

# “Sound” alternatives to visual graphics for exploratory data analysis

JOHN H. FLOWERS and TERRY A. HAUER  
*University of Nebraska, Lincoln, Nebraska*

Efficient exploratory data analysis (EDA) may be aided by succinct, but informative, graphical representations (e.g., Tukey plots) that convey information about central tendency, variability, and shape of distributions, and that permit detection of outliers. Using research strategies adapted from studies of cross-modal perceptual equivalence, we show how auditory analogies of such displays may offer an effective alternative to visual plots for EDA.

A casual inspection of current advertisements and marketing strategies for both desktop computers and workstations will show that “multimedia capabilities,” which include interfaces and output devices for auditory signals, have become a widely promoted feature. For perceptual psychologists, who have studied the constraints on human ability to process multiple sources of information, the dazzle of such innovations as dynamic multimodal presentation graphics is tempered by some concern that such developments may be taking place without adequate evaluation of whether such display techniques are optimally matched to the architecture of human attention. If development efforts are directed exclusively toward increasing the “data bandwidth” of the human user, human-factors history suggests that neglect of the constraints of attention may cause some developments to fall short of expectations. On the other hand, the multimodal flexibility of modern computer design holds potential for discovering innovative and effective *alternatives* to traditional visual graphics and displays, suited to particular tasks and/or categories of users, such as the visually impaired (e.g., Lunney & Morrison, 1990). Our present research investigates one potential application of an alternative mode for data display in a task of importance to researchers in the social and behavioral sciences. That task is assessing the distribution characteristics of numeric data samples during exploratory data analysis.

Previous research from our laboratory (Flowers & Hauer, 1992) demonstrated that information about statistical parameters of a univariate numeric data sample (central tendency, variability, and shape) could be conveyed by an auditory graphical scheme that we called an *auditory histogram*. These displays consisted of strings of musical notes in which numeric values of observations (the

midpoint of a class interval or bin) was represented by musical pitch, and the frequency of scores in each interval or bin was represented by the number of repetitions of a given note in the display. Subjects made paired comparisons of such auditory histograms, producing matrices of dissimilarity judgments that were highly similar to those produced by comparisons of visual histograms of the same numeric distributions. The dissimilarity judgments for distribution pairs of auditory and visual displays were predicted by very similar multiple-regression models in which central tendency, standard deviation, skewness, and kurtosis were used as predictors. Multidimensional scaling (MDS) solutions obtained from auditory and visual histogram judgments led to very similar configurations (each of which adequately depicted the differences in the stimulus parameters). Our present research extends that initial investigation by addressing the following issues: (1) can *combined* auditory and visual presentation enhance sensitivity to differences in stimulus parameters, and (2) can *simplified* depictions of distributions (auditory analogies of Tukey box-whisker plots) provide adequate information about distributional properties?

## EXPERIMENT 1

### Visual, Auditory, and Bimodal Histograms

#### Method

**Subjects.** Twenty-five advanced undergraduate psychology majors, all of whom had completed a course in experimental design and statistics, served in two 45-min sessions. Each subject received \$10 for participating. All subjects had normal vision and hearing.

**Stimulus materials.** Nine samples of 50 numbers each were computer generated to produce simulated data samples that differed in central tendency and shape. These nine distributions, for which the parameters are shown in Table 1, consisted of three groups of three distributions each, for which distribution shape was the defining characteristic. Stimuli 1-3 were essentially normal in shape, Stimuli 4-6 were positively skewed, and Stimuli 7-9 were negatively skewed. Within each of these three shape categories, the three distributions varied in central tendency, having means of approximately 44, 50, and 55. The variability of the distributions remained relatively constant across distributions, with standard deviations ranging from 8.1 to 9.4.

---

This research was supported in part by a grant from the Stanton Memorial Fund from the University of Nebraska Foundation and by the Layman Fund from the University of Nebraska Foundation. Experiment 1 is based on a Senior Honors Thesis submitted to the University of Nebraska-Lincoln, College of Arts and Sciences by the second author. Correspondence should be addressed to J. H. Flowers, Department of Psychology, 209 Burnett, University of Nebraska, Lincoln, NE 68588-0308 (e-mail: jflowers@unl.edu).

**Table 1**  
**Statistical Parameters of Distribution Stimuli for Experiments 1 and 2**

Stimulus	Shape	M	SD	Kurtosis	Skewness	Range
1	Normal	44.3	8.3	.06	.12	39
2	Normal	51.2	9.0	-.47	-.09	38
3	Normal	54.4	8.1	.01	.17	39
4	Pos. Skewed	43.5	8.8	2.09	1.55	40
5	Pos. Skewed	49.8	8.9	.71	1.23	36
6	Pos. Skewed	54.7	8.3	.30	1.03	35
7	Neg. Skewed	45.1	9.3	.68	-1.22	39
8	Neg. Skewed	51.1	8.9	.72	-1.20	37
9	Neg. Skewed	55.3	9.4	.25	-1.10	38

These nine numeric distributions served as the basis for generating three different types of data displays. *Visual frequency polygons* consisted of projected 35-mm color slides of plots containing pairs of distributions that were generated by Quattro Pro (Borland Software). Figure 1 displays an example of a frequency polygon stimulus, showing a "normal" distribution having a high level of central tendency simultaneously with a negatively skewed distribution having a medium-level central tendency (Distributions 3 and 8 from Table 1). On each trial, the stimulus pair was projected for 35 sec (the approximate time required to present the auditory histograms described below), following which the subject made a judgment of dissimilarity.

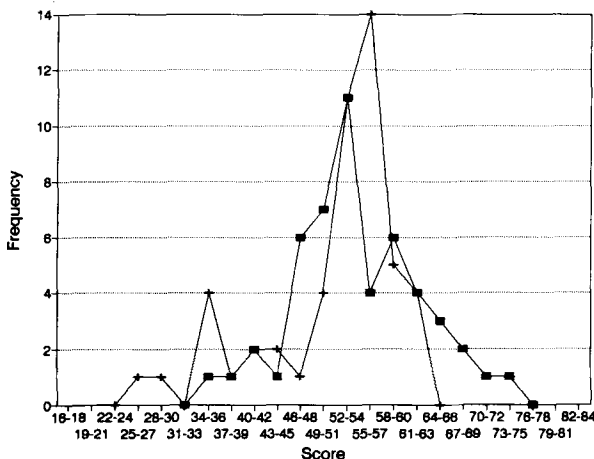
*Auditory histograms* consisted of mapping each of the numeric values in a distribution in numeric bins and assigning a musical pitch value to each bin. The number and "size" of the bins for each distribution were the same as those used in plotting the visual frequency polygons. The bins and pitch assignments for the auditory histograms are shown in Table 2.

The presentation of an auditory histogram consisted of a string of 50 musical notes, lasting a total of 5.2 sec, in which the pitch representing the lowest bin was presented first, followed by the pitch of the next higher bin, and so on. Frequency of values in each bin was represented by the number of successive repetitions of a pitch. Thus, to represent the normal high central tendency distribution (Distribution 3), an auditory histogram would consist of a single C#, followed by one D, two D's, one E, six F's, and so on. As with the visual frequency polygons, our task required subjects to

compare pairs of these histograms on each trial. Since ample evidence exists that human observers cannot effectively attend to simultaneous presentation of two musical passages, pairs were presented successively, spaced by a 0.83-sec pause. On each trial, the subjects were given three presentations of a pair prior to making a dissimilarity judgment. There was a 1.67-sec pause between each pair presentation within a trial. Stimulus presentation on each trial thus lasted about 35 sec.

*Combined auditory and visual* displays involved the simultaneous visual display of a distribution pair, along with three presentations of the same pair as auditory histograms. For these displays, the subjects were instructed to attend to both the visual graph and the auditory histogram in order to make their assessment of dissimilarity between the distributions.

**Procedure.** The subjects were tested in groups of 5. They were seated in a dimly lit laboratory room, approximately 2.9 M from a projection screen. At this distance, the visual angle of the slide displays that contained the visual frequency polygons was approximately 19°. The auditory histogram was presented through the speaker of a Gateway 386 IBM-compatible computer located on a table to the right of the projection screen. Each subject was given a booklet containing three answer sheets (one each for the visual, auditory, and combined tasks) upon which they recorded dissimilarity judgments. The subjects were instructed to judge the overall dissimilarity between the two distributions in each pair, taking into account differences in central tendency, variability, and shape, using a 10-point scale, for which 1 indicated *extremely similar* and 10 indicated *extremely dissimilar*. The subjects were given examples of visual frequency polygons and auditory histograms to study, together with detailed instructions about how the auditory histograms and visual frequency polygons were constructed. Before beginning data collection in the first of the two sessions, the subjects were given three practice blocks of 10 trials each (one block for each of the three presentation modalities) in order to familiarize them with the task and the range of similarities and differences among the stimuli. Following these practice trials, the subjects participated in three blocks of 36 trials each—one block for each presentation modality, within which each of the nine distributions was paired with each of the other distributions once. The second experimental session for each group of subjects was identical to the first session, except that no instructions or practice blocks were administered. Thus, each subject produced a total of two dissimilarity judgments for each possible distribution pairing for each of the three distribution modalities. Order of presentation modality was determined by block randomization for each group of subjects and each session.



**Figure 1.** Example of a frequency polygon display, comparing a "normal" sample with a "high" central tendency (Stimulus 3 = square symbols) with a negatively skewed sample with a "moderate" central tendency (Stimulus 8 = + symbols). As shown in color to subjects, Stimulus 3 was plotted in green, and Stimulus 8 was plotted in red.

**Results and Discussion**

**Reliability of judgments.** Since the subjects judged each distribution pair twice for each modality (once in each of the two sessions), we computed the correlation between judgments for each of the 36 pairs for each subject as a measure of judgment reliability across presenta-

Table 2  
The Actual Assignment of Musical Pitches to Numeric Values  
for Each of the Nine Distributions for Experiments 1 and 2

Note	Interval	Norm. Low	Norm. Med.	Norm. High	Pos. Low	Pos. Med.	Pos. High	Neg. Low	Neg. Med.	Neg. High
O2G#	19-21							1		
O2A	22-24							1		
O2A#	25-27	1						2	1	
O2B	28-30	1						2	1	1
O3C	31-33	2	1		1			1	0	0
O3C#	34-36	4	3	1	6			3	4	7
O3D	37-39	6	0	1	14	2		0	1	4
O3D#	40-42	7	4	2	11	6		3	2	0
O3E	43-45	16	7	1	6	13	3	6	2	4
O3F	46-48	3	3	6	2	10	8	5	1	2
O3F#	49-51	6	6	7	1	5	13	13	4	1
O3G	52-54	6	9	11	2	2	6	9	13	2
O3G#	55-57	1	4	4	2	1	5	3	14	5
O3A	58-60	1	5	6	2	3	4	1	5	13
O3A#	61-63	1	3	4	1	3	3		4	12
O3B	64-66	1	3	3	1	2	1			4
O4C	67-69		1	2		1	4			2
O4C#	70-72		1	1		0	1			
O4D	73-75			1		1	1			
O4D#	76-78						1			

tions. For visual presentation, the mean Pearson correlation coefficient between the sessions was .74 (range .41 to .84). For auditory presentation, the correlation was .42 (range -.23 to .73); for combined presentation, the mean correlation was .66 (range .24 to .88). In terms of between-session reliability of the dissimilarity judgments, visual frequency polygons outperformed both auditory histograms [ $t(24) = 6.57, p < .01$ ] and combined presentation [ $t(24) = 2.19, p < .05$ ]. The better reliabilities of judgments for the visual displays is consistent with the findings of Flowers and Hauer (1992) and is not particularly surprising given the subjects' far greater familiarity with frequency polygon graphs than with auditory displays of any variety. More important, however, is that these data offer no evidence that combined presentation led to greater consistency of judgment, since the *addition* of the auditory displays led to *less* consistency in judgments across sessions than did visual presentation alone.

It should be noted, however, that despite the lower overall between-session reliability of the dissimilarity judgments for the auditory histograms, considerable individual differences in reliability existed. One subject produced greater reliability with auditory presentation than with either of the other modalities (.73 for auditory vs. .63 for visual and .71 combined). For 2 subjects, on the other hand, small negative correlations existed in the judgments between sessions for auditory presentation (-.23 and -.14). The extent to which these large individual differences may reflect different experiences with auditory tasks in general (such as musical experience), the prevalence of more pervasive differences in frequency discrimination than in visual shape discrimination, or simply differences in ability to learn a novel symbolic task involving quantitative relationships cannot be determined. Individual differences do, however, raise an issue of concern for future applications of auditory display technology.

Inspection of the dissimilarity matrices obtained for each session did not, however, reveal any systematic changes in the ratings of particular stimulus pairs. Thus, the remainder of our analyses, which focus on the *perceptual structure* of the displays and its relationship to the statistical parameters of the distributions, was based upon the average of the dissimilarity judgments for the two sessions collapsed across subjects.

**Regression analysis.** Since the individual distributions in the stimulus set were constructed to differ in the statistical features of central tendency and skew, we performed regression analyses to see how well differences in these two parameters of the numeric distributions predicted the matrix of dissimilarity judgments for each presentation mode. Thus, for each of the 36 pairings of stimulus distributions, we computed the absolute values of differences in the mean and differences in the skew, and we used those values as predictors with the mean dissimilarity rating for each stimulus pair as the dependent variable in a regression model.

For visual presentation, standardized regression (beta) coefficients were .17 for mean and .84 for skew, with a multiple correlation ( $R$ ) of .84. For auditory presentation, the corresponding values were .14 for mean, .88 for skew, and  $R = .87$ . For combined presentation, the values were .12 for mean, .83 for skew, and  $R = .82$ . It should be noted that the average dissimilarity matrix across subjects was predicted about equally well for all three display types, with the percentage of variance accounted for ranging from about 67% for combined presentation to 76% for auditory presentation. It should also be noted that differences in skew accounted almost exclusively for those judgments. In fact the contribution of the mean was only marginally significant for visual and auditory presentation ( $ps = .08$  and  $.11$ ) and nonsignificant for combined presentation ( $p > .20$ ). Given that sen-

sitivity to central tendency differences was found to be more salient than was sensitivity to differences in shape in our previous research using auditory and visual histogram displays (Flowers & Hauer, 1992), we attribute this dominance of shape differences to the selection of our stimuli (which included symmetrical, positively skewed, and negatively skewed distributions in the same set).

More important, however, is that these data, like the reliability data, do not show any advantage for joint auditory and visual presentation. However, the strongest evidence indicating that combining auditory and visual graphics is not a useful technique for conveying information about statistical parameters of data samples came from subjective comments of the subjects. When informally asked which task they considered to be the most difficult, nearly every subject found the combined presentation to require the most effort, followed by the auditory mode, whereas the more familiar visual format required the least effort. Several subjects stated that they found attempting to divide attention between viewing the polygon displays and listening to the histograms to be virtually impossible, so they adopted the strategy of concentrating attention on the visual display alone. Several subjects also stated that the extra effort required in the bimodal task caused them to focus on the single most salient dimension (shape) and virtually abandon attempts to evaluate differences in central tendency and variability.

**Multidimensional scaling solutions.** With the overwhelming perceptual dominance of the differences in skew in determining the dissimilarity ratings revealed by the regression analysis, one might expect an MDS configuration produced by the dissimilarity matrix to collapse to a one-dimensional solution. Stress values for one-dimensional ALSCAL solutions were in excess of .150 for each of the three modalities; therefore, two-dimensional solutions were obtained (see Figures 2-4).

While considerable caution is warranted in interpreting these solutions given the regression data and small num-

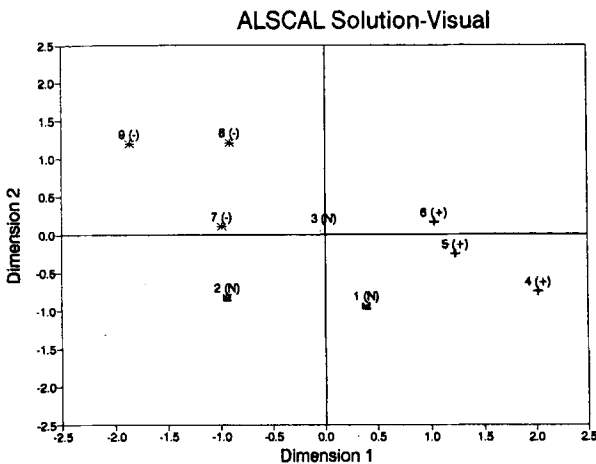


Figure 2. ALSCAL MDS solution for visual frequency polygon comparisons in Experiment 1.

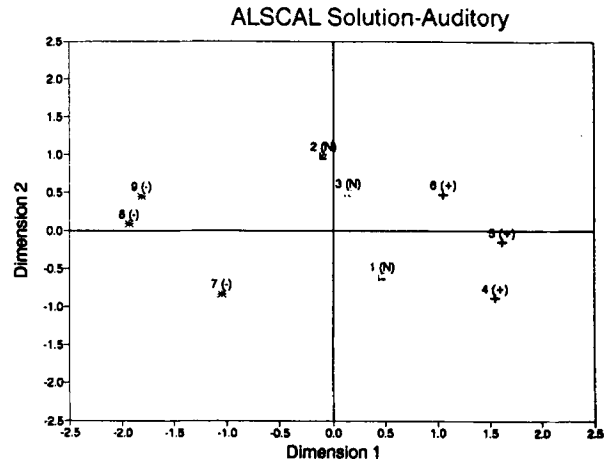


Figure 3. ALSCAL MDS solution for auditory histogram comparisons in Experiment 1.

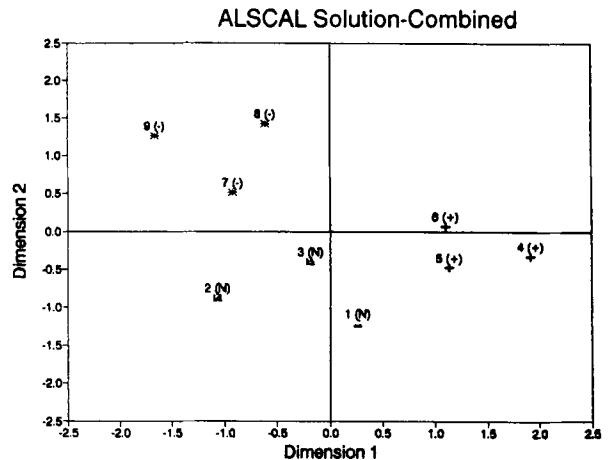


Figure 4. ALSCAL MDS solution for combined auditory and visual comparisons in Experiment 1.

ber of stimuli, it should be noted that the auditory solution appears to offer the best "reproduction" of the stimulus structure, with the horizontal dimension representing skew and the vertical dimension representing central tendency. All the negatively skewed stimuli are on the left, the normal stimuli are in the center, and the positively skewed stimuli are on the right. With only one exception (the reversal of Stimuli 2 and 3), the distributions having higher means are located higher on Dimension 2 than are those having lower means. With visual and combined presentation, the solutions essentially degenerate into three shape groupings in an equilateral triangular configuration. These findings suggest that auditory histogram presentation does seem to be capable of conveying information about distributional properties of data samples, even though it is a much less familiar format, and perhaps more effortful to process than traditional visual graphics.

## EXPERIMENT 2

### Auditory and Visual Box-Whisker Plots

A recent trend in computer software designed for exploratory data analysis is the use of highly simplified visual depictions of distributions, such as box-whisker plots (Tukey, 1977). By showing the relative position of the median, quartile boundaries, and extrema of distributions (often in conjunction with a routine for identifying and marking outliers), such displays presumably facilitate initial inspection of data or "quick and dirty" visual comparisons of distributions. Flowers and Hauer (1992) investigated a primitive auditory analogy of such plots—an ascending "arpeggio" of musical notes that mapped the extrema, quartiles, and median onto a musical scale. That study indicated that auditory quartile displays, consisting of a simple 5-note sequence, were found to be much less effortful to compare than were either visual or auditory histograms, although somewhat less information about distribution shape was conveyed by these simplified displays than by plots of the entire distributions. As a follow-up to that investigation, we conducted an experiment that required subjects to make paired comparisons among the same nine stimulus distributions that were used in Experiment 1. This experiment compared the use of visual box-whisker displays and a slightly modified version of the auditory box-whisker displays used by Flowers and Hauer (1992).

#### Method

**Subjects.** Seven graduate students who had completed the graduate course sequence in experimental design and data analysis served as subjects in two experimental sessions lasting about 0.5 h each. Each subject was paid \$10 for participating. All subjects had normal vision and hearing.

**Stimulus materials.** The numeric distributions used in Experiment 2 were identical to those of Experiment 1. Two different forms of displays were used to depict these distributions. Figure 5 (top)

shows an example of a visual box-whisker display, portraying the same pair of stimulus distributions shown in Figure 1 from Experiment 1—a normal, high central tendency distribution with a negatively skewed, medium central tendency distribution (Distributions 3 and 8 from Table 1). The asterisks mark the positions of the medians of the distributions, the edges of the boxes indicate the 25th and 75th percentiles, and the horizontal lines at the end of the whiskers indicate the positions of the extreme scores in each distribution. Thirty-six plots were constructed, pairing each of nine stimulus distributions with each of the eight other distributions.

**Auditory box-whisker displays** provided equivalent information to the visual box plots in auditory format (an example of which is shown in musical notation at the bottom of Figure 5). Again, Stimuli 3 and 8 are depicted. An auditory box plot consisted of a quarter note marking the median of the display, followed by a sixteenth note upward arpeggio marking the lower bound, 25th percentile, median, 75th percentile, and upper bound. The note values for the first distribution of a pair were scaled to fit into a single octave, beginning with middle C denoting the lowest score in that distribution. The note assignments for the second distribution of each pair were made on the same scaling. Thus, the note range of the second distribution could be more or less than one octave, depending on the relative variability of the two distributions. Except for the addition of the quarter note "median prefix" and the timing of the presentation, this method of constructing auditory box-whisker displays was identical to that of Flowers and Hauer (1992). Each arpeggio required approximately 1.5 sec to present. A 0.5-sec pause separated the presentation of the pairs. On every trial, a pair was presented three times before the subjects rated the dissimilarity, and a 1.5-sec pause separated these three presentations. Thus, the entire display sequence of a trial required approximately 13.5 sec to complete.

**Procedure.** As in Experiment 1, the subjects were given two blocks of 10 practice trials (one each for visual and auditory presentation) for familiarization with the task and range of variation among the stimuli. On experimental trials (two blocks of 36 trials each per session), dissimilarity judgments (using a 10-point scale) were entered onto an answer sheet similar to that used in Experiment 1. Visual presentation of stimuli was done using a hand-held booklet depicting each stimulus pair. Auditory presentation was done using the computer speaker as in Experiment 1.

#### Results and Discussion

**Regression analysis.** Since the box-whisker displays directly depict the median, quartile range, and range as features of the display, these stimulus parameters were used in regression analyses, along with differences in skewness, as predictors of the dissimilarity judgments. However, for both auditory and visual presentation, only two of these variables remained after stepwise removal of variables not significantly contributing to the model. These were, as might be expected from the outcome of Experiment 1, median and skew. For visual presentation, the standardized regression coefficient for median was .11, and the value for skew was .91. Multiple  $R$  was .94, indicating that the joint contribution of median and skew differences accounted for about 88% of the variance in dissimilarity judgments. For auditory presentation, a somewhat different pattern emerged. Attention was apparently divided more equally between median (beta = .43) and skew (beta = .45) differences. However, the lower multiple  $R$  (.69) indicated that somewhat less information was contributed by both variables, relative to vision (only about 48% of the variance of the dissimilarity judgments). While the initial presentation of the me-

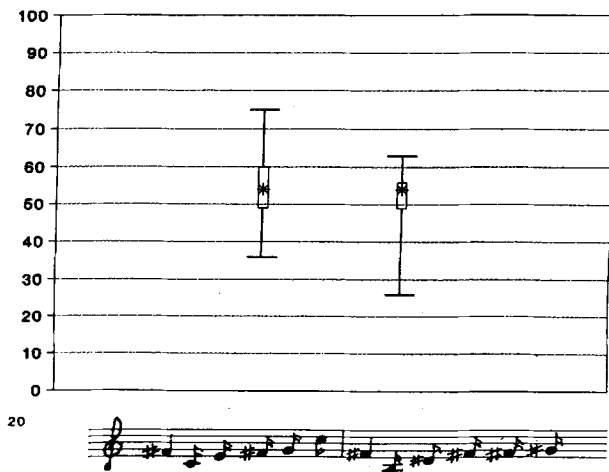


Figure 5. Example of a visual box-whisker display, comparing the same stimuli depicted in Figure 1 (top) and the note profiles corresponding to the auditory box-whisker comparisons of these same stimuli (bottom).

dian tone might have greatly increased the salience of central tendency, this perceptual emphasis might have been at the expense of attention to other parameters of the distributions. Given that *for this stimulus set* skew differences were paramount in their potential to allow easy discrimination of stimulus dissimilarity, this attentional shift may have come at a cost to overall performance. However, it should be noted that for most real world exploratory data analysis tasks, using actual data, it is usually the case that central tendency differences are of greater interest than shape differences, and it is somewhat unusual to encounter data sets from a single survey or experiment that include positive and negative skew differences of the magnitude included in the present stimulus set. For these reasons, we conducted two more experiments (Experiments 3A and 3B), using a new set of stimuli that included less variation in shape, together with greater variation in score variability. For these final experiments, we directly compared the effectiveness of two types of auditory box-whisker displays: displays that included a leading quarter-note median tone, and displays that presented only the arpeggio.

**EXPERIMENTS 3A AND 3B  
Auditory Box-Whisker Design**

**Method**

**Stimulus materials.** In order to provide a range of differences among simulated data samples that was less extreme and less dominated by a single perceptual dimension than were those of the previous two experiments, 12 new distributions of 50 numeric values each were computer generated. The distributional characteristics of each of these 12 samples are summarized in Table 3. Stimuli 1-3 were essentially symmetrical “normally” distributed distributions. Stimuli 4-6 were positively skewed distributions. Stimuli 7-8 were essentially “normal” in shape characteristics, but of somewhat larger dispersion than Stimuli 1-3 (larger standard deviations, quartile ranges, and ranges). We will refer to these three distributions as the *wide* distributions. For the three distributions within each of these categories (normal, skewed, and wide), central tendency, as measured by the median, varied from about 40 to over 50. Thus, Stimuli 1-9 can be distinguished on the basis of three statistical properties: central tendency, dispersion, and shape. Stimulus 11 was added to produce a distribution that was both wide and skewed; Stimuli 10 and 12 were added to produce two stimuli that were quite similar to two other stimuli in the set (Stimuli 2 and 8, respectively). These extra stimuli served as “validity” checks

on use of perceptual dimensions, since their placement in MDS solutions should help verify the extent to which subjects are actually “using” dimensions of central tendency, shape, and dispersion in their dissimilarity judgments of the displays.

Two types of displays were used: auditory box-whiskers *with a leading tone* (constructed using the same procedure as for the auditory box-whisker stimuli of Experiment 2), and auditory box-whiskers *without a leading tone* that were identical except for the omission of the leading quarter-note median tone. We hypothesized that inclusion of the initial median tone would cause the subjects to use more central tendency information in their dissimilarity judgments, perhaps at the expense of attending to shape or variability.

**Subjects.** Two different groups of subjects were used. For Experiment 3A, the subjects were 6 graduate psychology students, all of whom had completed a graduate level course in statistics and experimental design. Each subject received \$10 for participating in two sessions lasting about 45 min each. These subjects were fully informed about the purpose of the experiment and were given detailed instructions about the construction of the stimuli, similar to those provided to the subjects in Experiments 1 and 2. In Experiment 3B, 12 introductory psychology students, none of whom had completed a formal course in statistics, served as subjects. Their participation served as one alternative way of satisfying a “research exposure” requirement for the introductory psychology course. These subjects were instructed that the task involved judgments of “musical passages” (i.e., they were *not* told that these displays had a relationship to numerical values or statistics, until debriefing following the experiment). Our purpose in including perceptual judgments from a naive group of subjects, not aware of the experimental purpose or context, was to see whether the “natural” perceptual structure contained within the tone sequences would lead to perceptions of dissimilarity that corresponded to the actual statistical properties of the underlying numeric distributions.

**Procedure.** Except for the differences in instructions to the two groups described above, procedures for stimulus presentation and recording of dissimilarity ratings were identical to that for the box-whisker stimuli in Experiment 2. On each trial, a stimulus pair was presented three times before the dissimilarity rating was made. Each subject participated in two blocks of trials during each of the two sessions. One block included stimuli with leading tones; the other block had no leading tones. Over the two sessions, the subjects judged each possible distribution pair twice, and the average of these judgments was used as a dissimilarity rating.

**Results and Discussion**

**Regression analyses: leading-tone box-whiskers.** Dissimilarity judgments of both graduate students (Experiment 3A) and introductory students (Experiment 3B) were analyzed by a regression model in which differences in median, skew, range, and quartile range were used as

**Table 3  
Statistical Parameters of Distribution Stimuli for Experiments 3A, 3B, and 4**

Stimulus	Shape	Median	SD	Kurtosis	Skewness	Range	QRRange
1	Normal	43	9.0	.15	.38	42	11
2	Normal	51	10.2	-.55	.02	43	14
3	Normal	56	8.0	-.31	-.23	35	9
4	Skewed	39	11.0	-.05	1.09	40	11
5	Skewed	45	9.0	-.15	1.07	33	7
6	Skewed	51	9.0	.84	1.26	37	8
7	Wide	40	12.8	.11	.13	61	16
8	Wide	48	14.0	-.63	.11	57	21
9	Wide	53	13.1	.38	.62	64	16
10	Normal (like Stimulus 2)	49	9.0	.66	.43	43	10
11	Wide and Skewed	44	13.0	.90	1.24	52	8
12	Wide (like Stimulus 8)	45	13.0	.34	.13	65	16

predictors. Quartile range did not offer significant additional independent prediction from the other three variables, and thus the solutions shown for both subject groups include the median, skew, and range only. For the graduate students, the beta weights were .58 for median, .20 for skew, and .31 for range, indicating that all three of these differences in stimulus parameters contributed to the assessment of dissimilarity between display pairs. The multiple  $R$  was .67, indicating that these predictors accounted for about 45% of the variance of the dissimilarity judgments. Interestingly, the regression solution obtained for the introductory students, who were unaware that the stimuli they were comparing were depictions of numeric information, was nearly identical to that obtained for the informed, and presumably statistically knowledgeable, graduate students. The regression beta weights were .57 for median, .20 for skew, and .31 for range, and  $R = .63$ . This provides rather compelling evidence that the "natural" perceptual structure of the auditory box-whisker displays corresponds well to the statistical properties that it is intended to depict.

**Regression analyses: leading-tone-absent box-whiskers.** For the graduate students (Experiment 3A), regression weights obtained for the leading-tone-absent stimuli clearly reflect less attention to the central tendency differences in the displays. The beta weight for median was only .12; weights for skew and range were .39 and .41, respectively. Of equal importance, however, is that the multiple  $R$  (which can be viewed as an overall measure of information transmitted by the displays) was *less* than that for the leading-tone displays (.57), accounting for about 32% of the variance. Again, the regression solutions for the uninformed introductory psychology students (Experiment 3B) led to a nearly identical solution. Beta weights for median, skew, and range were .20, .38, and .41, respectively, and  $R = .58$ .

This pattern of results suggests that the leading-tone version of the auditory box-whisker displays is more effective in providing an appropriate balanced evaluation of central tendency, shape, and range for comparing data sets varying in all three of these characteristics. Moreover, the highly similar judgment patterns obtained for informed subjects (who may be considered reasonably knowledgeable *users* of statistical software) and for subjects making purely perceptual evaluations of tone sequences suggest that this very simple type of display technology represents a natural fit between auditory sensory experience and depiction of quantitative information.

**Multidimensional scaling solutions.** As might be expected from the strong similarity of the regression solutions for both the graduate students and the introductory psychology students, the MDS configurations obtained with these groups are highly similar. For leading-tone stimuli, stress values for three-dimensional solutions were .126 ( $R^2 = .80$ ) for the graduate students and .134 ( $R^2 = .79$ ) for the introductory students. Because of the overall similarity of these configurations (essentially they dif-

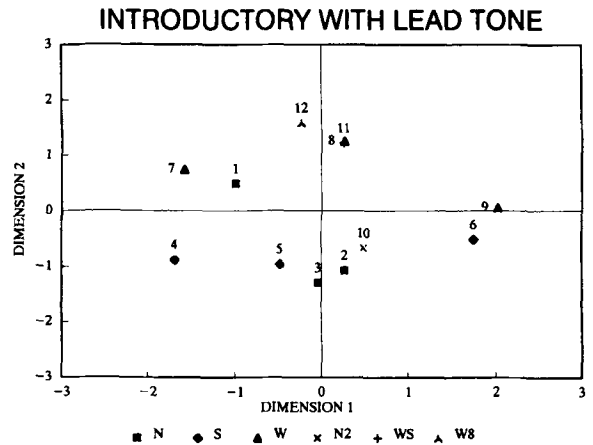


Figure 6. ALSCAL MDS solution obtained from the introductory psychology subjects in Experiment 3B, showing Dimension 1 (central tendency) versus Dimension 2 (dispersion).

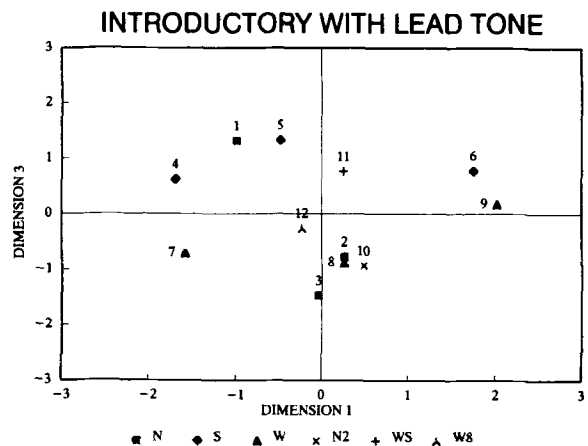


Figure 7. ALSCAL MDS obtained in Experiment 3B, showing Dimension 1 (central tendency) plotted against Dimension 3 (skew).

fered in reflection and rotation), we will only present the configuration obtained with the introductory students as Figures 6 and 7.

In Figure 6, the wide distributions 7, 8, 9, 12 are grouped at the top of the display, and the remaining stimuli (having less dispersion) are located in the lower portion of the display. With the exception of Stimuli 2 and 3, the stimuli having the larger central tendencies are located toward the left of the display. Thus, the horizontal dimension appears to represent perceived central tendency, and the vertical dimension represents perceived dispersion or variability. In Figure 7, the skewed stimuli appear at the top of the display separated from the normal stimuli (with the exception of Stimulus 1, which we will discuss presently). This third dimension thus appears to represent perceived skew. In general, these MDS plots confirm that dissimilarity judgments of the auditory box-whisker plots

were indeed based on perceptual differences in skew, spread, and central tendency. The only anomalously classified stimulus was Stimulus 1, which was rated as similar both to Stimulus 7 (a wide normal stimulus of about equal central tendency) and Stimulus 4 (a skewed stimulus of about equal central tendency). Inspection of the note assignments given to Stimulus 1 revealed that the median note was identical to that played for Stimulus 7, and that the major difference in note profile between the two sequences occurred at the beginning of the sequence in which Stimulus 7 began with the lowest note used in the entire set of displays, the A below middle C. Unfortunately, the speaker used on our computer greatly attenuated notes at those lower frequencies, so the greater spread at the low end of Stimulus 7 was not highly audible. We suspect that with better display equipment these stimuli might have been rated as somewhat less similar. A similar comparison of the note profiles for Stimulus 1 and Stimulus 4 revealed a similar problem, in that the primary differences in note profile that signify the greater skew of Stimulus 4 occurred at the beginning of the arpeggio. The median tones differed by only a half step (F vs. E), and the interval jump for the last two notes (the notes corresponding to the 75th percentile and the highest score for each distribution) only differed by a half step (G to C vs. G to C#). We suspect that primacy and recency effects for auditory sequences would normally create greater perceptual salience for the initial note and final notes of sequences such as these, and differences that are contained in the second, third, and fourth notes were probably not weighted as strongly in their contribution to judgments of dissimilarity between stimuli. One possible design modification that might help equalize ability to detect differences at both ends of the distributions in auditory box-whisker displays would be to require subjects to make comparisons for displays presented in both ascending and descending pitch order.

## GENERAL DISCUSSION

Our findings basically confirm our earlier (Flowers & Hauer, 1992) conclusions that auditory analogies to visual depictions of numeric distributions can convey equivalent quantitative information. However, our follow-up suggests that use of auditory signals *in conjunction with* traditional visual graphics is *inadvisable*, owing to basic limitations in human attentional systems. In addition, our results from Experiments 2, 3A, and 3B, suggest that "simpler is better" when it comes to conveying basic distributional information of the type important to exploratory data analysis. This seems apparent for both visual and auditory depictions; however, since time, in the auditory modality, is the domain that substitutes for visual space, simplicity is of particular relevance to auditory displays.

In summary, we feel the use of displays similar to our auditory box-whisker plots holds potential for facilitating initial grasps of distributional characteristics of data samples. The close correspondence between statistical parameters upon which these displays were constructed and "pure" perceptual judgments of dissimilarity from subjects uninformed of the purposes of these displays suggests that similar simple types of auditory displays might be useful quantitative education tools, particularly with the young and with visually impaired individuals.

## REFERENCES

- FLOWERS, J. H., & HAUER, T. A. (1992). The ear's versus the eye's potential to assess characteristics of numeric data: Are we too visuocentric? *Behavior Research Methods, Instruments, & Computers*, **24**, 258-264.
- LUNNEY, D., & MORRISON, R. (1990). Auditory presentation of experimental data. *Proceedings of the Society of Photooptical Instrumentation Engineers Conference*, **1259**, 140-146.
- TUKEY, J. W. (1977). *Exploratory data analysis*. Reading, MA: Addison-Wesley.