6-2007

Perceiving the relationship between discrete and continuous data: A comparison of sonified data display formats

Carina M. McCormick
University of Nebraska-Lincoln, cmccormick@buros.org

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

Perceiving the relationship between discrete and continuous data: A comparison of sonified data display formats

Carina M. McCormick & John H. Flowers

University of Nebraska-Lincoln
Departments of Educational Psychology
And Psychology
Lincoln, Nebraska
carina@bigred.unl.edu
jflowers1@unl.edu

ABSTRACT

This study compared the effectiveness of two auditory display designs for conveying the relationship between discrete and continuous data. Participants judged the relationship between simulated data representing "sea temperature," (a continuous variable) and "storm occurrence" (a categorical variable) by rating the strength of covariation between these variables and qualitatively describing the relationship for one of two types of auditory displays. One format integrated the representation of storms and sea temperature into a single pitch-varying "stream" by signaling storms occurrence by momentary amplitude and timbre changes. The other format presented the storm occurrence information as atonal percussive events separate from the pitch-varying stream that represented temperature. While both formats led to statistically equivalent proportions of verbal descriptions of the temperature-storm relationships present in the simulated data samples, the integrated display produced higher correlations between ratings of the strength of the temperature-storm relationship and the actual storm-temperature covariation present within each data sample.

[Keywords: Sonification, perception of covariation, auditory graphs, multivariate data display]

1. INTRODUCTION

Advancements in technology have provided researchers with a wide variety of complex multivariate data samples, ranging from astronomical measurements to time-series medical information. Such data samples often include combinations of measurements of continuous numeric quantities and discrete, categorical "events". Important scientific discoveries may depend upon research techniques that lead to the detection and eventual description of pattern of covariation among such measurements.

Both visualization procedures and emerging sonification techniques provide potentially useful tools for displaying multivariate data. However, relatively few studies have specifically examined the use either display modality to represent the combination of continuous numeric data and categorical events. In our study, subjects judged the relationship between simulated time series data representing "sea temperature" and "storm occurrence" in sonified data displays. This investigation relates previous research regarding perceptual principles with the goal of realizing novel and effective display design.

1.1. Perception of Covariation and Randomness

A number of experimental studies have demonstrated that people have systematic problems with judging covariation, i.e. when two variables' values change together [definition from 1]. People often make errors in identifying structure, especially in covariation judgment [review in 2] and randomness [reviews in 3 and 4]. When judging relationships between either co-occurring or sequential values, people act as if the correlation is stronger than it actually is and detect more structure in positively related values than in negatively related values [1]. Research has repeatedly shown that people judge strings of binary data with high repetition rates as nonrandom [5]. Stimuli with an alternation rate of about .6 -- surpassing the truly statistically random .5 alteration rate -- are perceived as most random [6]. However, researchers do not agree on why people have such problems with judging structure and randomness.

Memory constraints do impose some control over judgments of structure and randomness. Researchers have found support for the idea that people base judgments of randomness on the difficulty of mentally encoding the sequence [6], and memory demands detrimentally affect people's ability to accurately judge covariation [7]. Given that people make so many errors in judging covarying relationships and even in distinguishing between random and dependent or covarying events, it is necessary to present information in ways that minimize these errors. To the extent that factors, such as limitations of selective attention and working memory affect people's ability to assess covariation, the design of display formats that reduce these limitations is important goal.

1.2. Auditory Display and Sonification

Sonification has a wide variety of applications for data exploration, analysis, and presentation. Under certain circumstances, especially when visual attention is required elsewhere, presenting information in an auditory display may be an efficient solution. Auditory representation of data is as pervasive as Morse Code, sonar, Geiger counters, and as displays in airplanes, but also as new and innovative as replacement for tactile feedback in spacesuits, financial analysis, monitoring seismic data, and geographic mapping [8], as well as complex multivariate medical monitoring [9]. Judgments of visual and

auditory scatterplots have produced statistically equivalent results, though the auditory condition had more variability [10].

Variations among auditory displays may make certain elements of the data seem more salient. Consequently, researchers must determine the best design to properly convey information with sound. In a study examining the use of sonification for exploratory data analyses, subjects who judged differences between the auditory version of bar-whisker charts included consideration of the median more than those viewing traditional visual bar-whisker charts, who predominately relied on skew [11]. In addition, certain components of the auditory display itself can affect judgment, with subjects rating structure of auditory displays more accurately when the presentation was faster [12]. In other sonification uses such as medical monitoring, researchers must be aware of the way users perceive the data displayed and create sonification techniques to maximize judgment accuracy [13].

In auditory displays such as those used in the current study, pitch changes work well for demonstrating changes in the quantity represented by the data (e.g. temperature) across time because people can easily learn to recognize the qualities of the data with this type of graph [Mansur as cited in 8]. Kramer [14] called using higher pitch to represent more of a quantity a logical metaphor.

Although most of the literature discussed does not specifically address the types of complex data studied in this experiment, judgments of multivariate covariation in data displays could be affected by the same biases as in previous studies. In the current study, a continuous variable predicted a discrete event. Display technology for this type of data is not standardized. If researchers can better understand people's perceptions of structure in combinations of continuous and categorical data, they can learn to present the data through sonification in ways that better convey their statistical properties.

1.3. The Current Study

The present experiment compared a display design that uses an integrated display approach (in which indication of storm occurrence is embedded in the data stream depicting temperature) with a design that represents storm and temperature in separate streams. The researchers hypothesize that the condition that integrates the pattern for the discrete event will be more accurately evaluated. The use of integrated rather than separate streams for continuous and discrete event data (particularly for the purpose of discovering the relationship between the streams) is consistent with a visualization principle offered by Tufte [15], recommending the use of multifunctional graphical elements. The use of integrated displays in sonified data also follows Kramer's [14] endorsement of using redundant display characteristics in a single stream to offset attention being divided between streams or drawn to a new stream when an event occurs. In the present experiment, one condition integrates the representation of storms with the temperature when the storm occurred, creating a complimenting pattern and emphasizing the temperature when storms occurred.

Gestalt psychology studies human's propensity for recognizing patterns for recognizing patterns, using the principle that 'the whole is greater than the sum of its parts.' While much research on this research has historically investigated visual perception, the same principles can be applied to perception of sound as well [16]. The integrated method utilizes the Gestalt principles of similarity and belongingness, which are characteristics of effective display design [16, 17], and should yield better performance.

In the second condition, a woodblock sound (played at the same pitch every time) represented storms. Because the listener must attend to both channels simultaneously to hear the relationship, this may be more difficult than the condition in which information about temperature is embedded in the representation of storms. For the integrated design, storms were indicated by embedding an accented guitar note into a synthetic string stream, in which pitch represented temperature, while the separate stream design used a woodblock strike to indicate storms.

2. METHOD

2.1. Participants

Forty-six university students participated in the experiment and received credit for a research option of their course work. All subjects reported normal hearing and normal or corrected vision. Fifteen subjects were excluded from analysis because their answers suggested they did not understand the directions, either because the inter-item correlations of their ratings compared to ratings from all subjects was extremely low or negative or because their descriptions of the relationship did not describe what was happening in regard to temperature when storms occurred. Thus, for the purpose of analysis, there were 31 subjects in the experiment, with 16 subjects in the first (Guitar) condition and 15 subjects in the second (Woodblock) condition. The differences between these conditions are explained in detail in the remainder of the method section.

2.2. Data Simulation

The samples of simulated data (representing continuous measurements of 'sea temperature' and the discrete occurrence of 'storms') were constructed using Microsoft Excel spreadsheets. This first step was to create several streams of smoothly rising and falling values that could reasonably be construed as fluctuations in sea temperature. This was accomplished by addition sets of arbitrarily selected harmonically unrelated sine waves to produce four prototype patterns from which segments could be "cut" from arbitrarily selected starting points. These segments each consisted of 149 values (i.e. 'temperature' points).

This segment was then pasted into a column of a spreadsheet that would create a corresponding column of binary data (zeros and ones), where 'one' indicated the presence of a storm. The determination of where in that vector storms (ones) were placed was done by using Excel's random number function, according to one of several algorithms that simulated a particular relationship (statistical correspondence) between sea temperature and storms. These relationships included (1) storms were more likely when sea temperature was increasing, (2) storms were more likely when sea temperature was decreasing, (3) storms were more likely any time the sea temperature was above its median value, and (4) storms were more likely any time the sea temperature was below its median value.

Each of the four relationships varied in "strength" of covariation from deterministic (e.g., storms *only* occurred when sea temperature was rising, with 50% of the observations when

sea temperature was rising), "strong" (e.g., storms occurred with 40% of the sea temperature observations when sea temperature was rising, and with a random 10% of all observations) and "weak" (e.g., a 25%-15% probability split). In addition to these four generation rules that produced simulated data in which there was a relationship between sea temperature and storm probability, a fifth generation rule produced generated storms with equal probability throughout the sequence (i.e., independently of sea temperature).

Samples drawn from these various constructed patterns produced the 34 sets of data from which the MIDI files were constructed, in addition to the MIDI files used for the instructions, which were created the same way. However, it should be noted that these probabilities only provided parameters for the pseudorandom number generator in Excel to produce strings of one's and zero's. Thus, the actual proportions of storms, in "rule consistent" and "rule inconsistent" segments of the simulated data differed somewhat from the proportions specified to the stimulus generation spreadsheet.

Ten different data samples depicted relationships in which temperature change (slope) was associated with storm occurrence. For four of these, the relationship was totally predictive; storms only occurred in segments of decreasing temperature for two samples, and only in segments of increasing temperature for the other two. For the remaining six samples illustrating a relationship between temperature change (slope) and storms, the proportion of storm occurrences that were consistent with the temperature slope trend were 0.88, 0.86, 0.83, 0.76, 0.73 and 0.73. Twelve data samples illustrated temperature level dependent relationships with storms. Three of these were deterministic, storms only occurred above the median temperature for two of these, and only below the median for the other. The proportions of storms occurring in the above or below median temperature were 0.83, 0.81, 0.77, 0.77, 0.76, 0.75, 0.71, 0.69 and 0.65.

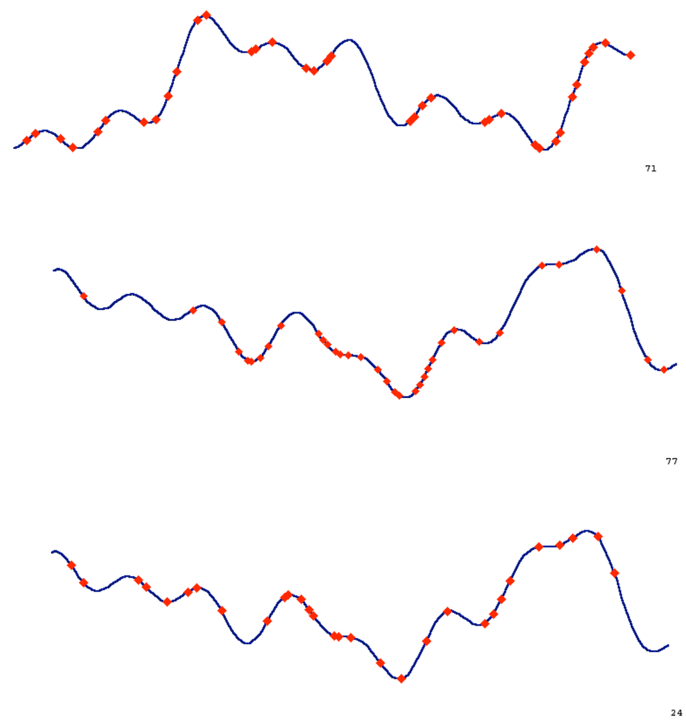

Figure 1. *An analogous visual example of stimuli. In these examples, storms occur (from top): more often when temperature is increasing, more often when temperature is below the median temperature, and randomly. To compare to MIDI files used in the study, see http://www-class.unl.edu/psyc463/*

2.3. Display Construction

In the *integrated* (Guitar) condition, the display embedded the storm and temperature information into a single pitch-varying stream, while in the *separate stream* (Woodblock) condition a separate atonal percussion stream specified the temporal occurrence of storms. Musical Generator software was used to create 17-second duration audio files within which pitch (representing "temperature") varied within a three-octave range, from c4 to c7. For example MIDI files of both display methods, see http://www-class.unl.edu/psyc463/

In the integrated condition, one MIDI channel played the temperature continuously with synthetic strings, and a second channel played the sound of an accented guitar at the same pitch as temperature briefly when a 'storm' occurred. This grouping should lead to the perceptual effect of the two channels becoming part of the same auditory stream, in which momentary timbre changes and loudness increases indicated storm occurrence.

In the separate stream condition, temperature was mapped to synthetic strings as in the integrated condition, but instead of the guitar sound, a MIDI constant-pitch woodblock played every time a storm occurred. This combination was intended to lead to the perceptual effect of separate streams – one varying in pitch, the other represented by a rhythmic percussion line. Each stimulus sequence was saved as a separate MIDI file.

2.4. Procedure

During a single experimental session lasting approximately 45 minutes, each subject examined 34 MIDI files using a personal computer by listening to each file through headphones using Microsoft PowerPoint. The informed consent form and instructions stated that the data were simulated. Instructions asked the participants to judge the relationship between sea temperature and storm occurrence based only on what they heard and not on any previous knowledge of weather. Subjects listened to one example to become acquainted with what elements of the chart represented. They then examined three examples of varying relationship strength and type and then read what the relationship was for each example stimulus in order to practice.

In the data collection component, subjects indicated the perceived strength of the relationship on a 1-to-10 scale (where '1' represented "no relationship" and '10' represented "extremely strong") after listening to each stimulus and described the relationship displayed in a free response question. These subjects wrote their answers on paper forms and were able to listen to each stimulus as many times as they wanted. Three different random presentation orders were used for each of the two display conditions. Subjects were randomly assigned to one of the two display conditions and were then assigned one of the three presentation order sets.

3. RESULTS

Analyses will consider differences between the two conditions by evaluating those stimuli in which the frequency of storms was dependent on temperature increasing or decreasing (referred to as 'slope-dependent') and in which it was dependent on the temperature relative to the median temperature ('level-dependent'). In all of the comparisons, the integrated (Guitar) condition performed at least nominally better than the separate steam (Woodblock) condition.

3.1. Perception of relationship strength

First, for the slope-dependent displays in the Guitar condition, the correlation between strength of the relationship in the display and subjects' mean rating of the display was $r(13) = .87$, $p < .001$, whereas the Woodblock condition produced a correlation of $r(13) = .82$, $p < .001$. (See Figure 2.) Analyses also considered judgments without the stimuli in which storms *only* occurred according to the rule (i.e. those stimuli of which the relationship strength was 1.0, or the far right of the scatterplots). For this calculation, the correlations were weaker, with the Guitar stimulus remaining significant and the Woodblock condition approaching the designated alpha level of .05: Guitar, $r(9) = .72$, $p = .012$; Woodblock, $r(9) = .61$, $p = .048$. Here, the integrated condition clearly outperformed the integrated condition.

The correlation between actual display strength and subjects' strengths were also significant for the level-dependent displays in both the Guitar condition, $r(13) = .93$, $p < .001$, and the Woodblock condition $r(13) = .82$, $p < .001$. (See Figure 1.) Without the strongest stimuli, the correlations were again weaker, with the Guitar stimulus remaining significant and the Woodblock condition just reaching the designated alpha level of .05, Guitar $r(9) = 0.78$, $p = .003$; Woodblock, $r(9) = .58$, $p = .05$.

Because these figures are essentially correlations of means, traditional tests of statistically significant differences between correlations are not appropriate. However, all results showed that the integrated (Guitar) display format led to higher correlations between perceived strength and actual correlations between storm dependence and temperature change or level.

3.2. Accuracy of relationship descriptions

Inspection of verbal descriptions of the temperature storm relationships revealed that, overall, participants provided "correct" labels for the statistical storm-temperature relationship in the displays at nearly identical rates in each display format. For the temperature change dependent samples the proportion of participants providing accurate descriptions were .56 and .57 for the integrated and separate stream formats, respectively. For the temperature level dependent formats these values were .49 and .43.

However, for *all* seven of the data samples depicting a deterministic (no exceptions) relationship, the integrated display produced a greater proportion of correct responses – mean proportion of correct relationship identification responses for these stimuli were .80 for the integrated and .69 for the separate stream displays.

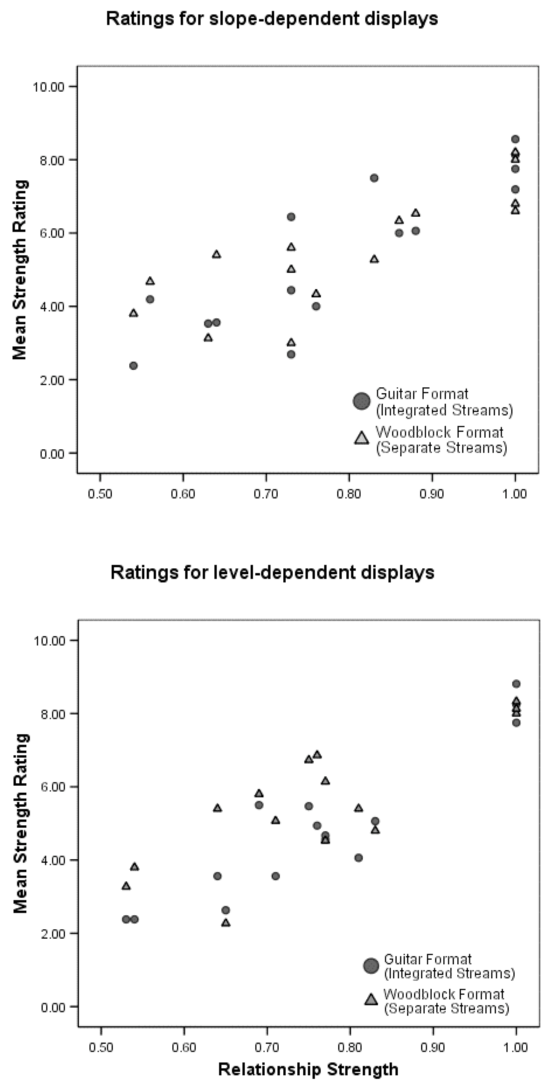

Figure 1. *Rated versus actual dependency between temperature change and storm occurrence for the slope-dependent displays (top panel) and the level-dependent displays (bottom panel).*

4. DISCUSSION

The purpose of this study was to evaluate different auditory display formats for combinations of discrete and continuous data. To do this, subjects utilized one of two auditory data display formats and made judgments of the relationships displayed. The first method (Guitar) integrated the discrete and continuous data more than the second method (Woodblock). Both display styles in the current study were effective for conveying the strength of the relationship; still, researchers need to consider changes such as method of presentation and amount of information integration when designing displays for various purposes.

To understand people's ability to perceive structure in different displays, researchers should consider both how well their ratings of strength corresponded with actual strength and

how well they were able to detect the strongest relationship displayed in each stimulus. Measures of perceived strength without reference to actual strength are not appropriate because effective displays allow the viewer or listener to understand quantitative differences in strength. Thus, a high rating for a display with only a weak relationship is undesirable.

For both display formats, perceived strength of the relationship was positively correlated with the actual strength of the relationship. These correlations were stronger in the integrated condition than in the separate stream condition. When the stimuli with the strongest relationships were removed, the correlations were weaker but still positive. This is meaningful because when using real-life data, absolute relationships are unlikely. For instance, it sometimes rains on even sunny days, though storms are associated with cloudy days.

In addition to judging the strength of the relationship, people should be able to describe the appropriate relationship if one exists. In this regard, the two conditions produced statistically equivalent results. Many subjects demonstrated difficulty in perceiving the correct relationship when the relationship was fairly weak. This dilemma further emphasizes the need for improved methods of auditory display when conveying complex data. For the purposes of experiments, it may be more effective for participants to select the relationship they perceive from a listing of possible relationships, instead of writing an open-ended response. In the present experiment, participants often described relationships that could not easily be classified according to the guidelines set in advance.

Subjects may be able to improve their judgments when using auditory displays if they practiced specifically or if such displays were generally more prevalent. Subjects can improve performance when given feedback on their judgments of covariation when using visual displays [5]. Additionally, training improved subjects' interpretations of sonified financial data [18]. Researchers could perform a study similar to this one in which participants would receive feedback after giving answers or undergo training before making judgments of structure. Such studies would help determine the role of experience in the perception of relationships in multivariate auditory data displays.

Another significant constraint in judgments based on auditory displays is reliance upon memory for decisions. It could be argued that the integrated display condition lessened the strain on working memory because it was easier for subjects to simultaneously attend to both temperature and storm information in a single pitch stream. Conversely, the separate stream condition required that attention be shared between two perceptually separated streams, without the second stream adding information about the temperature itself. An investigation of a related design [18] found that, without training, adding tones which provided dynamic y-axis context improved judgments of sonified data better than a static reference tone did.

When subjects view visual data displays, they are able to direct attention at will and divert attention when appropriate. Auditory data display should replicate as many benefits from visual data display as possible. Errors caused by the difficulty of mentally encoding reveal necessary procedural guidelines for auditory display design. Subjects must be able to replay the stimulus as many times as desired, and the subjects should be able to control timing and frequency of playback, as they were in this study. They must be able to listen repeatedly in order to encode the data and make judgments, but they must also be able to turn the sound off in order to attend to what they have encoded rather than what is playing.

Considering the challenges of unfamiliarity with auditory display that participants faced, they performed well in rating the structure and in describing the strongest relationships. More research should examine effective uses of sonification for displays of complex data. This study showed sonification to be a promising method to display combinations of discrete and continuous data.

5. REFERENCES

[1] Kareev, Y. (1995) Positive bias in the perception of covariation. *Psychological Review, 102(3),* 490-502.

[2] Shaklee, H. (1983) Human covariation judgment: accuracy and strategy. *Learning and Motivation, 14*, 433-448.

[3] Nickerson, R. (2002) The production and perception of randomness. *Psychological Review, 109,* 330-357.

[4] Bar-Hillel, M., & Wagenaar, W. (1993) The perception of randomness. In G. Keren & C. Lewis (Eds.), *A handbook for data analysis in the behavioral sciences: methodological issues.* Hillsdale, NJ: Earlbaum, *pp. 369-393*.

[5] Lopes, L., & Oden, G. (1987) Distinguishing between random and nonrandom events. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 13,* 392-400.

[6] Falk, R., & Konold, C. (1997) Making sense of randomness: implicit encoding as a basis for judgment. *Psychological Review, 104(2)*, 301-318.

[7] Shaklee, H., & Mims, M. (1982) Sources of error in judging event covariations: effects of memory demands. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 8(3)*, 208-224.

[8] Kramer, G. (1994 a) An introduction to auditory display. In G. Kramer (Ed.), *Auditory display*. Reading, MA: Addison-Wesley, *pp*. 1-77.

[9] Anderson, J. (2004) Sonification design for complex work domains: Streams, mappings and attention. Unpublished PhD thesis, School of Psychology, The University of Queensland, St. Lucia, Qld.

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

[11] Flowers, J., & Hauer, T. (1993) "Sound" alternatives to visual graphics for exploratory data analysis. *Behavior Research Methods, Instruments, & Computers, 25(2),* 242-249.

[12] Flowers, J., Whitewer, L., Kotan, C., & Fairchild, K. (2001) Perceiving sequential constraint in simulated categorical data streams: Influence of display format. Paper presented at the meeting of the Midwestern Psychological Association, Chicago, IL, May.

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

[14] Kramer, G. (1994 b) Some organizing principles for representing data with sound. In G. Kramer (Ed.), *Auditory display*. Reading, MA: Addison-Wesley, *pp*. 185-221.

[15] Tufte, E. (1990) Envisioning Information. Cheshire, CN: Graphics Press.

[16] Williams, Sheila. (1994). Perceptual principles in sound Grouping. In In G. Kramer (Ed.), *Auditory display*. Reading, MA: Addison-Wesley, *pp*. 95-125.

[17] Colet, E., & Aaronson, D. (1995) Visualization of multivariate data: human-factors considerations. *Behavior Research Methods, Instruments, & Computers, 27(2)*, 257-263.

[18] Smith, D., & Walker, B. (2005) Effects of auditory context cues and training on performance of a point estimation sonification task. *Applied Cognitive Psychology, 19,* 1065-1087.