Chapter 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 inline 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.

<table>
<thead>
<tr>
<th><strong>Reading</strong></th>
<th>The most direct representation is the one with the shortest psychological description [Norman D.A. (1991)]</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Type</strong></td>
<td>An appropriate representation provides the information required by the task: neither more nor less [Norman D.A. (1991)]</td>
</tr>
</tbody>
</table>

**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.

<table>
<thead>
<tr>
<th><strong>Level</strong></th>
<th>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)]</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Organisation</strong></td>
<td>Useful information involves regrouping. The interactive reorganisation of relations between elements can uncover information in the interplay of the data [Bertin J. (1981)]</td>
</tr>
<tr>
<td><strong>Range</strong></td>
<td>Any undetectable element is useless. Utilise the entire range of variation [Bertin J. (1981)]</td>
</tr>
</tbody>
</table>

Table 6-1: Some principles for information design
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.
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.

<table>
<thead>
<tr>
<th>Information Type</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>boolean</td>
<td>2 qualitatively different categories</td>
</tr>
<tr>
<td>nominal</td>
<td>qualitative difference without order</td>
</tr>
<tr>
<td>ordinal</td>
<td>qualitative difference with order</td>
</tr>
<tr>
<td>ordinal-with-zero</td>
<td>qualitative difference, with order and a zero</td>
</tr>
<tr>
<td>ordinal-bilateral</td>
<td>qualitative difference, with order and a central zero</td>
</tr>
<tr>
<td>interval</td>
<td>quantitative difference, with order and metric</td>
</tr>
<tr>
<td>ratio</td>
<td>quantitative difference, with order, metric, and a zero</td>
</tr>
</tbody>
</table>

**Table 6-4: TaDa Information types**

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.

<table>
<thead>
<tr>
<th>Perceptual Relation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>difference</td>
<td>all perceptual elements are detectably different</td>
</tr>
<tr>
<td>order</td>
<td>the perceptual elements have a discernable order</td>
</tr>
<tr>
<td>metric</td>
<td>there is a unit of equal perceptual difference</td>
</tr>
<tr>
<td>zero</td>
<td>an absolute point of reference for variations at any scale</td>
</tr>
</tbody>
</table>

**Table 6-5: Elementary information 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 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
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.

#### Table 6-7: Auditory difference

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

#### Table 6-8: Auditory 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.

#### Table 6-8: Auditory order

<table>
<thead>
<tr>
<th></th>
<th>vibrato.orc</th>
<th>order.sco</th>
</tr>
</thead>
<tbody>
<tr>
<td>sr</td>
<td>8000</td>
<td>f1 0 8193 10 1</td>
</tr>
<tr>
<td>kr</td>
<td>800</td>
<td>;instr start dur</td>
</tr>
<tr>
<td>ksmps</td>
<td>10</td>
<td>parameter 0.0-1.0</td>
</tr>
<tr>
<td>instr</td>
<td>1</td>
<td>i1 0 1 0.67</td>
</tr>
<tr>
<td>k1</td>
<td>oscil 50,p4*10.1</td>
<td>i1 + . 0.10</td>
</tr>
<tr>
<td>aout</td>
<td>pluck 10000,220+k1,220,0.1</td>
<td>i1 + . 0.23</td>
</tr>
<tr>
<td>out</td>
<td>aout</td>
<td></td>
</tr>
<tr>
<td>endin</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
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.

<table>
<thead>
<tr>
<th>pluck.orc</th>
<th>metric.sco</th>
</tr>
</thead>
<tbody>
<tr>
<td>sr = 8000</td>
<td>f1 0 8193 10 l</td>
</tr>
<tr>
<td>kr = 800</td>
<td>:sin</td>
</tr>
<tr>
<td>ksmcps = 10</td>
<td>;instr start dur parameter 0.0-1.0</td>
</tr>
<tr>
<td>instr 1</td>
<td>i1 0 1 0.0</td>
</tr>
<tr>
<td>icps = cpsoct(6.0+p4)</td>
<td>i1 + . &gt;</td>
</tr>
<tr>
<td>aout pluck 10000,icps,icps,0,1</td>
<td>i1 + . 1.0</td>
</tr>
<tr>
<td>out aout</td>
<td>endin</td>
</tr>
</tbody>
</table>

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.

<table>
<thead>
<tr>
<th>nat0.sco for geiger.orc</th>
<th>orig0.sco for vibrato.orc</th>
</tr>
</thead>
<tbody>
<tr>
<td>f1 0 8193 10 l ; sin</td>
<td>f1 0 8193 10 l ; sin</td>
</tr>
<tr>
<td>;instr start dur parameter 0-9</td>
<td>;instr start dur parameter 0.0-1.0</td>
</tr>
<tr>
<td>i1 0 1 4</td>
<td>i1 0 1 0.4</td>
</tr>
<tr>
<td>i1 + . 0</td>
<td>i1 + . 0.0</td>
</tr>
<tr>
<td>i1 + . 8</td>
<td>i1 + . 0.8</td>
</tr>
</tbody>
</table>

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 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.

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

Table 6-11: Elementary characterisation of some sounds
<table>
<thead>
<tr>
<th>Sound relation</th>
<th>Difference qualitative/quantitative</th>
<th>Order 1D, 2D, 3D, nD</th>
<th>Metric ratio/difference unit</th>
<th>Zero natural original conventional</th>
</tr>
</thead>
<tbody>
<tr>
<td>music tempo</td>
<td>qualitative</td>
<td>1D</td>
<td>-</td>
<td>natural</td>
</tr>
<tr>
<td>machine rate</td>
<td>qualitative</td>
<td>1D</td>
<td>-</td>
<td>natural</td>
</tr>
<tr>
<td>machine work</td>
<td>qualitative</td>
<td>1D</td>
<td>-</td>
<td>conventional</td>
</tr>
<tr>
<td>pitch class</td>
<td>qualitative</td>
<td>1D</td>
<td>difference Semitone</td>
<td>conventional</td>
</tr>
<tr>
<td>event force</td>
<td>quantitative</td>
<td>1D</td>
<td>-</td>
<td>natural</td>
</tr>
<tr>
<td>drum stretch</td>
<td>quantitative</td>
<td>1D</td>
<td>-</td>
<td>original</td>
</tr>
<tr>
<td>fuzz level</td>
<td>quantitative</td>
<td>1D</td>
<td>-</td>
<td>original</td>
</tr>
<tr>
<td>reverb wetness</td>
<td>quantitative</td>
<td>1D</td>
<td>-</td>
<td>original</td>
</tr>
<tr>
<td>vibrato rate</td>
<td>quantitative</td>
<td>1D</td>
<td>-</td>
<td>original</td>
</tr>
<tr>
<td>vibrato depth</td>
<td>quantitative</td>
<td>1D</td>
<td>-</td>
<td>original</td>
</tr>
<tr>
<td>tremolo rate</td>
<td>quantitative</td>
<td>1D</td>
<td>-</td>
<td>original</td>
</tr>
<tr>
<td>tremolo depth</td>
<td>quantitative</td>
<td>1D</td>
<td>-</td>
<td>original</td>
</tr>
<tr>
<td>phasor depth</td>
<td>quantitative</td>
<td>1D</td>
<td>-</td>
<td>original</td>
</tr>
<tr>
<td>phasor rate</td>
<td>quantitative</td>
<td>1D</td>
<td>-</td>
<td>original</td>
</tr>
<tr>
<td>brightness</td>
<td>quantitative</td>
<td>1D</td>
<td>ratio Acum</td>
<td>original</td>
</tr>
<tr>
<td>object size</td>
<td>quantitative</td>
<td>1D</td>
<td>-</td>
<td>conventional</td>
</tr>
<tr>
<td>filling a bottle</td>
<td>quantitative</td>
<td>1D</td>
<td>-</td>
<td>conventional</td>
</tr>
<tr>
<td>rolling marble</td>
<td>quantitative</td>
<td>2D</td>
<td>-</td>
<td>conventional</td>
</tr>
<tr>
<td>granular density</td>
<td>quantitative</td>
<td>1D</td>
<td>-</td>
<td>conventional</td>
</tr>
<tr>
<td>pitch scale</td>
<td>quantitative</td>
<td>1D</td>
<td>difference Semitone ratio Mel</td>
<td>conventional</td>
</tr>
<tr>
<td>repetition rate</td>
<td>quantitative</td>
<td>1D</td>
<td>ratio B = 1.0</td>
<td>natural</td>
</tr>
<tr>
<td>white noise duration</td>
<td>quantitative</td>
<td>1D</td>
<td>ratio B = 1.1</td>
<td>natural</td>
</tr>
<tr>
<td>binaural loudness</td>
<td>quantitative</td>
<td>1D</td>
<td>ratio B = 0.6</td>
<td>natural</td>
</tr>
<tr>
<td>monaural loudness</td>
<td>quantitative</td>
<td>1D</td>
<td>ratio B = 0.54</td>
<td>natural</td>
</tr>
</tbody>
</table>

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)].

<table>
<thead>
<tr>
<th>Level</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>local</td>
<td>can answer questions about a single element</td>
</tr>
<tr>
<td>intermediate</td>
<td>can answer questions about subsets and groups of elements</td>
</tr>
<tr>
<td>global</td>
<td>can answer questions about the entire set of elements as a whole</td>
</tr>
</tbody>
</table>

**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.

<table>
<thead>
<tr>
<th>Level</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>local</td>
<td>answered by listening to an element within a stream</td>
</tr>
<tr>
<td>intermediate</td>
<td>answered by listening to a stream</td>
</tr>
<tr>
<td>global</td>
<td>answered by listening to an auditory scene</td>
</tr>
</tbody>
</table>

**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 straightforward 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>
<thead>
<tr>
<th>schema.orc</th>
<th>schema.sco</th>
</tr>
</thead>
<tbody>
<tr>
<td>sr = 8000</td>
<td>f1 0 8193 10</td>
</tr>
<tr>
<td>kr = 800</td>
<td>t 0 200</td>
</tr>
<tr>
<td>ksmps = 10</td>
<td></td>
</tr>
<tr>
<td>; tone</td>
<td></td>
</tr>
<tr>
<td>instr 1</td>
<td></td>
</tr>
<tr>
<td>kamp</td>
<td></td>
</tr>
<tr>
<td>linen 3000</td>
<td>i0 0 1 8.04</td>
</tr>
<tr>
<td></td>
<td>i0 1 1 8.02</td>
</tr>
<tr>
<td>out</td>
<td>i0 2 2 8.00</td>
</tr>
<tr>
<td>oscil kamp</td>
<td></td>
</tr>
<tr>
<td>, cpspch(p4), 1</td>
<td></td>
</tr>
<tr>
<td>out aout</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>; noise</td>
<td></td>
</tr>
<tr>
<td>instr 2</td>
<td></td>
</tr>
<tr>
<td>kamp</td>
<td></td>
</tr>
<tr>
<td>linen 5000</td>
<td></td>
</tr>
<tr>
<td>i0 =</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>out</td>
<td></td>
</tr>
<tr>
<td>oscil kamp</td>
<td></td>
</tr>
<tr>
<td>, kamp+i1, 1</td>
<td></td>
</tr>
<tr>
<td>out aout</td>
<td></td>
</tr>
<tr>
<td>endin</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>; uncomment the next line to remove noise bursts</td>
<td></td>
</tr>
<tr>
<td>:e</td>
<td></td>
</tr>
<tr>
<td>i2 3.5 1 8.00 ; interference</td>
<td></td>
</tr>
<tr>
<td>i2 5 1 8.00 ; missing</td>
<td></td>
</tr>
<tr>
<td>i2 8 1 8.00 ; missing</td>
<td></td>
</tr>
<tr>
<td>i2 9 0.5 8.00 ; interference</td>
<td></td>
</tr>
<tr>
<td>i2 10 2 8.00 ; missing</td>
<td></td>
</tr>
<tr>
<td>i2 13 0.5 8.00 ; missing</td>
<td></td>
</tr>
</tbody>
</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 reor-
ganise 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.

<table>
<thead>
<tr>
<th>range.orc</th>
<th>range.sco</th>
</tr>
</thead>
<tbody>
<tr>
<td>sr = 8000</td>
<td>f1 0 8193 10 1 ; sin</td>
</tr>
<tr>
<td>kr = 800</td>
<td>; instr start dur answer</td>
</tr>
<tr>
<td>ksmpps = 10</td>
<td>i1 0 1 1 ; light</td>
</tr>
<tr>
<td>instr 1</td>
<td>i1 + . 2 ; medium</td>
</tr>
<tr>
<td>kamp linen</td>
<td>i1 + . 3 ; heavy</td>
</tr>
<tr>
<td>0.01</td>
<td></td>
</tr>
<tr>
<td>k1 rand aout oscil kamp, 400+k1, 1</td>
<td></td>
</tr>
<tr>
<td>1600</td>
<td></td>
</tr>
<tr>
<td>out aout</td>
<td></td>
</tr>
<tr>
<td>endin</td>
<td></td>
</tr>
</tbody>
</table>

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 \( s=(x_{21}, x_2, x_3, x_4, x_5, x_6) \) belongs to Set 2 (dirt) if

\[
\begin{align*}
x_2^* x_2 + x_3^* x_3 + x_4^* x_4 + x_5^* x_5 + x_6^* x_6 & \leq 1.5^* 1.5 \\
x_1^* x_1 + x_3^* x_3 + x_4^* x_4 + x_5^* x_5 + x_6^* x_6 & \leq 1.5^* 1.5 \\
x_1^* x_1 + x_2^* x_2 + x_4^* x_4 + x_5^* x_5 + x_6^* x_6 & \leq 1.5^* 1.5
\end{align*}
\]

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 \( x_1, x_2, x_3 \) 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.
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 uncom-
menting them in the score file.

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

### 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.

<table>
<thead>
<tr>
<th>#Correct</th>
<th>10</th>
<th>9</th>
<th>8</th>
<th>7</th>
<th>6</th>
<th>5</th>
<th>4</th>
<th>3</th>
<th>2</th>
<th>1</th>
<th>0</th>
</tr>
</thead>
<tbody>
<tr>
<td>icad1 #subjects</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>3</td>
<td>7</td>
<td>3</td>
<td>-3?</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>icad2 #subjects</td>
<td>0</td>
<td>1</td>
<td>6</td>
<td>7</td>
<td>4</td>
<td>4</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>icad3 #subjects</td>
<td>3</td>
<td>7</td>
<td>7</td>
<td>5</td>
<td>-1?</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
</tbody>
</table>

### 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 de-
signs, as shown in Figure 6-2, indicating that the principles are of benefit in practice.

<table>
<thead>
<tr>
<th>#Correct</th>
<th>10</th>
<th>9</th>
<th>8</th>
<th>7</th>
<th>6</th>
<th>5</th>
<th>4</th>
<th>3</th>
<th>2</th>
<th>1</th>
<th>0</th>
</tr>
</thead>
<tbody>
<tr>
<td>gold1 #subjects</td>
<td>1</td>
<td>7</td>
<td>10</td>
<td>6</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>gold2 #subjects</td>
<td>1</td>
<td>7</td>
<td>8</td>
<td>6</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>gold3 #subjects</td>
<td>1</td>
<td>4</td>
<td>4</td>
<td>6</td>
<td>7</td>
<td>3</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

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

![Figure 6-2](image)

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.

```plaintext
levelok.sco - goes with fm.orc
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 0.198 0.358 0.065
dirt01
i1 1 1 0.130 0.096 1.686 0.108 1.020 1.198
gold01
gold1
i1 2 1 0.292 0.699 0.153 0.303 0.064
dirt02

g01
gold1
dirt01

gold2
gold3
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.

<table>
<thead>
<tr>
<th>gold.orc</th>
</tr>
</thead>
<tbody>
<tr>
<td>sr = 8000</td>
</tr>
<tr>
<td>kr = 800</td>
</tr>
<tr>
<td>ksmps = 10</td>
</tr>
<tr>
<td>nchnls = 2</td>
</tr>
<tr>
<td>gir = 100</td>
</tr>
<tr>
<td>gis = 1000</td>
</tr>
<tr>
<td>instr 1</td>
</tr>
<tr>
<td>kamp linen 10000, 0.01, p3, 0.1</td>
</tr>
<tr>
<td>ao0 oscili gir, gis*p4, l</td>
</tr>
<tr>
<td>ao1 oscili gir, gis*p5+ao0, l</td>
</tr>
<tr>
<td>ao2 oscili gir, gis*p6+ao1, l</td>
</tr>
<tr>
<td>ao3 oscili gir, gis*p7+ao2, l</td>
</tr>
<tr>
<td>ao4 oscili gir, gis*p8+ao3, l</td>
</tr>
<tr>
<td>ao5 oscili kamp, gis*p9+ao4, l</td>
</tr>
<tr>
<td>; separate d3 and d4 in space</td>
</tr>
<tr>
<td>a1,a2,a3,a4 pan ao5, p6, p7, 2, 1, 1</td>
</tr>
<tr>
<td>outs a1, a2</td>
</tr>
<tr>
<td>endin</td>
</tr>
</tbody>
</table>

**Table 6-20:** Gold detector
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.

### Table 6-21: Handfuls of dirt and gold

<table>
<thead>
<tr>
<th>dirt.sco</th>
<th>gold.sco</th>
</tr>
</thead>
<tbody>
<tr>
<td>f1 0 8193 10 1 ; sin</td>
<td>f1 0 8193 10 1 ; sin</td>
</tr>
<tr>
<td>f2 0 129 7 0 128 1 ; ramp 0 to 1</td>
<td>f2 0 129 7 0 128 1 ; ramp 0 to 1</td>
</tr>
<tr>
<td>t 0 200 ; tempo</td>
<td>t 0 200 ; tempo</td>
</tr>
<tr>
<td>; d1 d2 d1 d2 d3 d4 d5 d6</td>
<td>; d1 d2 d1 d2 d3 d4 d5 d6</td>
</tr>
<tr>
<td>i1 0.830 1 0.830 0.812 0.170 0.064 0.417 0.897</td>
<td>i1 1.394 1 1.394 0.775 0.080 0.446 0.541 1.631</td>
</tr>
<tr>
<td>i1 0.170 1 0.170 1.062 0.085 0.716 0.080 0.871</td>
<td>i1 0.807 1 0.807 1.411 0.515 0.851 0.681 0.749</td>
</tr>
<tr>
<td>i1 0.671 1 0.671 0.237 0.380 0.689 0.899 0.600</td>
<td>i1 1.355 1 1.355 0.099 0.235 2.054 0.656 2.127</td>
</tr>
<tr>
<td>i2 3.80 1 2.380 0.690 0.355 0.006 0.691 0.437</td>
<td>i2 0.498 1 0.498 2.482 0.370 0.743 1.580 0.415</td>
</tr>
<tr>
<td>i1 0.551 1 0.551 0.377 0.007 0.112 0.051 0.246</td>
<td>i1 1.197 1 1.197 0.411 3.198 0.370 1.241 0.373</td>
</tr>
<tr>
<td>i1 0.242 1 0.242 1.417 0.275 0.405 0.590 0.810</td>
<td>i1 1.955 1 1.955 0.433 1.390 0.067 1.314 0.858</td>
</tr>
<tr>
<td>i1 0.404 1 0.404 1.750 0.831 0.247 0.060 0.767</td>
<td>i1 0.330 1 0.330 0.572 0.815 0.574 0.778 1.606</td>
</tr>
<tr>
<td>i1 0.443 1 0.443 1.472 0.297 1.050 0.046 0.455</td>
<td>i1 2.019 1 2.019 0.317 0.737 1.144 0.193 0.705</td>
</tr>
<tr>
<td>i1 0.180 1 0.180 1.112 0.245 0.620 0.782 0.547</td>
<td>i1 0.996 1 0.996 0.426 0.613 1.612 0.157 0.711</td>
</tr>
<tr>
<td>i1 0.007 1 0.007 0.286 1.223 1.339 0.300 0.475</td>
<td>i1 0.775 1 0.775 0.359 0.589 0.874 0.281 1.618</td>
</tr>
<tr>
<td>i1 0.640 1 0.640 0.393 0.642 0.941 0.670 0.123</td>
<td>i1 1.335 1 1.335 0.280 1.014 0.407 2.072 0.138</td>
</tr>
<tr>
<td>i1 0.737 1 0.737 0.427 0.796 0.723 0.194 0.149</td>
<td>i1 0.251 1 0.251 2.937 0.902 0.905 1.296 0.118</td>
</tr>
<tr>
<td>i1 0.565 1 0.565 0.571 0.005 0.340 0.507 0.168</td>
<td>i1 1.182 1 1.182 0.186 0.503 1.302 0.807 1.188</td>
</tr>
<tr>
<td>i1 0.976 1 0.976 0.342 0.090 0.536 0.532 0.461</td>
<td>i1 0.782 1 0.782 0.134 1.671 2.251 2.005 0.678</td>
</tr>
<tr>
<td>i1 1.165 1 1.165 0.510 0.212 0.900 0.736 0.574</td>
<td>i1 1.206 1 1.206 1.025 0.267 0.584 1.056 0.116</td>
</tr>
<tr>
<td>i1 0.821 1 0.821 0.291 1.185 0.465 0.061 0.392</td>
<td>i1 0.674 1 0.674 0.867 0.567 1.283 1.531 1.046</td>
</tr>
<tr>
<td>i1 2.195 1 2.195 0.776 1.027 0.510 0.517 0.135</td>
<td>i1 0.207 1 0.207 0.244 0.969 0.523 1.060 1.342</td>
</tr>
<tr>
<td>i1 0.430 1 0.430 1.427 0.183 0.050 0.370 0.563</td>
<td>i1 0.907 1 0.907 1.190 0.834 0.281 1.129 0.029</td>
</tr>
<tr>
<td>i1 0.243 1 0.243 0.717 0.531 0.597 0.376 1.170</td>
<td>i1 0.453 1 0.453 0.733 1.138 0.840 1.078 0.324</td>
</tr>
<tr>
<td>i1 0.042 1 0.042 0.199 0.745 0.364 0.399 1.086</td>
<td>i1 0.086 1 0.086 0.699 0.167 1.241 0.980 0.222</td>
</tr>
</tbody>
</table>

### 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.

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.

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.

<table>
<thead>
<tr>
<th>Colour relation</th>
<th>Difference category/ continuous</th>
<th>Order 1D, 2D, 3D, nD</th>
<th>Metric ratio/ difference</th>
<th>Zero natural/ original/conventional</th>
<th>Just noticeable differences</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hue</td>
<td>category</td>
<td>difference</td>
<td>conventional</td>
<td>~ 24</td>
<td></td>
</tr>
<tr>
<td>Saturation</td>
<td>continuous</td>
<td>1D</td>
<td>ratio</td>
<td>original</td>
<td>~ 5</td>
</tr>
<tr>
<td>Lightness</td>
<td>continuous</td>
<td>1D</td>
<td>ratio</td>
<td>natural</td>
<td>~ 100</td>
</tr>
</tbody>
</table>

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)](image)

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:

<table>
<thead>
<tr>
<th>Elementary relation/ISS dimension</th>
<th>Difference</th>
<th>Order</th>
<th>Metric</th>
<th>Zero</th>
</tr>
</thead>
<tbody>
<tr>
<td>Angle</td>
<td>qualitative</td>
<td>8 steps</td>
<td></td>
<td>conventional</td>
</tr>
<tr>
<td>Radius</td>
<td>quantitative</td>
<td>yes</td>
<td>8 steps</td>
<td>natural</td>
</tr>
<tr>
<td>Vertical</td>
<td>quantitative</td>
<td>yes</td>
<td>100 steps</td>
<td>conventional</td>
</tr>
</tbody>
</table>

**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.

<table>
<thead>
<tr>
<th>Info Type</th>
<th>Angle category</th>
<th>Radial zero</th>
<th>Vertical metric</th>
<th>IPS Mapping</th>
</tr>
</thead>
<tbody>
<tr>
<td>Boolean</td>
<td>X</td>
<td></td>
<td></td>
<td>Opposite angles</td>
</tr>
<tr>
<td>Nominal</td>
<td>X</td>
<td></td>
<td></td>
<td>Circle</td>
</tr>
<tr>
<td>Ordinal</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>Coil</td>
</tr>
<tr>
<td>Ordinal and zero</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>Spiral</td>
</tr>
<tr>
<td>Ordinal bilateral</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>Sloped Line</td>
</tr>
<tr>
<td>Interval</td>
<td></td>
<td>X</td>
<td></td>
<td>Vertical Line</td>
</tr>
<tr>
<td>Ratio</td>
<td></td>
<td>X</td>
<td>X</td>
<td>Radial Line</td>
</tr>
</tbody>
</table>

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 de-
rived 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.

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 7-5: Caivano’s organisation

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 Hear-say principles. The investigation involved the generation of prototype sound sequences for testing the perceptual characteristics of the space. The equipment was a Sun SPARCstation™ 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

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 in-
strument 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](image)

### 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.
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.

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

**Table 7-4:** Pitch Circle rated against Pedestal criteria
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.

![Formant Diagram](image)

**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.

\[\text{F1 = abscissa; F2 = ordinate. The data points represent the measured formant frequencies of the vowels that were correctly identified by a panel of listeners. (From Peterson and Barney 1952.)}\]
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.
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)].
Chapter 7 • Information-Sound Space: a cognitive artefact for auditory design 105

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.

Fig. 3. Average ratings of both groups of subjects for all sounds (sound numbers as in Fig. 2) on the scales dull-sharp and compact-scattered, which represented the two most important related factors of the musicians. Full lines indicate the effect of increasing the upper limiting frequency, dashed lines that of increasing the slope of the spectral envelope. Dash-dotted lines lead from harmonic complex tones to noise with equal spectral envelopes.
(a) Musicians,
(b) non-musicians.
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 (0,1,2), (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.
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-violin-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 resynthesising 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 synchro-
nicity in the development of upper harmonics and the presence of low-energy high frequency noise during the attack segment.

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 al-
located 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™ 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.
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 rechoosing 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.

<table>
<thead>
<tr>
<th>Point</th>
<th>Grey’s timbre</th>
<th>UltraSound patch</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>flute</td>
<td>flute</td>
</tr>
<tr>
<td>1</td>
<td>cello (muted sul tasto)</td>
<td>cello</td>
</tr>
<tr>
<td>2</td>
<td>bass clarinet</td>
<td>bass clarinet</td>
</tr>
<tr>
<td>3</td>
<td>saxophone (mf)</td>
<td>tenor saxophone</td>
</tr>
<tr>
<td>4</td>
<td>soprano sax</td>
<td>soprano saxophone</td>
</tr>
<tr>
<td>5</td>
<td>English horn</td>
<td>English horn</td>
</tr>
<tr>
<td>6</td>
<td>bassoon</td>
<td>bassoon</td>
</tr>
<tr>
<td>7</td>
<td>trombone (muted)</td>
<td>trombone</td>
</tr>
</tbody>
</table>

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

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

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.
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

<table>
<thead>
<tr>
<th>Pedestal</th>
<th>Zero</th>
<th>Difference</th>
<th>Order</th>
<th>Metric</th>
<th>Level</th>
<th>Range</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pitch</td>
<td>ok</td>
<td>ok</td>
<td>fail (4)</td>
<td>ok</td>
<td>fail (X,X)</td>
<td>ok</td>
</tr>
<tr>
<td>Formant</td>
<td>ok</td>
<td>ok</td>
<td>fail (2)</td>
<td>ok</td>
<td>fail (110,X)</td>
<td>ok</td>
</tr>
<tr>
<td>Static Timbre</td>
<td>ok</td>
<td>ok</td>
<td>fail (4)</td>
<td>fail (3)</td>
<td>ok</td>
<td>(150,110)</td>
</tr>
<tr>
<td>Timbre</td>
<td>ok</td>
<td>ok</td>
<td>fail (1)</td>
<td>ok</td>
<td>ok</td>
<td>(150,150)</td>
</tr>
</tbody>
</table>

**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.
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 Caivano’s 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-pass 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.

<table>
<thead>
<tr>
<th>Criteria</th>
<th>Result</th>
<th>Null hypothesis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Zero</td>
<td>repeating cycle of 8 steps can pick the bright point in the cycle</td>
<td>reject</td>
</tr>
<tr>
<td>Difference</td>
<td>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</td>
<td>accept</td>
</tr>
<tr>
<td>Order</td>
<td>repeating cycles as for Difference</td>
<td>accept</td>
</tr>
<tr>
<td>Metric</td>
<td>regular triplet the levelling off found in the repeating cycles supports the null hypothesis of non-uniform change</td>
<td>accept</td>
</tr>
<tr>
<td>Level</td>
<td>4 = far = 2 sounds 4 = mid = 2 sounds 4 = near = 1 sound</td>
<td>reject</td>
</tr>
<tr>
<td>Range</td>
<td>regular triplet 1 = count 3 2 = count 3 3 = count 2 4 = count 2 5 = count 2 6 = count 1 7 = count 1</td>
<td>accept</td>
</tr>
</tbody>
</table>

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 har-
monics. 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 disappear-
ance 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.

<table>
<thead>
<tr>
<th>Criteria</th>
<th>Result</th>
<th>Null hypothesis</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Zero</strong></td>
<td>repeating cycles silence at the start of each cycle</td>
<td>reject</td>
</tr>
<tr>
<td><strong>Difference</strong></td>
<td>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</td>
<td>accept</td>
</tr>
<tr>
<td><strong>Order</strong></td>
<td>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</td>
<td>accept</td>
</tr>
</tbody>
</table>

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 in-

<table>
<thead>
<tr>
<th>Criteria</th>
<th>Result</th>
<th>Null hypothesis</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Metric</strong></td>
<td>regular triplets</td>
<td>accept</td>
</tr>
<tr>
<td>0 = unequal 2,3 are similar</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1 = equal</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2 = equal</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3 = equal</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4 = equal</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5 = equal</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6 = equal</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7 = equal</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Level</strong></td>
<td>3 = far = 2 sounds</td>
<td>reject</td>
</tr>
<tr>
<td>3 = mid = 2 sounds</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3 = near = 1 sound</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Range</strong></td>
<td>repeating cycle of 5% steps</td>
<td>accept</td>
</tr>
<tr>
<td>0 = count 2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1 = count 3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2 = count 3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3 = count 3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4 = count 3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5 = count 3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6 = count 3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7 = count 3</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

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 in-

Chapter 7• Information-Sound Space: a cognitive artefact for auditory design 119
Chapter 7• Information-Sound Space: a cognitive artefact for auditory design

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.

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

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.

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

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


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 by device frequency characteristics.
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.
7.9 The 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

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

Table 7-14: Elementary representation mappings in TBP ISS
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 axe.

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.