STEREOPHONIC AND SURFACE SOUND GENERATION FOR EXPLORATORY DATA ANALYSIS

Stuart Smith
Computer Science Department
University of Lowell
Lowell, MA 01854
stu@ulowell.edu

R. Daniel Bergeron
Computer Science Department
Kingsbury Hall
University of New Hampshire
Durham, NH 03824
rdb@unh.edu

Georges G. Grinstein
Computer Science Department
University of Lowell
Lowell, MA 01854
grinstein@ulowell.edu

ABSTRACT
The analysis and interpretation of very high dimensional data require the development and use of data presentation techniques that harness human perceptual powers. The University of Lowell's Exploratory Visualization project (Exvis) aims at designing, implementing, and evaluating perceptually-based tools for data presentation using both visual and auditory domains. This paper describes several auditory data presentation techniques, including the the generation of stereophonic sound with apparent depth and sound that appears to emanate from a two-dimensional area. Both approaches can produce sound with auditory *texture*.

KEYWORDS: Exploratory data analysis, sound perception, multi-dimensional data perception.

INTRODUCTION
The Exploratory Visualization (Exvis) project at the University of Lowell is a multi-disciplinary effort to develop new paradigms for the exploration and analysis of data with very high dimensionality. The fundamental philosophy behind Exvis is that data presentation tools should be driven by the perceptual powers of the human. In addition, the interpretation of data of very high dimensionality will be maximized only when we learn how to capitalize simultaneously on multiple domains of human perceptual capabilities. In particular, we have already begun experimentation with the presentation of data using visual perception capabilities including line texture, color, and motion (see Grinstein et al. [5] and Pickett and Grinstein [12]), and auditory perception based on one-dimensional *probing* of two-dimensional visual space (see Smith and Williams [14]).

ICONOGRAPHIC DISPLAY OF DATA
The critical requirement of an effective data display is that it stimulate spontaneous perceptions of structure in data. The scatterplot, for example, is effective because it stimulates a natural capac-

ity to sense the clustering of points in space. The Exvis project employs a novel "iconographic" approach that extends the concept of scatterplot and exploits another spontaneous perceptual capacity: the ability to sense and discriminate texture in line drawings (see Gibson [3] and Pickett [11]).

To stimulate visual texture perception, many small but discriminable elements must be displayed over a relatively small area. Such an image is experienced, for example, when one looks at an expanse of wall-to-wall carpeting. With deliberate visual analysis, one can obtain specific kinds of information about the lengths and materials of individual fibers, but one also receives, without any deliberate effort, an impression of the overall texture of the carpeting. It is just such natural impressions of texture that Exvis seeks to exploit.

In Exvis, each data sample is represented by a small graphical unit called an "icon." An icon's attributes, such as the sizes and positions of its parts, are data driven. Many icons deployed densely over the screen create an "iconographic display." The icons in an iconographic display are analogous to the fibers in the carpet. Just as footprints or the sweep of a vacuum cleaner will create variations in the texture of the carpet, so changes in the shape, size, spacing, and orientation of icons will create gradients or contours in the visual texture of the iconographic display. The challenge, of course, is to design icons that create highly distinctive textures. The iconographic display technique is described in detail in Grinstein, Pickett, and Williams [5]. Figure 1 shows an iconographic display representing census data take from the Public Use Microsample-A (PUMS-A) of the 1980 United States Census. Data such as this is particularly interesting because many of the fields contain discrete data values, such as *sex* and *highest educational degree* which may not map naturally to a continuous domain. On the other hand, such data can produce individual data icons that stand out as distinct entities within an overall visual pattern. For example, the choice of icon and data mapping used in figure 1 causes females with high education to be represented by an icon containing a small triangle. Figure 2 shows an iconographic display representing a solution of a partial differential equation that describes the flow field, temperature field, electromagnetic field, and reaction kinetics inside an inductively coupled plasma chemical reactor.

The iconographic approach has been extended to provide support for auditory textures: every icon has auditory, as well as visual, attributes. Depending on the values assigned to each of these attributes, an icon can produce a single tone or other sound on command. Multiple icons sounding simultaneously produce an auditory texture analogous to visual texture. The term "texture" is intended here in more or less its conventional musical sense, that is, the overall quality or surface character of music which arises from the relationships among its harmonic and melodic components. The auditory textures generated by our system would hardly be mistaken for traditional music, however. Typically the sounds are reminiscent of musique concrete, atonal music, aleatoric music, and other forms of "modern" music.

An analyst may elect to use only vision, only sound, or both. The same fields of the data sample that are used to determine the auditory attributes of an icon may also be used to determine the visual attributes. Alternatively, some or all of the data fields that determine the visual attributes may be different from those that determine the auditory attributes. Thus, the visual and auditory attributes may be used redundantly to reinforce each other or independently to maximize the number of data dimensions the icon can represent. The flat regions in the upper and lower portions of figure 2, for example, look similar; however, the sounds generated by the icons in each region are dramatically different. The auditory textures of these regions are clearly discriminable while the visual textures are not.

Presently, the auditory data presentation is triggered by the position of the mouse on the visual display. As the user moves the mouse, the icons at the successive mouse positions generate sound based on the data represented by the icons. If the user sweeps the cursor over many icons, the resulting multiple attacks and overlapping tones can produce auditory textures analogous to the visual textures

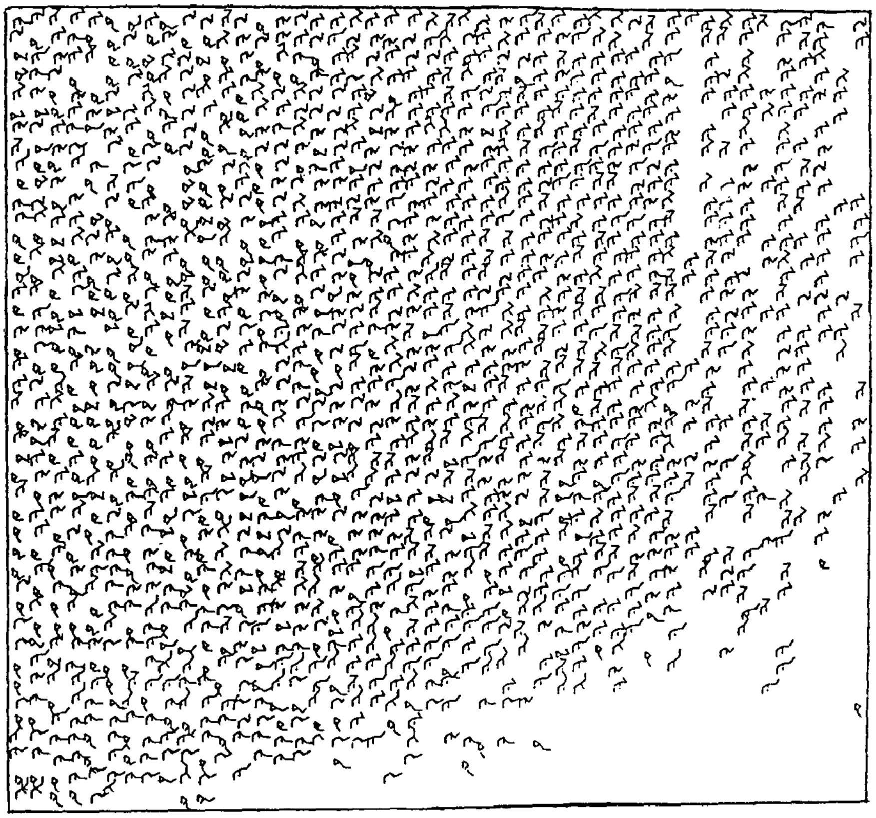

Figure 1: Iconographic display of census data.

produced by the icons. These textures are affected by the way the user moves the cursor: fast or slow, in a linear or circular motion, over a small or wide area of the screen, etc. Cursor movement determines, for example, how many notes start per unit of time, how many notes sound simultaneously, and how many different timbres may be sounding at once. In a sense, then, an iconographic display is a two-dimensional musical score which the user can play as he or she wishes.

This presentation of sound is effectively one-dimensional, where the dimension is time. Although this approach clearly does access real human perceptual capabilities (see Williams et al. [16]), it has significant limitations. Consider a visual display in which a user can see only the portion of the image that appears under a small (10x10 pixels) window at the current mouse position. It is possible to gain some sense of the image by moving the mouse cursor around the screen, but this technique requires remembering the parts of the image and integrating them in order to process or interpret the overall image. This limited approach certainly does not capitalize on our very significant visual power to integrate and interpret large areas of visual information at once. Our auditory system can also process and interpret sound information coming from many different directions simultaneously. We need to access this perceptual power in the presentation of complex data. The remainder of this paper describes the concept and design of a hardware/software environment to produce sound that appears to be generated by multiple sound sources.

REPRESENTING DATA IN SOUND

The psychophysical basis for the use of sound as a medium for presenting data is that sound has a number of perceptual properties which correspond in well-understood (though not necessarily simple) ways to its physical properties. For example, pitch corresponds closely to frequency, and loudness corresponds closely to intensity. While there are some interactions between pitch and loudness, these attributes can nonetheless be varied independently

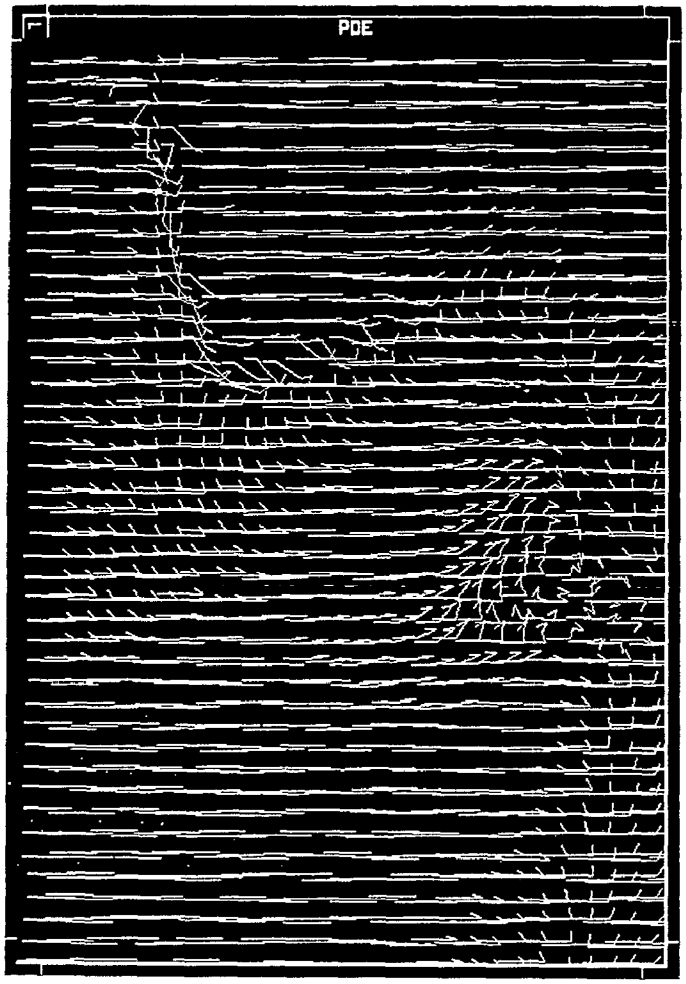

Figure 2: Iconographic display of a solution to a partial differential equation.

over broad ranges with predictable results. Among other sound attributes which can be manipulated independently are attack rate (the rate at which a sound grows to maximum loudness) and decay rate (the rate at which a sound dies away). Attack rate can be varied from nearly instantaneous, like the impact of a hammer striking an anvil, to slow and gradual, like the beginning of a tuba note. Decay rate can be varied in a similar manner.

To present multidimensional data in sound, it is necessary to encode data as the values of properties of sound. One way to do this is to establish a mapping between the dimensions of each data point and the properties of a discrete sonic event, such as a tone or burst of noise. For example, one data dimension can be mapped to pitch, another to attack rate, and a third to loudness. The measure on each dimension of the data point determines the value of the corresponding sound property of the event. The values of all the measures together determine a specific sonic event. In this way a set of data samples can generate a set of sonic events which may be heard individually, sequentially, or simultaneously.

The perception of the various properties of sound is described in a vast literature drawn from physiology, psychoacoustics, cognitive psychology, the psychology of music, musical acoustics, and music theory. Roederer [13] provides a highly readable summary of the pertinent results in these fields. The use of sound in interactive dialogues appears to be an effective mechanism for providing an additional or alternative channel of user feedback (see, for example, [4]). Buxton [2] offers a thoughtful discussion of the advantages and disadvantages of using specific sound attributes for presenting data.

RELATED RESEARCH

Four studies suggest the potential of sound as a medium for presenting data. Yeung [17] explored the use of sound as an alternative to graphic presentation of data. In Yeung's experiment, test subjects were asked to classify mineral samples from four sites in California using 7-dimensional chemical data encoded in sound. Each dimension of the data was mapped to a different property of sound and the resulting tones were presented to the subjects. The subjects were able to classify the samples with 90% accuracy without training, and with greater than 98% accuracy after training.

Lunney and Morrison [7] mapped the high frequency peaks of infrared spectra onto high-pitched tones, low-frequency peaks onto low-pitched tones, and assigned durations to the tones according to the intensities of the corresponding peaks. The resulting notes were played both sequentially as "melodies" and simultaneously as "chords." In informal tests that required subjects to match tone patterns produced by unknown substances with patterns produced by about a dozen simple organic compounds, subjects rarely failed to match identical patterns.

Bly [1] investigated the use of sound to add dimensionality to a two-dimensional graphic display. For her experiment, Bly synthesized 6-dimensional data samples and separated them into two sets according to a group of inequalities involving all six variables. Test subjects were to discriminate samples from the two sets when the data were presented visually, aurally, and in a combination of both. All six dimensions were mapped to properties of sound, but only two dimensions were mapped to the x-y scatterplot display. Trained subjects averaged 62% correct identification of samples with graphics alone, 64.5% correct with sound alone, and 69% correct with the combined graphics and sound presentation. Several months after the original experiment, subjects who had participated in the sound-only trials were given additional training. In a second experiment, these subjects correctly identified 74% of the test samples.

Mezrich, Frysinger, and Slivjanovski [8] used completely redundant visual and auditory displays to present multivariate time series data. Visually, each variable was represented by a pair of vertical lines which got longer and further apart as the value of the variable increased, shorter and closer together as it decreased. Simultaneously, each variable was represented by a tone whose pitch rose or fell according to the value of the variable. The set of lines and tones corresponding to a point in time was treated as a frame of a movie; the set of frames displayed in sequence at a suitable rate constituted the movie. In the experiment, test subjects had to discriminate correlated and uncorrelated data for four different time-series variables. The data were presented in four different formats: four separate graphs, four graphs aligned vertically, four graphs overlaid on the same pair of axes, and the movie format — four pairs of moving lines with sound accompaniment. The subjects performed this task significantly better with overlaid graphs and the movie technique than with separate graphs and vertically-aligned graphs. For sequences of ten or so frames, the subjects performed significantly better with the movie technique than with overlaid graphs.

THE STEREOPHONIC AUDITORY DISPLAY

Our initial sound implementation essentially produced monaural non-directional sound activated by the position of the cursor. Though effective, this technique utilizes only a small portion of our auditory perceptual abilities. Our first extension of this facility utilizes traditional stereophonic sound generation techniques which provide not only a left-to-right perceptual dimension, but also cues for the apparent distances of sound sources (see Olson [10], for example).

Figure 3 shows the hardware components of our current auditory display. The components are mostly standard MIDI (Musical Instrument Digital Interface) devices. The auditory display associates an independent musical "voice" with each icon of an iconographic display. The large number of logical voices represented by all the icons in a typical iconographic display are realized by eight physical voices belonging to a Roland D-110 tone generator. The physical voices are shared in round-robin fashion by the icons. Each physical voice is capable of producing a time-varying audio waveform under data control. The kinds of sounds that can be gen-

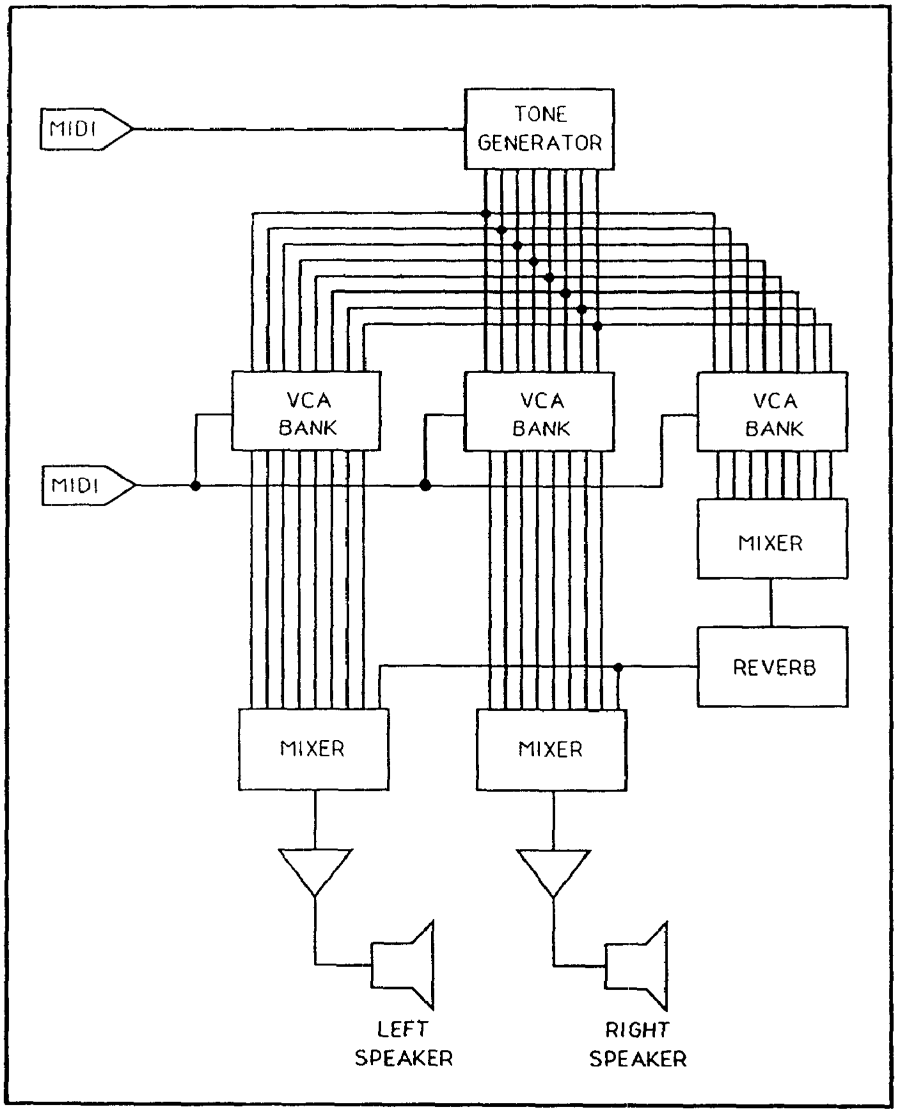

Figure 3: Hardware components of the auditory display.

erated range from musical tones, to more complex sonorities reminiscent of gongs and bells, and, at the extreme, to apparently random noise. The auditory display has the ability to place the apparent source of each icon's sound anywhere horizontally in front of the user and in apparent depth relative to the user, that is, it can create a conventional stereo sound image. Note in figure 3 that there are separate mixes of the eight sound sources for the stereo left channel, the stereo right channel, and the reverb unit. The contribution of each source to each of the three mixes is determined by the settings of the corresponding voltage-controlled amplifiers (VCA's). The relative level of each source in the left channel vs. the right channel determines the left/right placement of that source in the stereo image, and the level of the source in the reverb channel determines its apparent distance from the listener. These capabilities permit clusters of icons to be heard as a group in some region of the stereo sound image. Because the user can map any desired data fields to the stereo-left, stereo-right, and reverberation parameters, the user can choose whether or not the apparent location of each icon's sound source is correlated with the icon's horizontal position in the visual display. The stereo image generated by this auditory display could be said to be "1.5-dimensional." It has one real dimension defined by the left and right stereo channels. The distance effect created by reverberation is an aural illusion analogous to perspective in the visual arts, hence not entitled to full status as a dimension.

We have already accumulated some experience in the use of sound in an iconographic visualization environment. Some preliminary results obtained in perception experiments by Marian Williams may be found in Williams, Smith, and Pecelli [16].

GENERATION OF SURFACE SOUND

The visualization workstation we are developing presents the user with a rectangular visual display window and a concentric rectangular auditory display window. The visual and auditory display spaces may extend beyond the the physical visual and auditory windows. When this is the case, the workstation allows the user to pan over the display space. The workstation also provides both visual and auditory zoom capability. As the user zooms in on a region of the display, the icons in that region come acoustically to the foreground at the same time that they grow visibly larger. It is possible to pan and zoom within the auditory and visual displays independently.

Each icon in a display is able to generate its sound simultaneously with, and independently of, all other icons. By default, each icon's sound emanates from a direction correlated with the icon's position in the visual display space; however, because the user is free to map any desired data fields to the sound parameters which determine the direction of the sound source, spatial realism is a user-selected characteristic of the auditory display.

The sounds of the icons are turned on and off according to a user-selected pattern. The pattern may be an ordering based on the values in a given data field, a geometric scheme such as a left-to-right/top-to-bottom scan, or a random sequence. The repetitions of groups of icons related under a particular pattern are synchronized, but not necessarily simultaneous (the attacks may be staggered by a fixed time interval). Turning the icons on and off in this way increases the likelihood that clusters of related icons will be heard as an area or surface of sound. This effect is especially important when sound calls attention to structures not evident in the visual representation of the data.

Implementing a true two-dimensional auditory display, which allows placement of the apparent sound source anywhere in an imaginary vertical window in front of the user, is problematic. The key problem is synthesizing the cues for the *elevation* of a sound source. There is as yet no standard technique for synthesizing such cues. Kendall, Martens, and Decker [6] and Moore [9] have developed elegant models for controlling the apparent spatial locations of sound sources. Both models, however, require lengthy computations that cannot be accomplished in real time with general-purpose computers. A promising technique developed by Wenzel, Wightman, and Foster [15] uses the direction-dependent filtering characteristics of the outer ear to synthesize a spatially-realistic three-dimensional auditory display. Crystal River Engineering of Groveland, CA offers the "Convolvotron," a two-card set for the IBM PC or PC-AT that synthesizes a 3D auditory display according to this method. The Convolvotron can place 1-4 sound sources in auditory 3-space in almost real time (there is a delay of 30-40 milliseconds for directional controls and audio to propagate through the system).

We believe that powerful special-purpose devices like the Convolvotron will prove useful for the incorporation of auditory representations within visual displays. The MIDI devices we are currently using, which are designed to meet the needs of the popular music market, are not well suited to scientific work. Eliciting the full sonic resources of these devices requires extensive use of the "System Exclusive" capability, which is nothing more than a device-dependent bypass of the MIDI standard.

SUMMARY

Experimentation with sound as a medium for representing data has been underway since at least the beginning of the 1980's. The Exvis project brings several new ideas to this work: interactive manipulation of sound, iconographic data representation incorporating sound attributes, and the integration of auditory and visual displays into a single unified data exploration facility.

REFERENCES

1. Bly, S. A. Presenting Information in Sound. *Proceedings of the CHI '82 Conference on Human Factors in Computer Systems*, pp. 371-375, 1982.

2. Buxton, W. The Use of Non-Speech Audio at the Interface. *Tutorial #10, CHI '89*, pp. 2.1-2.15, 1989.

3. Gaver, William W. The SonicFinder: an interface that uses auditory icons. *Human-Computer Interaction*, Volume 4, Number 1, pp. 67-94 (1989).

4. Gibson, J.J. *The Perception of the Visual World.* Boston: Houghton-Mifflin, 1950.

5. Grinstein, Georges, Ronald Pickett, and Marian G. Williams. EXVIS: An Exploratory Visualization Environment. *Graphics Interface '89.* London, Ontario, 1989.

6. Kendall, Gary S., William L. Martens, and Shawn L. Decker. Spatial Reverberation: Discussion and Demonstration. in M.V. Mathews and J.R. Pierce, *Current Directions in Computer Music Research.* Cambridge: MIT Press, 1989.

7. Lunney, D. and R.C. Morrison. High technology laboratory aids for visually handicapped chemistry students. *Journal of Chemical Education,* 58(3):228-231, 1981.

8. Mezrich, J.J, S. Frysinger, and R. Slivjanovski. Dynamic Representation of Multivariate Time Series Data. *Journal of the American Statistical Association,* 79(385):34-40, 1984.

9. Moore, F.R. Spatialization of Sounds over Loudspeakers. in M.V. Mathews and J.R. Pierce, *Current Directions in Computer Music Research.* Cambridge: MIT Press, 1989.

10. Olson, Harry F. Electronic music synthesis for recordings. *IEEE Spectrum*, pp. 18-30, April 1971.

11. Pickett, R.M. Visual Analyses of Texture in the Detection and Recognition of Objects. in B.S. Lipkin and A. Rosenfeld, *Picture Processing and Psycho-Pictorics.* New York: Academic Press, 1970.

12. Pickett, R.M. and G.G. Grinstein, Iconographics Displays for Visualizing Multidimensional Data. *Proceedings of the 1988 IEEE Conference on Systems, Man and Cybernetics.* Beijing and Shenyang, People's Republic of China, 1988.

13. Roederer, Juan. *Introduction to the Physics and Psychophysics of Music.* New York: Springer-Verlag, 1973.

14. Smith, Stuart and Marian G. Williams. The Use of Sound in an Exploratory Visualization Environment. *University of Lowell Computer Science Department Technical Report No. R-89-002,* 1989.

15. Wenzel, Elizabeth M., Frederic L. Wightman, and Scott H. Foster. Development of a three-dimensional auditory display system. *SIGCHI Bulletin,* 20(2):52-57, 1988.

16. Williams, Marian G., Stuart Smith, and Giampiero Pecelli. Computer-Human Interaction Issues in the Design of an Intelligent Workstation for Scientific Data Visualization (Phase 1). *ACM SIGCHI Bulletin,* to appear. (Preprints available as University of Lowell Computer Science Department Technical Report No. R-89-006.)

17. Yeung, E.S. Pattern Recognition by Audio Representation of Multivariate Analytical Data. *Analytical Chemistry,* 52(7):1120-1123, 1980.