2005

Thirteen Years of Reflection on Auditory Graphing: Promises, Pitfalls, and Potential New Directions

John H. Flowers
University of Nebraska-Lincoln, jflowers1@unl.edu

THIRTEEN YEARS OF REFLECTION ON AUDITORY GRAPHING: PROMISES, PITFALLS, AND POTENTIAL NEW DIRECTIONS

John H. Flowers

Department of Psychology,
University of Nebraska -- Lincoln
jflowers1@unl.edu

ABSTRACT

While developments in sound production hardware now make the creation of auditory graphs possible for casual users of personal computers, some of the same pitfalls to effective auditory display development that arose in the early 1990's continue to impede effective applications of this promising technology. Most of these pitfalls stem from lack of adequate understanding about key properties of auditory perception and attention and from inappropriate generalizations of existing data visualization practices. At the same time, however, we now know about some strategies that appear to work and offer promise for making sonification a useful and accepted tool for data exploration and decision making. The present paper summarizes several selected examples in each of these categories, along some suggestions for future research directions.

1. INTRODUCTION

The author and his students first began to explore auditory graphs in early 1990's, initially on an informal basis as a teaching tool, and eventually as alternative data exploration/display technique that appeared to have potential for a variety of research purposes. In those days, sonification (and even advanced visualization) of complex multivariate data was in its early stages of development. Pursuit of sonification research was almost exclusively limited to institutions with access to advanced supercomputer facilities and highly specialized hardware for auditory synthesis – resources that those of us in a psychology department on the Great Plains (who used our early generation PC's for square wave audio synthesis) observed with awe and envy. But, coming from the perspective of cognitive psychology, as opposed to engineering, music, or computer science, we also observed design recommendations, and claims about the potential of sonification that seemed to violate basic principles of human perception, attention and memory. These included a view that the "data bandwidth" of humans could be vastly augmented by sound (either alone or in combination with other modalities), and that given the right mappings, we could gain immediate insight into exceptionally complex multivariate data, or instantly pick out critical events or a single data stream of interest among a large number of other streams. There also seemed to be a tendency to make choices about format of auditory data displays that were based on accepted practices for visual graphics, some of which neglected important differences about how vision and audition operate. And finally, there often seemed to be little consideration that, like visualization, optimal display formats may be highly dependent on the specific *task context* to which they are to be applied.

More than a decade later, the capability to control pitch, timbre, loudness and temporal properties of several auditory channels by relatively complex data streams is now possible from our desktops. In addition, technology innovations have stimulated the collection of enormous amounts of complex data in many scientific and business related domains, and sonification is finally being explored as a viable option for addressing that "data glut." There appears to be increasing consensus about some strategies of sonification that actually "work" for a variety of applications and tasks. However, as a reviewer of sonification studies for conference presentations and articles, and a direct observer of applications that failed to work as hoped (including those from my own laboratory), it is clear to me that some of the same categories of pitfalls observed thirteen years ago continue to be repeated in sonification attempts. In particular, there continues to be an over-optimism about human "data bandwidth" and inadequate consideration of some key differences between auditory and visual perception that impact optimal display design.

2. SOME THINGS THAT "WORK"

The following (non-exhaustive) list describes a sample of sonification design principles that previous and current research and applications have used with some success. While the usefulness of some of these may vary according to the specific task context, it is the author's view that they provide a good starting point for thinking about a sonification design for making an auditory graph.

2.1. Pitch Coding of Numeric Value

Pitch profiles are a compelling dimension for representing changes in numeric values. Mapping pitch height (essentially log frequency) to numeric magnitude affords perception of function shape or data profile changes, even for relatively untrained observers. Some data exploration tasks, such as those requiring point estimation [1], may also require additional context to calibrate pitch with data values. There may also be situations where metaphoric or cultural stereotypes for magnitude of specific continua might be better supported by the reverse mapping of pitch to quantity [2], but for representing numeric quantity in scientific data exploration these exceptions are likely to be infrequent.

2.2. Exploiting Temporal Resolution of Human Audition

The temporal resolution of the auditory system is exquisite. The duration of sound streams can be successfully mapped to numeric quantities, as in the case of auditory histograms. In addition, the temporal patterning of sound events can also provide information about changes in the *distributional properties* of numeric values. For example, auditory scatterplots can reveal fine-grain details of bivariate distributions such as gaps, outliers, or even differences in the roundoff procedures for different data samples -- features that may be missed in inspection of visual scatterplots [3]. This suggests that for data observations where changes in short-tem variability are an important issue (e.g. sudden changes in frequency of time differences between seismic events that could be predictive of an eruption or quake, chaotic excursions of solar emission related variables predictive of planetary geomagnetic events) the auditory system might provide considerable insight.

2.3. Manipulating Loudness Changes in a Pitch Mapped Stream to Provide Contextual Cues and Signal Critical Events

Temporal or rhythmic patterning of loudness levels, especially when integrated into pitch and timbre defined data streams may be highly useful. Rhythmic patterning of a pitch mapped stream can carry time or X-axis information, and use of such a time patterned stream may be more efficient than placing timing markers in a separate stream of clicks, beeps or percussion instruments [1]. Loudness changes can indicate key events in a pitch coded stream such as axis crossing, reaching of some criterion value, etc. If one treats loudness changes as a means of emphasizing or marking categorical changes or temporal markers, as opposed using loudness for continuous quantitative mapping, concerns about dimensional interactions between pitch and loudness changes that might bias quantitative estimation [4] are minimized. It should be noted that momentary timbre changes, perhaps in conjunction with loudness changes can be used in the same manner.

2.4. Choosing Distinct Timbres to Minimize Stream Confusions and Unwanted Grouping

Timbre differences can be useful for minimizing unwanted perceptual grouping of separate continuous data streams when multiple continuous variables are required to be plotted, but there are serious (not always well understood) principles that need to be considered when doing so. Timbre changes due to onset envelope differences in note streams probably allow better separation than timbre differences due to harmonic content per se. Conversely, harmonically impoverished continuous sine waves, aside from not being particularly pleasing to the ear, are probably *not* good choices for use in pitch coding of numeric values, particularly if the auditory display contains multiple data streams [5].

2.5. Using Time to Represent Time

Multivariate time series datasets that include combinations of continuous data and discrete events, are often good candidates for sonification. Such displays can exploit the auditory system's sensitivity to changes in temporal patterning, thereby calling attention to data features that may be less apparent in traditional visual graphs or charts. Climate data, longitudinal health records, and various types of remote sensing data are examples of time series data for which sonification could be useful. However as the next section will indicate, judicious selection of variables and mappings can be an issue; plotting "everything at once" in a dense multivariate data sample can lead to "a mess" in either sound or sight.

2.6. Sequential Comparisons of Sonified Data

For many data exploration activities, a sonification design that permits *sequential* comparisons of short data streams may be more efficient than one that presents the data streams simultaneously. This may be particularly important for tasks in which detection of differences in pitch coded data profiles or function shapes is important for a research or decision making activity. While profile differences may "pop-out" in visual line graphs when two or more profiles are printed on the same axes (e.g. interaction effects in factorial experiments), such is not the case with pitch-defined data profiles. Melodic (and also temporal pattern or rhythm) comparisons are often best done sequentially. Software tool development should thus be sensitive to the need to allow continuous listening to data over time, but also to allow selection of segments of interest for sequential comparison. The duration of display segments selected for sequential comparisons is, however, a critical issue that must take into account the limits of working memory and auditory sensory memory [6].

3. APPROACHES THAT DO *NOT* WORK

This is also a non-exhaustive list. Other participants in this symposium will undoubtedly have a good many examples to contribute to this category! The three that are presented here reflect examples attributable to common misunderstandings of auditory perception and attention.

3.1. Simultaneous Plotting of Numerous Continuous Variables -- Particularly via Pitch Mapping.

While the claim that submitting the entire contents "dense and complex" datasets to sonifciation will lead to the "emergence" of critical relationships continues to be made, I have yet to see it "work." This is not surprising since a visualization with dense overlaying of line graphs does not work well either, except for selected instances such as visual clustering of event makings or sparse graphs (e.g. plotting on a map of disease outbreaks or tornado paths over time). Listening to simultaneously plotted multiple continuous pitch mapped data streams, even when attention is given to timbre choice for different variables to reduce unwanted grouping, is probably not productive. It is possible that with levels of consistent practice that are well beyond those of most sonification evaluation studies, we might do somewhat better at listening to multiple sonified streams than is currently apparent. But it is generally the case that attending to three or more continuous streams of sonified data is extremely difficult even when care is given to selection of perceptually distinct timbres or temporal patterning. There may be some auditory analogies to visual plot clustering (auditory scatterplots offer a primitive example) but it will take a great deal of display engineering, and judicious variable selection to make this principle apply to more complex multivariate data. Insightful data groupings are not likely pop out by simply submitting a multivariate set to a sonification engine and "plotting everything."

3.2. Simultaneous Plotting of "Just a Few" Continuous Variables with Similar Timbres.

Simultaneous presentation of streams with only moderate degrees of similarity can create confusion possibilities, and unwanted grouping. Use of onset envelope differences in timbre and patterned or modulated streams will help keep them separate, but at tradeoff for observing value crossings, etc. This is another reason to design sonification software that can easily switch from simultaneous to sequential presentation of segments of data streams.

3.3. Using Loudness Changes to Represent an Important Continuous Variable.

Loudness is a function of surrounding auditory context, not just the intensity programming of the auditory stream, due to the complexity of masking and related issues. It is not independent of pitch changes. Even with isolated presentation of a single auditory stream of constant pitch, ability to discriminate different loudness levels for reliable to mapping numeric values is far more limited than for pitch (log frequency) mapping to quantity, or temporal auditory changes such as modulation rate, pulse or note rate, etc. In addition, there is a major non-perceptual factor that makes loudness unsuitable for carrying fine-grained quantitative information – limitations of sound reproduction equipment, and differences in the dynamic ranges and general quality of such equipment from setting to setting. If auditory graphs are to become a a widespread technique for data exploration and other purposes, it would be best to choose dimensional mappings that do not become distorted when rendered on different equipment of with different loudspeakers or earphones. Pitch, pulse rate, etc., are far more robust than loudness in this regard. As stated previously loudness manipulation for signifying categorical events, or modulation of streams carrying pitch coded information for providing time scale context, etc. may be extremely useful. However, for most auditory graph sonifications, one should pick auditory dimensions other than loudness level to represent continuous quantitative variables.

4. THINGS WE NEED TO KNOW MORE ABOUT

4.1. Expertise and Extended Practice.

One limitation of most research, development, and usability testing of auditory graphs is that it is generally conducted using relatively inexperienced participants. While "successful" human interfaces are likely to be those which exhibit immediate ease of use, rather dramatic quantitative and well as qualitative changes in performance can occur with extensive practice, in both visual and auditory tasks. For example, few novice users of a software application would accept high speed compressed speech as an audio display option for any purpose. Nevertheless, blind users of products such as JAWS learn and soon *prefer* to use high speech rates for tasks such as spreadsheet navigation, that are incomprehensible to normally sighted observers. There is a good bit of anecdotal evidence concerning selective listening skills of symphony conductors who are able detect features that stem from a single instrument in a large orchestra. We thus cannot rule out that some of our concerns about display complexity, or optimal number of data streams could be considerably modified in situations where auditory graphs are used consistently for a given application by experts. It is also reasonable to expect that some features of auditory graphs, such a providing reference or context cues to calibrate scales and facilitate point estimation, might be useful for new users, but become distractions for experts who not longer need the context. The effect of listening expertise is an extremely important, but difficult-to-research topic, since it is difficult to recruit motivated participants for extended periods. However, this is issue shared with developers of sonification designs for industrial and medical monitoring, an area for which specific applications are more well-defined, and testing of motivated experts more pervasive. Researchers interested in auditory graphs development can benefit from communication with auditory display researchers working in those domains.

Because one can expect expertise dependent changes in perception to occur with auditory display use, it is important to envision display designs that allow users to have control of formats that allow the *discarding* of context features, in addition to having flexible control of speed, display duration, mappings, numbers of variables simultaneously displayed. Optimal formats will vary as a function of expertise as well as that nature of the task for which auditory graphs are being used. This should be the case for both displays used for assistive purposes, and those used by normally sighted users for specific data exploration tasks.

4.2. Effects of Stream Timbre and Patterning on Perceptual Grouping.

For many auditory graph applications, avoiding confusions between simultaneous data events and streams is important. However, there is little basic psychoacoustic research that directly relates to the attention and perceptual demands of listening to auditory mappings of data. Some of the concepts of basic perceptual organization and of Auditory Scene Analysis [7] are relevant to this issue, but more empirical research needs to be conducted. This is another issue that auditory graph development shares with displays for on-line process monitoring. A specific topic of interest is the relative importance of harmonic content, modulation or patterning, and onset/offset envelopes in keeping streams perceptually segregated.

4.3. Representing Multiple Variables in a Single Auditory Stream.

This is a strategy frequently and successfully applied in displays for process control monitoring, where it is often crucial to monitor two or more processes that are co-occurring in real time (e.g., pulse oximetry) [8] [9]. Depending on the type of task and information represented, integrating two (possibly more) sources of data into a single stream may be preferable to presenting them in different streams. For either auditory or visual graphs, where discovery or comprehension of data patterns, comparisons among variables, etc., are the primary issues, there tends to be less use of either multivariate visual objects, or mutltivariate auditory streams. In part, this comes from concern that integrated or configural objects (in sound or vision) may bias or distort the information in the original univariate data sources that are combined. However there are probably some very good uses for combining two variables on an auditory stream. One that has been already mentioned is rhythmic or amplitude modulation to carry time or "X-axis markings" in a stream for which pitch is representing a numeric measurement [1]. Continued research on this approach should be pursued to see the extent to which, for example, amplitude modulating more than one stream with the same time base information might convey graph positions and point estimation with greater fidelity. A second area worthy of exploration for which would seem, on the basis of preliminary research from our laboratory [10] concerns to portrayal of covariation between discrete event occurrences and continuous data (e.g. ocean temperature and storms, solar flux and geomagnetic storms). For such displays, representing the discrete events as a momentary loudness change and or momentary timbre shift within the pitch coded stream representing the continuous variable may provide greater ease of identifying the existence of a statistical association than plotting the continuous and discrete data in separate streams.

4.4. Flexible Tool Development for Sonification Research and Actual Data Exploration.

This has been restated frequently for the past several years by numerous investigators, so I will not elaborate on this ongoing need. However, I do think that in addition to the control of basic mappings and timing, we need to assure that tools allow flexible search and inspection of subsections of a sonification, ease in removing streams, and ease in combining sounds that represent continuous data with momentary sounds representing discrete events. We also need greater flexibility in adding features such as amplitude modulation or rhythmic patterning to pitch coded streams to explore or exploit some of the display principles mentioned previously.

5. CONCLUSIONS

In addition to giving a brief overview of "things that work, things that don't, and things we need to work on", based on my observation of sonification research specifically related to auditory graphs, I have reiterated my longstanding concern about overemphasis on the use of sonification to provide insight about structure of dense and complex datasets [11]. I do not wish to imply that sonification unlikely to be useful in complex data rendering, however. My concern has been more with trusting the processes of the auditory system to do things that it is not suited to do. If our data sonification tools permit the user to interact with a complex data to select subsets of variables, and to select mappings based on "things that work", and to change selections quickly and easily to gain multiple auditory viewpoints, I believe our ears can indeed help us gain insight into complex data.

6. REFERENCES

[1] Smith, D. R. & Walker, B. N. "Effects of auditory context cues and training on performance of a point estimation." *Applied Cognitive Psychology,* in press.

[2] Walker, B.N. "Magnitude estimation of conceptual data dimensions for use in sonification." *Journal of Experimental Psychology' Applied, 8,*

[3] Flowers, J. H., Buhman, D.C., and Turnage, K.D. "Cross-modal equivalence of visual and auditory scatterplots for exploring bivariate data samples." *Human Factors, 39,* 341-351, 1997

[4] Neuhoff, J.G., Kramer, G., & Wayand, J. "Pitch and loudness interact in auditory displays: Can the data get lost in the map?" *Journal of Experimental Psychology: Applied. 8 (1),* 17-25, 2002.

[5] Brown, L.M., Brewster, S.A., Ramloll, R., Burton, M. & Riedel, B. "Design guidelines for audio presentation of graphs and tables". *Proceedings of ICAD 2003 Workshop on Auditory Displays In Assistive Technologies* (University of Boston, MA).

[6] Flowers, J.H. & Grafel, D. C "Perception of sonified daily weather records." *Proceedings of the Human Factors and Ergonomic Society 46th Anuual Meeting,* 1579-1583, 2002.

[7] Bregman, A. *Auditory Scene Analysis.* Cambridge, MA: MIT Press, 1990.

[8] Anderson, J. E. "Sonification design for complex work domains: Steams, mappings and attention." Doctoral dissertation, University of Queensland, 2004.

[9] Watson, M. & Sandeson, S. "Sonification supports eyes-free respiratory monitoring and task time sharing." *Human Factors, 46,* 497-517, 2004

[10] McCormick, C. M. & Flowers, J. H. "Perceiving the relationship between discrete and continuous data: An evaluation of auditory and visual display formats." Manuscript under review.

[11] Flowers, J. H., Buhman, D.C., and Turnage, K.D. "Data sonfication from the desktop: Should sound be part of standard data analysis software?" *Proceedings of ICAD 1996.*