# A Brief History of Auditory Data Representation to the 1980s

*Steven P. Frysinger*

College of Integrated Science and Technology,
Departments of Computer Science & Integrated Science and Technology,
James Madison University,
Harrisonburg, Virginia
USA.
frysinsp@jmu.edu

**ABSTRACT**

The field of Auditory Data Representation, which addresses the representation of quantitative data through the use of auditory, rather than visual, displays, has seen considerable activity in the last twenty years. On the occasion of the first Symposium on Auditory Graphs it is well to consider the roots of this field. This paper presents a brief history of the field, leading up to the beginning of the 1980s, and accompanies a demonstration of a multivariate time series representation developed by the author and his colleagues in 1980.

## 1. INTRODUCTION

Auditory Data Representation is a term used to reflect the use of sound to display quantitative information. The present work attempts to document the early history of this field, and accompanies a demonstration of a combined auditory/visual display of multivariate time series data developed in 1980 by the author and his colleagues. Much of this paper is extracted with modification from an unpublished thesis written by the author [1].

## 2. EARLY EXPLORATION

One of the earliest investigations of quantitative auditory displays to be found in the open literature was conducted by Pollack and Ficks [2], in the wake of the birth of information theory by Shannon [3], and was primarily concerned with evaluating the information transmission properties of auditory stimuli. Although Pollack and Ficks did not allude to applications of auditory displays, they did evaluate two different mappings of multidimensional data onto the parameters of sound. In the first display type, they presented the subject with a tone and noise in alternation, and represented eight variables as binary parameters:

1. Frequency range of the noise (100-500 or 5000-8000 Hz)
2. intensity of the noise (40 or 105 db)
3. frequency of the tone (100 or 6000 Hz)
4. intensity of the tone (40 or 105 db)
5. alternation rate (0.4 or 4.0 interruptions per second)
6. temporal ratio of tone to noise (10% or 90%)
7. total duration of the display (5 or 17 seconds)
8. apparent direction of origination of the display (-90 or +90 degrees)

In the second display type, the noise-related parameters of the first type were excluded, leaving an interrupted tone described by the last six parameters. These parameters were assigned either two, three, or five levels, with the limits held constant for all variations, instead of only two as in the first display type.

Using these two display types, the second with its three variations, Pollack and Ficks [2] measured the information transmitted to subjects as the sum of the number of bits in each correctly identified dimensional level. Their results indicate that multidimensional displays, that is displays using multiple parameters of sound, in general outperformed unidimensional displays measured elsewhere, and that subdivision of display dimensions into finer levels does not improve information transmission as much as increasing the number of display dimensions does. This result will very likely have a great influence on the nature of displays used for Auditory Data Representation.

Another early example of Auditory Data Representation experimentation was published by Speeth [4], who was searching for improved ways to discriminate earthquakes from underground bomb blasts based upon seismic measurements. Given the extremely complex vibration patterns measured by the seismometer, this task was apparently very difficult to perform using visual plots of the data. So Speeth sped up the playback of data recorded by seismometers to place the resultant frequencies in the audible range, and then set human subjects to the task of determining whether the stimulus was a bomb blast or an earthquake, after an appropriate training program.

In Speeth's [4] experiment, subjects were able to correctly classify seismic records as either bomb blasts or earthquakes for over 90% of the trials. Furthermore, because of the time-compression required to bring the seismic signals into the audible range, an analyst could review twenty-fours hours worth of data in about five minutes, making the technique extremely attractive as a monitoring and surveillance tool.

Chambers, Mathews, and Moore [5] also investigated the use of sound to represent quantitative data, this time using multiple parameters of sound to encode those dimensions of multidimensional data which were not displayed on a conventional scatter plot. Their auditory display was based on three parameters:

1. Frequency (150-700 Hz, quantized chromatically)
2. Spectral content (an additive formant frequency, 50-8000 Hz, also chromatic)
3. Amplitude modulation (amplitude of 15 Hz modulator proportional to data)

Without formal experimentation, they found that their auditorily-enhanced scatter plot display system promoted the classification of multivariate data.

## 3. BATTLE SONGS AND IRISES

A more comprehensive study of Auditory Data Representation was published by Bly [6] in her thesis, wherein she evaluated auditory displays for three classes of data: multivariate, logarithmic, and time-varying.

In considering multivariate data, Bly [6] was interested specifically in discriminating non-ordered sets of multivariate data points, and in attempting to classify an unknown data point as belonging either to one set or the other. In her multivariate-display system, one data point (either an unknown or a representative of a set) would be sounded at a time, with its various dimensions (up to seven) mapped onto the following parameters of sound:

1. Frequency (48 chromatic levels, from 130 to 2000 Hz);
2. Intensity (12 levels from "very soft to very loud");
3. Duration (201 levels from 50 to 1050 msec);
4. Fundamental waveshape (128 levels from pure sinusoid to random noise);
5. Attack envelope (15 levels from "long attack" to constant amplitude);
6. Fifth harmonic (128 levels from pure sinusoid to random noise added to the fundamental);
7. and ninth harmonic (128 levels from pure sinusoid to random noise added to the fundamental).

One data set to which she applied her auditory display technique was the Iris data set of Fisher [7]. These data characterize samples of three different species of flower using four measurements per plant (sepal length, sepal width, petal length, and petal width). Although one species is easily distinguishable, the other two have some overlap with each other, and thus present a problem to the analyst attempting to classify an individual plant as belonging to one species or the other. Displaying the 4-dimensional data auditorily, Bly [6] found that most observers could correctly classify all but one or two of the samples.

In representing logarithmic data, Bly [6] was motivated by the logarithmic relationship between frequency and pitch, and therefore encoded the exponential variable in pure frequency without conversion to a chromatic scale. She found that the resulting displays were useful in highlighting features in seismic records of earthquakes.

Finally, Bly [6] represented time-varying multivariate data with the frequency and intensity of multiple tones. To help distinguish the tones, different waveforms were used for each, though waveform itself did not correspond to a dimension of the data. She applied this technique to simulated, two-sided military battles by assigning one tone (sinusoidal or noisy) to each side. The frequency of each tone represented the number of units that side had at the front, and the intensity represented the number of units in transit to the front. In the resulting "battle songs" listeners were able to distinguish battles which had the same outcome but which evolved differently, although they apparently had difficulty tracking the tones for each side independently.

To validate her approach to Auditory Data Representation, Bly [6] conducted a series of formal experiments on multivariate data displays. The experiments considered two 6-dimensional data sets which differed by translation, scaling, or correlation, and tried sound only, graphics only, and bimodal displays. She also experimented with changes in the mapping of data values to sound parameters and changes in training methods. In all experiments, the subjects were classifying an unknown test sample as belonging either to one set or the other, which sets differed in a well-defined way made known to the subjects beforehand by unrestricted training.

The most telling experiment used data sets which were completely non-overlapping only in six-space, thus representing a clear multivariate discrimination problem. Bly [6] compared her auditory display scheme to a visual scatterplot, and also to a combined (redundant) auditory/visual representation. The results of this experiment indicate that the auditory display was at least as effective as the visual display, and that the combined display outperformed them both.

## 4. ACOUSTIC CHROMATOGRAPHY

Morrison and Lunney [8], interested in presenting analytical chemistry data to visually impaired students, developed a somewhat more elaborate scheme for representing infrared spectral data in sound. In one of their representations, the pitch of a tone is proportional to the frequency location of the infrared peak it represents. These are first played sequentially in descending pitch order, with note durations proportional to the intensity of the represented infrared peak, producing a descending arpeggio of varying member note intensity. Then the same data are played sequentially in descending order of peak intensity with equal note durations. Finally, a chord (usually highly-dissonant) is formed by sounding all of the peak-notes at once with equal intensity. They informally found that identical matches were reliably made from a set of spectra produced by approximately twelve organic compounds.

Similarly Yeung [9], in preparing an audible display for experimental data from analytical chemistry, sought auditory parameters exhibiting continuity in scaling and relative independence from each other. His parameters were

1. Frequency (two dimensions, 100-1000 and 1000-10,000 Hz, logarithmically indexed)
2. Intensity
3. Damping
4. Direction (left to right)
5. Duration/Repetition
6. Rest

His display consisted of data vectors, each dimension of which corresponded to the detected levels of various metals in a given sample, with one vector per sample. The analysis task involved classifying a given vector as belonging to one of four sets, after having been trained with vectors from those four sets. Although Yeung did not compare the performance of his subjects using the auditory display with that of any other display, he noted that all of his subjects achieved the 98% correct classification rate after (at most) two training sessions.

## 5. ECONOMIC INDICATORS

Mezrich, Frysinger, and Slivjanovski [10] developed a dynamic representation employing both auditory and visual components for multivariate time-series displays. Such data play an extremely important role in human decision making.

Time-series data are best characterized as discrete functions of an ordered independent variable which is often, though not always, a quantized representation of time. Multivariate time-series are multiple functions of the same independent variable. These functions may be independent, correlated, or exhibit some other relationship or pattern.

Time-series data are often displayed visually in x-y plots, with the multiple dimensions either overlaid, stacked, or displayed on separate axes. Such visual displays are almost always static; all of the available data are drawn on the display at once and then examined by the analyst.

The dynamic representation developed by Mezrich et al. [10] represents multivariate time-series redundantly, employing both auditory and visual components. Although this representation was intended for oil well log data, a type of multivariate time-series common to the oil exploration industry, the proprietary nature of such data made it difficult to gain access to meaningful examples. Thus, they turned to economic indicators, which are statistically similar to well logs, and are generally in the public domain.

Individual economic indicators are univariate time-series describing the temporal fluctuations of such things as car sales and housing starts. They are not usually of interest individually, but when combined they form a window onto the state of the economy. The difficulty is that there are no widely-accepted models for the interactions among these indicators; in practice they are simply co-plotted on a visual display and analyzed for "interesting behavior". Unfortunately, the visual displays are too complex for meaningful visual inspection, especially when the number of indicators grows. As a result, the indicators are often grouped by weighted linear combination into a single index, which is then used as an economic predictor. The problem with this is that, lacking well-founded models of interaction among the indicators, the weighted combinations have questionable validity and tend to throw away information which was available in the individual time-series.

The representation developed by Mezrich et al. [10] allows the analyst to "view" (i.e. hear and see) the unperturbed indicators without experiencing "sensory overload", and further permits interaction with the data display which was not previously available. In their scheme, the analyst is confronted at any moment with one multivariate sample from the time-series, rather than the whole data set. These samples are displayed in succession, forming "frames" of data analogous to frames in a movie. Each frame consists of a collection of visual objects whose position and size correspond to the dependent variable values, and a collection of simultaneously-sounding musical notes whose frequencies correspond to the same dependent variable values.

The auditory representation used for these multivariate time-series assigned one "voice" to each variable. The chromatic frequency of the voice was proportional to the value of each variable, and all other parameters (such as intensity and attack) were held constant. Because the voices for all variables used the same range of frequencies, they could overlap as they progressed in succession. The default condition assigned the same waveform to all voices, so that they were essentially indistinguishable; however, the analyst could interactively "enhance" one or more variables by assigning its voice a brighter waveform (i.e. one with more harmonics), "mute" a variable by assigning its voice a pure sinusoidal waveform, or simply remove a variable from the display altogether. The homogeneous assignment seemed to be effective at promoting global pattern recognition, while the interactive enhancement facility permitted local scrutiny at the analyst's discretion, although this observation has never been tested formally.

The temporal nature of the representation permitted novel interactions with the display. For example, the analyst could "play" the data either forward or backward in the independent variable, providing two distinct (though not independent) "views" of the data. Furthermore, sub-series could be marked, saved, and played in temporal juxtaposition with each other to facilitate comparisons.

To formally evaluate the effectiveness of this bimodal data representation, Mezrich et al. [10] compared it to three commonly-used static visual representations, manipulating the degree of correlation of the randomly generated stimulus time-series, and measuring the psychophysical threshold of correlation detection for each display technique. A secondary manipulation was the number of samples in the stimulus time-series.

The results of their experiment indicated that the dynamic auditory/visual display outperforms the static visual displays with one notable exception. When a visual display was constructed of overlaid time series plots, it was outperformed by the dynamic display only for small sample sizes (i.e. series lengths), and became essentially equivalent to the dynamic display for longer series. Mezrich et al. [10] conjectured that this result reflects the fact that both the dynamic display and the overlaid static display facilitate global pattern recognition, while the stacked and separate-axis displays required local pattern recognition by feature scrutiny. Within the realm of global pattern recognition, the dynamic display allowed the subject to detect correlation with fewer points than were required when using the overlaid display.

In subsequent experiments, Frysinger [1] further examined the performance implications of different data set sizes, as well as different detection tasks.

## 6. CONCLUSIONS

The foregoing illustrates two important points about the field of Auditory Data Representation. The first of these is that the idea of displaying data through sound has been with us for quite some time. The second, though, is that relatively little progress was made in the field until the 1980s and early 1990s (e.g. when the International Conference on Auditory Display series commenced). There is at least one technological reason for this, in that ready access to digital sound generation technology didn't become available until the advent of sound generation cards for personal computers in the mid-1980s, as well as the development of the MIDI standard.

While Auditory Data Representation is still not a standard feature of commercial spreadsheet packages, the historical foundation suggests that such an outcome is possible. What is lacking is scientifically-informed guidance in the use of sounds for data representation. While visual displays have been used for centuries without a good psychophysical framework, auditory displays are somewhat less intuitive, and the field will therefore benefit greatly from rigorous and accessible research.

## 7. REFERENCES


[1] S. P. Frysinger, *Pattern Recognition In Auditory Data Representation*, Unpublished Thesis, Stevens Institute of Technology, 1988.

[2] I. Pollack and L. Ficks, "Information of Elementary Multidimensional Auditory Displays," Journal of the Acoustical Society of America, vol. 26, pp. 155-158, 1954.

[3] C. E. Shannon, "A Mathematical Theory of Communication," Bell System Technical Journal, vol. 27, pp. 623-656, 1948.

[4] S. D. Speeth, "Seismometer Sounds," Journal of the Acoustical Society of America, vol. 33, pp. 909-916, 1961.

[5] J. M. Chambers, M. V. Mathews, & F. R. Moore, "Auditory Data Inspection," Technical Memorandum 74-1214-20). Murray Hill, NJ: AT&T Bell Laboratories, 1974.

[6] S. Bly, "Sound and Computer Information Presentation," Unpublished doctoral dissertation, University of California, Davis, 1982.

[7] R. A. Fisher, “The Use of Multiple Measurements in Taxonomic Problems,” Annals of Eugenics,vol. 7, pp. 179, 1936.

[8] S. Bly, S. P. Frysinger, D. Lunney, D. L. Mansur, J. J. Mezrich, J. J., and R. C. Morrison, “Communication with Sound,” In R. Baecker & W. A. S. Buxton (Eds.), *Readings in Human-Computer Interaction: A Multidisciplinary Approach,* Morgan Kaufmann, 1987.

[9] E. S. Yeung, “Pattern Recognition by Audio Representation of Multivariate Analytical Data,” Analytical Chemistry, vol. 52, pp. 1120-1123, 1980.

[10] J. J. Mezrich, S. P Frysinger, and R. Slivjanovski, “Dynamic Representation of Multivariate Time-series Data,” Journal of the American Statistical Association, vol. 79, pp. 34-40, 1984.