Chapter 3:

AURALIZATION: Representing Data as Sound

Introduction

One of the earliest uses of non-speech audio in the interface was the representation of complex data. Consider the many parameters associated with stock market data or the variables related to seismic activity. Even with sophisticated analysis, exploring such data to find meaningful patterns is still often a problem requiring human intervention. Computer graphics has revolutionized the ability to obtain information from computers, but graphics typically offers a 2-dimensional projection of what is often many-dimensional data. For example, consider the following set of values which represent three different species of iris flowers [ref Fisher]. For each flower, there is a sepal length and width and a petal length and width. In fact, these four variables are generally sufficient to classify each plant into the appropriate class of species: *Iris setosa*, *Iris versicolor*, and *Iris virginica*. However, this classification is difficult if one only looks at the numerical values. Now, notice the graph in Figure 3.1 (and listen to sound example 3.1). It is easy to identify examples 1, 3, 4, and 7 as

one set (*Iris setosa*), examples 2 and 5 as a second set (*Iris versicolor*), and examples 6, 8, and 9 as a third set (*Iris virginica*).

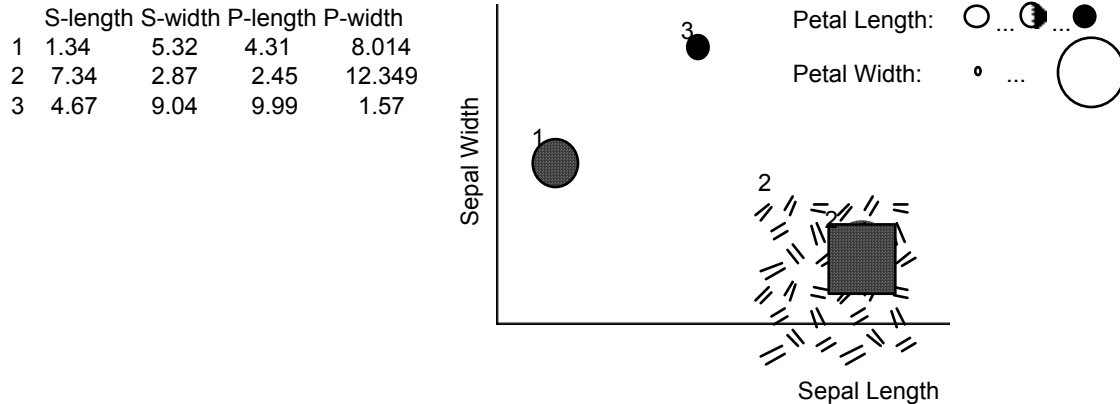|	S-length	S-width	P-length	P-width
1	1.34	5.32	4.31	8.014
2	7.34	2.87	2.45	12.349
3	4.67	9.04	9.99	1.57

Figure 3.1: It is difficult to distinguish the three sets of data by looking only at the numeric data. In both graphical and aural representations, the three sets are easily differentiated.

Representing complex data in sounds might be considered the acoustic equivalent of scientific visualization; we call it *scientific auralization.* Typically, representing scientific data as non-speech audio involves mapping the parameters of the data to the parameters of sound. In the example above, the sepal length was mapped to pitch, the sepal width to volume, the petal length to duration, and the petal width to the fundamental waveshape (a variation between a pure sine and a square waveform). The acoustic properties of each sound are thus shaped by the values of the parameters of the underlying data.

History

Although the field of study of auditory interfaces is relatively young and undeveloped, the notion of representing data values in sounds has arisen frequently over the years. One of the first examples comes from the arts in the form of a musical composition. In 1970, the American composer Charles Dodge wrote a piece of computer music entitled "The Earth's Magnetic Field" (Dodge 70). In this composition, the sounds correspond to the magnetic activity for the Earth in the year 1961. Dodge and colleagues used the Kp index for magnetic activity. A new value for the index is calculated every three hours and may take any of 28 distinct values. The musical interpretation is a correlation between the succession of indices and pitches over four octaves.

A scientific version of representing data in sound was presented by S. D. Speeth in 1961 (Speeth 61). Speeth was particularly interested in distinguishing underground explosions from natural earthquakes. Seismograms were digitized at a sampling rate of ten samples/sec and then time compressed to 1000, 2000, 4000, and 8000 samples/sec. These samples were read through a digital-to-analog convertor and written on magnetic tape for audio playback. Listeners were quite successful in separating the explosions from the earthquakes.

In the early 1970s, Max Mathews and colleagues (Chambers 74) used both graphics and sound to present up to five dimensions of data. Two variables provided a visual x-y scatter plot, while frequency, timbre, and amplitude modulation provided a corresponding note of three dimensions for each point. A user interactively selected a sequence of points to hear. Thus, the user sees the entire scatter plot while hearing a sequence of sounds. Using examples of actual scientific data (including the Fisher iris data), Mathews et al demonstrated that auditory representations reveal structure in the data.

Mathews proposed a number of aural dimensions: loudness, pitch, vibrato, rate of modulation, aspects of timbre, and tempo. While these parameters are suited for discrete or

continuous variables, Mathews also was interested in the use of chords, stereo, and localization for discrete events. He suggested that eye tracking devices might be useful in integrating the visual and aural representations of sounds so that what one hears is what one sees. Mathews identified several problems for data auralization: The range of most of the sound dimensions is relatively small, thereby limiting the range of the data values represented. Sounds provide a gestalt of the data; other methods of analysis are necessary for quantitative results. Graphics has the advantage of offering a simultaneous view of all the data while sound is inherently a stream of events.

In 1980 Yeung, a chemist, also recognized that sound could be used in addressing human pattern recognition in data analysis (Yeung 80). He was concerned that sound dimensions be based on quantitative standards, continuity in scaling, independence, resolution, and relative ease in perception. The nine parameters he chose were pitch (two dimensions, one in the 100-1000 Hz range and the other in the 1000-10000 Hz range), loudness (in increments of 3 dB), amplitude damping, direction (as determined by three variable values to locate a position in a cube thus requiring six loudspeakers, at the top, bottom, front, back, left, and right of the listener), duration/repetition, and rest (the time between repeats of the entire sequence). Yeung felt that it would be possible to use at least these nine dimensions in sound and perhaps as many as twenty.

Yeung took a data set of elemental concentrations of ten metals in 63 samples from four sites and mapped seven of the ten variables to his primary sound parameters (neglecting the 3-d location information). Four analysts heard 40 samples from the four data sets and were subsequently able to achieve 90% to 100% correct classification after one or two training sessions.

In all cases of early work of representing data in sounds, scientists were concerned with the human ability to find patterns in a complex set of data. They were interested in making use of the fact that we easily distinguish a variety of sounds and that this *gestalt* listening could be applied to data exploration. Furthermore Speeth, Mathews et al, and Yeung were concerned about verifying their results and ran experiments to show that the sound did yield appropriate and useful information relative to the data being represented. Their methods involved mapping the values of the data itself to various parameters of sounds. They raised issues of the perceptual abilities of the human ear, the need for a variety of data representation methods, and the necessity for user interaction with the data presentation.

Case Studies

Studying the representation of data in sounds explores not only the accessibility of complex xdata but also the forms of audio representations. Four pieces of work will illustrate several characteristics of the research. One of the early studies was the dissertation work of Bly who undertook to verify that sounds could yield useful information about the underlying data. About the same time, Mezrich, Frysinger, and Slivjanovski focused on time-varying data and the potential for integrating sound and graphics to promote pattern recognition. Grinstein, Smith, and colleagues have created a workstation environment for exploratory data analysis using scientific visualization and auralization. Lunney and Morrison concentrated on applying the methods for the visually handicapped.

Case Study 1: Presenting Information in Sound

Bly was interested in the problems of portraying multi-dimensional data for exploratory data analysis. Although a variety of graphical methods exist for data representation (Everitt 78, Tukey 70), the analyst is nevertheless limited to the number of dimensions that can be perceived visually and to the constraints of a visual display such as focused visual attention. Bly's work made two contributions to the study of auralization: one is that she tried a

variety of data types and mappings and two is that she conducted an experiment to show that the sound representation did, in fact, yield useful information for exploratory analysis.

Bly attacked three types of data: discrete multi-dimensional samples, logarithmic data, and time-varying data. With sets of discrete data samples, the particular problem Bly addressed was that of discriminant analysis. Given known samples from each of the three sets, can an analyst correctly classify an unknown sample? Bly represented each data sample as a single note of up to six characteristics: pitch, volume, duration, waveshape, attack, and harmonic. The pitch was selected from the 128 musical notes ranging over four octaves. The volume was dependent on the equipment available and ranged across twelve levels ("very soft" to "very loud"). The duration of a single note varied in 5 msec steps from 50 msec to 1050 msec. For timbre, a sine waveshape of 128 values between -1.0 and 1.0 was the base parameter. For each of the original 128 values, one was randomly selected and modified until a completely random waveshape resulted. This provided 128 waveshape variations. The attack envelope was varied in fifteen increments between a gradual attack (slope of 45 degrees) and a constant attack. Finally, an odd harmonic (the fifth) was chosen and varied in waveshape as the fundamental. Bly recognized but ignored the problems of psychoacoustics in the hopes of gaining some basic results with respect to the usefulness of a straightforward application of sound to data.

She first applied her technique to the Fisher iris data as described in the introduction to this chapter. Representing the data aurally, Bly found the three species of iris flowers to group as accurately as with most graphical methods (nearly all listeners could classify all but one or two samples correctly). Similarly, spectra data was encoded into sounds. A few individuals were given twenty known samples from each of four sets and then asked to classify twenty unknown samples. The results were quite successful and compared favorably to a similar classification test using Chernoff's FACES as a graphical representation (Chernoff 73).

In representing logarithmic data, Bly was motivated by the logarithmic relationship between frequency and pitch and by the frequent difficulty in reading logarithmic data more generally. LIke Speeth, Bly took advantage of seismic records of earthquakes. However, she was interested in finding patterns within the earthquake data itself and in representing logarithmic data generally. Given the longitude, latitude, depth, magnitude, and start time for a series of quakes, Bly encoded the magnitude (the exponential variable) in pure frequency while displaying the location. A magnitude of 0.0 mapped to 5120 Hz; 8.0 mapped to 20.8 Hz. Depth was not used but patterns and extremely large events were easily observable suggesting that sound could be used to highlight features that are most relevant to seismologists.

More interesting, perhaps, was an attempt to play logarithmic plots musically. Frequently a two-dimensional logarithmic plot could not be easily distinguished from another but audio representations of the same data were distinguishable. The plots consisted of scatter points and the corresponding best fit line of the form

$$y = Ae^{Bx}$$

By varying the values of A and B, different plots were obtained which appeared visually quite similar. A chirp designated a sound plot in which the x-value was encoded as time and the y-value provided the frequency. A warbel designated a sound plot in which the slope (y/x) prvided the frequency base (thus a linear plot has a constant warble). Listening to the plots made them much more distinct than the graphs alone.

For time-varying data, Bly took information from battlefield simulations. These simulations were run with different starting parameters and then analysed based on the results of the battle. However, the length of the battle and the amount of data collected over time made it difficult for the analyst to gain any appreciation for the differences that occurred in the process of the battle itself. Therefore, Bly created "battle songs" for the simulations. The intuition was that similar battles would create similar songs so that an analyst could quickly find those battles that differed substantially from the norm. An analyst could subsequently attend to more careful examination of the distinguished battles.

To validate the notion of auralization for exploratory data analysis, Bly ran a series of experiments. The data sets were discrete data samples of six dimensions generated in well-

defined ways. Subjects listened to a few sounds from each of two sets and then decide for each subsequent sound whether it belonged in Set 1 or in Set 2. The training samples and test samples were chosen randomly but did not overlap. The study consisted of three phases: one, the data was translated, scaled, and correlated to ensure that users appropriately recognized the differences in data sets. Two, the subjects performed a discriminant analysis task on 6-dimensional data sets. Three, different mappings and training further substantiated the results.

In Phase 1, a set of data was transformed to create a second, related set of data. In this way, the expectation was that as the data sets were further apart (a greater transformation), participants could more easily determine into which set a sound sample belonged. The second set of data was created from the first set by translating, scaling, or correlating the data values. As expected, subjects were able to identify an increasing number of samples as the translation factor increased and the data sets became more widely separated. Similarly, as the scaling factor increased, the ability to discriminate between two sets increased. However, for some value of increased scaling, the ability to discriminate became more difficult (although the sets were even further "apart"). Overall the scaled data was more difficult to distinguish but notice that Set 1 essentially lies within Set 2. In the final experiment of Phase 1, the data in Set 1 was strongly correlated and that in Set 2 was a random mix of samples with no correlation. Subjects did not perform particularly well on the correlation differences. Note that in none of the cases were subjects given information about how the data sets differed.

Despite some mixed results, Phase 1 did indicate that participants obtained useful information about data from aural represention. The goal of Phase 2 was to determine if sound could add information to a graphical method of data presentation. In this case, the two sets of data were generated so that they were only distinct in 6-space. That is, data was obtained from a normal random deviate generator and then separated into two sets such that a sample belonged to Set 2 if and only if

$$x_2^2 + x_3^2 + x_4^2 + x_5^2 + x_6^2 <= 1.5^2 \text{ OR}$$

$$x_1^2 + x_3^2 + x_4^2 + x_5^2 + x_6^2 <= 1.5^2 \text{ OR}$$

$$x_1^2 + x_2^2 + x_4^2 + x_5^2 + x_6^2 <= 1.5^2$$

Thus, at least five of the six variables in each sample of Set 2 had value less than 1.5 and at most one of the variables x_1, x_2, or x_3 could have value greater than 1.5. In a complete positive and negative space, Set 2 lies somewhat within Set 1 and a projection plot onto any two axes has overlap.

Each six-dimensional sample was represented by a note consisting of pitch, volume, duration, waveshape, attack, and harmonic as described earlier. For the experiment, the mapping was

    x1 -> waveshape,
    x2 -> overtone waveshape,
    x3 -> pitch,
    x4 -> envelope,
    x5 -> duration, and
    x6 -> volume.

The sounds in Set 2 were generally shorter and softer than Set 1 sounds because variables x4, x5, and x6 were always each less than 1.5. If a Set 2 note were buzzy (random waveshape), then it was low in pitch since both x1 and x3 couldn't both be large. Similarly, if it were high in pitch, then it was a more pure tone. For the visual display, two-dimensional plots were generated of the data. A particular sample was indicated by highlighting that point.

Seventy-five subjects participated in Phase 2, 25 in each of three groups. The first group had only aural representation of the data, the second group had only visual representation of

the data, and the third group had both aural and visual representation. With sound only, the first group correctly identified 64.5% of the samples, the second group with graphics only identified 62% correctly, and the third group using both sound and graphics identified 69% of the samples. All the results verified significantly better discrimination between sets than if the subjects were responding by chance. In particular, the identification with both sound and graphics was significantly better than with either sound or graphics alone.

In Phase 3, the same data was used as in Phase 2, however the mapping was changed. From discriminant analysis calculations on the data, it was determined that the importance of the contribtuion of each variable could be ranked as x2, x1, x3, x6, x4, and x5. Subjectively listening to the sound variables suggested a ranking of pitch, duration, waveshape, volume, envelope, and overtone. Theses rankings determined a new mapping (x2:pitch, x1:duration, x3:waveshape, x6:volume, x4:envelope, and x5:overtone). The results did not indicate a significant difference with this change in mapping.

Another aspect of Phase 3 was to provide additional training. A set of subjects who had participated in Phase 2 again performed the discriminant analysis task. This time, they were given examples and explanations of the data differences in Phase 1 to acquaint them with ways sounds might differ based on data values. They could listen to each of the six dimensions varying alone, they were allowed to listen to the training samples at any time throughout the experiment, and they could listen to the training samples in any order. The average number of correct responses was 73.8% compared to 64.5% originally for sound-only representation. Unfortunately there was an experimental flaw and training samples were not deleted from the larger set of test items. This potential repetition of samples could contribute to the better performance.

Bly's work confirmed her hypothesis that sound offers useful information for multivariate data presentation. In addition, she offered a number of early explorations of various data types and possible auralizations. The weaknesses of the work indicate a number of issues for further study. The interdependencies among the sound parameters (the psychoacoustics) cannot be ignored. The question of which sound parameters to use and how to map the data to these parameters is very much open. The integration of more sophisticated graphical displays deserves attention. Furthermore, this work gives very little indication of the effects of learning and what types of data are best suited for sound encoding.

Sound Example 3.3: Illustrate Bly's work with experimental 6-d data.

Case Study 2: Dynamic Representation of Multivariate Time Series Data

While Bly's work demonstrated that multi-dimensional data samples can be usefully represented in sound, the work of Mezrich, Frysinger, and Slivjanovski took a systematic approach to the design of animated graphics with audio representation (ref). They were interested in finding patterns in time-varying data. Using economic indicators, they mapped the data to musical notes so that each indicator has its own melody over time. The resulting data presentation takes advantage of the listener's ability to pick out similar sequences of notes to locate trends in the data.

Typically, time-series data is depicted graphically as a 2-dimensional plot, the horizontal dimension being time. Multiple variables are overlaid, stacked, or even displayed on separate axes so that seeing correlations in the data is difficult. Even more difficult is the problem of finding patterns in the data as a whole. The usual statistical methods often just reduce the multiple variables to a single composite index. Mezrich et al hoped to use audio and graphical animation to provide insight into dynamic information while preserving all the information available. Their goal was to facilitate global pattern recognition, to allow an

analyst to find relationships among variables. They did not intend to portray precise quantitative data or to replace traditional graphs as one means of data representation.

Mezrich et al explored their ideas using seven economic indicators descriptive of the cyclic state of the economy. For each of the seven indicators or variables, they assigned both a note and a vertical line pair. At any point in time, the pitch of the note was determined by the value of the economic indicator at that time; all other characteristics of the note, such as loudness, were held constant. Similarly, the height and position of the vertical line was also dependent on the value of the variable. For small values, the note had a low pitch and the corresponding vertical line pair was short and near the center of the graphical display. For large values, the note had a high pitch and the corresponding vertical line pair was tall and positioned at the edges of the display (see Figure something). Each of the seven economic indicators was assigned a different color for the respective line pair. The seven corresponding notes did not differ except in pitch as determined by the values of economic indicators. Figure 3.3 illustrates the mapping from economic data to notes.
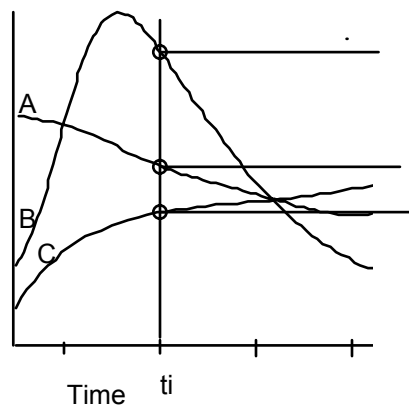
Figure 3.3: At a time, ti, the value for each economic indicator determines the length of the vertical line pair and the pitch of the corresponding note.

An analyst could play a single variable over a period of time, listening to the notes rise and fall while watching the pair of lines come forward and retreat. A combination of variables provided an animation of line pairs and a tune of the corresponding pitches or voices. A combination of variables could be explored for correlations among the indicators; time slices could be explored for patterns of behavior.

There are few, if any, experimental methods for measuring the amount of insight a data representation provides. In order to test the usefulness of their approach, Mezrich et al set up a series of experiments to determine a user's ability to discern correlations among variables. The experiment was designed to measure how easy it is to perceive a relationship among a set of variables, i.e. how easily a correlation among a set of variables can be perceived. Using four variables or dimensions, the experiment compared three graphical techniques (overlaid, stacked, and separated graphs) and the dynamic technique. The experiment measured the threshold correlation for each of these four techniques. Six subjects were tested on all experimental conditions with the order of representation balanced across subjects. Subjects recognized correlations more readily with the dynamic presentation than with most graphical representations. The partial exception, overlaid graphs, appears to be related to series length, suggesting that temporal effects or data magnitude are important.

In a subsequent study (Frysinger 88), Frysinger explored the issue of correlations in overlaid graphs as well as other issues raised by Mezrich et al. He conducted an investigation of auditory/visual representations of multivariate time-series data to determine the sensitivity of the previous results of Mezrich et al to time-series dimensionality and

signal dectection tasks as well as to determine the degree to which an auditory display was enhanced by a simultaneous visual display. Two forced-choice experiments were conducted in which subjects were to determine which of two data sets was correlated (Task 1) or contained a perturbed version of a previously trained data set (Task 2).

A modified up/down procedure was used to estimate the psychophysical threshold of signal recongition for each of three display types: overlaid visual line graphs, a dynamic auditory/visual representation, and an auditory-oly representation. This threshold was taken as a measure of display effectiveness. Subjects' data interpretation performance was found to depnd upon the detection task. For correlation detection, time-series dimensionality was a significant variable in display performance, and the combined auditory/visual display proved superior to the auditory-only display. For trained-pattern detection, dimensionality was apparently not a factor, and the performance of the auditory/visual display was esentially the same as the auditory-only display. These results illustrate the futility of attempting to characterize the properties of auditory data representation independent of the data analysis task.

The work of Mezrich, Frysinger, and Slivjanovski was important both in terms of the presentation offered, an integration of animated graphics and sound for time-varying data, and in the experimental data to test their approach. They also discovered informally that mapping the frequencies to the chromatic scale was more effective than using pure frequencies. Furthermore, they offered a system that was user-tailorable. The real-time interactive capabilities provided an important opportunity to highlight, or tease out, patterns in the the data that might convey information.

Sound Example 3.2: Illustrate Mezrich's work with a pattern and then the pattern repeated but embedded in a larger context.

Case Study 3: Stereophonic and Surface Sound Generation

A third type of data, to be distinguished from discrete samples and time-varying data, is that of a space of *coherent,* or inherently visual, data such as a surface or solid. Although the data might be sampled at discrete points in the space, those data points have a meaningful relationship as one moves among them in the space. Such data occurs, for example, in brain scans, solid volume measurements, and geographic tracking data. Smith, Grinstein, Williams, Bergeron, Pickett, and Pecelli [refs] have been tackling the problem of designing a workstation for scientific perceptualization, providing specifically for inherently coherent or visual databases. One particular domain of data has been magnetic resonance imaging scans of a cross-section of a human brain. The typical resolution of these scans is 256x256, with 8 bits of information per pixel. Note that each of the 256x256 pixels can be considered as a data sample of 8 dimensions. However, unlike the discrete samples in Bly's domain, these samples are related in the x,y space. Thus one is interested not only in examining the individual data points but also in examining the relationships among neighboring points. It is important to note, however, that their techniques may certainly be used for inherently non-visual databases (such as Bly's samples) in which the user specifies the x and y axis parameters.

As with other work on scientific data representation using nonspeech audio, Smith et al are interested in capitalizing on multiple domains of human perceptual capabilities and in finding new paradigms for exploratory data analysis. Their work is further motivated by the desire to design and build a workstation environment that allows exploration of multivariate data interactively and with a multitude of data sets. A third goal is to provide support for experimentation in the use of iconographic displays. The Exvis (Exploratory Visualization) project underlying their work is aimed at providing a visualization *(we might say perceptualization?)* workstation that supports these goals.

Exvis data is represented both graphically and aurally as *icons*. Each data point (or pixel) has an (x,y) position on a graph. At each point, the variables at that point are displayed as a graphical *icon* of an unlimited number of dimensions, typically at most 20. The graphical icons are stick-figure icons consisting of a body and some number of limbs (typically four) arranged in some topological configuration. (Limbs are restricted to being attached only to the ends of the body or to the ends of other limbs and only two attachments can be made at any point.) Each limb has length, width, angle, and color. Thus, a five-limb icon can represent 20 data dimensions. In addition, each icon has an associated auditory icon/note consisting of 5 dimensions or sound characteristics: pitch, loudness, waveshape (as determined by the depth of frequency modulation), rate of attack, and rate of decay. Note that the mapping of the data values to the graphical and auditory dimensions may be independent or redundant. The premise is that "interesting gradients and contours in the auditory or visual texture of an iconographic dispay reflect structures in the underlying data." Also, an analyst may chose to use only the visual representation, only the aural representation, or both.

The auralization is entirely interactive, in that the auditory data presentation is triggered by the mouse position within the graphical display. For example, if one selects the point illustrated in Figure x, one might hear the note in sound example y. In addition to listening to the samples as discrete notes, an analyst can "play" a sequence by sweeping along a line in the space. Depending on the rate of attack and decay of the various notes, the auditory sound becomes a texture of sounds just as the graphical display is a texture of icons. The sound texture varies depending on the speed and direction in which the user moves the cursor. Up to eight *voices* are used in sequence so that each sound icon or note is unique. The repretoire of sounds include but are not limited to musical tones, bells, and even apparently random noise. Although changeable, the vibrato, tremolo, and ratio of carrier frequency to modulation frequency are held constant for a given display or sound representation. On-going work includes a focus on using left/right stereo and distance (created by reverberation)to offer clusters of auditory icons. Within a cluster or region, the sounds may be heard simultaneously with or independently of other sounds in that region. The ordering of the presentation of the sounds in the region is a user-defined ordering such as left-to-right/top-to-bottom. Another scheme for activiating sounds of several icons as a cluster is to use sound "paintbrushes" which specify an area of the display.

A pilot study was run to test the hypothesis that sound attributes, in addition to visual attributes, would improve the performance of subjects in classifying regional patches. The data set was four-dimensional with two regions, A and B, specified for the experiment. Regions A and B had a single point of intersection. Test subjects were shown representative samples from A and B and were then shown a display consisting of 16 different patches of icons. The task was to determine for each of these patches whether it belonged to Region A or to Region B. Each patch was created from a data set clustered around a point selected at random from Region A or Region B. A patch from Region A and a patch from Region B might have non-empty intersection. Two dimensions of the data were mapped to the visual icon (one-limb icons with variable length and variable angle) and two to auditory attributes. Eight subjects performed the task twice, once using only the visual texture of the patches and once using both the auditory and visual attributes. For each task, the screen was created anew. The results are shown in Figure x, indicating that test subjects did perform better with the addition of sound.

As with the work of Bly and Mezrich et al, Smith et al used the chromatic scale for determining sound pitch.

> *Sound Example 3.3: I suggest some various notes highlighted on a graph visually and a sequence/sweep, also shown on the graph. Stu says he has a suggestion about some examples and that he will generate whatever we want on high quality audio tape.*

Case Study 4: Auditory Presentation of Experimental Data

Another major area of early work is the use of auralization for those who are visually impaired. Lunney and Morrison have attacked problems similar to those of Bly, Mezrich et al, and Smith et al, i.e. multi-dimensional data representation for exploratory analysis, but with the need to represent the entire work in the aural domain. Lunney and Morrison are interested in the presentation of infrared spectra, both because it is complex data (multi-dimensional) and because it is an important tool in the identification of organic compounds. Their goal is to provide improved access to data for visually impaired scientists and students. Their techniques are intended to allow the identification and classification of samples; the recognition of similarities and differences in data patterns.

To obtain infrared spectra, scientists use spectrophotometry to measure the absorption of a light by a given sample. For each wavelength, the spectrophotometer detects the intensity of the light passing through the sample. Typically a graph is created in which the intensity of the transmitted light is plotted as a function of the wavelength. The method is particularly important in the identification of organic (carbon-containing) compounds because almost every organic compound has a unique spectrum.

To translate a visual spectrum to one in the aural domain, the continuous infrared spectrum is considered as a set of discrete events. That is, each absorption peak in the spectrum is replaced by a vertical line having location (frequency) and height (intensity). Thus a single spectrum consists of some number of ordered pairs of values. The frequency of each peak (the location) is mapped to pitch and the intensity to time durations. Figure x shows an example of a spectrum graph, the lines indicating the absorption peaks, and a mapping from the stick representation to a set of notes.
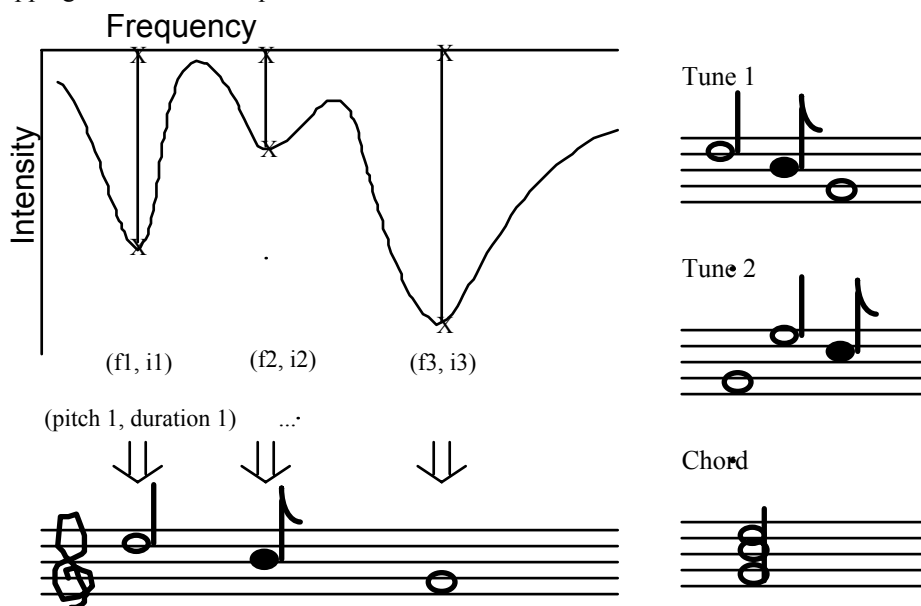

Figure 3.5: Absorption peaks are represented as pitch and duration pairs.

Lunney and Morrison present the resulting set of notes in three different ways. One, the spectrum is represented by a "tune" consisting of the notes played in order of descreasing frequency. Since peak frequencies are mapped to pitch, the "tune" is monotonically decreasing (not right!). Two, the notes are played in order of descreasing intensities. In this case, all notes in this sequence are played with the same duration so that only the order in the "tune" indicates the relative value of the intensity of the peak. This representation is most like a "spectral melody". Finally, the six peaks with the greatest intensities are played together as a chord.

To present a chemical infrared spectrum, the first pattern (or tune) is played twice, then the second is played three times, and finally the chord is played. In general, the second

pattern is heard as a somewhat syncopated piece with multiple parts. Because the pitch sequence is unconnected, the perception is often of one part taking a rest while another part enters. The chords are most often dissonant. Figure x shows 3 different chemical compounds and their corresponding spectra. You may listen to each of these in Sound Example 3.4.

Sound Example 3.4: Listen to 3 or 4 different spectrum.

FIGURE: Show the corresponding graphical spectra with stick peaks drawn.

Lunney and Morrison have run informal tests in which there are very few failures in matching identical patterns. The chords themselves allow rapid screening and matching; a scientist can easily listen to known compounds until the match is found for the unknown. Also, the perceived syncopation in the second part of the sound sequence adds an element of rhythmic interest, making the recognition task somewhat less dependent on pitch discrimination. The method can be applied to any continuous (x,y) function such as chromatograms or nuclear magnetic resonance spectra.

This work is interesting not only because it is motivated by eliminating the visual representation totally but also because it uses a combination of "melodies" and chords to present the data. Rather than offering graphics and sounds, they offer different perceptions within the aural representation. Also, the data analysis problem differs somewhat in that the scientist is not looking primarily for patterns in the data but rather for recognition of known data. Thus, memory plays a particularly important role in this work. Lunney and Morrison suggest that the sequences of notes are more memorable than single chords. Yet, for quick comparisons, the chords provide a useful tool.

Lunney et al also use speech output, sometimes in combination with simple tone-varying representations, to give data information to visually impaired scientists. They have developed their own hierarchy of auditory variables (ref the SPIE paper): pitch, duration, attack, waveform, loudness, and decay rate. However, they note that attack and waveform interact so strongly that it is not advisable to use them to represent independent variables. Also, loudness is affected by frequency so that it is most useful when used to distinguish among notes of the same frequency. Their experience is that decay rate is almost useless in conveying data information.

Other work

The four case studies were chosen because they represent a variety of data analysis problems and a range of techniques for presenting that data in sound. Several other researchers have been exploring auralization in equally interesting and provocative ways. Mansur provided an early notion of "sound graphs" to convey (x,y) data to visually impaired scientists (Mansur 84, Mansur, Blattner, and Joy 85). The pitch varied continuously as a function of the changes in the y-axis. His goal was to offer a "holistic" view of the curve (a typical graph only lasted a few seconds) and an independence in exploring typically visual data. Mansur produced 13 training graphs and 22 testing graphs for identification of linear slopes, curve classification (linear or exponential), monotonicity, convergence, and symmetry. His subjects were able to answer the questions with 79%-95% accuracy. Mansur raises the issue of linear data being mapped to frequencies which are inherently logarithmic. One approach of his was to increase the frequency in a logarithmic fashion.

Work at the National Center for Supercomputing Applications offers a number of data visualization problems and aural representations (Scaletti & Craig 91). The work is focused on combining data-driven sound with data visualizations to elucidate the data further. The general method is to map streams of time-varying data to parameters of sound. Examples

include the movements of two pendulums to illustrate the behavior of a Duffing oscillator, forest fire suppression and its effect on decreasing forest diversification, and models of simulations such as air pollution and the human arterial system. A significant aspect of the work is the implementation and use of a set of tools for data auralization. This provides a strong base for examining many different data sets with various aural mappings.

Air flow turbulence is a rich data set for exploring sound representation (ref Blattner, Greenberg, and Kamegai) with interesting comparisons to the time-varying economic data of Mezrich et al. Blattner et al characterize fluids as continuous sounds and then vary parameters based on the fluid vicosity, density, temperature, speed of motion, direction of motion, the vortices (size, velocity, and density), and the energy dissipation. Although they have not produced an auralization of the data, they do offer a rationale for mappings to sound parameters as well as the suggested mappings themselves. Unlike the fairly straightforward mapping of Mezrich et al where separate voices represent the several data parameters, Blattner et al propose the creation of complex sounds utilizing not only frequency changes but waveforms, frequency relations, and tempo of the various component sounds. The introduction and dissipation of the vortices poses a combination of exploratory pattern recognition and event notification.

A type of data not directly considered thus far is that of process flow. In previous examples of time-varying data, we have been concerned with detecting patterns in the data. In the case of process flow, sound can offer both an on-going awareness of the status of the process and a notification of particular events. Program control or computational processes is an example of process flow used as a basis for auralization (Sonnenwald, Haberman, Myers, Gopinath & Keese 89, and Francioni/Jackson/Albright). Sonnenwald's work was concerned with simulation and the execution cycle for parallel computation. Something more here; look at the paper. In Francioni et al, three different mappings showed ways that auralization can contribute to information about program execution in distributed memory parallel programs. They used trace data consisting of event identifications, time, processor, message type, and message length. For issues of load balance in the parallel programs, each processor was mapped to a frequency that became louder as idle time increased. They followed program flow-of-control by assigning a specific timbre to each processor and then using a note or pitch for each event. Finally, the process communication tracked send and receive events by incorporating stereo to aid in the movement from send states to receive states. In all cases, Francioni et al found that they gained useful insights into the program execution, that they could obtain several different perspectives with a single pressing of the trace data, and that they could easily synchronize the aural and visual representations. Their various representation schemes are noteworthy for depending on the information needed rather than on an obvious mapping of data to sound. For example in considering the load balance, the load on each machine might have been mapped to a pitch. However, they were more concerned with the idle time and thus not only mapped idle time per processor to a note but also increased the loudness to draw attention to lengthy idle times.

Although several of the research projects consider the use of sound itself as opposed to considering only the data analysis problem, the work at NASA in developing the Convolvotron is significant in pushing the hardware to support the auralization (ref Wenzel). The Convolvotron provides externalized 3-dimensional cues over headphones in real-time digitally. Up to 4 moving or static sources may be simulated as localized sounds in 3-space. Several types of auralization are being explored such as computational fluid dynamics. Here the fuel flow around a liquid oxygen post of the main shuttle engine is visualized with a fly-through model. Auditory representations of various particles track the interactions of those particles with the shuttle engine.

Issues

The early work was exciting in verifying that auralization has potential for multivariate data representation. However, despite a growing interest in auralization, little has been done beyond the demonstration of concept. No compelling evidence exists to indicate that

auralization is more than an interesting twist on visualization. It is imperative that the community of research and development scientists in auditory data representation a) show beyond doubt at least one example in which auralization yields information in exploratory data analysis beyond other methods, b) provide tools and environments for other workers in the field, and c) issue guidelines for mapping data to sound parameters.

The reminder of this chapter outlines three areas of work that are informed by the studies to date and that offer difficult questions for further research. The three areas of work are represented by ten research problems. These address the issues of going beyond a demonstration of the concept of auralization to 1) a better understanding of multivariate data, 2) the representation of that data in sound, and 3) the evaluation of the representation. Though the problems clearly have relationship to one another, progress on each will offer valuable insights on which to build a theory and framework of auralization.

Issues of Data

Understanding the nature of multivariate data is key to exploring and understanding auralizations of that data. Here are two issues that attack the types of multivariate data and how those data types relate to sound.

Auralization Problem #1. What implications does the data type have for the type of auralization?

Scientific multivariate data may be loosely described as discrete events, as time-varying, or as continuous over some space. Discrete events have no inherent ordering. Time-varying data is such that samples occur in time sequence. Generally continuous data can be described by a function. Somewhat similarly, sounds may be thought of as single notes, chords, or sequences of notes (tunes). A note is made up of sound characteristics such as frequency, amplitude, waveshape, etc. A chord is several frequences or voices sounded together. A tune is a sequence of notes or chords. Given a sample

$$s = (x_1, x_2, x_3, ..., x_n)$$

the sample s could be a note as x_1 to x_n are mapped to the note characteristics, a chord as x_1 to x_n are mapped to different frequencies, or a tune as x_1 to x_n are mapped one at a time to successive notes.

An approach to understanding more about the relationship of the data type to the auralization method is to find a framework for the multivariate data and sound categories. Figure x attempts to classify the work described in this chapter into such categories. In the work to date, discrete samples have been represented by notes, by chords, or by tunes. Time-varying data have been represented as tunes by varying pitch only or by complex notes or chords. Data continuous in n-space has been represented as discrete events with a visual attachment for the location in space. It would be useful to the field of auralization to have a more complete space of data types and implications.

	DISCRETE	TIME-VARYING	CONTINUOUS
SINGLE NOTE	Bly, 6-dimensional Kramer, 9-d Grinstein, 5-d Rabenhorst, 2-d Michigan State, 2-d		
CHORD			
TUNE (one sample sequence)	Lunney		
TUNE (sequence of samples)		Mezrich Francioni Sonnenwald	

Auralization Problem #2. What data is appropriate for hearing directly?

   The issue here is to distinguish between aural representations which involve mapping data parameters to sound characteristics or directly presenting the raw data as sound. Several researchers have used data directly as an aural signal. For example, seismic signals can be heard by shifting the raw data to the audible frequency spectrum (ref paper from ICAD). The approach presented in this chapter focuses on n-dimensional measurements of the data which are then mapped into sounds. The relationship between these two approaches and a systematic understanding of the implications of each would benefit the understanding of multivariate data and auralization.

Issues of Sound Parameters

Perhaps the most crucial aspect of auralization is the mapping of data values to sound parameters.   Five issues below present a range of problems from straightforward relationships among sound components to user navigation in the aural space.

Auralization Problem #3. What constitutes a useful mapping of data to sound?

   It is too early to generalize the ways in which data could or should be mapped to sounds or to predict what information will be obtained by various representations. In taking a systematic look at mappings requires not only deciding which data parameter to represent by which sound parameter but also the relationship among the data parameters (are they independent or dependent variables and if dependent, what is the relationship) and the relationship among sound parameters (refer to the chapter on psychoacoustics).
   As in Problem #1, an approach to this issue might be to begin finding a framework for the possible data parameter relationships, the available sound parameters, the relationships among those sound parameters, and the perceived changes as sound parameters vary. The following audio parameters are used in perceptualization work to date.

   Frequency/Pitch: all

   Duration: Bly, Lunney, Kramer, Grinstein[1], Evans, Francioni, Michigan State
   Intensity/Loudness/Amplitude/Volume: Bly, Kramer, Grinstein, Craig, Francioni, Sonnenwald[2]

   Waveshape/Timbre: Bly, Grinstein, Rabenhorst[3], Francioni, Sonnenwald
   Attack: Bly, Kramer, Grinstein

Detune: Kramer, Rabenhorst[3]

Harmonics/Brightness: Bly, Kramer

Flange (phase shifting): Kramer

Pan (left/right stereo)/Direction: Kramer, Rabenhorst[3], Craig, Francioni, Sonnenwald

Rhythm/Pulse speed: Kramer, Evans, Sonnenwald

Distance?: Grinstein?

Decay: Grinstein

Vibrato/Tremolo: no one

[1] Grinstein et al get duration from the combined time of the attack and decay

[2] Sonnenwald has something to say about this.

[3] Rabenhorst doesn't exactly use these....

Auralization Problem #4. Where am I?

One of the disadvantages of using sounds for data representation is its temporary nature. While it may be relatively easy to know how one sound differs from another when they're presented in sequence, it's very difficult to remember any baseline over time. With the possible exception of those with perfect pitch, most listeners need guideposts or reference tones.

Two possible approaches might be to provide a constant underlying tone or to provide a baseline tone before each other tone.

Auralization Problem #5. How do parameters of sound relate to mathematical or data parameters?

As yet there has been no systematic calibration of sound parameters and data (mathematical) parameters. Though we talk about the value of psychophysics, there has been no attempt to relate such findings to mathematical parameters. Such an understanding of the relationship between aural and mathematical structures is imperative for auralization.

One approach would be to start with a good list of psychophysics/psychoacoutics results and a good list of mathematical relationships. Suggest ways that one would expect mathematical relationships to be heard in aural parameters. Test the resulting hypotheses. Issue a set of guidelines so that given data that varies in some known way, a particular set of sound parameters would be an accurate representation. Note work of Bly in thesis plus Mezrich et al and Frysinger.

A related problem is that most auralization to date does not take advantage of statistical methods of data analysis. Statistical methods offer two primary benefits. One they can often reduce the data to a simpler data set. Two, they can often find relationships among the data that can be exploited in the sound parameters. The value of the first goe without saying and no excuse can be offered for not investigating statistical methods before applying auralization techniques.   The value of the second requires an understanding of sound parameters as suggested by Problem #1. Using statistical methods to find relationships in the data is the first step in mapping the data to sounds. Those relationships in the data then become the basis for determining which sound parameters will be most effective in representing the data.

Auralization Problem #6. How does the audio space meet the graphics space?

Just as it is important to understand the relationship between parameters of sounds and parameters of mathematical data, it is important to understand the relationship between sounds and graphics. One needs to ensure that patterns available in the graphics space are consistent in the aural space. More interestingly, consider the integration or merger of the two. How do sound parameters vary relative to graphics parameters? If one maps a dimension in a graphical way and another in an aural way, do the two vary appropriately?

Approaching this issue is much like approaching the mapping problem generally but with the addition of visual parameters. The field of scientific visualization offers a basis for study. Problems #1, #3, #5, and #8-#10 suggest particularly relevant issues to consider.

Auralization Problem #7. Where is an ideal environment for auralization?

Given a set of multivariate data, there must be a computing environment to support the exploration. This environment must provide tools for statistical analysis, for mapping data dimesions to sound parameters, and perhaps to use graphics and audio together.

Two very excellent steps have been made to provide environments for exploratory data analysis that includes auralization. One is Exvis from the University of Lowell in Massachusetts (ref). The other is xsy from asdf in Illinois. Futher work will build on these environments as well as providing toolkits for building future perceptualization systems.

Issues of Evaluation

Underlying all the issues for auralization is a need to evaluate the usefulness and the information provided by the auralization techniques. Three problems offer pointers to evaluations that need to be considered.

Auralization Problem #8. How do we know if we're making progress?

Not only should we include evaluations in our work in auraliations but we should use appropriate methods. These issues are not altogether different from evaluations in other fields of computing applications and depend on the application of psychology and social science.

Frysinger (ref) has addressed this problem.

Auralization Problem #9. Can we hear anything that we can't see?

A central evaluation for auralization is whether or not it is a contributor to scientific perceptualization. The strong motivation for the value of auralization is the fact that more dimensions and/or different data types may be represented in sounds than in visual displays. If so, it should be possible to find patterns or characteristics in data with auralization that have not been found with standard statistical or graphical methods. Can information, in fact, be uncovered using auralization that has not been uncovered otherwise?

Two approaches to this problem are a) systematic data generation that creates patterns that can't be detected otherwise and b) a close collaboration with a scientific exploration of multivariate data. Two types of data seem particularly suited for auralization: time-varying and highly dimensional data. Tackling this problem in a close collaboration with scientists using such data offers an opportunity to push auralization techniques.

Auralization Problem #10. How annoying are sounds?

User interface issues for the data exploration are important in ensuring that users have good listening capabilities without hindering colleagues and that the sounds generated are acceptable.

Certainly earphones provide a means of listening to data. The sounds themselves may be influenced by cultural as well as psychoacoustical features. Several researchers have suggested anecdotally that they have used only frequencies in the musical scale because sounds based on general frequencies were difficult to hear and distinguish.