A Perceptual Framework for the Auditory Display of Scientific Data

Stephen Barrass
CSIRO Division of Information Technology
PO Box 664, Canberra, ACT 2601, Australia
Email: stephen.barrass@cbr.did.csiro.au

Abstract

A principal requirement of the auditory display of scientific data is an accurate portrayal of the information contained in the data. The characteristics of the display device can significantly affect the faithfulness of the data presentation. Human hearing is a complex nonlinear process and the intuitive comprehension of the display requires consideration of perceptual interactions and the natural connection between the data and the perception of the data.

This chapter proposes a perceptual framework for observable and systematic specification and comprehension of sounds in a device-independent auditory display of scientific data. The framework consists of a perceptually scaled sound space, and a display model which maps points from the perceptual space to a display device parameter space while preserving the interrelationships between them.

The advantages of a perceptual space are described using the example of an established framework for applying perceptually uniform color models to scientific visualization. A perceptual sound space is defined and constructed by scaling a naturally ordered sound model derived from research studies in timbre perception.

An auditory display is implemented on a Sun Sparc10 workstation using Csound and samples of musical instruments. The display model consists of an interpolated mapping from perceptual space to display space and a description of the display gamut which is the boundary in perceptual space between points that the display can realize and those that it cannot. The display gamut can be used to analyse the display, compare displays and optimize data mapping sequences for the display.

The concepts are illustrated by a graphic visualization of the gamut of the implemented display as a geometric shape in the perceptual sound space.

1 Introduction

Auditory display is a relatively new medium for the exploration and presentation of scientific data. Innovative and novel paradigms are being developed to take advantage of the temporal, spectral, and spatial properties of sounds. Auditory representations have the potential to increase the communication bandwidth at the human-computer interface. Although the prospects are promising there are a number of difficulties which must be addressed if the auditory display of data is to become more than a curiosity.

Some obstacles to the growth of sonification were listed by Smith [1]. These included the need for a means of specifying timbre in terms of a comprehensive device-independent model, and the arbitrary mapping of data to sound parameters without considering any natural connection between data types and perceptual characteristics of sound.

The problems of achieving a balanced auditory display were described by Kramer [2], who found that interactions between display parameters can be confusing. He suggested that the data be mapped to perceptual parameters which do not interfere with each other when they are heard, and pointed out the need for orderliness and intuitive coherence of sequences in the auditory display.

The Sonify module for the AVS visualization system allows data to be mapped to the parameters of a physical model of an acoustic wave. In the user documentation Kaplan [3] notes that the parameters may have different effects depending on the waveform used, and that some combinations may have no apparent effect at all. This indicates the need for a clear distinction to be made between the subjective perceptual and objective physical aspects of sounds.

Important areas which require further research were nominated by Scaletti [4] as the perception and cognition of sounds, tools to get from an imagined sound to a realized sound, and more examples of successful data to sound mappings.

In science there is an imperative to maintain the integrity of the data so that the interpretation is not biased or distorted by processing artifacts. Bertin [5] makes the point that data values carry an elementary level of information, but it is the observable interrelationships between data values which carry intermediate and overall levels of information. The observation of informative structures relies on the perception of contrast, ratios, and order in the data display. The meaningful transfer of higher levels of information in the display requires that the interrelationships between the data points be preserved by the display process.

This chapter proposes a framework for auditory display of scientific data which addresses the issues of

- natural specification of the data display,
- intuitive comprehension of the data display,
- device independence, and
- device characterization.

2 Display and Perception

The perceptual issues in auditory display are fundamental to the human-computer interface, where a transfer of information occurs by transmission through a physical display device and reception through biological sense organs. As with any communication medium, the characteristics of the system can have significant influence on the information which passes through it.

2.1 Subjective/Objective

The subjective experience of sound is related to the objective world through the mechanical transduction and neural encoding of acoustic waves, generated by events in the world, as a spatial arrangement of nerve firing rates which are processed by a variety of centers in the auditory complex. The hearing system attempts to relate the neural pattern of activity back to events in the world, not to physical parameters of the acoustic wave generated by those events (e.g., frequency, intensity, spectrum). The subjective experience of sound covaries with physical parameters in complex nonlinear ways, as can be seen for example in the Fletcher-Munson [10] equal loudness contours which show how perceived loudness varies with frequency and intensity of a sinusoidal stimulus. Some of the many subjectively observable aspects of sounds include loudness, scratchiness, hardness, brightness, pitch, hollowness, roughness, duration, graininess, tenseness, etc. Different sounds may have different observable attributes and may exhibit more or less variation in some of these attributes. Musical sounds are often described by instrument, loudness, pitch, and duration. Everyday sounds have been described by material type, size, and hardness [7], and speech sounds by sex, maturity, and expression [8]. Some of these descriptions of aspects of sounds are more semantic, operating at the cognitive level of learned meanings and associations, while others are more abstract perceptions which have no meaning per se but which carry information for interpretation. This distinction is made by Bregman [9] as learned "schemas" and innate "primitive" perceptions, and he argues the case for a collaborative perceptual system in which the innate influences bootstrap the learning process, based on evidence from experiments on hearing in infants. The physics of acoustic waves are the same everywhere on earth and this leads to the hypothesis that pre-attentive perceptual mechanisms have evolved to take advantage of this consistency, while the learned schemas allow adaption to particular environments and their characteristic ecologies of sounds. Innate pre-attentive perceptions do not require learning and are valid across individuals, age groups and cultures, providing a basis for an intuitive, natural, and immediate sensory language [6].

An auditory display device produces sounds by causing acoustic waves in the air (or perhaps some other medium—e.g., water, earth, bone). The parameters of the display may be directly linked to the physical description of the wave, e.g., frequency, intensity, spectrum etc. However the hearing system does not include a mathematical model of an acoustic wave and does not measure the parameters of such a model. It would be advantageous to be able to control the display device in a more natural way in terms of human hearing, so that systematic variation of the control parameters has regular observable effects. A simple example of this principle is the logarithmic loudness to intensity mapping implemented on many sound output devices.

2.2 How similar problems have been addressed for color

The relationship between the subjective experience and the physical display of color has been a focus of the science of colorimetry and has resulted in the definition of perceptual color spaces. These geometric spaces arrange perceived colors in an observable and systematic manner in which euclidean distance is a measure of psychological dissimilarity. Perceptual color spaces are founded upon the identification of observable and separable perceptual aspects of color. A history of empirical studies has been dedicated to orthogonalizing and building a metric around these bases, resulting in a variety of spaces such as the Munsell, CIE, OSA, Coloroid, and NCS systems, which have different properties depending on the choice of axes and their arrangement, whether additive or subtractive color mixing is supported, the basis used for dissimilarity comparisons, and the weighting of local versus global orthogonality. The scientific visualization community has applied this research to the problems of representing data through color. Robertson [11] proposed a computational framework based on a perceptual color space and lists the advantages of this approach as:

- Intuitive addressability—the specification of data representation in perceptual terms (such as hue, saturation, and lightness).
- Uniformity—the regular representation of numerical data variations by gradations in perceived color, due to the perceptual metric of the space.

- Independent control of lightness and chromatic contrast—the opportunity to expand lightness contrast independent of chromatic contrast, or vice versa.
- Display device characterization in perceptual terms—device in intuitive terms, as a guide to choosing appropriate display representations and controlling their production.
- A basis for complex perceptual data descriptions—the opportunity to use natural color terms or construct models of physical processes in terms of their natural spectral descriptors, lead to more complex 3-D scene syntheses for multidimensional data displays.

2.3 A Brief History of Related Work

A framework for sound with these properties would address the problems of natural specification, perceptual interactions between parameters, intuitive ordering of sequences, and device independence. This has led to the proposal of several models of sound perception derived from color theory.

A physical ratio mapping from acoustic wavelength to electromagnetic wavelength was the basis for Pridmore's [12] device which translated sound input at a microphone into colored lights. The mapping was derived from a link between pitch circularity and the hue circle.

A distinction between objective and subjective analogies was made by Sebba [13], who noted that the notion of human perception of a physical relationship between acoustic and electromagnetic waves was rejected early in the 20th century. However, subjective analogies between color and sound persist and Sebba carried out an experiment to systematically examine whether some sort of psychological relationship does exist. The results show correlations between the observation of structure in sound and color sequences, due to the perception of order, ratios, and contrast.

The Hue, Saturation, Lightness (HSL) color model was used as a template for a model of sound by Caivano [14], who, like Pridmore, made a link between pitch circularity and the hue circle. This cylindrical polar arrangement has pitch angle, timbre radius, and loudness height as axes. The timbre axis is ordered from white noise at the center, through inharmonic spectra, through harmonic spectra, to a simple sinusoid at the extreme radius. This ordering is derived from consideration of the complexity of the physical spectrum but it is not clear that an observer would subjectively hear this sequence as ordered.

A scaled sound space was introduced by Padgham [15] for comparing the timbre of different pipe organs. This timbre assessment chart is a polar plot in which angle represents the characteristic tone caused by the formant region containing the first five 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 "colorfulness," 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. The resulting three dimensional solid characterizes important aspects of the sound gamut of a pipe organ, and can be used for calibration and comparisons. This cylindrical polar sound space has tone angle, complexity radius, and loudness height.

A perceptual sound space must have an observable and systematic basis. This premise underlies the Timbre-Brightness-Pitch model of sound, proposed by Barrass [16] using the results from studies of timbre perception by von Bismarck [17], Slawson [18], and Grey [19]. The geometry of the Timbre-Brightness-Pitch model, shown in Figure 1, is derived from the HSL color model and founded on the recognition that both timbre perception and hue perception are qualitative and categorical. An analogy is drawn between a complex timbre attribute, consisting of a timbre category modified by brightness, and color chromaticness, consisting of a hue modified by saturation.

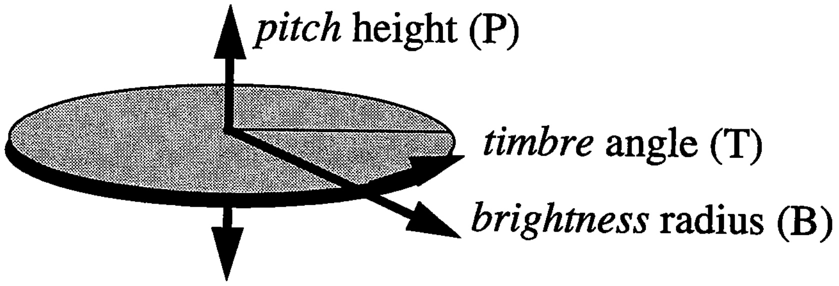

Figure 1: The naturally ordered TBP sound model.

Perceptually salient bilateral axes are used to arrange timbres in a diametrically complementary circle, modelled on the arrangement of hues in the color circle by the red-green and blue-yellow opponent axes. This means that timbres which are most observably similar in terms of the underlying bases of ordering lie close together while those which are most dissimilar lie opposite. This layout provides a means of ordering and showing relationships between qualitative perceptions. Bregman identifies timbre, brightness, and pitch as important for the formation of perceptual streams in auditory scene analysis [9], and it is conjectured that this geometric model may be helpful in the visualization 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 Timbre-Brightness-Pitch 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. The cylindrical polar system has a categorical timbre angle, a scalar brightness radius, and a scalar pitch height as axes.

3 Perceptual Sound Space

An understanding of structure allows higher levels of information to be extracted from a data set. Presentations which enable intuitive comprehension of data relationships can significantly improve access and retrieval of information from data. Sounds can carry structure through perceptions of order, ratios, and contrast. The portrayal of data structure by sounds requires the accurate perception of the data interrelationships.

A perceptual sound space can address the requirement for accurate perception of data interrelationships because it has the properties of:

- observable and systematic order, and
- a perceptual metric.

This chapter proposes to extend the TBP sound model to create a Perceptual Sound Space (PSS). The axes of the TBP sound model provide the necessary naturally ordered basis. The PSS is constructed by perceptually scaling each of the axes of the TBP sound model. Points in the space are perceptually defined, and distance is related to subjective difference between points. The resulting space is a cylindrical polar arrangement, consisting of an Equally Spaced Timbre Circle, a perceptually scaled brightness measure, and a perceptually scaled vertical pitch measure.

The PSS is akin to the Munsell color space [20], and is intended to align itself with Munsell's ideals of psychological equispacing and practical usefulness.

3.1 Equally Spaced Timbre Circle

The Timbre Circle is a cyclic sequence in which timbre is designated by angular displacement. The Equally Spaced Timbre Circle has the following requirements:

- adjacent timbres around the circle are subjectively equally different
- diametrically opposite timbres are subjectively most different

Timbre is a categorical perception and has no natural ordering. However the TBP sound model imposes an observable and systematic ordering based on salient underlying opponent axes. The advantage of this arrangement is that perceptual scaling of the underlying axes maximizes separability, and causes the most dissimilar timbres to lie opposite each other in the circle.

The hearing process allows sounds to be fused together or separated out from a complicated mixture, and so must preserve all available cues for grouping and segregation. Different aspects of timbre are dominant in different groups of sounds, so the best way to maximize separability is to scale the sounds according to the most salient underlying axes of the group. The properties of the Equally Spaced Timbre Circle do not depend on the subset of timbres in it, but on the relationships between those timbres. This has important and useful implications for learned schema mappings, providing the opportunity to substitute different Timbre Circles to signal context shifts without altering the structure contained in the interrelations between data points mapped into the space. This flexibility allows the actual timbres, and even the underlying axes of the Timbre Circle, to be changed without affecting the structure of mapping sequences.

3.1.1 Finding Some Equally Spaced Timbres

The problem in creating an Equally Spaced Timbre Circle is to find a set of psychologically equally spaced timbres. Techniques which have been used to make perceptual comparisons between sounds are (1) multidimensional scaling (MDS), and ((2) perceptual stream segregation. In this chapter the results from Grey's [19] MDS study of 16 re-synthesized musical instruments are used.

Grey found that a 3-dimensional cartesian space accounted for most of the perceived variation between the set of 16 timbres in the study. The axes which describe the dimensions of principal variation were analyzed in terms of the spectrograms of the data points, and it was found that the Y axis was related to spectral energy distribution, while the X and Z axes were related to temporal aspects of timbre, covarying with synchronicity in the development of upper harmonics and the presence of low-energy high frequency noise during the attack segment. The results were presented in a series of graphic visualizations which show where each data point lies in the 3-dimensional perceptual space. These results provide an opportunity to construct a first approximation to an Equally Spaced Timbre Circle, because they are perceptually scaled and arranged according to identified perceptually salient axes. This procedure is shown in Figure 2. A circle which encloses the projection of the data points in the temporal plane is divided into 8 segments of 45 degrees, and the position of each 45 degree increment around the circumference is nominated to represent the categorical timbre of that segment. Because distance is a measure of dissimilarity, the data point in the segment lying closest to each of the equally spaced points on the circumference is allocated to that point. There are only a limited number of data points available to choose from, so that in the 315 to 0 degree segment where there is no data, the closest point from the adjacent segment was used (i.e., TM).

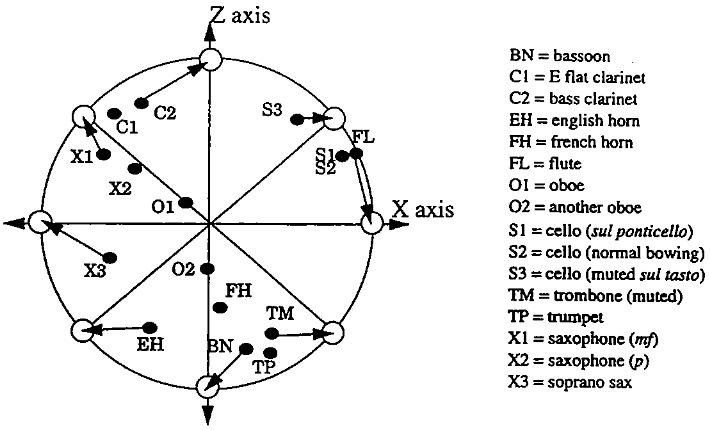

Figure 2: Equally spaced timbre circle constructed from Grey's temporal plane.

3.2 Perceptual Timbre Wheel

The Perceptual Timbre Wheel is an extension of the Equally Spaced Timbre Circle in which a radial component of timbre is used to fill in the disc.

The radial component has the following requirements:

- perceptual separability from temporally categorized timbre
- a perceptually scaled metric
- a common point of origin
- salience in perceptual stream segregation

Brightness is a perceptual dimension found consistently in a wide variety of timbre research. It has been roughly defined as the balance between the upper and lower partials of a sound spectrum [9]. The Y axis, which is orthogonal to the temporal plane in Grey's MDS study, corresponds with this definition of brightness. The natural 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 perceptual stream segregation, and it is observable in a wide range of musical, everyday and speech sounds.

In this chapter it is conjectured that 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 categorized 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. In Gaver's auditory icon [7] a wooden or metal bar of some length is hit with a mallet of variable hardness (implemented via brightness modification) without altering the category of material (implemented by temporal aspects of timbre development). These illustrations give 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 a filter cut-off frequency. At the dull end of the scale each spectrum tends toward a sinusoid at the fundamental frequency. This supports the

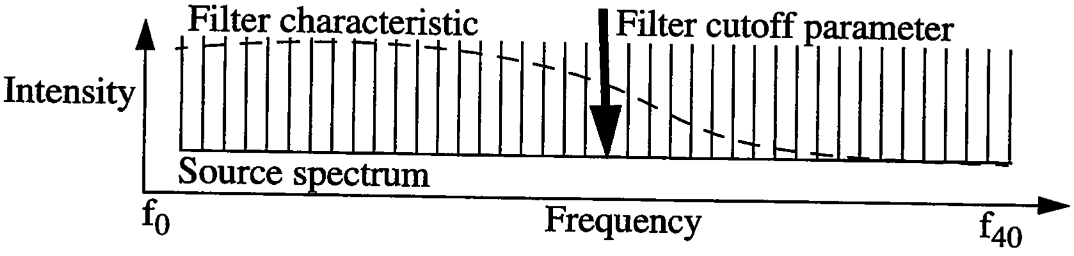

Figure 3: PSSReference instrument.

polar geometry of the timbre wheel by allowing seamless transitions across the center and also accords with Padgham's definition of the complexity radius in his timbre assessment chart.

To scale the space it is necessary to define a perceptual metric. The brightest timbre in Grey's study is the trombone (muted) (TM) with 34 harmonics and a gradually sloping spectral envelope, while the dullest is the French horn (FH) with 11 harmonics dominated by the first two. A subjective reference scale which provides an appropriate range and resolution for these data points can be created by defining maximum brightness by a spectrum consisting of equal intensity harmonics up to the 40th harmonic. The reference was implemented as a Csound [22] instrument called PSSReference. This instrument consists of a source spectrum to which a low pass filter is applied, as illustrated in Figure 3. The cutoff frequency of the filter is a control parameter. A graphical user interface (GUI) to the instrument was built which allows the filter cutoff parameter to be adjusted at each brightness increment. The results of the adjustments may be written to, or read from, an ASCII file. This tool was used to perceptually scale the PSSReference by the fractionation method to give 9 equally spaced brightnesses at a particular fundamental frequency..

3.3 The "Dull" Axis

The PSS is completed by transfixing the perceptual timbre wheel on a vertical axis from which all radial brightness scales originate. All the "dull" points in the space lie along this vertical axis, which is analogous to the "gray" axis sometimes referred to in color models.

The vertical axis has the following requirements:

- perceptual separability from timbre
- a perceptually scaled metric
- salience in perceptual stream segregation

Analyses of sound perception have traditionally begun by separating pitch and loudness from timbre. Although there is considerable interaction between these aspects of sounds, loudness and pitch are observable and separable across a wide range of timbres. Pitch is a very important factor influencing stream segregation, whereas loudness is a relatively weak factor (although co-modulations in loudness are important in simultaneous segregation). There are several subjective pitch scales available which are calibrated across a wide range of sound output devices. These considerations lead to the choice of pitch for the vertical axis in the sound space.

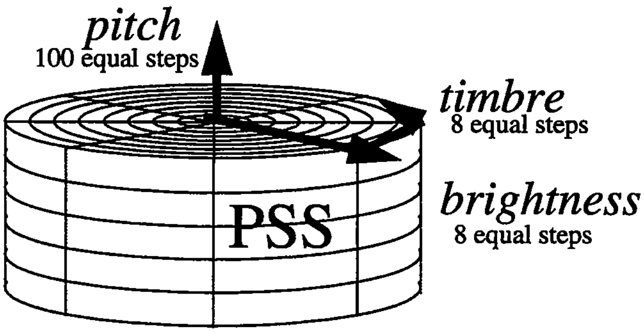

Figure 4: Perceptual sound space (PSS).

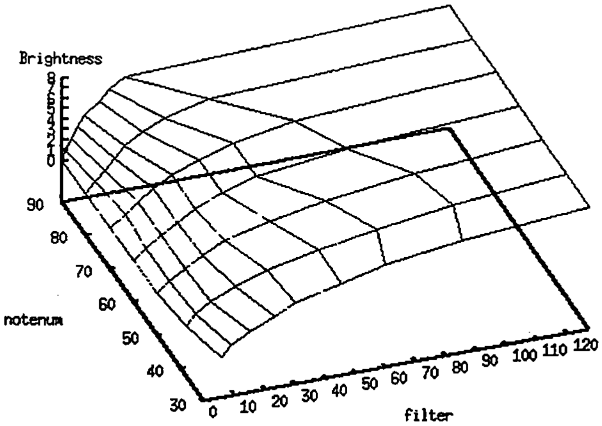

Figure 5: PSSReference(notenum, filter).

The mel scale [10] of subjective pitch was considered for the dull axis, but it is appropriate only for static sinusoidal tones. The western diatonic music scale has been developed for a wide range of musical instruments over a long history of practice. This scale was chosen for the pitch axis because it consists of more than a hundred discrete subjectively equal steps which are calibrated for many different timbres on a variety of devices. This calibration enables a pragmatic assumption that maps a semitone step offset (or *notenum* parameter) linearly to a pitch. The complete polar cylindrical PSS is shown in Figure 4.

The PSSReference instrument allows the brightness of sounds to be measured. Because brightness and pitch both depend on frequency and spectral composition it is necessary to calibrate the brightness scale at different pitches. A general PSSReference instrument was built by linearly interpolating between 7 brightness scales measured at 10 semitone intervals up the pitch axis, as shown in Figure 5.

4 Display Model

In this chapter an auditory display is considered to be any means of systematically producing sounds. This broad definition includes an orchestra, an electronic musical instrument, a row of bottles, or a computer with synthesis software and audio hardware. Each display is capable of producing some subset of all possible sounds. If the capabilities of these diverse displays overlap, then it is possible to produce the same sound in different ways on each of them. The set of sounds that each display can produce is limited by the characteristic properties of the display, and hence is a reflection of the characteristics of the display.

In order to use the PSS to specify sounds for the display it is necessary to build a display model. The display model consists of a mapping from PSS coordinates to display parameters and a description of the regions of the PSS which can be realized on the display.

An auditory display was created with the following components:

- Sun Sparcstation10 workstation with 16 bit, 44.8 kHz audio hardware
- Csound signal processing software
- samples of musical instruments described by Grey - 16 bit, 44.1 kHz, mono

The display was implemented as a Csound instrument called GreySun. The coordinate space of the display is defined by the parameters *patch* 0..7, *notenum* 0..127, and *filter* fp 0..127. These parameters covary strongly with the subjective dimensions of timbre, pitch, and brightness but use simple functions to control the physical production of a sound

Patch selects a source sample or synthesis algorithm. A one to one mapping is assumed between patches and timbres which have similar verbal descriptions. This assumption is derived from the categorical nature of timbre and the properties of the Equally Spaced Timbre Circle. If a *patch* resembles a timbre in terms of the underlying axes which arrange the Timbre Circle, then variation which doesn't alter the perceived category can be accommodated. This flexibility means that verbal descriptions of the most salient aspects of a group of sounds, such as instrument family, may be sufficient to practically match *patches* with categorical timbres while maintaining the properties of the PSS. The GreySun instrument consists of multi-pitch samples of the timbres nominated in the construction of the Perceptual Timbre Wheel from Grey's study of timbre.

The frequency of the sample playback is controlled by an interpolation technique. *Notenum* maps integers to fundamental frequencies of the diatonic scale according to the equation

$$f0(Hz) = 8.175 * 2^{notenum/12}.$$

A low pass filter is applied to the frequency adjusted sample to control the spectral centroid by attenuating the upper harmonics in the output sound. Filter fp controls the cut-off frequency fc of a low pass filter by linearly interpolating between f0 and f40

$$fc(Hz) = f0 + fp * (f40 - f0)/127.$$

4.1 PSS-to-Display Mapping

The PSS-to-display mapping is a nonlinear three dimensional transformation from perceptual sound space to the display parameter space, as illustrated in Figure 6.

Building the PSS-to-display mapping involves several steps:

- map a regular grid of sample points from the display parameter space to PSS
- invert the sample mapping to obtain a PSS-to-display mapping
- resample the PSS-to-display mapping onto a regular grid

Regularized linear spline interpolation [23] is used to create a continuous input space in the final PSS-to-display mapping.

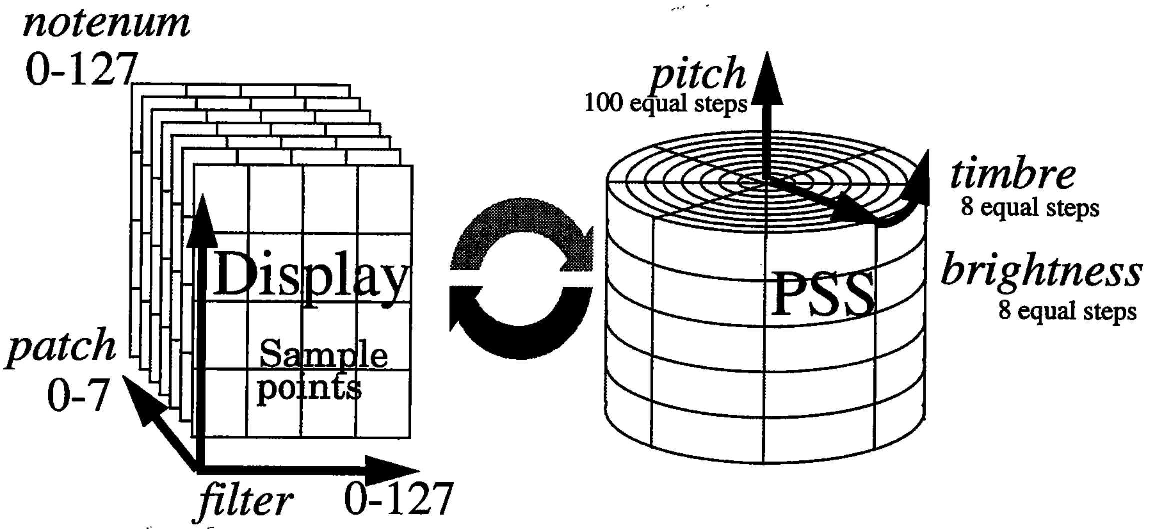

Figure 6: Display parameter space and perceptual sound space.

A display-to-PSS mapping was created by sampling the display parameter space with a grid of 7 *notenums* x 4 *filter* cutoffs for each *patch*, resulting in 7 x 4 x 8 = 224 sample points. The sound generated at each sample point is matched against the perceptually scaled reference sounds which define the PSS, implemented by the PSSReference instrument. The PSS coordinates of the best match to the display space sample point are recorded in the display-to-PSS mapping. The matching procedure is a variation of van Norden's galloping stream method [9], using the pattern XOXOX-XOXOX-XOXOX where X is the PSSReference, O is the timbre being measured, and—is silence. The brightness of the PSSReference is incremented from 0-8 over 40 repetitions of the sequence (which is faded in and out to minimize end effects). The sounds were 68 ms long and there was 75 ms between onsets. The XXX-XXX-XXX pattern is heard in one stream and OO-OO-OO in another. As X increases in brightness the two streams group together, then separate again. The length and strength of the galloping rhythm effect depends on the similarity between X and O. The initially dull XXX-XXX-XXX stream seems to start slightly ahead of the OO-OO-OO stream, but after the brightness glide passes through the fused region, and the streams separate again, the X stream seems to be delayed with respect to the O stream, and this cue helps in the matching process. The subject is directed to attempt to hear the sequence as a single stream with a galloping rhythm, and nominates the brightness value at which the rhythm effect starts and the value at which it stops. The midpoint of this region is taken as the temporal coherence boundary. This method is a partial match against pitch and brightness. The matching process may be carried out over a number of sessions and an extended period of time. The match is quasi-symmetric, and tends toward symmetric over an extended number of cycles [20].

4.2 Display Gamut

The regions of the PSS which can be realized on the display are called the display gamut. In the pitch dimension the valid range at each timbre angle will form an upper and lower boundary. The pitch range is characteristic of a particular musical instrument and the range over which samples were taken. In the brightness dimension the spectral composition of the sampled instrument causes a unique limit for each timbre.

The PSS is three dimensional so that the boundary representing a display gamut can be visualized as a geometric solid. Figure 7 shows a graphic image of the gamut of the GreySun auditory display in which the pitch and brightness limits for each timbre can be observed. This visualization allows displays to be compared in terms of their gamut shapes, and enables a geometric metaphor for prescribing mapping sequences as paths or planes inside the gamut. The gamut

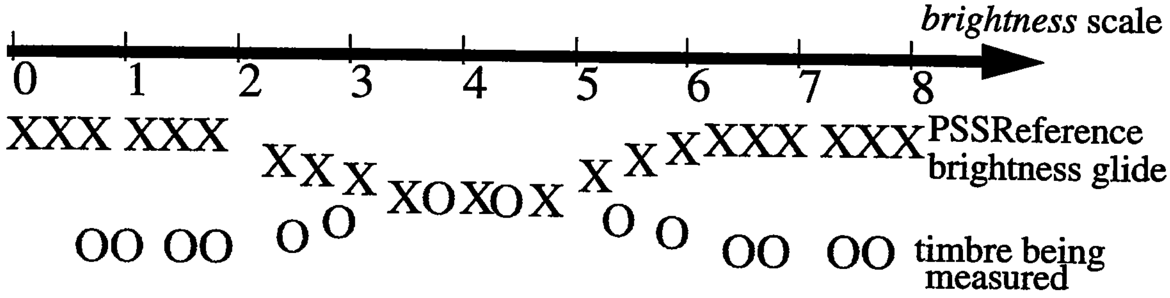

Figure 7: Brightness measurement by streaming.

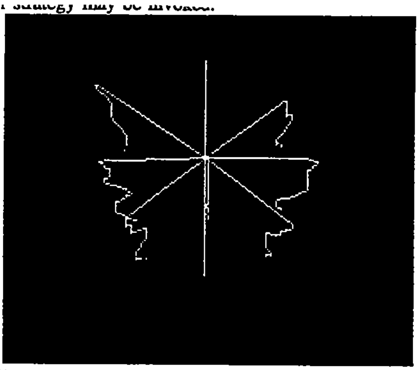

Figure 8: Graphic visualization of the GreySun display gamut in the PSS.

shape provides feedback for optimizing these sequences for a particular device or for several different devices (by choosing intersecting regions of their gamuts). The display model contains a run length encoding of the gamut which can be used to check PSS coordinates before they are mapped to the display. If an out-of-gamut coordinate is specified then it can be returned along a line to the nearest in-gamut point, or some other strategy may be invoked.

5 Summary

Some significant obstacles to the construction of auditory displays include the lack of a device-independent means of specifying sounds, interactions between parameters which obfuscate the perception of order, contrast, and ratios in mapping sequences, confusion between subjective and objective aspects of sound, idiosyncratic device characteristics and control parameters, and a lack of tools which enable nonexpert users to create effective sonification mappings.

This chapter addresses these problems through a perceptually based framework consisting of two components—a perceptual sound space which characterizes the human receiver in the system, and a display model which characterizes the physical output device or transmitter.

A perceptual sound space (PSS) was constructed by perceptually scaling the TBP sound model. The properties of the PSS are:

- Natural specification, comparison and matching—the specification of data representation in perceptual terms (such as timbre, brightness, and pitch).

- Intuitive order—the axes are observably coherent and ordered. Brightness and pitch have a natural unidimensional order. Timbre is arranged in a circle so that the most similar timbres are adjacent and the most dissimilar are diametrically opposite.
- Perceptual metric—euclidean distance is a measure of subjectively perceived difference.
- Transportable timbre model—the space is built on a model of timbre which allows substitution of timbres while maintaining the properties of the space.
- Dynamic sequences—timbre, brightness, and pitch are important in sequential stream integration, and the PSS may be used to visualize and control grouping in the specification of coherent dynamic sequences.

A display model which maps points in the perceptual space to the display parameter space was constructed. This model allows the display to be controlled in terms of the PSS, and includes a boundary representing the gamut of sounds which can be produced by the display device. The properties of the display model are:

- Device-independent interface—the coordinate system allows sounds to be specified in natural terms without reference to display device parameters. A point in this space will sound the same no matter what device is used to produce it. The interrelationships between data points in the perceptual space are preserved by the mapping to the display. The display model handles the mapping to the display space, and different models may be substituted depending on the display.
- Device characterization—the properties of a display are represented and can be used to compare displays, understand the capabilities of a display, and return out-of-gamut points to the display.
- Geometric interaction—the spatial representation of the gamut enables a geometric metaphor for interaction so that paths and planes may be used for specifying sounds and sequences on the display.

6 Limitations

The scaling each of the axes 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 parallel to the axes—caution should be exercized in mapping more complex sequences.

There is significant change in perceived loudness with changes in pitch and brightness. This could be addressed with a loudness model.

It is assumed that all sounds are of constant duration, and that dynamic sequences are presented at a constant rate. The PSS 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 PSS 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 PSS (perhaps in the form of Shephard's pitch helix) may help here. Other interactions which effect the perception of simultaneous sounds include spatial location, onset asynchrony, and shared modulations.

The psychoacoustic data used to construct the GreySun display has come from only one experimental subject—the author.

7 Further work

Meaningful sonification mappings require consideration of the connection between the data and the representation; for example a nominal variable would best be mapped to the categorical perception of timbre, while an interval variable would best be mapped to the ordered perception of brightness or pitch. A user interface which embodies meaningful mapping schemes for 1d, 2d, and 3d data of different types, and employs a geometric metaphor for specifying these mappings, is being built to provide a tool to enable nonexperts to construct successful sonification mappings. These mappings will be tested with data from a variety of application scenarios.

Acknowledgments

This work was funded by a postgraduate research scholarship from the CSIRO, Division of Information Technology, Australia. I would like to thank Dr. Phil Robertson of the CSIRO for his support, guidance, and expertise. I would like to thank Mr. David Worrall of the Australian Centre for Arts and Technology for his perspective and advice. Many exploratory experiments were carried out using equipment supplied by Advanced Gravis.

References

[1] Smith, S. "Representing Data with Sound." Proceedings of the Visualization 1990 Conference, IEEE Computer Society Press, 1990.

[2] Kramer G., ed. "Some Organizing Principles for Representing Data With Sound." In *Auditory Display: Sonification, Audification, and Auditory Interfaces*, edited by G. Kramer, 185–222. Santa Fe Institute Studies in the Sciences of Complexity, Proc. Vol. XVIII. Reading, MA: Addison-Wesley, 1994.

[3] Kaplan B., and J. MacCuish. "User Documentation." Sonify module for AVS, Center for Innovative Computer Applications (CICA), Indiana University, 1993.

[4] Scaletti C. "Sound Synthesis Algorithms for Auditory Data Representations." In *Auditory Display: Sonification, Audification, and Auditory Interfaces*, edited by G. Kramer, 223–252. Santa Fe Institute Studies in the Sciences of Complexity, Proc. Vol. XVIII. Reading, MA: Addison-Wesley, 1994.

[5] Bertin J. *Graphics and Graphic Information Processing.* Berlin: Walter de Gruyter, 1981.

[6] Ware C. "The Foundations of Experimental Semiotics: a Theory of Sensory and Conventional Representation." *J. Visual Lang. & Comp.* **4** (1993): 91–100.

[7] Gaver W. "Auditory Icons." *Human-Comp. Interaction* **2(2)** London: Lawrence Erlbaum Associates 1986.

[8] Gibson J. G. *The Senses Considered as Perceptual Systems.* Boston: Houghton Mifflin, 1966.

[9] Bregman, A. S. *Auditory Scene Analysis: The Perceptual Organization of Sound.* Cambridge: The MIT Press, 1990.

[10] Zwicker, E., and H. Fastl. *Psychoacoustics Facts and Models.* Berlin: Springer-Verlag, 1990.

[11] Robertson P.K. "Visualizing Color Gamuts: A User Interface for the Effective Use of Perceptual Color Spaces in Data Displays." *IEEE Computer Graphics and Applications* **8(5)** (1988): 50–64.

[12] Pridmore R.W. "Music and Color: Relations in the Psychophysical Perspective." *COLOR Research and Applications* **17(1)** (1992): 57–61.

[13] Sebba R. "Structural Correspondence Between Music and Color." *COLOR Research and Applications* **16(2)** (1991): 81–88.

[14] Caivano J.L. "Color and Sound: Physical and Psychophysical Relations." *COLOR Research and Applications* **19(2)** (1994): 126–132.

[15] Padgham C. "The Scaling of the Timbre of the Pipe Organ." *Acustica* **60** (1986): 189–204.

[16] Barrass S. "A Naturally Ordered Geometric Model of Sound Inspired by Colour Theory." Proceedings of Synaesthetica '94, Canberra: Australian Centre for the Arts and Technology, 1994.

[17] von Bismarck G. "Timbre of Steady State Sounds: A Factorial Investigation of its Verbal Attributes." *Acustica* **30** (1974): 146–159.

[18] Slawson A. W. "Vowel Quality and Musical Timbre as Functions of Spectrum Envelope and Fundamental Frequency." *J. Acoust. Soc. Am.* **43** (1968): 87–101.

[19] Grey J. M. "Exploration of Musical Timbre." Ph.D. Thesis, CCRMA Dept. of Music, Stanford University, Report No. STAN-M-2, 1975.

[20] Wyszecki G., and W. S. Stiles. *Color Science.* John Wiley and Sons, 1967.

[21] Handel S. *Listening: An Introduction to the Perception of Acoustic Events.* London: The MIT Press 1990.

[22] Vercoe B. CSOUND, A Manual for the Audio Processing System and Supporting Programs, Cambridge: MIT Media Laboratory, 1991.

[23] Bone, D. "Adaptive Color-printer Modelling Using Regularized Linear Splines." *Proceedings of Device Independent Color Imaging*, IS&T/SPIE Symposium on Electronic Imaging Science and Technology 93.