of elementary perceptual relations is shown in Table 6-6.

difference	all perceptual elements are detectably different
order	the perceptual elements have a discernable order
metric	there is a unit of equal perceptual difference
zero	an absolute point of reference for variations at any scale

Table 6-6: Elementary perceptual relations

To recap - we started with a description of TaDa information types which we characterised by lower level building blocks of difference, order, metric and zero. These building blocks were then used to describe perceptual relations with similar properties. This gives us the opportunity to construct a faithful mapping of an information type to a perceptual representation. If we can realise the perceptual building blocks as auditory relations then the process can be used to design faithful auditory representations for a general range of information types. The following examples demonstrate that the perceptual building blocks can be heard in auditory relations, as required by this process.

6.5.1 Difference

Qualitatively different sounds can be easily generated by plugging in parameters to a synthesis instrument. The demonstration is an FM instrument and a score that generates a sequence of three different sounds, shown in Table 6-7. You can hear that the sounds are different by listening to the sequence in a loop. It can take some fiddling with parameters to ensure that the sounds are more than a little different from one another. When the sounds are not very different you can hear a double sound in the loop. As you adjust the parameters you may get a feel for the folds, flat regions and non-linearities in the mapping from synthesis parameters to perceived sounds. This can be a problem if a display simply connects data values to synthesis parameters, because some differences may be undetectable, and others may be exaggerated.

fm.orc	diff.sco
sr = 8000 kr = 800 ksmps = 10 gir = 100 gis = 1000 instr 1 kamp linen 10000, 0.01, p3, 0.1 ao0 oscili gir, gis*p4, 1 ao1 oscili gir, gis*p5+ao0, 1 ao2 oscili gir, gis*p6+ao1, 1 ao3 oscili gir, gis*p7+ao2, 1 ao4 oscili gir, gis*p8+ao3, 1 aout oscili kamp, gis*p9+ao4, 1 out aout endin	f1 0 8193 10 1 ; sin t 0 60 ; 1 beat per second ;instr start dur p4 p5 p6 p7 p8 p9 i1 0 1 0.333 0.692 0.176 0.138 0.354 0.058 i. + . 1.028 0.576 0.070 0.077 2.401 0.162 i. + . 0.217 0.885 0.259 1.087 0.739 1.005

Table 6-7: Auditory difference

6.5.2 Order

Changes in a sound are sometimes described by words like buzziness, or squelchiness or heaviness that indicate a degree of order in the sounds. The demonstration, shown in Table 6-8, is a vibrato at three different rates.

vibrato.orc	order.sco
sr = 8000 kr = 800 ksmps = 10 instr 1 k1 oscil 50,p4*10,1 aout pluck 10000,220+k1,220,0,1 out aout endin	f1 0 8193 10 1 ; sin ;instr start dur parameter 0.0-1.0 i1 0 1 0.67 i1 + . 0.10 i1 + . 0.23

Table 6-8: Auditory order

When you listen to the sequence in a loop it is easy to hear an ordered change in the sound. Try to pick the middle sound, the one that lies between the other two in the amount of vibrato. Even if the middle sound is closer to one side or the other the sequence is ordered.

6.5.3 Metric

A metric variation has a unit of equal perceptual difference. This can be heard by a unit step in difference no matter where the step occurs. The units that are available by default in Csound are semitones and decibels. The example in Table 6-9 demonstrates equal steps in pitch. The size of the steps can be heard by listening to the sequence in a loop. The difference between the middle sound and those on either side should seem equal.

pluck.orc	metric.sco
sr = 8000 kr = 800 ksmps = 10 instr 1 icps = cpsoct(6.0+p4) aout pluck 10000,icps,icps,0,1 out aout endin	f1 0 8193 10 1 ; sin ;instr start dur parameter 0.0-1.0 i1 0 1 0.0 i1 + . > i1 + . 1.0

Table 6-9: Auditory metric

6.5.4 Zero

A perceptual zero can be detected no matter where it occurs in a sequence, and no matter what the scale of variation. There are three types of zero that can be listened for

- a natural zero where the sound disappears altogether
- an original zero where an observable aspect of the sound disappears
- a conventional zero, such as middle c, which may be compared against a reference, or perhaps learnt

The natural zero is demonstrated by varying the density of grains in the geiger counter, from Table 6-2, with nat0.sco from Table 6-10. Generate the sequence and listen to it in a loop. As the density goes to zero the sound disappears. The zero can be detected anywhere in the sequence and at any scale of variation.

nat0.sco for geiger.orc	orig0.sco for vibrato.orc
f1 0 8193 10 1 ; sin ;instr start dur parameter 0-9 i1 0 1 4 i1 + . 0 i1 + . 8	f1 0 8193 10 1 ; sin ;instr start dur parameter 0.0-1.0 i1 0 1 0.4 i1 + . 0.0 i1 + . 0.8

Table 6-10: Auditory zeros

The original zero is demonstrated with the vibrato instrument, from Table 6-8, and orig0.sco from Table 6-10. The vibrato disappears at both the lower and upper extremes of its range, although the sound remains. The lower point is the zero because timbre of an upper extrema can be heard to change with scaling.

6.5.5 An elementary characterisation of some sounds

There are a many auditory variations which we might harness in an auditory display. These include everyday sounds, musical sounds, synthetic sounds, vocal sounds, and verbal sounds. Some sounds of each of these types have been characterised in terms of the elementary relations of difference, order, metric and zero, as shown in Table 6-11. The table shows how some sounds can be described in these terms, but is not meant to be definitive or complete. The characterisation of sounds in this way can help select an appropriate representation for a display element.

Sound relation	*Difference* qualitative/ quantitative	*Order* 1D, 2D, 3D, nD	*Metric* ratio/ difference unit	*Zero* natural original conventional
door knocks	qualitative	-	-	-
object material	qualitative	-	-	-
event type	qualitative	-	-	-
rhythm	qualitative	-	-	-
harmonicity	qualitative	-	-	-
tune	qualitative	-	-	-
musical key	qualitative	-	-	-
phasor pattern	qualitative	-	-	-
binaural cohesion	qualitative	-	-	-
temporal order hiss, tone, buzz, 'ee'	qualitative	-	-	-
vowels a,e,i,o,u	qualitative	-	-	-
animals moo, woof, meow, baa	qualitative	-	-	-
formants	qualitative	nD	-	-
timbre	qualitative	nD	difference MDS	-
squeakiness	qualitative	1D	-	natural
flapping	qualitative	1D	-	natural
popcorn popping	qualitative	1D	-	natural

Table 6-11: Elementary characterisation of some sounds

Sound relation	*Difference* qualitative/ quantitative	*Order* 1D, 2D, 3D, nD	*Metric* ratio/ difference unit	*Zero* natural original conventional
music tempo	qualitative	1D	-	natural
machine rate	qualitative	1D	-	natural
machine work	qualitative	1D	-	conventional
pitch class	qualitative	1D	difference Semitone	conventional
event force	quantitative	1D	-	natural
drum stretch	quantitative	1D	-	original
fuzz level	quantitative	1D	-	original
reverb wetness	quantitative	1D	-	original
vibrato rate	quantitative	1D	-	original
vibrato depth	quantitative	1D	-	original
tremolo rate	quantitative	1D	-	original
tremolo depth	quantitative	1D	-	original
phasor depth	quantitative	1D	-	original
phasor rate	quantitative	1D	-	original
brightness	quantitative	1D	ratio Acum	original
object size	quantitative	1D	-	conventional
filling a bottle	quantitative	1D	-	conventional
rolling marble	quantitative	2D	-	conventional
granular density	quantitative	1D	-	conventional
pitch scale	quantitative	1D	difference Semitone ratio Mel	conventional
repetition rate	quantitative	1D	ratio B = 1.0	natural
white noise duration	quantitative	1D	ratio B = 1.1	natural
binaural loudness	quantitative	1D	ratio B = 0.6	natural
monaural loudness	quantitative	1D	ratio B = 0.54	natural

Table 6-11: Elementary characterisation of some sounds

6.6 Principle of level

The power of a graphical display is that it allows us to summarise general behaviour and at the same time to examine details [Cleveland W.S. (1985)].

Higher level information is contained in the groupings, clusters, trends, correlations, outliers and other relations between data elements. The level of the display can be determined by the level of question that can be immediately answered from the display, as shown in Table 6-12. A poor display can only answer questions about individual elements [Bertin J. (1981)].

local	can answer questions about a single element
intermediate	can answer questions about subsets and groups of elements
global	can answer questions about the entire set of elements as a whole

Table 6-12: Levels of information

The different levels of information defined by Bertin correspond well with the description of acoustic grouping used by Bregman in his theory of auditory scene analysis [Bregman A.S. (1990)]. This theory has two levels of listening processes- a global level of overall analysis, and a local level of attention to details. These processes group and segregate acoustic elements into coherent sounds or "streams". Levels of information may be linked with levels of auditory scene analysis as shown in Table 6-13.

local	answered by listening to an element within a stream
intermediate	answered by listening to a stream
global	answered by listening to an auditory scene

Table 6-13: Levels of auditory information

Bregman first noticed streams whilst investigating phonemes in speech. He wondered whether non-speech sounds would hold together the way spoken phonemes do. To find out he concatenated 26 sounds such as water splashing in a sink, a doorbell, and a dentist's drill to form a sentence of mock phonemes. He found that the sentence did not sound like speech at all, even when it was played at speech rates of 10 phonemes per second. Parts of the sentence seemed to pop-out, and the order of the phonemes was disrupted. We can repeat this experiment with the FM sounds from Table 6-7. Play the looped sequence at the slow rate of 1 sound per second and notice that you can easily write down the order of the three sounds. Now speed it up to 10 sounds per second by changing the tempo in diff.sco from 60 beats per minute (t 0 60) to 600 beats per minute (t 0 600). Play the loop and try to write down the order again - this time it will be very difficult to tell which sound comes after which. This is because the sounds have segregated into different auditory streams. The segregation of elements into streams can make simple tasks like counting much harder. Some consequences of streaming for auditory display are

- streams are categorical and exclusive
- judgements involving elements in the same stream are easy
- judgements involving elements in different streams are difficult

An understanding of the factors that influence streams can guide the design of a higher level display. As mentioned earlier, there is a global level and a local level. The global level, or primitive process, is a default bottom-up grouping by acoustic factors such as spectral similarity. The local level, or schema process, allows the listener to alter the default grouping by mental effort. Mental schemas detect familiar acoustic patterns and draw attention to them. Schemas are a top-down process that explains why what we hear depends so much on attention and previous listening experience. In this view, the characteristics of sounds are calculated from streams, not directly from the acoustic array. This is very different from a straight forward signal processing model of auditory perception.

6.6.1 Primitive grouping in auditory displays

The factors that influence the primitive process group operate sequentially in time, and simultaneously across the spectrum. The perception of a new sound depends on the streams that exist when it is introduced. Parts that are acoustically similar to an existing stream will be grouped with it, leaving the residue to be heard as the new sound. This is called the old-plus-new heuristic by Bregman, and it implies that the sequential factors that capture recent auditory context tend to override the moment-to-moment spectral factors. A listing of factors in order of influence can be made from results of experiments that have placed various factors in competition. The sequential factors are toward the top of the list. This listing may provide a basis for controlling the primitive grouping of elements in a higher level auditory display.

- temporal rate measured by the separation of onsets in the range 60-150 ms
- the difference between spectral centroids
- difference in fundamental frequency in the range 4-13 semitones
- binaural harmonic correlation
- correlated frequency modulations
- correlated amplitude modulations
- harmonic relations
- parallel spectral movement
- synchronous onsets

6.6.2 Schemas in auditory displays

Schemas are important in auditory design because attention and previous learning have a marked influence on what is heard. We can take advantage of familiar patterns to improve the detection of information elements, and to improve the coherence of information in a mixture. The semantics of familiar sounds can also be used to improve the interpretation of the display in a particular task - for example rain sounds can be easily related to rainfall records. Some consequences of schemas in auditory display include

- improved coherence and separation of figure from ground
- the selection of streams and material from streams
- recognition of familiar patterns
- restoration of hidden material

The effects of a schema can be demonstrated by the restoration of a damaged tune. Generate the example from the orchestra and scores in Table 6-14, and listen to the sequence in a loop. Can you identify the tune, despite all the noise and interference? Some notes are actually missing, can you tell which ones? To find out uncomment the “e” in schema.sco so that the noises aren't included when you generate the sequence again. The interesting thing is that the noises don't provide any acoustic clues, yet you can hear the correct notes! If you can't hear the tune then it may be because "3 blind mice" isn't familiar, and you could try again with a tune that is.

<table>
<tr><th>schema.orc</th><th>schema.sco</th></tr>
<tr><td>
<pre>
sr =   8000
kr =   800
ksmps = 10

; tone
instr 1
kamp   linen   3000, 0.05, p3, 0.01
aout   oscil   kamp, cpspch(p4), 1
out   aout
endin

; noise
instr 2
kamp   linen   5000, 0.01, p3, 0.01
i1 =   cpspch(p4)
k1   rand   i1
aout   oscil   kamp, k1+i1, 1
out   aout
endin
</pre>
</td><td>
<pre>
f1 0 8193 10 1   ; sin
t 0 200
; tune with missing notes
i1   0   1   8.04
i1   1   1   8.02
i1   2   2   8.00
i1   4   1   8.04
i1   6   2   8.00
i1   9   0.5   8.05
i1   9.5   0.5   8.05
i1   12   1   8.07
i1   13.5   0.5   8.05
i1   14   2   8.04
; uncomment the next line to remove noise bursts
;e
i2   3.5   1   8.00   ; interference
i2   5   1   8.00   ; missing
i2   8   1   8.00   ; missing
i2   9   0.5   8.00   ; interference
i2   10   2   8.00   ; missing
i2   13   0.5   8.00   ; missing
</pre>
</td></tr>
</table>

Table 6-14: Restoration of a tune by a schema

6.7 Principle of organisation

Useful information involves regrouping. The interactive reorganisation of relations between elements can uncover information in the interplay of the data [Bertin J. (1981)].

Bertin demonstrated the way regrouping alters the information in a display by moving cards with simple marks, such as spots of different sizes, around on a table. In one example he transcribes the length of stay of hotel guests on these cards then physically reorganises them on the table to show different information about peak booking periods which was not evident in the original graph. The useful information does not correspond with the values of the individual elements but with the structures formed by the interplay of these elements with each other as a whole. Only the spatial organisation of the elements was permuted, not any of the other visual variables such as lightness or size. This is because you can only see two distinct cards if they have different positions on the table, or are in the same place at different times. This is why space and time are called the "indispensable" dimensions of a visual display. Elements that use-up an indispensable dimension constrict the options for permutation. For example a time-series plot uses-up the horizontal dimension, leaving only the vertical for permutation. A map uses-up both the horizontal and vertical dimensions and so cannot be permuted.

Streaming experiments have shown two sounds can occupy the same space and time but still be heard as separate identities when they occupy different parts of the spectrum. It seems that the auditory display designer has a great deal of freedom to organise and reorganise display elements. The degree of freedom depends on the capabilities of the display device. A display that uses Csound may employ multiple synthesis parameters to reorganise spectral relations. The score events can be organised to separate elements in time. A quadraphonic pan is available that can provide a degree of separation in space. Interactive permutation and exploration is supported by real-time input sensors. A limiting factor is the amount of computation required to generate the sounds in real time. Any apparent lag in reaction can compromise the usability of an interactive display. One way to address this problem is to design the synthesis to be as computationally simple as possible. Another way is to take advantage of fast hardware for audio synthesis.

6.8 Principle of range

Any undetectable element is useless. Utilise the entire range of variation [Bertin J. (1981)].

The number of elements that can be differentiated in a display depends on the range of perceptual variation available on the display. Most people can't hear the pitch of frequencies below about 80 Hz in which case human hearing is the limiting factor. Some devices can't play frequencies above 4 kHz, in which case the device is the limiting factor. The range of perceptual variation on a device is called the display gamut. The knowledge of a gamut allows the designer to optimise the display for the device. A transportable display must be designed to fit in the intersection of the gamuts of the target devices. The effect of available range is demonstrated by the orchestra and score in Table 12. The sequence is 4 levels of rainfall {none, light, medium, heavy} mapped to loudnesses (0, 40,60,80} dB. This demonstration assumes that you can easily change the loudness setting of your audio equipment. Generate the sequence and turn the loudness knob down low to avoid the risk of an uncomfortably loud sound. Start playing the sequence in a loop and adjust the loudness knob to a comfortable level. Can you hear the 3 sounds that correspond with light, medium and heavy rain? You might notice that the equal differences in dB of loudness do not really sound equal at all, an observation that lead Stevens to propose a new law of psychophysics which states that psychological judgements are based on ratios in stimulus, rather than differences [Stevens S.S. (1962)]. Turn the loudness knob down further until the light rain can no longer be heard. This display can no longer provide the required information. A display that relies on loudness will need to be calibrated to ensure the that all the elements are discriminable. Some attributes, like duration are not so susceptible to device characteristics, and may be a better option for a portable display.

range.orc	range.sco
sr = 8000 kr = 800 ksmps = 10 instr 1 kamp linen ampdb(20+p4*20), 0.01, p3, 0.01 k1 rand 1600 aout oscil kamp, 400+k1, 1 out aout endin	f1 0 8193 10 1 ; sin ;instr start dur answer i1 0 1 1 ; light i1 + . 2 ; medium i1 + . 3 ; heavy

Table 6-15: Auditory range

6.9 Putting the principles to work

So far we've borrowed some principles of information design from graphic design, and tried some simple demonstrations to get a feel that they may be of benefit in the design of auditory displays as well. The next step is to try them out in the design of an actual display. In this instance I have chosen Bly's dirt and gold scenario because it is a good example of the type of activity where sounds may provide information that is difficult to obtain visually, and it is a reference for other investigators. Here is the problem scenario described by Bly [Bly S. (1994)]...

Can you find the gold ?" It is hypothesised that six different aspects of the land in which gold may be found are determinative of whether or not gold is there. The first 20 data variables (each 6-d) are from sites known to have gold; the second 20 data variables are from sites known not to have gold. For each of the remaining 10 data variables, decide for each whether or not it is from a site with gold.
The data set consists of 100 samples generated from a normal random deviate generator and then separated into two distinct sets. Only samples in which all six variables had positive values between 0.0 and 3.5 were included. A sample s=(x21,x2,x3,x4,x5,x6) belongs to Set 2 (dirt) if

$x2*x2+x3*x3+x4*x4+x5*x5+x6*x6 <= 1.5*1.5$ *or*
$x1*x1+x3*x3+x4*x4+x5*x5+x6*x6 <= 1.5*1.5$ *or*
$x1*x1+x2*x2+x4*x4+x5*x5+x6*x6 <= 1.5*1.5$

At least five of the six variables in each sample of Set 2 have a value less than 1.5 and at most one of x1,x2,x3 have a value greater than 1.5. The 2 sets are completely distinct only in six-space; any representation in fewer dimensions will overlap Set 1 and Set 2.

Before we begin the design we must be clear about the information that is required by the task, and the data elements that are involved. The TaDa analysis, shown in Figure 6-1, begins by recasting the scenario as a question, based on the observation that useful information is the answer to a question. This scenario is already summarised by a question - "Can you find the gold?", which has the useful answers {yes or no}. However the participants answered this question one sample at a time - indicating that the actual question was more like "is this sample gold?". The purpose of this question is to identify whether a particular sample is gold or not, which is a local question. The display should allow the interactive selection of information elements from the display. The analyst does not know what to expect so the task is explorative. The level of information is local because it only involves one sample at a time. The answers {yes or no} are boolean categories. The answer that may be given is not affected by prior answers or other concurrent answers - so does not depend on other answers. Likewise, the samples have no intrinsic organisation in space or time. The soil samples are comprised of 6 independent ratio data measurements that range between 0.0 and 3.0.

Scenario

Title	Gold
Storyteller	Sarah Bly

STORY

Can you find the gold ?" It is hypothesised that six different aspects of the land in which gold may be found are determinative of whether or not gold is there. The first 20 data variables (each 6-d) are from sites known to have gold; the second 20 data variables are from sites known not to have gold. For each of the remaining 10 data variables, decide for each whether or not it is from a site with gold

KEYS

Question	is this sample gold ?
Answers	yes, no
Subject	soil samples
Sounds	???

TaDa Analysis

TASK

Generic	is this it ?
Purpose	identify
Mode	interactive
Type	discrete/procedure
Style	exploration

INFO

Reading	direct
Type	boolean
Level	local
Organisation	none
Range	2

DATA

Type	6d ratio
Range	0.0-3.0
Organisation	none

Figure 6-1: TaDa analysis for the Gold scenario

Having analysed the information requirements we can set about the design of a useful display from principles of reading, type, level, organisation and range. A direct reading will allow the listener to answer quickly, correctly and confidently. The first step in the design of the direct display is an appropriate representation of the display elements. Each sample is comprised of 6 ratio variables characterised by quantitative difference, order, metric and a zero. We could map these to perceptual building blocks with the same characteristics. However it is not the individual measurements that are of interest, but the samples themselves. The appropriate display is one that allows the difference between gold and dirt to be heard by an auditory difference, not one that allows us to hear measurements that are part of the samples. Hence the display elements should be cohesive or integrated wholes that show difference at the level of the samples. In the demonstration of the principle of difference we found that the FM algorithm in Table 6-7 could produce different sounds from 6 auditory parameters. Of course there are many other synthesis algorithms that could also be tried, the important point is that they have 6 parameters that all cause perceptual difference. It is likely that not all the parameters can be heard to cause a difference throughout the range, preventing discrimination in some regions of the display. However we are not really interested in telling gold from gold, just gold from dirt, so flat spots and even folds are ok as long as they don’t cross the boundary between yes and no. This is a case where the information characterisation has significantly reduced the complexity of the display. If the requirement had involved the separation of different types of gold then the perceptual metric of the space would have become a major concern. As it is we can make a simple first pass by maximising the discrimination of representative pairs of gold and dirt. The FM instrument may be calibrated with a global scaling factor gis, or individual factors may also be adjusted. Some representative pairs to tune the range against are in rangeok.sco shown in Table 6-16. Different pairs are selected by uncommenting them in the score file.

rangeok.sco - goes with fm.orc	identity
; uncomment dirt and gold pairs ;i1 0 10.554 1.232 0.074 0.198 0.358 0.065 i1 0 10.292 0.699 0.076 0.153 0.303 0.064 ;i1 0 10.462 1.383 0.047 1.059 0.344 0.798 ;i1 1 10.130 0.096 1.686 0.108 1.020 1.198 i1 1 10.886 0.366 0.570 1.571 2.040 1.357 ;i1 1 11.224 1.481 1.835 0.318 0.510 1.500	 dirt01 dirt02 dirt03 gold01 gold02 gold03

Table 6-16: Range calibration

After listening to some representative pairs I found that it is easy to tell the pairs of sounds apart with this display, which indicates that the range is adequate. It is also easy to tell different gold samples apart, which is more information than we need to answer the question. The test of the effectiveness of the display is the ability of the listener to answer the question “is this sample gold?” quickly, correctly and confidently. This design was put to the test in an experiment modelled on Bly’s investigation of different designs for the dirt and gold scenario [Bly S. (1994)]. Her experiment involved the testing of three displays on participants at the First ICAD conference. The first display was a granular synthesis, the second a mapping to balance, timbre, sustain, pitch, duration and volume, and the third was a mapping of the sum of squares to pitch. The effectiveness of each was measured by the number of listeners who could correctly identify the samples more than half the time. The results varied widely, as shown in Table 6-17, with 50%, 75% and 95% of the respondents correctly identifying more than half of the test samples in each case. The superior results of the third mapping rested on a preliminary data analysis that showed that the sum of squares was a reasonable classifier for this data. A vocalisation of “yes” to the gold and “no” otherwise could even have been used in this display.

#Correct	10	9	8	7	6	5	4	3	2	1	0
icad1 #subjects	0	0	0	3	3	7	3	5	-3?	0	0
icad2 #subjects	0	1	6	7	4	4	2	0	0	0	0
icad3 #subjects	3	7	7	5	-1?	1	0	0	0	0	0

Table 6-17: Results of 3 different designs and 24 subjects

The ‘dirt and gold’ experiment provides a standard for comparing the effectiveness of different designs. I repeated the experiment for the FM gold detector that has been developed in this chapter so far. The aim of the experiment was to find out whether the principles were helping to produce an effective display. The subjects were 27 unpaid volunteers of both sexes, between 20 and 50 years of age, working as administration, engineering and research staff at CSIRO Mathematics and Information Sciences. The procedure was altered so that each person classified all three test sets with the same mapping, in isolation, without time constraints. The results, in Table 6-18, show that 100%, 96% and 80% of subjects correctly identified more than half the samples. The ability of the listeners to obtain the required information from this display compares favourably with the ICAD designs, as shown in Figure 6-2, indicating that the principles are of benefit in practice.

#Correct	10	9	8	7	6	5	4	3	2	1	0
gold1 #subjects	1	7	10	6	3	0	0	0	0	0	0
gold2 #subjects	1	7	8	6	2	1	0	1	0	0	0
gold3 #subjects	1	4	4	6	7	3	1	1	0	0	0

Table 6-18: Results for the FM gold detector with 27 subjects

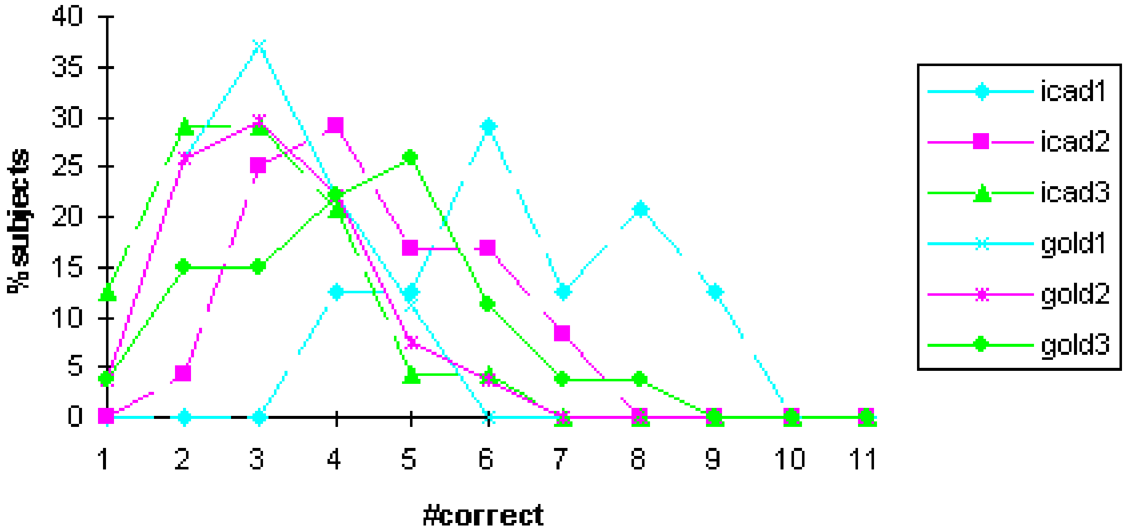


Figure 6-2: Results from the ICAD designs and the Fm gold detector.

6.9.1 A more useful high level display

Although the FM gold detector is an effective display, it only answers the local question “is this sample gold?” A more useful higher level display would enable the listener to quickly and directly answer the question “is there any gold in all this dirt?”. The principle of level raised the possibility of streaming as a way to construct a higher order auditory display. In the demonstration we found that the FM sounds tend to segregate if they are played at speech rates of 10 per second. Perhaps we can calibrate the FM parameters so that gold tends to segregate from the dirt. This can be done with the sequence of dirt-gold-dirt-silence, shown in levelok.sco in Table 6-19. If you listen to this sequence in a loop you will hear the gold as a high stream that is distinctly separate from the low stream made of the two different dirt samples. The gold clearly segregates for these samples, but should this not be the case then the mapping parameters may be tuned to increase the segregation.

levelok.sco - goes with fm.orc	identity
f1 0 8193 10 1 ; sin t 0 600 ; tempo f0 4 ; 1 beat of silence at the end i1 0 1 0.554 1.232 0.074 0.198 0.358 0.065 i1 1 1 0.130 0.096 1.686 0.108 1.020 1.198 i1 2 1 0.292 0.699 0.076 0.153 0.303 0.064	dirt01 gold01 dirt02

Table 6-19: Level calibration

According to the principle of organisation we may also take advantage of spare indispensable dimensions to improve the separation of the answers. At the moment we are only using the spectral dimension of the display. However the samples have no intrinsic order in space or time, leaving these indispensable dimensions spare. The Csound pan command can provide spatial separation for two data measurements. The temporal separation is not so easy in Csound because a start and duration for each event has to be specified up front in the score, and cannot be varied by the measurement fields. My solution was to copy one of the measurements into the start field. The result is a redundant mapping of 3 of the 6 measurements in the samples to spare indispensable dimensions which may improve the discrimination between yes and no. The display is presented as a Csound instrument called gold.orc in Table 6-20. You can listen to a handful of dirt and a handful of gold (which are in fact Bly's training sets) with the scores in Table 6-21. Listen to each to get a feel for how some typical dirt and gold samples sound.

gold.orc

```
sr =        8000
kr =        800
ksmps = 10

nchnls = 2
gir = 100
gis = 1000
instr 1
kamp    linen   10000, 0.01, p3, 0.1
ao0     oscili gir, gis*p4,        1
ao1     oscili gir, gis*p5+ao0, 1
ao2     oscili gir, gis*p6+ao1, 1
ao3     oscili gir, gis*p7+ao2, 1
ao4     oscili gir, gis*p8+ao3, 1
ao5     oscili kamp, gis*p9+ao4, 1
; separate d3 and d4 in space
a1,a2,a3,a4   pan ao5, p6, p7, 2, 1, 1
outs a1, a2
endin
```

Table 6-20: Gold detector

dirt.sco - 20 dirt samples	gold.sco - 20 gold samples
f1 0 8193 10 1 ; sin	f1 0 8193 10 1 ; sin
f2 0 129 7 0 128 1; ramp 0 to 1	f2 0 129 7 0 128 1; ramp 0 to 1
t 0 200 ; tempo	t 0 200 ; tempo
; d1 d2 d1 d2 d3 d4 d5 d6	; d1 d2 d1 d2 d3 d4 d5 d6
i1 0.830 1 0.830 0.812 0.170 0.064 0.417 0.897	i1 1.394 1 1.394 0.775 0.080 0.446 0.541 1.631
i1 0.170 1 0.170 1.062 0.085 0.716 0.080 0.871	i1 0.807 1 0.807 1.411 0.515 0.851 0.681 0.749
i1 0.671 1 0.671 0.237 0.380 0.689 0.899 0.600	i1 1.355 1 1.355 0.099 0.235 2.054 0.656 2.127
i1 2.380 1 2.380 0.690 0.355 0.006 0.691 0.437	i1 0.498 1 0.498 2.482 0.370 0.743 1.580 0.415
i1 0.551 1 0.551 0.377 0.007 0.112 0.051 0.246	i1 1.197 1 1.197 0.411 3.198 0.370 1.241 0.373
i1 0.242 1 0.242 1.417 0.275 0.405 0.590 0.810	i1 1.955 1 1.955 0.433 1.390 0.067 1.314 0.858
i1 0.404 1 0.404 1.750 0.631 0.247 0.060 0.767	i1 0.330 1 0.330 0.572 0.815 0.574 0.778 1.606
i1 0.443 1 0.443 1.472 0.297 1.050 0.046 0.455	i1 2.019 1 2.019 0.317 0.737 1.144 0.193 0.705
i1 0.180 1 0.180 1.112 0.245 0.620 0.782 0.547	i1 0.996 1 0.996 0.426 0.613 1.612 0.157 0.711
i1 0.007 1 0.007 0.286 1.223 1.339 0.300 0.475	i1 0.775 1 0.775 0.359 0.589 0.874 0.281 1.618
i1 0.640 1 0.640 0.393 0.642 0.941 0.670 0.123	i1 1.335 1 1.335 0.280 1.014 0.407 2.072 0.138
i1 0.737 1 0.737 0.427 0.796 0.723 0.194 0.149	i1 0.251 1 0.251 2.937 0.902 0.905 1.296 0.118
i1 0.565 1 0.565 0.571 0.005 0.340 0.507 0.168	i1 1.182 1 1.182 0.186 0.503 1.302 0.807 1.188
i1 0.976 1 0.976 0.342 0.090 0.356 0.532 0.461	i1 0.782 1 0.782 0.134 1.671 2.251 2.005 0.678
i1 1.165 1 1.165 0.510 0.212 0.900 0.736 0.574	i1 1.206 1 1.206 1.025 0.267 0.584 1.056 0.116
i1 0.821 1 0.821 0.291 1.185 0.465 0.061 0.392	i1 0.674 1 0.674 0.867 0.567 1.283 1.531 1.046
i1 2.195 1 2.195 0.776 1.027 0.510 0.517 0.135	i1 0.207 1 0.207 0.244 0.969 0.523 1.060 1.342
i1 0.430 1 0.430 1.427 0.183 0.050 0.370 0.563	i1 0.907 1 0.907 1.190 0.834 0.281 1.129 0.029
i1 0.243 1 0.243 0.717 0.531 0.597 0.376 1.170	i1 0.453 1 0.453 0.733 1.138 0.840 1.078 0.324
i1 0.042 1 0.042 0.199 0.745 0.364 0.399 1.086	i1 0.086 1 0.086 0.699 0.167 1.241 0.980 0.222

Table 6-21: Handfuls of dirt and gold

Then try listening to the mixture of dirt and gold from Bly's first test listed in goldtest1.sco in Table 6-22. You can immediately tell that there is gold in this test set, demonstrating that the display answers the question at a high level. The local question can be answered by slowing down the display rate, either by editing the tempo in the score, or playing the sequence with a tool that allows variable rate playback. Intermediate structure may be heard by altering the rate of presentation so that different streams pop-out. As the rate slows the streams become broader and encompass more elements. The effect is akin to a visual zoom that shows detail at different scales.

goldtest1.sco	identity
f1 0 8193 10 1 ; sin	
f2 0 129 7 0 128 1; ramp 0 to 1	
t 0 200 ; tempo	
; d1 d2 d1 d2 d3 d4 d5 d6	
i1 0.333 1 0.333 0.692 0.176 0.138 0.354 0.058	dirt
i1 0.554 1 0.554 1.232 0.074 0.198 0.358 0.065	dirt
i1 1.028 1 1.028 0.576 0.070 0.077 2.401 0.162	gold
i1 0.217 1 0.217 0.885 0.259 1.087 0.739 1.005	gold
i1 0.507 1 0.507 0.862 0.198 0.309 0.578 1.107	dirt
i1 0.570 1 0.570 0.866 0.301 0.434 0.526 0.674	dirt
i1 0.554 1 0.554 1.232 0.074 0.198 0.358 0.065	dirt
i1 0.530 1 0.530 0.042 0.852 0.250 0.053 0.185	dirt
i1 0.497 1 0.497 1.590 1.365 1.175 1.278 0.101	gold
i1 0.167 1 0.167 1.261 0.031 0.658 0.306 1.289	dirt

Table 6-22: A mixture of dirt and gold

If you have an ear for gold you can mix up your own sets from the training sets, or you can generate new samples from the equations in Bly's scenario. The orchestra may be simple enough to allow interaction from score statements generated in real-time. A Perl program, called Goldmaker.prl, that can generate data with any mix of gold and dirt in it is listed in Appendix 6-1. Using this program I generated 12,000 data values, with various amounts of gold in them. The data-set was rendered as 1 minute of audio. I was able to hear regions of gold as distinct masses of higher, brighter material. Even when the proportion was only 5% the individual gold samples pop-out from the background of dirt sounds. The ability to hear low proportions of outliers may be useful in data-mining (rather than geo-mining) applications, where the interesting data are rare and multidimensional. I do not know of any other display technique that allows a person to detect single outliers in a mass of 12000 6D data points per minute. However these observations need to be empirically validated.

6.10 Summary

Sounds are useful in everyday activities and they can also be useful in abstract information processing activities. The advent of faster audio synthesis hardware has potential to allow auditory displays to become common in many as-yet-unforeseen computer-based activities. However auditory display is still a new field and the designer is faced with some difficult challenges, due to the type of information to be represented, the need for consistent comprehension, and the need to support interactive exploration. The lack of a systematic approach for mapping data into sounds has been identified as a gaping hole that is impeding progress in this field of practice. One way to bridge this gap is to borrow some principles that have been developed by graphic designers faced with similar issues. Principles of Reading, Type, Level, Organisation and Range have appeared consistently and been applied broadly, and may also be helpful in auditory display. The integration of these principles with psychoacoustic observations may provide the systematic approach that has been called for.

This suggestion was investigated by a demonstration of each principle in the context of auditory display. The demonstrations show that the principles could apply to sounds. The next step was to find out whether they were of any benefit in practice. Bly's dirt and gold scenario was chosen as a test-bed because it is an example of information that may be difficult to display and understand visually, and because it is a reference for other designers. The resulting display enables a listener to quickly, correctly and confidently answer the question "can you find the gold?" at local and global levels. The important thing is that these principles for auditory information design have been shown to produce a useful and effective high level display, and are demonstrably of benefit in practice.

7 • Information-Sound Space: a cognitive artefact for auditory design

Some classifications of sound events tend to be categorical...[cut]...These simple categorical distinctions can potentially be exploited in auditory presentations to communicate important distinctions in the data. Beyond these categorical distinctions, the essential goal is that perceptually continuous auditory attributes are scaled and mapped to data attributes in a way that is meaningful to the observer [Kendall G.S. (1991)].

The previous chapter introduced the Hearsay principles for auditory information design that summarise some knowledge that can help in the design process. Although they are helpful, principles and guidelines can be unwieldy in practice because of the need to keep referring back to them. Principles cannot be simply applied by rote, they have to be learnt and understood. This chapter describes an alternative representation of the Hearsay principles in the form of an Information-Sound Space (ISS). The ISS is a three dimensional spatial organisation of auditory relations that bridges the gap from theory to practice by changing the way a designer can think about and manipulate relations between sounds. Rather than having to follow written principles the designer is able to think in terms of simple spatial structures that represent information relations. The following sections describe the development of the ISS, which is based on the HSL colour space that has been applied in many areas of design including scientific visualisation of data sets. The feasibility of implementing an ISS is investigated in several experiments that draw upon psychoacoustic observations made by Von Bismarck, Slawson, Grey, and Bregman.

7.1 What is a cognitive artefact?

Written rules are not the only way that principles can be represented. A more direct and accessible form of representation may better support design practice. A cognitive artefact is a tool that leverages knowledge to enhance design skills by altering the design task [Barnard P. (1991)]. An example is the Hue, Saturation, Lightness (HSL) colour model which is commonly found in paint brochures to assist in the choice of colour schemes. The arrangement of colours as a 3 dimensional solid makes it easy to understand concepts such as colour complementarity, similarity, and contrast. The relations between colours can be understood by geometric distance, lines, slices and planes. The selection of schemes with specific properties is very direct and simple. Examples of colour schemes are complementary pastel shades, or a high contrast saturated scheme. Rather than looking through principles of colour theory to realise these schemes the painter can choose them directly from the colour solid.

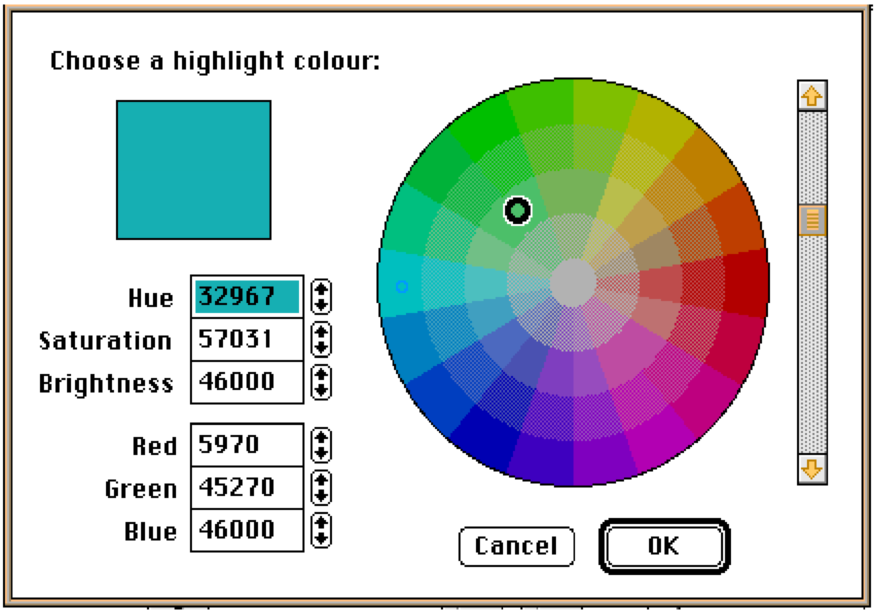


Figure 7-1: A colour choosing tool that uses the HSL colour space

Colour models have many applications in architecture, painting, dyeing, decoration, and anywhere else that people need to specify and understand colours, and groups of colours. Different colour models have been specialised to support different applications. However they share a similar 3 dimensional organisation based on observable and separable perceptual aspects of colour. Some spaces have been perceptually scaled so that there is a regular change in the amount of variation in each aspect of the colour. There has been about a hundred years of empirical studies dedicated to orthogonalising and building colour metrics. Colour spaces such as the Munsell, Ostwald, CIE, OSA, Coloroid and NCS systems have different properties depending on the choice of axes and their arrangement, whether additive or subtractive colour mixing is supported, the type of perceptual comparisons that are supported, and the weighting of local versus global orthogonality [Hunt R.G.W. (1987)].

7.2 Information properties of colour spaces

The colour solid has a polar-cylindrical coordinate system which has a circular hue dimension, a radial saturation axis, and a vertical lightness axis, as illustrated by sequences of 8 equal steps in Figure 7-2. The hue angle varies from 0 through 360 degrees. The number of hues that are used in various colour systems ranges from 5 for Munsell to 24 for DIN to 40 for the NCS system. Lightness is a ratio (prothetic) perception which has a natural zero where an absence of lightness causes the colour to disappear. Lightness has been ratio scaled with a psychophysical constant B = 1.2 for reflectance of grey papers, and B = 0.5 for a point source [Stevens S.S. (1966)]. The lightness difference in the CIE colour space equations are based on measurements of just noticeable differences (JNDs) from a white reference. This scale has 100 equal steps in the lightness dimension. Saturation is ordered and also scaled by just noticeable differences in the CIE colour space. It has an original zero starting at grey which is the same point for all scales of saturation no matter the hue. The saturation range varies with lightness, reaching a widest point of about

8 JNDs at mid lightnesses.

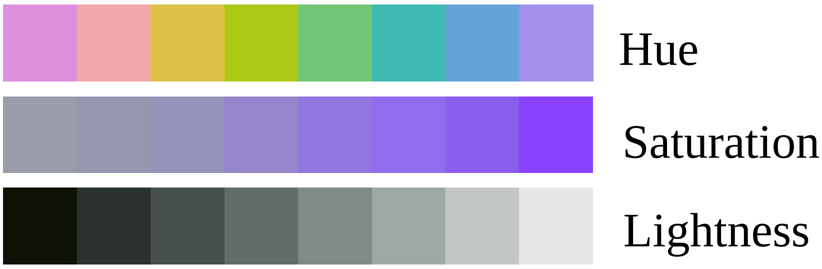


Figure 7-2: 8 steps in hue, saturation and lightness

The hues are different from the other dimensions because hue can be organised into a circle. The most similar hues, such as yellow and orange, are next to each other, and the most dissimilar, or complementary hues, are opposite each other in this circle. The arrangement reflects a theory of colour vision called the colour-opponent theory which has independent red-green and blue-yellow axes of hue variation.

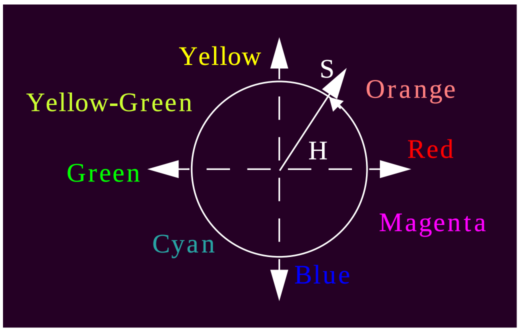


Figure 7-3: Opponent hue axes

The placement of hues in the hue circle is organised so that the red and green opponents form the 0-180 degree axis, whilst yellow and blue are on the 90-270 degree axis. The intermediate colours are placed between them. Although the hue circle is locally continuous, the perception of hue is globally categorical. This is important in the application of colour to represent data in maps, graphs and visualisations where hue is used to separate distinct regions or display classifications, rather than for smooth gradients or continuous variables. The order of hues in the rainbow may seem natural, but people do not have a consistent intuition of what the order is.

The use of colour to represent information is of particular interest to us in sonification, because it involves issues of perceptual representation which may also apply to sounds. Colour is commonly found in graphs, maps and road signs. Colour spaces have been used to control colour relations in satellite imagery and scientific visualisations where data structure is perceived by colour structure [Robertson P.K. and O'Callaghan J.F. (1986)]. Trumbo proposed four principles of colour representation for univariate and bivariate statistical maps [Trumbo B.E. (1981)]

- *Order - If levels of a statistical variable are ordered, then the colours chosen to represent them should be perceived as preserving the order - for example from dark to light, or pale to saturated.*
- *Separation - important differences in the levels of a statistical variable should be represented by colours clearly perceived as different - for example distinct lightness levels or changes in hue*
- *Rows + Columns - if preservation of univariate information or display of conditional distribution is a goal, then the levels of the component variables should not interact to obscure one another*
- *Diagonal - if display of positive association is a goal, scheme elements should resolve themselves visually into three classes: those on or near the principal diagonal, those above it and those below it.*

Trumbo realised that colour spaces had the properties that could support the application of these principles in the design of colour displays. He implemented colour mapping schemes as paths and planes within the Ostwald colour solid, which has hue, saturation and lightness axes like the HSL model. The Ostwald colour solid is comprised of triangular leaves that characterise the variation of saturation with changes in lightness. The dynamic range of saturation is greatest at mid lightnesses, and collapses to zero at the light and dark points of the solid. When manipulating dependent parameters it is be difficult to know when the limit of a range in one parameter had been exceeded, due to variation in another. A visualisation of the colour solid allows a designer to understand the gamut of variation and keep to valid regions of the colour space.

The colour solid supports the application of Trumbo's principles to the design of a colour display, as shown in Table 7-1. The principle of order is supported by the perceptually ordered lightness and saturation axes. Questions like "what is the mean income?" with answers { <$5000, $5000-$9000, $10000-$14000, $15000+} can be shown by 4 changes in lightness. The principle of separation is supported by the perceptual metric of the axes. This metric can ensure that the visual answers are have consistent perceived difference and weighting in the display. Trumbo warns against the choice of a hue sequence to answer this question because it does not allow the perception of order, and so does not satisfy the principal of separation. The principle of rows+columns is supported by the orthogonality of the lightness and saturation dimensions. This principle enables answers to questions which involve two variables at the same time - for example "are all districts with both low income and low education near the centre of the city?" by colours that can be reliably imagined from a specification in terms of two aspects of variation. The principle of the diagonal can be supported by the categorical difference from a hue to grey to a complementary hue that occurs in a vertical slice through the space. These categories allow the perception of bivariate distributions as having three distinct hues classes that indicate above, correlation, and below. This enables the capability to answer questions such as "is there a positive association between higher education and income?" and "whereabouts are their groupings of exceptional cases?". The organisation of the colours in the bivariate sequences can enable the perception of higher level information in the display.

Colour relation	Difference category/ continuous	Order 1D, 2D, 3D, nD	Metric ratio/ difference	Zero natural original conventional	Just noticeable differences
Hue	category		difference	conventional	~ 24
Saturation	continuous	1D	ratio	original	~ 5
Lightness	continuous	1D	ratio	natural	~ 100

Table 7-1: Information characteristics of the colour solid

7.3 Blueprint of an Information-Perception Space

The colour solid has proven very helpful for understanding and specifying colour relations in information displays. This gives a motivation for a similar method for understanding and specifying auditory relations. There are many different ways we could organise a space of auditory relations, each perhaps tailored to a particular type of representation task. However the architecture of the HSL space provides us with a familiar and well-known starting point that has proven capability to support a general range of information representations. This architecture may serve as a blueprint for a general purpose Information-Perception Space (IPS), and provide an abstract foundation for an auditory display. Just as the colour solid makes colour easier and more direct to use, the IPS makes the Hearsay principles more direct and easy to apply in auditory design practice.

The Information-Perception Space (IPS) is a cylindrical polar organisation that has a cyclic dimension of 8 categories, a radial of 8 equal, ordered differences, and a vertical axle of 100 equal, ordered differences. Each principle is addressed by the organisation of the IPS as follows:

7.3.1 Reading

The IPS focuses the design on direct perceptual relations between elements. This is a very different focus from the usual design of conventional symbols that must be learnt and read from the display. The IPS can also help to select conventional symbols with prescribed perceptual properties. An example might be to select some symbols which will be perceived as distinctly separate in a conventional display, by selecting perceptual points that are equally spaced around the categorical circle.

7.3.2 Type

The combination of a circle of globally unordered difference, an axis with order, an axis with a metric and a zero covers all of the elementary perceptual relations {difference, order, metric, zero}. This enables the IPS to support the TaDa information types {boolean, nominal, ordinal, ordinal-with-zero, ordinal-bilateral, interval, ratio} that have been defined in terms of these relations.

The pedestal of categories

The other properties of the IPS rest upon the pedestal of categories. The circle can be divided into 8 regular categories to accord with the limits of short term memory. The equally spaced pedestal has the following properties

- differences do not have an observable order
- adjacent points are subjectively equally different
- the circle has a conventional zero, so that cycles are perceptually seamless

The disc of radial spokes

The disc of radial spokes is an extension of the pedestal of categories by a radial variation within each category. It is important that this radial component does not cause a perceptual change in category. In the colour solid the radial saturation of a hue can vary without the hue changing, for example pink can change to red. The original zero is an anchor for absolute judgments along any radius. For example grey is a zero in the colour space, independent of hue or lightness. The radial component has the following characteristics:

- observable variation throughout each category
- a perceptual metric
- an original zero which is a common point of origin, independent of category

The vertical axle

The IPS is completed by transfixing the disc of radial spokes on a vertical axle. The variation in this dimension must be observable and ordered everywhere in the space. This dimension has a natural zero, which marks the absence of a perceptual element. The zero anchors absolute judgments on this axis. The axle contains all the original zero points of the radial scales. In colour models it is sometimes called the “grey” axis because all the desaturated points lie along it, stretching from the dark point to the light point.
The vertical axle has the following characteristics:

- observable variation throughout each category
- perceptual independence from the radial dimension
- a perceptual metric
- a natural zero for absolute judgements

7.3.3 Level

The combination of the three different types of perceptual axes into an orthogonal basis provides the opportunity to construct bivariate and trivariate representations. The axes are scaled so that euclidean distance corresponds with perceptual difference. To be a truly uniform the scaling needs to account for the perceptual interactions between dimensions - for example the CIE perceptually uniform colour space is scaled by just noticeable differences in each dimension at each point.

A factor not included in the colour space is the control of perceptual grouping. This capability is important for designing higher level (intermediate and global) information displays that depend on perceptions of grouping and segregation. The selection of perceptual attributes for each axis may be made from factors that have an influence on grouping. The categorical factor should be particularly strong to maintain cohesiveness whilst other aspects vary.

7.3.4 Organisation

The organisation by a scheme {category, time, location, alphabet, continuum} or reorganisation for exploration can only be done in an indispensable dimension that preserves the separation between objects necessary for the perception of “twoness”. Examples are space and time in vision, neither of which are a part of the colour spaces. Audio and video sequencers are common tools for organising sounds and pictures in time. These aspects are not a part of the IPS but may be organised with another tool, called Personify, that is developed in Chapter 9 of this thesis.

7.3.5 Range

The dynamic range of each perceptual axis in the IPS constrains how representation schemes can use the space. Representations that require discrimination between large numbers of elements may have to be oriented differently from those with only a few. The range of each dimension can be set in many ways, but the colour solid is the framework we are starting with, and so we will set the ranges accordingly.

- the pedestal has 8 categories
- the radial spokes have 8 steps
- the vertical axle has 100 steps

The Information-Perception Space with information properties of Reading, Type, Level, and Range modelled on colour spaces is shown in Figure 7-4.

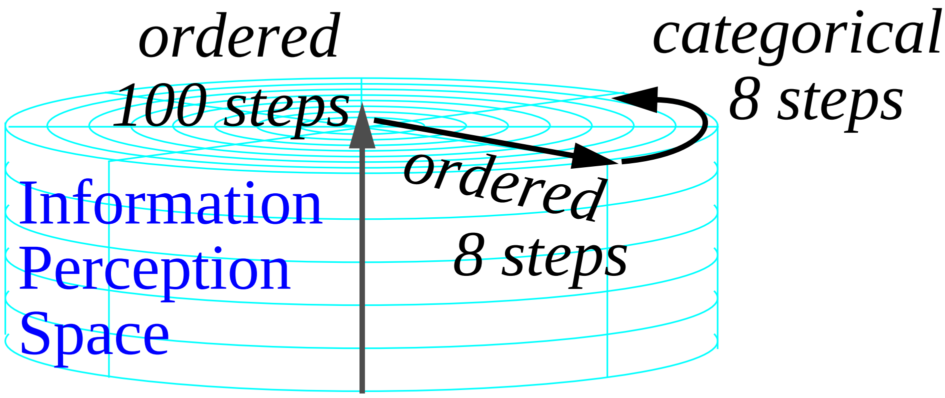


Figure 7-4: Blueprint for an Information-Perception Space (IPS)

The boundaries of available variation of device parameters can be shown in the space in a way that makes the device seem like a solid object. The designer is able to manoeuvre within the boundaries to maximise dynamic range and ensure separation without exceeding the limits of the device. If the device parameters are not identical with the perceptual parameters then a mapping can be constructed from the IPS to the device parameter space. This allows the designer to work in perceptual terms, and enables the specification of the display in device-independent perceptual parameters.

7.3.6 Representations in IPS

The design of representations in the IPS can be specified by paths through the space that have particular properties due to the organisation of the space. The angular axis is qualitative difference, the radial axis is quantitative difference with order, metric and a zero, and the vertical axis is quantitative difference with order and a metric. The elementary information relations in the IPS are listed in Table 7-2

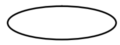
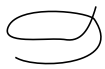
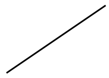
Elementary relation/ ISS dimension	Difference	Order	Metric	Zero
Angle	qualitative		8 steps	conventional
Radius	quantitative	yes	8 steps	natural
Vertical	quantitative	yes	100 steps	conventional

Table 7-2: Elementary information relations in the IPS

A direct representation preserves the characteristics of the information when it is heard in the display. The rules for choosing a direct representation in the IPS are

- Rule 1 - if the Information Type has qualitative difference vary the qualitative Angular axis
- Rule 2 - if the Information Type has quantitative difference vary the quantitative Vertical axis
- Rule 3 - if the Information type has a zero vary the Radial axis starting from the origin

These rules specify paths through the IPS that directly represent different types of information relations, as shown in Table 7-3.

Info Type	Angle category	Radial zero	Vertical metric	IPS Mapping
Boolean	X			Opposite angles
Nominal	X			Circle 
Ordinal	X		X	Coil 
Ordinal and zero	X	X	X	Spiral 
Ordinal bilateral	X	X	X	Sloped Line 
Interval			X	Vertical Line
Ratio		X	X	Radial Line

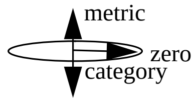


Table 7-3: Information paths in IPS

7.4 Specialising the IPS for auditory display

This section specialises the abstract IPS for auditory display. The specialisation is made by assigning auditory dimensions to each dimension in accordance with the Hearsay principles. The result is a 3 dimensional space of auditory relations that can represent information relations, called the Information-Sound Space (ISS).

The selection of aspects of sound for each axis can be informed by some other efforts to construct geometric sound spaces. The Hue, Saturation, Lightness (HSL) colour model was used as a template for a model of sound by Caivano [Caivano J.L. (1994)], who made a link between the circularity of pitch classes and the hue circle. This cylindrical polar arrangement has pitch angle, timbre radius and loudness height as axes, as shown in Figure 7-5. The timbre axis is ordered from white noise at the centre, through inharmonic spectra, through harmonic spectra, to a simple sinusoid at the extreme radius. This ordering is derived from consideration of the complexity of the physical spectrum but it is not clear that an observer would quickly, correctly or confidently hear ordered information mapped into this sequence as being ordered.

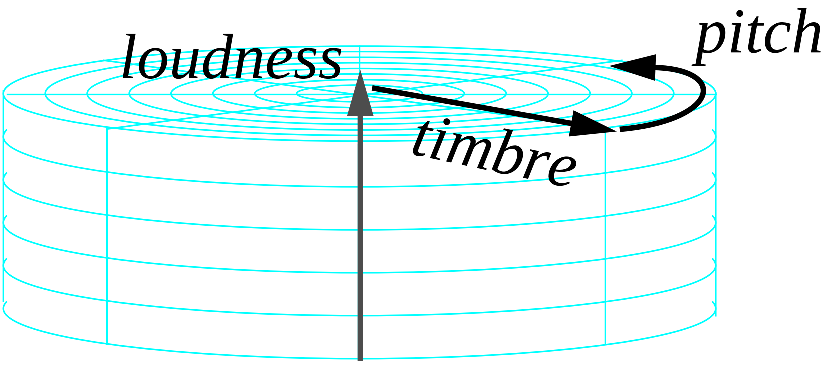


Figure 7-5: Caivano's organisation

Another circle of sounds is found in Padgham's sound chart for comparing and calibrating pipe organs [Padgham C. (1986)], shown in Figure 7-6. This chart is a polar plot in which angle represents the characteristic tone caused by the formant region containing the first 5 harmonics, and radius represents the complexity of the spectrum, or number of harmonics in the second formant region. Padgham linked increased spectral complexity with increased "colourfulness", which is similar to Caivano's timbre radius, but opposite in direction. The relationship between timbre and loudness is represented by stacking together a pile of these charts measured at different loudnesses. This cylindrical polar sound space has tone angle, complexity radius and loudness height. The four opponent timbres in the chart are analogous with the opponent hues in the colour space. In a series of experiments Padgham found that listeners were as consistent in plotting the sounds in this chart as they were at plotting hues in the colour model. The results suggest a correspondence between the flute position and first harmonic, the string with the second harmonic, and the trumpet and the third harmonic. However not all the results agree, for example one experiment shows an inverse relation between the third harmonic and the trumpet position. After concluding that there is evidence of a regular perceptual relationship between the angular dimension and the shape of the first formant Padgham doesn't explicitly define what this relationship is. However the idea of a circle of timbres organised by opponent axes provides a way to organise a cycle of sounds that can be extended to other organisations.

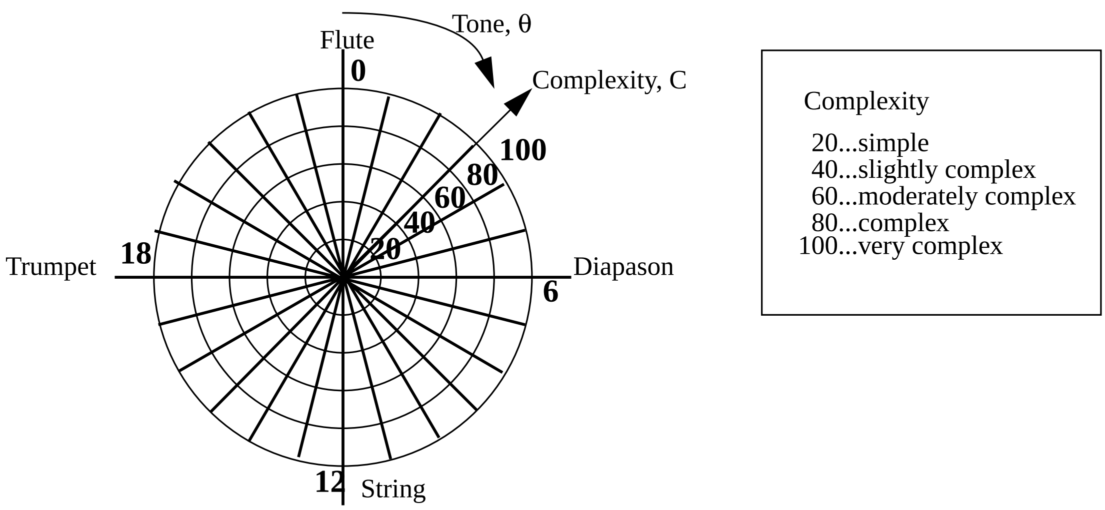


Figure 7-6: Padgham's timbre assessment chart for the pipe organ

These variations on a sound space are just the beginning of the possible organisations that can be chosen. The next sections investigate some of the possibilities in terms of the Hearsay principles. The investigation involved the generation of prototype sound sequences for testing the perceptual characteristics of the space. The equipment was a Sun SPARCstation^{TM} 10 workstation which includes 16 bit, 44.1 kHz audio as standard hardware. The prototypes were generated with the Csound audio synthesis software which is a freely available for research and has many sound generation and processing functions. The Csound programs for each synthesis instrument are included in Appendix 7-1 at the end of the chapter.

7.5 The pedestal of auditory categories

The starting point for an Information-Sound Space is a circular pedestal of categories upon which the space revolves. The pedestal is made up of auditory relations that have difference but no order or zero. This section is a pilot study to investigate candidate pedestals. Some candidate auditory relations with suitable characteristics are listed in Table 6-11, for example musical key, tunes, rhythms, material type, event type, vowels, and timbres. The models of Caivano and Padgham are also considered.

7.5.1 The Pedestal criteria

A set of criteria that must be satisfied to realise a categorical pedestal are proposed. A null hypothesis for each criterion is given. A criterion is supported if its null hypothesis is rejected. The tests were carried out by listening to auditory sequences having prescribed relations in terms of Zero, Difference, Order, Metric, Level and Range.

Zero - categories do not have a perceptual zero

Null hypothesis: the sequence has a perceptually singular point.
Zero sequence: a repeating cycle of points regularly spaced around the circle.

The listener hears repeating cycles of the sequence. The task is to indicate the start of the sequence. The null hypothesis is accepted if a starting point is consistently identified.

Difference - each category sounds different

Null hypothesis: two or more elements of the sequence are identical
Difference sequences: complementary and adjacent triplets

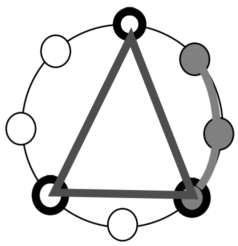


Figure 7-7: Complementary and adjacent triplets

The task is to find two points that sound the same. The search through all pairwise comparisons is very large. A limited analysis can be obtained by listening to all complementary and adjacent triplets, as shown in Figure 7-7. Complementary triplets are sets of three elements chosen at equal intervals around the circle. Adjacent triplets are three points at successive positions around the circle. The null hypothesis is supported by the consistent identification of two identical elements in a triplet.

Order - categories are not heard to have a simple order

Null hypothesis: subsets in the sequence have a simple order.
Order sequence: complementary triplets

Complementary triplets are sets of three points chosen at equal intervals around the circle.

The listener hears each complementary triplet in a repeating cycle. The null hypothesis is accepted if a repeating triplet is consistently heard to have a simple unidimensional variation.

Metric - the difference between categories is regular

Null hypothesis: adjacent elements do not have regular spacings
Metric sequence: adjacent triplets

The listener hears sets of adjacent triplets from around the circle. The task is to identify the most similar pair in each triplet. The null hypothesis is accepted if there is a consistent pairing that indicates irregular spacing in a triplet.

Level - categories segregate into different auditory streams

Null hypothesis: the categories do not segregate into different streams
Level sequence: adjacent pair in Van-Noorden's XOX-XOX galloping sequence, at rates from 500 ms to 50 ms

The listener hears a pair of categories in a sequence of the form XOX-XOX where X is one category and O is an adjacent category, and - is silence. This is the sequence that Van Noorden used to measure the temporal coherence of sounds [Bregman A.S. (1990)]. The task is to hold each triplet together as a single unit, and indicate the rate at which this can no longer be done. The point of segregation is signalled by a galloping rhythm where the X-X-X-X is heard in one stream and O---O---O is heard in the other. The typical cohesion threshold is between 50 ms for very similar sounds to 150 ms for dissimilar sounds. The null hypothesis is supported if the triplet is very cohesive as indicated by segregation only occurring at fast rates with onsets of less than 100 ms.

Range - there are 8 discriminable steps

Null hypothesis: two of the categories are the same
Range sequence: complementary and adjacent triplets

This test is identical to the difference test. If the difference test fails then there are less than the requisite 8 categories, and a listener may hear two different categories as the same.

7.5.2 Pitch circle

Let us begin the investigation with Caivano's model in which pitch class is the polar dimension. In the development of this model Caivano proposes a variety of discrete pitch scales and analogies with various hue segmentations - for example a pentatonic scale is linked to Munsells five way division of hues, and the chromatic scale of 12 semitones is linked with the colour opponent model divided into 12 equal segments. However it is immediately evident that these scales fail the criteria of both order and zero. Pitch order is a fundamental property of each scale. A repetition of each scale will be heard to have a zero at the octave discontinuity. An exception is Shephard's pitch illusion, in which pitch seems to rise forever, but never actually leaves the octave [Shephard R.N. (1964)]. This continuous pitch illusion satisfies the Zero criterion, but still does not satisfy the Order criterion, because the ordering of pitch from low to high is still plainly heard in this sequence. The Order criterion can be addressed by reorganising the pitches in some other manner. For example the circle-of-fifths is an arrangement of pitch classes in which angular proximity is a measure of musical similarity. The most harmonically similar pitch classes (separated by fifths) are next to each other, and the most dissimilar lie opposite.

The difference between neighbouring pitch classes is perceptually equal so the sequence may satisfy the metric criterion.

The circle-of-fifths and the Shephard pitch illusion were amalgamated into a Csound instrument to test whether this combination could satisfy the Pedestal criteria. The synthesis algorithm is a comb of 20 partials, or teeth, generated by individual sinusoidal oscillators. The teeth of the comb are spaced at intervals of the fundamental frequency, but the comb itself can be shifted up and down the spectrum by adjusting the frequency location of the lowest tooth. As it slides upward a rising pitch is heard, even though the partials maintain fixed intervals. A continuous cycle is created by passing the comb through a formant that attenuates the edges, removing the discontinuity that occurs at each octave as the lower and upper partials double. The FFT spectra at four opponent locations around the circle are shown in Figure 7-8.

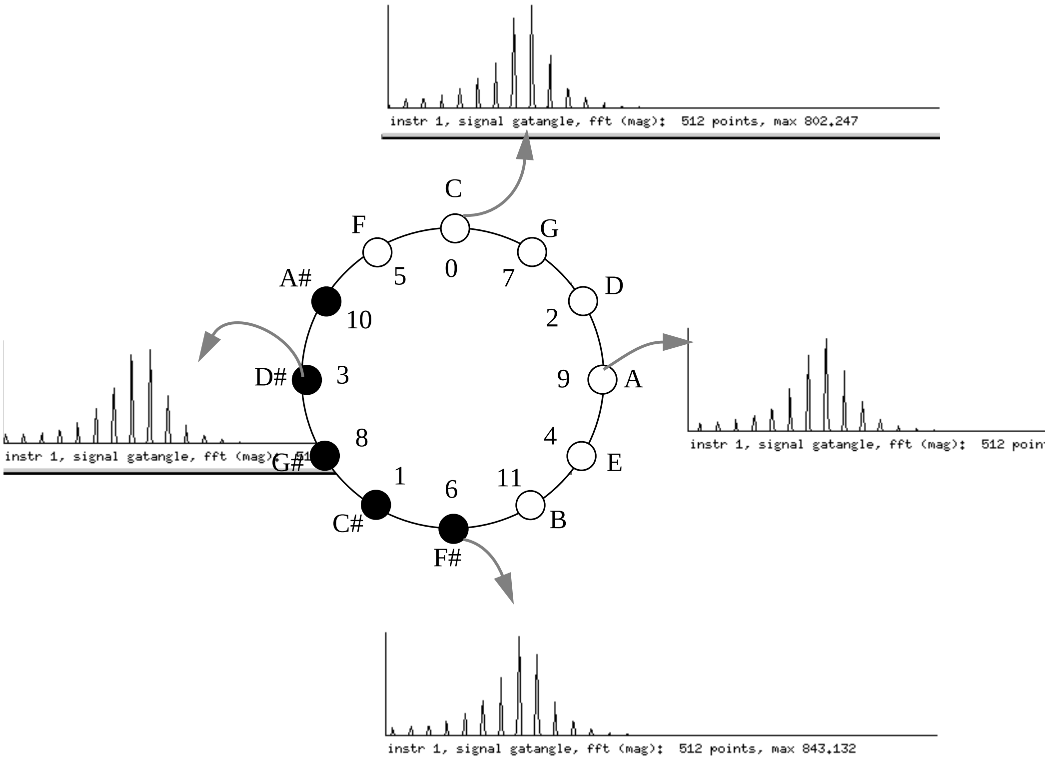


Figure 7-8: The continuous circle of fifths

7.5.3 Pitch Circle rated against Pedestal criteria

The test sequences that were generated were a continuous circle, the categorical circle of fifths with 12 points, four sets of adjacent triplets (0,5,10), (3,8,1), (6,11,4), (9,2,7), four sets of complementary triplets (0,8,4), (6,10,2), (3,7,11), (9,5,1), and the complementary pair (0,6,0). I tested the null hypothesis for each of the Pedestal criteria by listening to these sequences, as shown in Table 7-4.

Criteria	Result	Null hypothesis
Zero	cycle of 12 steps no singular point could be heard	reject
Difference	adjacent triplets 0,5,10 = different 3,8,1 = different 6,11,4 = different 9,2,7 = different complementary triplets 0,8,4 = different 6,10,2 = different 3,7,11 = different 9,5,1 = different	not supported
Order	complementary triplets 0,8,4 = ordered 6,10,2 = ordered 3,7,11 = ordered 9,5,1 = ordered	accept
Metric	adjacent triplets 0,5,10 = equal 9,2,7 = equal 6,11,4 = equal 3,8,1 = equal	reject
Level	complementary gallop 0,6,0 = no segregation up to 50 ms adjacent gallop 0,5,0 = no segregation up to 50 ms	accept
Range		not supported

Table 7-4: Pitch Circle rated against Pedestal criteria

The investigation shows that the null hypothesis for order and for level were both accepted, indicating the failure of the Pitch Circle to satisfy the Pedestal criteria. The problem is that subsets can be heard as ordered, so that a mapping of categorical relations may be perceived to have order where none exists. The strength of pitch and brightness in sequential grouping is shown by the grouping of tones even at the fastest rates. The Csound instrument design involves pitches that are within an octave, and a static formant which forces the brightness of all the elements to be very similar.

7.5.4 Formant circle

The shape of localised regions of the spectrum can be described by parameterised formants. Formants provide a means to characterise complex spectra more simply than a description of individual spectral components. Spoken vowels fall into unique locations in this space, as shown in Figure 7-9, and formants are established principle components in speech research.

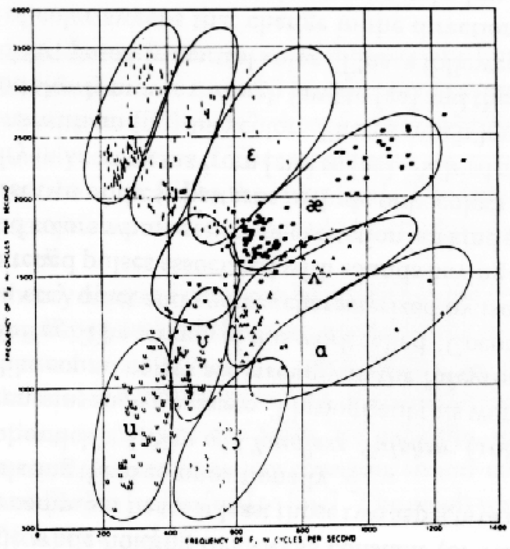


FIGURE 32: *Vowel diagram with ovals enclosing the vowel areas. F1 = abscissa; F2 = ordinate. The data points represent the measured formant frequencies of the vowels that were correctly identified by a panel of listeners. (From Peterson and Barney 1952.)*

Figure 7-9: Location of vowels in terms of F1-F2 formant axes

The two dimensional plane constructed by the first two formants was used as a compositional device by Slawson, who proposed that musical timbres are perceptually ordered in this space. He investigated the transposition of sequences selected from this space, and suggests that the transposed sequences maintain perceptual coherence and order. Following on from Padgham's opponent organisation of timbres, and using Slawson's formant basis, we can propose a cycle of sounds which revolves around the F1-F2 axes, as shown in Figure 7-10. The Csound algorithm for this sequence is shown in the FormantCircle.orc and FormantCircle.sco files in Appendix . The F1-F2 coordinates individually manoeuvre the centre frequency of a formant region. The F1 axis linearly positions the peak of the formant of the first 5 harmonics. The F2 axis linearly positions the peak of the formant of the harmonics from 6 to 10. The algorithm consists of a pair of bandpass filters, one for each formant. The centre frequency of each filter varies between the extremes of its range in accordance with the values of the F1 and F2 parameters. The filters are applied to a harmonic spectrum which extends from the fundamental to the Nyquist. The bandwidth of each filter is twice the fundamental. The spectrum at four opponent points in the space are shown in the FFT frames in Figure 7-10.

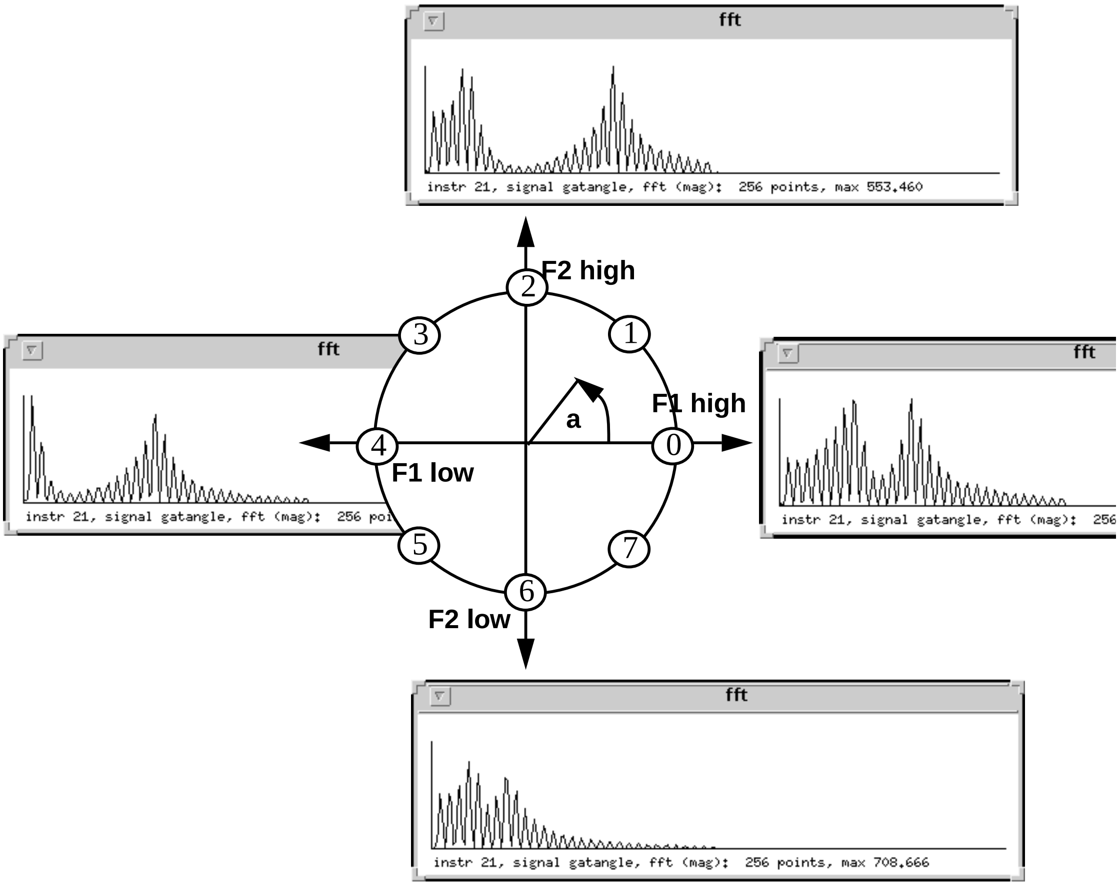


Figure 7-10: Formant Circle

7.5.5 Formant Circle rated against Pedestal criteria

The test sequences that were generated were the categorical circle of eight elements at 45 degree angles, four sets of adjacent triplets (7,0,1), (1,2,3), (3,4,5), (5,6,7), four sets of complementary triplets (0,3,5), (2,5,7), (4,7,1), (6,1,3), and the complementary pair (1,5,1). I tested the null hypothesis for each of the Pedestal criteria by listening to these sequences, as shown in Table 7-5. The investigation shows that the Formant Circle does not satisfy the criteria of Order, because two of the complementary triplets sounded ordered in brightness. However all other criteria were satisfied. Support for the Level criteria was good for complementary points, but there was no segregation for adjacent points. This result may reflect the sensitivity of the ear to speech like sounds, and the importance of the formant space in hearing perception. Strong grouping for similar sounds, and strong segregation of dissimilar sounds may be a very useful characteristic for supporting higher level displays.

Criteria	Result	Null hypothesis
Zero	cycle of 8 steps repeating cycle = no singular point	reject
Difference	adjacent triplets 7,0,1 = different 1,2,3 = different 3,4,5 = different 5,6,7 = different complementary triplets 0,3,5 = different 2,5,7 = different 4,7,1 = different 6,1,3 = different	not supported
Order	complementary triplets 0,3,5 = ordered brightness 2,5,7 = unordered 4,7,1 = ordered brightness 6,1,3 = unordered	accept
Metric	adjacent triplets 7,0,1 = equal 1,2,3 = equal 3,4,5 = equal 5,6,7 = equal	reject
Level	complementary gallop 1,5,1 = segregation at 110 ms adjacent gallop 0,1,0 = no segregation up to 50ms	accept
Range	as for Difference	reject

Table 7-5: Formant Circle rated against the Pedestal criteria

7.5.6 StaticTimbreCircle

Early studies in timbre perception were of steady sounds which had spectral components that did not change over time. Ramps, blocks, trapezoids and humps were among the spectral envelopes that Von Bismarck [Von Bismarck G. (1974a)] applied to harmonic and white noise sources. Subjects were asked to rate the resulting sounds against 30 verbal scales consisting of opposite meaning pairs such as hard-soft, sharp-dull, violent-gentle, dark-light, rough-smooth, coarse-fine, dirty-clean, thin-thick, compact-scattered, empty-full, solid-hollow. Four main dimensions which spanned 90% of the variation were found by factor analysis. Sharpness was the most dominant factor, followed by compactness. The compactness dimension showed a clear discrimination between sounds with harmonic sources and those with noise sources. Sharpness was related to the centre of gravity of the spectral envelope, with a progression from sounds with dominant low harmonics to sounds where the upper harmonics were emphasised. In further experiments Von Bismarck demonstrated that sharpness can be doubled and halved in a similar fashion to loudness and pitch [Von Bismarck G. (1974b)].

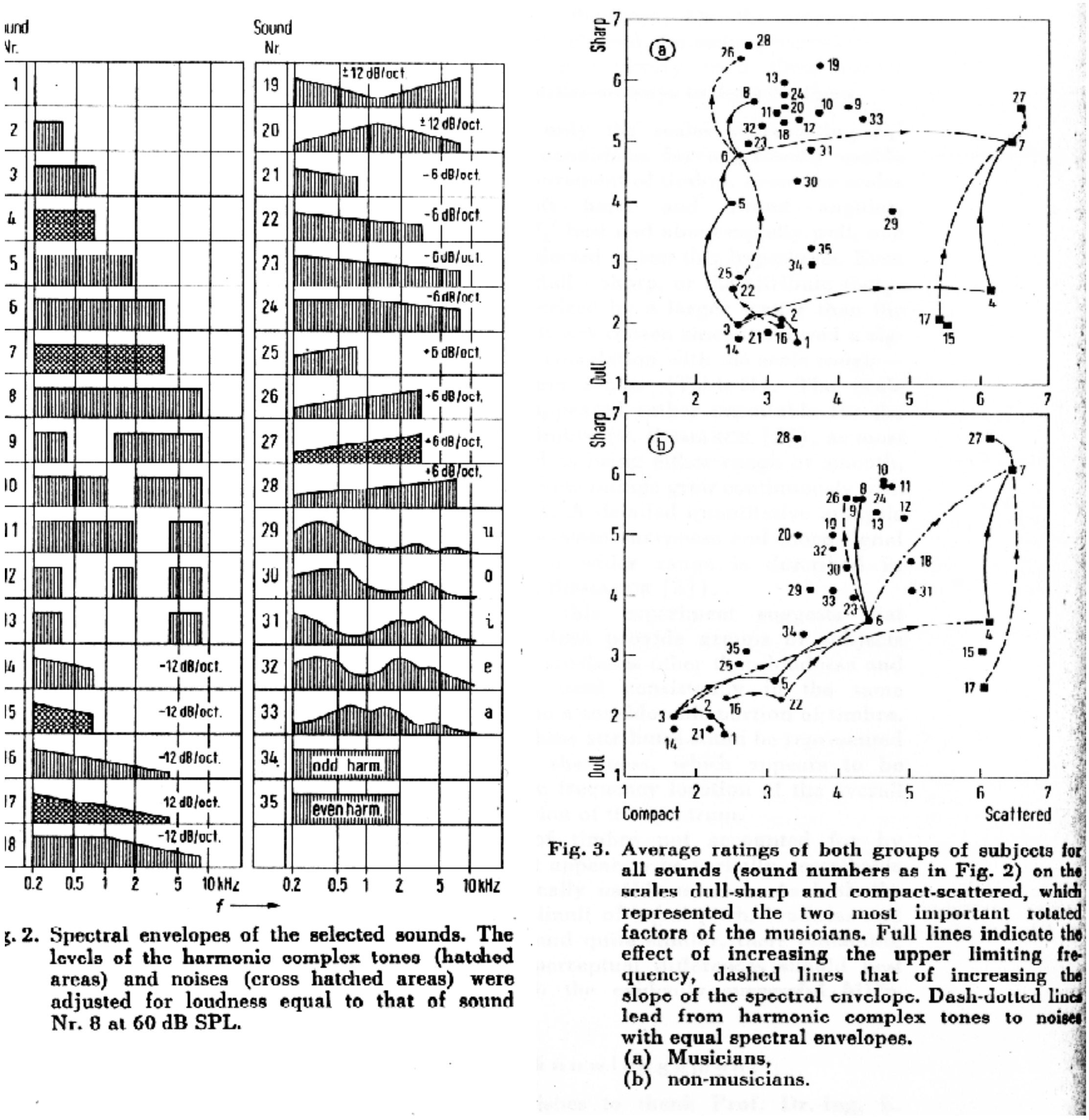


Figure 7-11: Figures from Von Bismarck

The ordering axes for the StaticTimbreCircle are compact/scattered and dull/bright. The instrument is implemented as a source/filter. The source consists of bands of noise centred at the first 20 harmonic frequencies. The X axis linearly controls the width of each noise band in the range 0.1 Hz to $f_0/2$ Hz, so that at the compact end the trend is toward a pulse train and at the scattered end it is a band of noise. The brightness of the source is adjusted by the Y axis which linearly controls the centre frequency of a 2nd order bandpass filter with bandwidth bw = $5f_0$ in the range f_0 to f_{20}. The Csound algorithm is shown in the StaticTimbreCircle.orc and StaticTimbreCircle.sco files in Appendix . The spectrum at four opponent points in the space are shown in the FFT frames in Figure 7-10.

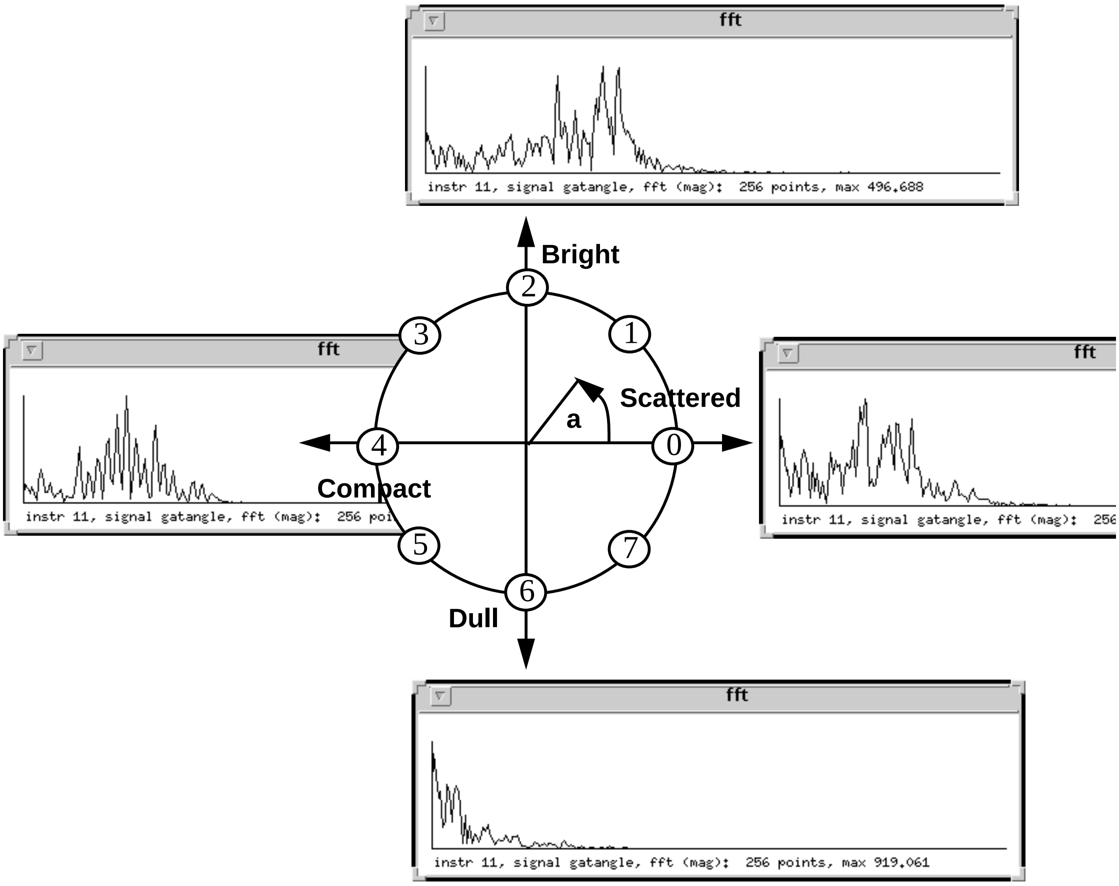


Figure 7-12: Static Timbre Circle

7.5.7 Static Timbre rated against Pedestal Criteria

The test sequences that were generated were the categorical circle of eight elements at 45 degree angles, four sets of adjacent triplets (7,0,1), (1,2,3), (3,4,5), (5,6,7), four sets of complementary triplets (0,3,5), (2,5,7), (4,7,1), (6,1,3), and the complementary pair (1,5,1). I tested the null hypothesis for each of the Pedestal criteria by listening to these sequences, as shown in Table 7-5. The investigation shows that the Static Timbre Circle does not satisfy the criteria of Order, or Metric. All of the complementary triplets sounded ordered in brightness or noisiness or both. This indicates that the underlying opponent axes are perceptually separable and observable, and that the listener is able to make direct judgements about the coordinates in terms of these aspects of sound. This is very interesting, because it indicates a need to use more subtle variations that are integral to the timbre and do not cause a perception of order. However this observability allows us to understand that the axes need scaling to ensure that the circle is not skewed by listening to the apparent brightness ordering of points with the same brightness coordinate in the complimentary triplets. This non-uniformity is reflected in the metric ratings where regular intervals are not heard to be regular.

Criteria	Result	Null hypothesis
Zero	cycle of 8 steps repeating cycle = no singular point	reject
Difference	adjacent triplets 7,0,1 = different 1,2,3 = different 3,4,5 = different 5,6,7 = different complementary triplets 0,3,5 = different 2,5,7 = different 4,7,1 = different 6,1,3 = different	not supported
Order	complementary triplets 0,3,5 = ordered brightness 2,5,7 = ordered brightness 4,7,1 = ordered noisiness 6,1,3 = ordered brightness and noisiness	accept
Metric	adjacent triplets 7,0,1 = equal 1,2,3 = much closer 1,2 3,4,5 = a bit closer 3,4 5,6,7 = closer 6,7	accept
Level	complementary gallop 1,5,1 = segregation at 150ms adjacent gallop 0,1,0 = segregation at 100ms	reject
Range	as for Difference	reject

Table 7-6: Static Timbre Circle ratings against Pedestal criteria

7.5.8 TimbreCircle

Most sounds are not static. A temporal and spectral dimension were found to be most important in an MDS study of complete musical instrument samples made by Wessel [Wessel D. (1985)]. The temporal dimension was grouped by instrument family: trumpet-trombone-French horn, oboe-bassoon-clarinet, and violin-viola-cello. In the spectral dimension sounds with most energy in the low harmonics were at one extreme, and those with energy concentrated in the upper harmonics at the other. In another MDS study Grey [Grey J.M. (1975)] equalised the loudness, pitch and duration of each timbre by re-synthesising 16 musical instruments. He found that a 3 dimensional space was required to explain the results. The cartesian space consisted of a temporal plane defined by two orthogonal temporal dimensions, and a vertical spectral dimension. The results are shown in graphic visualisations of the instruments positioned relative to each other in a 3D space. Grey analysed the results with respect to the amplitude, frequency, time spectrograms of the data points. His conclusion was that the Y axis related to spectral energy distribution, whilst the X and Z axes relate to temporal properties of timbre, covarying with synchronicity in the development of upper harmonics and the presence of low-energy high frequency noise during the attack segment.

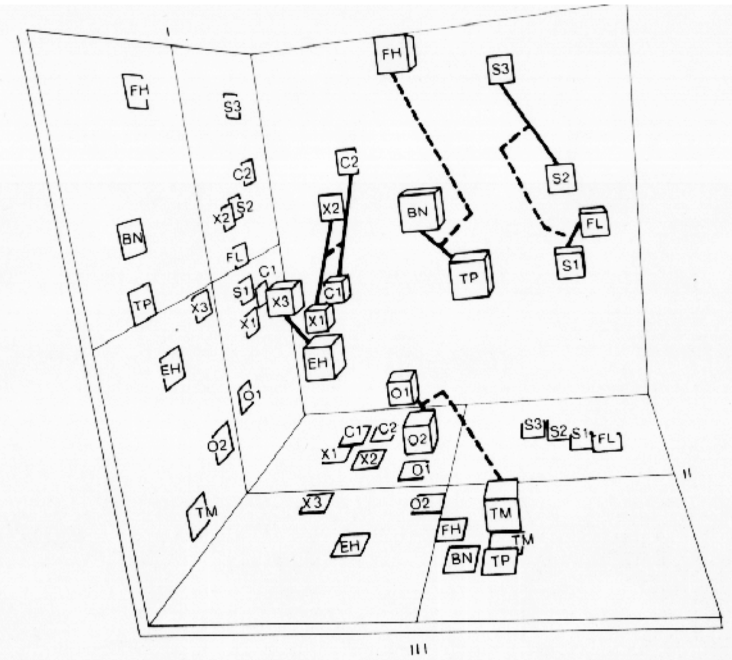


Figure 8.4
A three dimensional representation of the similarities among 12 instruments (16 notes). Instruments (or notes from one instrument) that are highly related are connected by solid lines, and instruments that are strongly related are connected by dashed lines. The values on each of the three dimensions can be determined by looking at the position on the "floor" and the position on the "left" wall. Key: *O1*, *O2* = oboes (different instruments and players); *C1* = E-flat clarinet; *C2* = bass clarinet; *X1* = saxophone, medium loud; *X2* = saxophone, soft; *X3* = soprano sax; *EH* = English horn; *FH* = French horn; *FL* = flute; *TM* = muted trombone; *TP* = trumpet; *BN* = bassoon; *S1* = cello, bowed near bridge producing nasal brittle sound (sul ponticello); *S2* = cello, normal bowing; *S3* = cello, bowing very lightly over the fingerboard producing a flutelike effect (sul tasto). (Adapted from Grey 1977.)

Figure 7-13: A Figure from Grey's MDS study

I used the temporal dimensions identified by Grey's MDS study as opponent axes for a dynamic Timbre Circle to build a Csound instrument. The X axis has endpoints synchronous/spread and linearly controls the rise times of the upper harmonics in the range 0 to 0.3 seconds. This control changes the sweep rate of the centre frequency of a bandpass filter which is applied to a static harmonic series. The Y axis linearly controls the intensity of 0.1 seconds of an inharmonic high frequency onset noise which is mixed with the x axis source. This algorithm was used to generate the sequences as per the previous tests of the pedestal criteria. However it was not possible to complete the tests because the generated sounds did not hold together. The onset noise segregates as a distinct and unrelated sound. The resulting timbres did not exhibit much variation compared with the differences between real musical instruments. Timbre perception is multidimensional and the reduction to two components does not capture the qualities of realistic dynamic sounds.

Digital samples of musical instruments capture the multidimensional variation that makes each timbre unique, and identifiable. The instruments in Grey's study can be used to define a palette of instrument samples. The procedure for selection of a timbre circle from this palette is shown in Figure 7-14. A circle which encloses the projection of the data points in the temporal plane is divided into 8 segments of 45 degrees, and the position of each 45 degree increment around the circumference is nominated to represent the categorical timbre of that segment. Because distance is a measure of similarity, the data point in the segment lying closest to each of the equally spaced points on the circumference is allocated to that point. There are only a limited number of data points available to choose from, so that in segment 7 where there is no data, the closest point from the adjacent segment 6 was taken (i.e. TM). This is only a first approximation to equal spacing as can be seen by the small difference in distance between FL and its neighbour S3 in segments 0 and 1, and the much greater distance between FL and its other neighbour TM in segment 7. This is a consequence of the sparsity and unevenness of the palette, which might be addressed by the use of a different palette of source sounds

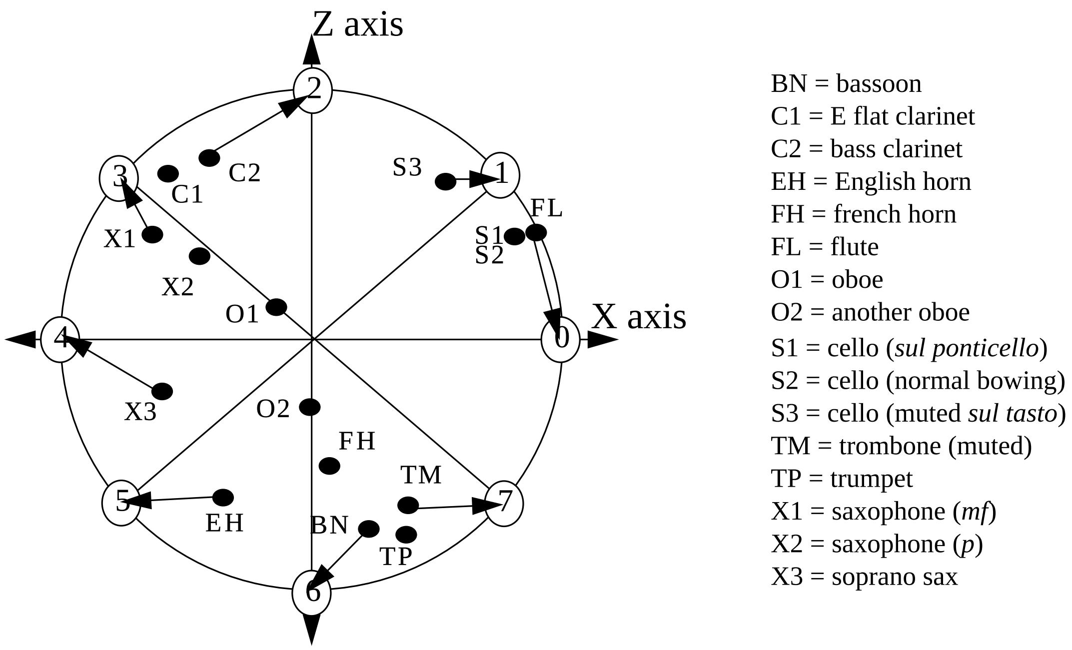


Figure 7-14: Timbre Circle constructed from Grey's temporal plane

A Timbre Circle was implemented but substituting Grey's timbres with digital samples of musical instruments, as shown in Table 7-7. The samples were part of a palette provided with the Gravis UltraSound^{TM} soundcard. The soundcard has a 1 megabyte memory so the samples have been modified by cutting out portions of the steady state portion and using looping to extend the duration of these portions as necessary. These modified samples are called patches. A patch is typically pitch shifted by altering the playback rate of the looped portion, to cover a range of several semitones. The variation beyond this range becomes noticeable due to artefacts that cause timbral changes. The validity of the substitution of Grey's resynthesised instruments with UltraSound patches might be questioned since a repetition of the MDS experiment using these sounds would likely have different results. However this does not invalidate a categorical substitution where the criterion is relaxed so that the primary structuring is on equal similarity between neighbours rather than on euclidean distance between all pairwise comparisons. The goal of the exercise is to enable the representation of categorical data relationships using the categorical nature of timbre perception, and the substitution is not of timbres but of timbre categories. The categorical substitution can be justified under the assumption that sounds which originated from similar physical sources played in the same way (e.g. two different cellos bowed normally) are more perceptually similar than sounds from sources as physically different as musical instrument families (e.g. a cello and a flute) which are also activated or excited in different ways. This remains satisfactory because categorical difference is the essential characteristic required for representing nominal and ordinal data.

Point	Grey’s timbre	UltraSound patch
0	flute	flute
1	cello (muted sul tasto)	cello
2	bass clarinet	bass clarinet
3	saxophone (mf)	tenor saxophone
4	soprano sax	soprano saxophone
5	English horn	English horn
6	bassoon	bassoon
7	trombone (muted)	trombone

Table 7-7: Matching Grey’s timbres to UltraSound patches

A Csound instrument was built to generate the test sequences for the sample-based Timbre Circle. However the generation of the test sequences highlighted some interesting properties of ensembles of musical instruments. The first sequence is a circle of all the instrument sounds, which must be made at a constant pitch in order not to introduce a zero by pitch discontinuity. A problem is that each musical instrument has a unique range of pitches which is physically constrained, and there is no one pitch at which all of these ranges overlap. The best that can be done is to select MIDI note-number 48, where the pitch ranges of seven of the instruments intersect - these are (0,1,2,3,5,6,7). The adjacent triplet (3,4,5) cannot be generated at a constant pitch because there is no overlap in pitch range between the tenor sax and the soprano sax in the MUMS samples. Similarly the complementary triplet (4,7,1) cannot be generated because there is no overlap between the soprano sax and either of the trombone or cello. It may be possible to choose a more compatible set of timbres by revolving the selection circle to a new angular position and re-choosing the category nodes based on information about the pitch ranges of each instrument.

Another problem occurred in the generation of galloping sequences, due to interaction between the presentation rate and the temporal evolution of the flute which is very slow. This was overcome by selecting an adjacent pair on the other side of the circle where the onset time is shorter, these being the bass clarinet and the tenor sax.

7.5.9 Dynamic Timbre rated against Pedestal criteria

Test sequences were all selected at a constant pitch of 48. The sequences consist of the categorical circle of seven elements (0,1,2,3,5,6,7), three sets of adjacent triplets (7,0,1), (1,2,3), (5,6,7), three sets of complementary triplets (0,3,5), (2,5,7), (6,1,3), and the galloping pairs (1,5,1) and (2,3,2). The null hypothesis for each of the Pedestal criteria was tested by listening to these sequences, as shown in Table 7-5.

Criteria	Result	Null hypothesis
Zero	cycle of 8 steps repeating cycle = no singular point	reject
Difference	adjacent triplets 7,0,1 = different 1,2,3 = different 5,6,7 = different complementary triplets 0,3,5 = different 2,5,7 = different 6,1,3 = different	not supported
Order	complementary triplets 0,3,5 = unordered 2,5,7 = ordered brightness 6,1,3 = unordered	accept
Metric	adjacent triplets 7,0,1 = equal 1,2,3 = equal 5,6,7 = equal	reject
Level	complementary gallop 1,5,1 = segregation at 150 ms adjacent gallop 2,3,2 = segregation at 150 ms	reject
Range	as for Difference	reject

Table 7-8: Timbre Circle rated by Pedestal criteria

The investigation shows that the Timbre Circle does not satisfy the criteria of Order, because one of the complementary triplets sounded ordered in brightness. It does however satisfy all the other criteria. The timbres segregate at very low rates, and the categories seem to be strongly dissociated so that complementary and adjacent classes are equally different. This indicates that all of the categories are equally different from one another, and so should clearly segregate in a higher level display.

7.5.10 Comparing the Pedestals

The results of the investigations of each candidate pedestal are summarised in Table 7-9. The tests where the null hypothesis was rejected are indicated by ‘ok’, which means that the criteria is supported. Tests where the null hypothesis was accepted are shown by ‘fail’ and a number in brackets indicating the number of sub-tests that failed. This allows a comparison of degree of failure across the different pedestals. In the Level test a fail rating of X indicates that the categories did not segregate even at the fastest rate of 50 ms onset.

From this summary we can see that the dynamic Timbre Circle had the best overall rating against the Pedestal criteria. All of the prototypes passed the Zero, Difference and Range tests. This indicates that each variation could produce 8 discriminably different sounds, in which no one sound was heard to be an outlier that could indicate a zero in a repeating cycle.

Pedestal	Zero	Difference	Order	Metric	Level	Range
Pitch	ok	ok	fail (4)	ok	fail (X,X)	ok
Formant	ok	ok	fail (2)	ok	fail (110,X)	ok
Static Timbre	ok	ok	fail (4)	fail (3)	ok (150,110)	ok
Timbre	ok	ok	fail (1)	ok	ok (150,150)	ok

Table 7-9: Comparison of the Pedestals

The Order test strongly ruled out the Pitch circle by highlighting the likelihood that categorical data would be heard as ordered with this configuration. The underlying opponent axes of the Static Timbre Circle were observable and separable from the timbre to such an extent that the ordering in terms of the axes themselves could be plainly heard. This allowed an appreciation of the need to scale the circle because the metric of the underlying axes could be clearly heard. The Formant circle seems to be a good candidate, and perhaps a scaling could address the problems of order that were heard with 2 triplets. The Level test showed that adjacent Formant categories grouped very strongly, whilst complementary categories segregated strongly. This is an interesting property of distance relating to grouping strength that could be useful in some types of higher level displays. However it does not fit the criteria of equal difference between categories that we have proposed. Only the dynamic Timbre Circle performed well on the Level test, and it also performed best on the order test. For this reason the Timbre Circle will be used as a basis for the rest of the investigation. The variations in other parameters need to be investigated in terms of the timbres from this pedestal.

7.6 The radial spokes

The pedestal provides a platform for the rest of the Information-Sound Space to sit on. The next stage of investigation is to fill in the pedestal with a disc of radial variation. This variation can be thought of as spokes that radiate from the centre of the pedestal out to the categorical node on the perimeter of each segment. This section is a pilot study to investigate the radial spokes. The investigation is framed by the theory that a radial variation can represent quantities for comparison, without altering the perceptual category. The radial dimension has the characteristics of difference, order, metric and an original zero. The spokes can only exist if they can be observed throughout every category in the pedestal. Therefore the radial variation needs to be observable across a general range of timbres, or else highly specialised to the particular timbres in the pedestal. The straightness of the spokes relies on the selection of an aspect of variation that does not cause a change in category as it traverses its range. The choice is constrained by the common point of origin and the need to support smooth transitions through the origin.

7.6.1 Criteria of the radial spokes

A set of criteria that must be satisfied to realise the radial spokes of a Timbre Disc are proposed. A null hypothesis for each criterion is given. A criterion is supported if its null hypothesis is rejected. The tests were carried out by listening to auditory sequences having prescribed relations in terms of Zero, Difference, Order, Metric, Level, and Range.

Zero - radial zero is a common point of origin across categories.

Null hypothesis: the radial zeros in each category are not similar
Zero sequence: the zero from each category + random 50% point

The listener hears repeating cycles of the sequence. The task is to identify the most dissimilar point. The null hypothesis is accepted if there is consistent choice of the point with 50% brightness.

Difference - different elements sound different

Null hypothesis: two or more elements of the sequence are identical
Difference sequences: repeating cycles of 8 steps along the radial.

The listener hears repeating cycles of the radial sequence. The task is to listen for level regions or turning points in the sequence. These points are where different values sound the same. The null hypothesis is supported by the consistent identification of a level point or a turning point.

Order - ordered subsets sound ordered

Null hypothesis: ordered subsets do not sound ordered
Order sequence: repeating cycles of 8 steps along the radial.

The listener hears repeating cycles of the radial sequence. The task is the same as the difference task - to listen for level regions or turning points in the sequence. These regions indicate the possibility that an ordered subset will not be heard as ordered. The null hypothesis is supported by the consistent identification of a level point or a turning point.

Metric - regular intervals sound regular

Null hypothesis: regular intervals along the radius do not sound regular
Metric sequence: ordered triplets with regular spacing 25%,50%,75%

The listener hears sets of ordered triplets with regular spacings. The task is to identify the most similar pair in each triplet. The null hypothesis is accepted if there is a consistent pairing that indicates irregular spacing in a triplet.

Level - difference influences sequential grouping strength.

Null hypothesis: radial difference does not influence sequential grouping strength
Level sequence: pairs of tones in the XOX-XOX galloping sequence.

Three test sequences are generated at 100 ms rate. The first has maximum difference, the second is 50% maximum difference and the third is 5% maximum difference. The listener is asked to try to hear the sequences as single repeating sounds. The task is to answer whether there is more than one sound perceived in each sequence. The null hypothesis is supported if the answers to all three tasks are consistent- either they all had only one sound, or all had more than one sound.

Range - there are 8 discriminable steps

Null hypothesis: there are not even 3 discriminable steps
Range sequence: triplet with regular spacing 25%,50%,75%

The listener hears a slowly repeating triplet with regular spacings. The task is to identify whether there are one, two or three different sounds. The null hypothesis is accepted if less than 3 different sounds are consistently heard.

7.6.2 Selecting an auditory relation as a radial spoke

The choice of an auditory relation that can satisfy the criteria of the radial spokes is restricted to those that have order, a metric and an original zero. This section will investigate the choice of a radial component for an Information-Sound Space. Some candidates are listed in Table 6-11, and include drum stretch, fuzz level, vibrato rate, tremolo depth, and brightness. However let us start with the previous models of Caivano and Padgham.

The radial component in Caivano's model is spectral complexity that varies from a white noise at the origin to a pure tone at the other extreme. This variation is closely related to the dull/bright and compact/scattered bases of the Static Timbre Circle, tracing a path from point 5 to point 3 in that space. The segregation of categorical points in this space is an indication that a change in timbre category occurs in this variation. Caivano's example of a radial sequence is white noise, percussion, kettledrum, guitar, oboe, trumpet, flute, tuning fork. This series of categories is a replacement relation, rather than a transparent modification of the category that we are looking for.

Another candidate is Padgham's radial component which he, like Caivano, calls complexity. However, whereas Caivanos complexity alters both spectral structure and spectral shape, Padgham's radius only affects the spectral shape, covarying with an increased weighting of the upper partials. This variation is closely related to brightness variation - roughly defined as the balance between the upper and lower partials of a sound spectrum. Brightness is a perceptual dimension found consistently in a wide variety of timbre research. The Y axis which is orthogonal to the temporal plane in Grey's MDS study corresponds with the definition of brightness. The order and metric of brightness were demonstrated by von Bismarck, who built a subjective brightness scale using the fractionation technique of doubling and halving perceived values. Bregman identifies brightness as a significant factor influencing sequential grouping, and it is observable in a wide range of musical, everyday and speech sounds [Bregman A.S. (1990)].

In a multidimensional timbre space it may be possible to vary a single aspect and keep the identity of the timbre category stable due to the other unchanged aspects which hold the conservative perceptual classification process in place. Anecdotal support for the separability of brightness from temporally categorised timbre can be found in the common use of brightness filters in recording studios to adjust the timbre of an instrument sound without altering the identity of the instrument. This gives us some insight into how brightness of a timbre category may be modulated. Timbres containing more upper harmonics tend to be brighter. A low pass filter can be applied to reduce the brightness of sounds by attenuating some of the upper harmonics. In this passive filter model each spectrum has a maximum and characteristic brightness when all the harmonics are present, and can be made duller by controlling the filter cut-off frequency. At the dull end of the scale each spectrum tends toward a sinusoid at the fundamental frequency. This supports the polar geometry of the Timbre Disc by allowing seamless transitions across the centre and also accords with Padgham's definition of the complexity radius in his timbre assessment chart.

7.6.3 The criteria of radial spokes applied to brightness

A radial brightness parameter was added to the Csound TimbreCircle instrument. The brightness variation was implemented by linearly adjusting the cut-off frequency of a first order low-past filter. The parameter range [0,127] moves the cut-off from the fundamental to the Nyquist. The test sequences that were generated were an eight-step change in brightness for each timbre category, and galloping triplets of close, mid and far radial difference at onsets spacings of 100ms for timbre 3. The ratings of these sequences against the criteria are shown in Table 7-10.

Criteria	Result	Null hypothesis
Zero	repeating cycle of 8 steps can pick the bright point in the cycle	reject
Difference	repeating cycle of 8 steps 0 = levels off at the end 1 = no level or turning 2 = no level or turning 3 = no level or turning 4 = levels off at the end 5 = no level or turning 6 = levels off in the middle 7 = no level or turning	accept
Order	repeating cycles as for Difference	accept
Metric	regular triplet the levelling off found in the repeating cycles supports the null hypothesis of non-uniform change	accept
Level	4 = far = 2 sounds 4 = mid = 2 sounds 4 = near = 1 sound	reject
Range	regular triplet 1 = count 3 2 = count 3 3 = count 2 4 = count 2 5 = count 2 6 = count 1 7 = count 1	accept

Table 7-10: Brightness rated against Radial criteria

The brightness radius failed on the criteria of Order, Difference and Metric because many sequences were heard to level-off in brightness. However the fact that none of the sequences turned back on itself suggests that to a first approximation the variation is simply ordered. The levelling-off does point to an important problem of linearity and scaling in perceptual spaces. The brightness variation of the synthesis algorithm is not perceptually uniform and different acoustic spectra have different brightness characteristics. The difference and order criteria may be satisfied by scaling of the brightness variation to ensure equal steps in brightness along the radius.

The Brightness did not fare well against the Range criteria either. Differences in the sound due to brightness steps were very subtle, and could mainly be heard as squeaky upper harmonics. These observations run counter to the wide range of brightness that can be heard when the source is a broad spectrum. The problem may be in the limited spectral spread of the samples which contracts further at lower pitches. The Range criteria could be addressed by providing information about the number of equal steps available at each pitch of each sample. The designer could then ensure the necessary range is available for a particular sound. Brightness can also be affected by the frequency response of an output device which can reduce the available dynamic range. The definition of device characteristics can help prevent saturation effects caused by extending the sequence out of the available range.

The Zero test was passed because it was easy to hear the non-zero outlier in the sequence. However the brightness zero points at each timbre did sound different, due to the temporal variations. These points all had a characteristically dull sound, but they were not identical.

In the Level test the segregation was heard as an extra sound like a high squeak that becomes more pronounced with difference in brightness. This effect can be explained by the old+new heuristic. The common portion of the sounds group leaving the extra brightness components in a the higher squeaky stream of their own. At faster rates there is an interaction between the onset transient and the onset rate. The flute has a long attack and is difficult to use in the Van-Noorden type of sequences.The instruments with short attack segments are best for investigating grouping by streaming.

7.7 The vertical axle

The IPS is completed by fixing the disc of radial spokes on a vertical axle. The variation in this dimension must be observable and ordered everywhere in the space. This axle also contains all the original zero points of the radial scales. In colour models it is sometimes called the “grey” axis because all the desaturated points lie along it stretching from the dark point to the light point. The vertical axle has the following requirements:

- observable separability throughout each category
- perceptual orthogonality to the radial dimension
- a perceptually scaled metric
- a natural zero

7.7.1 Criteria of the vertical axle

A set of criteria that must be satisfied to realise the vertical axle of the complete Information-Sound Space are proposed. A null hypothesis for each criterion is given. A criterion is supported if its null hypothesis is rejected. The tests were carried out by listening to auditory sequences having prescribed relations in terms of Zero, Difference, Order, Metric, Range and Level.

Zero - the variation has a natural zero

Null hypothesis: the vertical variation does not have a natural zero.
Zero sequence: repeating cycles of the vertical variation.

The listener hears repeating cycles of a vertical sequence. The task is to identify the point at which the sequence cannot be heard. This is a natural zero that is an absolute anchor point for all vertical sequences. The null hypothesis is accepted if the point of disappearance cannot be consistently identified.

Difference - difference is heard as difference

Null hypothesis: two or more elements of the sequence are identical
Difference sequences: repeating cycles of a vertical sequence.

The listener hears repeating cycles of the vertical sequence. The task is to listen for level regions or turning points that indicate points of repetition. The null hypothesis is supported by the consistent identification of a level point or a turning point.

Order - ordered subsets sound ordered

Null hypothesis: ordered subsets from the vertical do not sound ordered
Order sequence: repeating cycles of a vertical.

The listener hears repeating cycles of the vertical sequence. The task is the same as the difference task - to listen for level regions or turning points in the sequence. These regions indicate the possibility that an ordered subset will not be heard as ordered. The null hypothesis is supported by the consistent identification of a level point or a turning point.

Metric - regular intervals sound regular

Null hypothesis: regular intervals up the axle do not sound regular
Metric sequence: ordered triplets with regular spacing

The listener hears sets of ordered triplets with regular spacings. The task is to identify the most similar pair in each triplet. The null hypothesis is accepted if there is a consistent pairing that indicates irregular spacing in a triplet.

Level - difference influences sequential grouping strength.

Null hypothesis: vertical difference does not influence sequential grouping strength
Level sequence: pairs of tones in the XOX-XOX galloping sequence.

Three test sequences are generated at 100 ms rate. The first has maximum difference, the second is 50% maximum difference and the third is 5% maximum difference. The listener is asked to try to hear the sequences as single repeating sounds. The task is to answer whether there is more than one sound in each sequence. The null hypothesis is supported if the answers to all three tasks are consistent- either they all had only one sound, or all had more than one sound.

Range - there are of the order of 100 discriminable steps

Null hypothesis: there are less than 20 discriminable steps
Metric sequence: A triplet with 5% spacing

The listener hears a slowly repeating triplet with regular spacings. The task is to identify the whether there are one, two or three different sounds. The null hypothesis is accepted if only one or two sounds are consistently heard.

7.7.2 Selecting an auditory relation as a vertical axle

The choice of an auditory relation that can satisfy the criteria of the vertical axle is restricted to those that have difference, order, a ratio metric, and a natural zero. This section will investigate the choice of a vertical component for an Information-Sound Space. Some candidates are listed in Table 6-11, and include repetition rate, tempo, duration, force and loudness. The vertical dimension chosen by both Caivano and Padgham is loudness. Loudness is related to the overall energy of the spectrum, and was ratio-scaled by Stevens in a similar manner to lightness. Loudness can be observed in all sounds. Therefore loudness seems like a good candidate for a vertical axle.

7.7.3 The criteria of vertical axle applied to loudness

An intensity parameter was added to the TimbreCircle instrument. The intensity parameter varies linearly from [0, 90] dB. This is only a first approximation to loudness because the timbre, duration, pitch and other aspects can significantly influence how loud a sound is heard to be. The test sequences that were generated were an eight step change in loudness for timbre 3, a triplet of loudnesses at regular spacings for each timbre, and galloping triplets of differences at 100ms onsets for timbre 3. The ratings of these sequences against the criteria are shown in Table 7-10.

Criteria	Result	Null hypothesis
Zero	repeating cycles silence at the start of each cycle	reject
Difference	repeating cycle of 5% steps 0 = levels off 1 = no level or turning 2 = no level or turning 3 = no level or turning 4 = no level or turning 5 = no level or turning 6 = no level or turning 7 = no level or turning	accept
Order	repeating cycles of 5% steps 0 = levels off 1 = no level or turning 2 = no level or turning 3 = no level or turning 4 = no level or turning 5 = no level or turning 6 = no level or turning 7 = no level or turning	accept

Table 7-11: Loudness rated against Axle criteria

Criteria	Result	Null hypothesis
Metric	regular triplets 0 = unequal 2,3 are similar 1 = equal 2 = equal 3 = equal 4 = equal 5 = equal 6 = equal 7 = equal	accept
Level	3 = far = 2 sounds 3 = mid = 2 sounds 3 = near = 1 sound	reject
Range	repeating cycle of 5% steps 0 = count 2 1 = count 3 2 = count 3 3 = count 3 4 = count 3 5 = count 3 6 = count 3 7 = count 3	accept

Table 7-11: Loudness rated against Axle criteria

Loudness failed against Difference, Order, Metric and Range criteria. This failure was entirely due to one point in the repeating cycle of 5% loudness variation of timbre 0 (the flute). This point indicates that some caution should be used with assuming a metric of loudness difference within and across timbres. Nevertheless the result for all the other timbres indicates that loudness has good potential as a vertical axis. The zero criteria was satisfied because a silent point can be consistently detected in a repeating sequence that spans the range. The difference test illuminated a potential problem with loudness in a display situation. Although the difference steps were set in absolute units of dB the actual loudness that is heard depends on the volume setting of the display device. A detectable difference at one volume setting becomes undetectable at another.
The order of loudness was clear in most of the repeating cycles, which had an obvious start, middle and end. The metric test found that steps in loudness were of similar size within each timbre, though identical dB levels were heard as different loudnesses across the timbres. The unusual flute point indicates the non-linear perceptual response to intensity difference.

The Level criteria was passed because intensity difference did affect streaming. This effect was unexpected, as Bregman has concluded that there is little evidence for primitive grouping by loudness [Bregman A.S. (1990)]. The effect was observed at 100 ms onset rate with timbre 3 (tenor saxophone), but can also be heard with the other timbres. The sequence is heard as a single sound when the loudness difference is nominally 4.5 dB. At medium (18 dB) and large (36 dB) differences a squeaky sound segregates. This effect is similar to the brightness segregation, and may also be explained by the old+new heuristic. Turning the volume of the display down did not affect the streaming. This is an indication that it is not intensity difference that is causing the effect, but an interaction between intensity and frequency spectrum that is preserved at different playback volumes. This interaction may be due to a spread of energy into higher partials at higher intensity levels.

The range of loudness is an order of magnitude greater than brightness. This range provides the possibility to enable discrimination of fine differences in a wide range of variation which may be necessary for some representations. Although loudness is a good match to the requirements of the vertical axle, this investigation highlighted two problems. The first is a serious issue of ergonomics that has not previously been considered in the TaDa design approach at all. During the course of the experiment some sudden loud sounds were generated that were very uncomfortable, and potentially dangerous, particularly if wearing headphones. The range of loudness should probably be kept small, to prevent unexpected shocks when a parameter approaches an extreme. The second issue is the lack of calibration on a display device. The ability to easily change the loudness of most devices from the front panel is a necessary ergonomic feature, but this adjustment changes the characteristics of a loudness sequence dramatically, perhaps distorting difference and metric relations.

7.7.4 The criteria of vertical axle applied to duration

Time-based relations are independent of the spectral characteristics of a display device that can affect loudness relations. The duration and repetition rate of sounds are time-based relations that were ratio-scaled by Stevens and Galanter [Stevens S.S. and Galanter E.H. (1957)]. They found very good correspondence between the perceived factor and the physical stimulus - indicated by psychophysical constants of 1.1 for duration of white noise, and 1.0 for repetition rate of a tone. This section investigates duration as a candidate for the vertical axle. Duration can occur at different scales from milliseconds to hours. In this investigation the order of magnitude is 0.1 second, which allows for interactive queries at a normal human rate. The test sequences that were generated were an eight step change in duration for timbre 3, a triplet of durations at regular 5% increments, and galloping triplets of far, mid and close differences at 100ms onsets for timbre 3. The ratings of these sequences against the criteria are shown in Table 7-10.

Criteria	Result	Null hypothesis
Zero	repeating cycle of 8 equal steps in duration silence at the start of each cycle	reject
Difference	repeating triplet of 25% steps 3 = difference	reject
Order	repeating triplet of 25% steps 3 = ordered	reject
Metric	repeating triplet of 25% steps 3 = steps 2 and 3 seem slightly closer	accept
Level	3 = far = 2 sounds - a flap and a tone 3 = mid = 1 sounds 3 = near = 1 sound	reject
Range	repeating cycle of 5% steps 3 = count 1	accept

Table 7-12: Duration rated against the Axle criteria

Duration in the 0.1 to 1 second range failed against the Metric and Range criteria. Durations of 0.5 and 0.75s seemed more similar than the 0.25s and 0.5s pair. However the main problem with duration was the difficulty of making fine judgments. The Range test showed that differences of 0.05 seconds could not be heard. Stevens' ratio measure was obtained for durations in the range 0.25 to 4 seconds. This is perhaps the scale at which a higher dynamic range of duration perception could be available.

7.7.5 The criteria of vertical axle applied to pitch

The vertical axle in the IPS has 100 ordered steps and a natural zero. However the radial axis provides an axle of original zeros that can satisfy the zero requirement of the space. Therefore it may be more important to include a dimension which can provide a good level of resolution. The pitch dimension has been scaled with a suitable number of equal steps, and may be a substitute that allows the ISS to still fulfil the IPS requirements of difference, order, metric and zero. This section will investigate pitch as a candidate for the vertical axle. The test sequences that were generated were a triplet of pitches at regular 5% increments, and galloping triplets of far, mid and close differences at 100ms onsets for timbre 3. The rating of these sequences against the criteria are shown in Table 7-10.

Criteria	Result	Null hypothesis
Zero	silent pitch is not possible	accept
Difference	repeating triplet of 5% steps 3 = difference	reject
Order	repeating triplet of 5% steps 3 = ordered	reject
Metric	repeating triplet of 5% steps 3 = equal	reject
Level	6 = far (30 semitones) = 2 sounds 6= mid (10 semitones) = 2 sounds 6= near (2 semitones) = 1 sound	reject
Range	repeating cycle of 5% steps 3 = count 3	reject

Table 7-13: Pitch rated against the Axle criteria

The pitch axle necessarily fails the criteria of a natural zero. However a conventional zero can be heard as an original zero in a pitch cycle because of the ordered nature of pitch. All of the other criteria of difference order, metric and range are satisfied. In particular it is very easy to discriminate the 5 semitone steps that indicate a useful range. A problem arises with the use of sampled instruments because each ranges over a different part of the pitch axis, and none of them extend over the entire audible pitch range. Further the timbre of a musical instrument can change significantly over its pitch range, and may not be recognisable at extremes in its pitch range. The characteristics of the display samples can make it difficult to select ranges, and may constrain the available variation. The designer needs to be aware of the pitch range of each instrument to make the most of its dynamic range. The segregation due to pitch was very strong in the Level test. Only at very close range did the pitches group. This indicates that pitch difference can cause categorical effects that may override the categorical circle. The cohesion of categories may require that the dynamic range of pitch be limited to only a few semitones. This limitation may in turn influence the range of steps in pitch that can be perceived.

Pitch is quite robust to device characteristics, although very high pitches may be affected by device frequency characteristics.

7.8 A prototype of an ISS

The previous section investigated the potential of various auditory candidates for the dimensions of an Information-Perception Space. A subset of the most satisfactory candidates may be combined to form an Information-Sound Space (ISS) which can represent a general range of information relations in sound. This ISS will be organised to have perceptual characteristics of difference, order, metric, zero, level and range that are necessary to represent the TaDa information types.

The Timbre Circle was the most effective of the four Pedestals that were tested. The Timbre Circle consists of a subset of subjectively equally spaced musical instrument timbres. It provides a platform for data mappings that preserve the unordered difference between elements mapped to sounds, so that it may veridically represent categorical data. The properties of the Timbre Circle rely on the relationships between the component timbres rather than on absolute identities, so that alternative timbre schemas may add semantic connectivity between an application and a data set, to build for example a “medical” scheme or an “underwater” scheme from sounds that are familiar and may have an association in these contexts.

The radial axis that was tested was Brightness, which was found to be a separable and observable across timbres in the investigation of the Static Timbre Circle. Brightness is related to Padgham’s radius, and is a principal component of static and dynamic timbres that has been widely reported in experiments by Von Bismarck [Von Bismarck G. (1974a)], Plomp [Plomp R. (1976)], Wessel [Wessel D. (1985)], Grey [Grey J.M. (1975)] and others. There is a problem with choosing brightness in that it did not fare particularly well against the radial criteria. The main problem was the non-uniform variation of the brightness, and the saturation of the brightness in different timbres. These problems can be addressed by perceptually scaling the brightness dimension, as demonstrated by Von Bismarck [Von Bismarck G. (1974b)].

Three candidates were tested for the vertical axis - loudness, duration and pitch. Loudness satisfied all of the criteria, although there was a hint that some scaling might be required in the case of the flute sample which has a long slow attack that noticeably influences the relation between loudness and intensity. There were two problems with loudness which make it less attractive than other aspects - the ergonomic need to prevent startling and dangerous loudnesses, and the variability of the output range due to user control of the volume knob. An alternative natural zero is the duration of an auditory stimulus. However it was found that duration could not provide the 100 levels of discrimination needed for this axis, because differences of the order of 0.1s were not discriminable. The final test was of pitch, which does not have a natural zero, but does have a good range. The pitch axis may be an acceptable compromise because the zero that is a necessary characteristic of the IPS is provided by the brightness radius which has an original zero.

The Timbre Circle, Brightness Radius, and Pitch Axle can be combined to form a prototype Information-Sound Space. This Timbre-Brightness-Pitch (TBP) space, shown in Figure 7-15, has perceptual properties derived from the abstract Information-Perception space by assigning aspects of sound perception to the perceptual axes in accordance with the organisational principles. The TBP model is a polar cyclindrical space that revolves around a pedestal of 8 equally different timbres. The radial axis is 8 equal steps in brightness, and the vertical axle is 100 equal steps in pitch. There are other permutations that may also satisfy the ISS criteria, but the TBP basis is an initial point for further investigation. Note that the motivation for the space is not to describe hearing perception but to support a method of data-sensitive auditory display. The properties of the TBP model are intended to mirror Munsell's ideals of psychological equispacing and practical usefulness.

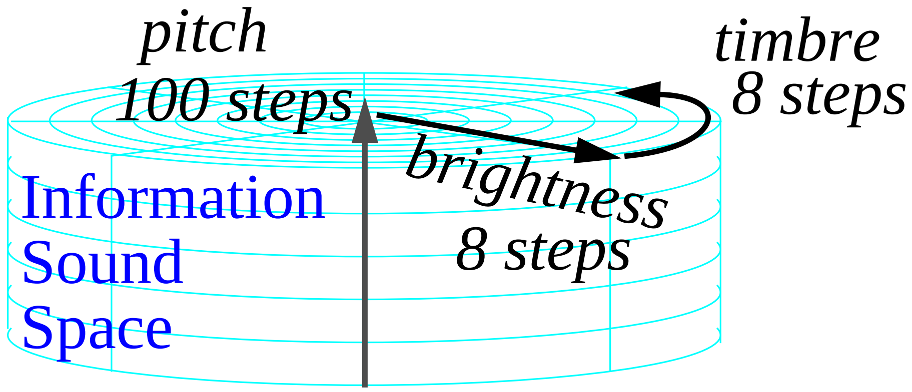


Figure 7-15: The TBP prototype of an Information-Sound Space (ISS)

The TBP sound model was developed to have properties similar to those of the HSL colour model. The advantages of the TBP model are:

- Natural specification, comparison and matching - Timbre, Brightness, and Pitch are perceptually separable attributes of sounds.
- Natural order - the Timbre Circle is ordered by an underlying perceptually orthogonal basis which arranges complementary timbres diametrically opposite each other. The Brightness and Pitch axes both have a natural order.
- Independent control of perceptually aligned parameters - Timbre, Brightness, and Pitch can be changed independently.
- Geometric interface - the 3D sound solid provides the opportunity for spatial interaction with sounds.
- Transportability - the TBP model may be used to specify sounds in natural terms rather than device coordinates.

7.8.1 Representational Mappings in TBP ISS

The various information paths in the Information-Perception Space become auditory representations when they are mapped to the TBP Information-Sound Space. The angular axis is qualitative timbre difference, the radial axis is quantitative brightness difference with order, metric and a zero, and the vertical axis is quantitative pitch difference with order and a metric. The elementary auditory relations in the TBP ISS are shown in Table 7-14.

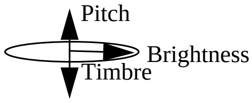
Info Type	ISS Mapping 	TBP description
Boolean	Opposite angles	2 very different timbres
Nominal	Circle	up to 8 categorically different timbres
Ordinal	Coil	categorically different timbres ordered by pitch
Ordinal and zero	Spiral	categorically different timbres, ordered by pitch, with dull zero
Ordinal bilateral	Sloped Line	dull central zero, -ve and +ve category timbres, ordered by pitch
Interval	Vertical Line	ordered change in pitch
Ratio	Radial Line	ordered change in brightness, starting from a dull zero

Table 7-14: Elementary representation mappings in TBP ISS

7.9 The SoundChooser

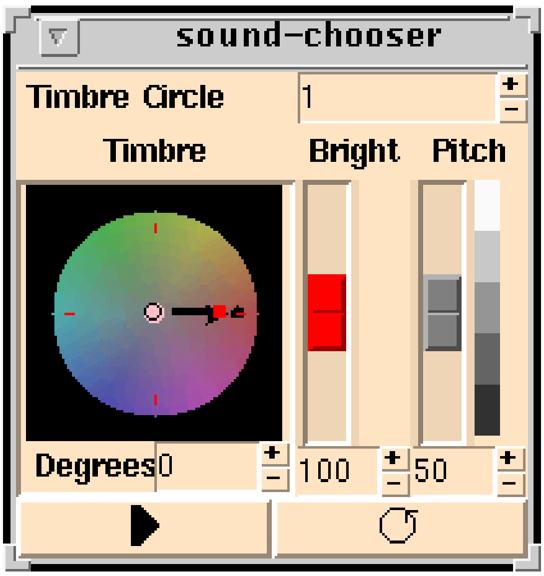


Figure 7-16: SoundChooser

The usual interface to sound in computer and electronic music systems is a list of around 200 verbally described timbres (e.g. muted trumpet, pizzicato strings), and a number of parameters which allow local variations in the timbre (e.g. reverb depth, vibrato rate). However the list-style interface only provides local information about a timbre and does not allow an overview of the range of timbres or the relations between them. This makes it difficult to find a particular timbre in the list, and to structure timbre relations in terms of similarity and difference. The SoundChooser is a graphical user interface for selecting timbres that is modelled on the familiar Colour Chooser. The interface, shown in Figure 7-16, consists of a dial with an arm which rotates through 360 degrees. This arm can be directly manipulated to select a timbre angle, or can be set with a numeric entry box. On the dial arm is a bead which is the radial brightness. This bead may be directly manipulated or set using a slider or a numeric entry. Pitch height is controlled with a vertical slider, or a numeric entry. A “play” button activates the current sound and a “cycle” button causes the dial to rotate and generate a sequence. Different Timbre Palettes can be selected with the numeric entry widget at the top of the panel, for example the Static Timbre Circle is 1, the Formant Circle is 2 and the Timbre Circle is 3. The SoundChooser makes it easy to search for particular sounds, to remember where sounds are, and to compare sounds. The user interface of the SoundChooser was implemented with tk/tcl [Ousterhout J.K. (1994)] and the coordinates were sent to the Csound instrument through a Unix™ pipe.

7.10 Summary

The HSL colour model makes it easy to choose colours and colour schemes, without an in depth knowledge of colour theory or principles of colour design. This type of tool is called a cognitive artefact, because it can help a person to think about a problem in a very direct manner. This chapter recasts the Hearsay principles of auditory design in the form of a cognitive artefact, modelled on the HSL colour model. This cognitive artefact for auditory design is called the Information-Sound Space (ISS) because it aligns perceptual structure with information structure in a spatial organisation.

A blueprint for an ISS was proposed, which has properties that can support the general range of TaDa information types. The criteria of Zero, Difference, Order, Metric, Range, and Level were proposed to rate different aspects of auditory perception as candidates for a prototype realisation of an ISS. The Timbre Circle was the best candidate for the polar pedestal upon which the space revolves. Other polar candidates were, in rating order, the Formant Circle, The Static Timbre Circle, and the Pitch Circle. Brightness was selected as the radial dimension of Timbre that was most separable and observable. It was found that to be acceptable the brightness radius would have to be perceptually scaled to equal units of difference and to prevent saturation effects. There were three candidates for the vertical axle - loudness, duration and pitch. Although loudness was very promising it was ruled out by ergonomic problems and the difficulty of calibration when the user can easily change the dynamic range. Pitch was chosen because it can support 100 steps of difference. The three axes were combined to form a complete prototype, called the Timbre-Brightness-Pitch (TBP) model. The cylindrical polar system has a categorical timbre angle, an ordered brightness radius, and an ordered pitch axle.

This model of auditory relations has properties that have been derived from the HSL colour model, and may have similar benefits in making the design with sounds more direct and easy to understand. The TBP model embodies both qualitative and quantitative aspects of sound perception and provides a framework for data sensitive mappings which connect data characteristics with perceptual characteristics. Bregman identifies timbre, brightness and pitch as important for the formation of perceptual streams in auditory scene analysis, and it is conjectured that this geometric model may be helpful in the visualisation of streaming, for example selecting points with opposite timbres and large pitch separation would indicate a high likelihood of stream segregation in sequential presentations.

The SoundChooser interface was built to allow interactive selection of sounds from the TBP model. This interface is similar to the familiar colour chooser, and allows the user to quickly select and modify timbres. The observations made in the course of the development of the TBP are encouraging enough to motivate further research in this direction.

7.11 Limitations

The observations and ratings of the auditory relations against the criteria of the Information-Perception Space all came from a single subject - the author. However the nature of the exercise is an exploration, and the methods and ideas are only just coalescing. There will need to be a stage of iteration and consolidation before empirical studies of validity with other subjects can be justified.

The colour model provides access to the entire range of perceptible colours. This is not the case with the TBP sound model. This limitation is due to the multidimensional nature of timbre. The selection of the axes which underly the Timbre Circle constrains the range of timbres to those which can be described in terms of those axes. Access to the greatest possible range of sounds can be enabled by using the most perceptually salient axes. The value of the Timbre Circle is that it allows timbres which are spanned by the nominated axes to be ordered in terms of those axes, irrespective of how they vary in other aspects. The scaling of each axis in isolation is only a first approximation to a perceptually uniform space because it does not take into account interactions between them. The linearity of the space is strongest in the directions aligned to the axes - caution should be exercised in mapping more complex sequences.

Although the axes of the TBP model are ideally orthogonal none of these aspects of sound are truly independent. Even pitch and loudness have influence on each other - so for example a change in pitch can affect perceived loudness even though the intensity remains constant [Zwicker E. and Fastl H. (1990)]. Loudness will vary considerably throughout the model, and this could be corrected using loudness calculation algorithms such as that described in Zwicker and Fastl as ISO standard 532B [Zwicker E. and Fastl H. (1990)].

It is assumed that all sounds are of constant duration, and that sequences are presented at a constant rate. The ISS addresses temporal aspects of sound in terms of timbre. Variations in the durations of sounds in sequences could be addressed by a model of rhythm perception.

The framework does not address harmonicity of simultaneous sounds. For example two sounds separated by an octave in pitch are quite distant in the TBP model, yet are difficult to hear separately when they occur together (due to their spectral harmonicity). The incorporation of the circle of fifths, which describes pitch similarity, into the pitch axis of the TBP model (perhaps in the form of Shephard’s pitch helix) may help here.

The TBP model does not accommodate sounds in that do not have distinct ordered pitches. For example inharmonic bells and gongs can produce sounds with ambiguous pitches, and percussive instruments can produce sounds where the pitch is very weak or non-existent.

The TBP space makes an assumption that timbre is a categorical aspect of sound that causes it to be identified with some musical instrument. However there is no clear definition of timbre given. This is not unusual because timbre is a very ill-defined concept.

8 • GreyMUMS: realisation of an Information-Sound Space

Beyond these categorical distinctions, the essential goal is that perceptually continuous auditory attributes are scaled and mapped to data attributes in a way that is meaningful to the observer. Relevant changes in data should insure a change in what is perceived. Changes in what is perceived should signify meaningful changes in the data. The appropriate scaling functions will probably not exist a-priori. Psychophysical scaling experiments may be needed in order to create perceptual scaling functions through which collections of auditory stimuli are mapped. This is made feasible only by utilizing a limited number of auditory tokens with well-understood perceptual properties. This suggests that sets of tokens be developed and scaled in advance.[Kendall G.S. (1991)]

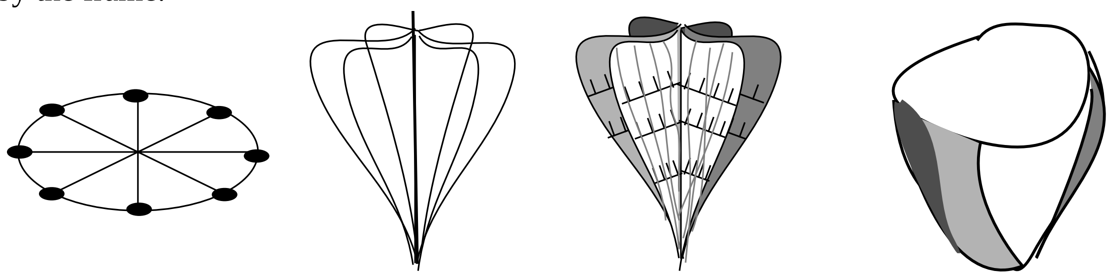
This chapter describes the realisation of an Information-Sound Space (ISS) organised in accordance with the TBP Sound Space developed in the previous chapter. The raw material for the construction is the McGill University Master Samples (MUMS) reference palette of musical samples that is specifically intended for research into musical timbre. The ISS was constructed in 4 stages - the pedestal, the frame, the grating and the plasticine. The pedestal is 8 equally spaced timbre steps organised in a circle by similarity. The frame is an arrangement of brightness profiles which define the limits of dynamic range in pitch and brightness for each timbre. The grating consists of grids of equal differences in brightness and pitch for each timbre. The grating grids are joined on a central axis and radiate outward like segments of a mandarin. The plasticine is a continuous medium moulded to the grating and frame to model the behaviour of the overall space. The resulting sculpture has the property that there is a relationship between distance and the strength of perceptual grouping between points in the space. A vertical line is a pitch scale, and may be used to represent continuous data. A radial line is a scale of equal brightness increments for a timbre, and may also be used for continuous data. A circle of constant radius is a contour of constant brightness across the range of timbres which can be used to represent categorical data. These properties are a rich area for further experiment with data mappings.

8.1 Raw materials

The ISS is constructed from sounds that have been perceptually organised and scaled. The construction process requires psychoacoustic data about the relations between the sounds that are used. The measurement of this data can be a very time-consuming process and requires expert knowledge. However there already exists some measurements of this kind for musical instrument timbres - for example Grey's multidimensional scaling (MDS) study. By choosing a musical palette based on these studies we can use the existing results to build a proof-of-concept ISS. The McGill University Master Samples (MUMS) are a set of compact discs which include a sample of each note in the pitch range of every instrument in a modern orchestra [Opolko F. and Wapnick J. (1995)]. These samples are a reference resource specifically intended for timbre research. The value of this resource is enhanced by the SHARC timbre database which contains a spectral analysis of each MUMS sample that can be used to calculate brightness and other factors [Sandell G.J. (1995)]. This data and the palette of samples are the raw materials needed to realise a proof-of-concept ISS.

8.2 Stages of construction

The Concrete Sound Space is built in four stages, as shown in Figure 8-1 - a “pedestal” of support, a “frame” which defines the boundary, a “grating” which defines the internal behaviour, and the “plasticine” which is a continuum moulded to the grating and contained by the frame.



pedestal
8 equally spaced timbres arranged in a circle

frame
Brightness profiles of each timbre

grating
Brightness scales which define the internal behaviour

plasticine
Interpolation between grating points to form a continuum

Figure 8-1: Stages of construction of the Concrete Sound Space

8.3 The pedestal

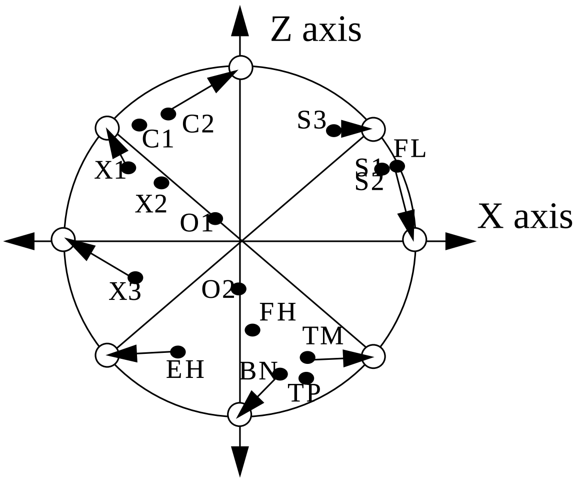
The pedestal is the central organising structure of the space. It is a circle of 8 equally discriminable timbre categories which can represent categorical data types. Different timbres may be substituted into the arrangement if they meet the criteria of equal perceived difference between neighbours. In this instance the MUMS collection of musical instrument samples is the source palette. To construct the pedestal it is necessary to measure the similarity between the timbres in some way. There have been quite a few MDS studies of timbre which provide perceptually scaled results depicting similarity relationships between particular subsets of timbres. MDS has been criticised because the results are based on small sets of stimuli, the scales of measurement are prescribed and therefore biased, it is difficult to identify the physical correlates of the dimensions of principal variation, and new stimuli cannot be inserted into the results without a global restructuring [Pachella R.G. Somers P. and Hardzinski M. (1981)]. Despite these problems there have been some convincing correspondences between MDS and other methods of timbre analysis in terms of identifying principle components of timbre perception in the onset and spectral distribution of a discrete musical sound.
One of the best known MDS studies is Grey’s scaling of 16 re-synthesised instruments. The MUMS palette has sampled instruments which correspond with all of those used in Grey’s experiment, as shown in Table 8-1. The validity of this substitution might be questioned since a repetition of the MDS experiment using MUMS samples in place of Grey’s resynthesised timbres would likely produce different results. However this does not invalidate a categorical substitution where the criterion is relaxed so that the primary structur-

ing is on equal similarity between neighbours rather than on euclidean distance between all pairwise comparisons. The goal of the exercise is to enable the representation of categorical data relationships using the categorical nature of timbre perception, and the substitution is not of timbres but of timbre categories. The categorical substitution can be justified under the assumption that sounds which originated from similar physical sources activated in the same manner (e.g. two different cellos) are more perceptually similar than sounds from sources as physically different as musical instrument families (e.g. a cello and a flute) which are activated in different ways by bowing and by blowing. This remains satisfactory because categorical difference is the essential characteristic required for representing nominal and ordinal data.

Grey's timbre	MUMS sample
flute	flute, vibrato
cello (muted sul tasto)	cello, muted, with vibrato
bass clarinet	bass clarinet
saxophone (mf)	tenor saxophone
soprano sax	soprano saxophone
English horn	English horn
bassoon	bassoon
trombone (muted)	tenor trombone, muted

Table 8-1: Matching Grey's resynthesised timbres with MUMS samples

The results from Grey's MDS study of 16 re-synthesised musical instruments were presented in a series of graphic visualisations which showed where each data point lies in the 3D perceptual space. Equally discriminable timbres are scaled according to identified perceptual axes. The dimensions of principal variation were analysed in terms of the spectrograms of the data points, and it was found that the Y axis was related to spectral energy distribution, whilst the X and Z axes were related to temporal aspects of timbre, covarying with synchronicity in the development of upper harmonics and the presence of low-energy high-frequency noise during the attack segment. The procedure for selection of a timbre circle from this perceptually measured palette is shown in Figure 8-2. A circle which encloses the projection of the data points in the temporal plane is divided into 8 segments of 45 degrees, and the position of each 45 degree increment around the circumference is nominated to represent the categorical timbre of that segment. Because distance is a measure of similarity, the data point in the segment lying closest to each of the equally spaced points on the circumference is allocated to that point. There is only a limited number of data points available to choose from, so that in the segment where there is no data, the closest point from the adjacent segment was used (in this case TM). This is only a first approximation to equal spacing as can be seen by the small difference in distance between FL and its neighbour S3 on one side, and the much greater distance to its other neighbour TM. This is a consequence of the sparsity and unevenness of the palette, which might be addressed by the use of a different set of sounds.



BN = bassoon
C1 = E flat clarinet
C2 = bass clarinet
EH = English horn
FH = french horn
FL = flute
O1 = oboe
O2 = another oboe
S1 = cello (*sul ponticello*)
S2 = cello (normal bowing)
S3 = cello (muted *sul tasto*)
TM = trombone (muted)
TP = trumpet
X1 = saxophone (*mf*)
X2 = saxophone (*p*)
X3 = soprano sax

Figure 8-2: Equally Spaced Timbre Circle in Grey's temporal plane

8.4 The frame

The frame defines the boundary of available variation within the space. The frame is constructed by joining the brightness profile of each timbre to the central seam of the pedestal to form a hollow irregular ball-like shape. This shape represents the upper limits of pitch and brightness for each timbre in the pedestal. The raw material for the frame is the brightness profile which describes how brightness varies with pitch for the outer limits. Each musical instrument has a unique brightness profile which reflects the spectral variations at each pitch due to its physics.

In order to make the frame for the TBP Sound Space, the brightness profile of each GreyMUMS timbre in the pedestal must be obtained in some way. Relative brightness may be measured using perceptual streaming, and this method was used in a previous experiment with a different palette of musical instrument samples [Barrass S. (1994b)]. However these measurements are very time consuming, and require more formal statistical verification across a general survey of individuals. This led to the use of the psychoacoustic model, shown in Equation 8-1 [Zwicker E. and Fastl H. (1990)], to calculate the brightness of an energy distribution. This model has the advantage that the results are calibrated in terms of a standard unit, called the Acum, which is defined as the brightness of a 1 critical bandwidth noise centred at 1 KHz with a sound pressure level of 60 dB.

The calculation requires spectral data for the brightness profiles of each GreyMUMS timbre. The SHARC timbre database contains this data for most of the MUMS samples. The process of selecting a representative spectrum from a time varying sound is described in the SHARC documentation [Sandell G.J. (1995)]

- *The sound file was converted from 44100 Hz to 22050 Hz, and analysed with a Phase Vocoder (PV).*
- *The longest continuous stretch of time in which the note was at 75% or more of its maximum amplitude was identified from the PV information. This located the steady portion of the tone.*
- *An average spectrum was calculated from all the PV frames identified in the previous step. Then least squares was used to find the actual PV frame most closely resem-*

bling this average spectrum. The point in time corresponding to this PV frame was designated the "representative point".

The brightness of representative spectral frames was calculated with Equation 8-1. The upper integral is the first moment of specific loudness over the critical band-rate spectrum, where Bark is the unit of critical bandwidth [Zwicker E. and Fastl H. (1990)]. The lower integral is the overall loudness of the spectrum. N is the summed energy in each critical band, and g(z) is a weighting factor which accounts for the strong increase in brightness at high frequencies.

$$Acum = 0.11 \frac{\int_{0}^{24Bark} N\mathrm{g(z)dz}}{\int_{0}^{24Bark} N dz}$$

Equation 8-1: Acum calculation

A graph for g(z) is provided by Zwicker, but no equation. In order to make calculations a first order approximation was made by measuring values from Zwicker's graph at 1 Bark intervals in the region from 17 to 24 Bark where the weighting takes effect, as shown in Table 8-2.

Bark	0-16	17	18	19	20	21	22	23	24
g(z)	1.0	1.1	1.2	1.3	1.5	1.7	2.0	3.0	4.0

Table 8-2: Critical band rate weighting factor g(z)

The calculation requires the spectral data to be in band-rate form, which models frequency spacings along the basiliar membrane. The linear frequency spectra of the SHARC data can be converted to band-rate, prior to the brightness calculation, with Equation 8-2 [Zwicker E. and Fastl H. (1990)]. The unit of band-rate is the Bark, frequency is in KHz, the angles returned from the arctan expressions are in radians. When Barks is an integer, f is the frequency of the dividing line between critical bands.

$$Barks = 13arc\tan(0.76f) + 3.5arc\tan\left(\frac{f^2}{56.25}\right)$$

Equation 8-2: Bark calculation

The brightness profile calculated for each GreyMUMS timbre is shown in Figure 8-3. Note that the saxophones have dummy profiles set to 1 Acum because, as yet, there is no SHARC data for these particular instruments.

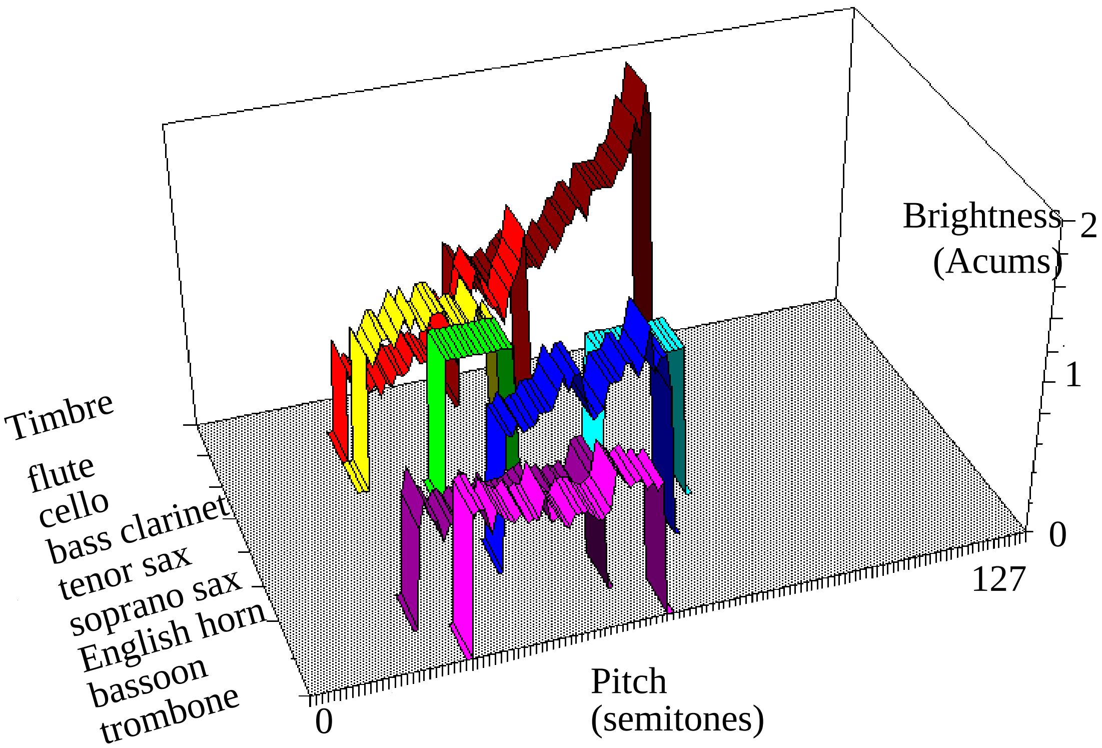


Figure 8-3: Brightness profiles of GreyMUMS timbres

8.5 The grating

The grating defines the internal behaviour of the sound space. It is a grid of points which define regular steps in brightness, pitch and timbre inside the frame of the Sound Space 'sculpture'. The construction is carried out by measuring a brightness scale for each timbre in the pedestal. A 2D grating segment is extrapolated for each timbre by calibrating this scale to the brightness profile over the range of pitches. Each segment has a different shape due to its particular pitch range, brightness profile and brightness scale. The 3D grating is completed by arranging the segments together inside the frame, like pieces of a mandarin. This section will describe the process of measuring the brightness scale to build a grating segment.

The fractionation technique is a way to build a perceptual scales from judgements of half and double of some perceptual continua. Scales that have been built this way include the Mel scale of pitch and the Sone scale of loudness. Fractionation has also been used to build an ordered scale of equal brightness steps for various static spectra [Von Bismarck G. (1974b)]. Through informal observation it was found that the brightness of the time varying MUMS samples can also be systematically varied by applying a low pass filter to attenuate the upper frequencies. This led to the application of the fractionation technique to create brightness scales for the GreyMUMS samples. All scales were measured at MIDI note-number 48, except for the soprano sax which was measured at the beginning of its pitch range at MIDI note-number 61. The maximum brightness of each scale differs for each timbre. This maximum brightness was divided into 8 equal steps by fractionation. The dependent parameter had a range of [0 to 127] which adjusted the cutoff frequency of a first order low pass filter in a linear manner between the fundamental frequency and the Nyquist frequency (22.5 KHz) of each sample. The results of the experiment are shown in Table 8-3 and in Figure 8-4 where each line is the filter cut-off for equal brightness steps of a GreyMUMS sample. These results are merely indicative, and were obtained for only one subject (the author). The purpose in showing the data here is to illustrate the construction process. The results were further validated by a repetition of the experiment a week later with a high correlation with the previous data.

Timbre	Bright max	Pitch	*0*	1	2	3	4	5	6	7	8
flute_ vibrato	0.93	48	*0*	8	20	30	50	65	85	100	127
cello_muted _vibrato	1.1	48	*0*	15	25	40	50	70	85	108	127
bass_ clarinet	0.8	48	*0*	25	35	55	70	85	100	110	127
tenor_sax	1	48	*0*	8	20	38	50	65	80	100	127
soprano_sax	1	61	*0*	5	10	15	20	35	50	80	127
English_ horn	1	48	*0*	10	20	30	45	65	80	100	127
bassoon	0.63	48	*0*	7	15	30	60	75	90	110	127
trombone_ muted	0.76	48	*0*	2	6	10	25	35	50	90	127

Table 8-3: Filter position vs Brightness step for GreyMUMS timbres

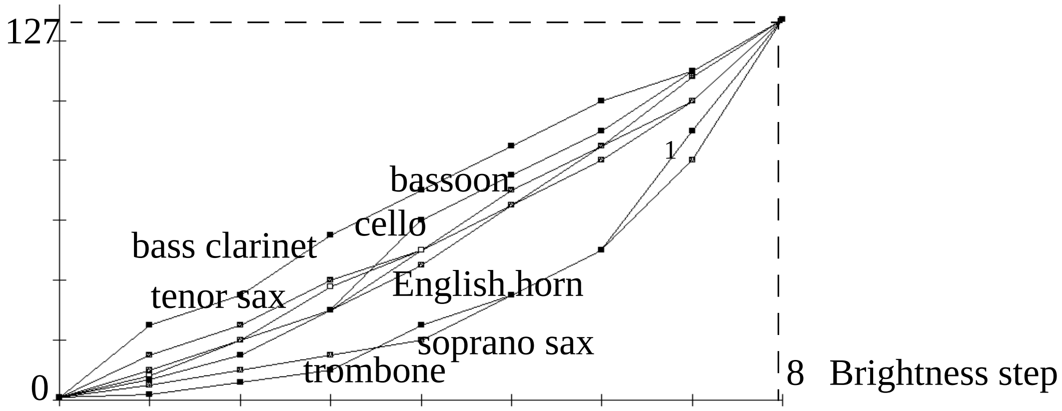


Figure 8-4: Filter position vs Brightness step for GreyMUMS timbres

The scales are internally consistent but the maximum brightness (level 8) is not calibrated across the scales. This calibration can be carried out using the calculated brightness profiles which correspond with brightness 8, and are in referenced units of Acums. The grating segments are built by interpolating the calibrated brightness scale over the pitch range of the brightness profile. This results in a 2D grid of equal brightness and pitch steps at each timbre. These grids are arranged in a 3D polar layout to form the 3D grating of measured data points.

8.6 The plasticine

The plasticine is a continuous ‘material’ moulded to the grating which defines the internal behaviour of the space. Its function is to create a continuum of sounds between the measured points.

The modelling method was designed to represent the behaviour of a colour output device from discrete measurements of its behaviour in a perceptual colour space [Bone D. (1993)]. It consists of a 3D thin plate spline fit to the sparse data points. A 1D mechanical analog of the spline fitting method is shown in Figure 8-5, and can be imagined as a bit like a spring mattress.

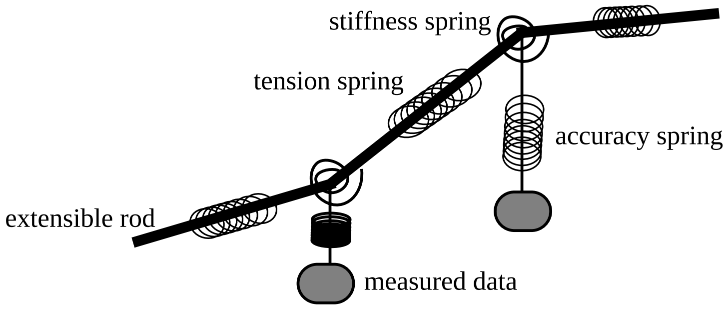


Figure 8-5: 1D mechanical analog of the spline fitting technique

This system of extensible rods joined to form a regular mesh with a spring at each intersection node that affects the amount of curvature. Each node is constrained to move along a vertical line. The sparse data values are attached to this flexible mesh by another set of springs, and the system minimises a global measure of the curvature of the space. This model of the behaviour can be used to calculate intermediate values between the original samples using a 3D linear interpolation. The technique is effective for fitting continuous non-linear multidimensional data - but the sound space grating posed a problem because of the discontinuity between the categorical segments which have to be glued together. The stretchiness of the rods and energy of the springs are parameters which can be adjusted globally at the start of the process to affect the behaviour of the spline material. It was necessary to carry out a systematic search on this parameter space to find a material with properties which could accurately represent the frame and grating behaviour of each segment. The size of the margins between timbres was increased so the interpolation technique could flow more smoothly across them.

The results of the modelling are shown in the following sequence of figures, which depict vertical slices through the sound space together with the calculated frame profile for each segment. It can be seen that there is a good match with the calculated shape.

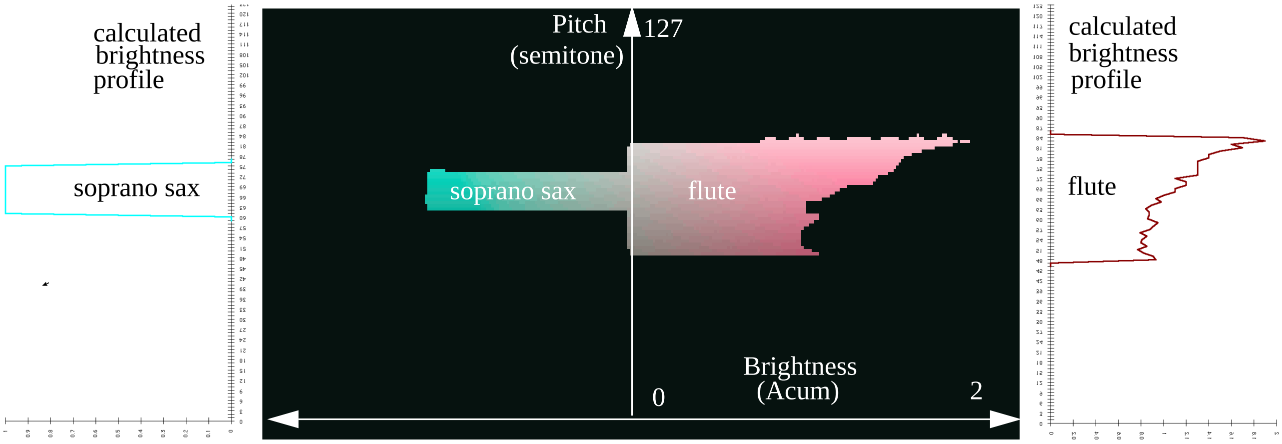


Figure 8-6: Timbre Leaf - soprano sax / flute

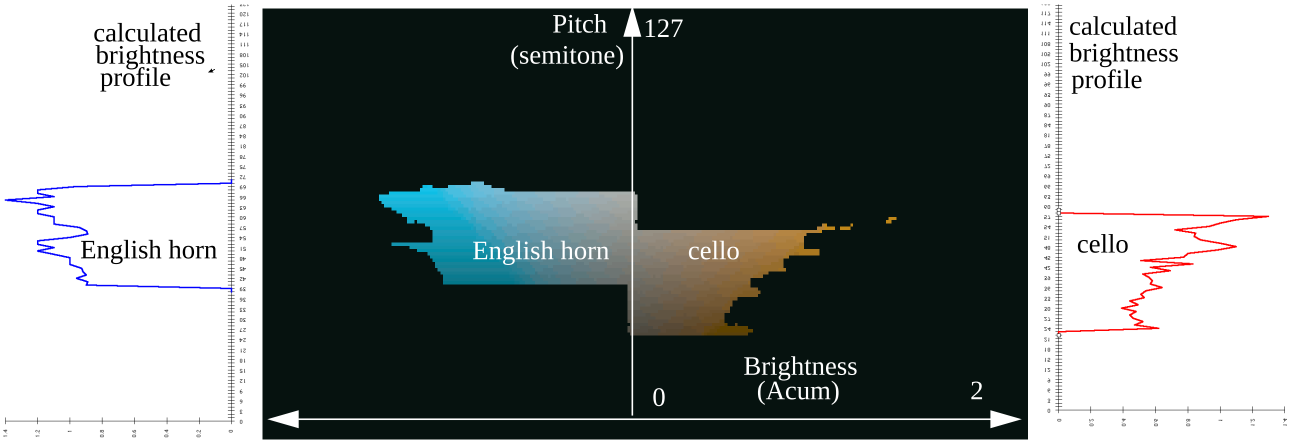


Figure 8-7: Timbre Leaf - English horn / cello

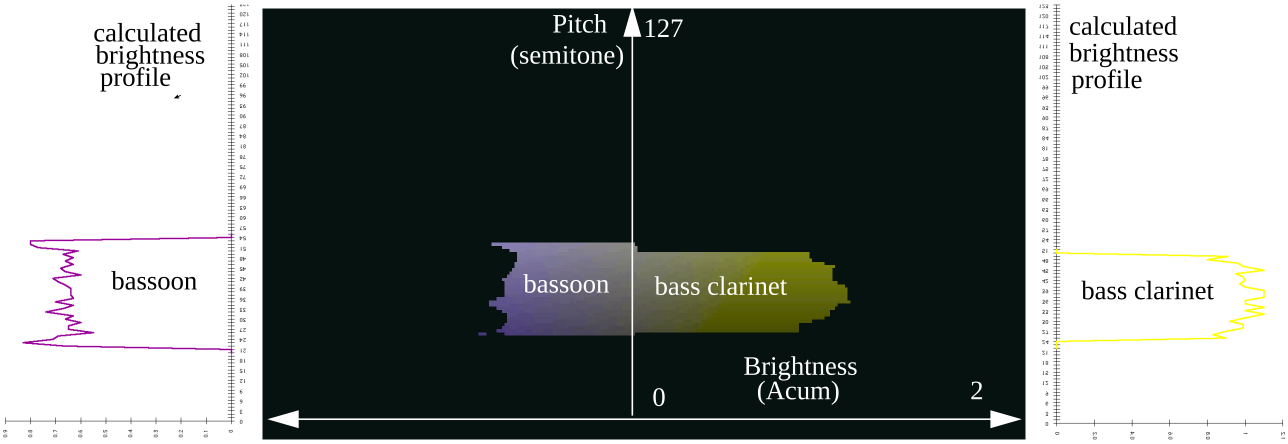


Figure 8-8: Timbre Leaf - bassoon / bass clarinet

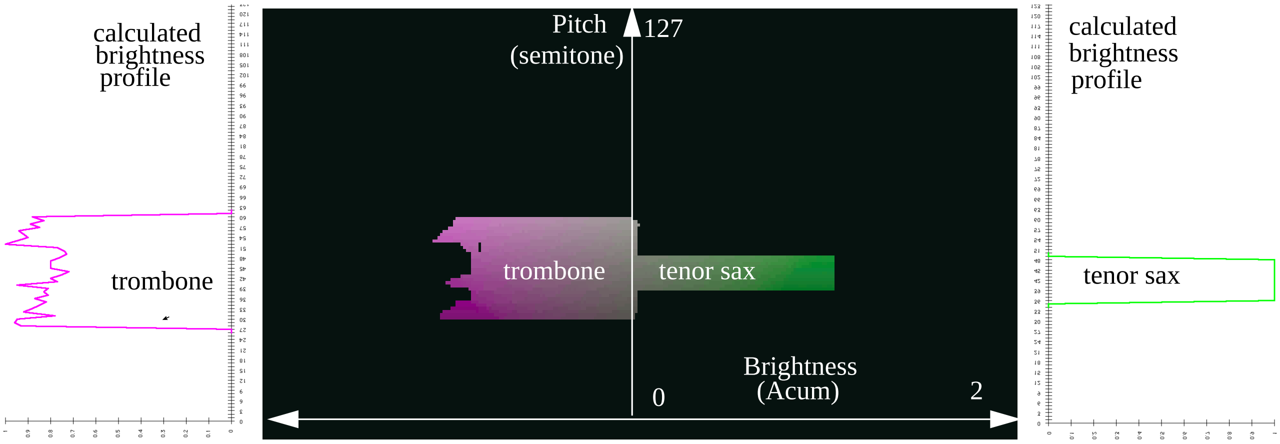


Figure 8-9: Timbre Leaf - trombone / tenor sax

The abstract TBP Sound Space has properties which can be used to verify the internal behaviour of the concrete realisation. A vertical line should be a pitch scale with constant brightness and timbre. The pitch scale and timbre constancy were checked both numerically and by listening. A radial line should be a brightness scale of constant pitch and timbre. This was also checked numerically and by listening. A horizontal slice through the space should have constant pitch but will vary in timbre and brightness. A representative subset of points was checked numerically and by listening. A circle of constant radius should vary in timbre but remain constant in brightness and pitch. The change in timbre and constancy of pitch were checked numerically and by listening. The testing of brightness constancy by listening proved to be a more difficult problem. The method used was to repeat the sound sequence at a rate of 1 sound per second for 50 cycles. After a few repetitions a "tune" becomes quite distinct (musically known as a Klangfarbenmelodie), even though all pitches are equal. When the brightnesses of the timbres are similar the tune becomes more difficult to perceive. It may be that the tune effect is related to the variation in the brightnesses of the sequence of timbres - a flatter more constant profile results in less variation making the tune more difficult to discern. This effect was used to check the brightness constancy for regularly spaced brightness contours, by listening for a "constant profile" Klangfarbenmelodie in each of them.

8.7 GamutExplorer

The GamutExplorer, shown in Figure 8-10, shows the realised Information-Sound Space in 3 viewing panels - a 3D wireframe, and 2D plan and elevation views. The 3D wireframe, on the top-left, can be rotated and tilted to gain an impression of the overall shape of the sound gamut. The plan view, on the top-right, is a horizontal slice through the space at a constant pitch. Each timbre is a coloured segment in which the radial dimension is brightness. If the timbre segment is not visible then it is unavailable at this pitch, as is the case for the flute and soprano sax in the Figure 8-10. The elevation view, on the bottom-right, is a vertical cross section through the space showing a pair of opposed "timbre-leaves" in complementary colours. The vertical dimension is pitch and the horizontal dimension is brightness. The brightness axis is bilateral, with the zero in the middle of the panel. The increase in brightness, to the left and right, is shown by an increase in colour saturation. This view also shows the unique pitch range and brightness profile of each timbre, and the region of pitch overlap between the timbres can be seen immediately. Different slices can be chosen by directly dragging a "cutting" line in each view, or by typing into a text widget. The sound at any point in the plan and elevation views can be heard by tapping it with the mouse. If the sound at that point is unavailable, or out-of-gamut, then a spoken error message such as “out-of-pitch range” is heard instead.

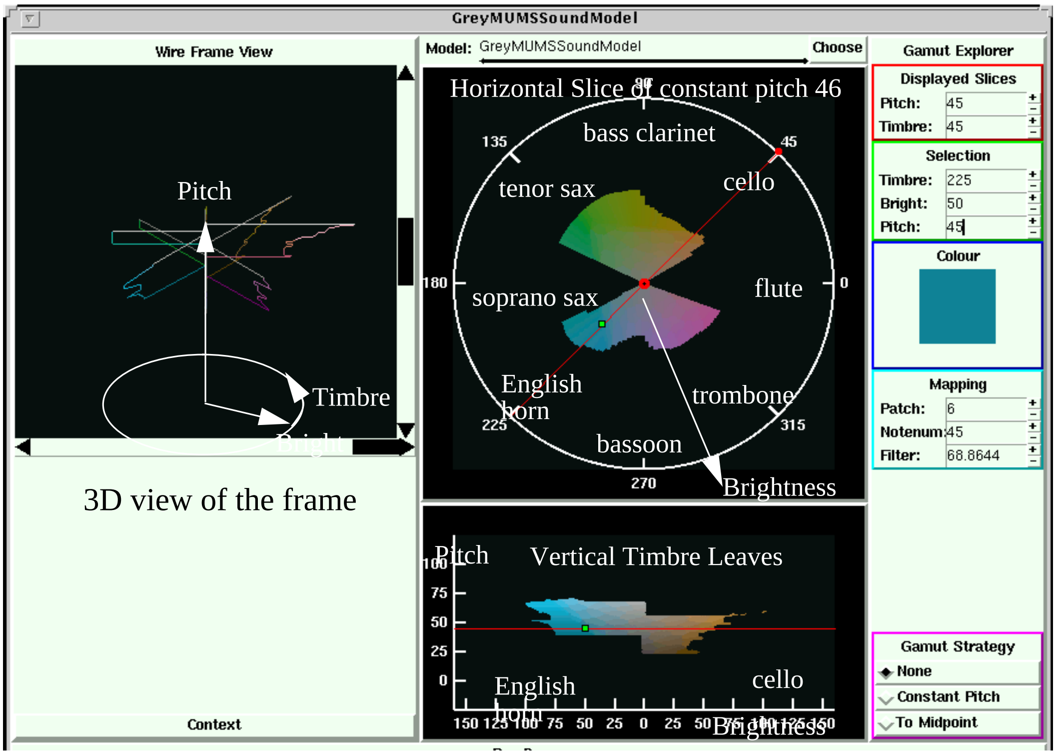


Figure 8-10: The GamutExplorer

8.8 Summary

This chapter describes the construction of a perceptually organised sound space to support information representations. The construction was carried out in four stages - the pedestal, the frame, the grating and the plasticine. The raw material for the construction was the MUMS palette of musical instrument samples, which is a reference palette for other researchers in the area of auditory perception. The pedestal consists of 8 equally spaced timbre steps organised in a circle by similarity. The frame is an arrangement of brightness profiles which defines the limits of dynamic range for variation in pitch and brightness at each timbre. The grating is made up from grids of equal differences in brightness and pitch for each timbre. These grids are joined on a central axis and radiate outward like segments of a mandarin. The plasticine is a continuous medium moulded to the grating and frame to model the behaviour of the overall space. The resulting sculpture has the property that there is a relationship between distance and the strength of perceptual grouping between points in the space. A vertical line is a pitch scale, and may be used to represent continuous data. A radial line is a scale of equal brightness increments for a timbre, and may also be used for continuous data. A circle of constant radius is a contour of constant brightness across the range of timbres which can be used to represent categorical data. These properties are a rich area for further experiment with data mappings and simultaneous and sequential presentations of the selections. The GamutExplorer is an interface that allows you to look at the GreyMUMS space as though it were a 3D coloured solid. You can hear points in the space by picking them from the sound solid with the mouse.

9 • Personify: computer-aided design for auditory display

Needs for future hardware and software include: integrated sonification/visualisation languages, tools for getting from an imagined sound to a realised sound, the integration of sonification tools into mass market software like spreadsheets or statistical analysis packages, and more and better tools for exploratory sonification [Scaletti C. (1994)].

Multi-modal interfaces are becoming increasingly important, and designing sounds for the human-computer interface is something more people are going to want to do. This chapter describes the Personify tool that can assist in the design of useful and effective sounds. The tool integrates a principled design approach with a direct manipulation interface. The guidance provided by the tool makes it quick and easy to use, and improves the likelihood of producing an effective display.

The chapter begins with an overview of interfaces for handling audio material found in musical tools. This is followed by an overview of tools that are specialised for auditory display design, with attention to the way these tools allow you to handle sounds. The Personify tool is then described in two parts - the Requirements part and the Representation part. The meshing of the parts as a system is demonstrated in a design scenario that involves resource monitoring by satellite data.

9.1 Tools for handling sounds

Sounds are intangible, insubstantial and hard to get hold of. But computer tools let us handle audio material like clay. We can bend it, fold it, stretch it and mould it. We can shred it, dice it, slice it and splice it. Some of the most common ways for handling sounds, shown in Figure 9-1, are a) music notation in composition tools b) caterpillar shadows in audio editors and c) landscapes in signal processing.

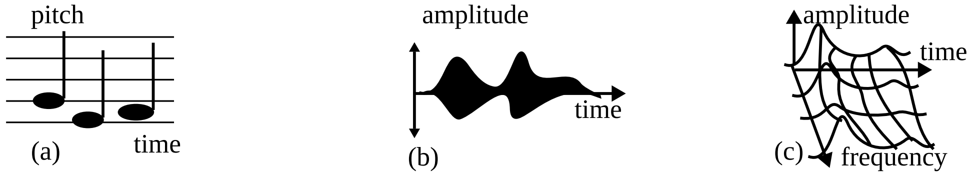


Figure 9-1: Common ways to handle sounds

The computer music community actively investigates new and innovative ways to handle sounds. An example of the types of interfaces that have been developed is the MAX music programming tool with which knobs, delay lines, lookup tables and other components can be wired together to build a music machine. MAX is a visual programming environment that is simple to learn and quick to produce results, so programming is less of a burden on musical activity. However MAX is an interface to devices rather than sounds, and the realisation of an imagined sound takes experience and know-how. The Intuitive Sound Editing Environment (ISEE) was developed to make the realisation of an imagined sound less complicated. The description of sounds is based on principal components of perceptual variation that have been consistently identified in research on timbre by Wessel, Grey, Plomp, and others. [Wessel D. (1985)], [Grey J.M. (1975)], [Plomp R. (1976)]. ISEE has 4 dimensions - Overtones related to harmonic content, Brightness related to spectral distribution, Articulation related to onset transient, and Envelope related to temporal envelope speed. A point in the 4D space can be chosen with a pair of 2D coordinate grids with axes labelled Overtones vs Brightness and Articulation vs Envelope. As the selection point is moved with the mouse the coordinates are mapped to whatever device is present to produce a timbre with the requested features.

Another direct interface is found in the Chroma3D tool for experimenting with pitch scales [Moreno E. (1994)]. The central element is a graphic wire-frame of Shephard's pitch helix that can be rotated and tilted to gain an appreciation of relations between the pitches. Selections of groups of pitches can be made in terms of the geometric properties of the model, and are shown as beads on the wire frame view. The selected points are automatically mapped to the audio device. Chroma3D provides a way to think and interact with pitch that can lead to new understanding and new scales.

Sounds become even more solid in virtual reality interfaces. An example is the Audio Browser for navigating a library of sound samples. This interface lets you float through families of instruments hanging in a 3D Virtual Reality (VR) space [Whitehead J.F. (1994)]. When an instrument is selected the corresponding sound is heard coming from the location where it is hanging. You can zoom in and out to get an overall impression of the library and its layout which can assist in searches and explorations. Another VR interface that has a similar idea is Horry’s tool for real time composition [Horry Y. (1994)]. In this interface there are coloured shapes hanging around that you can pick up and move with a data glove. The location of each shape affects aspects of the associated sound - such as frequency, amplitude or stereo position. Paths through the space are compositional operators that vary the sound in specific ways. A single gesture can alter several aspects of the sound at once.

9.2 Tools for designing auditory displays

Musical tools are very useful and widely used to design auditory displays. However there are many aspects of auditory display that are not catered for by these tools. Special tools for auditory display are being developed to address issues such as the faithful representation of data structure, the support for interaction in user interfaces, and the use of sounds in information processing applications. Some of the tools include programming libraries, reusable examples, synthesis algorithms, and rule-based displays. What sets them apart is the specialist expertise that these tools bring to a particular application area.

Application programming interfaces (APIs) are interfaces to software libraries and hardware drivers that can be linked into a computer program to generate sounds. These APIs have generally been developed for programming music and multimedia samples, but provide a platform for auditory display applications too. A popular area of application in auditory display turns the idea of a program for controlling sound on its head by purposefully using the sounds to assist in programming. Debugging computer software can be a tedious process that may involve the reconstruction of events from printouts or stepwise execution of the code. By placing auditory routines into the code you can get a higher level impression of the looping, branching and pathways through the code by listening to it as it executes. An API that was developed specifically for this purpose is the Listen Specification Language (LSL) which has routines that generate MIDI events from your program - for example a piano note might trigger every time a particular statement is reached [Mather A.P. Boradman D.B. and Khandelwal (1994)]. Another example is the Auditory Domain Specification Language (ADSL) in which programming symbols are associated with auditory symbols - so for example all *for* loops may trigger the piano note [Bock D.S. (1994)].

A different approach is found in the API for ENO, which is a tool that supports the use of sounds in computer interfaces [Beaudouin-Lafon M. and Gaver W.W. (1994)]. The API in ENO is based on the auditory icon method that lets you design sounds in terms of everyday objects and events that cause them [Gaver W.W. (1994)]. The server automatically handles the synthesis and playback of impacts, scraping events and machine sounds that are generated by interactions between objects in the interface. Another server-based tool is the Centralised Audio Presentation System (CAPS) which has an API that supports the earcon method for designing interface sounds from short musical motifs [Papp A.L. and Blattner M.M. (1994)]. CAPS has the extra capability that you can prioritise earcons and rely on the server to automatically maximise the perception of the most important information. The interference and masking that can occur when two or more sounds occur at once are dealt with by a rule-base that dynamically adjusts the display according to perceptual guidelines. The guidelines are observations that were made in an empirical investigation into the effectiveness of earcons [Brewster S.A. Wright P.C. and Edwards A.D.N. (1994)].

Programming is usually a text-based activity, so text-based APIs fit in with the way a programmer normally works. However visual programming environments are an increasingly popular way to program, because they can be easier to learn and quicker to produce results. Sonnet is program auralisation tool that has a visual programming environment founded upon the MAX music environment. With Sonnet you can build an auralisation algorithm by wiring components, such as a Pitch box, NoteOn box, and Counter, together and connect them to the code [Jameson D.H. (1994)]. Another tool based on MAX has been developed for data sonification. The Audification Programming Environment (APE) supports Kramer's "parameter nesting" method for audifying multivariate data. You can connect data dimensions to the auditory parameters in an environment that supports rapid iteration and prototyping of the design. A different kind of visual programming is found in Kyma, which is a tool for both music and auditory display. Kyma has an object-oriented language which encourages the reuse of components through inheritance. Sonification expertise is encapsulated in objects that support Scaletti's "parameter mapping" method [Scaletti C. (1994)]. The sonification objects are examples that can be reused and adapted in new designs.

The tools that have been developed for designing auditory displays demonstrate that the specialised focus and knowledge contained in them can improve the likelihood of producing a useful and effective display. However something that is missing from these tools is the intuitive and direct manipulation style of the interfaces that have been developed in music. These observations are summarised in Table 9-1.

Interface style	Music Composition	Auditory Display Design
Text	APIs for Audio and MIDI	LSL, ADSL, ENO, CAPS
2D Graphic	Notation, Caterpillars, MAX	Sonnet, APE, Kyma
3D Graphic	Landscapes, Chroma3D	Personify
VR	Horry’s VR MIDI	

Table 9-1: Audio tools

9.3 Introducing Personify

Personify is a tool that integrates the TaDa approach to auditory display design with a direct manipulation interface to sounds. This tool fits in the empty cell at the intersection of 3D Graphic and Auditory Display Design in Table 9-1. The tool is described in 2 parts - the Requirements part and the Representation part, as shown in Figure 9-2.

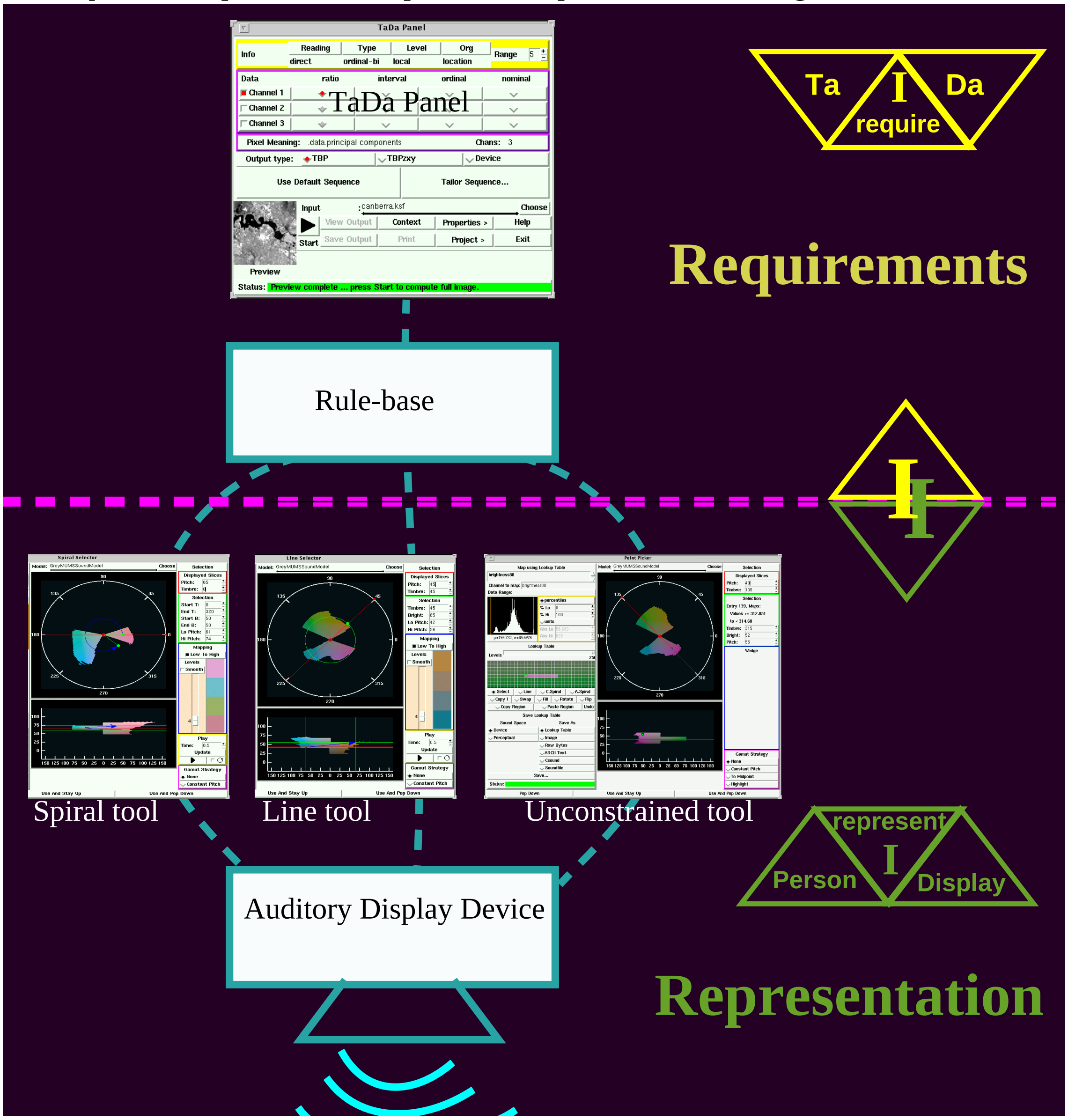


Figure 9-2: Introducing Personify

The Requirements part has 2 main components

- the TaDa Panel for capturing information requirements
- a rule-base that configures a default representation from the requirements

The Representation part also has 2 main components

- interactive tools for tailoring the representation to a device
- an auditory display device

Personify is an extension of a tool for choosing colour mappings for scientific visualisation and satellite imagery applications [Robertson P.K. Hutchins M. Stevenson D. Barrass S. Gunn C. and Smith D. (1994)]. The extensions include the TaDa Panel, the rule-base, the auditory display device, and the audio feedback. The tool is implemented in the tk/tcl language [Ousterhout J.K. (1994)], and C++, on a Sun^{TM} workstation

9.3.1 The TaDa panel

The TaDa panel is the interface for specifying a representation in Personify. The TaDa panel has 3 main windows - Info, Data and Compute shown in Figure 9-3. The Info window is where you enter the information requirements of the problem. The Data window is where you can browse and load data sets, and read a description of the loaded data-set. The Compute window maps the loaded data-set through the representational mapping into a file of sound specifications.

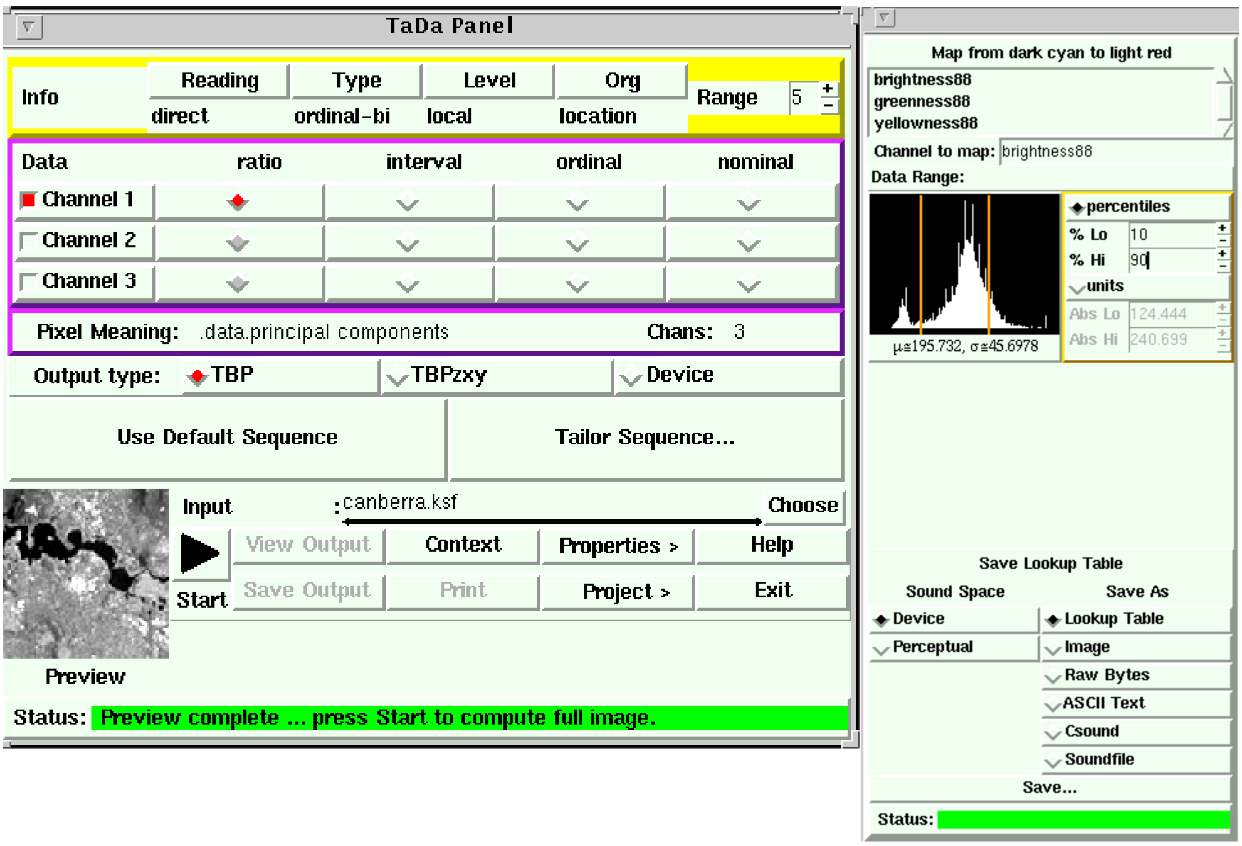


Figure 9-3: The TaDa panel

The Info window has pull down menu buttons for each of the TaDa information descriptors - Reading, Type, Level and Organisation, and a numeric entry for the Range field.

Under each button is a menu of options as shown in Table 9-2..

Reading	Type	Level	Organisation
direct	boolean	local	category
conventional	nominal	intermediate	time
	ordinal	global	location
	ordinal-with-zero		alphabet
	ordinal-bilateral		continuum
	interval		
	ratio		

Table 9-2: Info menu options

The Data window has a description of the data set which is currently loaded for processing. You choose a data set with a pop-up file browser, and the selected set is shown in a 100x100 postage stamp view. The data is expected to consist of 2D arrays of floating point numbers, called channels, that are stacked like a pile of paper. Tags in the data-file describe the number of channels and size of the arrays. As each channel is read a histogram is computed, as shown in the Histogram Panel in Figure 9-3. The histogram can be used to optimise the mapping of the data into the representation, or to select subsets of the data for mapping. Something that is missing from the data-file is a description of the data type in terms of {nominal, ordinal, interval or ratio}. Since the satellite data is always floating point there is an assumption that only ratio data is ever involved. However this may not always be the case - for example classification data is often encoded numerically by floating point numbers which act as surrogate categories. This type of data can be observed in the histogram as discrete spikes. Mistaking this data for continuous data can lead to misleading and hard-to-interpret representations. The Data window has a set of Type descriptors where the ratio default can be changed if you know more about the data. It may be possible to set this field automatically by analysing the histogram, although telling ordinal from nominal, and interval from ratio may be impossible.

The Compute window has controls for mapping from the data-file to a file of sound specifications. The Start button begins computing the mapping for all the data, and the Stop button can stop the process at any time. A status bar progressively shows the amount of data that has been processed. There is a default representation, or you can tailor the representation to maximise the range on the display device. The default representation is configured from the information and data descriptions by rules that are described in the next section. The tools for tailoring the design are described after that.

9.4 The rule-base

The rule-base takes the requirements captured by the TaDa Panel and configures a representation to meet them. The rules adhere to the Hearsay principles developed in Chapter 6. The representation is selected from among the templates developed in Chapter 7. The rule-base is described in detail in the following subsections.

Info Reading

The Info reading can be {direct or conventional}. If the Reading is {direct} then only the constrained selection tools are available, and the constraints are switched on. These tools guarantee that the display adheres to the Hearsay principles. Setting the Reading to {conventional} activates the Unconstrained tool that lets you choose a set of sounds in any way you please, perhaps for experiments, music or whatever.

If InfoReading is direct then enable the constrained tools
If InfoReading is conventional then enable the Unconstrained tool

Info Type

The Info-Type field is used to look-up a mapping that has the required perceptual properties of difference, order, metric and zero. The mapping schemes are geometric paths through Information-Sound Space (ISS). The conjunction of the path and the organisation of the ISS forms a sequence with specialised information properties. There are two basic paths - a spiral and a line, that can be configured to form coils, circles, arcs, sloped lines, vertical lines, and radial lines. The configuration is set by the InfoType according to the following rules

If InfoType is boolean then configure the Line tool for opposite timbres.
If InfoType is nominal then configure the Spiral tool as a circle
If InfoType is ordinal then configure the Spiral tool as a coil
If InfoType is ordinal-with-zero then configure the Spiral tool as a spiral
If InfoType is ordinal-bilateral then configure the Line tool as a bilateral line
If InfoType is interval then configure the Line tool as a vertical line
If InfoType is ratio then configure the Line tool as a radial line

Info Level

The Info Level can be {local, intermediate or global}. This setting changes the duration of auditory elements in a mapped file. When the level is local the sounds in the mapped file are all 1 second long, which is enough time to analyse and judge aspects such as timbre or brightness. Setting the level to global makes the sounds 10 times shorter so that individual elements cannot be easily analysed, but groups can be heard as a granular texture. The influence of timbre, brightness and pitch on auditory grouping cause the texture to contain streams that correspond with clusters of data with similar values. You may immediately answer questions like “are there any outliers in this data?” by listening for elements that pop-out from the main texture. Between global and local is the intermediate setting where a wider range of data values stream together, and fewer streams are heard.

If InfoLevel is local then set duration to 1 second
If InfoLevel is intermediate then set duration to 0.2 second
If InfoLevel is global then set duration to 0.1 second

Info Organisation

The Info Organisation can be {category, time, location, alphabet, continuum}. For example planning a driving route can be done with a map, which is organised by location. However once you start driving and your eyes are on the road it is better to have a navigator tell you when to turn and which roads to take. You can get where you are going either by remembering the roads from a visual map, or hearing directions as you need them. The task and information is the same but the organisation is different.

Ideally the organisation of the information in the display could be changed interactively, as required. Personify operates on files, and display reorganisation is left to the display interface. However Personify can reorganise the mapped file so that the most important dimension is the primary index for accessing elements. The reorganisation from the input data organisation to a different representational organisation occurs as the data is mapped to the output file. Each dimension of the output file has an identification tag for decoding its organisation.

If InfoOrg is category then tag the first dimension as category
If InfoOrg is time then tag the first dimension as time
If InfoOrg is location then tag the first 1,2, or 3 dimensions as spatial x,y, or z
If InfoOrg is alphabet then tag the first dimension as category and organise the categories according to a convention.
If InfoOrg is continuum then tag the first dimension by the continuous property

Info Range

The Info Range is the number of answers to the scenario question. For example a question such as “has the vegetation increased in this region” may have the answers {decreased a lot, decreased, no change, increased, increased a lot} which has an InfoRange of 5. The InfoRange sets the number of discriminable auditory elements in the representation. The data is automatically mapped into representation elements, by segmenting the data into InfoRange bins, and mapping each bin to a representational element.

Segment the representation path into InfoRange equally spaced elements
Segment the data into InfoRange bins
Map each bin to the corresponding representation element

9.5 Tailoring tools

The tailoring tools let you handle sounds as though they were a tangible coloured material of some kind. The coloured solid shows the range of sounds that can be produced by a display device, and is called the display gamut. The gamut can be rotated and sliced to get a feel for its 3D shape. Categorical timbre differences are shown by categorical differences in colour hue, so a flute timbre might be red and a saxophone timbre might be green. Ratio differences in sound brightness are shown by ratio differences in colour saturation. Metric differences in pitch are shown by metric differences in lightness. The tailoring tools have two views that are horizontal and a vertical slices through the gamut. Paths can be drawn on the slices to specify sets of sounds to be produced by the device. The paths should not stray outside the gamut because this means that the display cannot produce that sound. The paths can be adjusted with the mouse, to make sure they fit the display. The tailoring tools behave a lot like familiar drawing tools, but the behaviour is constrained to adhere to the Hearsay principles. Three tools have been implemented - a Spiral tool, a Line tool and an Unconstrained tool, which are described in the following sub-sections.

9.5.1 Spiral tool

The Spiral tool, shown in Figure 9-4, has 3 windows where you can design a 3D spiral path through the display gamut. These are the horizontal view at the top of the panel, the vertical view below it, and the collection of control widgets to the right of the slice views. The horizontal view shows a coloured cross-section through the device gamut, inside a circle marked with 45 degree segments that correspond with different timbre categories. The sounds of all the points in this slice have the same pitch, but vary in timbre and brightness. You can listen to any point in the slice by tapping it with the mouse. If you tap a region outside the gamut you will get a spoken message that says “out-of-gamut”. The gamut slice is overlaid by a blue arc that is the selection spiral. The spiral can be directly manipulated by green handles on the ends to change the angular extent, and hence the range of timbres that are selected. If the endpoints are moved to different radial positions the path is a flat spiral that selects timbres with an ordered change in brightness. The red line that passes through the centre of the window can be rotated to choose which slice is shown in the vertical view in the other window.

The vertical view shows a pair of complementary "timbre leaves" in complementary colours. The vertical dimension is pitch and the horizontal is bilateral brightness that radiates to the left and right from the central axis. This view shows how the maximum brightness of each timbre changes with pitch, and also the unique pitch limits of each musical instrument that the timbres have been sampled from. The spiral path is shown as a blue curve overlaid on the timbre leaves. In this view there are two green sliding bars that change the pitch height of each endpoint to spread the spiral in the vertical direction. The spread spiral is a coil that selects timbres at different pitches. There is a horizontal red line that can be moved to change the slice that appears in the horizontal view above it.

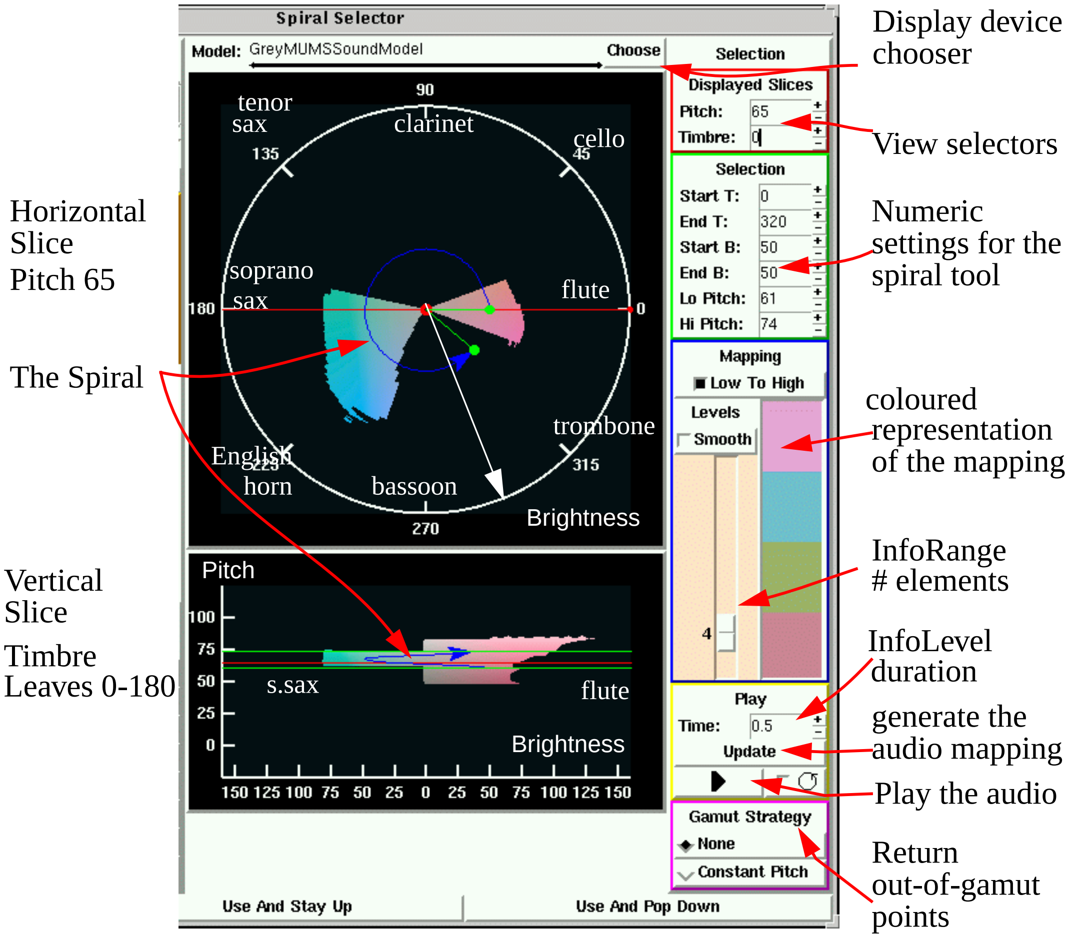


Figure 9-4: Spiral tool

The control window has 6 subsections - Device, Slice, Selection, Mapping, Sound, and Strategy. The Device subsection has a file browser that you can pop-up to choose a display device. Different devices have different characteristics that are reflected by the shape of the gamut as seen in the slice views. The gamut of each device constrains what auditory relations may be realised on that device. The Slice section has fields where you can numerically set the coordinates of the horizontal and vertical slices. In the Selection section you can also numerically specify the coordinates that shape the selection path, such as the coordinates of the endpoints of the spiral. The Mapping section has a slider where you can set the number of elements that the path is divided into. This slider is automatically set from the TaDa Panel by the InfoRange rule. The selected set of elements is shown in a coloured legend to the right of the slider. Individual elements from this legend can be heard by tapping them with the mouse. Several elements can be highlighted with the middle mouse for simultaneous playback. This is so you can listen to how well 2 or more elements separate when they are heard at the same time. Below the Mapping is the Sound section. At the top of this section is a duration entry which affects how long each sound will be when it is played. The duration is automatically set from the TaDa Panel by the InfoLevel rule. Beneath duration is the Update button that causes the selected set of sounds to be generated by the display device. The results can be heard by pressing the play button, and played over and over by pressing the loop button next to it. Finally the Strategy section affects what happens to parts of the representation path that stray outside the range of the device. The {None} option causes “out-of-gamut” to be spoken for these points. The {Pitch} option maps out-of-gamut points back into gamut along a radial path of constant pitch.

9.5.2 The Line tool

The Line tool, shown in Figure 9-5, differs from the Spiral tool only in the path that may be chosen. The line is shown in blue in the vertical view. In this tool the path is a straight line with either its mid-point or one end tied to the central vertical axis.

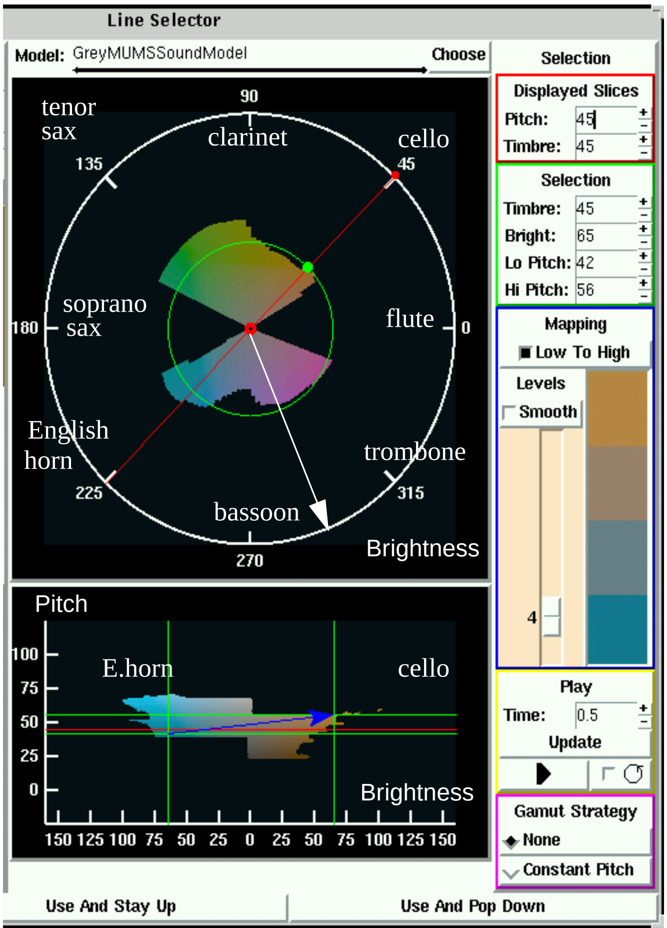


Figure 9-5: Line tool

When the mid-point is tied it is a bilateral line that guarantees that a central zero will be heard through brightness and timbre variation. When the end-point is tied it is a radial line of constant timbre with a brightness zero at one end. The length of the line is adjusted by the radius of the green circle in the horizontal view, or the vertical green bars in the vertical view. You can adjust the slope of the line with the two horizontal green handle-bars in the vertical view. This causes the selection to have a pitch order.

9.5.3 Unconstrained tool

The Unconstrained tool allows you to choose an array of up to 256 sounds in any way you wish. The sounds can be selected from the space individually, or using paths between pairs of points. You may highlight regions of a mapping by a change in pitch or timbre that really stands out from the other parts of the set, or experiment with new representations, or use the interface to select sounds for music, or other purposes.

The interface, shown in Figure 9-6, has the horizontal and vertical slice views just like the other tools. There is an extra window on the left of these panels that provides the space for the extended set of 256 sounds that are shown as cells in an array. Beneath the array are a variety of buttons for selecting different paths between the points in the array, for example a line, a spiral, a flat region and so on. Some editing operators let you flip, rotate, cut and paste parts of the array. Below the editing panel are some buttons for setting the format for writing the array to disc. There are 6 options - a lookup table of coordinates, an image of coloured pixels, raw bytes, ascii text, Csound score specifications, and an audio sequence. This range of formats is for exporting the selected array of sounds into other tools.

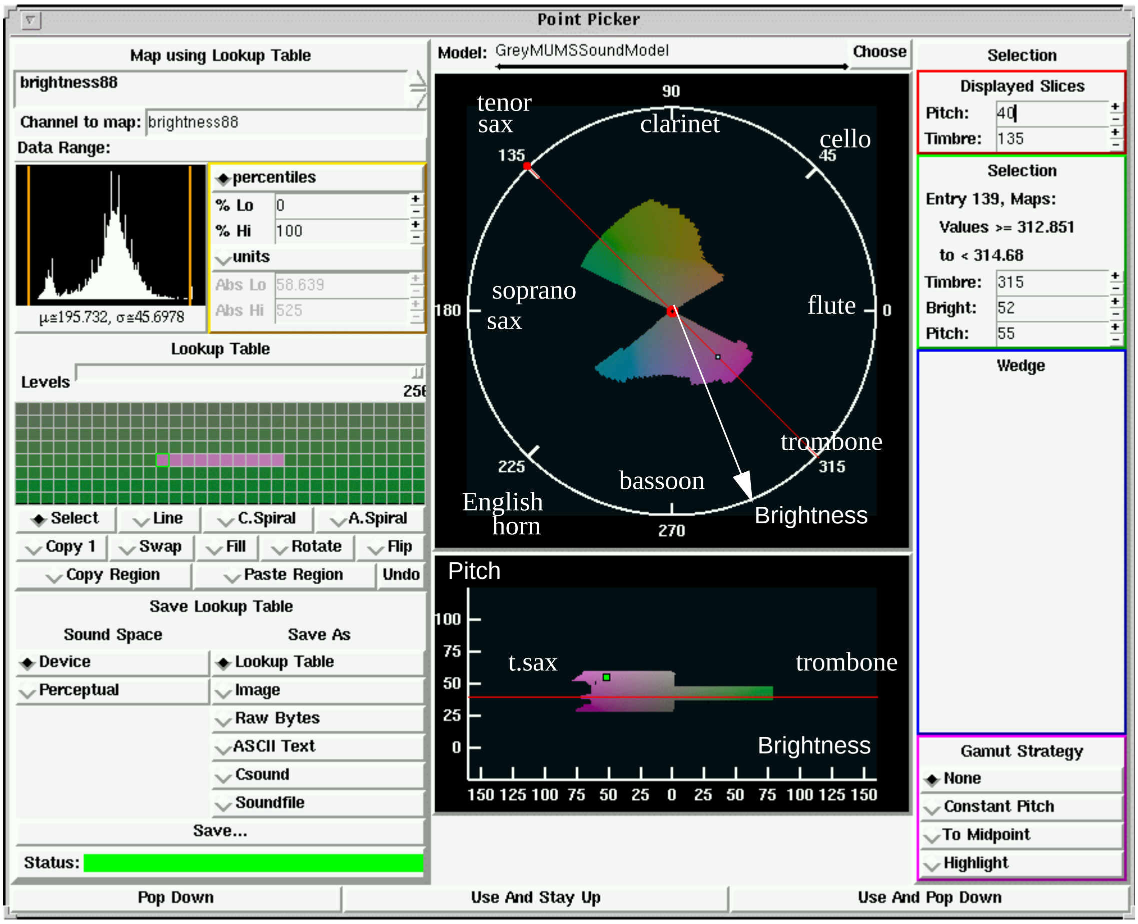


Figure 9-6: Unconstrained tool

9.6 Display device

The auditory display device, shown in Figure 9-7, is a compound of software and hardware that is capable of producing auditory relations specified by device-independent ISS coordinates. The device is glued together by a Display object which coordinates and communicates with the components through Unix pipes. There are 4 components - a mapping from ISS coordinates to device parameters, a gamut that characterises the range of sounds on a particular display, a sound engine, and an audio output system. The Display object passes the ISS coordinates that specify a set of sounds through the mapping to device parameters, sends these parameters to the sound engine, and then plays the result on the audio output system. The mapping and gamut were described in Chapter 8. The sound engine is the Csound audio synthesis and processing software [Vercoe B. (1991)]. The audio output system is the Sun 16 bit, 44.8 kHz stereo dbri audio device.

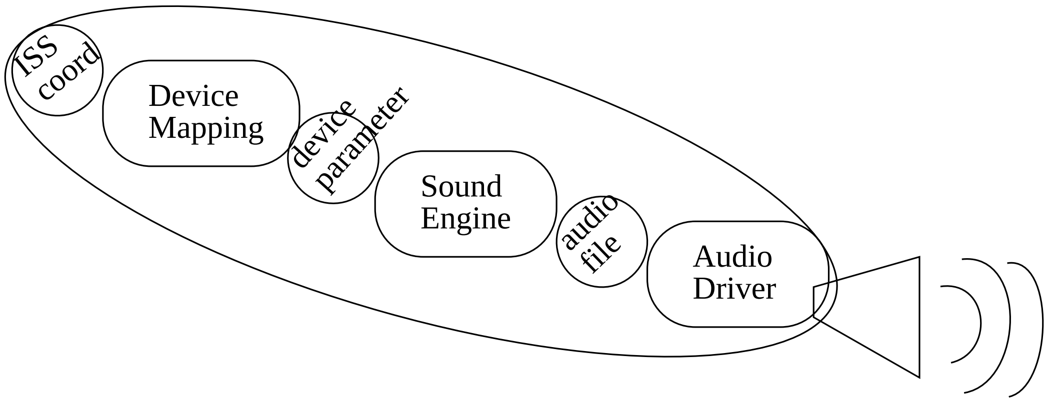


Figure 9-7: The auditory display device

9.7 Personify in action

The parts of Personify are demonstrated working together as a system in a design scenario. The scenario is drawn from the domain of resource monitoring by remote-sensed satellite data ...

Remote sensed satellite data can provide information for urban planning, resource management and environmental impact studies. The Landsat image of Canberra, in Figure 9-8, shows change in landcover by colour variation, where grey=no change, red= loss of vegetation, blue= increase in vegetation, dark=landcover loss, light=landcover increase [Graetz D. Fisher R. and Wilson M. (1992)]. New suburbs, where houses replace grassland on the perimeter of the city, are seen as light red areas, whilst pastures transformed by forest plantations appear light green. The large grey blob in the centre is Lake Burley-Griffin which provides a reference for “no change”. These images can help an analyst answer questions such as “where did the greatest change in landcover occur?”, “is the change due to vegetation?”, “has there been an increase in the landcover overall?” and so on. These are questions at the overall and intermediate levels. However it is difficult to answer questions about a single point, particularly in a highly variegated regions, because high spatial frequencies in hue can exceed the resolving capability of the eye [Rogowitz B.E. and Treinish L.A. (1993b)]. Can we design a display that allows the user to hear information about the landcover that is difficult to see in the image? Since it is the local information about vegetation that is most difficult to see we may improve the display by providing this information with sounds.

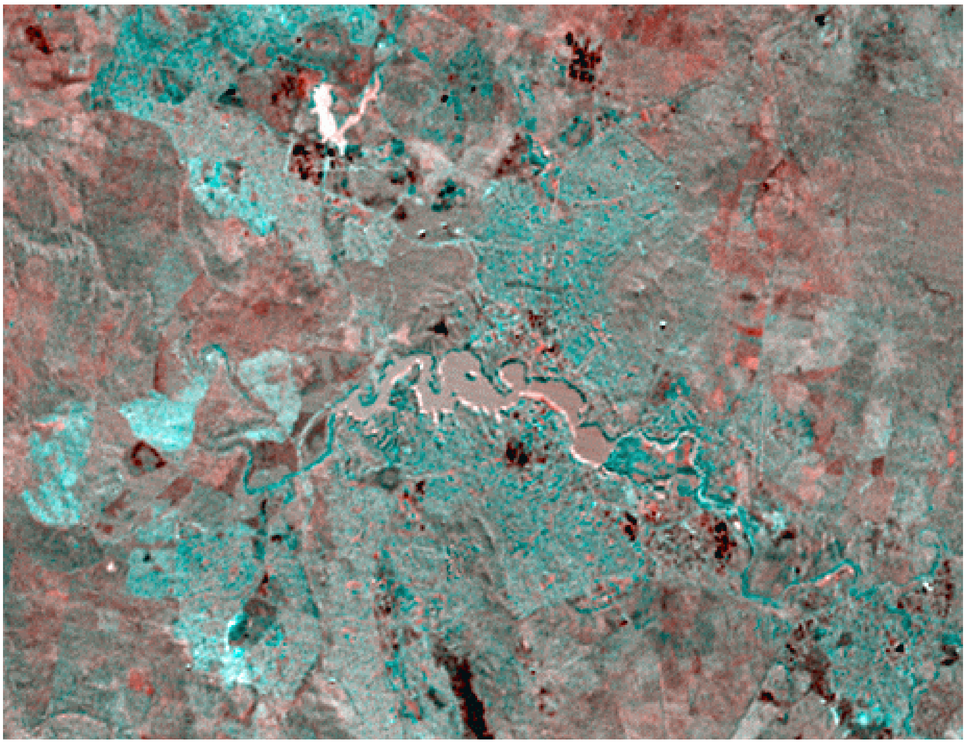


Figure 9-8: Landcover difference for Canberra between 1972 and 1988

The design process begins with a TaDa analysis of information requirements, shown in Figure 9-9. The scenario is recast as a Question “how much change in vegetation at this location?” with answers {large decrease, decrease, no change, increase, large increase}. The Subject of the question is the change in vegetation.

Scenario

STORY

Title	LandCover
Storyteller	S.K.

Satellite images show regions well but the colour of a single pixel or very small area is hard to see, especially in highly variegated regions. What we would like to do is design a display that allows the user to hear local information about the vegetation that is difficult to obtain visually from the image.

KEYS

Question	how much change in vegetation at this location ?
Answers	--, - 0, +, ++
Subject	change in vegetation
Sounds	???

TaDa Analysis

TASK

Generic	how much here ?
Purpose	judge
Mode	interactive
Type	discrete/procedure
Style	exploration

INFO

Reading	direct
Type	ordinal-bilateral
Level	local
Organisation	location
Range	5

DATA

Type	ratio
Range	??
Organisation	location

Figure 9-9: .TaDa for LandCover

The Question key is the bridge from the scenario to the Task analysis, which has fields {Generic, Purpose, Mode, Type, and Style}. The Generic question is obtained by removing the Subject from the Question to obtain "how much at this location?". The purpose is to judge how much the change in landcover is due to plant growth. The mode is interactive because it involves a shift in attention between different parts of the display as new questions arise. The task type is discrete/procedural because a new query can only be made after the previous one has been answered. Although the demonstration only involves a single data-set the aim is to design a display that can answer the same questions for other data-sets as well. This makes the Style exploratory because we don't know what information will appear when other data-sets are loaded.

The Answers key is analysed in the Info section, which has the fields {Reading, Type, Level, Organisation and Range}. The Reading is direct which indicates that the listener should be able to hear an answer quickly, correctly, and confidently without much training. The answers are an ordered sequence with "no change" in the middle and increase and decrease either side that is typically ordinal-bilateral. The Level is local because the answer is about individual points. The Organisation of the information is by location in a map. The Range is 5 because this is the number of answers that must be clearly discriminated.

The Subject key is analysed in the Data section. Vegetation is measured by electromagnetic reflectance of the earths surface in a particular band (in the infra-red). The measurements are ratio data to start with, because there is a zero at no reflectance. In the example the measurements for 1972 have been subtracted from 1988 to get the difference which is still ratio data, but now has a signed component. At this stage we do not know the Range of the data. The Organisation is by location - each data element is related to its neighbours by location in a 2D array with spatial coordinates.

The results of the TaDa analysis can be entered into the TaDa Panel in Personify, as shown in Figure 9-10. Setting the Reading to {direct} enables the constrained tools. Setting the Type to {ordinal-bi} configures the Line tool for a sloped line with a central zero. Setting the Level to {local} configures the duration of the mapped sounds to be 1 second so you can listen to individual elements. Setting the Organisation to {location} causes the mapped elements to be organised by their spatial location. The satellite image, called D01_Canberra.ksf is loaded with the file browser, and shown as a small postage stamp image. As the file is loaded a histogram of the data is computed. The Data section indicates 3 channels of principal components data. The channels in the histogram window are labelled brightness, ch_bright (change in brightness) and ch_green (change in greenness). It is the 3rd channel that contains the vegetation landcover information that we are interested in. Channel 3 has been selected and set to ratio in the Data window, causing the histogram for the channel to be shown in the histogram view where we see that the Data Range is -62.5905 to 131.0.

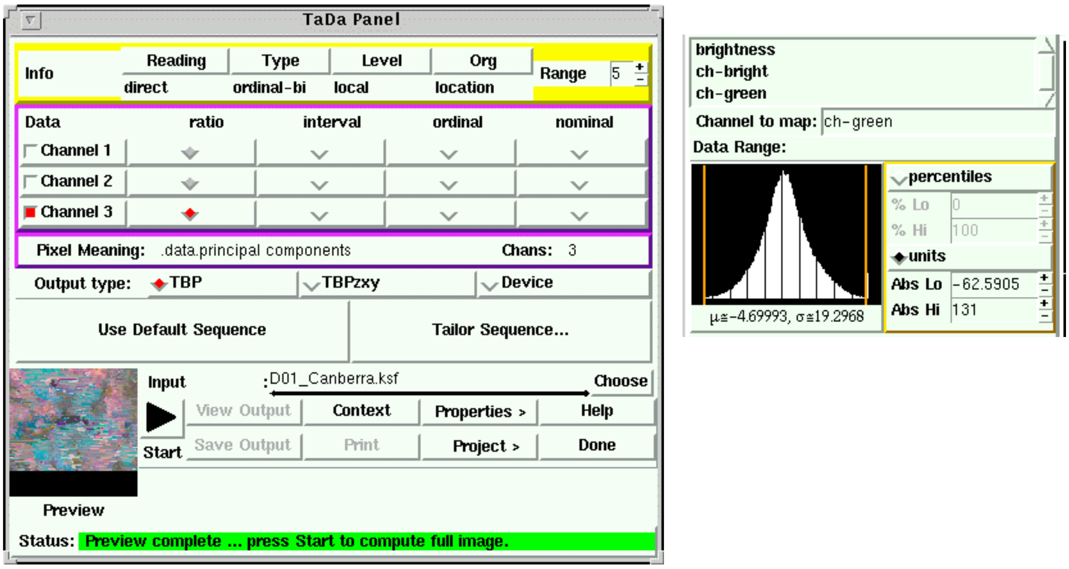


Figure 9-10: The TaDa Panel for the LandCover design

Once the TaDa Panel has been filled-in the design process moves from the Requirements to the Representation stage. This stage involves consideration of how people can hear the required information in sounds, and what sounds the display device can produce. This is where the gamut view and graphic selection tools can help. The best tool for the job is automatically chosen by the rule-base, and is activated by pressing the Tailor Sequence button below the Data window in the TaDa Panel. In this instance the Line tool pops-up configured for a bilateral-ordinal sequence. The default display device is the GreyMUMSSoundModel. Another auditory display could be chosen, if there was one, but for now GreyMUMS is all there is. Initially the tool shows timbre leaf 180-0 which is the soprano-sax and flute. The pitch range of the soprano sax is limited to about 15 semitones, so I chose to move to the 90-270 leaf which has a wider pitch range that maximises the difference between {large decrease} and {large increase} answers. The snapshot in Figure 9-11 shows the pitch going from a low of 31 to a high of 52 - a range of 21 semitones. The path is automatically divided into 5 equally spaced pitches, one for each answer. The difference of 5.2 semitones between each is easy to discriminate in comparisons. If you look at the vertical view in the Figure you can see that the upper endpoint of the path extends beyond the gamut of the display - which means that this sound cannot be produced on this display. The solution is to grab the vertical green handle and pull the endpoint into gamut - though the brightness of the lower extrema is also reduced by the bilateral constraint. Pitch does not have a natural zero, so the zero that corresponds with no change is provided by brightness instead. The change in timbre between the decrease and increase segregates these categories, so that you can discriminate between them.

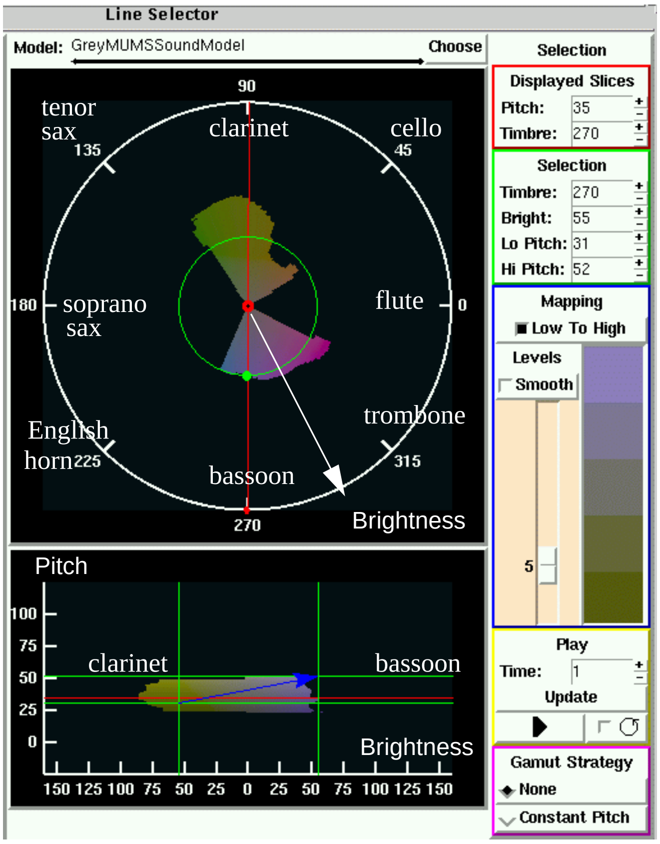


Figure 9-11: Line tool for the LandCover design

The final set of sounds and the answers they provide are shown in Table 9-3. There are many other palettes of sounds that could have been chosen like this. However, they would all share the same information properties in terms of difference, order, metric and zeros, because the tool is constrained to align the relations between the sounds with the relations between the information elements.

answer	sound palette
{large decrease}	clarinet, pitch 31, brightness 40
{decrease}	clarinet, pitch 36, brightness 20
{no change}	clarinet, pitch 41, brightness 0
{increase}	bassoon, pitch 46, brightness 20
{large increase}	bassoon, pitch 51, brightness 40

Table 9-3: LandCover palette

9.7.1 The ImageListener

Once the set of auditory information elements has been chosen the data loaded in the TaDa Panel can be mapped into sounds by pressing the Start button. The output is saved as a 2D array of sound specifications or “audio pixels”. These specifications can be sent to be realised on the auditory display device. You could load the mapped file into an image browser to look at it, but since the pixels are sounds rather than colours that doesn’t make much sense. The ImageListener, shown in Figure 9-12, is an integrated audiovisual display that lets you see the image and hear the sounds at the same time. The tool works by loading an image file and displaying it in a window, and also loading the file of audio pixels. Because both files are organised by location you can tap a point in the image and use this coordinate to retrieve a pixel from the sound file. The audio pixel is passed to the Display object which generates the sound.

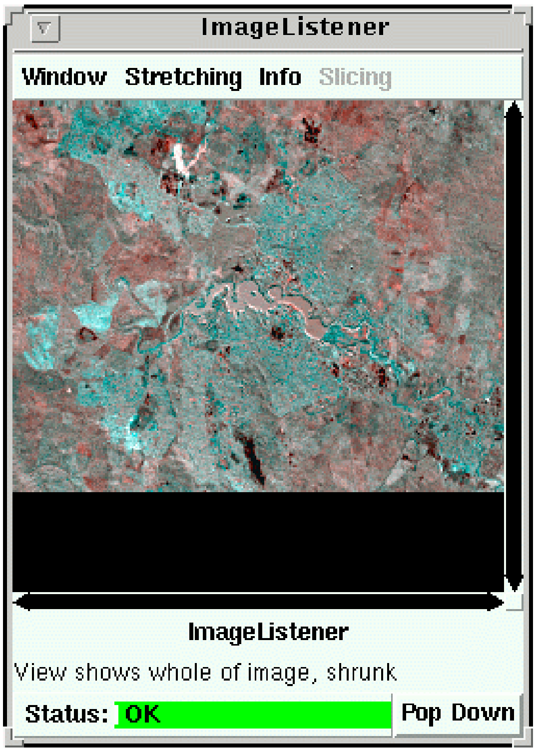


Figure 9-12: The ImageListener

The image provides overall and intermediate information, while local information is obtained by tapping on points of interest with the mouse. You can make interactive local queries and comparisons between points that are hard to see. I found that in regions of apparent uniformity, such as the green patch just north of Lake Burley-Griffin, there are scattered anomalies that are much easier to hear than see. I found the auditory answers were clear and easy to make in comparisons. The order of the sounds is immediately heard from the pitches. The relative brightness differences were not so easy to hear, and the perception of order was dominated by pitch. For example a question like “which is the more extreme” required careful listening to judge the relative brightnesses when one point was {increase} and the other was {large decrease}. Although not as immediate as expected I was able to attend to the brightness of a sound better after several minutes of experience with the display.

Although the representation works for comparisons between elements, the absolute answers were not as direct. I found it quite difficult to identify the pitches, even though they differed by 5 semitones which is a large amount in relative terms. I could tell that a pitch was in the higher or lower part of the range, but could not say with confidence that it was second from the top or top. The identification of which side of the zero an element was from was made much easier by the two timbres. It only took a few trials before I could confidently identify the sax and bassoon, and associate each with decrease and increase. The {no change} answer could be listened for by a characteristic dullness, but it was not an intuitive “zero”.

It was hard to ear the ordered difference been timbres of different brightness, and the brightness zero was not easy to hear either. These observations lead me to conclude that perhaps one of the other candidates for the zero axis, such as loudness, or duration, might be more suitable for representing local information. Brightness may be better for higher level displays that involve streaming. On the other hand I did find that within a few minutes of using the display I could always answer even the absolute questions correctly from the sounds.

9.8 Summary

Sounds are becoming common in human-computer interfaces, and designing these sounds is something more people are going to want to do. Musical tools can be helpful but they do not address issues of data representation that are important in auditory display. Tools specifically for auditory display are developing, but the programming style approach that is current is not as direct as some interfaces for composing music. Personify is a tool that integrates a rule-based method for auditory display design with a direct manipulation interface. The tool has 2 main parts - the Requirements part and the Representation part. The Requirements part is made up of the TaDa Panel that captures the information requirements, and the rule-base that uses the requirements to configure a default representation. The Representation part is made up of Tailoring tools for fitting the representation to the display, and the display device for realising the sounds. Personify was demonstrated in a scenario about resource monitoring by satellite data. The resulting audiovisual display shows higher level information by a colour image, and provides local information that is hard to see, in sound. The required information was easy to hear and understand from the display, indicating that Personify can provide assistance for designing an effective display.

9.9 Limitations and further work

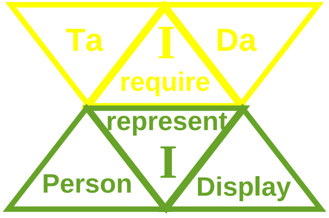
The informal observations about the effectiveness of the display could be used to specialise an Information-Sound Space to provide absolute answers to local questions. After improvements have been made and tested a more formal validation of the display could be made.

The image shows both the landcover and the vegetation at the same time, by ordered variations in two separable dimensions of colour - lightness and saturation. However this bivariate representation does not adhere to the Hearsay principles. Two zeros are needed - one for {no change} in landcover and one for {no change} in vegetation. However the zero in the lightness series is a mid-grey colour which is not directly perceptible as a zero. The lake in the middle of Canberra is the reference that is needed to judge whether the landcover has increased or decreased in other regions of the display. Without the lake it would only be possible to judge extreme cases. The problem is not too bad because the information in the image comes from relations between elements. However the problem is very important at the local level, because only individual elements are involved in a judgement, and so no comparison can be made with a reference (if it is to be a direct display). The same problem occurs with the Information-Sound Space which has its characteristics modelled on the colour space. There is only one natural zero, so bivariate ratio information cannot be directly represented at the local level. This problem raises the issue of restructuring the Information-Perception Space to provide capabilities that cannot be realised by colour but may be possible in other perceptual representations. For example the IPS could be configured so the vertical dimension has a central zero. Another possibility is to organise the display to present two pieces of information at the same location. In the auditory display this may be done in time, spectrum or space. We could represent both landcover and vegetation with the same auditory set by playing pairs in unison or series.

The ImageListener is a 2D arrangement of mapped data values presented as coloured pixels, which preserves the spatial relationships in the data as spatial dimensions of the image. The image provides the context for selection of individual points which may be heard by tapping them with the left mouse button. Currently the way to compare two points is by tapping them one after the other. This is an organisation in time. An organisation of answers by spectrum could be enabled by allowing the simultaneous sounding of 2 points at once. One way to do this could be Sonic Pins which may be placed into the image, much like map pins. Local elements with pins of the same colour would repeat in synchrony, allowing intermediate level comparison and correlation type queries.
Another way to provide higher levels of auditory information could be to take the technique from Chapter 6 of changing the duration of the sounds. A region of elements around the mouse cursor could all be tapped at the same time and heard as a group by a granular synthesis. The timbre, brightness and pitch of the sounds are strong factors in perceptual grouping, so that regions comprised of different elements may be perceptually discriminable.

The ImageListener is a display where the image provides the global information and the sounds provide local information. However this is not the only organisation that is possible. In Virtual Reality applications the sounds may provide global information about the surroundings in an immersive interface, and the visuals may provide for local detail. Display interfaces for information with other organisations also need to be built - for example for information organised in time, or in concurrent series. Again we can look to musical interfaces where there are many tools for organising concurrent channels of information in time.

Interactive queries need to be answered immediately. The delay between tapping a location and hearing the sound is about two seconds which is noticeably sluggish. A faster Display could be implemented for the ImageListener to provide more interactive response. This Display might have a lookup table of pre-computed audio files. The audio file could then be sent straight to the audio device. This technique bypasses the mapping from ISS to device parameters, and the synthesis by the Sound Engine, which are both computationally expensive processes.



10•TaDa! demonstrations of auditory information design

Broad categories for further study include further and more sophisticated examples of sonification applied to specific problems. [Scaletti C. (1994)]

This chapter describes the design of auditory displays for 4 information processing scenarios drawn from mining exploration, resource planning and climatology applications. The RiverAndRain scenario is about the siting of a new sewerage treatment works to minimise the environmental impact on a river system. The PopRock scenario involves the assessment of risk in digging a mineshaft. The cOcktail scenario is about modelling climate change from measurements of oxygen isotopes in sea-bed drill-core sites. The LostIn-Space scenario is about navigating back to an interesting place through irregular structures in a 3D visualisation of geology. Experiences with the multimedia interfaces that were implemented shows that the sounds can provide information that is difficult to obtain visually, and can improve the usefulness of the display.

Besides showing ways that sounds can be useful, the demonstrations also show how the TaDa approach to design works in practice. The hybrid system of case-based and rule-based methods and tools that support the approach are shown in Figure 10-1. Facets of the system, and how they fit together, are highlighted in the demonstrations, which follow.

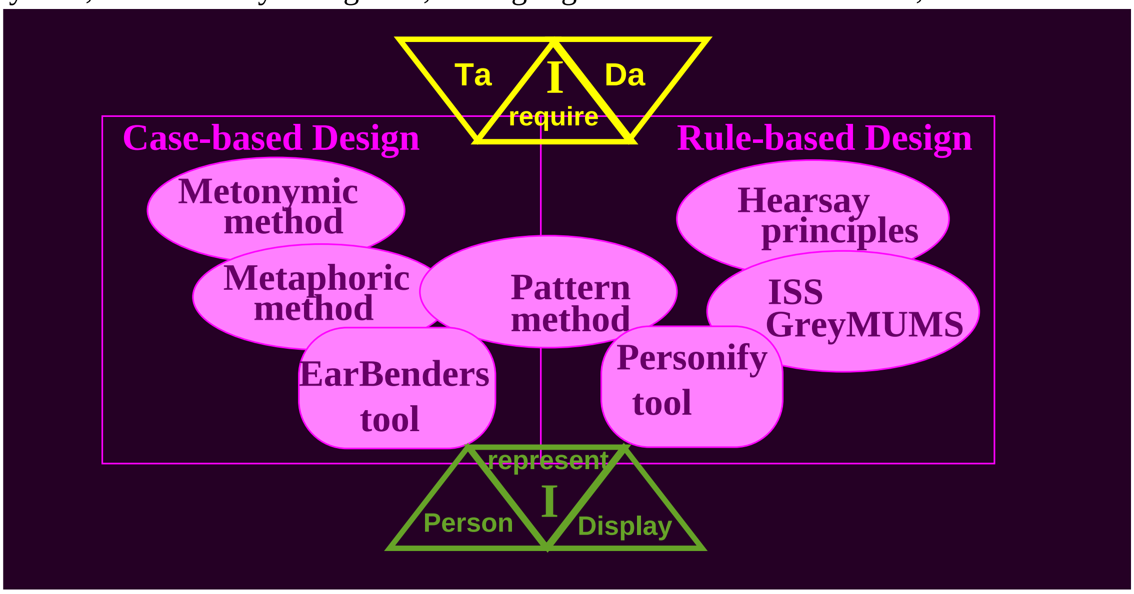


Figure 10-1: A system of methods and tools to support the TaDa approach

10.1 RiverAndRain

The Department of Water Resource is planning a new treatment plant on a river system. An important consideration in the plan is to minimise the environmental impact of the plant on the river system. Engineers have developed hydrological models to predict chemical concentrations and sediment levels along the river [Booty W. and Wong I.W.S. (1994)]. These predictions are to be presented to an audience who may not have in-depth expertise but who need to understand the implications quickly, correctly and confidently. The need for direct presentation led to the development of a visualisation of the river system over a one year period. Changes in chemical concentrations and sediment are seen by changes in the colours of segments of the river in an animated sequence that lasts about 3 minutes, a frame of which is shown n Figure 10-2. Different animations show different chemicals. Rainfall is an important influence on the river, and a bar graph of rainfall levels is in the upper right corner of the display. A yellow cursor moves along under the graph so you can look up the rainfall at the current date as the animation progresses. To see the influence of the rain on the river you can watch the cursor, lookup the graph and then look at the section of the river that interests you. In this way you may notice correspondences between the rainfall record and the chemical concentration at some site. However it can be easy to miss important relations whilst looking back and forward between elements of the display. An improvement would reduce the amount of time spent shifting visual attention to-and-for to obtain the information.

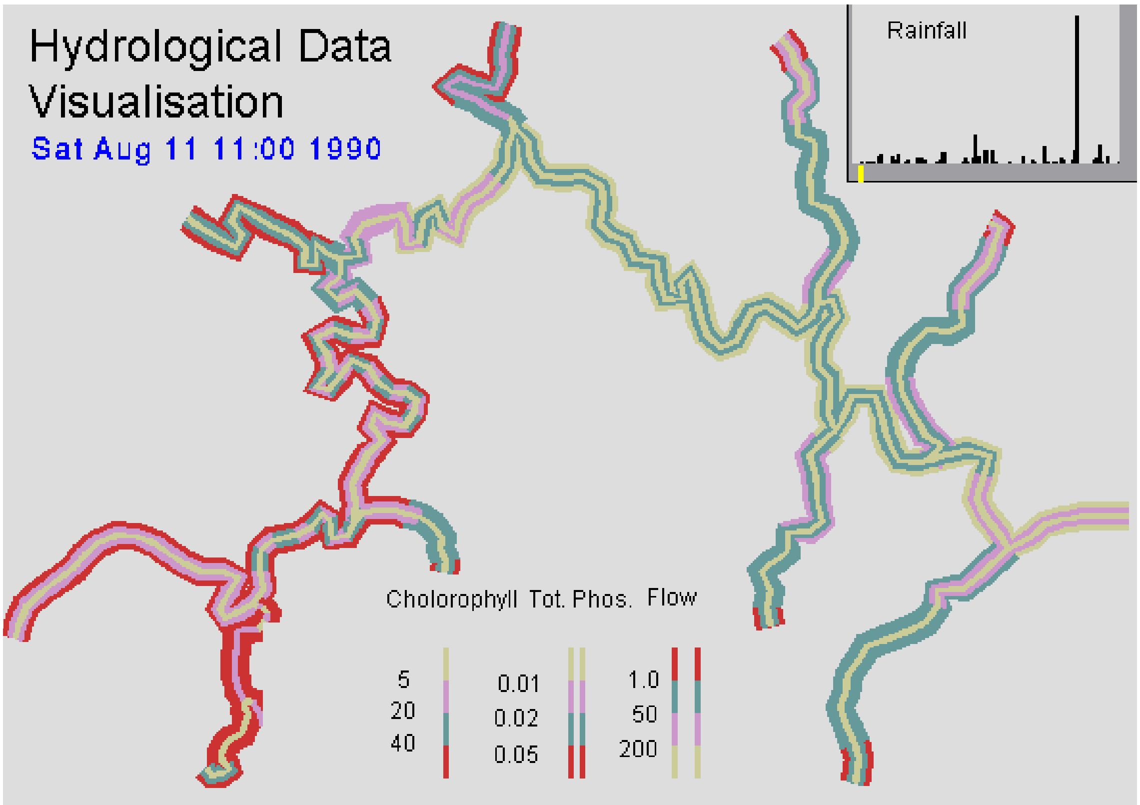


Figure 10-2: A frame from a RiverAndRain animation

10.1.1 TaDa analysis

The first stage of the design process is always a TaDa analysis of the information requirements. The information required by the RiverAndRain scenario can be recast as the Question {what is the rainfall level now?}. The Answers to the Question are {none, low, medium, high}. The Subject of the Question is the rainfall level. The TaDa analysis of these Keys is summarised in Figure 10-3.

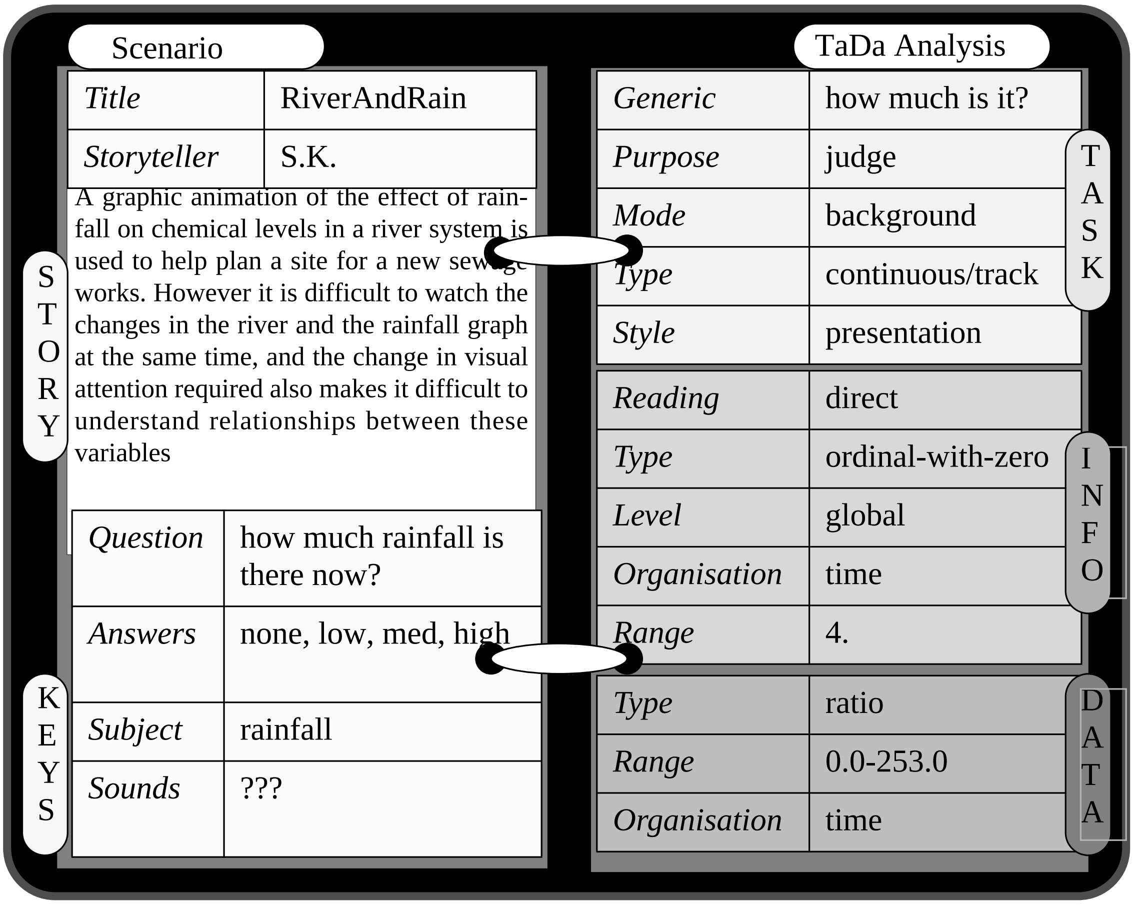


Figure 10-3: TaDa analysis for RiverAndRain

The Task section is an analysis of the Question Key. The Generic question obtained by removing the Subject from the Question is {how much is it?}. The Purpose is to {judge} how much rainfall there is. The Mode is {background} because the primary question in the display is {what is the influence of rainfall on the river?} which requires the correlation of both visual and auditory information. The Type is {continuous/tracking} because the animation is 3 minutes of continuous information. The Style is {presentation} because the display is designed specifically for this particular data set, and will not change after it has been realised.

The Information section is an analysis of the Answers Key. The Reading is {direct} because the presentation should be quickly understood without training, and the observers should be confident of the answers they understand from the display. The Type is {ordinal-with-zero} because there are 4 ordered answers that include {none}. In a {direct} display these answers are perceptually ordered, and the {none} answer doesn’t require a categorically different symbol to be learnt. The Level is {local} because the answers are about the rainfall at a single place in the region. The Organisation is {time} because each answer occurs at a unique time that separates it from other answers. The Range is {4} because there are 4 appropriate answers.

The Data section is an analysis of the Subject Key. The Type is {ratio} because the rainfall measurements have difference, order, metric and a natural zero. The Range of the rainfall is from 0.0 to 253.0 mm in a 3 day period. The Organisation of the rainfall is {time} because each measurement was made over a unique 3 day interval that separates it from other measurements made at the same weather station.

10.1.2 Rule-based design with Personify

The TaDa analysis of the RiverAndRain scenario was entered into the TaDa Panel of the Personify Tool (see Chapter 9), as shown in Figure 10-4. The Info selections were set to Reading ={direct}, Type = {ordinal-with-zero}, Level = {local}, Organisation = {time}, and Range = {4}. The Data is described as {1 channel} of {ratio} data in the Panel. The default tailoring tool selected by the rule-base is the Line tool configured with one end tied to the central axis to give a perceptual zero.

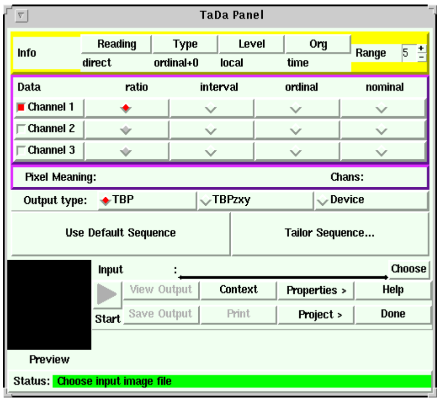


Figure 10-4: TaDa Panel for RiverAndRain

There is a browser at the top of the Line tool where you can select a display device to tailor the design to. The target display is a laptop PC with a 486DX chip running at 33 MHz, a VGA 16 colour monitor, and 16 bit, 44.1 kHz audio. The software component is Macromind Director™. The animation rate is 2 frames per second for the images, and the audio is limited to 8 kHz sample rate to prevent drop-outs due to CPU and I/O load. These display characteristics must be considered for a successful realisation of the design. However a display model has not been perceptually measured for this display, so there is no gamut visualisation to provide guidance in the tailoring for the display. A first pass design was made with the existing GreyMUMS display, which has a gamut of sounds with a 44.1 kHz sample rate (see Chapter 8). The Line tool was adjusted to a maximum brightness of 20% of the range of the GreyMUMS display, as shown by the radius of the selection circle in the horizontal slice view in Figure 10-5. This range constraint is an approximation to the maximum brightness for 8 kHz samples on the target display.

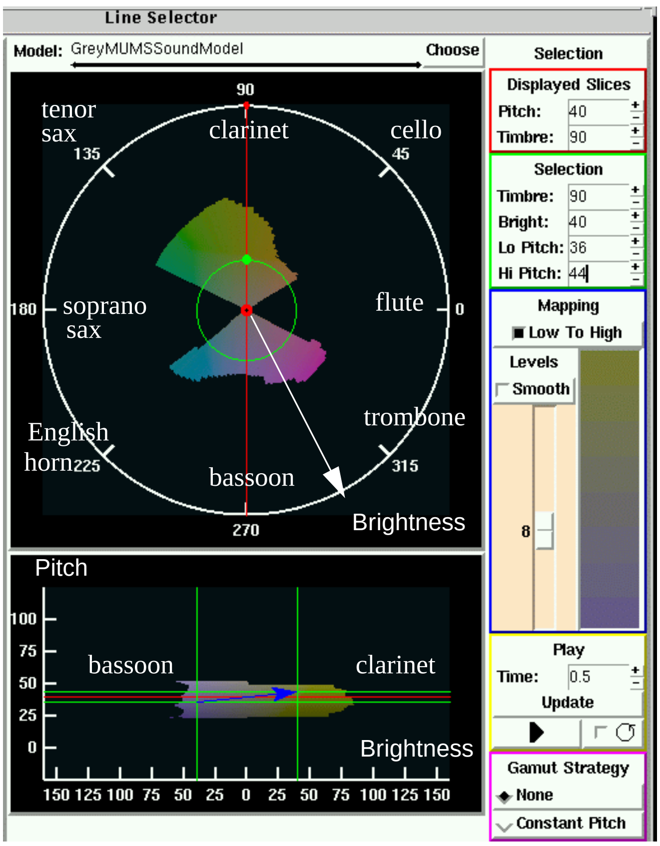


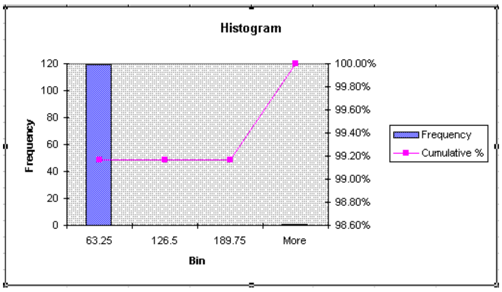
Figure 10-5: Sound design for RiverAndRain

The sound palette that was selected is shown in Table 10-1. This palette was generated at 44.1 kHz Sun™ .au format, then down-sampled to 8 kHz .wav format for the PC. The down-sampling does not alter the brightness of the sounds because they were constrained to be 8 kHz in the first place. The sound files were down-loaded to the PC and linked to appropriate image frames in the animation sequence to produce the final multimedia display.

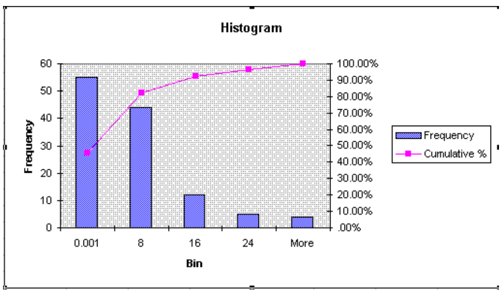
answer	sound palette
{none}	clarinet, pitch 40, brightness 0, duration 0.5s
{low}	clarinet, pitch 41, brightness 10, duration 0.5s
{medium}	clarinet, pitch 42, brightness 20, duration 0.5s
{high}	bassoon, pitch 43, brightness 30, duration 0.5s

Table 10-1: RiverAndRain palette 1

The first test of the display was rather disappointing - all the sounds were identical except for a solitary outlier. The one interesting thing was that this unusual point (a torrential rainfall event) signals a sudden significant change in the river. The problem in the soundtrack is explained by looking at a histogram of the data. The torrential event is so extreme that it has caused all other rainfall events to be binned together into the {none} category, as shown in Histogram 1 of Figure 10-6. In a case like this it is common to remove the influence of the upper and lower 5% of elements on the selection of the bins. However this would hide the significance the torrential outlier amongst all the other elements in the {high} bin. Instead a new answer of {torrential} was added to preserve the significance of the upper 5% of points. The zero values (45% of the elements) were binned in the {none} category, and the rest of the range was linearly distributed to {low, medium and high}, as shown in Histogram 2 of Figure 10-6.



Histogram 1- linear binning



Histogram 2- linear with 5% outliers

Figure 10-6: Mappings from data to information

The introduction of a new answer requires a new sound to be added to the palette. This sound cannot be simply appended by continuing the sequence because the {high} sound is already at the maximum brightness of the display gamut. We could go back to Personify and change the Info Range from 4 to 5 to rebuilt a palette of 5 sounds. However, in the testing of the first display I observed that the {none} answer, given by all the points except the outlier, was not a very direct zero. The original zero in brightness (see Chapter 6) is too subtle for a display which may only be heard one or two times in presentations. The problem was addressed by reinforcing the information zero with a zero in duration that varies linearly from 0.0s for {none} to 0.5s for {torrential}. Since the {none} element has no duration the current palette of 4 sounds was able to be reused, as shown in Table 10-2.

answer	sound palette
{none}	silence
{low}	clarinet, pitch 40, brightness 0, duration 0.12s
{medium}	clarinet, pitch 41, brightness 10, duration 0.25s
{high}	clarinet, pitch 42, brightness 20, duration 0.38s
{torrential}	bassoon, pitch 43, brightness 30, duration 0.5s

Table 10-2: RiverAndRain palette 2

10.1.3 Discussion

The second soundtrack was much more interesting. The order of the sounds is immediate to understand in terms of more and less, and the sounds allow visual attention to remain fixed on parts of the river. The RiverAndRain presentation has been shown several times within the CSIRO Mathematics and Information Sciences. The response has been very encouraging, with people making comments that indicate that the sounds do provide information about the rainfall level, and that this information is easy to understand and relate to the visual variables. Over the course of several presentations the rainfall soundtrack became very familiar, and provided a frame of reference for observations about relationships between visual variables in different animations - for example the torrential event is an auditory milestone which coincides with, and draws attention to, a simultaneous increase in sediment and phosphate levels that might otherwise go unnoticed because they are seen in different animations.

A frequent and significant comment was that “the rain should sound like rain”. The lesson learnt was that the sound palette in a presentation interface should adhere to a familiar sound schema if at all possible. The expectations of a listener can influence what they hear and how they understand the sounds. If the sounds are familiar in the context of the scenario then this may improve the understanding of a display which may only be experienced once or twice. This observation is mirrored in the visual representation of sediment levels by a sequence of browns increasing in darkness which is quickly understood to be an increase in the muddiness (sediment level) of the water. The choice of green or purple to show muddiness is less connected with the everyday world. However green or purple can represent chlorine and other chemicals that aren’t usually seen as coloured, and might be considered an “abstract” palette in this context. The comments and observations made during this demonstration suggest that the immediate understanding of a presentation by a general audience may be improved if the design takes advantage of expectations.

10.2 PopRock

The plan to dig a mineshaft includes the need to minimise the risk of collapse. When the mineshaft is being dug the stresses in surrounding rocks are registered by seismic sensors and logged. The MicroSeismic Animator, shown in Figure 10-7, is a tool for visualising the seismic event log as an aid for planning further mining. The visualisation is an animation of the mineshaft that shows its progression over time. The seismic log is displayed as graphic spheres that appear at the time and place where each event was registered. The effects of the mining can be observed in an animation that shows 90 days in 3 minutes.

However the spheres that mark seismic events appear only very briefly, and even large isolated events can be missed. When many events occur close together at the same time it can be difficult to see how many there were because the spheres cover each other.

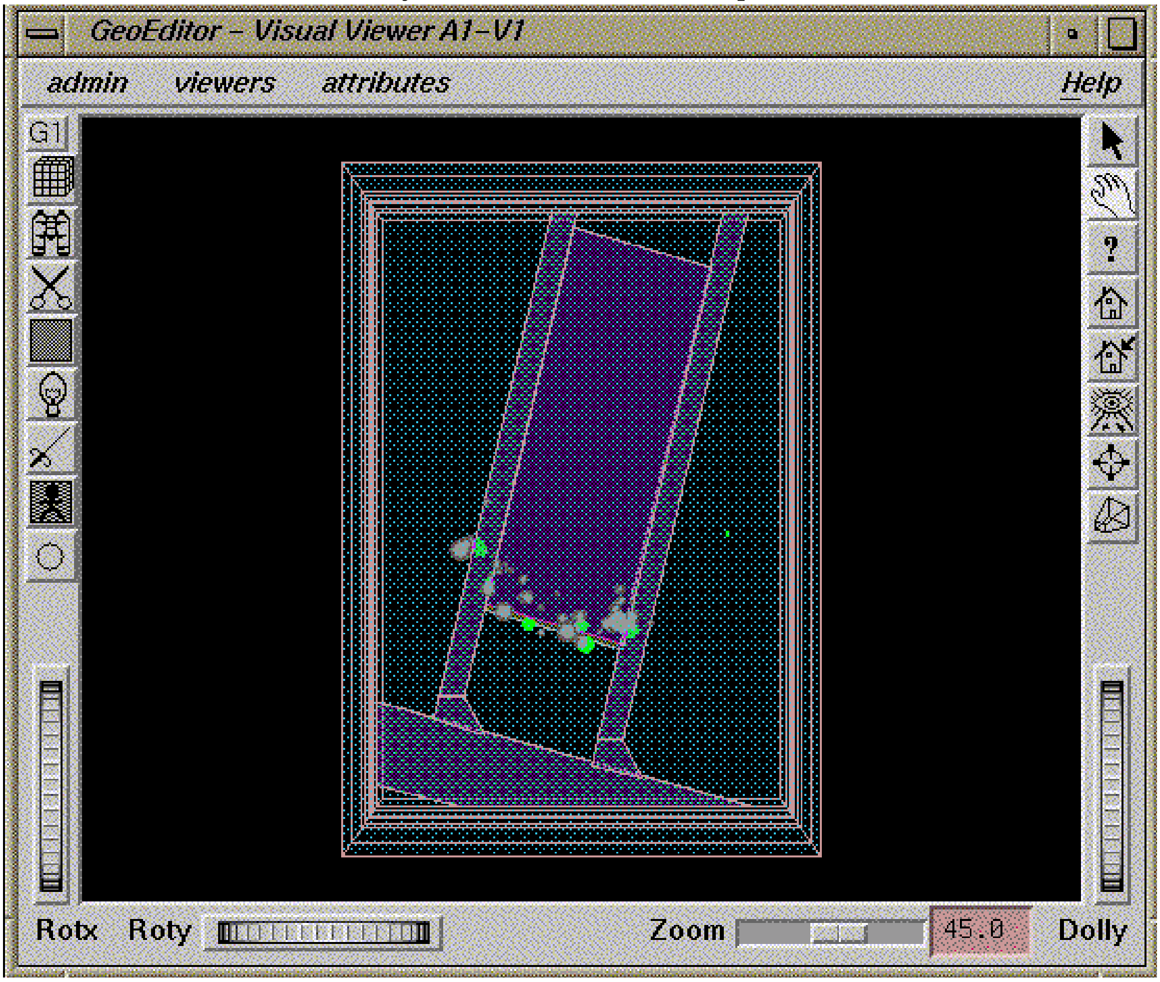


Figure 10-7: Screensnap of the MicroSeismic Animator

10.2.1 TaDa analysis

The information required by the MicroSeismic Animator scenario can be recast as the Question {how much microseismic activity is there now?}. The Answers are {none, one, little, lots, extreme}. The Subject of the Question is the micro-seismic activity. The analysis of the Question, Answers and Subjects keys is summarised in Figure 10-8. The Task section is an analysis of the Question Key. The Generic question obtained by removing the Subject is {how much is there now?}. The Purpose is to {judge} the level of activity throughout the animation. The Mode is {focus} because the relation between the seismic activity and the mineshaft is the primary information in the display. The Type is {continuous/tracking} because the animation and its soundtrack are 3 minutes of continuous information. The Style is {exploration} because the tool has the option to load data sets from different mines so the design cannot be specialised for this one data set.

The Information section is an analysis of the Answers Key. The Reading is {direct} because the display should allow someone who has not had much experience with it to give a correct answer immediately and confidently. The Type is {ordinal-with-zero} because the answers have a simple unidimensional order starting from {none}. The Level is {intermediate} because the answer is about some subset of the entire data set, and is about groups of elements rather than single elements. The Organisation is {time} because the answers are separated by time. The Range of {5} is the number of appropriate answers.

The Data section is an analysis of the Subject Key. The Type is {ratio} because each measurement is an amplitude with a zero level. The Range of measurements is from 0.0 to 2.5. The Organisation is {time and location} because both these dimensions are necessary to separate the measurements. The TaDa analysis for the MicroSeismic Animator is summarised in Figure 10-8.

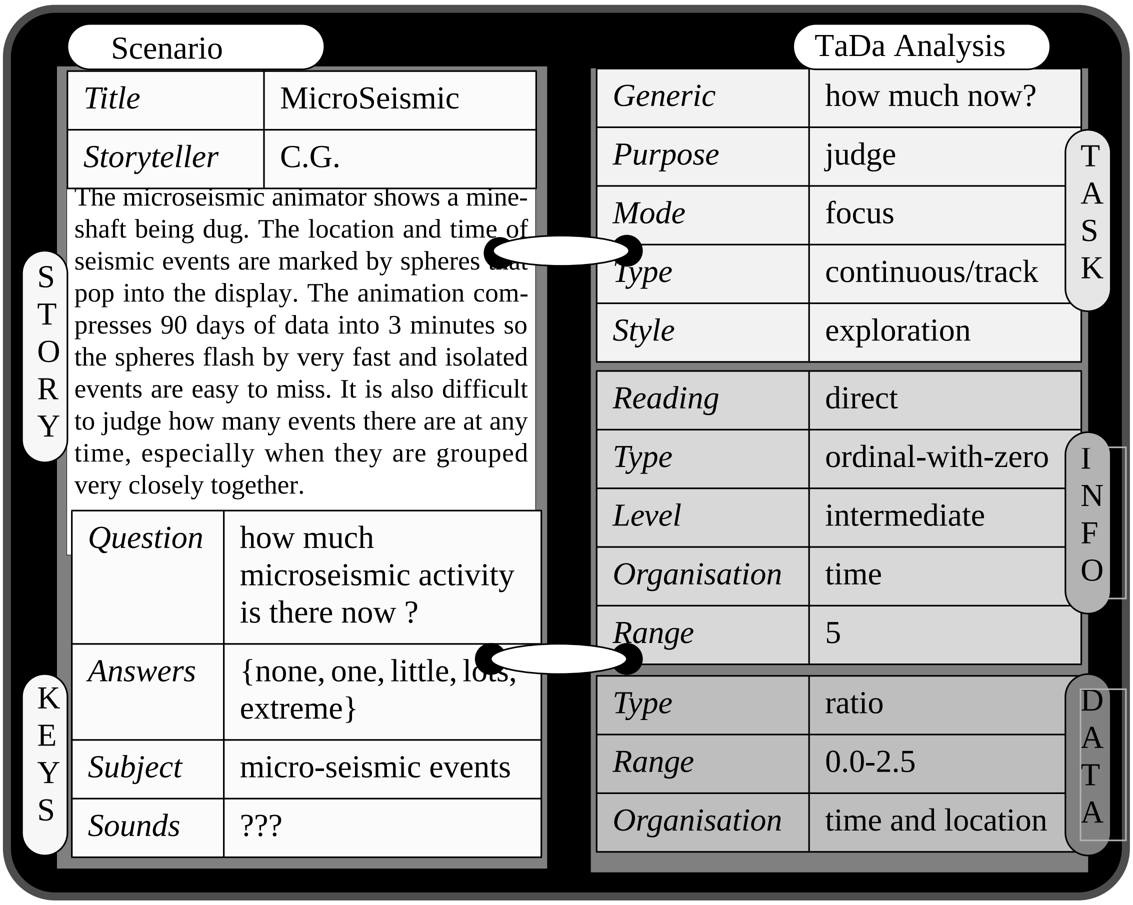


Figure 10-8: TaDa analysis for the MicroSeismic Animator scenario

The TaDa analysis of the MicroSeismic Animator scenario highlights the difference between the information requirements of the display, and the data characteristics of the measurements that were made. The information that is required is the amount of seismic activity during the digging process. Although the data contains amplitudes, these are not of primary importance in this task.

10.2.2 Design with Personify

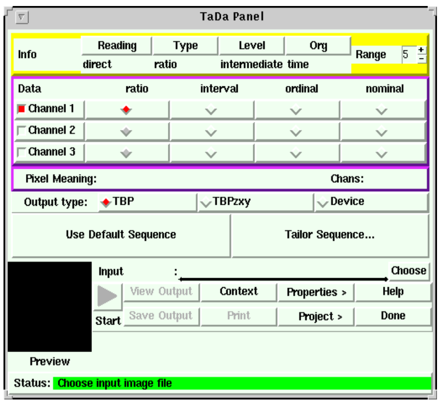


Figure 10-9: TaDa Panel for MicroSeismic Animator

The TaDa analysis of the MicroSeismic Animator scenario was entered into the TaDa Panel shown in Figure 10-9. The Info was set to Reading ={direct}, Type = {ratio}, Level = {intermediate}, Organisation = {time}, and Range = {5}. The Range setting automatically configures the tool to segment the data into 5 equal bins. However, the information we require is not the level of an individual data element, but the level of activity indicated by groups of elements that occur close together in time. The interaction between Level and Range may lead to the proposal of a new Hearsay principle in the future. For now we will accept the segmentation into 5 answers about the level of an individual element that is set by the rule-base. The Data is described as {1 channel} of {ratio} data. The rule-base selects the Line tool and configures it with one end tied to the central axis to give a perceptual zero. The sequence is divided into 5 equal steps in brightness and pitch. The Level is {intermediate} so the duration has been configured to 0.2s. The default is the flute timbre leaf, but the flute has a long slow attack that makes it a poor choice for higher level displays where the sounds must be very short. The instrument with the fastest attack in the GreyMUMS palette is the soprano sax, as indicated by its position opposite the flute in the timbre circle (see Chapter 8). The selection in the Line tool was manually configured to the soprano sax leaf, as shown in Figure 10-10. The observation about the effect of Level on the choice of timbre may be fed-back into the TaDa system as an amendment to the Hearsay principle of Level which might go something like - higher level information is best shown by rapid sounds with temporal features less than 0.1s in duration. The amended principle can be migrated into the Personify rule-base so that the choice of a higher Level in the TaDa Panel automatically configures the tailoring tools to the timbre with the shortest attack segment.

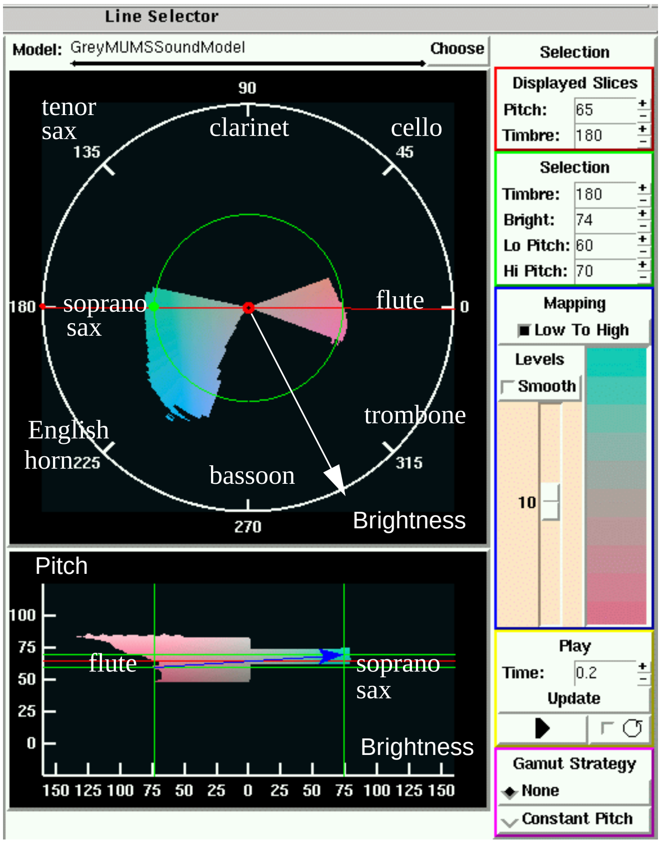


Figure 10-10: Line tool for MicroSeismic Animator

The palette of sounds selected to represent a seismic event is shown in Table 10-3.

answer	sound palette
{0.5}	soprano sax, pitch 65, brightness 0, duration 0.2s
{1.0}	soprano sax, pitch 65, brightness 18, duration 0.2s
{1.5}	soprano sax, pitch 65, brightness 36, duration 0.2s
{2.0}	soprano sax, pitch 65, brightness 54, duration 0.2s
{2.5}	soprano sax, pitch 65, brightness 72, duration 0.2s

Table 10-3: Palette for MicroSeismic Animator

10.2.3 Realisation

The target display is a Silicon Graphics Indy workstation with 4 channels of CD quality (44.1 kHz sample rate) sound. The palette generated with Personify was saved as 44.1 kHz .aif files. The MicroSeismic Animator was modified to play a soundtrack during the visualisation of a mining operation. This involved a change to the rendering function of the SeismicEvent object so that it plays a sound file just before the sphere marker is painted to the display. This change showed up a problem with the synchronisation of sound and graphics in the ViewKit programming library. The sounds play when the rendering function is called on a SeismicEvent, but the sphere does not appear until the processing thread returns to the main event loop, at which point all the graphics are flushed to the screen at the same time. This causes the sounds to precede the graphics, which all appear at the same time. The problem may be solved by flushing each sphere to screen when the rendering function is called, but this requires lower level programming at the level of Xlib that hasn't been attempted at this stage.

10.2.4 Discussion

Despite the synchronisation problem the sounds do provide information about the amount of activity in each frame. The sounds draw attention to isolated events that might otherwise go unnoticed, by preparing you to scan the display for the marker that shows where the event occurred. Frames that look as if they have similar levels of activity are heard to be quite different, because overlapping sounds have additive properties that overlapping spheres do not.

In the RiverAndRain demonstration it was observed that listeners may have everyday expectations about sounds. The soprano sax is not a sound that you would normally expect to hear while digging a mine, and perhaps a more connotative sound would be better. However a search through EarBenders for stories containing keywords {mining, fracture, seismic} from the MicroSeismic scenario didn't retrieve any relevant stories. If we are particularly interested in designing displays for mining scenarios then it may be worthwhile collecting stories from miners and others who have experience in the mining domain. In this way the EarBenders case-base can be developed for a particular application domain. Another way to find a connotative palette is the metaphoric method. This time the query did retrieve some cases - Insect aggregation in Figure 10-11, Croaking frogs in Figure 10-12, and Popcorn Figure 10-13. These cases are about judging the size of populations of insects and frogs, and judging the amount of popcorn popping. These judgements are all made by listening to the sum of many similar sounds. However the connotations of frogs, insects or popcorn may interfere with the understanding of the scenario. In this design the “abstract” musical palette may not connote mining, but may be less incongruous than these other metaphors.

Scenario

STORY		
Title	Insect aggregation
Storyteller	W.B.

Calling by aggregated insects - the group selection theory was that chorusing groups evolved as displays that allowed individuals to gain information as to the size of the population, and adjust their behaviour accordingly.

KEYS	
Question	how many insects around here ?
Answers	none, one, several, a lot
Subject	insects
Sounds	species specific chirps, buzzes etc.

TaDa Analysis

TASK	
Generic	how many ?
Purpose	judge
Mode	focus
Type	continuous/track
Style	exploration

INFO	
Reading	direct
Type	ordinal-with-zero
Level	global
Range	4
Organisation	location

DATA	
Type	ratio
Range	?
Organisation	location

Figure 10-11: Insect aggregation

Scenario

STORY	
Title	Croaking frogs
Storyteller	S.B.

You can hear that there are a lot of frogs in the ponds during the wet season. Zoologists estimate frog populations by listening to tape recordings made of ponds at night.

KEYS	
Question	how many frogs are there around here ?
Answers	none, one, several, a lot
Subject	frogs
Sounds	croaking

TaDa Analysis

TASK	
Generic	how many ?
Purpose	judge
Mode	background
Type	continuous/track
Style	exploration

INFO	
Reading	direct
Type	ordinal-with-zero
Level	global
Range	4
Organisation	location

DATA	
Type	ratio
Range	?
Organisation	location

Figure 10-12: Croaking frogs

Scenario

Title	Popcorn
Storyteller	S.B.

STORY

when you pop popcorn it starts off slowly popping then builds up in rate and loudness, then dies down again as most of the corn has popped. you can tell how many popcorns are going off all through the cooking process, and you can tell when its finished

Question	how many popcorns are popping?
Answers	{none, hardly any, some, lots}

KEYS

Subject	popcorn
Sounds	popping

TaDa Analysis

TASK

Generic	how many ?
Purpose	judge
Mode	background
Type	continuous/track
Style	exploration

INFO

Reading	direct
Type	ordinal-with-zero
Level	intermediate
Range	4
Organisation	time

DATA

ratio	ratio
Range	?
Organisation	location

Figure 10-13: Popcorn

10.3 cOcktail

Paleoclimatologists are interested in understanding the Earth's climate in prehistoric times. The amount of O_{18} in sea-shells varies with the temperature of the water the creature was living in, and cores that cut through layers of sediment provide a record of temperature changes over the past 400,00 years. [Zubakov V.A. and Borzenkova I.I. (1990)]. The O_{18} measurements from 12 cores drilled at sites around the Atlantic is available from the specmap archive [Imbrie J. McIntyre A. and Mix A.C. (1990)]. This data is shown as a set of overlaid time-series plots in Figure 10-14. Changes in the climate over time across the region can be seen as trends, groupings and patterns in these traces. However it can be difficult to follow a single record through the spaghetti of plots, or track relationships between records.

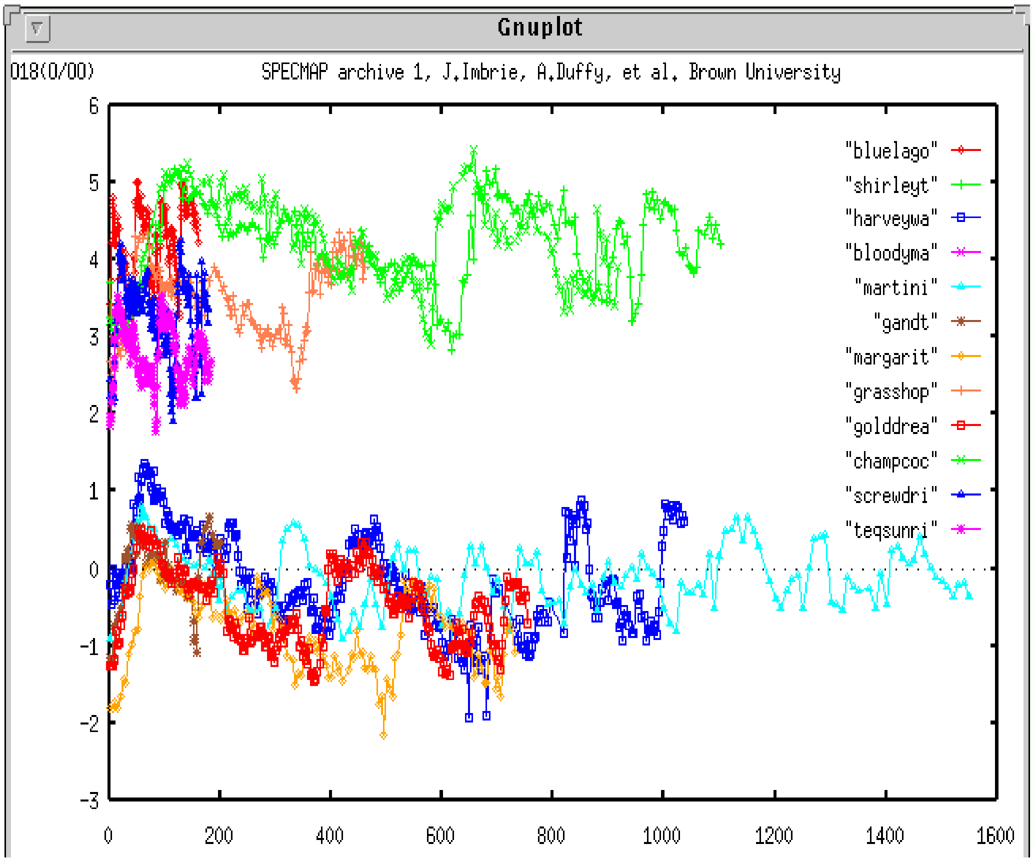


Figure 10-14: O_{18} levels from 12 deep-sea drill-cores

10.3.1 TaDa analysis

The information required in the scenario can be recast as the Question {what is happening to O_{18} at site X?}. The Answer is a combination of {low, medium and high} and {decreasing, stable, increasing}. The Subject of the Question is the O_{18} at site X.
The Task section is an analysis of the Question Key. The Generic question obtained by removing the Subject is {what is happening?}. The Purpose is to {analyse} what is happening to the level of O_{18} at some site. The Mode is {focus} because there is information from many sites available at the same time, but we are interested in a particular site. The Type is {continuous/tracking} because we would like to be able to track the information about a particular site through the information about all the other sites in the display. The Style is {presentation} because the display is to provide specific information which will not change after the display is realised.

The Information section is an analysis of the Answers Key. The Reading is {direct} because the presentation should be quickly understood without training, and the observer should be confident of the answers they understand from the display. The Type for the information about the O_{18} is {ordinal} and the Type for the site is {nominal}. The Level is {local} because the answer is about a specific site. The Organisation is {time / location} because answers are separated by when and where they occur. The Range for O_{18} is {3} and for site is {12}.

The Data section is an analysis of the Subject Key. The Type is {ratio / nominal} because the O_{18} measurements are ratio and the site locations are nominal. The Range of the measurements is from {-3.0 to 5.0} for $O_{18,}$ and there are {12} sites. The Organisation of the measurements is by {time} down the drill hole, and by {location} across the sites. The TaDa analysis is summarised in Figure 10-3.

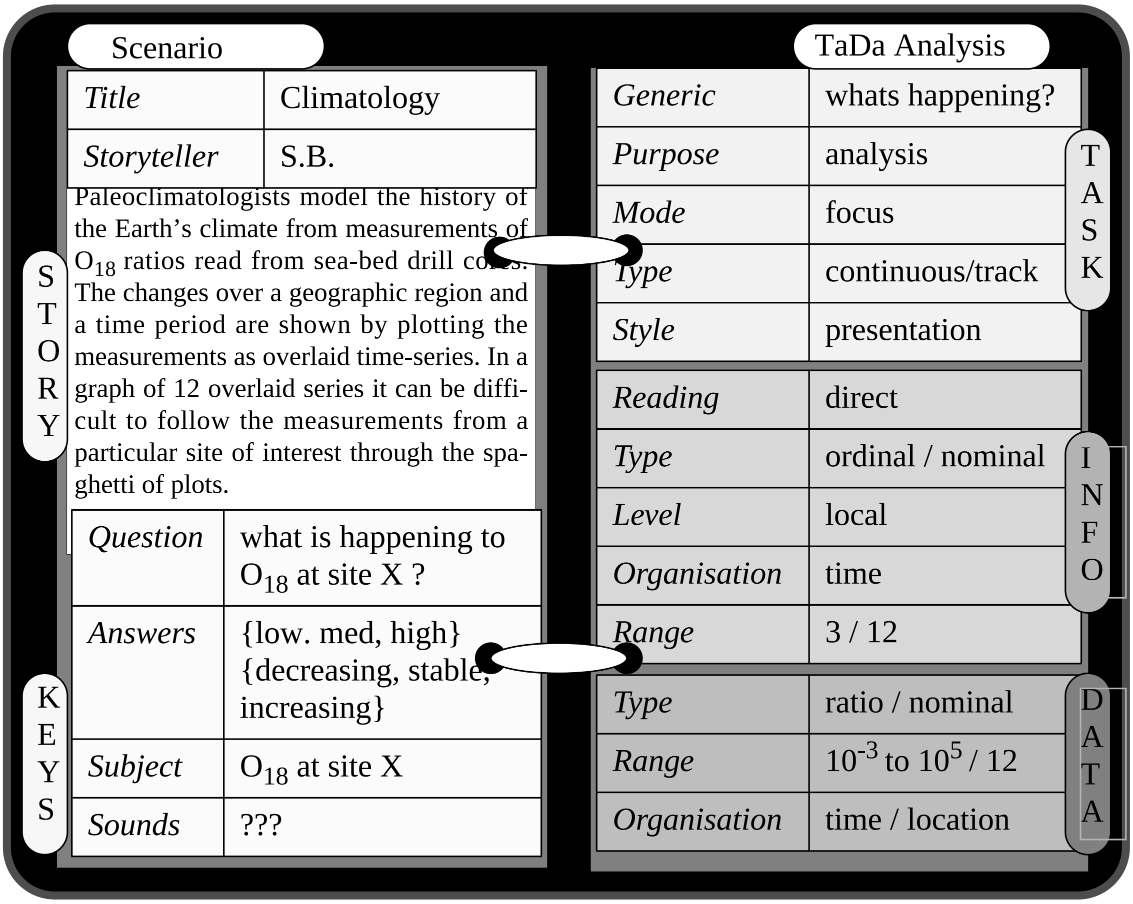


Figure 10-15: TaDa analysis for the Climatology scenario

10.3.2 Pattern method

The 12 drill-sites are nominal variables that need to be separately observable in the display. The rule-based method in Personify is built on an Information-Perception Space (IPS) (see Chapter 7) which supports the display of up to 8 nominal variables. This rules out Personify as a tool for designing this display. This is where the flexibility of a multi-faceted system is valuable - if a problem is not amenable to one method then there is an alternative. In this instance we can fall back on the pattern method (see Chapter 5). We begin the pattern method by entering the TaDa analysis of the scenario into the EarBenders case-base as a query. The 3 best matching cases retrieved for the Climatology query were Walking-behind in Figure 10-16, Bicycling-in-traffic in Figure 10-17, and Cocktail-party Figure 10-18. Walking-behind is about footsteps approaching from behind as you walk along the pavement. Bicycling-in-traffic is about the importance of listening for maintaining awareness of traffic. Cocktail-party is about listening to snatches of conversation in a room full of chattering people, clinking glasses, and music. Although the sounds in each case are very different, they have quite similar auditory characterisations. In Walking-behind the sounds are continuous and repetitive, local to a particular source, moving smoothly in location, and changing in distance. In Bicycling-in-traffic the sounds are also continuous, moving smoothly, and changing in distance. However this story involves many sources moving around at once. The sounds in Cocktail-party are similarly smoothly moving, continuous sources that vary in distance from the listener. The auditory specification synthesised from trends in these characterisations is a global display of many continuous, smoothly moving, sound sources that vary in distance from the listener, as shown in Figure 10-19.

.

Scenario

	Title	Walking behind
	Storyteller	S.B.
STORY	when walking home you can hear someone behind you approaching as their footsteps get louder and also by the rate of their steps which is faster than your own	
	Question	what is the person walking behind me doing ?
	Answers	dropping back, staying level overtaking
KEYS	*Subject*	person behind
	Sounds	footsteps getting louder or softer, faster or slower

TaDa Analysis

	Generic	whats happening ?
TASK	*Purpose*	analysis
	Mode	background
	Type	continuous/track
	Style	exploration
	Reading	direct
INFO	*Type*	ordinal
	Level	local
	Range	3
	Organisation	time
DATA	*Type*	ratio
	Range	?
	Organisation	time

Figure 10-16: Walking behind

Scenario

	Title	Bicycling traffic
	Storyteller	P.V.
STORY	riding a bike you listen all the time - Crossing blind intersections. - Pedestrian road crossings. - Trucks, cars traffic behind your bicycle. - Bicycle coming along footpath	
	Question	what are other vehicles doing ?
	Answers	{car, bus, truck} is {behind, level, in front}
KEYS	*Subject*	other vehicles
	Sounds	location of cars, trucks, buses, zooming by

TaDa Analysis

	Generic	whats happening ?
TASK	*Purpose*	analysis
	Mode	background
	Type	continuous/track
	Style	exploration
	Reading	direct
INFO	*Type*	nominal / ordinal
	Level	global
	Range	? / 3
	Organisation	category+location
DATA	*Type*	nominal / ratio
	Range	?
	Organisation	time + location

Figure 10-17: Bicycling in traffic

Scenario

Title	Cocktail party
Storyteller	S.B.

STORY

at a cocktail party you can listen to the person you are conversing with, whilst monitoring another conversation elsewhere in the room, or follow someones voice or identify people and groups of people, all against the hubbub of voices, music and other background noise which indicates the degree of animation of the proceedings...

KEYS

Question	who is here and what are they saying?
Answers	various people
Subject	who
Sounds	chattering voices

TaDa Analysis

TASK

Generic	whats happening?
Purpose	analyse
Mode	background
Type	continuous/track
Style	exploration

INFO

Reading	direct
Type	nominal
Level	global
Range	?
Organisation	category

DATA

Type	nominal
Range	?
Organisation	category+location

Figure 10-18: Cocktail party

Behind

Sounds	footsteps getting louder or softer,faster or slower
Nature	everyday
Level	local
Streams	single
Occurrence	continuous
Pattern	cycle
Movement	smooth
Type	metathetic
Compound	separable
Descriptors	footsteps, be

Occurrence	continuous
Pattern	unpredictable
Movement	smooth
Type	categories
Compound	separable
Descriptors	voices, chatter

Bicycling

Sounds	location of cars, trucks, buses, closer, zooming by
Nature	everyday
Level	global
Streams	many
Occurrence	continuous
Pattern	unpredictable
Movement	smooth
Type	metathetic / categories
Compound	separable
Descriptors	cars, trucks, buses, closer, zooming

SOUND

Climatology

Sounds	???
Nature	everyday 2
Level	global 2
Streams	many 2
Occurrence	continuous 3
Pattern	unpredictable 2
Movement	smooth 3
Type	metathetic 3
Compound	separable 3
Descriptors	footsteps, louder, softer, faster, slower, trucks, buses, zooming, moving, voices, chatter

SOUND

Figure 10-19: Pattern method for the Climatology scenario

The auditory characterisation is a low-level specification of the perceptual structure of the design, but it doesn’t prescribe the high-level palette of sounds to realise those characteristics. As we have previously observed, the best choice of palette is one that connotes the design scenario. Unfortunately a metonymic search of EarBenders for the keywords {oxygen, isotope, drill-core, climate} did not retrieve any cases. It isn’t too surprising really that there aren’t any stories about listening to oxygen isotopes in sea-bed drill cores. Another way to find sounds with a connection to the design domain is with the metaphoric method. A metaphor is a familiar schema that may improve the understanding and perception of the auditory structure. Bregman observes that familiarity and expectation can significantly improve a listeners ability to hear elements and reject interference in a complex auditory scene [Bregman A.S. (1990)]. We have already found the best matching cases as part of the pattern method, which were about footsteps, traffic sounds, and a cocktail party. Metaphorical sounds need to be understood in the unusual context of a computer interface, and should connote the metaphor strongly enough to allow the leap in imagination needed for a metaphor to be recognised. Voices are perhaps the most unambiguously connotative of the 3 metaphors. The cocktail party is a familiar cliche that may improve recognition and transposition. This metaphor is also an opportunity to explore verbal sounds as a representation of non-verbal information in an auditory display.

The metaphorical sound palette must have the perceptual structure specified by the low-level auditory characterisation. A major requirement is that the listener be able to hear each nominal variable as a separable element in the display. We distinguish voices by sex, size, maturity, emotion, and location, and these human factors are related to auditory factors that affect perceptual streaming, such as timbre, pitch, brightness, and spatial position [Bregman A.S. (1990)]. By maximising the differences between the human factors we also maximise the perception of separate categorical streams. Each drill-core site was allocated a unique voice and a unique location in the display. The voices could speak nonsense words, but the words are an opportunity to build a semantic link to the design scenario. The name of a drill-core is an immediate reference that can be understood by all English speakers. Words have advantages for displaying large catalogues of nominal data, because they are discriminable and unambiguous. Unfortunately the identities of the drill-cores are alpha-numeric codes, like RC12-294 which take a long time to say, and are not very memorable unless you have been using them a lot. In the interests of investigating the design I substituted the names of cocktails in place of drill-core codes. Although the cocktails connote the metaphor rather than the application, they can help in testing the feasibility of this design. The sound palette for the Climatology scenario is shown in Table 10-4.

drill-core site	word, speaker, sex, location
RC12-294	“blue-lagoon”, S.B, male, 0 degrees
RC13-228	“shirley-temple”, D.S., female, 8 degrees
RC24-16	“harvey-wallbanger”, S.K., male, 16 degrees
V12-122	“bloody-mary”, M.C., male, 24 degrees
V22-174	“martini”, A.L., male, 32 degrees
V25-21	“gin-and-tonic”, R.U., male, 40 degrees
V25-56	“margarita”, D.L., female, 48 degrees

Table 10-4: Climatology palette

drill-core site	word, speaker, sex, location
V25-59	“grasshopper”, O.B, male, 56 degrees
V30-40	“golden-dream”, S.V.,female, 64 degrees
V30-97	“champagne-cocktail”, K.S., female, 72 degrees
V30-49	“screwdriver”, D.B., male, 80 degrees
RC11-120	“tequila-sunrise”, T.L., male, 88 degrees

Table 10-4: Climatology palette

The variation in the oxygen isotope at each site also needs to be heard in the display. This presents complications at both the structural and metaphorical levels. A variation in the auditory properties of sounds in a mixture can cause regroupings that may merge or disintegrate the perception of some nominal elements. The variation of the voices also needs to maintain the connotation of the display. A variation found in each of the metaphorical cases retrieved by the pattern method was the smooth change in the distance of the sound source from the listener. A change in the distance of each voice maintains the cocktail party metaphor whilst allowing variation in the information about each site. The distance of a sound influences loudness, which can represent ordinal information as required. Bregman observes that similarity in loudness has only weak effects on the perceptual grouping of sounds, which supports the notion that we can vary the loudness of a voice without causing a regrouping of the elements of the display. The auditory variation in the oxygen isotope is shown in Table 10-5.

O18 level	ordinal variation
low	far away
medium	mid distance
high	close

Table 10-5: Climatology palette

10.3.3 Realisation

The target display is a Sun™ workstation with stereo dbri audio device. Samples of people of different ages and sexes speaking the names of cocktails were recorded in mono at 8 kHz sample rate with the Audio Tool and microphone. The samples were looped and placed in a spatial audio display, built with Csound. This display, listed in Appendix 10-1, provides a spatial audio capability through stereo headphones. It is a model of a head shadow implemented by an inter-aural delay, and frequency and amplitude attenuation. Each sound can be placed at a unique angular location with this effect. The change in distance of the sound is an attenuation of the loudness by 10dB for every doubling in distance, based on observations of free field attenuation of sounds made by sound engineers. A spectral attenuation of the upper harmonics is also included as a cue for distance. There are much more sophisticated spatial audio displays available, but this display was adequate for the task. In retrospect I realised that the display implementation could have been greatly simplified by recording the location cue for each voice with a stereo sample. Although the angular position of the voices would be fixed, this recording would provide the spatial cues for perceptual segregation of the voices which was all that was required.

The data values were mapped from the data archive format to a Csound score with a Perl script, called cOcktail.prl, listed in Appendix 10-2.

10.3.4 Discussion

The display clearly connotes a roomful of voices. You can hear the chatter of people speaking from different directions, and voices rising and falling as time progresses. At first it is hard to understand what some of the voices are saying, and it may be necessary to introduce each word individually at the start of the display, or provide a list of the cocktails to prepare the listener to hear them. You can listen for a particular cocktail, and track what happens to it during the course of the display. For example “margarita” spoken by a female voice is located in the middle of the aural space. By listening to it at the same time as watching the graph you can see the correspondence between the variation in loudness and variation in the height of the line. In the initial part of the graph “margarita” is overlaid by many other plots and you can’t see what is happening to it in places. However, if you listen while watching this plot you realise that it dips downward in this section, and you can tell where to look for parts of it that appear here and there. The ears lead the eyes. If midway through the display you want to switch attention to “champagne cocktail” you can mentally select it and listen for it. The ability to mentally switch between information streams keeps the focus on the task and avoids a diversion to a control panel to select local elements from the display. This is only a very preliminary investigation but it demonstrates that an auditory display of 12 time-series records is feasible. This demonstration also shows that a verbal display can provide non-verbal information. Spoken words may be particularly effective for representing large catalogues of nominal data, and have immediate advantages for novice users.

10.4 LostInSpace

The GeoViewer is a 3D visualisation of geological structure to support mining applications, shown in Figure 10-20. This tool enables you to move around exploring the shapes and intersections of rock layers.

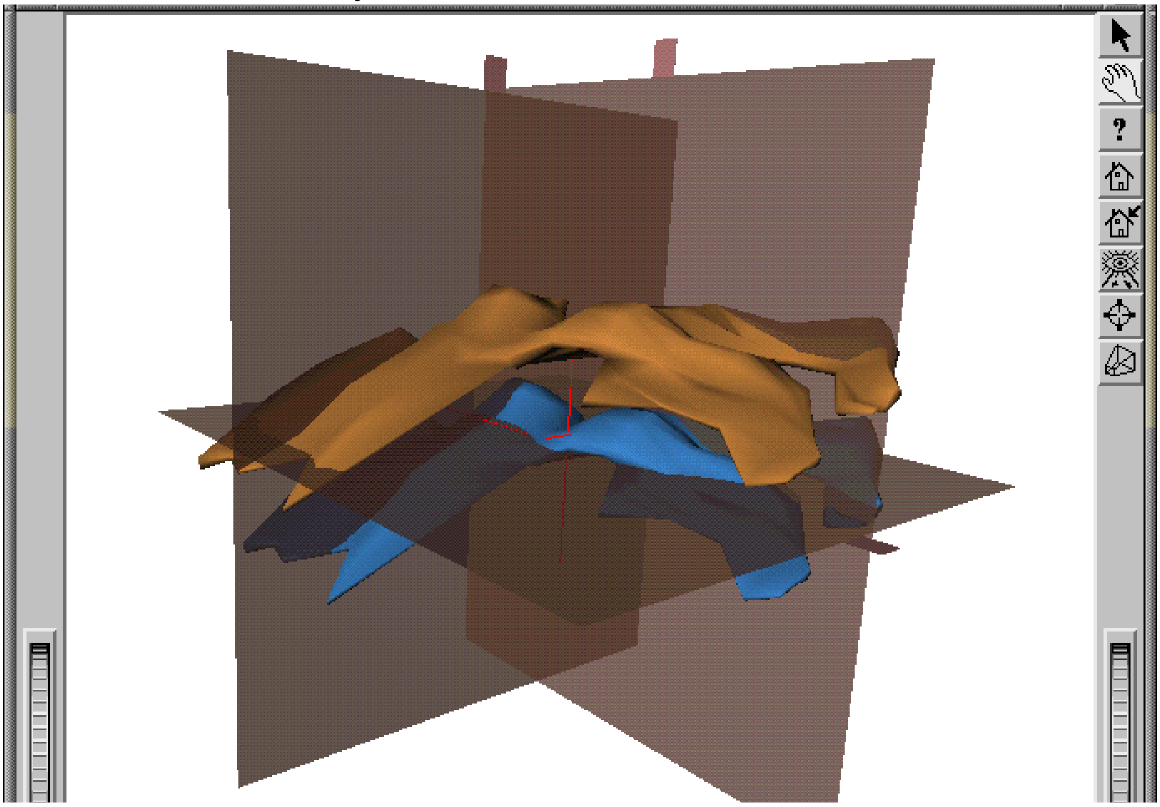


Figure 10-20: Screen snapshot of the GeoViewer

You can move anywhere in the space, and look in any direction. Although the visual freedom is useful, you can easily lose track of where you are, where you are looking, where locations of interest are, and even which way is up. The viewpoint is shown as x,y,z coordinates in an interface panel, but the calculation of a heading to previously recorded coordinate is laborious and can slow down exploration activities. The geological structures are visually irregular and it can take some time before you learn to recognise landmarks that can help when you want to return to previous location.

10.4.1 TaDa analysis

The information required by the scenario can be recast as the Question {where was that place again?}. The Answers are a heading {forward, back, left, right, up, down} paired with a distance such as {stop, a bit more, a lot more}. The Subject of the Question is the location of a place that has previously been visited. The TaDa analysis is summarised in Figure 10-21.

Scenario

Title	Navigate
Storyteller	D.L.

STORY

The GeoViewer viewpoint can be moved around in 3 spatial dimensions. It can be hard to remember where you have been and how to get back to previous viewpoint. The location in the space is shown as x,y,z coordinates in an interface panel, but calculating the direction to a point slows the interactive navigation. Can we use an auditory display to also assist in navigation back to an interesting location?

KEYS

Question	where was that place again ?
Answers	{forward, back, left, right, up, down} {stop, bit more, lot more}
Subject	a location
Sounds	???

TaDa Analysis

TASK

Generic	where is it?
Purpose	navigate
Mode	interactive
Type	branching/decisio
Style	exploration

INFO

Reading	direct
Type	nominal / ordinal
Level	local
Organisation	location
Range	6 / 3

DATA

Type	interval / ratio
Range	?
Organisation	location

Figure 10-21: TaDa analysis for the GeoViewer Navigation scenario

The Task section is an analysis of the Question Key. The Generic question obtained by removing the Subject is {where is it?}. The Purpose is to {navigate} back to a location of interest that has been previously visited. The Mode is {interactive} because navigation is done by steering the viewpoint with the mouse. The Type is {branching/decision} because the sounds influence moment to moment steering decisions. The Style is {exploration} because different data-sets to can be loaded so the design cannot be tailored to a specific data-set.

The Information section is an analysis of the Answers Key. The Reading is {direct} because the listener should be able to Answer the Question immediately, confidently and with little training. The Type is {nominal / ordinal} because the heading is a set of cate-

gories, and the distance is classified into ordered categories. The Level is {local} because the answer is about a particular location. The Organisation is {location} because different answers occur at different locations. The Range is {6} headings by {3} distances.
The Data section is an analysis of the Subject Key. The Type is {interval / ratio} because the heading has a conventional zero, and a distance has a natural zero. The Range of vectors is unlimited, and depends on the scale of the data loaded into the GeoViewer. The Organisation is {location} because each data measurement has a unique location that distinguishes it from all other measurements.

10.4.2 Metonymic method

The metonymic method can help find a familiar schema semantically linked with the application domain. Some keywords from the GeoViewer Navigation scenario were listed as {remember, find, location, position, and distance}. The EarBenders stories retrieved by these keywords are summarised in Table 10-6.

Keyword	Title	Story Summary
remember	Diabetes-test	have a test if you remember this tune from your teens
	Troubadours	remember and spread news by singing tunes
	Marimbas	easier to remember pitch of keys than visual length
find	Dropped object	you can hear where to find a dropped coin
	Rescue	rescuers listen to find people in collapsed buildings
	Cordless-phone	you can page a cordless phone to find the receiver
	Dog-bell	you can find ziggy walking in the bush from her bell
	Cooee	bushwalkers find each other by shouting “cooooeee”
	Dropped-key	finding a dropped door key in the dark by ear
	Finding-studs	knock on the wall to find where the studs are
	Party-game	children find a hidden toy by warmer/colder directive
location	Car-speed	you can hear the location of a car going past
	Wire-pairs	telecom techs locate wire pairs with beeping device
position	Songlines	landmarks were sung into position in the Dreamtime
	Dolphins	a dolphin triangulates position from echoes off reef
distance	Lightning	distance of lightning calculated from time til thunder
	Wolf-whistle	someone at a distance wants to get your attention
	Male-spacing	male insects space themselves by loudness
	Courtship	insects call to attract mates from a distance
	Insect-distance	insects hear distance by frequency attenuation

Table 10-6: Metonymic stories for the GeoViewer Navigation scenario

All the “remembering” stories involve a tune. The “find” stories mostly involve listening for the location where a sound came from, except for 2 stories - Finding-studs and Party game. Finding-studs involves listening for a change in brightness or hollowness, and Party-game involves someone telling verbal instructions. Several stories with “location”, “position”, and “distance” in them involve hearing the spatial location where the sound came from. However there were exceptions that show the spatial cue is only one of a number of ways we hear where things are. Wire-pairs is about locating a connecting wire by stroking the ends of a bundle with a probe. Songlines is about the aboriginal Dreamtime songs where the relations between places in a journey are represented by relations between references in the song. Dolphins and Lightning tell how distance can be heard by auditory delays. Insect-distance tells how insects have specially tuned auditory filters that detect distance by the attenuation in upper frequency bands.

10.4.3 Metaphorical method

The number and diversity of stories retrieved with the keyword search highlights many different ways that sounds can help us remember something or find something. Now that we have some seed designs, which to choose? We can turn to the pattern method to search for auditory regularities in the stories that best match the information requirements of the design scenario. The TaDa requirements for the GeoViewer Navigation scenario were entered as a query into EarBenders to generate the auditory characterisation in Figure 10-22.

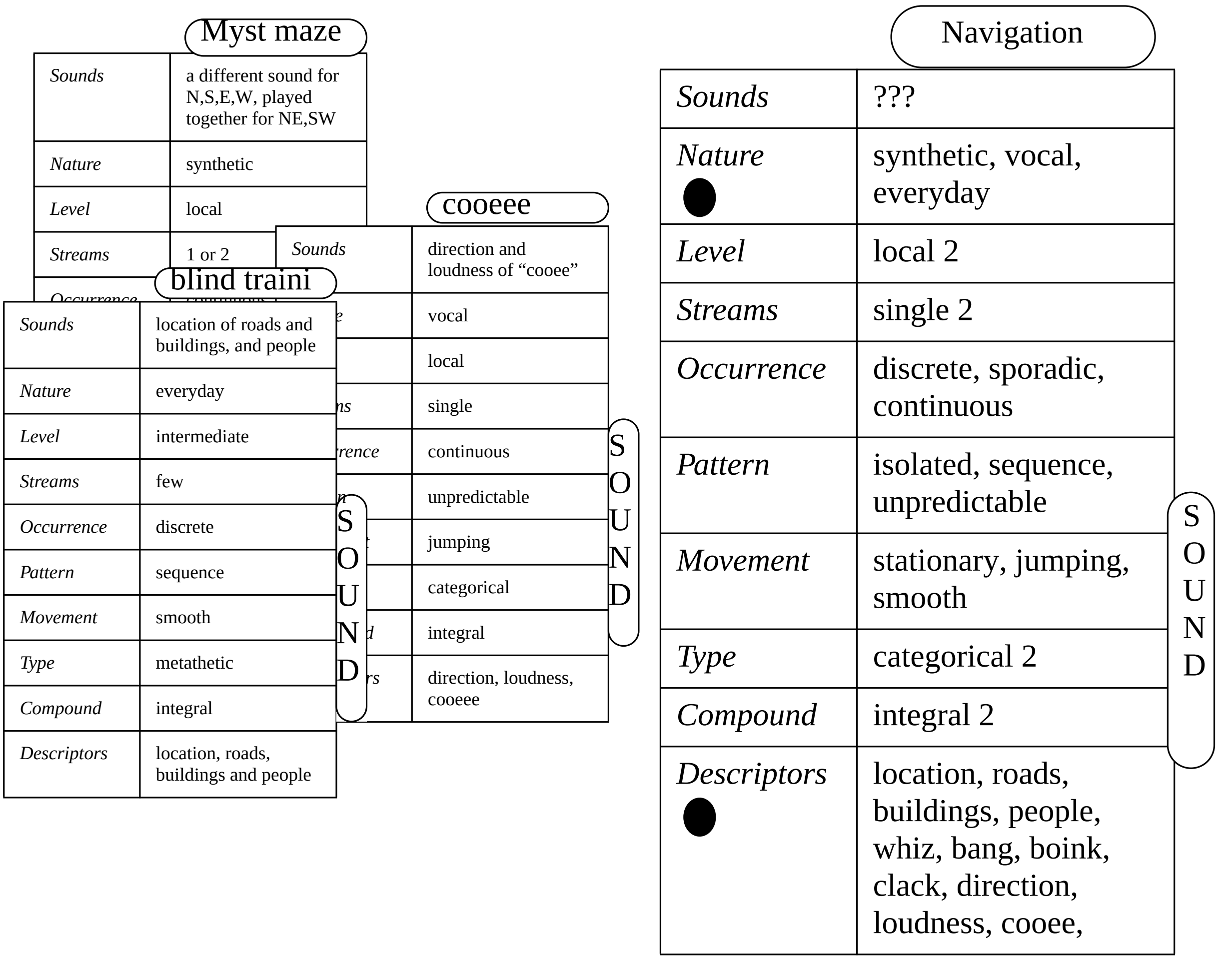


Figure 10-22: Pattern method for the Navigation scenario

The characterisation is not very regular. Of the 8 fields there were 4 without any trend, and the others only had 2/3 majorities. The weak pattern is for a local, integral, single stream, categorical sound. The nature, occurrence, pattern, and movement of the sound is different in each case, and there are many ways we could choose them. An examination of the 3 cases that were retrieved by the TaDa query can help us to understand why it is so hard to narrow down the design space for this scenario,

Scenario

Title	Myst maze
Storyteller	S.V.
STORY	the Myst game has a train maze where the sounds tell you which direction to head. whiz is north, bang is south, boink is east, and clack is west. In between directions are combinations played at the same time - so NW is a whiz and clack at the same time.
Question	which direction do we go next ?
Answers	{N,NE,E,SE,S, SW,W,NW}
Subject	direction
Sounds	whiz, whiz-boink, boink, boink-bang, bang, bang-clack, clack, clack-whiz

KEYS

TaDa Analysis

TASK	*Generic*	where is it?
	Purpose	navigate
	Mode	interactive
	Type	branching/decisio
	Style	presentation
INFO	*Reading*	conventional
	Type	nominal
	Level	local
	Range	8
	Organisation	location
DATA	*Type*	nominal
	Range	8
	Organisation	location

Figure 10-23: Myst-maze

Scenario

Title	Cooee
Storyteller	S.B.
STORY	bushwalkers call "coooeee" to find each other when they get separated, or when there are two groups who are heading to a meeting spot
Question	where is the other bushwalker?
Answers	{N,NE,E,SE,S, SW,W,NW} {near, far}
Subject	other bushwalker
Sounds	direction and loudness of the cooeee

KEYS

TaDa Analysis

TASK	*Generic*	where is it ?
	Purpose	navigate
	Mode	interactive
	Type	branching/decisio
	Style	exploration
INFO	*Reading*	direct
	Type	nominal / ordinal
	Level	local
	Range	8 / 2
	Organisation	location
DATA	*Type*	interval / interval
	Range	360degrees / 1km
	Organisation	location

Figure 10-24: Cooee

Scenario

Title	Blindness Trainin
Storyteller	S.R.

STORY

...we really try and look at is focusing on sounds that are useful and filtering out sounds that are not useful. So sounds that are useful are sounds that can help you maintain a straight line, and help you to orient so that you know where you are. Sound that are not useful in that situation are sounds that are moveable for example, such as the sounds of people talking.

KEYS

Question	where am I?
Answers	next to a road, or building
Subject	landmarks
Sounds	location of road and building sounds

TaDa Analysis

TASK

Generic	where is it
Purpose	navigate
Mode	interactive
Type	branching/decisio
Style	exploration

INFO

Reading	direct+convention
Type	ordinal
Level	global
Range	?
Organisation	location

DATA

Type	interval
Range	?
Organisation	category+location

Figure 10-25: Blindness-training

The cases were Myst-maze in Figure 10-18, Cooee in Figure 10-24, and Blindness-training in Figure 10-25. Myst-maze is about a computer game where sounds tell you which way to head to escape from a maze. Cooee is about calling to someone so they can find you in the bush. Blindness-training is about learning to hear roads and other stationary sources of sounds as landmarks for navigating the city. The sounds in each of these cases has a different point of reference. In Myst-maze the sounds are *telling* you where to head, like someone sitting next to you giving instructions. In Cooee the sounds are *calling* you from the location where you want to go, and the information comes from a spatial audio cue. In Blindness-training the sounds are *landmarks* where the spatial location is a familiar reference point that helps you find where you want to go. Telling, calling and landmarks are all ways that we can navigate by listening, The main difference is the Reading. The most {direct} navigation is by calling because most people can immediately answer the question “where are you?” by the direction the call comes from. Telling is {conventional} because the meaning of instructions like {“north”, “south”, “left”, “right” etc.} must be learnt. Landmarks are a mixture - the location of the landmark can be heard directly but the relation between the landmark and the destination must be learnt. The Hearsay principle of directness suggests the most direct design will be most effective and should be the first choice if there is an option. The options will be influenced by the target display. For example conventional telling is best for navigation in a car, because the distances are too great for calling to be heard.

10.4.4 Realisation

The GeoViewer is implemented on a Silicon Graphics workstation with 4 channels of audio built in. Spatial audio devices built on the SGI range from a simple stereo pan, to quadraphonic ambisonics, to Head Related Transfer Function (HRTF) based displays. These displays can move a CD quality sound source smoothly in real-time. This provides the capability to realise the calling design, which is the most direct of the options. A call-button has been added to the GeoViewer interface to store a location, and to activate a calling sound from that location. When you are in an interesting place you can store the coordinates by pressing the button with the right mouse. When you want some guidance back to the stored coordinate you press the call-button with the left mouse. The calling sound has a spatial cue to indicate the heading and distance from the current coordinate to the stored coordinate. At this stage the call-button has been added to the interface, and triggers a sound file. However the spatial display has not yet been fully realised.

10.5 Summary

This chapter demonstrated the TaDa approach in practice on some application scenarios. The approach is focused on the design of a display that is useful for a task and true to the data. TaDa integrates task analysis, a database of sound examples, a rule-based design aid, and interactive sound design tools, in a multifaceted system that can address practical design problems. The descriptions of the design process given in each demonstration show how the array of methods and tools fit together as a flexible and functional system.

The application scenarios were drawn from mining exploration, resource planning and climatology. A display was designed for information that was difficult to see or understand visually in each scenario. RiverAndRain is a display to help in the siting of a new water treatment works on a river system. The soundtrack was a binder that holds together a set of animations that show the influence of rainfall on chemicals in the river system. This binder enabled comparisons and correlations across variables that were in different animations and so could not be seen together. The sounds drew attention to features of the rainfall record that might otherwise have been missed, and enabled a more direct perception of relations between the variables. PopRock is a display to assist planning a mineshaft. The sounds are information about the amount of seismic activity that indicates faults and weaknesses in the surrounding geology. You can hear isolated events that may otherwise go by unnoticed, and differences in the activity of groups of events that look the same. cOcktail is a display of oxygen isotopes in 12 sea-bed drill-core sites. The soundtrack helps the listener to single out a particular time-series in the graph and watch it even when it is obscured by other traces. The display also demonstrates that a listener can switch attention between simultaneous time-series information in an auditory display. The ability to mentally switch between information elements without manual interaction may be useful in other human-computer interfaces. LostInSpace is an auditory display that helps you to navigate back to a previous location in a 3D visualisation of geological structures. It was found that there were many different ways that sounds could assist in navigation, and 3 techniques with differing degrees of directness were identified - calling, telling and landmarks.

Experience with the multimedia interfaces that were built in these demonstrations showed that the sounds can provide the required information, and provide information that is difficult to obtain visually. The demonstrations have stimulated discussions at CSIRO about other possible uses of sounds in monitoring traffic, data mining and software debugging - indicating an appreciation that sounds can provide useful information that is difficult to obtain in a purely visual interface.

10.6 Further work

The design process for the LostInSpace scenario identified calling, telling and landmarks as ways that sounds could support navigation in a 3D space. If all of these options are available, which should be taken? An evaluation and comparison of the designs could be made in terms of usability, learnability, expressiveness, effectiveness and other HCI factors measured by task times, error rates, etc. However we must beware of “the curse of empiricism” which Cleveland says has been the cause of a great waste of effort in graphical display [Cleveland W.S. (1985)]. The experimental comparison of techniques without a concomitant development of predictive principles is not very helpful to design practice, because without principles all design decisions must be resolved by an expensive experiment. The TaDa approach is an example of a principled framework that may help auditory display to progress by avoiding the pitfalls of purely comparative empirical approaches. TaDa is open ended so that results and observations made from experiments can be fed-back as new Hearsay principles, and new EarBenders cases. For example we could carry out an experiment to validate the principle of directness that caused us to choose the calling display over the telling and landmark designs. The observations from this experiment can then reinforce or improve the principle.

11 • Summary

Although there is a well developed practice and culture of movie sound, computer applications are a new challenge because of the types of information to be conveyed and the interactivity between the user and the sounds. This thesis develops an approach to *the design of sounds to support information processing activities.* The design approach is supported by a system of case-based and rule-based methods and tools. This hybrid system addresses key design issues in a flexible manner necessary for real world design practice. The approach, methods and tools are demonstrated in a series of scenarios that show how sounds can provide information that is difficult to obtain visually, and how they can provide extra affordances in the human-computer interface.

The following sections summarise the TaDa approach, the methods and tools that were developed to support it, the various demonstration scenarios, and suggestions of further work that arise out this thesis.

11.1 The TaDa approach

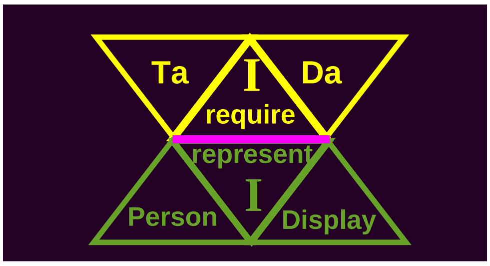


Figure 11-1: TaDa diagram

The TaDa approach, described in Chapter 3, answers the call for a principled approach to auditory display made by pioneers in the field. The approach identifies key design issues of usefulness in a task and faithfulness to the data. These issues inform the specification of information that is to appear in the display. TaDa is summarised by an arrangement of triangular tiles shown in Figure 11-1. The Task analysis (Ta) and Data characterisation (Da) tiles sit upon and specify the information requirements tile (Irequire). The Irequire tile rests upon the Information representation (Irepresent) tile. The Irepresent tile is supported by the (Person) and (Display) tiles. These tiles point to the need to consider how people hear information in sounds, and the way the display device influences the range of available sounds.

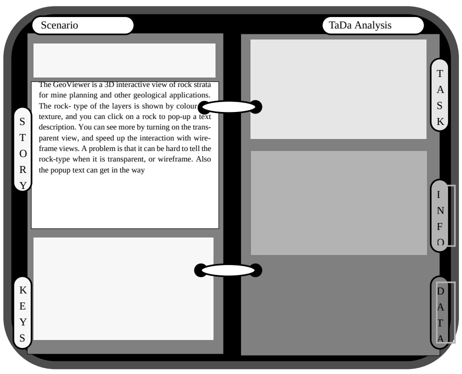


Figure 11-2: TaDa analysis

The information requirements are obtained by a TaDa scenario analysis, described in Chapter 4. The scenario analysis begins with a short story describing the problem. The scenario is analysed by recasting it as a Question, some Answers, and a Subject. These keys are the bridge between the scenario and the analysis of the information required to support the scenario. The analysis is an amalgam of task analysis methods from HCI and data characterisation methods from scientific visualisation.

11.2 Methods and tools

The design of a representation to meet the TaDa information requirements occurs at the boundary between the Irequire and Irepresent tiles. Design is typically messy - inductive, lateral, top-down, bottom-up and iterative in nature, involving issues of semiotics, style, and culture that are not amenable to straight-forward deductive methods. The TaDa approach is supported by a system of case-based and rule-based methods and tools, shown in Figure 11-3. These methods have different strengths that make them helpful in different facets of design. The system is principled yet flexible to accommodate the complexity of real world design practice.

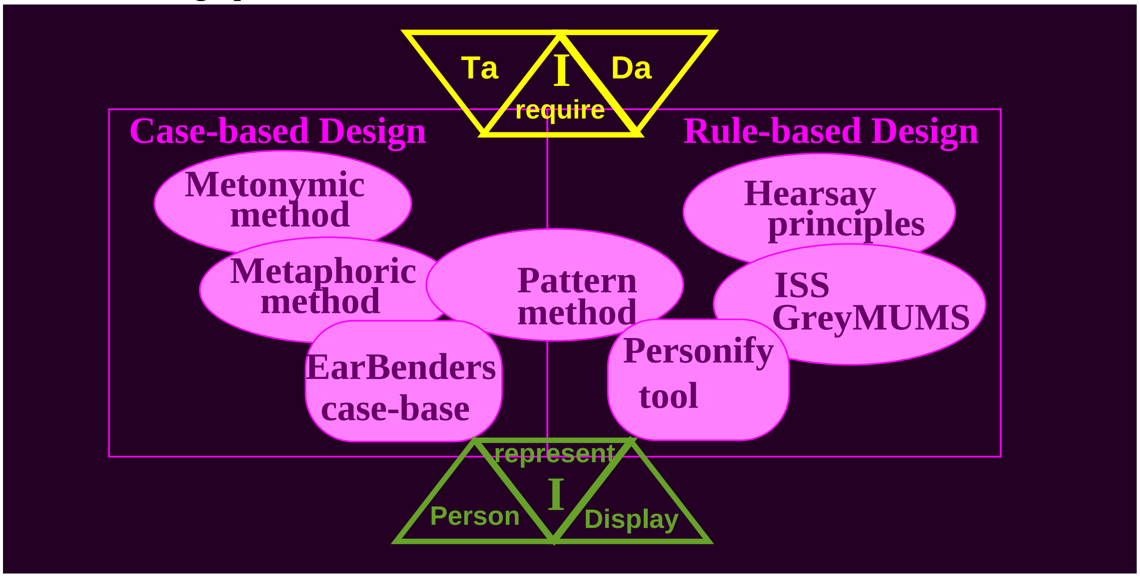


Figure 11-3: Methods and tools to support the TaDa approach

11.2.1 EarBenders case-based methods and tools

The EarBenders case-based method, developed in Chapter 5, helps to identify a familiar auditory schema based on everyday experiences with sounds. A schema can improve the comprehension and discrimination of elements in the display, and can connect the semantics of the sounds in the display with the application domain. The methods rely on the EarBenders case-base of 200 stories about everyday listening experiences. The cases can be retrieved from a database by a similarity match on TaDa analysis fields. Three case-based design methods were developed using EarBenders -called metonymic, metaphoric and pattern.

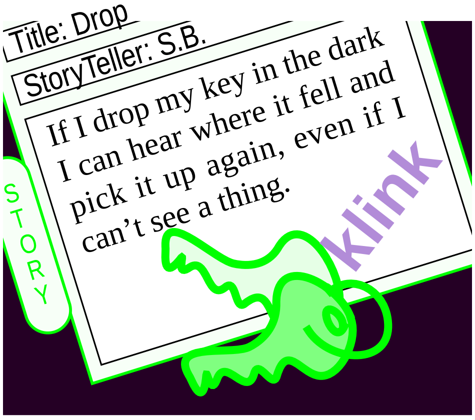


Figure 11-4: EarBenders story

The metonymic method uses sounds that are a part of the design scenario. A metonym connotes the whole by a part of that whole - for example the sound “woof” may be a metonym for a dog. The sounds in a medical scenario might be heartbeats, breathing, or instrument beeps. Any sounds mentioned in a design scenario may be the beginnings of a metonymic palette. The metonymic method is a search of EarBenders for stories that include some part of the design scenario. The sounds in these stories can connote the application domain at the same time as providing perceptually structured information.

A metaphoric design can help when sounds are not a natural part of the design scenario, which is often the case in computer-based applications. A metaphor expresses the unfamiliar in terms of the familiar, for example a tree may be a metaphor for a filing system. The metaphoric method retrieves stories that have TaDa information structure similar to the design scenario. These stories are a source of familiar sounds that can represent information about stock prices or internet traffic or other things that don't normally make sounds.

The pattern method identifies features shared by different solutions to the same problem that may capture patterns in the mapping between domains. A pattern is a regularity in the mapping between the problem and solution domains. The pattern method was enabled by appending a description of auditory characteristics to each EarBenders case. When a TaDa query retrieves cases with similar information structure it also retrieves the auditory characterisation of each case. The characterisations are scanned for regularities identified by the majority trend in each auditory characteristic. The weighting of each trend is calculated from the number of cases which follow the trend. The result is a synthesis of the auditory characteristics of the retrieved cases. Regularities in the auditory characteristics of the retrieved cases may capture patterns in the mapping between information and sounds. The design of an auditory display from these regularities may support the information requirements of the design scenario.

11.2.2 Hearsay rule-based methods and tools

The Hearsay rule-based method, developed in Chapter 6, integrates broad principles of information design from graphic display with psychoacoustic observations relevant to auditory display. These principles encapsulate expert knowledge in a manner that can be communicated, applied and developed by other designers. The Hearsay principles were tested in the design of a display for Bly's 'dirt and gold' soil classification scenario. The resulting display allows a listener to immediately answer questions about the data set that are difficult to answer from a visual representation, demonstrating that the principles are of benefit in practice.

Principles, guidelines and rules can be unwieldy in practice because they must be looked up or learnt. The Information-Sound Space, developed in Chapter 7, moulds the Hearsay auditory principles into a form that allows a designer to think in terms of spatial relations such as lines and planes, rather than rules. The ISS is structured so that a general range of information relations can be represented by auditory relations. This tool allows the designer to think about the mapping from information relations to auditory relations as though it were a 3 dimensional object.

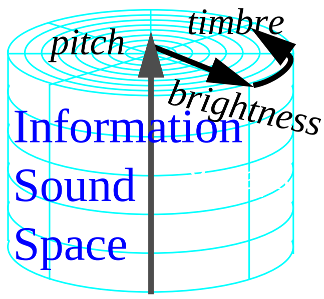


Figure 11-5: ISS blueprint

The GreyMUMS ISS, built in Chapter 8, is a proof-of-concept ISS built from musical instrument samples. The construction was carried out in four stages - the pedestal, the frame, the grating and the plasticine. The raw material for the construction was the MUMS palette of musical instrument samples, which is a reference palette for other researchers in the area of auditory perception. The resulting sculpture has the property that there is a relationship between distance and the strength of perceptual grouping between points in the space. A vertical line is a pitch scale, and may be used to represent continuous data. A radial line is a scale of equal brightness increments for a timbre, and may also be used for continuous data. A circle of constant radius is a contour of constant brightness across the range of timbres which can be used to represent categorical data. These properties are a rich area for further experiment with data mappings and simultaneous and sequential presentations of the selections. The GamutExplorer is an interface that allows you to look at the GreyMUMS space as though it were a 3D coloured solid. You can hear points in the space by picking them from the sound solid with the mouse.

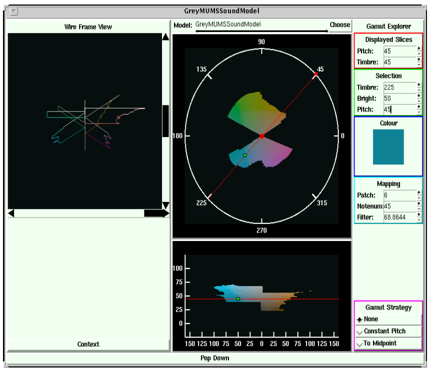


Figure 11-6: GamutExplorer

Personify is a constrained tool, developed in Chapter 9, for selecting representational mappings from the GreyMUMS solid. Personify integrates the TaDa analysis with a direct manipulation interface similar to interfaces found in advanced musical tools. There are 2 main parts that mirror the organisation of the TaDa tiles - the requirements part and the representation part. The requirements part is made up of the TaDa Panel that captures the information requirements, and a rule-base that automatically configures a default auditory representation according to Hearsay principles. The representation part is made up of tailoring tools for fitting the representation to the display, and a device model that maps coordinates from the GreyMUMS space to sounds on the display device. The tailoring tools allow you to select representations by drawing paths through ISS space, in device independent terms.

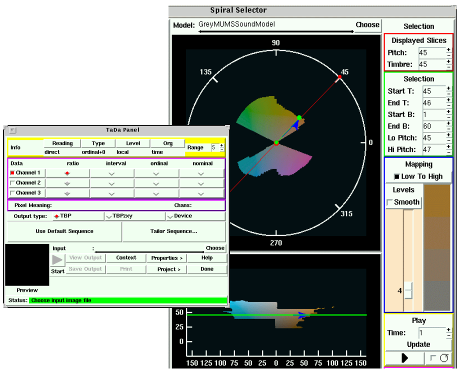


Figure 11-7: Personify

11.3 Demonstrations

The TaDa approach, methods and tools were demonstrated on some application scenarios drawn from mining exploration, resource planning and climatology. A scenario describing each problem was recast as a question that the sounds in the display need to answer to support the information processing activity. You can hear some of the demonstrations at the ICAD web site (http://www.santafe.edu/~icad/ICAD96/proc96/barrass.htm).

11.3.1 GeoViewer

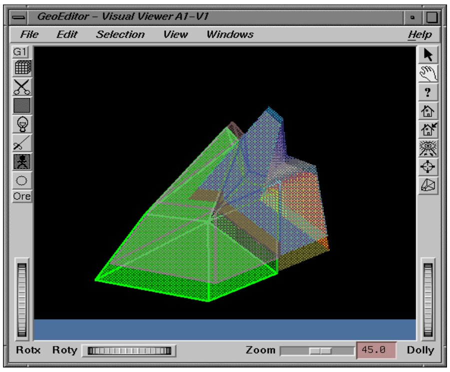


Figure 11-8: GeoViewer

The GeoViewer, in Chapter 5, is a 3D interactive view of rock strata for mine planning and other geological applications. A problem is that it can be hard to tell what type of rock you are looking at in the various transparent and wireframe modes, and popup text descriptions can obstruct the view. The question is "what rock is this?" The case-based methods were used to design 3 different palettes with different semantic connotations - impacts, musical sounds, and spoken words. You can answer the question by tapping on a rock with the mouse. The sounds allow the front most surface at an intersection of strata to be disambiguated by tapping there, saving on a distracting and computationally expensive change of viewpoint operation. A development could provide information about the number and material of hidden overlapping layers.

11.3.2 Gold

The Gold demonstration, in Chapter 6, applies the Hearsay principles to Bly's multidimensional sonification challenge ... *Can you find the gold ? It is hypothesised that six different aspects of the land in which gold may be found are determinative of whether or not gold is there.* The display allows an immediate answer to the question for groups of data as well as individual elements. Regions of gold are heard as distinct masses of higher, brighter material. Even when the data is only 5% gold the gold samples pop-out from the background of dirt sounds. The ability to hear low proportions of outliers may be useful in data-mining applications where interesting elements are hard to display visually due to multidimensionality, and hard to see because they are rare. If the data is presented in an appropriate form then a human analyst may be able to perceive information that a programmed algorithm cannot detect.

11.3.3 LandCover

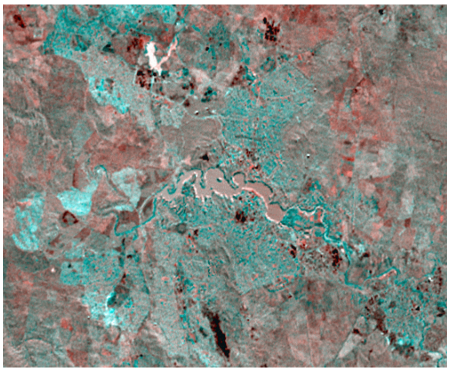


Figure 11-9: LandCover

The LandCover demonstration, in Chapter 9, is a scenario about resource monitoring by satellite data. The level of vegetation can be seen by coloured regions in a satellite image, but the colour of small regions is hard to make out due to limits in hue resolution of the eye. The design question is "what is the seasonal variation in landcover here ?" An auditory display was designed to provide this local information. The design was specified with the TaDa Panel and interactively tailored to the GreyMUMS display with the Personify tool. The audiovisual display shows global information about vegetation as a colour image, and local information by tapping on individual elements with a mouse to hear a sound.

11.3.4 RiverAndRain

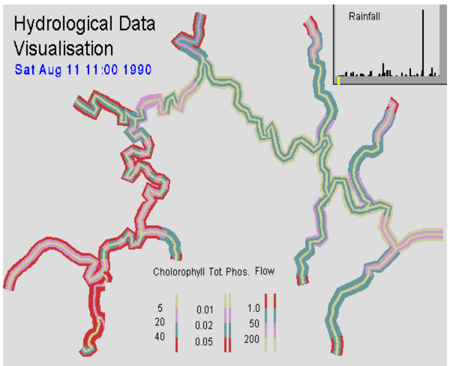


Figure 11-10: RiverAndRain

RiverAndRain, in Chapter 10, is a display to help in the siting of a new water treatment works on a river system. The relations between the rainfall and the level of chemicals in the river were difficult to see in the original visualisation because there was a need to continually shift visual attention between different elements in the display. The question is “what is the rainfall now?” A soundtrack was designed to provide the rainfall information and allow visual attention to remain fixed on changes in the river. With use the soundtrack becomes a binder that enables the perception of relations between variables that appear in different animations and so can’t be seen together. The sounds draw attention to features of the rainfall record that might otherwise be missed, and allow a more direct perception of relations between the rain and river variables.

11.3.5 PopRock

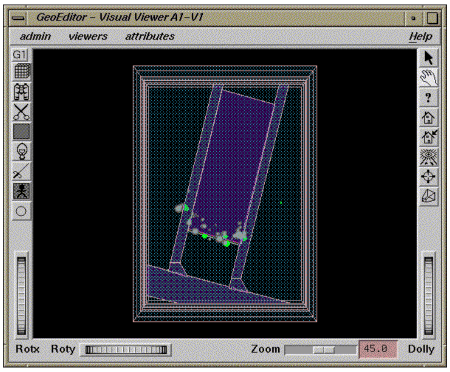


Figure 11-11: PopRock

PopRock, in Chapter 10, is a display to assist planning a mineshaft. The visualisation of seismic events as coloured spheres provides information about the levels of seismic activity that may indicate faults and weaknesses in the surrounding rocks. However the events happen so quickly that isolated events are easy to miss, and the spheres overlap making it difficult to perceive events that occur at the same time and place. The question is “how much seismic activity is there at the moment?”. The case-based method was used to select a sound that is semantically linked to the application domain - a small explosion like “pop”. You can hear isolated events that may otherwise go by unnoticed, and differences in the activity of groups of events that look similar.

11.3.6 cOcktail

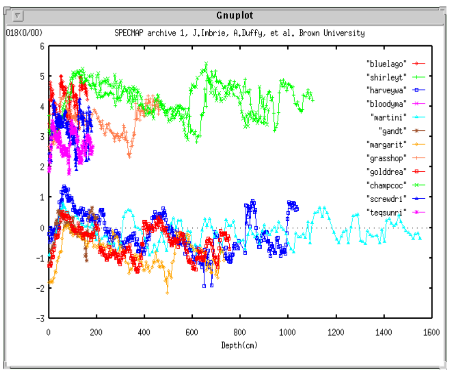


Figure 11-12: cOcktail

cOcktail, in Chapter 10, is a display of oxygen isotopes in 12 sea-bed drill-core sites. The individual records are difficult to track as they overlap and intertwine in the visual display. The question is “what is the level at site X?” The soundtrack is designed to help the listener to single out a particular record in the graph and watch it even when it is obscured by other traces. The listener is able to mentally switch attention between co-occurring time-series information in this auditory display. The ability to make a mental, rather than manual, switch between elements may have consequence in other human-computer interfaces.

11.3.7 LostInSpace

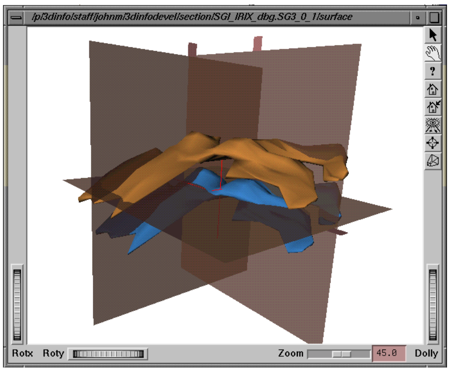


Figure 11-13: LostInSpace

LostInSpace, in Chapter 10, is an auditory display that helps you to navigate in a 3D visualisation of geological structures. These structures are irregular and it can be hard to remember where you have been or find a way back there. The question is “where was I?”. The case-based method identified 3 auditory techniques with differing degrees of directness that could provide a schema for navigation by ear - calling, telling and landmarks. Calling is when someone calls to you from somewhere and you can head in that direction to find them. Telling is when someone sitting next to you gives you directions, as in when driving a car. Landmarks is when you can hear a familiar sound, such as the clock chiming in the village square, and can use it to orient yourself and work out where to head.

11.4 Further work

This thesis develops an approach that tackles a range of important issues such as usefulness, directness, learnability, transportability. Other issues that have been touched-on but require more attention are aesthetic connotations and ergonomics of sounds.

The demonstrations show that the TaDa approach can help in the design of an auditory information display. However the claim that the sounds improve an interface needs to be backed by evidence of some sort. In the HCI community usability measurement has been developed as a way to quantify and compare interface designs. A usability method could be used to test whether the sounds are of benefit. by comparing task performance in auditory and non-auditory conditions. Further work might also integrate usability measurements into the design process as a way to iteratively improve an existing design.

Many real-world problems are more complex than the demonstration scenarios tried so far. Further work might involve development of the TaDa approach against scenarios which involve more than one question, more than a one kind of answer, and more than one type of subject. This might involve further development of the TaDa requirements analysis.

The methods and tools that support the TaDa approach could also be developed further. The EarBenders case-base could be expanded by collecting more stories from the internet, or by making it publicly available as a design resource. This expansion could allow scope for culturally specific schemas, and provide more relevant schemas for a wider range of design scenarios.

New configurations of the Information-Perception Space could allow perceptual structures that go beyond those that can be achieved with the current structure that is based on a colour space with a single zero. These configurations would also require the development of new mapping schemes and tailoring tools in Personify. If a particular configuration proves broadly useful then it would be worthwhile to experimentally verify its information properties across a sample of individuals.

GreyMUMS is a proof-of-concept of an Information-Sound Space. However the palette of musical sounds constrains the design of a schema and it would be good to build other ISS's based on everyday sounds, or synthetic sounds or some other connotative palette. This is not a trivial exercise because the psychoacoustic data required is not currently available. Hence there would also be a need to develop fast, robust techniques and readily available tools for measuring and building an ISS.

The perceptual scaling of the GreyMUMS sound space could be improved on in a number of ways. The scaling of each axis in isolation is only a first approximation to a uniform space because it does not take into account interactions between them. The problem of harmonicity might be addressed by including Shephard's pitch helix into the Pitch axis. Variation in loudness could be normalised by a loudness model. It would be good to allow continuity around the timbre circle by some form of continuous synthesis. This may allow the construction of a sound space that is perceptually uniform in all directions.

Bibliography

Arons B. (1994) Efficient Listening with Two Ears: Dichotic Time Compression and Spatialization, *Proceedings of the Second International Conference on Auditory Display ICAD `94*, Santa Fe, New Mexico.

Alty J. (1995) *Multimedia Interface Design Tutorial Notes*, Australian Conference on Computer-Human Interaction OZCHI'95, University of Wollongong.

Bailey W.J. (1991) *Acoustic Behaviour of Insects: an Evolutionary Perspective*, Chapman and Hall, London, 1991.

Ballass J.A. (1994) Effect of Event Variations and Sound Duration on Identification of Everyday Sound, *Proceedings of the Second International Conference on Auditory Display ICAD `94*, Santa Fe, New Mexico.

Ballass J.A. (1994) Delivery of Information Through Sound, in Kramer G. (ed) *Auditory Display : Sonification, Audification and Auditory Interfaces,* SFI Studies in the Sciences of Complexity, Proceedings Volume XVIII, Addison-Wesley Publishing Company, Reading, MA, U.S.A.

Barnard P. (1991) Bridging Between Basic Theories and the Artifacts of Human-Computer Interaction, in Carroll J.M. (ed), *Designing Interaction : Psychology at the Human-Computer Interface*, Cambridge University Press.

Barrass S. (1994a) A Naturally Ordered Geometric Model of Sound Inspired by Colour Theory, *Proceedings of Synaesthetica `94*, Australian Centre for the Arts and Technology.

Barrass S. (1994b) A Perceptual Framework for the Auditory Display of Scientific Data. *Proceedings of the Second International Conference on Auditory Display ICAD `94*, Santa Fe Institute, New Mexico.

Barrass S. and Robertson P.K. (1995) Data Exploration with Sound using a Perceptually Linearized Sound Space, *IS&T/SPIE Symposium on Electronic Imaging:Science and Technology,* San Jose, California.

Barrass S. (1995) Personify: a Toolkit for Perceptually Meaningful Sonification, *Proceedings of Australian Computer Music Conference ACMA'95,* Australian Computer Music Association, Inc.

Barrass S. (1996a) Sculpting a Sound Space with Information Properties, *Organised Sound*, 1, 2, Cambridge University Press.

Barrass S. (1996b) EarBenders: Using Stories About Listening to Design Auditory Interfaces, *Proceedings of the First Asia-Pacific Conference on Human Computer Interaction APCHI'96,* Information Technology Institute, Singapore.

Barrass S. (1996c) TaDa! Demonstrations of Auditory Information Design, *Proceedings of the Third International Conference on Auditory Display ICAD'96,* Xerox PARC, Palo Alto, California.

Barrass S. (1997) Some Golden Rules for Designing Auditory Displays, in Vercoe B. and Boulanger R. (eds) Csound Textbook, MIT Press, in press.

Barthes R. (1977) *Image, Music, Text*, Hill and Wang, New York.

Bateson G. (1972) *Steps to an Ecology of Mind*, Ballantine Books, New York, 1972

Beaudouin-Lafon M. and Gaver W.W. (1994) ENO: Synthesizing Structured Sound Spaces, *Proceedings of User Interface Software and Technology UIST `94*, ACM Press.

Beck M. (1996) Micro-narratives in Sound Design: Context, Character and Caricature in Waveform Manipulation, in *Proceeedings of the Third International Conference on Auditory Display ICAD'96*, Palo Alto.

Bergman L.D. Rogowitz B.E. and Treinish L.A. (1995) A Rule-based Tool for Assisting Colormap Selection, *Proceedings of IEEE Visualization '95*, Atlanta, Georgia.

Bertin J. (1981) *Graphics and Graphic Information Processing*, Walter de Gruyter.

Blattner M. Sumikawa D. and Greenberg R. (1989) Earcons and Icons: Their Structure and Common Design Principles, *Human Computer Interaction*, 4, 1, Lawrence Erlbaum Associates, London.

Blattner M.M. Papp A.L. and Glinert E.P. (1994) Sonic Enhancement of Two-Dimensional Graphics Displays, in Kramer G. (ed) *Auditory Display : Sonification, Audification and Auditory Interfaces*, SFI Studies in the Sciences of Complexity, Proceedings Volume XVIII, Addison-Wesley Publishing Company, Reading, MA, U.S.A.

Bly S. (1994) Multivariate data analysis, in Kramer G. (ed) *Auditory Display : Sonification, Audification and Auditory Interfaces*, SFI Studies in the Sciences of Complexity, Proceedings Volume XVIII, Addison-Wesley Publishing Company, Reading, MA, U.S.A.

Bock D.S. (1994) ADSL: An Auditory Domain Specification Language for Program Auralisation, in *Proceedings of the International Conference on Auditory Display ICAD '94*, Santa Fe, New Mexico.

Bone D. (1993) Adaptive Color-printer Modelling Using Regularized Linear Splines, *Proceedings of Device Independent Color Imaging*, IS&T/SPIE Symposium on Electronic Imaging Science and Technology '93.

Booty W. and Wong I.W.S. (1994) Water Quality Modelling Within the RAISON Expert System, *Journal of Biological Systems*, 2, 4, pp 453-466.

Bregman A.S. (1990) *Auditory Scene Analysis*, MIT Press.

Brewster S.A. Wright P.C. and Edwards A.D.N. (1994) A Detailed Investigation into the Effectiveness of Earcons, in Kramer G. (ed) *Auditory Display : Sonification, Audification and Auditory Interfaces*, SFI Studies in the Sciences of Complexity, Proceedings Volume XVIII, Addison-Wesley Publishing Company, Reading, MA, U.S.A.

Brewster S.A. (1994) Providing a Structured Method for Integrating Non-Speech Audio into Human-Computer Interfaces, Ph.D. thesis, University of York.

Cain W.S. and Marks L.E. (1971) *Stimulus and Sensation*, Little Brown and Company, Boston.

Caivano J.L. (1994) Color and Sound: Physical and Psychophysical Relations, *COLOR Research and Applications*, 19, 2, pp 126-132.

Caivano J.L. (1994) Towards an Order System for Visual Texture, *Languages of Design*, 2, 1, March, 1994.

Carlow International Inc. (1992) *Human-Computer Interface Guidelines*, http://groucho.gsfc.nasa.gov/Code_520/Code_522/Documents/HCI_Guidelines/.

Carterette E.C. and Friedman M.P. (ed) (1974) *Historical and Philosophical Roots of Perception, Handbook of Perception*, 1, Academic Press, New York.

Casner S. (1992) A Task Analytic Approach to the Automated Design of Graphic Presentations, *ACM Transactions on Graphics*, 10, pp 111-151.

Chandler D.C. (1997) Semiotics for Beginners, http://www.aber.ac.uk/~dgc/semiotic.html.

Cleveland W.S. (1985) *The Elements of Graphing Data*, Wadsworth Advanced Books and Software, Monterey, California.

Crist E.P. and Cicone R.C. (1994) Application of the Tasselled Cap Concept to Simulated Thematic Mapper Data, *Photogramm. Eng. and Remote Sensing*, 50, pp 343-52.

Deatherage B.H. (1972) Auditory and Other Sensory Forms of Information Presentation, in Van Cott H.P. and Kinkade R.G. (eds) (1972) *Human Engineering Guide to Equipment Design (Revised Edition)*, McGraw-Hill.

Dougherty T. (1996) What Does Pink Sound Like? Designing the Audio Interface for the TalOS, *in Proceedings of the Third International Conference on Auditory Display ICAD'96*, Palo Alto.

Dowling W.J. (1991) Pitch Structure, in Howell P. West R. and Cross I. (eds) *Representing Musical Structure*, Academic Press.

Draper S.W. (1995) More Than We Think: Facing up to the Plurality of Goals, Methods, Needs and Resources in HCI, *Proceedings of the Australian Conference on Human-Computer Interaction OZCHI'95*, University of Wollongong.

Ehresman D. & Wessel D. (1978) *Perception of Timbral Analogies*, IRCAM, Centre Georges Pompidou.

Erickson T. (1996) Design as Storytelling, *ACM Interactions*, July 1996, ACM Publications.

Fletcher N.H and Rossing T.D. (1991) *The Physics of Musical Instruments*, Springer-Verlag, New York.

Flowers J.H. Buhman D.C. and Turnage K.D. (1996) Data Sonification from the Desktop: Should Sound be part of Standard Data Analysis Software, in *Proceeedings of the Third International Conference on Auditory Display ICAD'96*, Palo Alto.

Frysinger S.P. (1990) Applied Research in Auditory Data Representation, *SPIE Extracting Meaning from Complex Data: Processing, Display Interaction*, 1259.

Gaver W.W. (1986) Auditory Icons, *Human-Computer Interaction*, 2, 2, Lawrence Erlbaum Associates, London.

Gaver W.W. Smith R.B. and O'Shea T. (1991) Effective Sounds in Complex Systems : The Arkola Simulation, *Proceedings of CHI '91 Human Factors in Computing Systems*, ACM Press.

Gaver W.W. (1993) Synthesizing Auditory Icons, *Proceedings of InterCHI `93*, ACM SigCHI.

Gaver W.W. (1994) Using and Creating Auditory Icons, in Kramer G. (ed) (1994) *Auditory Display : Sonification, Audification and Auditory Interfaces*, SFI Studies in the Sciences of Complexity, Proceedings Volume XVIII, Addison-Wesley Publishing Company, Reading, MA, U.S.A.

Geissler H.-G. and Petzold P. (eds) (1980) *Psychological Judgement and the Process of Perception*, North-Holland Publishing Company, Amsterdam.

Getty D.J. and Howard Jr. J.H. (eds) (1981) *Auditory and Visual Pattern Recognition*, Lawrence Erlbaum Assosciates, New Jersey.

Gibson J.J. (1966) *The Senses Considered as Perceptual Systems*, Houghton Mifflin Company, Boston.

Goldstein J. Roth S.F. Kolojejchick J. and Mattis J. (1994) A Framework for Knowledge-based Interactive Data Exploration, *Journal of Visual Languages and Computing*, 5, 4, December 1994, Academic Press Limited.

Graetz D. Fisher R. and Wilson M. (1992) Looking Back: the Changing Face of the Australian Continent, 1972-1992, CSIRO Office of Space Science and Applications.

Grayson J. (ed) (1975) *Sound Sculpture*, A.R.C. Publications, Vancouver.

Grey J.M. (1975) *Exploration of Musical Timbre*, Phd. Thesis, Report No. STAN-M-2, CCRMA Dept. of Music, Stanford University.

Handel S. (1990) *Listening: An Introduction to the Perception of Acoustic Events*. MIT Press.

Hayward C. (1994) Listening to the Earth Sing. in Kramer G. (ed) (1994) *Auditory Display : Sonification, Audification and Auditory Interfaces*, SFI Studies in the Sciences of Complexity, Proceedings Volume XVIII, Addison-Wesley Publishing Company, Reading, MA, U.S.A.

Horry Y. (1994) A Graphical User Interface for MIDI Signal Generation and Sound Synthesis, *Proceedings of the International Computer Music Conference ICMC `94*, Aarhus.

Howell P. Cross I. and West R. (eds) (1991) *Representing Musical Structure*, Academic Press.

Hunt R.G.W. (1987) *Measuring Colour*, Ellis Horwood Limited, Chichester.

Ignatius E. Senay H. and Favre J. (1994) Visualization Assistant, *Journal of Visual Languages and Computing*, 5, 4, December 1994, Academic Press Limited.

Imbrie J. McIntyre A. and Mix A.C. (1990) *SPECMAP Archive #1*. IGBP PAGES/World Data Center-A for Paleoclimatology Data Contribution Series # 90-001. NOAA/NGDC Paleoclimatology Program (ftp://ngdc1.ngdc.noaa.gov/paleo/paleocean/specmap/specmap1/), Boulder CO, U.S.A.

Jackson J.A. and Francioni J.M. (1994) Synchronization of Visual and Aural Parrallel Program Performance Data, in Kramer G. (ed) (1994) *Auditory Display : Sonification, Audification and Auditory Interfaces*, SFI Studies in the Sciences of Complexity, Proceedings Volume XVIII, Addison-Wesley Publishing Company, Reading, MA, U.S.A.

Jameson D.H. (1994) The Run-Time Components of Sonnet, *Proceedings of the Second International Conference on Auditory Display ICAD `94*, Santa Fe, New Mexico.

Kaplan B. and MacCuish J. (1993) *User Documentation of the Sonify module for AVS*, Center for Innovative Computer Applications (CICA), Indiana University.

Kaplan B. and Goodsen J. (1995) *Task Analysis Method and Process*, http://www.jagunet.com/dalmation/techpapers.html.

Kendall G.S. (1991) Visualization by Ear: Auditory Imagery for Scientific Visualization and Virtual Reality, *Computer Music Journal*, 15, 4, MIT Press.

Klausen T. and Aboulafia A. (1995) An Empirical Study of Professional Software Designers' Use of Scenarios, *Australian Conference on Computer Human Interaction OZCHI '95*, CHISIG publications.

Kramer G. and Ellison S. (1991) Audification: The Use of Sound to Display Multivariate Data, *Proceedings of the International Computer Music Conference ICMC `91*, Montreal.

Kramer G. (1994a) Some Organizing Principles for Representing Data With Sound, in Kramer G. (ed) (1994) *Auditory Display : Sonification, Audification and Auditory Interfaces*, SFI Studies in the Sciences of Complexity, Proceedings Volume XVIII, Addison-Wesley Publishing Company, Reading, MA, U.S.A.

Kramer G. (ed) (1994b) *Auditory Display : Sonification, Audification and Auditory Interfaces*, SFI Studies in the Sciences of Complexity, Proceedings Volume XVIII, Addison-Wesley Publishing Company, Reading, MA, U.S.A., 1994.

Laurel B. (ed) (1991) *The Art of Human Computer Interface Design*, Addison-Wesley Publishing Company Inc.

Lehrman P.D. and Tully T. (1993) *MIDI for the Professional*, Amsco Publications.

Lewis C. and Rieman J. (1994) *Task Centered User Interface Design : a Practical Introduction*, http://www.cis.ohio-state.edu/~perlman/uidesign.html.

Lucas P.A (1994) An Evaluation of the Communicative Ability of Auditory Icons and Earcons,*Proceedings of the Second International Conference on Auditory Display ICAD `94*, Santa Fe, New Mexico.

MacKinlay J. (1986) Automating the Design of Graphical Presentations of Relational Information, *ACM Transactions on Graphics*, 5, 2, April 1986.

Madhyastha T.M and Reed D.A. (1992) A Framework for Sonification Design, in Kramer G. (ed) (1994) *Auditory Display : Sonification, Audification and Auditory Interfaces*, SFI Studies in the Sciences of Complexity, Proceedings Volume XVIII, Addison-Wesley Publishing Company, Reading, MA, U.S.A.

Maher M.L. Balachandran M.B. and Zhang D.M. (1995) *Case-based Reasoning in Design*, Lawrence Erlbaum Assosciates Inc.

Mather A.P. Boradman D.B. and Khandelwal (1994) LSL: A Specification Language for Program Auralization, *Proceedings of the Second International Conference on Auditory Display ICAD `94*, Santa Fe Institute, New Mexico.

May J. Scott S. and Barnard P. (1996) *Designing Displays: a Psychological Guide*, Medical Research Council Applied Psychology Unit.

McAdams M. (1995) Information Design and the New Media, *ACM Interactions*, 2, 4, October 1995.

McCormick E.J. and Sanders M.S. (1983) *Human Factors in Engineering and Design*, McGraw-Hill Book Co, Singapore.

Mereu, S.W. (1995) *Improving Depth Perception in 3D Interfaces with Sound*, Unpublished Masters Thesis, University of Waterloo.

Miller G.A. (1971) The Magical Number Seven, Plus or Minus Two: Some Limits on Our Capacity for Processing Information", in Cain W.S. and Marks L.E. (eds) *Stimulus and Sensation: Readings in Sensory Psychology,* Little Brown and Company, Boston.

Moreno E. (1994) A Visual Model for Embedded Chroma Spaces, *Proceedings of the International Computer Music Conference ICMC `94*, Aarhus.

Morris C.W. (1955) Foundations of the Theory of Signs, in Neurath O. Carnap R. and Morris C. (1955) *International Encyclopedia of Unified Science*, 1, The University of Chicago Press.

Moyes J. (1995) *Putting Icons in Context: the Influence of Contextual Information on the Usability of Icons,* Ph.D. Thesis, Department of Computing, University of Glasgow.

Mynatt E.D. (1994) Designing with Auditory Icons, *Proceedings of the Second International Conference on Auditory Display ICAD `94*, Santa Fe Institute, New Mexico.

Norman D.A. (1991) Cognitive Artifacts, in Carroll J.M. (ed), *Designing Interaction : Psychology at the Human-Computer Interface*, Cambridge University Press.

Opolko F. and Wapnick J. (1995) *McGill University Master Samples*, http://lecaine.music.mcgill.ca/MUMS/Html/Mums.html.

Ousterhout J.K. (1994) *Tcl and the Tk Toolkit,* Addison-Wesley, Reading, Mass.

Pachella R.G. Somers P. and Hardzinski M. (1981) A Psychophysical Approach to Dimensional Integrity. In Gett D.J. and Howard Jr. J.H. (eds) *Auditory and Visual Pattern Recognition*, Lawrence Erlbaum Assocs, Hillsdale.

Padgham C. (1986) The Scaling of the Timbre of the Pipe Organ, *Acustica*, 60, pp 189-204.

Papp A.L. and Blattner M.M. (1994) A Centralized Audio Presentation System, *Proceedings of the Second International Conference on Auditory Display ICAD `94*, Santa Fe Institute, New Mexico.

Patterson R.D. (1982) *Guidelines for Auditory Warning Systems on Civil Aircraft*, Paper No. 82017, Civil Aviation Authority, London.

Plomp R. (1976) *Aspects of Tone Sensation*, Academic Press, London.

Pollack I. and Ficks L. (1954) Information of Elementary Multidimensional Auditory Displays, *The Journal of the Acoustical Society of America*, 26, 2, March 1954.

Portigal S. and Carey T. (1994) Auralization of Document Structure, *Proceedings of the Second International Conference on Auditory Display ICAD `94,* Santa Fe Institute, New Mexico.

Pridmore R.W. (1992) Music and Color: Relations in the Psychophysical Perspective, *COLOR Research and Applications*, 17, 1, pp 57-61.

Rabenhorst D.A. Farrell E.J. Jameson D. II. Linton T.D. and Mandelman J.A. (1990) Complementary Visualization and Sonification of Multidimensional Data, *Proceedings of SPIE Extracting Meaning from Complex Data: Processing, Display, Interaction 1990*, Vol 1259, pp 147-153, SPIE, California.

Robertson P.K. and O'Callaghan J.F. (1986) The Generation of Color Sequences for Univariate and Bivariate Mapping, *IEEE Computer Graphics and Applications*, Feb, pp 24-32.

Robertson P.K. (1988) Visualizing Color Gamuts: A User Interface for the Effective Use of Perceptual Color Spaces in Data Displays, *IEEE Computer Graphics and Applications*, 8, 5, pp 50-64.

Robertson P.K. (1991) A Methodology for Choosing Data Representations, *IEEE Computer Graphics and Applications*, 11, 3, pp. 56-67.

Robertson P.K. Hutchins M. Stevenson D. Barrass S. Gunn C. and Smith D. (1994) Mapping Data into Colour Gamuts: Using Interaction to Increase Usability and Reduce Complexity, *Computers & Graphics*, 18, 5, pp 653-665.

Robertson P.K.and De Ferrari L. (1994) Systematic Approaches to Visualization: Is a Reference Model Needed?, *Scientific Visualization Advances and Challenges*, Academic Press.

Rogowitz B.E. Ling D.T. and W.A. Kellogg (1992) Task Dependence, Veridicality, and Pre-Attentive Vision: Taking Advantage of Perceptually-Rich Computer Environments, *SPIE Human Vision, Visual Processing and Digital Display*, 1666, pp. 504-512.

Rogowitz B.E. and Treinish L.A. (1993a) Data Structures and Perceptual Structures, *Proceedings of the SPIE Human Vision, Visual Processing and Digital Display IV*, SPIE Vol. 1913.

Rogowitz B.E. and Treinish L.A. (1993b) An Architecture for Rule Based Visualization, *Proceedings of IEEE Visualization '93*, San Jose, California.

Rogowitz B.E. and Trienish L.A. (1994) Using Perceptual Rules in Interactive Visualization, *Proceedings of IS&T/SPIE Symposium on Electronic Imaging: Science and Technology*, San Jose, California.

Roth S. and Mattis J. (1990) Data Characterisation for Intelligent Graphics Presentation, *Proceedings of Computer Human Interaction CHI'90*, pp. 193-200.

Sandell G.J. (1993) Perception of Concurrent Timbres and Implications for Orchestration, *Proceedings of the International Computer Music Conference ICMC `94*, Aarhus.

Sandell G.J. (1995) *SHARC Timbre Database*, http://www.parmly.luc.edu/sandell/sharc/SHARC.homepage.html.

Scaletti C. (1994) Sound Synthesis Algorithms for Auditory Data Representations, in Kramer G. (ed) *Auditory Display : Sonification, Audification and Auditory Interfaces*, SFI Studies in the Sciences of Complexity, Proceedings Volume XVIII, Addison-Wesley Publishing Company, Reading, MA, U.S.A.

Sebba R. (1991) Structural Correspondence Between Music and Color, *COLOR Research and Applications*, 16, 2, pp 81-88.

Shannon C.E. and Weaver W. (1962) *The Mathematical Theory of Communication*, The University of Illinois Press, Urbana.

Shedroff N. (1995) Interaction Design Tutorial Notes, http://www.vivid.com/high/unified.html, *Proceeding of CHISIG OZCHI'95*, University of Wollongong.

Shephard R.N. (1964) Circularity in Judgments of Relative Pitch, *Journal of the Acoustical Society of America*, 36, pp 2346-2353.

Slaney M. (1995) A Critique of Pure Audition, *Proceedings of the Computational Auditory Scene Analysis Workshop, Joint International Conference on Artificial Intelligence*, Montreal.

Slawson A.W. (1968) Vowel Quality and Musical Timbre as Functions of Spectrum Envelope and Fundamental Frequency, *Journal of the Acoustical Society of America*, 43, pp 87-101.

Smith S. (1990) Representing Data with Sound, *Proceedings of the Visualization 1990*, IEEE Computer Society Press.

Smith S. Picket R.M. and Williams M.G. (1994) Environments for Exploring Auditory Representations of Multidimensional Data, in Kramer G. (ed) *Auditory Display : Sonification, Audification and Auditory Interfaces*, SFI Studies in the Sciences of Complexity, Proceedings Volume XVIII, Addison-Wesley Publishing Company, Reading, MA, U.S.A.

Smith S, Levkowitz H., Pickett R.M. and Torpey M. (1994) A System for Psychometric Testing of Auditory Representations of Scientific Data, *Proceedings of the Second International Conference on Auditory Display ICAD'94*, Santa Fe, New Mexico.

Stevens S.S. and Galanter E.H. (1957) Ratio scales and category scales for a dozen perceptual continua, *Journal of Experimental Psychology*, 54, pp 377-411.

Stevens S.S. (1962) The surprising simplicity of sensory metrics, American Psychologist, 1962, 17, pp 29-39.

Stevens S.S. (1966) Matching Functions Between Loudness and Ten Other Continua, *Perception and Psychophysics*, 1, pp 5-8.

Soorgard P. (1988) A Discussion of Computer Supported Collaborative Work, Ph.D. thesis, Aarhus University, Denmark.

Swan N. (1996) Ways of Seeing, The Health Report, *Radio National Transcripts*Monday 19 February 1996, http://www.abc.net.au/rn/talks/8.30/helthrpt/hstories/hr190201.htm.

Tkaczevski A. (1996) Auditory Interface Problems and Solution for Commercial Multimedia Products, *Proceedings of the Third International Conference on Auditory Display ICAD'96*, Palo Alto.

Truax B. (1992a) Composing with Time Shifted Environmental Sound, *Leonardo Music Journal 2(1).*

Trumbo B.E. (1981) A Theory for Coloring Bivariate Statistical Maps, *The American Statistician*, 35, 4.

Tufte E.R. (1983) *The Visual Display of Quantitative Information*, Graphics Press, Cheshire, Connecticut.

Van Cott H.P. and Kinkade R.G. (eds) (1972) *Human Engineering Guide to Equipment Design*, U.S. Government Printing Office, Washington D.C.

Vercoe B. (1991) *CSOUND, A Manual for the Audio Processing System and Supporting Programs*, Media Laboratory, M.I.T, Cambridge, Mass.

Vertegaal R. and Bonis (1994) ISEE: The Intuitive Sound Editting Environment, *Computer Music Journal*, 18, 2 , Summer, 1994.

Vickers P. and Alty J.L. (1996) CAITLIN: a Musical Program Auralisation Tool to Assist Novice PRogrammers with Debugging, *Proceeding sof the International Conference on Auditory Display ICAD'96*, Palo Alto.

Von Bismarck G. (1974a) Timbre of Steady State Sounds : a Factorial Investigation of its Verbal Attributes, *Acustica*, 30, pp 146-158.

Von Bismarck G. (1974b) Sharpness as an Attribute of the Timbre of Steady Sounds, *Acustica*, 30, pp 159.

Wall L. and Schwartz R. (1991) *Programming Perl*, O'Reilly and Associates.

Wallschlaeger C. and Busic-Snyder C. (1992) *Basic Visual Concepts and Principles for Artists, Architects and Designers*, Wm. C. Brown Publishers.

Ware C. (1993) The Foundations of Experimental Semiotics: a Theory of Sensory and Conventional Representation, *Journal of Visual Languages and Computing*, 4, pp 91-100, Academic Press Limited, London.

Watson C.S. and Kidd G.R. (1994) Factors in the Design of Effective Auditory Displays, *Proceedings of the Second International Conference on Auditory Display ICAD `94*, Santa Fe Institute, New Mexico.

Wessel D. (1985) Timbre Space as a Musical Control Structure, in Roads C. and Strawn J (eds) (1985) *Foundations of Computer Music*, MIT Press.

Whitehead J.F. (1994) The Audio Browser, *Proceedings of the International Computer Music Conference ICMC `94*, Aarhus.

Wiersma W. (1991) Research Methods in Education, Allyn and Bacon, Sydney.

Williams S.M. (1994) Perceptual Principles in Sound Grouping, in Kramer G. (ed) *Auditory Display : Sonification, Audification and Auditory Interfaces*, SFI Studies in the Sciences of Complexity, Proceedings Volume XVIII, Addison-Wesley Publishing Company, Reading, MA, U.S.A.

Wittgenstein L. (1953) *Philosophical Investigations*, Basil Blackwell, Oxford.

Wong, Sallis and O'Hare (1995) Information Portrayal for Decision Support in Dynamic Intentional Process Environments, *Australian Conference on Computer Human Interaction OZCHI '95*, CHISIG publications.

Wurman R.S. (1989) *Information Anxiety*, Doubleday, New York.

Wyszecki G. and Stiles W.S. (1967) *Color Science*, John Wiley and Sons Inc.

Zubakov V.A. and Borzenkova I.I. (1990) *Global Paleoclimate of the Late Cenozoic*, Elsevier, Amsterdam.

Zwicker E. and Fastl H. (1990), *Psychoacoustics Facts and Models*, Springer-Verlag.

Appendix

Appendix 5-1: EarBenders Stories

Title	Story
marimba keys	OK; Am I first? There seems to be something about short term memory that makes it easier to compare 2 sounds one after the other; than it is to compare 2 visual objects (distances) one after the other. your brain remembers the sound more accurately than the distance. e.g. If you show me a marimba key; then hide it; then show me another; I would have some difficulty deciding which was longer. But if you played it; then played another; i would have no trouble. Could be a problem if you are tone deaf! Chris
phone clues	police sometimes use the background sounds heard in a phone call to locate the phone box where the call was made - for example near a railway station or a restaurant may have unique background sounds ....
call to prayer	churches have a bell that is rung on sundays as a call to prayer mosques have loudspeakers through which the call to prayers is sung
school bell	in schools a bell or siren is often used to call the students to an assembly or to announce the end of lunchtime
squeaky boards	the bedroom of the japanese emperor was made with flor-boards that squeaked so that he could hear anyone who might be sneaking up on him in his sleep
tapping shoes	there was a medieval fashion of shoes with heels that tapped on the group to attract attention to the wearer
windscreen wipers	the windscreen wipers start squealing when there is not enough rain sprinkiling on the windscreen and this tells you to slow them down or turn them off
geese guards	roman armies placed geese on the camp periphery at night to warn against sneak attacks - the geese would reat to a disturbance by loud honking calls
irritating chewing	the sound of someone chewing gum and blowing bubbles in the seat behind me on the bus nearly drove me crazy - sounds can be very irritating - another one is someone eating an apple.
healthy car	I am always sensitive to the “sound” my car makes. When the car is “healthy” I don’t hear any noises; but if a new noise develops I will immediately hear it and think “oh no; what’s wrong now...”
mechanics with stethoscopes	I am sure many mechanics use sound to diagnose engine problems; in fact most mechanics have stethoscopes.
the car engine has a familiar ‘normal’ sound	Example in mechanics. The engine of my (rather old) car makes a complex but predictable ‘sound’. Just recently probably due to the cold weather there has been a new component added when the engine is cold. This is a bit of a worry; because if its not caused by the cold it means something might be getting ready to break ! Now that I think about it there are numerous examples of ‘normal’ sounds where a change from the ‘normal’ pattern triggers awareness.
Dropped item	Last night Matt and I were removing a perspex thingey-me-jig off the car window and we kept dropping bits and pieces (cause it was cold and dark); but we found every piece because we heard where it fell.
Baby talk	Also; I always know what Keely wants from the different cries she gives (ie. hungry; tired; uncomfortable; lonely; wet or bored). You can also tell different things from her happy sounds. Baby’s are very expressive with sounds.
Games - Myst Space Ship Tune Puzzle	have you noticed that the computer game Myst gives sounds as major part of the game. Also to get to other worlds you may need to use and manipulate the sounds you have discovered into a sequential pattern. Some can be tricky (cause the sounds sound similar; but are different ie. on the organ in the spaceship).

Title	Story
Games - Myst Elements KlangFarben Puzzle	have you noticed that the computer game Myst gives sounds as major part of the game. To get to other worlds you may need to use and manipulate the sounds you have discovered into a sequential pattern. Some can be tricky (cause the sounds sound similar; but are different ie. on the organ in the spaceship).
Games - Myst Train Navigation Puzzle	have you noticed that the computer game Myst gives sounds as major part of the game. They are used for direction guides (ie when moving around the maze etc).
Overloaded electric tools	Here is something maybe. When using an electric drill one would like to know when it is being overloaded. Presumably tradesmen learn by experience with particular tools what a cry of pain sounds like; but not the average citizen. Similar of course for the redline rpm in cars.
Getting to sleep	Dear Stephen; Just one point about noise. I think that noise in a modern society is something that you are more likely to want to blot out than listen to. I live in a flat; and when I first moved in it took ages before I could sleep without being woken up by noises around me. I wired my micro-moog (love that machine) into a really BIG speaker (12" woofer and all that) with white noise and left it on 24 hours a day in my bedroom. The result was a peacefull and blissful nights sleep (after I sto
Waking up	You are interested in presenting noise as something useful. It might be interesting to explore the other side - that noise is sometimes a penalty - indeed it is designed that way to wake us up before being mauled my sabre-tooth tigers and the like. I believe you have the same problem with Bettina listening to those story tapes about "and then the horse jumped the gate; and raced out of site over the hill .."? You have my symphony .. oops; I mean my sympathy.Good luck with the write-up. Scott.
striking matches	striking matches I can hear how dan is doing it - the pressure of the match head on the box - the fluidity of motion - the flare as it lights, even though concentrating on another activity with my other son timmy
Estimating Numbers of Frogs or Cicadas	estimating frog populations by listening to tape recordings made of ponds at night, estimating the numbers of cicadas in the vicinity in the middle of the day, the number of animals contributing to the sound changes the loudness and timbre of the overall sound
Species identification	identifying frog species by tape recordings made by ponds at night, identifying cicada species by their distinctive chirpings,identifying bird species from their calls
Dishwasher	gone to bed and have turned on the dishwasher - can hear if something is going wrong. like bad stacking of the plates, or the water tap is off.
shaving	shaving with electric razor - starts off noisy but is finished when it goes quiet as hairs get chopped
Dropped object	finding dropped screws; coins; tools
Car diagnosis	renault manual - gear; brakes; transmission; engine diagnosis
Climbing Vines	life on a vine - sbs tv - thais climbing cliffs for birds eggs pull on a vine and listen to hear a 'bad' one
finding people	someone is there - physical presence,finding people in the rubble of collapsed buildings
finding people	finding people in the rubble of collapsed buildings or in the bush
outside tap	tap is on or off, tap is running, hose outside is running
footsteps	a persons walking style, recognise someone by their footsteps
visitors	a visitor has arrived - a car coming up the drive; people on the porch; doorbell - housemate has arrived has key in the lock instead of doorbell.

Title	Story
Syncronous revs of plane engines	difference between engine revs - pilots use beating between engines to adjust revs to be synchronous for comfort since the beating is annoying
Door key card	door key-card reader beeps on error
PIN keypresses	beep on successful keypress keypress recognised - eftpos & autobank input beep
pick up card	autobanks sometimes double beep to remind you to retrieve your card at the end of a session
Bikes in Traffic	When you are riding your bike you use your hearing all the time - Crossing blind intersections. - Pedestrian road crossings. - Trucks, cars traffic behind your bicycle.- Bicycle coming along footpath
bells and horns	bike bells and car horns indicate that someone is trying to get your attention - could be telling you to watch out, or saying hello as they drive by - depends on the situation...
bicycle maintenance	bike maintenance - chain lubrication; getting gears in; bell ringing; pinging spoke; pin in the tyre; puncture
cat talk	cat-talk - want food; get off my tail; let me out; pat me; purring with contentment
rafting rapids and waterfalls	rafting rapids ahead, waterfall is nearby
weather	wind gusts roaring toward tent thru trees at island bend (snowies). rain bursts on tin roof; absence of noise when its snowing.
sickness	coughing = disease (stay away) - a person is sick with the flu and they sniff or blow their nose
siren	air raid siren at 12 noon = first monday of the month in holland (in the 70s?)
timer alarm	alarm clock; oven timer; egg-timer
door muffle	when the lounge door is open the tv and conversation is loud; when it is closed it is soft
window or door is open	an open window in the car whistles, when the back screen-door is unlatched it flaps in the wind
Coffee percolator	the coffee percolator makes a burbling sound while it is pumping through the filter and then a louder brighter burble when it is finished, then eventually the burbling stops
boiling water	hot water kettle boiling starts off silent, then starts bubbling with a low dull sound and gradually gets higher and brighter until it reaches a stable boiling state. You can usually tell where it is up to in the sequence and how much further there is to go - e.g. 10 seconds or 1 minute etc.
operating bug in the burners	“Thumping” noise in the roof above the lobby corridor level 3 is from the plant room above. This noise is from incorrectly adjusted gas burners on the boilers. The burners have some operating bug which prevents correct adjustment and causes burner dropouts. Turning the boiler off fixes the noise but causes severe shivering to occupants. The suppliers have been working on the problem and have promised to have the bug fixed by Friday this week ?? Arch
Rubbish	I mentioned my shaving use-of-sound the other day. I was talking to my partner Robyn about your stuff; and she had a good one as a scavenger. She said the quickest; best way to see if a bag of garbage contains something interesting is to give it a bit of a kick. The combination of the weight of the bag and sound that it makes tells you what’s inside. Normal household rubbish sounds a bit plasticy; bottles clink; crockery clatters; disposable nappies squelch.... So a lot of the time her job involves wanderin
Voice identification	identifying a criminal by voice in a police line-up

Title	Story
car noises fit like a glove	car noises fit the car like a glove - most powerful diagnostic in mechanics toolkit. jon lilleyman
Helicopter Gearbox	Helicopter gearboxes are very difficult to remove and have stringent maintenance requirements. Maintentance diagnosis is done using a stethoscope to listen for worn gears, broken teeth etc.
Airspeed Indicator	aviation legend - after ww2 when cockpit instruments were hard to get glider pilots strapped a harmonica to the wing as a stall warning device - no music / no fly !
stall warning	contemporary cockpits use a buzzer to alert the pilot of an imminent stall
slow down warning	speed stripes on the road - rythmic tyre sound warns you to slow down
cereals	Dear Ear-Benders; I often rattle things (particularly in the kitchen) to see what's in 'em. This can give visual; muscular and auditory feedback. The other day I had two opaque film canisters; one with 200 speed film and the other with 400 - so I rattled them to find the one I wantedbut they both sounded the same
dog talk	Like Sharri's baby; my dog is quite expressive with noise. I have grown particularly alert to her (the dog's) "I'm throwing up on the carpet now" sound. Matthew.
footsteps	absence of noise - as in when the footsteps stop outside andrew lochley's door which makes him cringe because someone wants something
aural debug by disc access clicks	OK; I have two more 1/ Computer programmers used to be able to tell when there was in infinite loop in their program by listening to the disk accesses - a regular click ...click ...click ...click ... where there was none before was not considered a good sign.
internet break in alert	2/ My notebook is on the Internet with no fire-walling; but I know that I will hear the disk spin up if someone tries to break in (assuming that I am in my office of course).
filling bottles	Whenever I have to fill an opaque bottle with liquid I listen to the sound that the liquid makes and the pitch tells me how much liquid is in the bottle. This is particularly handy when filling hot water bottles for the kids. It means you can pour the hot water full blast out if the tap and not have too great a risk of scalding yourself. Cheers Don B.
jungle war	jungle warfare in a movie called "the last bullet" - a japanese soldier remembers "by the time you see your enemy it is too late - you must listen to the sounds"
faulty dishwasher	washing machine mechanic - first thing he does is listen - diagnosed faulty relay - matthew
pedestrain crossing beeper	when the traffic light starts beeping it means you can walk across the pedestrian crossing
phone	telephone you are calling is engaged or ringing or dead, modem monitoring - ringing; answer or engaged
depth of a hole	drop a rock down a hole to tell how deep it is - neale
lightning distance	Distance of lightning by counting til thunder arrives
fishing cast	fishing cast - a good cast makes a nice whizzing sound, and you can tell how far it has gone by the amount of time it whizzes before it "plonks" into the water. Also you can hear where it landed.
horse	how fast is a horse going - clippitty clop; galloping etc.
car speed	you can tell how fast a car is as it passes by by the rate of change of the location of the sound, the loudness and timbre of the engine; and for racing cars you hear a change in pitch as it passes by due to doppler shift as well.

Title	Story
hitchhikers guide	>But I still think the Earth got blown up before Ford’s house got ‘dozed! >Richard. In the book (and radio play I believe); the front of Arthur’s house gets bulldozed while he is in the pub with Ford Prefect. The crashing noise prompts him to rush outside and back to the house (with Ford running after him). In the TV series - due to low budget no doubt - you just see the front fence being knocked over. Of course the Vogons then destroy the earth; which I suppose completes the job of knocking down Arthurs ho
CLC stuffed	The CLC still appears to be stuffed! i.e. It does not print anything and seems to be recognising only black. It sounded very rough as well.
air rescue beacon	in an air rescue an emergency radio search beacon transmits a tone at a certain radio frequency and searchers home in on it by loudness of the radio signal - neale
ice floe	the condition of a glacier ice floe is learnt by nearby residents who listen to the cracking noises it makes. neale
power station	maintenance men in power stations can hear the possibility of the transformer exploding and have been known to make a quick exit when things start sounding abnormal - neale
microwaving	microwave food pops and crackles when hot enough
gas leak	when the gas is on or leaking in the stove it makes a hissing sound.
gas cylinders	which gas cylinder has more in it ? you tap them and the fuller one has a low dull sound and the emptier one is higher and has a brighter hollower tone.
tracking marbles	18 month old boy tracking marbles by ear on the wooden floor when he cant see them because they are behind him - neale
Lorikeets	rainbow lorikeets calling at Batemans Bay golf-course reminded me of holidays 20 years ago when my family camped next to a bird sanctuary at Curumbin
Seagulls	seagulls squawking at Lake Gininderra makes it feel like being at the seaside
a good hit	a good golf hit or tennis hit sounds “sweet”. You can improve your game with this feedback, or anticipate the speed of an opponents tennis return.
barcode scanner	when the librarian or supermarket checkout pass the barcode scanner over an item it beeps to signal “ok”
alarms	fire alarm; burglar alarm; car alarm
sirens	police; ambulance; fire truck sirens
dripping tap	a tap or shower is dripping - you hear this especially when trying to go to sleep at night, even when it is out in the laundry.
road surface	different road surfaces make different sounds as you drive over them - concrete roads make regular clacks due to the cracks; asphalt is even but noisy, dirt has loud thumps.
radio phone	The remote phone can be paged from the base so you can find it under a cushion in the lounge or amongst pots and pans on the kitchen bench.
doctors with stethoscopes	I am sure many mechanics use sound to diagnose engine problems; in fact most mechanics have stethoscopes. Now that leads on to the medical profession and how they use sound to diagnose a patient...
Engine Load	Another motoring example With familiarity, I can tell the load being applied to the engine of my car. The ‘load level’ can be sensed even at constant speed. It must be a change in harmonic content since the basic frequency is speed related. This principle could be used as an audio system load monitor to replace the visual perfmeter !

Title	Story
surfing	surfing - you know that you have caught a good wave and are going fast because there is a harmonic vibration in the fins which you can hear when you exceed a certain speed
Getting up to attend a baby	Sometimes Keely wakes at night and I can tell from bed whether I should check on her or leave her to cry herself back to sleep.
applause count	in debates the winning team is determined by the amount of applause given each team by the audience
internal or external phone call	On our phones at work any internal phone calls have a single ring, while external phone calls are signalled by a double ring. Sometimes you are too busy to deal with an internal call so you ignore it.
driving in fog	Driving up Bulli pass in thick fog we had to cross 3 lanes of the Sydney expressway at a T-junction. Visibility was down to 30 m but by rolling the windows down and listening it was possible to hear invisible cars approaching and so make a less risky crossing.
squeaky dog toys	Dogs love to play with squeaky or rattly toys - these noisy toys keep their interest.
eggs and bacon	bacon and eggs sizzle and pop when they are frying - this is how you tell the heat is high enough
popcorn	when you pop popcorn it starts off slowly popping then builds up in rate and loudness, then dies down again as most of the corn has popped. you can tell how many popcorns are going off all through the cooking process, and you can tell when its finished
Baby identification	You can recognise the cry of your own baby from those of others in a group of babies.
in tune	you can tell if a guitar is out of tune by strumming a familiar chord - if ok it sounds sweet, otherwise it sounds rough
in tune	you can also tell if a guitar is out of tune by plucking a familiar chord from low to high as an arpeggio - there is a rgular pattern if it is ok, otherwise the out of tune string sticks out like a sore thumb and the arpeggio sounds “off”
micro-tuning	You tune a guitar by listening for “beats” between adjacent strings when the same note on each is plucked together. When the beats disappear the strings are in tune.
caving	when you are in the dark in a small tunnel there is no reverb and the sound is very “dry”, when you enter a larger room you automatically register the increased size of the room by the increase in the “wetness” or reverb of voice and dripping and scraping sounds
mowing the lawn	When you are mowing you can monitor the thickness of the grass by the loudness of the motor soundm plus you can tell that you are mowing over sticks, stones or other obstacles from the clackety sounds.
customer buzzer	Shops sometimes have a bell on the door or an electronic buzzer which goes off as a customer passes through the doorway of the shop. Even if the shop keeper is serving someone or is stocking the shelves elsewhere in the shop they are able to monitor comings and goings. Actually the sound will be the same for both coming or going, but the context is different - so you might discern a going by a connection with preliminary sounds such as footsteps or talking ... maybe coming and going are actually different c
wind chimes	Wind chimes make a pleasant realxing sound when the wind blows. However they are not usually used to inform that the wind is blowing but as a form of aural decoration. Other garden decorations that come to mind are fountains and bubbling brooks and japanese water knockers.

Title	Story
vacuum cleaner	What Ducted vacuum cleaner with 9 metre hose -Action suck suck suck, then something gets blocked (in fact the hose twists and kinks blocking off the air flow.-Sound pitch of whine of electric motor rises alarmingly as the load becomes excessive -Interpretation hose is blocked
hammering a nail	Sound solid sounding knock, then a softer slightly clangy sound -Interpretation Nail is bent -Sound solid sounding knock, then overly solid (higher pitch) -Interpretation Nail has hit something (pipe ? concrete ?)
meditation mantra	In meditation you mentally repeat a mantra sound (a word without meaning) to help focus your mind and distance your connection with the external world.
music to study by	It is common for students to listen to music turned up very loud when studying. Rather than being distracting this actually helps some people focus on their work (well it did for me !). I think it is helpful because it blocks out other distracting sounds (noises) such as the tv in the next room, or people talking etc which would likely lure you away from the textbooks...also it can help you settle into your study habitat, acting as a psychological conditioner like a bell causing a dog to salivate.
pavlovs dogs	Pavlov trained dogs to assosciate the sound of a bell with food by ringing it when they were given a meal. Once this conditioning had occurred the dogs would salivate at the sound of a bell, even though food was not present. A sound which does not naturally occur in the presence of food could trigger a reflex which is only useful for eating. This is strong evidence against the ecological premise that information is structured by the environment, and undermines some of the claims made for the Auditory Icon m
Computer	What Waiting for something to happen as per instruction -Sound nothing then a busy clicking of the hard disc heads -Interpretation Something is happening
laser printer maintenance	printing a page on the laser printer -Sound graunch, grind -Interpretation something needs maintenance, maybe oil
electric kettle	boiling water in the electric kettle -Sound rumble gets louder, bubbling, then ‘click’ -Intepretation auto switch-off when boiled
Raining Outside	in a house with a tin roof you can hear how bad the rain is outside - sprinkling is a soft pitter-pat, but when its raining hard it can be so loud it drowns out your voice
cocktail party	at a cocktail party you can choose to listen to the person you are conversing with, whilst monitoring another conversation elsewhere in the room, or follow someones voice or identify people and groups of people, all against the general hubbub of voices, music and other background noise which indicate the level of animation of the proceedings...
Diabetes advert	TV advertisements for diabetes campaign Do you remember this song ? “chewie, chewie etc.” then you are over 40 years old and need a diabetes test.
wind chime	my family in Queensland have tuned windchimes on their porch and when the wind blows in a northerly it plays a distinctly different chord to when it blows east.
tennis	when you play tennis in a pressurised dome you can’t hear the ball when you hit it and it makes playing a very strange becuase you can’t tell when the ball has been hit by your opponents.
macro-tuning	to tune strings play the same note on adjacent strings, one close after the other. If the pitch of the tuned string is too high then release the string tension a little and test again. If too low then increase the string tension. Keep doing this until there is no gross pitch difference. Now use the beats as in the micro-tuning story.

Title	Story
Irrigation	ACT Parks and Gardens is responsible for maintaining the green spaces in and around Canberra, which is a national capital with a strong tourist industry. Irrigation is a major investment and it is important to ensure that the system is in good working order and is performing efficiently. Satellite images can be used to assess vegetation growth and soil wetness. An image of greeness difference is overlaid over cultural features but it is difficult to show soil wetness on the same display. Can sounds be used
walking at night	walking home in the dark you can tell when you are walking over leaves, grass, concrete gravel etc. by the sound of your footsteps
walking speed	when walking home you can hear someone behind you apporaching as their footsteps get louder and also by the rate of their steps which is faster than your own
breathing	you can tell if someone is exhausted by their breathing - heavy, fast or shallow wheezy, - sharp short means sick, - you can’t normally hear someone breathing if they aren’t somehow distressed
flag in the wind	a flag flapping in the wind makes a louder faster flappier sound as the wind speed increases
ducted central heating	action - turn-up heat on thermostat, sound - slight delay then satisfying click, interpretation - relay has clicked and it will start soon
ducted heating 2	action - turn-up heat on thermostat, sound - distant soft ‘explosion’, interpretation - gas burner has switched on and heat will start soon
cold start	action - cold start of car, sound - engine clatter gradually settles down to a hum after 5 mins, interpretation - engine was cold
power saw	as the saw cuts into the saw the whine gets deeper in pitch ad duller, as you pull it out it raises in pitch to its normal whine again - I guess its something to do with the speed of the saw.
sprinkler system	small drip sprinklers make a hissing sound when they are operating properly and though you cant usually see them you can walk along the row of bushes and listen to make sure the waterinng is going ok at each bush
wolf whistle	a wold whistle is someone trying to get your attention at a distance in an outdoor environment
counting	a person can communicate a quantity using spoken numbers
objects = nouns	nouns are usually perceptually distinctive sounds, especially if commonly used in similar contexts - e.g. car, bike, van, bus, train, ute
japanese ladies	japanese women continually flush the toilet to cover other noises
relaxation tapes	Rain is natures steady, cleansing symphony - Rainfall in the Mist relaxation tape helps you relax and drift away to a place of beauty and serenity. Other tapes have ocean or waterfall scenes.
dog collar bell	walking my dog ziggy on Mt. Ainslie I stopped for a breather and Ziggy went off into the bush sniffing around. I couldn’t see her but couyld tell she was moving away from me by the sound of the bell on her collar. In the end when she wouldne come I had to go and find her - I think she was lost !
cicadas in heat	cicadas only start to sing when the temperature rises above 18 degress celsius, though different species may vary several degrees
bumble bee	bumblre bees use a special buzzing key to open a particular type of flower
egg collectors	a tv documentary showed thai egg collectors who sail to remote islands and climb hundreds of metres up cliff-faces to reach bird nests using vines which grow there. The collectors pull on a vine and listen to the sound it makes before trusting it with their lives, apparently they can hear a good from a bad vine this way...

Title	Story
dropped keys	you can usually find tour keys straight away if you drop them in the dark by the sound of where they fell
camera shutter	you know you have taken the photo when the shutter clicks
camera timer	a camera timer makes a buzzing noise until it clicks off the shot
cassette rewind	as a cassette rewinds the sound gets faster and faster unti it clunks to a stop
phone dial tunes	when you dial a push button phone consecutive numbers have increasing pitch - with familiar numbers you know that you have mis-dialled because of the wrong tune
phone dial press	when you press a phone dialling button it makes a characteristic pitch that lets you know that you pressed it hard enough
broken cicada	a damaged male cicada cannot sing well and will not be able to attract mates
insects attract mates	crickets, cicadas and many other insects chirp, rub their legs together or hammer their abdomens to call mates. Each cicada species has a distinctive call and the females are very selective in their response so overlapping territories do not cause disruptions
insects frighten predators	beetles make explosions using gas to scare off predators, masses of cicadas chirp loudly (>100dB) to scare off birds
insects stake territory	as a grasshopper flies it clicks its legs to let others know that this is its territory
pair identification	a telecom technician in civic was using a beeping device to identify wire pairs between his location and a site across the road
continuity checker	my multimeter has a continuity option which makes a beep when there is no resistance between the leads
mice,rats or possums in roof	whilst staying at the coast house I heard something running around in the roof, maybe a rat or possum because it sounded bigger than a mouse
tuning a radio	when searching for a radio station you get a noise or whine which diminishes around the a station frequency which is a kind of notch in the noise band
microwave or oven timer	the microwave timer goes bing when the time has run out
collision	We may predict the potential collision of two objects by observing their paths with our eyes, but it is the sound of the collision that best reveals how the structure of the objects has been affected by the collision.
image projection	teenagers cruise around civic with their music blaring from their mobiles to project their particular image - hip-hop, gothic, thrash whatever...same with partioes and personal collections - play it loud to let people know who you are !
earcons	earcons are short musical motifs consisting of 2 or 3 changes in timbre, pitch or rhythm which are used to signal interface events such as opening a file or scrolling a menu.
seismic audification	seismic data consists of ? channels of continuous time series ratio pressure measurements from an array of pressure transducers buried in the ground. The analyst looks at visual traces to dentify event signatures which can be used to identify the time of unusual events and discriminate between nuclear tests and earthquakes at remote locations.
ripe fruit	you can tell whether a water melon is ripe by tapping it and listening, *this also works for apples, it must have something to do with the the damping of the sound when the fruit isn't ripe, because it sort of rings when its ripe*

Title	Story
songlines	youll hear the expression aquiring ritual knowledge. All this meant was that the man was extending his son-map...*The next point, he said, was to understand that every song cycle went leap-frogging through language barriers, regardless of tribe or frontier. A Dreaming track might start in the north-west, near Broome; thread its way through twenty languages or more; and go on to hit the sea near Adelaide. And yet, I said, its still the same song. ...Does that mean, I asked, that a young man on Walkabout c*
geese navigation	Certain ducks and geese can 'record' the choruses of frogs beneath them, and 'know' that they are flying over marsh.
birds migrations	songlines - *Other night fliers bounce their calls on to the ground below, and, catching the echo, fix their altitude and the nature of the terrain.*
dolphin triangulation	Dolphins flash echo-locating clicks on to submarine reefs, in order to steer a safe passage through..It has even occurred to me that, when a dolphin 'triangulates' to determine its position, its behaviour is analogous to our own, as we name and compare the 'things' encountered in our daily lives, and so establish our place in the world.
hula hoops	from the movie "the hudsucker proxy" *then we put a little bit of sand inside the hula hoop that shooooshes to make it - well to make it more fun*
a clue	in a tv cop show *the hunted man brushes past a hanging mobile that tinkles as he passes throgh the doorway. He hides in an adjoining room and waits. Subtly we hear the tinkling of the mobile again and the man opens fire through the fibro wall into the next room, killin gthe assassin in the dioorway.*
oil drilling	sometimes there is a risk of accidentlaly intersecting another oil well when drilling for oil. One way of monitoruing the progress is to lower a microphone down the oil well at risk and listen ffor warning sounds whilst drilling
mr whippy	the mr whippy ice-cream van plays greensleeves and you can hear it coming several blocks away, and tell when its getting close to your house in time to get som echanges and track it down
troubadors	before the printing press the troubadors were paid to sing the news as they travelled the countryside - they were the source of news. they would trade songs with each other, and were able to remember songs that lasted hours after only one hearing
gold	Can you find the gold? It is hypothesised that six different aspects of the land in which gold may be found are determinative of whether or not gold is there. The first 20 data variables (each 6-d) are from sites known to have gold; the second 20 data variables are from sites known not to have gold. For each of the remaining 10 data variables, decide fro each whether or not it is from a site with gold.
entomology	when working at entomology arch was asked to build a listening device to detect weevils on the conveyor belt on their way into grain trucks. you could hear these insects rustling in the grain even though you couldn't see them - especially in a big silo.
finding studs	you can find where the studs that the plasterboard walls of the house are fixed to by knocking on the walls. the walls sound hollow *but where the studs are it is more solid and duller*
male spacing behaviour	As in the bladder cicada, most insects use the distant cue of sound to indicate their presence to would be competitors, and maintain a safe distance between agressors. If the males adopt this tactic, and the vegetation is homogenous, the males will space themselves quite evenly, and spacing will be a consequence of acoustic rivalry. But the distance at which the male calls may be under selection, depending on the number of males present, the intensity of the call, the hearing sensitivity of neighbouring ma

Title	Story
size and reproductive success	In those insects using sound as the primary long-distance cue for mate attraction, females do prefer larger, more loudly calling males e.g. mole crickets, bush crickets, fruit flies. Even under water intensity is an overriding factor infleuncing the behaviour of female corixids. Deep croaking frogs repeatedly get more mates than higher pitched rivals, and the louder calls of larger natterjack toads attrack more mates. Bailey W.J. (1991) Acoustic Behaviour of Insects: an evolutionary perspective, Chapman and
species pattern	Males call with a species pattern and the call must have a quite resticted frequency range. The female must recognise this signal as its own species, and its hearing system would do well to be tuned to the males call. Bailey W.J. (1991) Acoustic Behaviour of Insects: an evolutionary perspective, Chapman and Hall, London
aggregations of calling insects	Calling by aggregated insects became central to the group versus individual slection controversyof the mid 1960s where the group selection theory was that chorusing groups evolved as displays that allowed individuals to gain information as to the size of the population, and adjust their behaviour accordingly. Bailey W.J. (1991) Acoustic Behaviour of Insects: an evolutionary perspective, Chapman and Hall, London
oversinging and masking	Perhaps males use their song to mask the call of another to minimise the influence of an intruder on a searching female. Resident male O.nigripes over-sing their neighbours in this way. The song consists of a long series of buzzes, and with this style of song masking could be quite effective. The interval between buzzes is adjusted to fit over the intruder’s song, essentially synchronisign the sound bursts. Bailey W.J. (1991) Acoustic Behaviour of Insects: an evolutionary perspective, Chapman and Hall, Lon
war propaganda	Amphiacusta maya is a phalangopsid cricket from Central America which occurs in groups in tree hollows. It is unusual because the males song has no calling function: females are not attracted to it. The primary role of the song appears to be for keeping other males away during copulation - war propaganda rathaer than courtship. Bailey W.J. (1991) Acoustic Behaviour of Insects: an evolutionary perspective, Chapman and Hall, London
insect thermometer	The snowy tree cricket produces its call as a nearly continuous sequence of chirps, with each containing 2-11 pulses, although 5-8 is more common. The chirp rate and structure of the chirp are highly temperature dependent, varying between 50 and 200 chirps per minute. To maintain synchrony in a chorus some members of the population in warm spots must slow down, whilst those at lower temperature must increase their singing rate. Bailey W.J. (1991) Acoustic Behaviour of Insects: an evolutionary perspective, C
distance information	Insects are frequency sensitive, with many showing specialised hearing structures capable of differentially filtering certain frequency bands. One of the features of sound as it travels through the medium is that certain frrequencies are lost faster than others and this loss may give information on the distance of the caller. Bailey W.J. (1991) Acoustic Behaviour of Insects: an evolutionary perspective, Chapman and Hall, London

Title	Story
defence	Many ants produce sounds with high frequencies clsoe to 10 kHz that function in defence. The production of sound in ants is throught the movement of a plectrum on the posterior part of the petiole, where it engages a file on the first segment of the gaster. Bailey W.J. (1991) Acoustic Behaviour of Insects: an evolutionary perspective, Chapman and Hall, London
call for help	High frequencies are severely attenuated by soil, favouring the transmission of low frequency elements. Thus the non-sexual signals by ants, such as the leaf cutter ant Atta sexdens are low frequency calls, and although the function of these may normally be between nest mates, the signal also acts as an alarm to members of the colony when one of their number is buried in an earth fall. Bailey W.J. (1991) Acoustic Behaviour of Insects: an evolutionary perspective, Chapman and Hall, London
popcorn done	when you pop popcorn it starts off slowly popping then builds up in rate and loudness, then dies down again as most of the corn has popped. you can tell how many popcorns are going off all through the cooking process, and you can tell when its finished

Appendix 5-2: Lookup.prl

```
#!/local/bin/perl5 -w
#
####################################################################################
#COPYRIGHT (C) CSIRO DIT CSIS 1994
#
#SOURCE FILElookup.prl
#
#MODULE      Sonify
#
#SYSTEM      Thesis
#
#ENVIRONMENTSun SPARC and SUNOS4.1.2
#
#AUTHORStephen Barrass, CSIS
#
#HISTORY
#            : First written October 1995
####################################################################################


#-------------------------------------------------------------------------------
# help message
#
sub help {
print STDERR “Synopsis : $0 < database > examples\n”;
print STDERR “EarBenders Lookup - ranks records by similarity to a lookup record\n”;
print STDERR “newline is the record separator\n”;
print STDERR “+ is the within field AND operator\n”;
print STDERR “; is the within field OR operator\n”;

print STDERR “[-h]\t\t\t\tthis help message\n”;
}

#-------------------------------------------------------------------------------
# main
#
print STDERR “$0\n”;

#-------------------------------------------------------------------------------
# parse command line
#

while ($_ = $ARGV[0], /^-/) {
  shift;
if (/^-h/)
     {
             help;
            exit 1;
     }
   else
     {
             help;
            exit 1;
     }
}

$i = 0;
while (<@ARGV>) {
print “$_\n”;

  shift;
  $i++;
```

```
}

# score the entries
$i = 0;
$max = 0;
$maxcnt = 0;
$TASK_OFFSET=6;

while (<>) {
        chomp;
# split the line into fields
        @fields = split(/\|/);
        if ($i == 0) {
                @query = @fields;
        print STDERR "#QUERY\n#FIELD#=$#fields\n$_\n\n";
              $i++;
              next;
        }
        print STDERR "#story $i, fields=#=$#fields\n@fields\n";

# for each field
        $score = 0;
        for ($j=$TASK_OFFSET; $j < $#fields; $j++) {
                @subfields = split(/[;+]/, "$fields[$j]");
                for ($k=0; $k <= $#subfields; $k++) {
                        print STDERR "\nQuery=$query[$j],Field=$subfields[$k]";
                   if ($query[$j] eq "???") {
                                        next;
                               }
                   elsif ($query[$j] eq "") {
                                        next;
                               }
                   elsif ($query[$j] eq "all") {
                                print STDERR "=MATCH";
                                $score++;
                               }
                   elsif ($subfields[$k] eq "all") {
                                print STDERR "=MATCH";
                                $score++;
                               }
                        elsif ((index $query[$j],$subfields[$k]) >= 0) {
                                print STDERR "=MATCH";
                                $score++;
                        }
                        elsif ((index $subfields[$k],$query[$j]) >= 0) {
                                print STDERR "=MATCH";
                                $score++;
                        }
                }
        }
print STDERR "SCORE=$score, MAX=$max\n";
        if ($score >= $max) {
                print "#story $i=$score, FIELDS=$#fields\n$_\n\n";
         $maxcnt++;
                if (($i > 0) && ($maxcnt > 1)) {
                        $max = $score;
                     $maxcnt = 0;
                }
        }
        $i++;
}

print "#QUERY -------\n";
print "@query\n";
print "#max=$max\n";
```

Appendix 5-3: Synthesis.prl

```
#!/local/bin/perl5
#
####################################################################################
#      COPYRIGHT (C) CSIRO DIT CSIS 1996
#
#      SOURCE FILE     synthesis.prl
#
#      MODULE          Sonify
#
#      SYSTEM          Thesis
#
#      ENVIRONMENT     Sun SPARC and SUNOS4.1.2
#
#      AUTHOR          Stephen Barrass, CSIS
#
#      HISTORY
#                    : First written December 1996
####################################################################################


#------------------------------------------------------------------------------
# help message
#
sub help {
print STDERR "Synopsis : $0 < database > examples\n";
print STDERR "EarBenders Design\n";
print STDERR "synthesise design from Sound Analyses\n";
print STDERR "[-i (3)]\t\t\tnumber of analyses to use\n";

print STDERR "[-h]\t\t\t\tthis help message\n";
}

#------------------------------------------------------------------------------
# Sound Design
#

format DESIGN =
------------------------------------------------
Sound Design
Nature        @<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<|
@SoundNature
Level         @<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<|
@SoundLevel
Streams       @<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<|
@SoundStream
Range         @<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<|
@SoundRange
Compound      @<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<
@SoundCompound
Occurence     @<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<
@SoundOccurence
Pattern       @<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<
@SoundPattern
Movement      @<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<
@SoundMovement
Relation      @<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<
@SoundRelation
Organisation  @<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<
@SoundOrganisation
.

sub SoundDesign {
```

```
# descriptions of the actual sounds to be used
@Sounds=();
# descriptions of principal variation
@Descriptors = ();

@SoundNatures = (‘everyday’, ‘synthetic’, ‘vocal’, ‘verbal’);
@SoundNature = ();

@SoundLevels = (‘local’, ‘global’);
@SoundLevel = ();

@SoundStreams = (‘single’,’few’,’many’);
@SoundStream = ();

@SoundOccurences = (‘isolated’);
@SoundOccurence = ();

@SoundPatterns = (‘discrete’);
@SoundPattern = ();

@SoundMovements = (‘stationary’,’jumping’,’smooth’,’blanket’);
@SoundMovement = ();

@SoundRelations = (‘category’, ‘continuum’, ‘order’, ‘metric’, ‘zero’);
@SoundRelation=();

@SoundOrganisations = (‘nominal’,’ordinal’,’interval’,’ratio’);
@SoundOrganisation = ();

@SoundCompounds = (‘integral’,’seperable’);
@SoundCompound = ();

#number of different sounds
@SoundRange = ();
}

#------------------------------------------------------------------------------
# Majority
# return the majority trend in a list of fields
#
sub Majority {
(@fields) = @_;
local($i,$j);
$majority = 0;
$matchesMAX = 0;
# for each but the last field
for ($i=0; $i<$#fields; $i++) {
# count matches along the fields
     $matches = 0;
    for ($j=$i+1; $j <= $#fields; $j++) {
            if ($fields[$i] eq $fields[$j]) {
                   $matches++;
                 }
            }
    if ($matches > $matchesMAX) {
            $matchesMAX = $matches;
            $majority = $i;
            }
    }
# if no majority then return the fields
if ($matchesMAX == 0) {
    return “no majority in (@fields)”;
    }
# return the majority trend
$m = ($matchesMAX+1);
$f = ($#fields+1);
```

```
$mTotal += $m;
$fTotal += $f;
return “$m\/$f ($fields[$majority]) in (@fields)”;
}

#-------------------------------------------------------------------------------
# main
#
print STDERR “$0\n”;

#-------------------------------------------------------------------------------
# parse command line
#
$iMAX = 3;                    # number of analyses to synthesise from
while ($_ = $ARGV[0], /^-/) {
  shift;
if (/^-h/)
     {
            help;
           exit 1;
     }
elsif (/^-i/)
     {
      if ($i=shift > 0) {
                 $iMAX = $i;
               }
     }
   else
     {
            help;
           exit 1;
     }
}


$SOUND_OFFSET=15;             # offset of the sound analysis in EarBenders
SoundDesign;                  # initialise the Sound Design

$i = 0;

print “#--------- query\n<“;
$q = <STDIN>;
$q = <STDIN>;
$q = <STDIN>;
$q = <STDIN>;
$q = <STDIN>;
print “$q\n”;

while (<STDIN>) {
if (/^ *#/) { next; }# skip comment lines
if (/^ *\n/) { next; }# skip blank lines
if ($i >= $iMAX) { last; }
chomp;
# split the line into fields
@fields = split(/\|/);
# get each field from each of the top entries
print STDOUT “#--------- case $i\n@fields\n”;
$j = $SOUND_OFFSET;
@SoundLevel = ($fields[$j++],@SoundLevel);
@SoundCompound = ($fields[$j++],@SoundCompound);
@SoundDescriptors = ($fields[$j++],@SoundDescriptors);
@SoundRelation = ($fields[$j++],@SoundRelation);
@SoundNature = ($fields[$j++],@SoundNature);
@SoundOccurence = ($fields[$j++],@SoundOccurence);
@SoundPattern = ($fields[$j++],@SoundPattern);
@SoundStream = ($fields[$j++],@SoundStream);
```

```
@SoundMovement = ($fields[$j++],@SoundMovement);

  $i++;
}

# design synthesis
# use majority trend for each field
# if no majority then leave it up to the designer
@SoundLevel = Majority(@SoundLevel);
@SoundCompound = Majority(@SoundCompound);
@SoundDescriptors = Majority(@SoundDescriptors);
@SoundRelation = Majority(@SoundRelation);
@SoundNature = Majority(@SoundNature);
@SoundOccurence = Majority(@SoundOccurence);
@SoundPattern = Majority(@SoundPattern);
@SoundStream = Majority(@SoundStream);
@SoundMovement = Majority(@SoundMovement);

open(DESIGN, “>-”) || die “can’t create”;
write DESIGN;                          # output

#close INFILE;

$percentScore = ($mTotal*100)/$fTotal;
print “Total Match Score = $mTotal\/$fTotal = $percentScore\n”;
```

Appendix 5-4: Casedesign.csh

```
#!/bin/csh -f
##################################################################################
#COPYRIGHT (C) CSIRO DIT CSIS 1994
#
#SOURCE FILEbatch.csh
#
#MODULE      Sonify
#
#SYSTEM      Thesis
#
#ENVIRONMENTSun SPARC and SUNOS4.1.2
#
#AUTHORStephen Barrass, CSIS
#
#HISTORY
#           : First written September 1994
##################################################################################
# Synopsis
#--------------------------------------------------------------------------------
# command line
#
set files =
while ($#argv > 0)

    switch ($1)
        case -h:
            goto synopsis
            breaksw

        default:
            set files = ($files $1)
            breaksw

    endsw
    shift
end

#--------------------------------------------------------------------------------
# main
# assume query is the last item in the earbend.txt
# need to move it to the top position
tail -1r < earbend.txt > x
cat x earbend.txt | lookup.prl > ranked.txt
# best stories are at the end
tail -r < ranked.txt | synthesis.prl > design.txt
rm x
goto end
#--------------------------------------------------------------------------------
# help message
#
synopsis:
echo “synopsis: $0”
echo “EarBenders case-based design synthesis”
echo “the earbenders case base is assumed to be in earbend.txt”
echo “the synthesised design goes to design.txt”
echo “lots of messages go to stderr”
end:
```

Appendix 6-1: GoldMaker.prl

```
#!/local/bin/perl5
#
################################################################################
#COPYRIGHT (C) CSIRO DIT CSIS 1994
#
#SOURCE FILEgoldbug.prl
#
#MODULE       Sonify
#
#SYSTEM       Thesis
#
#ENVIRONMENTSun SPARC and SUNOS4.1.2
#
#AUTHORStephen Barrass, CSIS
#
#HISTORY
#             : First written March 1997
################################################################################

# generate a mixture of dirt and gold
# there are a number of elements per handful and a number of handfuls that are thrown
$handful = 20;
$throws = 5;
$tempo = 600;
$goldlevel = 0.5;

#-----------------------------------------------------------------------------
# help message
#
sub help {
print STDERR “Synopsis : $0 \n”;
print STDERR “generate handfuls of Bly’s dirt and gold data\n”;
print STDERR “[-h]\t\t\t\tthis help message\n”;
print STDERR “[-e <$handful>]\t\t\telements in a handful\n”;
print STDERR “[-g <$goldlevel>]\t\t\tratio of gold 0.0-1.0\n”;
print STDERR “[-l <$tempo>]\t\t\trate at which elements occur\n”;
print STDERR “[-t <$throws>]\t\t\tnumber of handfuls that are thrown\n”;
}

#-----------------------------------------------------------------------------
# main
#
print STDERR “$0\n”;

#-----------------------------------------------------------------------------
# parse command line
#

while ($_ = $ARGV[0], /^-/) {
  shift;
if (/^-h/)
     {
             help;
            exit 1;
     }
elsif (/^-e/)
     {
      $handful = shift;
     }
elsif (/^-l/)
     {
      $tempo = shift;
```

```
      }
elsif (/^-t/)
      {
       $throws = shift;
      }
elsif (/^-g/)
      {
       $goldlevel = shift;
      }
   else
      {
              help;
             exit 1;
      }
}

# normal deviate by the Polar method

sub norm {
my $s = 1;
while ($s >= 1) {
      $u1 = (rand) * 2-1;
      $u2 = (rand) * 2-1;
      $s = $u1*$u1 + $u2*$u2;
     }
$r = abs($u1*sqrt(-2*log($s)/$s));
return($r);
}

sub norm1 {
my $r = 4;
while ($r > 3.5) {
          $r = norm;
     }
return $r;
}

sub score {
        printf "i1 %.3f 1 %.3f %.3f %.3f %.3f %.3f %.3f ",$t+$x1,$x1,$x2,$x3,$x4,$x5,$x6;
#         printf "i1 + . %.3f %.3f %.3f %.3f %.3f %.3f ",$x1,$x2,$x3,$x4,$x5,$x6;

}

sub handful {
my($goldlevel) = @_;
my $count = 0;
my $goldcount = 0;
while ($count < $handful) {
# first get a random number for each dimension
        $x1 = norm1; $x2 =  norm1; $x3 = norm1;$x4 = norm1;$x5 = norm1;$x6 = norm1;
        $type = gold;
# default is gold
# set 2 = DIRT must fit these criteria
        if (
                ($x2*$x2+$x3*$x3+$x4*$x4+$x5*$x5+$x6*$x6 <= 2.25) ||
                ($x1*$x1+$x3*$x3+$x4*$x4+$x5*$x5+$x6*$x6 <= 2.25) ||
                ($x1*$x1+$x2*$x2+$x4*$x4+$x5*$x5+$x6*$x6 <= 2.25)
        ) {
                $q = 0; $r=0;
                if ($x1 < 1.5) {$q++};
                if ($x2 < 1.5) {$q++};
                if ($x3 < 1.5) {$q++};
                if ($x4 < 1.5) {$q++};
                if ($x5 < 1.5) {$q++};
                if ($x6 < 1.5) {$q++};
                if ($q < 5) {next};
```

```
            if ($x1 > 1.5) {$r++};
            if ($x2 > 1.5) {$r++};
            if ($x3 > 1.5) {$r++};
            if ($r > 1) {next};
            if ($x4 > 1.5 || $x5 > 1.5 || $x6 > 1.5) {next};
                $type = dirt;
      }

# if its gold and theres not enough in the mix then add it
#print "G=$type,$goldlevel,$goldcount,$handful\n";
     if ($type eq gold) {
              if ($goldcount++/$handful < $goldlevel) {
                             score();
            print "; gold\n";
                            $count++;
                        }
                     next;
             }
          score();
            print "; dirt\n";
        $count++;
     }
# end of handful
#print "s\n";
}

#
# MAIN
#
# print the header
print "f1 0 8193 10 1; sin\n";
print "f2 0 129 7 0 128 1; ramp 0 to 1\n";
print "t 0 $tempo; tempo\n";
printf "i1 %.3f 1 %.3f %.3f %.3f %.3f %.3f %.3f\n",0,0,0,0,0,0,0;

srand;
$t = 0;
for (; $t <= $throws; $t++) {
handful($goldlevel);
}
```

Appendix 7-1: PitchCircle.orc

```
;################################################################################
;#     COPYRIGHT (C) CSIRO DIT CSIS 1993
;#
;#     SOURCE FILE     Shephard.orc
;#
;#     MODULE          Sonify
;#
;#     SYSTEM          PostGrad
;#
;#     ENVIRONMENT     Sun SPARC and SUNOS4.1.2
;#
;#     AUTHOR          Stephen Barrass, CSIS
;#
;#     HISTORY
;#                  : First written March 1994
;################################################################################
;
;
; sparc10
;       sr = 44100
;       kr = 441
;       ksmps = 100

; sparc2
sr = 16000
kr = 1600
ksmps = 10

nchnls = 1

gifnyq = sr/2    ; nyquist freq
giattack  = 0.01  ; prevent onset clicks
gidecay   = 0.005  ; time buffer (in seconds) to prevent clicks
giampmax = 70     ; max amplitude in dB

; timbre  - consists of timbre angle and timbre radius
gatangle init 0
gatradius init 0
gatimbre init 0

;
;// instrument                ////////////////////////////////////////////////
; pitch cycle - after Shephard
; p1p2      p3        p4        p5        p6        p7
; instrstartdur     comb    teeth     loud   ncycles
;                        oct.cl    oct.cl0.0..1.0    0-n

instr   1
print p1, p2, p3,p4,p5,      p6,        p7
idur=        p3
ifcomb=cpspch(p4)
ifteeth=cpspch(p5)
iamp  =ampdb(p6*giampmax)
incycles = p7
print idur, ifcomb, ifteeth, iamp, incycles

;-- loudness                ----------------------------------------------
kamp    linen  iamp, giattack, idur, gidecay  ; de-click envelope
print iamp

; build the cycle----------------------------------------------
; creat the src spectrum
```

```
;
; cycle up a ramp if continuous version is called by ncycles > 0
kcycleramp = 0
if (incycles < 1) kgoto discrete
kcycleramp line 0, idur, incycles    ; ramp
discrete:

; comb of 20 teeth----------------------------------------

; slight vibrato to bind the teeth together
ktrem oscil ifteeth*0.01, 15, 4

; controlling edge is sub-fundamental
kfh1 = ifcomb-ifteeth + ktrem + kcycleramp*ifteeth
axh1    oscil 1, kfh1, 3
; tooth
kfh2 = kfh1 + ifteeth
axh2    oscil 1, kfh2, 3
; tooth
kfh3 = kfh2 + ifteeth
axh3    oscil 1, kfh3, 3
; tooth
kfh4 = kfh3 + ifteeth
axh4    oscil 1, kfh4, 3
; tooth
kfh5 = kfh4 + ifteeth
axh5    oscil 1, kfh5, 3
; tooth
kfh6 = kfh5 + ifteeth
axh6    oscil 1, kfh6, 3
; tooth
kfh7 = kfh6 + ifteeth
axh7    oscil 1, kfh7, 3
; tooth
kfh8 = kfh7 + ifteeth
axh8    oscil 1, kfh8, 3
; tooth
kfh9 = kfh8 + ifteeth
axh9    oscil 1, kfh9, 3
; tooth
kfh10 = kfh9 + ifteeth
axh10    oscil 1, kfh10, 3
; tooth
kfh11 = kfh10 + ifteeth
axh11    oscil 1, kfh11, 3
; tooth
kfh12 = kfh11 + ifteeth
axh12    oscil 1, kfh12, 3
; tooth
kfh13 = kfh12 + ifteeth
axh13    oscil 1, kfh13, 3
; tooth
kfh14 = kfh13 + ifteeth
axh14    oscil 1, kfh14, 3
; tooth
kfh15 = kfh14 + ifteeth
axh15    oscil 1, kfh15, 3
; tooth
kfh16 = kfh15 + ifteeth
axh16    oscil 1, kfh16, 3
; tooth
kfh17 = kfh16 + ifteeth
axh17    oscil 1, kfh17, 3
; tooth
kfh18 = kfh17 + ifteeth
axh18    oscil 1, kfh18, 3
```

```
; tooth
kfh19 = kfh18 + ifteeth
axh19    oscil 1, kfh19, 3
; tooth
kfh20 = kfh19 + ifteeth
axh20    oscil 1, kfh20, 3

ax                                                                                                                =
axh1+axh2+axh3+axh4+axh5+axh6+axh7+axh8+axh9+axh10+axh11+axh12+axh13+axh14+axh15+axh16+axh17+a
xh18+axh19+axh20

; smooth the edges with a formant-------------------------------
dispfft ax, 1, 1024, 0, 0, 0

iformc= 2000
iformw= 1000
ay1reson  ax, iformc, iformw
ay2reson  ay1, iformc, iformw
ayreson  ay2, iformc, iformw

dispfft ay, 1, 1024, 0 ,0 ,0

;-- output                -------------------------------
gatangle    gain ay, kamp
dispfft gatangle, 1, 1024, 0, 0, 1
out     gatangle
endin


;
;################################################################################;#
;#      COPYRIGHT (C) CSIRO DIT CSIS 1993
;#
;#      SOURCE FILE     TWheel.sco
;#
;#      MODULE          Sonify
;#
;#      SYSTEM          PostGrad
;#
;#      ENVIRONMENT     Sun SPARC and SUNOS4.1.2
;#
;#      AUTHOR          Stephen Barrass, CSIS
;#
;#      HISTORY
;#                      : First written March 1994
;################################################################################
;
;-- tables    --------------------------------------------
;
; exponential curve
f2 0 128 5 0.001 128 1

; cosine wave
f3 0 8193 9 1 1 90

; sine wave
f4 0 8193 10 1

;-- rate       --------------------------------------------
;t 0 60

; streaming
; onsets 100 ms (range 50 -> 150 ms) = 60000/onset_in_ms
;t 0 1200; 5 = 50 ms onsets
;t 0 857; 4 = 70 ms onsets
;t 0 750; 80 ms onsets
```

```
;t 0 666; 3 = 90 ms onsets
;t 0 545; 2 = 110 ms onsets
;t 0 461; 1 = 130 ms onsets
;t 0 400; 0 = 150 ms onsets

;// instrument              ////////////////////////////////////////////////
; pitch cycle - after Shephard
; p1p2      p3       p4       p5       p6       p6
; instrstartdur    comb    teeth    loud   ncycles
;                  oct.cl   oct.cl0.0..1.0   0-n

;-- galloping----------------------------------------------
; adjacent
; 0 = 0,5,0
i1  0       1.0      8.00     8.00     1.0       0
i.   +        .      8.05       .        .        .
i.   +        .      8.00       .        .        .
i.   +        .        .        .        0        .
e

; complementary
; 0 = 0,6,0
i1  0       1.0      8.00     8.00     1.0       0
i.   +        .      8.06       .        .        .
i.   +        .      8.00       .        .        .
i.   +        .        .        .        0        .
e
;-- triplets   ----------------------------------------------
; split complementaries
; 0 = 0,8,4
i1  0       1.0      8.00     8.00     1.0       0
i.   +        .      8.08       .        .        .
i.   +        .      8.04       .        .        .
e
; 1 = 3,7,11
i1  0       1.0      8.03     8.00     1.0       0
i.   +        .      8.07       .        .        .
i.   +        .      8.11       .        .        .
e
; 2 = 6,10,2
i1  0       1.0      8.06     8.00     1.0       0
i.   +        .      8.10       .        .        .
i.   +        .      8.02       .        .        .
e
; 3 = 9,5,1
i1  0       1.0      8.09     8.00     1.0       0
i.   +        .      8.05       .        .        .
i.   +        .      8.01       .        .        .
e

;-- triplets   ----------------------------------------------
; adjacent
; 0 =  0,5,10
i1  0       1.0      8.00     8.00     1.0       0
i.   +        .      8.05       .        .        .
i.   +        .      8.10       .        .        .
e
; 3 = 9,2,7
i1  0       1.0      8.09     8.00     1.0       0
i.   +        .      8.02       .        .        .
i.   +        .      8.07       .        .        .
e
; 2 = 6,11,4
i1  0       1.0      8.06     8.00     1.0       0
i.   +        .      8.11       .        .        .
i.   +        .      8.04       .        .        .
```

```
e
; 1 = 3,8,1
i1  0        1.0      8.03     8.00      1.0        0
i.   +         .      8.08        .         .          .
i.   +         .      8.01        .         .          .
e

;-- circle of categories------------------------------------------------
; cycle of fifths complementary
i1  0        1.0      8.00     8.00      1.0        0
i.   +         .      8.06        .         .          .
i.   +         .      8.05        .         .          .
i.   +         .      8.11        .         .          .
i.   +         .      8.10        .         .          .
i.   +         .      8.04        .         .          .
i.   +         .      8.03        .         .          .
i.   +         .      8.09        .         .          .
i.   +         .      8.08        .         .          .
i.   +         .      8.02        .         .          .
i.   +         .      8.01        .         .          .
i.   +         .      8.07        .         .          .

e

;-- circle of categories------------------------------------------------
; cycle of fifths adjacent
i1  0        1.0      8.00     8.00      1.0        0
i.   +         .      8.05        .         .          .
i.   +         .      8.10        .         .          .
i.   +         .      8.03        .         .          .
i.   +         .      8.08        .         .          .
i.   +         .      8.01        .         .          .
i.   +         .      8.06        .         .          .
i.   +         .      8.11        .         .          .
i.   +         .      8.04        .         .          .
i.   +         .      8.09        .         .          .
i.   +         .      8.02        .         .          .
i.   +         .      8.07        .         .          .
e

;-- continuous -----------------------------------------------
i1  0       10.0     8.00      8.0       1.0        2
e

;-- discrete  ----------------------------------------------
; pitch order
i1  0        1.0      8.00     8.00      1.0        0
i.   +         .      8.01        .         .          .
i.   +         .      8.02        .         .          .
i.   +         .      8.03        .         .          .
i.   +         .      8.04        .         .          .
i.   +         .      8.05        .         .          .
i.   +         .      8.06        .         .          .
i.   +         .      8.07        .         .          .
i.   +         .      8.08        .         .          .
i.   +         .      8.09        .         .          .
i.   +         .      8.10        .         .          .
i.   +         .      8.11        .         .          .
e
```

Static Timbre, Formant and Timbre Circle

```
;
;#################################################################################;#
;#COPYRIGHT (C) CSIRO DIT CSIS 1993
;#
;#SOURCE FILEGreySun.orc
```

```
;#
;#MODULE      Sonify
;#
;#SYSTEM     PostGrad
;#
;#ENVIRONMENTSun SPARC and SUNOS4.1.2
;#
;#AUTHORStephen Barrass, CSIS
;#
;#HISTORY
;#               : First written March 1994
;################################################################################
;
;
;

sr = 44100
kr = 4410
ksmps = 10
nchnls = 1

gifnyq = sr/2.1    ; nyquist freq + some spare (as per Moore)
giattack  = 0.01; time buffer (in seconds) to prevent clicks
gidecay   = 0.01; time buffer (in seconds) to prevent clicks
giampmax = 90; max amplitude in dB

; gamut stuff----------------------------------------------------
; notenum gamut ranges are in the score tables
giminp = 1; patch min
gimaxp = 8; patch max
giminf = 0; filter min
gimaxf = 127; filter max

gioutofgamut = 600; out-of-gamut sample
gioogamutp = 601; out-of-patch gamut sample
gioogamutn = 602; out-of-notenumber gamut sample
gioogamutf = 603; out-of-filter gamut sample
gisinfunc = 604; sine function
gibuzzfunc = 605; buzz function

gicountbase = 610; words for counting


;# Device               ####################################
; CSOUND SAMPLE PLAYER
; this instrument plays samples
; patch selects the sample bank
; notenum selects the closest sample from the bank
;
;// Instrument params////////////////////////////////////////
;   p1// in : instrument 1
;   p2// in : start time
;
instr1
print p1, p2, p3, p4, p5, p6, p7
idur    =p3   ;// in : duration
ipatch =p4+1;// in : patch number0..127
inotenum =p5;// in : notenum     0..127
ifilter =p6      ;// in : filter          0..127
ivelocity = p7;// in : velocity0..127

;-- velocity ---------------------------------------------
iamp    = ampdb(ivelocity*giampmax/127); loudness max
;kamp linenr iamp, giattack, 0, .1  ;at noteoff, extend by 100 millisecs
kamplinen iamp, giattack, idur, gidecay
print iamp
```

```
;-- notenum---------------------------------------------
; calculate froot
; notenum 60 = middle C = csound oct 8.0
; range is from midi C0 to midi G10

ioct=3.0 + inotenum/12.0
ifroot=cpsoct(ioct)
print ioct, ifroot

;-- filter    ---------------------------------------------
if (ifilter > gimaxf) goto oogamutf
if (ifilter < giminf) goto oogamutf
; number of harmonics
; range is 0-32
inharm= ifilter/4.0
print inharm

; low pass filter with cutoff moving up the harmonics
ifcut =ifroot+ifroot*inharm

;check against nyquist - cut at nyquist if necessary
ifcut = (ifcut < gifnyq ? ifcut : gifnyq)
inharmnyq = gifnyq/ifroot
print gifnyq, ifcut, inharmnyq

;-- patch    ---------------------------------------------
print ipatch
if (ipatch == 500) goto refharmonic
if (ipatch == 501) goto refnoise
if (ipatch == 502) goto refacum
if (ipatch == 503) goto reftone
if (ipatch == gioogamutp) goto oogamutp
if (ipatch > gicountbase) goto count
; patch gamut range check for normal instruments
if (ipatch < giminp) goto oogamutp
if (ipatch > gimaxp) goto oogamutp

;---------
sample:
; get the sample bank index
;ibankindextableipatch, 1
; get the sample for this pitch
;isample tableinotenum, ibankindex
isample  tableinotenum, ipatch
print isample, ipatch, inotenum
if (isample == gioogamutn) goto oogamutn
asrcloscilkamp, ifroot, isample
goto filter

;----------
reftone:
; sine tone
afinoscilkamp, ifroot, gisinfunc
goto output

;----------
refharmonic:
; Harmonic reference instrument
asrcoscilkamp, ifroot, gibuzzfunc
goto filter

;----------
refnoise:
; Noise reference instrument
asrcrandhkamp, gifnyq
```

```
goto filter

;----------
refacum:
; Acum reference instrument
; get the filter position
ifcut = p6 * gifnyq/127
; critical bandwidth ~ 0.2*f for f > 500Hz
;                  100Hz for f < 500Hz
;anoiserandifcut*0.2
asrcrandhkamp, gifnyq
; centre the critical-bandlimited noise around the cutoff
;afinosciliamp,  ifcut+anoise
iwidth = (ifcut < 500 ? 100 : ifcut*0.2)
afinresonasrc, ifcut, iwidth
goto output

;----------
oogamutp:
; out of patch gamut
afinlosciliamp, 200, gioogamutp
goto output

;----------
oogamutn:
; out of notenumber gamut
afinlosciliamp, 200, gioogamutn
goto output

;----------
oogamutf:
; out of filter gamut
afinlosciliamp, 200, gioogamutf
goto output

;----------
count:
; count instrument
asrclosciliamp, 200, ipatch
; don’t apply the filter to words
afin = asrc
goto output

; apply the brightness filter-----------------------------------------------
filter:
dispfftasrc, 0.1, 1024; DEBUG - look at signal

; use lowpass
afilttoneasrc, ifcut
afilttoneafilt, ifcut
afilttoneafilt, ifcut
afintoneafilt, ifcut
afinbalanceafilt, asrc
;
;-- output the sound-----------------------------------------------
output:
dispfftafin, 0.1, 1024; DEBUG - look at signal
outafin
endin



;--------- flu
f1 0 128 -7 602 48 602 0 20 36 56 0 602 44 602
f20 0 0 1 “flu_48.aif” 0 4 0
f21 0 0 1 “flu_49.aif” 0 4 0
```

```
f22 0 0 1 “flu_50.aif” 0 4 0
f23 0 0 1 “flu_51.aif” 0 4 0
f24 0 0 1 “flu_52.aif” 0 4 0
f25 0 0 1 “flu_53.aif” 0 4 0
f26 0 0 1 “flu_54.aif” 0 4 0
f27 0 0 1 “flu_55.aif” 0 4 0
f28 0 0 1 “flu_56.aif” 0 4 0
f29 0 0 1 “flu_57.aif” 0 4 0
f30 0 0 1 “flu_58.aif” 0 4 0
f31 0 0 1 “flu_59.aif” 0 4 0
f32 0 0 1 “flu_60.aif” 0 4 0
f33 0 0 1 “flu_61.aif” 0 4 0
f34 0 0 1 “flu_62.aif” 0 4 0
f35 0 0 1 “flu_63.aif” 0 4 0
f36 0 0 1 “flu_64.aif” 0 4 0
f37 0 0 1 “flu_65.aif” 0 4 0
f38 0 0 1 “flu_66.aif” 0 4 0
f39 0 0 1 “flu_67.aif” 0 4 0
f40 0 0 1 “flu_68.aif” 0 4 0
f41 0 0 1 “flu_69.aif” 0 4 0
f42 0 0 1 “flu_70.aif” 0 4 0
f43 0 0 1 “flu_71.aif” 0 4 0
f44 0 0 1 “flu_72.aif” 0 4 0
f45 0 0 1 “flu_73.aif” 0 4 0
f46 0 0 1 “flu_74.aif” 0 4 0
f47 0 0 1 “flu_75.aif” 0 4 0
f48 0 0 1 “flu_76.aif” 0 4 0
f49 0 0 1 “flu_77.aif” 0 4 0
f50 0 0 1 “flu_78.aif” 0 4 0
f51 0 0 1 “flu_79.aif” 0 4 0
f52 0 0 1 “flu_80.aif” 0 4 0
f53 0 0 1 “flu_81.aif” 0 4 0
f54 0 0 1 “flu_82.aif” 0 4 0
f55 0 0 1 “flu_83.aif” 0 4 0
f56 0 0 1 “flu_84.aif” 0 4 0


;--------- cel
f2 0 128 -7 602 24 602 0 67 34 101 0 602 70 602
f67 0 0 1 “cel_24.aif” 0 4 0
f68 0 0 1 “cel_25.aif” 0 4 0
f69 0 0 1 “cel_26.aif” 0 4 0
f70 0 0 1 “cel_27.aif” 0 4 0
f71 0 0 1 “cel_28.aif” 0 4 0
f72 0 0 1 “cel_29.aif” 0 4 0
f73 0 0 1 “cel_30.aif” 0 4 0
f74 0 0 1 “cel_31.aif” 0 4 0
f75 0 0 1 “cel_32.aif” 0 4 0
f76 0 0 1 “cel_33.aif” 0 4 0
f77 0 0 1 “cel_34.aif” 0 4 0
f78 0 0 1 “cel_35.aif” 0 4 0
f79 0 0 1 “cel_36.aif” 0 4 0
f80 0 0 1 “cel_37.aif” 0 4 0
f81 0 0 1 “cel_38.aif” 0 4 0
f82 0 0 1 “cel_39.aif” 0 4 0
f83 0 0 1 “cel_40.aif” 0 4 0
f84 0 0 1 “cel_41.aif” 0 4 0
f85 0 0 1 “cel_42.aif” 0 4 0
f86 0 0 1 “cel_43.aif” 0 4 0
f87 0 0 1 “cel_44.aif” 0 4 0
f88 0 0 1 “cel_45.aif” 0 4 0
f89 0 0 1 “cel_46.aif” 0 4 0
f90 0 0 1 “cel_47.aif” 0 4 0
f91 0 0 1 “cel_48.aif” 0 4 0
f92 0 0 1 “cel_49.aif” 0 4 0
f93 0 0 1 “cel_50.aif” 0 4 0
```

f94 0 0 1 “cel_51.aif” 0 4 0
f95 0 0 1 “cel_52.aif” 0 4 0
f96 0 0 1 “cel_53.aif” 0 4 0
f97 0 0 1 “cel_54.aif” 0 4 0
f98 0 0 1 “cel_55.aif” 0 4 0
f99 0 0 1 “cel_56.aif” 0 4 0
f100 0 0 1 “cel_57.aif” 0 4 0
f101 0 0 1 “cel_58.aif” 0 4 0

;--------- cla
f3 0 128 -7 602 25 602 0 112 24 136 0 602 79 602
f112 0 0 1 “cla_25.aif” 0 4 0
f113 0 0 1 “cla_26.aif” 0 4 0
f114 0 0 1 “cla_27.aif” 0 4 0
f115 0 0 1 “cla_28.aif” 0 4 0
f116 0 0 1 “cla_29.aif” 0 4 0
f117 0 0 1 “cla_30.aif” 0 4 0
f118 0 0 1 “cla_31.aif” 0 4 0
f119 0 0 1 “cla_32.aif” 0 4 0
f120 0 0 1 “cla_33.aif” 0 4 0
f121 0 0 1 “cla_34.aif” 0 4 0
f122 0 0 1 “cla_35.aif” 0 4 0
f123 0 0 1 “cla_36.aif” 0 4 0
f124 0 0 1 “cla_37.aif” 0 4 0
f125 0 0 1 “cla_38.aif” 0 4 0
f126 0 0 1 “cla_39.aif” 0 4 0
f127 0 0 1 “cla_40.aif” 0 4 0
f128 0 0 1 “cla_41.aif” 0 4 0
f129 0 0 1 “cla_42.aif” 0 4 0
f130 0 0 1 “cla_43.aif” 0 4 0
f131 0 0 1 “cla_44.aif” 0 4 0
f132 0 0 1 “cla_45.aif” 0 4 0
f133 0 0 1 “cla_46.aif” 0 4 0
f134 0 0 1 “cla_47.aif” 0 4 0
f135 0 0 1 “cla_48.aif” 0 4 0
f136 0 0 1 “cla_49.aif” 0 4 0

;--------- tsx
f4 0 128 -7 602 36 602 0 147 13 160 0 602 79 602
f147 0 0 1 “tsx_36.aif” 0 4 0
f148 0 0 1 “tsx_37.aif” 0 4 0
f149 0 0 1 “tsx_38.aif” 0 4 0
f150 0 0 1 “tsx_39.aif” 0 4 0
f151 0 0 1 “tsx_40.aif” 0 4 0
f152 0 0 1 “tsx_41.aif” 0 4 0
f153 0 0 1 “tsx_42.aif” 0 4 0
f154 0 0 1 “tsx_43.aif” 0 4 0
f155 0 0 1 “tsx_44.aif” 0 4 0
f156 0 0 1 “tsx_45.aif” 0 4 0
f157 0 0 1 “tsx_46.aif” 0 4 0
f158 0 0 1 “tsx_47.aif” 0 4 0
f159 0 0 1 “tsx_48.aif” 0 4 0
f160 0 0 1 “tsx_49.aif” 0 4 0

;--------- ssx
f5 0 128 -7 602 61 602 0 171 14 185 0 602 53 602
f171 0 0 1 “ssx_61.aif” 0 4 0
f172 0 0 1 “ssx_62.aif” 0 4 0
f173 0 0 1 “ssx_63.aif” 0 4 0
f174 0 0 1 “ssx_64.aif” 0 4 0
f175 0 0 1 “ssx_65.aif” 0 4 0
f176 0 0 1 “ssx_66.aif” 0 4 0
f177 0 0 1 “ssx_67.aif” 0 4 0

f178 0 0 1 “ssx_68.aif” 0 4 0
f179 0 0 1 “ssx_69.aif” 0 4 0
f180 0 0 1 “ssx_70.aif” 0 4 0
f181 0 0 1 “ssx_71.aif” 0 4 0
f182 0 0 1 “ssx_72.aif” 0 4 0
f183 0 0 1 “ssx_73.aif” 0 4 0
f184 0 0 1 “ssx_74.aif” 0 4 0
f185 0 0 1 “ssx_75.aif” 0 4 0

;--------- ehn
f6 0 128 -7 602 40 602 0 196 30 226 0 602 58 602
f196 0 0 1 “ehn_40.aif” 0 4 0
f197 0 0 1 “ehn_41.aif” 0 4 0
f198 0 0 1 “ehn_42.aif” 0 4 0
f199 0 0 1 “ehn_43.aif” 0 4 0
f200 0 0 1 “ehn_44.aif” 0 4 0
f201 0 0 1 “ehn_45.aif” 0 4 0
f202 0 0 1 “ehn_46.aif” 0 4 0
f203 0 0 1 “ehn_47.aif” 0 4 0
f204 0 0 1 “ehn_48.aif” 0 4 0
f205 0 0 1 “ehn_49.aif” 0 4 0
f206 0 0 1 “ehn_50.aif” 0 4 0
f207 0 0 1 “ehn_51.aif” 0 4 0
f208 0 0 1 “ehn_52.aif” 0 4 0
f209 0 0 1 “ehn_53.aif” 0 4 0
f210 0 0 1 “ehn_54.aif” 0 4 0
f211 0 0 1 “ehn_55.aif” 0 4 0
f212 0 0 1 “ehn_56.aif” 0 4 0
f213 0 0 1 “ehn_57.aif” 0 4 0
f214 0 0 1 “ehn_58.aif” 0 4 0
f215 0 0 1 “ehn_59.aif” 0 4 0
f216 0 0 1 “ehn_60.aif” 0 4 0
f217 0 0 1 “ehn_61.aif” 0 4 0
f218 0 0 1 “ehn_62.aif” 0 4 0
f219 0 0 1 “ehn_63.aif” 0 4 0
f220 0 0 1 “ehn_64.aif” 0 4 0
f221 0 0 1 “ehn_65.aif” 0 4 0
f222 0 0 1 “ehn_66.aif” 0 4 0
f223 0 0 1 “ehn_67.aif” 0 4 0
f224 0 0 1 “ehn_68.aif” 0 4 0
f225 0 0 1 “ehn_69.aif” 0 4 0
f226 0 0 1 “ehn_70.aif” 0 4 0

;--------- bsn
f7 0 128 -7 602 22 602 0 237 31 268 0 602 75 602
f237 0 0 1 “bsn_22.aif” 0 4 0
f238 0 0 1 “bsn_23.aif” 0 4 0
f239 0 0 1 “bsn_24.aif” 0 4 0
f240 0 0 1 “bsn_25.aif” 0 4 0
f241 0 0 1 “bsn_26.aif” 0 4 0
f242 0 0 1 “bsn_27.aif” 0 4 0
f243 0 0 1 “bsn_28.aif” 0 4 0
f244 0 0 1 “bsn_29.aif” 0 4 0
f245 0 0 1 “bsn_30.aif” 0 4 0
f246 0 0 1 “bsn_31.aif” 0 4 0
f247 0 0 1 “bsn_32.aif” 0 4 0
f248 0 0 1 “bsn_33.aif” 0 4 0
f249 0 0 1 “bsn_34.aif” 0 4 0
f250 0 0 1 “bsn_35.aif” 0 4 0
f251 0 0 1 “bsn_36.aif” 0 4 0
f252 0 0 1 “bsn_37.aif” 0 4 0
f253 0 0 1 “bsn_38.aif” 0 4 0
f254 0 0 1 “bsn_39.aif” 0 4 0
f255 0 0 1 “bsn_40.aif” 0 4 0

```
f256 0 0 1 “bsn_41.aif” 0 4 0
f257 0 0 1 “bsn_42.aif” 0 4 0
f258 0 0 1 “bsn_43.aif” 0 4 0
f259 0 0 1 “bsn_44.aif” 0 4 0
f260 0 0 1 “bsn_45.aif” 0 4 0
f261 0 0 1 “bsn_46.aif” 0 4 0
f262 0 0 1 “bsn_47.aif” 0 4 0
f263 0 0 1 “bsn_48.aif” 0 4 0
f264 0 0 1 “bsn_49.aif” 0 4 0
f265 0 0 1 “bsn_50.aif” 0 4 0
f266 0 0 1 “bsn_51.aif” 0 4 0
f267 0 0 1 “bsn_52.aif” 0 4 0
f268 0 0 1 “bsn_53.aif” 0 4 0


;--------- tbn
f8 0 128 -7 602 28 602 0 279 32 311 0 602 68 602
f279 0 0 1 “tbn_28.aif” 0 4 0
f280 0 0 1 “tbn_29.aif” 0 4 0
f281 0 0 1 “tbn_30.aif” 0 4 0
f282 0 0 1 “tbn_31.aif” 0 4 0
f283 0 0 1 “tbn_32.aif” 0 4 0
f284 0 0 1 “tbn_33.aif” 0 4 0
f285 0 0 1 “tbn_34.aif” 0 4 0
f286 0 0 1 “tbn_35.aif” 0 4 0
f287 0 0 1 “tbn_36.aif” 0 4 0
f288 0 0 1 “tbn_37.aif” 0 4 0
f289 0 0 1 “tbn_38.aif” 0 4 0
f290 0 0 1 “tbn_39.aif” 0 4 0
f291 0 0 1 “tbn_40.aif” 0 4 0
f292 0 0 1 “tbn_41.aif” 0 4 0
f293 0 0 1 “tbn_42.aif” 0 4 0
f294 0 0 1 “tbn_43.aif” 0 4 0
f295 0 0 1 “tbn_44.aif” 0 4 0
f296 0 0 1 “tbn_45.aif” 0 4 0
f297 0 0 1 “tbn_46.aif” 0 4 0
f298 0 0 1 “tbn_47.aif” 0 4 0
f299 0 0 1 “tbn_48.aif” 0 4 0
f300 0 0 1 “tbn_49.aif” 0 4 0
f301 0 0 1 “tbn_50.aif” 0 4 0
f302 0 0 1 “tbn_51.aif” 0 4 0
f303 0 0 1 “tbn_52.aif” 0 4 0
f304 0 0 1 “tbn_53.aif” 0 4 0
f305 0 0 1 “tbn_54.aif” 0 4 0
f306 0 0 1 “tbn_55.aif” 0 4 0
f307 0 0 1 “tbn_56.aif” 0 4 0
f308 0 0 1 “tbn_57.aif” 0 4 0
f309 0 0 1 “tbn_58.aif” 0 4 0
f310 0 0 1 “tbn_59.aif” 0 4 0
f311 0 0 1 “tbn_60.aif” 0 4 0



;-- OUT-of-gamut------------------------------------------------------
f600 0 0 1 “out.aif” 0 4 0; out of gamut
f601 0 0 1 “outp.aif” 0 4 0; out of patch gamut
f602 0 0 1 “outn.aif” 0 4 0; out of notenum gamut
f603 0 0 1 “outf.aif” 0 4 0; out of filter gamut
f604 0 2049 11 1; sinusoid
f605 0 2049 11 32; harmonics



;-- Count-------------------------------------------------------
f610 0 0 1 “zero.aif” 0 4 0
f611 0 0 1 “one.aif” 0 4 0
```

```
f612 0 0 1 “two.aif” 0 4 0
f613 0 0 1 “three.aif” 0 4 0
f614 0 0 1 “four.aif” 0 4 0
f615 0 0 1 “five.aif” 0 4 0
f616 0 0 1 “six.aif” 0 4 0
f617 0 0 1 “seven.aif” 0 4 0
f618 0 0 1 “eight.aif” 0 4 0

;-- rate     ------------------------------------------------------------
;t 0 60        ; 1000 ms = 1s onsets

; streaming
; typical range is 50 to 150 ms repeat (from Bregman)
; t (beats/minute) =  60000 (ms/minute)/(repeat time) ms
;t 0 1200; 50 ms onsets
;t 0 1000; 60 ms onsets
;t 0 857; 70 ms onsets
;t 0 750; 80 ms onsets
;t 0 666; 90 ms onsets
t 0 600; 100 ms onsets
;t 0 545; 110 ms onsets
;t 0 461; 130 ms onsets
;t 0 400; 150 ms onsets
;t 0 333; 180 ms onsets
;t 0 300; 200 ms onsets
;t 0 240; 250 ms onsets
;t 0 218; 275 ms onsets
;t 0 200; 300 ms onsets
;t 0 150; 400 ms onsets
;t 0 120; 500 ms onsets

;------- add score events below this line --------------------------------------
; parameters
;device startdurpatchnotenumfiltervelocity
;1beats   beats   0..127   0..127   0..127   0..127

; max duration = 1s
; duration range
;0.2 - 2s = duration
; 0.2 - 0.05 s = streaming
; 0.01-0.05 = clicks

; vertical pitch  gallop ----------------------------------------------

; far
i1  0     1     6     32    127   127
i.  +     .     .     52    .     .
i.  +     .     .     32    .     .
i.  +     .     .     .     .     0
e

; near
i1  0     1     6     47    127   127
i.  +     .     .     52    .     .
i.  +     .     .     47    .     .
i.  +     .     .     .     .     0
e

; mid
i1  0     1     6     42    127   127
i.  +     .     .     52    .     .
i.  +     .     .     42    .     .
i.  +     .     .     .     .     0
e

; vertical pitch triplet ---------------------------------------------
```

```
; 5%
i1  0     1     3     38    127   127
i.  +     .     .     >     .     .
i.  +     .     .     48    .     .
i.  +     .     .     .     .     0
e

;-- brightness zeros+outlier---------------------------------------------
i1  0     1.0   0     48    0     127
i.  +     .     1     .     .     .
i.  +     .     2     .     .     .
i.  +     .     3     .     .     .
i.  +     .     5     .     .     .
i.  +     .     2     .     127   .
i.  +     .     6     .     0     .
i.  +     .     7     .     .     .
e

;-- brightness spiral-----------------------------------------------
i1  0     1.0   0     48    127   127
i.  +     .     1     .     .     .
i.  +     .     2     .     .     .
i.  +     .     3     .     0     .
i.  +     .     5     .     .     .
i.  +     .     6     .     .     .
i.  +     .     7     .     127   .
e

;-- brightness spiral triplet -----------------------------------------------
i1  0     1.0   1     48    0     127
i.  +     .     2     .     >     .
i.  +     .     3     .     80    .
e

;-- brightness bilateral ----------------------------------------------
i1  0     1.0   2     48    0     127
i.  +     .     2     .     127   .
i.  +     .     6     .     0     .
i.  +     .     6     .     127   .
e

; duration triplet-----------------------------------------------

; 25%
i1  0     0.25  3     48    127   127
i.  1     0.5   .     .     .     .
i.  2     0.75  .     .     .     .
i.  2.9   0.1   .     .     .     0
e

; 5%
i1  0     0.5   3     48    127   127
i.  1     0.55  .     .     .     .
i.  2     0.6   .     .     .     .
i.  2.9   0.1   .     .     .     0
e

; duration sequence----------------------------------------------

; 8 steps
i1  0     0.1   3     48    127   0
i.  1     0.1   .     .     .     127
i.  2     0.2   .     .     .     .
i.  3     0.3   .     .     .     .
i.  4     0.4   .     .     .     .
```

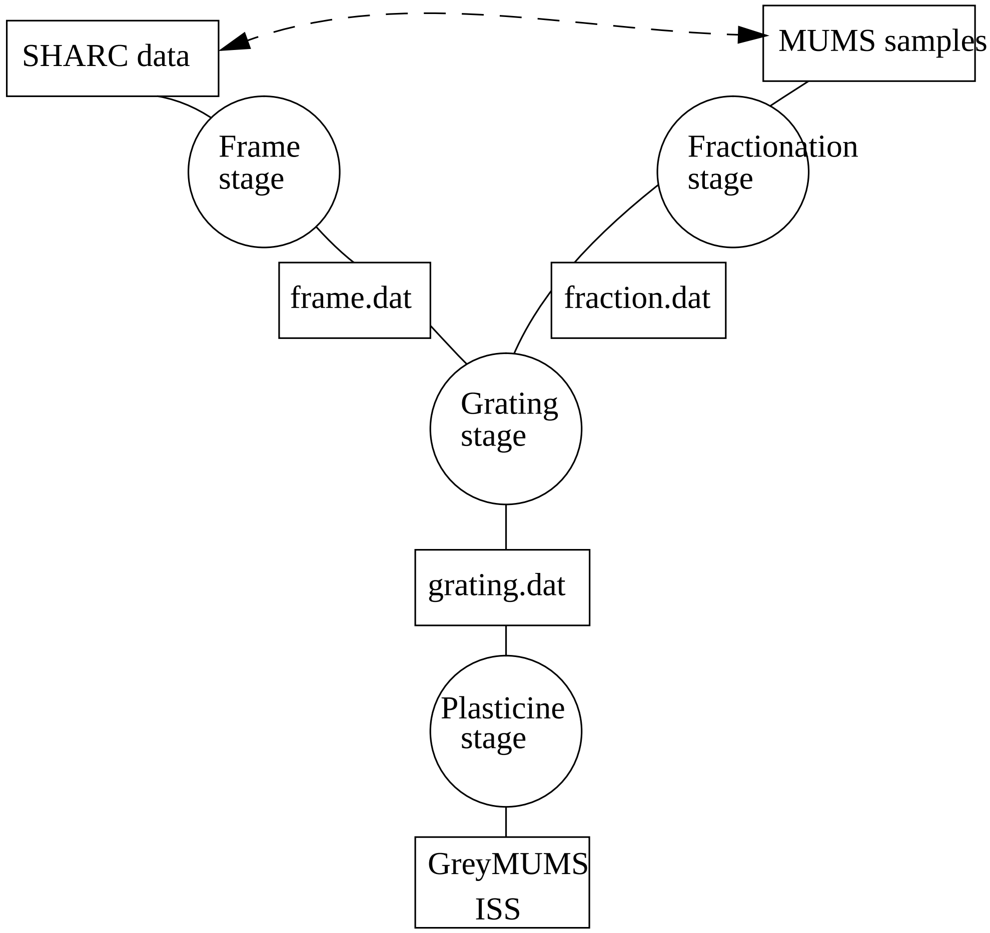
```
i.  5     0.5    .     .     .     .
i.  6     0.6    .     .     .     .
i.  7     0.7    .     .     .     .
i.  7.9   0.1    .     .     .     0
e

; duration gallop----------------------------------------------

; close
i1  0     0.5    3     48    127   127
i.  1     0.6    .     .     .     .
i.  2     0.5    .     .     .     .
i.  3     .      .     .     .     0
e

; far
i1  0     0.1    3     48    127   127
i.  1     1.0    .     .     .     .
i.  2     0.1    .     .     .     .
i.  3     .      .     .     .     0
e

; mid
i1  0     0.5    3     48    127   127
i.  1     1.0    .     .     .     .
i.  2     0.5    .     .     .     .
i.  3     0.5    .     .     .     0
e


;-- loudness gallop----------------------------------------------
; far
i1  0     1.0    3     48    127   50
i.  +     .      .     .     .     100
i.  +     .      .     .     .     50
i.  +     .      .     .     .     0
e

; mid
i1  0     1.0    3     48    127   60
i.  +     .      .     .     .     80
i.  +     .      .     .     .     60
i.  +     .      .     .     .     0
e
; close
i1  0     1.0    3     48    127   70
i.  +     .      .     .     .     76
i.  +     .      .     .     .     70
i.  +     .      .     .     .     0
e


; loudness triplet 5%----------------------------------------------
i1  0     1.0    4     65    127   80
i.  +     .      .     .     .     86
i.  +     .      .     .     .     92
e

;-- brightness triplet----------------------------------------------

; 25%,50%,75%
i1  0     1.0    7     48    32    127
i.  +     .      .     .     64    .
i.  +     .      .     .     96    .
e
```

```
;-- brightness gallop-----------------------------------------------

; mid
i1  0      1.0     4      65     70     127
i.   +       .        .       .      10       .
i.   +       .        .       .      70       .
i.   +       .        .       .       .        0
e

; close
i1  0      1.0     4      65     40     127
i.   +       .        .       .      45       .
i.   +       .        .       .      40       .
i.   +       .        .       .       .        0
e

; far
i1  0      1.0     4      65     10     127
i.   +       .        .       .     127       .
i.   +       .        .       .      10       .
i.   +       .        .       .       .        0
e


;-- brightness sequence-----------------------------------------------
; filter = 0..127
i1  0      1.0     3      48      0     127
i.   +       .        .       .       >       .
i.   +       .        .       .       >       .
i.   +       .        .       .       >       .
i.   +       .        .       .       >       .
i.   +       .        .       .       >       .
i.   +       .        .       .       >       .
i.   +       .        .       .       >       .
i.   +       .        .       .     127       .
e

;-- timbre   ----------------------------------------------
; galloping
; adjacent
; 0 = 2,3,2_
i1  0      1.0     2      48     100    127
i.   1       .       3       .       .       .
i.   2       .       2       .       .       .
i.   3       .       .       .       .      0.0
e
; complementary
; 0 = 1,5,1_
i1  0      1.0     1      48     100    127
i.   1       .       5       .       .       .
i.   2       .       1       .       .       .
i.   3       .       .       .       .      0.0
e


;-- triplets   ----------------------------------------------
; complementary
; 1 = 2,5,7
i1  0      1.0     2      48     100    127
i.   +       .       5       .       .       .
i.   +       .       7       .       .       .
e
; 2 = 4,7,1
;i1 0      1.0     4      48     100    127
;i.  +       .       7       .       .       .
```

```
;i.  +    .    1    .    .    .
;e
; 3 = 6,1,3
i1  0    1.0    6    48    100    127
i.   +    .    1    .    .    .
i.   +    .    3    .    .    .
e
; 0 = 0,3,5
i1  0    1.0    0    48    100    127
i.   +    .    3    .    .    .
i.   +    .    5    .    .    .
e


;-- triplets   ---------------------------------------------
; adjacent
; 3 = 5,6,7
i1  0    1.0    5    48    100    127
i.   +    .    6    .    .    .
i.   +    .    7    .    .    .
e
; 0 = 7,0,1
i1  0    1.0    7    48    100    127
i.   +    .    0    .    .    .
i.   +    .    1    .    .    .
e
; 1 = 1,2,3
i1  0    1.0    1    48    100    127
i.   +    .    2    .    .    .
i.   +    .    3    .    .    .
e
; 2 = 3,4,5
;i1 0    1.0    3    48    100    127
;i.  +    .    4    .    .    .
;i.  +    .    5    .    .    .
;e

;-- circle     ---------------------------------------------
i1  0    1.0    0    48    100    127
i.   +    .    1    .    .    .
i.   +    .    2    .    .    .
i.   +    .    3    .    .    .
;i.  +    .    4    .    .    .
i.   +    .    5    .    .    .
i.   +    .    6    .    .    .
i.   +    .    7    .    .    .
e
```

Appendix 8-1: Building an ISS

This appendix describes the process and tools used to build the GreyMUMS perceptually linearised information sound space. The process has the following stages, which are described in the following subsections.



Appendix 8-2: SHARC and MUMS

Written and maintained by Gregory J. Sandell, Sussex University
Release 0.90("beta"), November 1994

Contents of this README file

WHAT IS SHARC?

SHARC is a database of musical timbre information by Gregory Sandell. It stands for "Sandell Harmonic Archive." People for whom this dataset may be useful are Acousticians, Psychoacousticians, researchers in Music Perception and Cognition, researchers in Digital Signal Processing, Music Theorists, and Musicologists.
Over 1300 different notes have been analysed. Complete chromatic runs from the standard playing range of essentially all the non-percussive instruments of the modern orchestra have been included; for example, individual analyses of 32 different oboe notes (the chromatic scale from the pitches a#3 to f6) are available.
For each note, a short portion corresponding to the sustain or "steady state" portion of the tone was selected and analysed with a Fourier analysis. Each analysis consists of a list of amplitudes and phases for all the note's harmonics in the range 0-10,000 Hz.
The source of the musical notes were the orchestral tones from the McGill University Master Samples (MUMS) Compact Discs. These are digital recordings of live musical performers.

RELEASE VERSION, AVAILABILITY, AUTHOR INFORMATION

Version 0.90 ("beta", November 1994) is the first-ever release. The reason I am calling it a "beta" is that there are a number of changes which may occur in the short term: the name of the archive, the internet location or URL, availability of the database from a US ftp site, the availability of online graphics to view the database, and adding more instruments. Any of these changes will be announced in all the relevant internet bulletin boards and distributed email lists. Your comments are most welcome, especially now, as they will have the most impact on the nature of the database. Send email to sandell@epunix.sussex.ac.uk.
Here are instructions for installing the archive on a UNIX platform.

- Go to the directory where you want to install SHARC. You will need about 10 free megabytes to perform the installation. Once finished, the archive will occupy 4.5 megabytes space.
- ftp to ftp.ep.susx.ac.uk

- Enter user name "anonymous" and give your full email address as a password
- Type "binary" to set the transfer to binary mode
- Type "get sharc.tar.Z"
- When the file is finished transferring (it is about 0.8 megabytes large), leave ftp by typing "quit"
- Type "uncompress sharc.tar.Z". The result of this is that sharc.tar.Z will be replaced by sharc.tar
- Type "tar xf sharc.tar". This will put the archive in a directory called "sharc".
- Dispose of the sharc.tar file (type "rm sharc.tar") to reclaim about 4.5 megabytes space.

Gregory J. Sandell is a research fellow in the Hearing Research Group at the University of Sussex Experimental Psychology department in the U.K. Starting April 1995 I will be at Parmly Hearing Institute at Loyola University in Chicago, IL (USA), but my Sussex email address (sandell@epunix.sussex.ac.uk) will remain active for a while.

PERMISSIONS

The data contained in this directory are available for free with the following restrictions:

- It may be used and shared on a non-profit basis only. You may not sell it for money or use it for trade; you may not bundle it with a commercial product or use it to attract customers to buy a product.
- It must be identified as "SHARC" with Gregory Sandell identified as its author. The author appreciates proper acknowledgement when this data is referred to in published papers.

CONTENTS

The database consists of 39 directories, each corresponding to a particular instrument. Each directory consists of separate files for each note analysed for that instrument. The instruments are:

Bach_trumpet	bass_clarinet	altoflute_vibrato
Bb_clarinet	bass_trombone	piccolo
CB	bassflute_vibrato	trombone
CB_martele	bassoon	trombone_muted
CB_muted	tuba	violin_vibrato
CB_pizz	cello_martele	viola_martele
C_trumpet	cello_muted_vibrato	viola_muted_vibrato
C_trumpet_muted	cello_pizzicato	viola_pizzicato
Eb_clarinet	cello_vibrato	viola_vibrato
English_horn	contrabass_clarinet	violin_martele

French_horn	contrabassoon	violin_muted_vibrato
French_horn_muted	flute_vibrato	violin_pizzicato
alto_trombone	oboe	violinensemb

INFORMATION ABOUT THE SOURCE SOUNDS

The McGill University Master Samples (MUMS) is a library of compact discs made and sold by Frank Opolko and Joel Wapnick of McGill University. To obtain the CDs or obtain information about them, write to:

McGill University
Faculty of Music
555 Sherbrooke Street West
Montrel, Quebec
Canada H3A 1E3
phone: (514) 398-4548
email: CXJW@MUSICA.MCGILL.CA (Joel Wapnick)

The naming of instruments, the organisation of the directories, and the information in each directory's CONTENTS files reflects the contents of the MUMS CDs as precisely as possible. I would like to gratefully acknowledge the permission of the makers of MUMS use their product in this manner. Other than customer, I have no financial relationship to MUMS or McGill University.

HOW IS THIS DATA USEFUL?

One of the most mportant aspects of a musical instrument sound that determines its timbre are the spectrum of its steady state portion. Other critical features are the rapid spectral changes at attack and decay time, and slowly varying changes in spectrum during the steady state. The fact that timbre depends so critically on the latter three aspects makes the study of timbre a challenge, because of the increased complexity of including the temporal dimension. To create a database of all the instruments of the orchestra with complete spectrotemporal descriptions of individual notes is not currently feasible, not as an archive to be shared through current network means, at least; several gigabytes, or a few CD-ROM discs would be required. A library of steady state spectra, however, is feasable. Admittedly, the study of steady state spectra is nothing new. However, prior to the use of computer analysis of sound, spectral analyses were expensive and hard-won operations; the unfortunate practice of analysing one note and drawing conclusions about the entire instrument's timbre was sometimes seen in manuals of acoustics from that time. The balance struck in this collection between economy of representation (steady state spectra) and completeness (complete chromatic scales for each instrument) offers researchers new opportunity for timbral discovery. Specifically, it puts the study of the "macro timbre" of an instrument, i.e. its spectral content of its entire pitch range, within the grasp of the researcher.

Some of the ways in which this data might be used are:

- Calculating the spectral centroids of the notes, and plotting this as estimated "brightness" over the range of the instrument.
- Similarly, a algorithm for estimating "roughness" or acoustic dissonance may be applied to compare one note or instrument to another.
- Because the information on the relative amplitude of notes is available (in each instru-

ment's CONTENTS file), one can form hypotheses about the dynamic nature of an instrument's performing range.

- Spectra may be combined to simulate harmonies; dissonances of combined spectra could be calculated, and the database searched to find, for example, the most consonant example of a semitone interval, or the most dissonant perfect fifth.
- These analyses might be the basis for a particular orchestration used by a composer; or, they could be used by Music Theorists and Musicologists for analysis of orchestration.
- Acousticians may with to test propositions such as the claim that the oboe and English horn possesses "formants."

Perhaps most important of all is that users of this database understand how to interpret this data correctly. For users with backgrounds in acoustics and Digital Signal Processing, the section HOW THE ANALYSES WERE DONE may be sufficient; for others, the implications of this approach are more explicitly spelled out below. For each note analysed, the user should keep in mind that:

- The portion of the tone which has been analysed has been very carefully selected for "representativeness", but nevertheless the analysis represents the spectrum of the note at single, brief moment in time.
- While this data, by itself, cannot be used to synthesise realistic sounding musical instruments, it does represent a real moment in time from the instrument, and a resynthesis of the waveform from this data will faithfully reproduce that moment. Although such moments are usually rather flat and "electronic" sounding, there are many occasions in which the steady state alone produces a quite recognisable musical instrument sound. Wind instruments are a frequent example. In fact, with a few added features (such as adding a characteristic attack-sustain-decay-release envelope to the sound), I have even been able to generate a tone that can "fool the listener." However, as a general rule, this will be very poor strategy for musical instrument tone synthesis.
- Any given note, say, a flute c4, is by no means the "final word" on the spectrum of a flute c4. There are an infinite number of ways a flute c4 can be played, and the instance of it on the MUMS CD represents only one manner of playing from a particular player, instrument, and recording conditions. There are some who take a rather pessimistic stance and believe that, because of the infinite number of possibile performances of a note, there is no use to having information on a single note. I believe that so long as one interprets the data appropriately, data on one note is far better than data on no notes at all.

HOW THE ANALYSES WERE DONE

For each analysed note, the objective was to provide Fourier spectra for a portion of the tone that was maximally "representative" of the steady portion of the tone.
Tremendous care was taken in finding a "representative" portion of each note. The procedure was as follows:

- The samples for the note were taken from the CD and put in a computer sound file. Only one of the CD's stereo channels were saved. Leading and trailing silence was removed. The sampling rate of the file (44100) was converted to 22050.

- The sound file was analysed with a Phase Vocoder.
- The longest continuous stretch of time in which the note was at 75% or more of its maximum amplitude was identified from the PV information. This located the steady portion of the tone.
- An average spectrum was calculated from all the PV frames identified in step 3. Then least squares was used to find the actual PV frame most closely resembling this average spectrum. The point in time corresponding to this PV frame was designated the "representative point".

Analysis then proceeded as follows:

- A chunk of samples corresponding to five periods of the nominal fundamental (ie. according to the equal-tempered frequency of the note) were taken from the sound file, from a point in time symmetrically about the "representative point".
- Autocorrelation was used to estimate the actual fundamental frequency of the sample chunk. Once determined, the chunk was trimmed to four periods of this fundamental. The starting point of the four periods was selected to be at a zero crossing.
- The length of the sample chunk was changed to the next largest power of two by the method of bandlimited interpolation. This step was taken to make it possible to use an FFT.
- The sample chunk was Hamming-windowed.
- The samples were analysed with an FFT, and the real and imaginary values converted to power spectra in decibels. In order to save only partials at harmonic multiples of the fundamental, only every fourth bin was saved (because the sample chunk contained not just one period, but four). All bins greater than 10 kHz were discarded.

ORGANISATION OF THE DIRECTORIES

Each instrument has its own directory; within each directory is a separate file for each of the notes that were available for analysis. The organisation is such that in order to interpret each individual note file completely, you need to reference a file titled "CONTENTS" within the same directory. The individual note files have N rows, where N is the number of harmonics for that note (all possible harmonics below 10 kHz are included). There are two columns, one for the amplitudes (given in decibels relative to the amplitude of the loudest harmonic for that note) and one for the phases (-PI to +PI). The frequencies of the harmonics are integer multiples of the note's fundamental. The actual frequencies of the harmonics are simply the row number multiplied by the fundamental frequency for that note (as found in the "CONTENTS" file).

The CONTENTS file contains a line containing information about each of the notes in the directory. There are ten columns in each line:

- Column 1: The pitch (which identifies what file in the directory this line refers to). The pitch naming system is the Acoustical Society of America standard, i.e. c4 = middle C.
- Column 2: The note number of this pitch (where c4 = 48)
- Column 3: Number of harmonics in the file (hence the number of lines as well)
- Column 4: The maximum absolute value of the sample segment used in the analysis of this tone (i.e. the raw samples as read off of the CD). This is useful for compar-

ing the levels between notes. The possible range of samples on a CD are, of course, -32767 to 32768.

- Column 5: The nominal fundamental frequency for the pitch, according to equal-tempered tuning.
- Column 6: The actual fundamental frequency, as measured from the samples for this tone.
- Column 7: Volume number of the McGill University Master Samples (MUMS) CDs from which this note comes
- * Column 8: MUMS track number
- * Column 9: MUMS index number
- * Column 10: Total duration (in seconds) of the performed note on the CD, from onset to end of decay.
- * Column 11: The point in time (in seconds), relative to the onset of the note, from which the analysis was taken.
- * Column 12: the Spectral centroid in hertz

A BRIEF HISTORY OF THIS PROJECT

I began this project while a PhD student at Northwestern University (USA) in 1990. All the orchestral tones from the MUMS CDs were analysed. I reported on the project at the 1991 International Computer Music Conference in Montreal (Sandell, 1991, "A library of orchestral instrument spectra," Proceedings of the 1991 International Computer Music Conference, 98-101), but this data was never made publically available.

After a few years of looking at the data and thinking of ways in which the project could have been done better, and after finding a way in which to automate the task using a CD-ROM drive, I re-did the entire project from scratch. This was done in 1993-94 at Sussex University.

I would like to acknowledge the support of the graduate school of Northwestern University for a Dissertation Year Grant that helped fund the original project.

INFORMATION FOR WORLD WIDE WEB USERS

The top level URL for SHARC is ftp.ep.susx.ac.uk/pub/sandell/. Once there, you can read this documentation file by clicking on README.html, or browse a small example (just two of the 39 instruments) of what the database actually looks like when installed by clicking on examples. Note that the README.html is always the most up to date version of the documentation that is available.

My homepage is http://ep56c.ep.susx.ac.uk/Greg.Sandell.html. Be warned that all of these URLs may be changing in the near future.

BUGS, INACCURACIES, WISHLIST

- The autocorrelation method was not always successful in determining the fundamental frequency. The pizzicato and marcato string notes in particular seemed to pose the biggest problem. The most serious errors of this sort have been fixed, but a few modest errors (fundamentals off by several Hertz) remain to be corrected. In several cases the problem was due to a strong resonance or vibration from an open string that produced a second tone that competed with the nominal tone, or noisiness in the note such as a prominent bow scraping sound.

- This previous point highlights the fact that the analysis approach used in SHARC assumes all the instruments to produce harmonic spectra. Obviously this is not true in the case of the strings, in which the vibration of other strings may be an essential part of the timbre of the note in question. Inharmonicities in instruments can also occur when a strong native resonance for the particular instrument is active. The choice to save only harmonic information was made because

  + The vast majority of instruments do not evince strong inharmonic partials
  + A Fourier analysis will show energy at frequency locations that are not harmonic partials for any instrument; to decide whether to accept or reject a given inharmonic partial requires an ad hoc decision. Because of the large size of this database, this sort of attention to individual notes is not feasable.
  + I wanted to avoid having multiple formats for the files (one for harmonic notes, another for inharmonic notes)

- Some instruments have notes "missing" in the series (for example, tuba e2). This is because these notes are missing on the MUMS CDs as well.

- I have no idea how the MUMS engineers set the level from one day to the next over the course of the recordings. I have a hunch that, within one chromatic scale for each instrument, the same level and mike placement was used. Otherwise all bets are off: different instruments may have had different levels and mike placements, so it would be wise to practice caution in comparing the levels across instruments (which you can do by consulting the CONTENTS files for each instrument).

- A few instruments that I am analysing have not made their way into SHARC yet, but will soon: these include piano, celesta, harp, and some early instruments (all from MUMS).

- I plan to make graphic plots of the data available on the World Wide Web. Part of this will come very soon. Later, when advances in the Web make it possible, I hope to have a "graphical timbre server" where users are able to make individual requests for certain types of plots.

Gregory J. Sandell (sandell@epunix.sussex.ac.uk)

Appendix 8-3: Frame stage

The frame data is calculated from the SHARC database of spectroid measurements. The calculation is done by a csh script which expects a subdirectory for each instrument, with files for each note that have the naming convention instrument_notenum.spect, as is standard in the SHARC distribution. I have written a system of scripts to calculate brightness in Acums from the SHARC spectroids, and to plot the results. The process is automated by the script GreyAcumMap.csh which produces the frame as a text data file as follows

GreyAcumMap.csh > frame.dat

The programs in the system are listed below.

GreyAcumMap.csh	get the acumoid for Sample Pitches for the Grey timbres from SHARC files
GreyAcum.csh	get Acum data from Grey SHARC files
AcumPlot.csh	go through SHARC directories creating Acum plots
AcumData.csh	go through SHARC directories calculating Acumoids
Acumoid.cc	calculate Sharpness in Acums from SHARC timbre database files
midi2hz.cc	convert midi notenumber to Hz
name2midi.cc	convert note name (e.g. c4) to midi notenumber

```
#! /bin/csh -f
#/*
#*****************************************************************************
#*      COPYRIGHT (C) CSIRO DIT CSIS 1991
#*
#*      SOURCE FILE     GreyAcumMap
#*
#*      MODULE          Sonify
#*
#*      SYSTEM          research
#*
#*      ENVIRONMENT     Sun SPARC and SUNOS4.1
#*
#*      AUTHOR          Stephen Barrass, CSIS
#*
#*      HISTORY
#*      19/4/95: first written   s.b.
#*
#*****************************************************************************
#*/
# filenames are in form instr_nnn.acum

if ($1 == “-h”) then
goto synopsis
endif

set lo=$1
if ($lo == ““) then
set lo=0
endif

set hi=$2
if ($hi == ““) then
set hi=127
endif

foreach dir ( flute_vibrato cello_muted_vibrato bass_clarinet English_horn bassoon trombone_muted )
set notenum=$lo
while ($notenum < $hi)
      set notename=`midi2name $notenum`
      GetAcum.csh $dir $notename
       @ notenum++
   end
end

goto end

synopsis:
echo “$0 [lo <0>] [hi <127>]”
echo “lo = low notenumber”
echo “hi = high notenumber”
echo “gets the acumoid for Sample Pitches in range lo-hi for the Grey timbres from SHARC files”

end:

#! /bin/csh -f
#/*
#*****************************************************************************
#*      COPYRIGHT (C) CSIRO DIT CSIS 1991
#*
#*      SOURCE FILE     GreyAcum
#*
#*      MODULE          Sonify
#*
```

```
#*	SYSTEM		research
#*
#*	ENVIRONMENT	Sun SPARC and SUNOS4.1
#*
#*	AUTHOR		Stephen Barrass, CSIS
#*
#*	HISTORY
#*	19/4/95: first written   s.b.
#*
#*********************************************************************************
#*/
# filenames are in form instr_nnn.acum

if ($1 == "-h") then
goto synopsis
endif

foreach dir ( flute_vibrato cello_muted_vibrato bass_clarinet English_horn bassoon trombone_muted )
 cd $dir
foreach notename ($argv)
      set midi=`name2midi $notename`
      set hz=`midi2hz $midi`
      set acumoid=`Acumoid -b $hz < *_$notename.spect`
      if ($acumoid != "") then
                echo $dir $midi $acumoid
             else
                echo $dir $midi NULL
            endif
   end
  cd ..
end

goto end

synopsis:
echo "$0 notename (e.g. c4 c5 c6)"
echo "gets the acumoid for the Grey timbres from SHARC files"

end:


#! /bin/csh -f
#/*
#*********************************************************************************
#*	COPYRIGHT (C) CSIRO DIT CSIS 1991
#*
#*	SOURCE FILE	AcumPlot
#*
#*	MODULE		Sonify
#*
#*	SYSTEM		research
#*
#*	ENVIRONMENT	Sun SPARC and SUNOS4.1
#*
#*	AUTHOR		Stephen Barrass, CSIS
#*
#*	HISTORY
#*	19/4/95: first written   s.b.
#*
#*********************************************************************************
#*/
# filenames are in form instr_nnn.acum

if ($1 == "-h") then
goto synopsis
endif
```

```
set f=`ls *.spect`
set i=`expr index $f[1] '_'`
@ i--
set instr=`expr substr $f[1] 1 $i`
AcumData.csh > ${instr}_acumoid.data
graph -m 0 -g 1 -y 0 4 1 -l "SHARC $instr Acumoid" < ${instr}_acumoid.data | plot -Tdumb > ${instr}_acumoid.plot

goto end

synopsis:
echo "$0 produces a plot of acumoid data extracted from SHARC files"

end:

#! /bin/csh -f
#/*
#*********************************************************************************
#*     COPYRIGHT (C) CSIRO DIT CSIS 1991
#*
#*     SOURCE FILE     AcumData
#*
#*     MODULE          Sonify
#*
#*     SYSTEM          research
#*
#*     ENVIRONMENT     Sun SPARC and SUNOS4.1
#*
#*     AUTHOR          Stephen Barrass, CSIS
#*
#*     HISTORY
#*     19/4/95: first written   s.b.
#*
#*********************************************************************************
#*/
# filenames are in form instr_nnn.spect

if ($1 == "-h") then
goto synopsis
endif

foreach f ( *.spect )
set i=`expr index $f '_'`
set j=`expr index $f '.'`
@ i++
set l=`expr $j - $i`
set nnn=`expr substr $f $i $l`
set midi=`name2midi $nnn`
set hz=`midi2hz $midi`
set acumoid=`Acumoid -b $hz < $f`
echo $midi $acumoid $nnn
end

goto end

synopsis:
echo "$0 reads all SHARC analysis files with names instr_nnn.spect
and writes Acumoid data points to stdout"

end:
```

```
/*
********************************************************************************
*COPYRIGHT (C) CSIRO DIT CSIS 1995
*
*SOURCE FILEAcumoid.cc
*
*MODULE     research
*
*SYSTEM       Sonify
*
*ENVIRONMENTSun SPARC and SUNOS4.1.2
*
*AUTHORStephen Barrass, CSIS
*
*HISTORY
*:FirstwrittenApril1995.
********************************************************************************
*/
/* RCS history log
$Log$
*/
/* RCS revision identifier and equivalent string
$ID$
*/
static char rcsid[] = “$Id:$”;

/*****************************************************************************/
#include <stream.h>
#include <math.h>
#include <string.h>

//---------------------------------------------------------------------------
// critical band rate weighting factor
//
//          17      18      19      20      21      22      23      24
float w[] = {1.1,1.2,1.3,   1.5,      1.7,      2.0,      3.0,     4.0};

float
g(int i)
{
if (i < 17) return 1;
return(w[i-17]);
}

//---------------------------------------------------------------------------
// excitation mask
// Zwicker & Fastl pp. 151
const MASK_SIZE = 10;
//              -2   -1   0    12    3    4    56    77
float mask[MASK_SIZE] = {0.10, 0.50, 1.00, 0.80, 0.60, 0.40, 0.20, 0.10, 0.05, 0.02 };

//---------------------------------------------------------------------------
// synopsis
//
void
synopsis(char * prog)
{
cerr<<“Synopsis\n”;
cerr<<“-------- \n”;
cerr<<prog<<“ [-h help] -b baseHz < stdin > stdout” << endl;
cerr <<“Arguments[default]// description\n”;
cerr << “------------------------------------------------------------------\n”;
cerr<<“-h\t\t\t\tthis help message\n”;
cerr<<“-b baseHz\t <100>\t\t\tbase frequency in Hz\n”;
cerr<<“stin\t\t\tspectral analysis in SHARC format\n”;
cerr<<“stdout\t\t\tspectral centroid\n”;
```

```
cerr<<"Description\n";
cerr<<"----------- \n";
cerr<<"Calculate the Sharpness in ACUMs. (1 Acum is the sharpness of a narrow band noise centred at 1 kHz. Algorithm from Psychoacoustics by Zwicker & Fastl.\nThe input file consists of relative harmonic amplitudes wrt to peak in dB in column 1, and relative phase in column 2.\n";
cerr<<"Stephen Barrass\n";
cerr<<"CSIRO DIT\n";
cerr<<"stephen@cbr.dit.csiro.au\n";
cerr<<"\nCOPYRIGHT (c) CSIRO DIT 1995\n";
}

//---------------------------------------------------------------------------
// main
//
main(int argc, char * argv[])
{
  int i;
float baseKHz = 100;

// process args
for (i=1; i<argc; i++)
     {
      if (strcmp("-h", argv[i]) == 0)
               {
               synopsis(argv[0]);
                    return 0;
               }
      else if (strcmp("-b", argv[i]) == 0)
               {
               baseKHz = atof(argv[++i])/1000.0;
               }

     }

const BARK_MAX = 24;// maximum Bark
float relativeAmp;
float Ndz = 0;
float Nzdz = 0;
float dB, radians;
float bandrate[BARK_MAX];
for (i=0; i<BARK_MAX; i++)
      bandrate[i] = 0;

float f = baseKHz;
float Bark;
int bin,binI;
cin >> ws;
while (!cin.eof())
     {
      cin >> dB >> ws >> radians >> ws;
// ignore phase in power spectrum
// dB = 20 * log(Comparison/Reference)
// therefore Relative = Comparison/Reference = 10 ^ (dB / 20)
      relativeAmp = exp10(dB/20);
// convert kHz frequency of this harmonic into Bark 1..24
      Bark = 13*atan(0.76*f)+3.5*atan((f*f)/56.25);
// integer Bark is boundary between critical Bands
      bin = nint(floor(Bark))-1;
        binI = bin-2;
// replace each bin with a mask
// use simple triangle shape for Loudness = 60
// (Zwicker & Fastl, pp 151)
      for (i=0; i<MASK_SIZE; i++, binI++)
               {
                  if (binI < 0)
                              continue;
```

```
                bandrate[binI] += mask[i]*relativeAmp;
                }
      f += baseKHz;
     }
cerr << endl;

// find see-saw point (first moment)
// moment calculation

Ndz = Nzdz = 0;
cerr << "Bandrate = ";
float bandAmp;
// the barks have been band-rate binned
// apply a log function to bin amplitudes for saturation effect
for (i=0; i<BARK_MAX;)
     {
// want a logarithmic saturation effect
//    bandAmp = log(bandrate[i]+1);
      bandAmp = bandrate[i];
      cerr << bandAmp << " ";
      Ndz += bandAmp;
// weighting factor  (aka Zwicker)
             ++i;
      Nzdz += bandAmp * i * g(i);
     }
cerr << endl;

float Acumoid =  Nzdz / Ndz;

// scale factor from Zwicker pp 218
Acumoid *= 0.11;

cerr << "Sharpness = " << setprecision(2) << Acumoid << " Acum" << endl;
cout << setprecision(2) << Acumoid << endl;

return 1;
}



/*
*******************************************************************************
*COPYRIGHT (C) CSIRO DIT CSIS 1995
*
*SOURCE FILEmidi2hz.cc
*
*MODULE     research
*
*SYSTEM       Sonify
*
*ENVIRONMENTSun SPARC and SUNOS4.1.2
*
*AUTHORStephen Barrass, CSIS
*
*HISTORY
*:FirstwrittenApril1995.
*******************************************************************************
*/
/* RCS history log
$Log$
*/
/* RCS revision identifier and equivalent string
$ID$
*/
static char rcsid[] = "$Id:$";
```

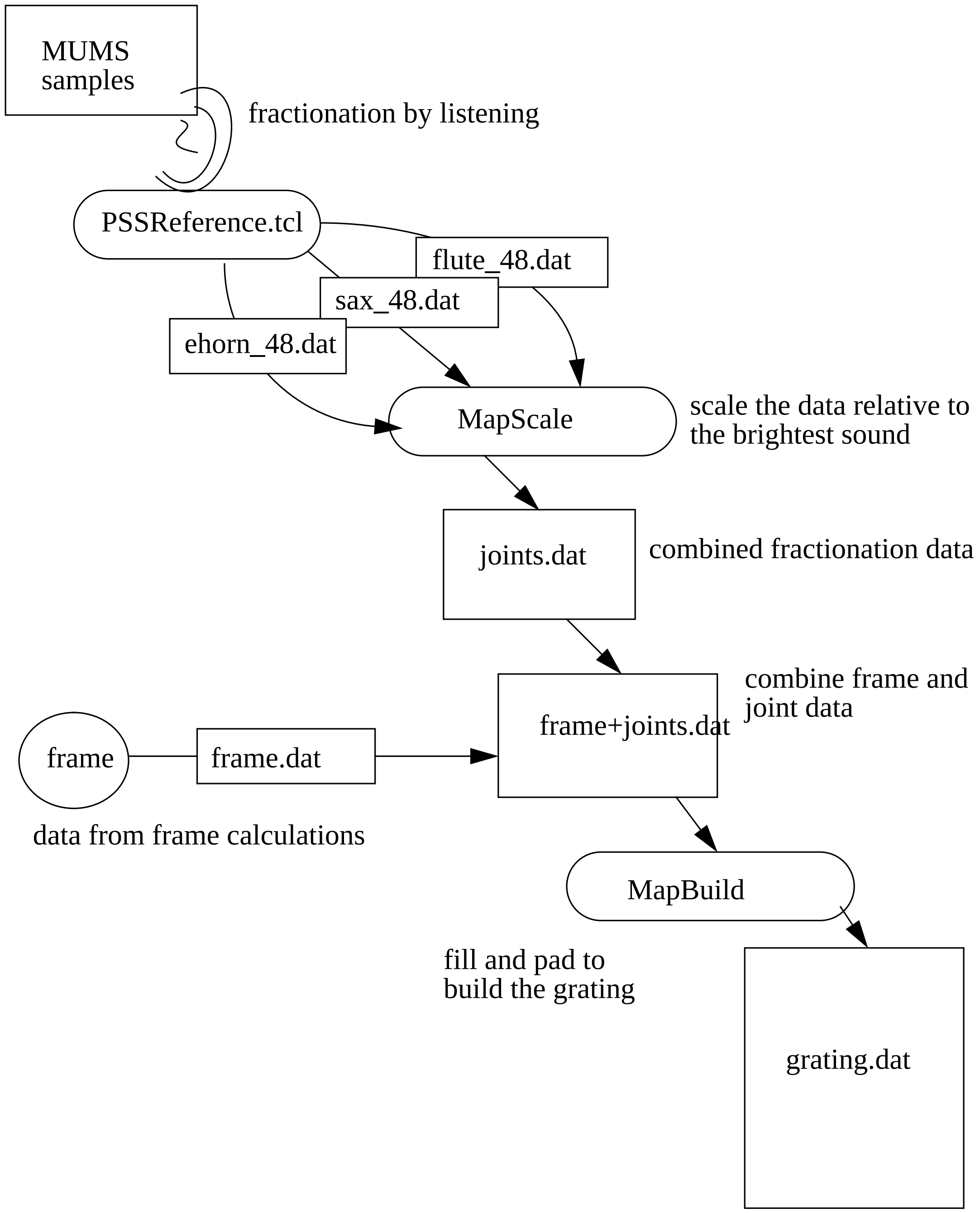
```
/******************************************************************************/
#include <stream.h>
#include <math.h>
#include <string.h>


//---------------------------------------------------------------------------
// synopsis
//
void
synopsis(char * prog)
{
cerr<<“Synopsis\n”;
cerr<<“-------- \n”;
cerr<<prog<<“ [-h help] midinote > stdout” << endl;
cerr <<“Arguments[default]// description\n”;
cerr << “-----------------------------------------------------------------\n”;
cerr<<“-h\t\t\t\tthis help message\n”;
cerr<<“midinote\t\t\t0-127 midi note number\n”;
cerr<<“stdout\t\t\tfrequency in Hz\n”;
cerr<<“Description\n”;
cerr<<“----------- \n”;
cerr<<“Convert midi notenumber to frequency in Hz\n”;
cerr<<“Stephen Barrass\n”;
cerr<<“CSIRO DIT\n”;
cerr<<“stephen@cbr.dit.csiro.au\n”;
cerr<<“\nCOPYRIGHT (c) CSIRO DIT 1995\n”;
}

//---------------------------------------------------------------------------
// main
//
main(int argc, char * argv[])
{
  int i;

// process args
for (i=1; i<argc; i++)
     {
      if (strcmp(“-h”, argv[i]) == 0)
               {
                synopsis(argv[0]);
                     return 0;
               }
     }
float notenum = atof(argv[argc-1]);
float hz = 16.351625 * pow(2, notenum/12);
cout << hz << endl;
return 1;
}
```

```
/*
********************************************************************************
*COPYRIGHT (C) CSIRO DIT CSIS 1995
*
*SOURCE FILEname2midi.cc
*
*MODULE     research
*
*SYSTEM       Sonify
*
*ENVIRONMENTSun SPARC and SUNOS4.1.2
*
*AUTHORStephen Barrass, CSIS
*
*HISTORY
*:FirstwrittenApril1995.
********************************************************************************
*/
/* RCS history log
$Log$
*/
/* RCS revision identifier and equivalent string
$ID$
*/
static char rcsid[] = "$Id:$";

/******************************************************************************/
#include <stream.h>
#include <math.h>
#include <string.h>


//---------------------------------------------------------------------------
// synopsis
//
void
synopsis(char * prog)
{
cerr<<"Synopsis\n";
cerr<<"-------- \n";
cerr<<prog<<" [-h help] notename > stdout" << endl;
cerr <<"Arguments[default]// description\n";
cerr << "-----------------------------------------------------------------\n";
cerr<<"-h\t\t\t\tthis help message\n";
cerr<<"notename\t\t\tnote name e.g. c4, g#2, D6, A#1 etc.\n";
cerr<<"stdout\t\t\tcorresponding midi notenumber 0..127\n";
cerr<<"Description\n";
cerr<<"----------- \n";
cerr<<"Convert note name to midi notenumber\n";
cerr<<"Stephen Barrass\n";
cerr<<"CSIRO DIT\n";
cerr<<"stephen@cbr.dit.csiro.au\n";
cerr<<"\nCOPYRIGHT (c) CSIRO DIT 1995\n";
}

//---------------------------------------------------------------------------
// main
//
main(int argc, char * argv[])
{
  int i;

// process args
for (i=1; i<argc; i++)
     {
```

```
      if (strcmp(“-h”, argv[i]) == 0)
               {
                synopsis(argv[0]);
                     return 0;
               }
      }
char * nnP = argv[argc-1];
int base;
switch(*nnP)
      {
           case ‘a’:
           case ‘A’:
                      base = 9;
                        break;
           case ‘b’:
           case ‘B’:
                     base = 11;
                        break;
           case ‘c’:
           case ‘C’:
                      base = 0;
                        break;

           case ‘d’:
           case ‘D’:
                      base = 2;
                        break;
           case ‘e’:
           case ‘E’:
                      base = 4;
                        break;
           case ‘f’:
           case ‘F’:
                      base = 5;
                        break;
           case ‘g’:
           case ‘G’:
                      base = 7;
                        break;
           default:
                cerr << “Error : Bad note name\n”;
                synopsis(argv[0]);
                     return 0;
      }
nnP++;
if (*nnP == ‘#’)
      {
           base++;
           nnP++;
      }
int offset = atoi(nnP)*12;
cout << base+offset << endl;
return 1;
}
```

Appendix 8-4: Grating stage

The grating is built by filling in the frame with perceptually measured joints that are equal steps in perceived brightness. The process has the following steps.



These programs are listed below

PSSReference.tcl	GUI for fractionating the brightness of a sound sample
MapScale.csh	scale the data relative to the maximum brightness
MapBuild.cc	fill and pad the data to build a grating

```
##################################################################################
#COPYRIGHT (C) CSIRO DIT CSIS 1993
#
#SOURCE FILEPSSReference.tcl
#
#MODULE      Sonify
#
#SYSTEM      PostGrad
#
#ENVIRONMENTSun SPARC and SUNOS4.1.2
#
#AUTHORStephen Barrass, CSIS
#
#HISTORY
#               : First written Feb 1994
##################################################################################

#

# Source file identification for debugging #####################################
#

if {[info exists DEBUG]} {
if {$DEBUG} {
     puts stderr {$Header: /proj/vis/mira/stephen/camroot/src/app/sonify/tk/RCS/PSSReference.tcl,v 1.1 1994/03/29
00:51:13 stephen Exp stephen $}
    }
}

#- public variables------------------------------------------------

proc PSSReference::public {} {
Debug::proc {PSSReference::public}

}

#- private variables------------------------------------------------

proc PSSReference::private {path} {
Debug::proc {PSSReference::private}

uplevel 1 "global $path.num"
}

#- PSSReference::init------------------------------------------------
# initialise privates
proc PSSReference::init {path} {
Debug::proc {PSSReference::init}
PSSReference::private $path

set $path.num 0
}

#- PSSReference------------------------------------------------
# create a PSSReference
#

proc PSSReference {path args} {

Debug::proc {PSSReference}
Environment::global
Perceptual::public
Device::public
```

```
# start Csound with the PSSReference
Csound $path.cs

toplevel $path
wm title $path Fractionation

PSSReference::public
PSSReference::private $path
PSSReference::init $path

# heading stuff
frame $path.head

FileName $path.head.file \
      -load "PSSReference::read $path" \
      -save "PSSReference::write $path" \
        -mask *.map


NumericEntry $path.head.notenum \
      -width 128 -height 22 -labwidth 80 \
      -text Notenum \
      -from [set Device::NOTENUM_MIN] \
      -to [set Device::NOTENUM_MAX] \
      -command "PSSReference::setNotenum $path"

button $path.head.refharmonic \
      -text Harmonic \
      -command "$path.head.patch set 500"

button $path.head.refnoise \
        -text Noise \
      -command "$path.head.patch set 501"

button $path.head.refacum \
        -text Acum \
      -command "$path.head.patch set 502"

NumericEntry $path.head.patch \
      -width 128 -height 22 -labwidth 80 \
        -text Patch \
          -from 0 \
         -to 9999 \
      -command "PSSReference::setPatch $path"

NumericEntry $path.head.dur \
      -width 128 -height 22 -labwidth 80 \
      -text Duration \
       -allowfloat 1 \
      -from [set Device::DUR_MIN] \
      -to [set Device::DUR_MAX] \
      -command "PSSReference::setDur $path"

pack append $path.head \
      $path.head.file {top expand fillx} \
      $path.head.notenum {left} \
      $path.head.refharmonic {left} \
      $path.head.refnoise {left} \
      $path.head.refacum {left} \
      $path.head.patch {left} \
      $path.head.dur {left}

# scale
frame $path.scale
frame $path.scale.label
```

```
label $path.scale.label.filter \
      -width 10 -height 1 \
         -anchor w \
        -text {Filter}

button $path.scale.label.play \
         -text Play \
      -bitmap @$CAMROOT/lib/bitmap/Play.xbm \
      -command "PSSReference::play $path"

pack append $path.scale.label \
      $path.scale.label.play {top expand fillx} \
      $path.scale.label.filter top

# create the Tuner bank
frame $path.scale.bank
set num 0
set timbre 0
set referenceInstrument 1

set bright [set Perceptual::BRIGHT_MIN]
while {$bright <= [set Perceptual::BRIGHT_MAX]} {
      set keynum [expr "$num + 1"]
      Tuner $path.scale.bank.tuner$num \
                  -timbre $timbre \
                  -bright $bright \
                  -timbreNoShow 1 \
                  -pitchNoShow 1 \
                  -notenumNoShow 1 \
                  -velNoShow 1 \
                  -brightNoShow 1 \
                        -inv 1 \
                   -key $keynum

      pack append $path.scale.bank $path.scale.bank.tuner$num {left}

      Tuner::setEqbw $path.scale.bank.tuner$num [expr "$bright*[set Device::EQBW_RANGE]/[set Perceptu-
al::BRIGHT_RANGE]"]
      Tuner::setInstrument $path.scale.bank.tuner$num $referenceInstrument

          incr bright
           incr num
     }

set $path.num $num
pack append $path.scale \
      $path.scale.label left \
      $path.scale.bank top

# pack it all
pack append $path \
      $path.head {top expand fillx} \
      $path.scale {top}

$path.head.notenum set 60
}

#- PSSReference::toplevel-----------------------------------------------
# create a toplevel pss reference
proc PSSReference::toplevel {path} {
Debug::proc {PSSReference::toplevel}

toplevel $path
PSSReference $path.pss
pack append $path $path.pss top
}
```

```
#- PSSReference::play-----------------------------------------------
# play the tagged sounds
proc PSSReference::play {path} {
Debug::proc {PSSReference::play}
PSSReference::private $path

set num 0
while {$num < [set $path.num]} {
    Tuner::play $path.scale.bank.tuner$num
        sleep 1
    Tuner::stop $path.scale.bank.tuner$num
        incr num
    }
}

#- PSSReference::setPitchH-----------------------------------------------
# set the Pitch
proc PSSReference::setPitchH {path {pitchH 5}} {
Debug::proc {PSSReference::setPitchH}
PSSReference::private $path

# change the pitch
PSSReference::private $path

set num 0
while {$num < [set $path.num]} {
    Tuner::setPitchH $path.scale.bank.tuner$num $pitchH
    Tuner::setNotenum $path.scale.bank.tuner$num [Perceptual::getPitchN $path.scale.bank.tuner$num.perceptual]
        sleep 1
        incr num
    }
}

#- PSSReference::setNotenum-----------------------------------------------
# set the Notenum
proc PSSReference::setNotenum {path {notenum 60}} {
Debug::proc {PSSReference::setNotenum}
PSSReference::private $path

# change the pitch
PSSReference::private $path

set num 0
while {$num < [set $path.num]} {
    Tuner::setNotenum $path.scale.bank.tuner$num $notenum
        incr num
    }
}

#- PSSReference::setPatch-----------------------------------------------
# set the Patch
proc PSSReference::setPatch {path {patch 0}} {
Debug::proc {PSSReference::setPatch}
PSSReference::private $path

# change the patch
PSSReference::private $path

set num 0
while {$num < [set $path.num]} {
    Tuner::setPatch $path.scale.bank.tuner$num $patch
        incr num
    }
}
```

```
#- PSSReference::setDur-----------------------------------------------
# set the Duration
proc PSSReference::setDur {path {dur 1.0}} {
Debug::proc {PSSReference::setDur}
PSSReference::private $path

# change the duration
PSSReference::private $path

set num 0
while {$num < [set $path.num]} {
    Tuner::setDur $path.scale.bank.tuner$num $dur
        incr num
    }
}

#- PSSReference::read-----------------------------------------------
# read from file
proc PSSReference::read {path filename} {
Debug::proc {PSSReference::read}
PSSReference::private $path

# open the file
set fd -1
set fd [open $filename r]
if {$fd == -1} {
          return
    }
# read the header
# timbre leaf
set header ““
set timbre 0
gets $fd header
gets $fd header
set notenum 0
scan $header “# Notenum %d” notenum
#
# read each tuner
set num 0
while {$num < [set $path.num]} {
    Tuner::read $path.scale.bank.tuner$num $fd
        incr num
    }
close $fd
$path.head.notenum set $notenum
}

#- PSSReference::write-----------------------------------------------
# write to file
proc PSSReference::write {path filename} {
Debug::proc {PSSReference::write}
PSSReference::private $path

# open the file
set fd -1
set fd [open $filename w]
if {$fd == -1} {
          return
    }
# write the header
puts $fd “# PSSReference”
puts $fd “# Notenum [$path.head.notenum get]”

# write each scale
# write each tuner
```

```
set num 0
while {$num < [set $path.num]} {
      Tuner::write $path.scale.bank.tuner$num $fd
          incr num
     }
close $fd
}

#- PSSReference::test-----------------------------------------------
# write all leaves to a file
proc PSSReference::test {path {filebase tleaf}} {
Debug::proc {PSSReference::test}
PSSReference::private $path

}

#!/bin/csh -f
#
#################################################################################
#COPYRIGHT (C) CSIRO DIT CSIS 1994
#
#SOURCE FILEMapScale
#
#MODULE       Sonify
#
#SYSTEM       Thesis
#
#ENVIRONMENTSun SPARC and SUNOS4.1.2
#
#AUTHORStephen Barrass, CSIS
#
#HISTORY
#              : First written May 1995
#################################################################################

# Synopsis
#
if ($1 == "-h") then
goto synopsis
endif

set Bmax = $1
set notenum = $2
set infile = $3
gawk -v Bmax=$Bmax -v notenum=$notenum -f MapScale.gawk < $infile

goto end

synopsis:
echo "$0 Bmax notenum mapfile"
echo "Bmax is maximum brightness in Acums"
echo "Create a pnf to TBP mapping from a PSSReference map"

end:

/*
;#################################################################################;#
;#COPYRIGHT (C) CSIRO DIT CSIS 1993
;#
;#SOURCE FILEMapBuild.cc
;#
;#MODULE      Sonify
;#
;#SYSTEM     PostGrad
;#
;#ENVIRONMENTSun SPARC and SUNOS4.1.2
```

```
;#
;#AUTHORStephen Barrass, CSIS
;#
;#HISTORY
;#		: First written May 1995
;####################################################################################
*/

#include <iostream.h>
#include <fstream.h>
#include <strstream.h>
#include <math.h>
#include <string.h>

int loNotenum = -100;
int hiNotenum = 200;
int hiFilter = 1000;
int OutOfGamutFilter = 140;
int extraB = 5;  // extra Brightness steps to confirm B boundary
float BMax = 2.0;
int PFillCnt = 0;// extra pitch steps to improve pitch accuracy
int BFillCnt = 6;// number of interpolated B points
int OutOfGamutPatch = 0;
int InGamutPatch = 100;
const PitchPadStep = 5;

char * helpString = "\n\
Arguments[default]// description\n\
------------------------------------------------------------------- \n\
-h				// print this help line \n\
-loP		0		// out-of-gamut lo pitch\n\
-hiP		127		// out-of-gamut hi pitch\n\
-hiFilter  200// out-of-gamut hi filter value\n\
-extraB  5   // number of B extrapolations\n\
-BMax		2.0		// maximum Brightness for extrapolation\n\
stdin		stdin		// read from stdin\n\
stdout  stdout// write to stdout\n\
\n\
Description \n\
----------- \n\
read Acum data and output pnfTBP \n\
\n\
Author \n\
----------- \n\
Stephen Barrass\n\
Visualisation \n\
CSIRO DIT \n\
stephen@csis.dit.csiro.au \n\
\n\
COPYRIGHT (C) CSIRO DIT CSIS 1995 \n\
";


//-BScale-----------------------------------------------------
// scale the equal brightness step points so that maximum brightness
// is calibrated with unfiltered brightness of this sample at this pitch
//
// the modelling software needs lots of points to do a good job,
// particulary for the non-linear (exponential) filter values
// fill-in extra filter values using interpolation between neighbours
// initially try simple linear interpolation
// extend filter values beyond gamut boundary to support the mapping
//
#ifdef TEST
int global_p = 0;
#endif
```

```
void
BScale
(
char * MeasureFile,// datafile with B for each P of this T
float Bacum127,// brightness with no filter (i.e filter = 127)
float Pitch, // pitch of this brightness measurement
int OutOfGamutNotenum = 0 // out-of-gamut Pitch sets this
)
{
  int p;
 float n;
int f = 0;
int fPrev;
int fFill;

float T,B;
B = 0.0;
// 8 equal steps up to the Brightness at this Pitch
float Bstep = Bacum127/8.0;
float Bdelta = Bstep/BFillCnt;

int i = 0;

// BScale
// open the Measurement file
ifstream MeasureStream(MeasureFile);
int startF = 1;

// skip whitespace
MeasureStream >> ws;
// end-of-file ?
while (!MeasureStream.eof())
     {
// skip commented lines
      if (MeasureStream.peek() == '#')
                {
                 MeasureStream.ignore(1024,'\n');
                      continue;
                }
// save previous f value
           fPrev = f;
// get a point
      MeasureStream >> p >> n >> f >> ws;
// skip to end of line
      MeasureStream.ignore(1024,'\n');
      MeasureStream >> ws;

// if first line get next line so we can start interpolating
           if (startF)
                {
// use patch to find Timbre angle
#ifdef TEST
                 T = (global_p-1)*45;
#else
                   T = (p-1)*45;
#endif
                      startF = 0;
                      continue;
                }
// ignore the notenum in the file for the moment since using just
// one measurement file for all pitches
// use Pitch instead
           n = Pitch;
// out-of-gamut Pitch
      p = InGamutPatch;
```

```
      if (OutOfGamutNotenum != 0)
                {
// XXX
//              n = OutOfGamutNotenum;
                p = OutOfGamutPatch;
//              f = OutOfGamutFilter;
                }
// heres the interpolated B fill
      for (i=0; i < BFillCnt; i++, B += Bdelta)
                {
                fFill = fPrev+i*(f - fPrev)/BFillCnt;
// write it out
                cout << p << " " << n << " " << fFill << " " << T << " " << B << " " << Pitch << endl;

                }
      }
// last f point
cout << p << " " << n << " " << f << " " << T << " " << B << " " << Pitch << endl;


// add extra B points to force gamut shape
int fStep = (hiFilter-f)/extraB;
float BStep = (BMax-B)/extraB;
f = OutOfGamutFilter;
p = OutOfGamutPatch;
//B = B*1.01;
B += BStep;
for (i=0; i<extraB; i++)
      {
//       f += fStep;
      cout << p << " " << n << " " << f << " " << T << " " << B << " " << Pitch << endl;
        B += BStep;
      }


}

//-PrePad------------------------------------------------------


void
PrePad
(
char * MeasureFile,// datafile with B for each P of this T
float Bacum127,// brightness with no filter (i.e filter = 127)
float PitchStart,//
float PitchEnd
)
{
float Pitch;
for (Pitch = PitchStart; Pitch <= PitchEnd; Pitch+=PitchPadStep)
      {
      BScale(MeasureFile, Bacum127, Pitch, loNotenum);

      }
}

//-AfterPad------------------------------------------------------
void
AfterPad
(
char * MeasureFile,// datafile with B for each P of this T
float Bacum127,// brightness with no filter (i.e filter = 127)
float PitchStart,//
float PitchEnd
)
```

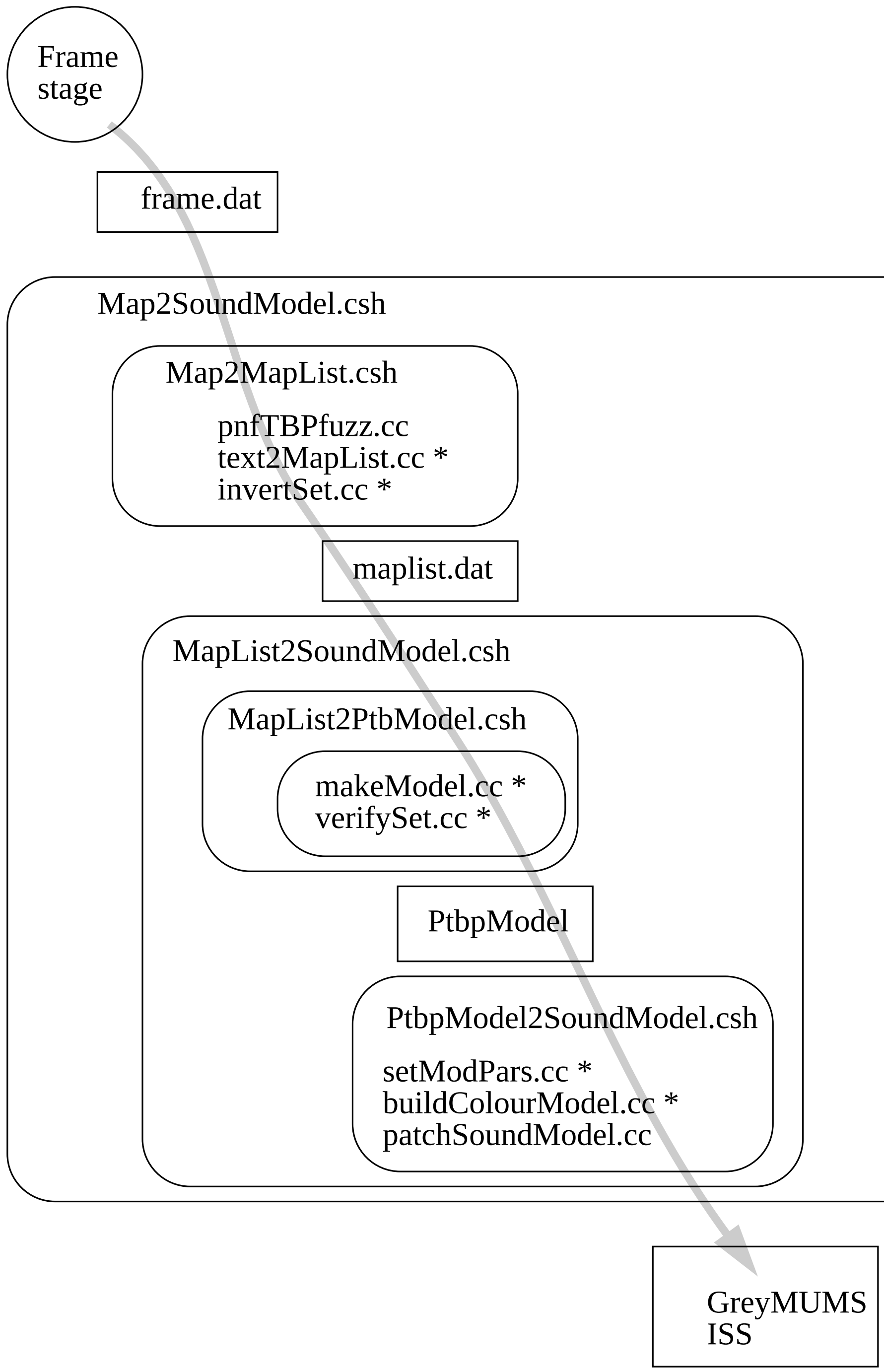
```
{
float Pitch;
for (Pitch = PitchStart; Pitch <= PitchEnd; Pitch+=PitchPadStep)
     {
      BScale(MeasureFile, Bacum127, Pitch, hiNotenum);

     }
}

//-main       -----------------------------------------------------
//

// this routine converts SHARC data to pnfTBP data suitable for the
// modelling routines. This process includes padding to help the
// modelling software smoothly drape flesh onto the bones
// which requires consideration of the hyper-linear interpolation algorithm
// The SHARC data consists of Brightness values for a Pitch range of a
// particular Timbre
// file format is
// stringfloatfloat
// Timbre pitchbrightness
//
// the Timbre data is arranged in increasing pitch order
// flute 48 0.5
// flute 49 0.6
// flute 50 0.45
// .. etc.
//
// when a new Timbre is started
// the last data from the previous Timbre is copied up the pitch axis
// to establish the upper half of the cylindrical contour
// then the new Timbre data is prepended is copied down the pitch axis
// to establish the lower cylindrical contour
//
//
main(int argc, char **argv)
{
  int i;
// parse the commandline
for (i=1; i < argc; i++)
     {
      if (strcmp(argv[i],"-h") == 0)
                {
                        i++;
                 cerr << argv[0] << " " << helpString;
                      return 1;
                }
      else if (strcmp(argv[i],"-loP") == 0)
                {
                 loNotenum = atoi(argv[++i]);
                }
      else if (strcmp(argv[i],"-hiP") == 0)
                {
                 hiNotenum = atoi(argv[++i]);
                }
      else if (strcmp(argv[i],"-hiFilter") == 0)
                {
                 hiFilter = atoi(argv[++i]);
                }
      else if (strcmp(argv[i],"-BMax") == 0)
                {
                 BMax = atof(argv[++i]);
                }
      else if (strcmp(argv[i],"-extraB") == 0)
                {
                 extraB = atoi(argv[++i]);
```

```
            }
    }

// output format
long formatF = cout.flags();
cout.flags(formatF | ios::fixed);
cout.precision(3);

char Timbre[64];
char TimbrePrev[64];
ostrstream(Timbre, sizeof(Timbre)) << ends;
float Bacum;
float BacumPrev = 0;
float Pitch;
float PitchPrev = 0;
float PitchMax = 127;
float p, pInc;

char MeasureFile[256];

unsigned char startFlag = 1;

// read SHARC database file from stdin
cin >> ws;
while (!cin.eof())
    {
// skip commented lines
     if (cin.peek() == '#')
            {
            cin.ignore(1024,'\n');
                continue;
            }

// read the line
// instrument_name pitch brigtness
     cin >> ws >> Timbre >> Pitch >> Bacum >> ws;

// is it a new Timbre ?
     if ((strcmp(Timbre, TimbrePrev) != 0))
            {
            if (strcmp(Timbre, "change") == 0)
                    {
                    cin >> ws >> Timbre >> Pitch >> Bacum >> ws;
                    }
            cerr << "change to " << Timbre << "," << Pitch << "," << Bacum << endl;

// change Timbre
// after-pad old Timbre
             if (!startFlag)
                    {
                    AfterPad(MeasureFile, Bacum, PitchPrev+1, PitchMax);
                    }
             startFlag = 0;

// get the new measurement file MeasureFile
// the measurement file contains the equal brightness filter values
// for this Timbre
// using just one file for all pitches at the moment
#ifdef TEST
             global_p++;
#endif
            ostrstream(MeasureFile, sizeof(MeasureFile)) << Timbre << "_" << 48 << ".map" << ends;

// pre-pad new Timbre
            PrePad(MeasureFile, Bacum, 0, Pitch-1);
// update old Timbre name
```

```
                ostrstream(TimbrePrev, sizeof(Timbre)) << Timbre << ends;

                }
        pInc = 1.0/(PFillCnt+1);
        p = Pitch-(PFillCnt/2.0)*pInc;
        for (i=0; i<=PFillCnt; i++)
                {
                BScale(MeasureFile, Bacum, p);
                    p += pInc;
                }
        PitchPrev = Pitch;
        BacumPrev = Bacum;
        }
// end of Timbre at eof
// after-pad
AfterPad(MeasureFile, Bacum, Pitch+1, PitchMax);

}
```

Appendix 8-5: Plasticine stage

The plasticine stage is a 3D spline fit to the grating data points. A set of scripts collects the necessary parameters from the user and automates the file conversion processes. In the process diagram below the colour modelling tools developed by Done Bone have an * next to them [Bone D. (1993)].



```
#!/bin/csh -f
#
#******************************************************************************
#	COPYRIGHT (C) CSIRO DIT CSIS 1992
#
#	SOURCE FILE	Map2SoundModel.csh
#
#	MODULE		personify
#
#	SYSTEM		Research
#
#	ENVIRONMENT	Sun SPARC and SUNOS4.1.2
#
#	AUTHOR		Stephen Barrass, CSIS
#
#	HISTORY
#			: First written May 1995
#******************************************************************************
#
#
echo "**** Map2SoundModel.csh"

source ModelRange

#------------------------------------------------------------------------------
# command line
#
set files =
set fuzz = false

while ($#argv > 0)

	switch ($1)
		case -h:
			goto synopsis
			breaksw

		 case -g:
			   shift
		set grid = $1
		 breaksw

		 case -f:
		set fuzz = true
		 breaksw

		 case -r:
			   shift
		set res = $1
		 breaksw

		 case -s:
			   shift
		set stiff = $1
		 breaksw

		 case -t:
			   shift
		set tight = $1
		 breaksw

	    default:
		  set files = ($files $1)
			breaksw
```

```
	endsw
	shift
end

#-----------------------------------------------------------------------------
# main
#
main:
if ($fuzz == true) then
Map2MapList.csh -f
else
Map2MapList.csh
endif
MapList2SoundModel.csh -g $grid -s $stiff -t $tight -r $res

goto end

#-----------------------------------------------------------------------------
# help
#
synopsis:
echo “$0 [-g grid <$grid>] [-f <$fuzz> (fuzz)] [-r res <$res>] [-s stiff <$stiff>] [-t tight <$tight>]”
echo “from pnfTBPGrey.map to SoundModel”

end:
```

```
#!/bin/csh -f
#
#*******************************************************************************
#	COPYRIGHT (C) CSIRO DIT CSIS 1992
#
#	SOURCE FILE	Map2MapList.csh
#
#	MODULE		personify
#
#	SYSTEM		Research
#
#	ENVIRONMENT	Sun SPARC and SUNOS4.1.2
#
#	AUTHOR		Stephen Barrass, CSIS
#
#	HISTORY
#			: First written May 1995
#*******************************************************************************
#
echo “**** Map2MapList.csh”

#
#-----------------------------------------------------------------------------
# command line
#
set files =
set fuzz = false
set BMax = 2.0

while ($#argv > 0)

	switch ($1)
		case -h:
			goto synopsis
			breaksw

		case -f:
```

```
                set fuzz = true
                breaksw

            case -BMax:
                    shift
                set BMax = $1
                breaksw

        default:
            set files = ($files $1)
                breaksw

    endsw
    shift
end

#-------------------------------------------------------------------------------
# main
#
main:

if ($fuzz == true) then
pnfTBPfuzz -o1 5 < pnfTBPGrey.map | pnfTBPtoDpnfPtbp -BMax $BMax | text2MapList | sed 1,10s/TRUE_CAM/
FALSE_CAM/ | invertSet > PtbpDpnf.ml
else
cat pnfTBPGrey.map | pnfTBPtoDpnfPtbp -BMax $BMax | text2MapList | sed 1,10s/TRUE_CAM/FALSE_CAM/ | in-
vertSet > PtbpDpnf.ml
endif

MapListView.csh

goto end

#-------------------------------------------------------------------------------
# help
#
synopsis:
echo “$0 [-f fuzz] [-BMax <$BMax>]”
echo “make DpnfPtbp.ml & PtbpDpnf.ml from pnfTBPGrey.map”

end:


#!/bin/csh -f
#
#*******************************************************************************
#     COPYRIGHT (C) CSIRO DIT CSIS 1992
#
#     SOURCE FILE     MapList2SoundModel.csh
#
#     MODULE          personify
#
#     SYSTEM          Research
#
#     ENVIRONMENT     Sun SPARC and SUNOS4.1.2
#
#     AUTHOR          Stephen Barrass, CSIS
#
#     HISTORY
#                     : First written May 1995
#*******************************************************************************
#
echo “**** MapList2SoundModel.csh”

source ModelRange
```

```
#
#-----------------------------------------------------------------------------
# command line
#
set files =

while ($#argv > 0)

        switch ($1)
                case -h:
                        goto synopsis
                        breaksw

                 case -g:
                            shift
                set grid = $1
                 breaksw

                 case -r:
                            shift
                set res = $1
                 breaksw

                 case -s:
                            shift
                set stiff = $1
                 breaksw

                 case -t:
                            shift
                set tight = $1
                 breaksw

             default:
                  set files = ($files $1)
                        breaksw

        endsw
        shift
end

#-----------------------------------------------------------------------------
# main
#

main:
MapList2PtbpModel.csh -g $grid -s $stiff -t $tight
PtbpModel2SoundModel.csh -r $res
SoundModelView.csh -small

goto end

#-----------------------------------------------------------------------------
# help
#
synopsis:
echo “$0 [-g grid <$grid>] [-r res <$res>] [-s stiff <$stiff>] [-t tight <$tight>]”
echo “from pnfTBPGrey.map to SoundModel”

end:


#!/bin/csh -f
#
#*********************************************************************************
#      COPYRIGHT (C) CSIRO DIT CSIS 1992
```

```
#
#     SOURCE FILE     MapList2PtbpModel.csh
#
#     MODULE          personify
#
#     SYSTEM          Research
#
#     ENVIRONMENT     Sun SPARC and SUNOS4.1.2
#
#     AUTHOR          Stephen Barrass, CSIS
#
#     HISTORY
#                  : First written May 1995
#******************************************************************************
#
#
echo "**** MapList2PtbpModel.csh"
source ModelRange

#
#------------------------------------------------------------------------------
# command line
#
set files =
set viewF = true

while ($#argv > 0)

    switch ($1)
        case -h:
            goto synopsis
            breaksw

         case -g:
              shift
        set grid = $1
         breaksw

         case -s:
              shift
        set stiff = $1
         breaksw

         case -t:
              shift
        set tight = $1
         breaksw

         case -Pt:
              shift
        set PtMin = $1
              shift
        set PtMax = $1
         breaksw

      case -Pb:
              shift
        set PbMin = $1
              shift
        set PbMax = $1
             breaksw

      case -Pp:
              shift
        set PpMin = $1
              shift
```

```
            set PpMax = $1
                  breaksw

        case -noV:
              set viewF = false
                  breaksw

          default:
              set files = ($files $1)
                  breaksw

      endsw
      shift
end

#-----------------------------------------------------------------------------
# main
# -C = monoticity constraint
#

main:
nice -20 makeModel -M PtbpDpnf.ml -o PtbpModel -a 0 $PtMin $PtMax -a 1 $PbMin $PbMax -a 2 $PpMin $PpMax
-W $stiff $tight -g $grid $grid $grid

verifySet -i PtbpDpnf.ml -m PtbpModel -o outliers.ml -d 10

# have a look at the results
if ($viewF == true) ModelView.csh -g $grid

goto end

#-----------------------------------------------------------------------------
# help
#
synopsis:
echo "$0 [-g <$grid>] [-Pt PtMin <$PtMin> PtMax <$PtMax>] [-Pb PbMin <$PbMin> PbMax <$PbMax>] [-Pp Pp-
Min <$PpMin> PpMax <$PpMax>] [-s stiff <$stiff>] [-t tight <$tight>] [-noV no view]"
echo "make a model from a maplist"

end:


#!/bin/csh -f
#
#*******************************************************************************
#      COPYRIGHT (C) CSIRO DIT CSIS 1992
#
#      SOURCE FILE     MakeSoundModel.csh
#
#      MODULE          personify
#
#      SYSTEM          Research
#
#      ENVIRONMENT     Sun SPARC and SUNOS4.1.2
#
#      AUTHOR          Stephen Barrass, CSIS
#
#      HISTORY
#                  : First written May 1995
#******************************************************************************
#
#
echo "****** MakeSoundModel.csh"

source ModelRange
```

```
#-------------------------------------------------------------------------------
# command line
#
set files =
set slices = false

while ($#argv > 0)

	switch ($1)
		case -h:
			goto synopsis
			breaksw

		 case -s:
		set slices = true
		 breaksw

		 case -r:
			  shift
		set res = $1
		 breaksw

	    default:
		   set files = ($files $1)
			breaksw

	endsw
	shift
end

#-------------------------------------------------------------------------------
# main
#

main:

setModPars -s .colour.UCS.Luv.Hsl -R 0 $DpMin $DpMax -R 1 $DnMin $DnMax -R 2 $DfMin $DfMax  < PtbpModel
| buildColourModel -r $res | patchSoundModel > SoundColourModel
cmodelVu -b -r -m SoundColourModel -o horizontal.ras -s 0
cmodelVu -b -r -m SoundColourModel -o vertical.ras -s 1

if ($slices == true) then
SoundModelView.csh
endif


goto end

#-------------------------------------------------------------------------------
# help
#
synopsis:
echo “$0 [-r <$res> gamut resolution] [-s = make slices]”
echo “make a SoundColourModel from a PtbpModel”

end:
```

Appendix 10-1: HeadSpace.orc

```
;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;
;;
;;  Stephen Barrass
;;  research
;;  June 14 1995
;;
;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;
;
;
; 24 tracks of 3d sound
; uses binaural delay, low-pass filter and amplitude
; to provide rough position and distance cues
;
; uses the zak gen to do array routing
;
sr = 22050
kr = 2205
ksmps = 10
nchnls = 2

; global        ------------------------------------------------
ginyq = sr*0.4
gipi = 3.14159265
giattack = 0.005
girelease = 0.005

; channels
gichannels = 24
zakinit gichannels,gichannels

; pseudo HRTF
gidelaymax = 0.001
gidelaymin = 0.0001
gifmin = 2.0/gidelaymax
gifscale = ginyq/2.171
; pinnae notch cue
ginotchf = 11000
ginotchw = 200

giamplut = 90                   ; distance-to-amplitude lookup table

; angle of the ear relative to the side of the head (in degrees)
giearangle = 30
; ellipse
gie = 0.6                       ; eccentricity
gik = (1-gie)/gie               ; normalise

;
; cable
; plug a sound into a channel
; instr start   dur     channel patch   pitch
;
;==========================================================
; channel
instr   1

; input parameters              --------------------------------------------
idur = p3
ich = p4                        ; channel
ipatch = p5                     ; patch
ipitch = p6                     ; pitch octave.class format 8.00 = middle C
; derived parameters            --------------------------------------------
```

```
ifroot = cpspch(ipitch)        ; root frequency in hz
print idur, ich, ipatch, ipitch, ifroot

kamp    linen  1, giattack, idur, girelease ; declicking envelope
; use vibrato to bind streams
ivib    iunirand 40; random vibrato frequency
iamp= 4                  ; amplitude of vibrato
print ivib
kvib    oscil  iamp,ivib+40,91     ; holds stream together
; use pitchbend to point to a stream
kheyzkr    ich
ach     loscil  1, ifroot+kvib+khey, ipatch
zaw     ach, ich                ; route to the z channel
endin

;==========================================================
; attract attention to a stream
; by a brief pitchbend
; this writes to a zak variable which is read by the stream instr
; POINTER
; i2     start   dur    channel
instr   2

; input parameters            -------------------------------------------
idur = p3
ich = p4                     ; channel

; derived parameters          -------------------------------------------
print idur, ich

iamp =30                   ; bend size
khey lineniamp, idur/3, idur, idur/3; bend
zkwkhey, ich          ; write to the zk variable

endin

;==========================================================
; this placer uses the declick envelope
; placer
; place a channel in the scene
; and write to stereo output
;
;
; instr start   dur     channel distance        angle   height

instr   3

; input parameters            ---------------------------------------
idur = p3
ach      zar p4                ; channel to play
idistance = p5                 ; distance 0 to 127 metres
iangle = p6                    ; angle 0_360 degrees where 0 = right hand
iheight = p7                   ; height -10_10 below_level_above
print idur,idistance,iangle,iheight

; derived parameters          ---------------------------------------
idb     table   idistance, giamplut     ; convert distance-to-amplitude
iamp = ampdb(idb)
irevtime = idistance/5                 ; reverb in seconds
iradians = iangle*2*gipi/360           ; convert degreees-to-radians
print idb, iamp, iradians, irevtime

; delay
ix = cos(iradians)
idelay = abs(ix)*gidelaymax + gidelaymin
```

```
irdelay = (ix < 0 ? idelay : gidelaymin)
ildelay = (ix < 0 ? gidelaymin : idelay)

printks “ix = %f, idelay = %f, ildelay = %f, irdelay = %f\n “, 10, ix, idelay, ildelay, irdelay

; rough up a HRTF using elliptical functions for each ear
; right ear
irearangle = iangle-giearangle
irhrtf = gik*gie/(1-gie*cos(irearangle*2*gipi/360))

; left ear
ilearangle = iangle-180+giearangle
ilhrtf = gik*gie/(1-gie*cos(ilearangle*2*gipi/360))

print ilhrtf,irhrtf

; k-rate          ------------------------------------------------
; de-clicking envelope
kamp linen iamp, giattack, idur, girelease

; a-rate          ------------------------------------------------

; distance related reverb
;arev   reverb ach, irevtime

; angle positioning by binaural delay
; note : there really should be frequency dependence here...
; used fixed delay lines because deltap is stuffed !
al0 = ach*ilhrtf
ar0 = ach*irhrtf

al1   delay   al0, ildelay
ar1   delay   ar0, irdelay

; attenuate upper frequencies for head shadow
al2  tone    al1, exp(ilhrtf)*gifscale
ar2  tone    ar1, exp(irhrtf)*gifscale

; put in a frequency notch as a pinnae cue
;al3    areson al2, ginotchf, ginotchw
;ar3    areson ar2, ginotchf, ginotchw

;dispfft al2, 0.3, 1024
;dispfft ar2, 0.3, 1024

outs    al2*kamp, ar2*kamp

endin
```

Appendix 10-2: cOcktail.prl

```
#!/local/bin/perl5 -w
#
################################################################################
#COPYRIGHT (C) CSIRO DIT CSIS 1994
#
#SOURCE FILEbatch.prl
#
#MODULE      Sonify
#
#SYSTEM      Thesis
#
#ENVIRONMENTSun SPARC and SUNOS4.1.2
#
#AUTHORStephen Barrass, CSIS
#
#HISTORY
#            : First written October 1995
################################################################################

$dur = 10;
$scorefile = ‘tada.sco’;
$backdistance = 70;
$backscale = 20;
$angleplus = 13;

#-------------------------------------------------------------------------------
# help message
#
sub help {
print STDERR “Synopsis : $0\nGenerate the tada cocktail score\n”;
print STDERR “[-h]\t\t\t\tthis help message\n”;
print STDERR “[-d duration <$dur>]\t\tduration of each section\n”;
}

################################################################################
# subroutines
#

#-------------------------------------------------------------------------------
# placer
# place a channel in the scene
# and write to stereo output
#
$channelInst = 1;
$placerInst = 3;

# start duration distance angle height
sub placerHead {
print “\n;circle\n;instr\tstart\tdur\tchannel\tdist\tangle\theight\n”;
}


sub placer {
local($timeBegin, $duration, $channel, $distance, $angle, $height) = @_;
print “i$placerInst\t$timeBegin\t$duration\t$channel\t$distance\t$angle\t$height\n”;

}

#-------------------------------------------------------------------------------
# circle
# do circle movement
# timeBegin timeDur channel arcBegin arcAngle steps distBegin distEnd scorefile
```

```
sub dec2 {
my($in)=@_;
return (int($in*100)/100);
}
sub remains {
my($nominator, $divisor)=@_;
while ($nominator > $divisor) {
      $nominator -= $divisor;
     }
return dec2($nominator);
}
sub circle {
local($timeBegin,$timeDur,$channel,$arcBegin,$arcAngle,$steps,$distBegin,$distEnd,$scorefile) = @_;
$angle = $arcBegin;
$angleStep = $arcAngle/$steps;
$durStep = $timeDur/$steps;
$height = 0;

placerHead;
placer $timeBegin,$durStep,$channel,$distBegin,$angle,$height;

for ($i = 1; $i < $steps-1; $i++)
     {
      $angle = remains($angle+$angleStep,360);
#     $angle = ($angle+$angleStep)%360;
      placer '+', '.', '.', '>', $angle, '.';
     }
$angle = remains($angle+$angleStep,360);
#$angle = ($angle+$angleStep)%360;
placer '+', '.', '.', $distEnd, $angle, '.';
}

#-------------------------------------------------------------------------------
# readData
#

$channel = 1;

sub readData {
print STDERR readData;

local($datafile) = @_;
local($cm, $amount);
print STDERR "readData file $datafile\n";
# get each data line from the file
open(F,"$datafile") || die "Can't open file $!\n";

#local($angle) = 0;
local($angleStep);
local($distance) = 30;
local($height) = 0;
local($cmPrev) = -10;
local(@dataline);
placerHead;
placer 0,0,$channel,$distance,$angle,$height;
  $i=0;

while (<F>) {
         next if /^#/;
       next if /FILE/;
      next if /COLUMN/;
             $i++;
      if ($i>$length) {last;}
      @dataline = split();
# just get the last 2 columns in cm, amount
```

```
      $cm = $dataline[$#dataline-1];
      $amount = $dataline[$#dataline];
#     print STDERR "$cm,$amount\n";
#     $angle = ($angle+$amount+360)%360.0;

#     $distance = 1/exp(-abs($amount)/5);
#     $distance = int($distance*5);
      $distance = $backdistance-$backscale*abs($amount);
      $dur = (int($cm - $cmPrev))/10.0;

      placer '+',$dur,$channel,1+abs($distance),$angle,$height;
      $cmPrev = $cm;
     }
close(F);
$channel++;
$angle += $angleplus;
}

#-------------------------------------------------------------------------------
# lounge
#
sub lounge {
local($dur, $scorefile) = @_;
print STDERR "LOUNGE,$dur,$scorefile\n";
#system 'ls c*'
# start duration channel arcBegin arcAngle steps distBegin distEnd scorefile
circle 0,$dur,1,0,360,10,5,6;
circle 0,$dur,2,90,360,10,3,6;
circle 0,$dur,3,180,360,10,4,8;

}

#-------------------------------------------------------------------------------
# main
#
print STDERR "Cocktail\n";

#-------------------------------------------------------------------------------
# parse command line
#
$length = 10000;
$angle = 0;
while ($_ = $ARGV[0], /^-/) {
  shift;
if (/^-l/) {
      $length = shift;
     }
elsif (/^-h/)
     {
            help;
           exit 1;
     }
   else
     {
            help;
           exit 1;
     }
}

foreach $file (@ARGV) {
print STDERR "$file\n";
readData($file);
}

exit;
```

```
# add background
$backdur = 140;
$arcAngle = 360;
$backstep = int($backdur*4);
$backchn = 16;
# timeBegin timeDur channel arcBegin arcAngle steps distBegin distEnd scorefile
$placerInst = 3;
#circle(20,$backdur,$backchn,90,$arcAngle,$backstep, $backdistance, $backdistance);
#circle(30, 20, 17, 30, 1000, 20, 50, 100);
```