Mappings and Metaphors in Auditory Displays: An Experimental Assessment

BRUCE N. WALKER
Rice University
and
GREGORY KRAMER
Clarity/Santa Fe Institute

Auditory displays are becoming more and more common, but there are still no general guidelines for mapping data dimensions (e.g., temperature) onto display dimensions (e.g., pitch). This paper presents experimental research on different mappings and metaphors, in a generic process-control task environment, with reaction time and accuracy as dependent measures. It is hoped that this area of investigation will lead to the development of mapping guidelines applicable to auditory displays in a wide range of task domains.

Categories and Subject Descriptors: [**General Literature—General**]—*Conference proceedings*

General Terms: Experimentation, Human Factors, Performance

Additional Key Words and Phrases: Sonification, auditory display, data mapping, metaphors, guidelines

---

1. INTRODUCTION

Sound has been used in human-system interfaces for many years [e.g., Patterson 1982; Pollack and Ficks 1954]. Until recently, however, the majority of these audible cues have been simple warning sounds. True auditory display, where actual data is represented directly by one of many possible sound attributes, or dimensions, is rapidly maturing [cf. Kramer 1994a], but is still at the technical and conceptual stage that visual display was a few decades ago. More and more applications use sound to convey information, but, just as was the case with early visual displays, there are currently no standards and interface designers have usually implemented what sounds "good" to them. In addition, few designers have tested their auditory displays within a rigid experimental setting.

The principles for designing effective visual displays are quite generic, in that they apply to displaying all sorts of information, across a wide variety of task domains [e.g., Shneiderman 1992; Tufte 1990]. We are now investigating whether generalizable guidelines for auditory displays can be determined as well. In particular, we are examining the actual mapping of data dimensions (e.g., temperature) onto display dimensions (e.g., pitch). Intuitively, representing a rising temperature with a rising pitch seems

---

Authors' current addresses: Bruce N. Walker, Sonification Lab, School of Psychology, Georgia Institute of Technology, Atlanta, GA 30332-0170; email: bruce.walker@psych.gatech.edu; Gregory Kramer, Metta Foundation, Portland, OR, 97229; email: gregory@metta.org.

ACM Transactions on Applied Perception, Vol. 2, No. 4, October 2005, Pages 407–412.

a "natural" choice, but are there other such natural mappings? An important consideration is whether a particular mapping choice has an effect on the performance of a task that relies on the auditory display. Thus, are there better ways to represent temperature? Would another mapping produce faster or more accurate responses? Are some mappings more pleasing or easier to understand? We are concerned with identifying situations where the choice of data-to-display mappings can have a significant effect on performance, especially for data dimensions that may be represented in a wide variety of auditory displays. There are many types of information that can be presented through sound. Along with the usual temperature, pressure, size, cost, and rate, we are also very interested in how to best display more subjective and affective variables such as "value," "goodness," "beauty," "risk," and so on [Kramer 1994b].

In addition to the decision of which data dimension to represent by which auditory feature, the *direction* of the mapping is often critical. The temperature-pitch mapping seems natural only as long as rising pitch signals a rise in temperature. We still do not know whether the "inverse" mapping (i.e., rising pitch signals a *drop* in temperature) would actually affect performance on a task that relied on that auditory display. Some mappings are based on very common or "dead" metaphors [Lakoff and Johnson 1980], and we can intuitively decide which direction makes more sense to us. There are many cases, though, where it is difficult to predict which direction of a mapping will produce superior results. If "voltage" were mapped onto "richness" (number of harmonics, for example), should an increase in voltage be represented by an increase or a decrease in the number of harmonics? To really find out which direction of this mapping is more effective, we need a performance measure based on a task that requires the auditory display.

2. PROCEDURE

In order to measure performance in a task setting, yet still pursue generalizable, task-independent mapping results, we have chosen to use a generic process control (a "widget" factory) as our experimental environment [cf. Gaver, Smith, and O'Shea 1991]. This way we can include virtually any type of data dimension (including affective variables) and have complete control over how the variables interact and how they are displayed.

Participants listened to the auditory display via headphones in a sound-attenuated room and they made responses using a response box consisting of buttons and sliders. Each participant received a basic description of the Widget Factory and was trained to associate each data dimension (e.g., the weight of a widget) with a dimension of the auditory display (e.g., "brightness" of a sound). This training involved both verbal description and auditory practice.

The actual environment involved four variables each of which controlled one aspect of the audio output. The data values all remained at their starting points for several seconds. One of the variables then began to increase or decrease. The listener heard this as a period of steady state in the factory process followed by a change in one of the process parameters. He or she was required to make an appropriate control action using the labeled response buttons. For example, if the temperature dropped (e.g., represented by an increased loudness of the sound), then the correct response would be to press the "heater" button [see Fitch and Kramer, 1994, for a similar design].

Subjects all heard the exact same actual sounds, but were required to make different responses depending on their training condition. There were several different trial types, varying the starting values of the variables, and the variable that changed. Each trial type was repeated 10 times within a block of trials and each participant completed four blocks of trials. The independent variables included the particular mapping that the listener had been trained to hear and the actual variable that changed on a given trial. The performance measures included response time (RT) and accuracy. In addition, after each trial, the participant was asked to say which parameter of the process changed to ensure that he or she was paying attention to the metaphor and not simply mapping the auditory display parameter directly to the response button.

3. PREDICTIONS

Note: Since this paper is being reprinted largely as it originally appeared, the data that were gathered and presented at the ICAD 1996 conference are not included here. However, those data and discussion are included in the Appendix.

Mappings that are based on stronger or more natural metaphors should result in faster and more accurate control reactions. They should also be learned faster, which would lead to a greater improvement in performance across the blocks of the experiment. For some mappings there should also be a particular direction that results in better performance (e.g., rising temperature mapped to rising, as opposed to falling, pitch). These results should complement the findings in the area of stimulus compatibility [e.g., Proctor and Reeve 1990] and cross-modality matching [Melara and O'Brien 1990; Walker and Ehrenstein 1996].

4. IMPLICATIONS

It is likely that a number of the most "successful" mappings will be the ones that have most often been used in auditory displays. However, we hope to discover other good mappings and, in particular, we will try to display variables that have great possibilities, but have not often been represented with sound. The strong emotive power of music [cf. Révész 1954] suggests that affective variables are perfect examples of information that may be difficult to describe with words or pictures, but will be easily recognized with sound.

This research is a big first step in attempting to quantitatively compare different auditory display setups. We are careful to note that the design of an effective auditory display will always require practice and good judgment. However, the extension of the present research may help to identify guidelines for representing data with sound, which will hopefully apply across a wide range of task domains.

APPENDIX: PRELIMINARY RESULTS AS PRESENTED AT ICAD 1996

For this project, the final data collection and analyses were completed after the paper was submitted, but before the ICAD conference. The following reflects the "preliminary results" as actually presented at ICAD 1996.

A.1 Mappings

The data dimensions that were employed included temperature, pressure, size, and rate. These were chosen both because they matched the cover story of the experiment and also because they are commonly used data dimensions in sciences ranging from physics to seismology to chemistry. These were mapped to the display (sound) dimensions of loudness, pitch, tempo, and onset sharpness. At the time, the sound dimensions we chose were all fairly easily manipulated. Note that this is not to suggest that these are the best, or even the only sound attributes one could or should use in such data sonifications. They were simply convenient. Sets or "ensembles" of data-to-display pairs were developed as shown in Table A1. Each listener performed the task using one of the mapping ensembles. The ensembles were designed to result in different levels of performance, based on how "natural" or "intuitive" the mappings seemed to the sound designers. Thus, the "Intuitive" mapping ensemble was supposed to have all the best pairings of data dimensions to sound attributes. The "Okay" ensemble was meant to be good, but not optimal. The "Bad" ensemble was designed to be counterintuitive and to yield poor performance. The "Random" ensemble was simply to balance out the 4×4 array of mappings and was not expected to lead to effective

Table A1. Sets ("Ensembles") of Data-to-Display Mappings

Data Dimension	Display Dimension (Sound Parameter)					
	"Intuitive"	"Okay"	"Bad"	"Random"	"Intuitive-Pitch-X"	"Bad-Pitch-X"
Temperature	Pitch	Loudness	Onset	Tempo	Pitch-X	Onset
Pressure	Onset	Pitch	Tempo	Loudness	Onset	Tempo
Size	Loudness	Tempo	Pitch	Onset	Loudness	Pitch-X
Rate	Tempo	Onset	Loudness	Pitch	Tempo	Loudness

Note: The labels "Okay", "Bad", and so on for the mapping ensembles were chosen by the experimenters, based on the intuitions of the sound designers, and before any performance data were gathered. These labels turned out not to be very good descriptors of the effectiveness of the various mapping ensembles.

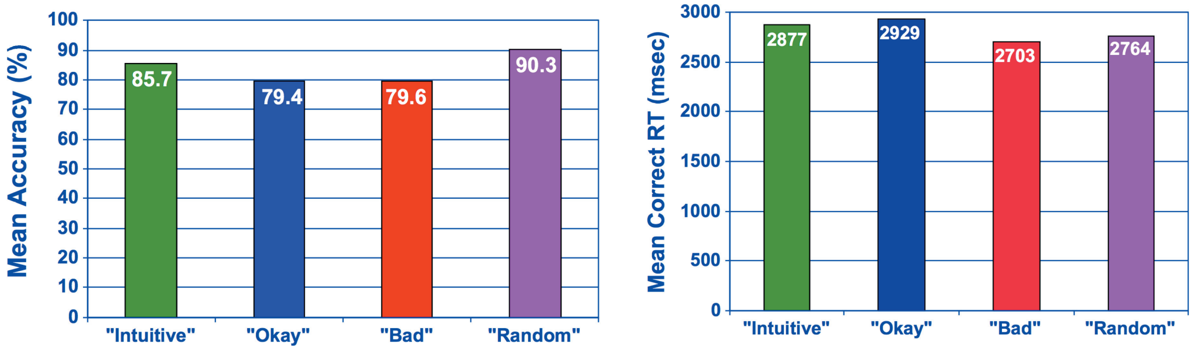

Fig. A1. Accuracy and reaction time (RT) results for the four primary mapping ensembles.

performance. The predictions in the original paper reflect the intended relation between "intuitiveness" of the mapping ensemble and performance. As will become clear, the labels "Intuitive," "Okay," and so on, did not turn out to be very applicable, once the actual performance was considered!

In addition to the four primary ensembles, we were also interested in the effect of the direction or "polarity" of the mappings. That is, we wanted to look at whether an increase in, say, temperature would be more effectively represented by an increase or a decrease in, say, pitch. Thus, we created two additional ensembles that used the same data-to-display mappings as the "Intuitive" and "Bad" ensembles, but simply had the pitch dimension's polarity reversed. Thus, for example, in the "Intuitive-Pitch-X" ensemble, as temperature increased, the pitch of the sonification *decreased* (denoted Pitch-X), opposite from the behavior of pitch in the "Intuitive" ensemble.

A.2 Results

Figure A1 summarizes the key results from the four primary ensembles. We were surprised to see that the "Bad" ensemble actually led to the fastest performance. The supposedly "Intuitive" and "Okay" ensembles led to the poorest performance, overall, while the "Random" ensemble led to the best performance overall, when both RT and accuracy were considered together.

Next, Figure A2 presents the RT and accuracy results for the ensembles where the polarities of the pitch dimension was flipped. For the "Intuitive" ensemble, the change of polarity had no effect, when both speed and accuracy are considered together. However, the simple change of one polarity in the "Bad" ensemble (which, as described above, was not really bad, after all) resulted in a dramatic increase in RT. Thus, performance was considerably worse in the "Bad-Pitch-X" condition.

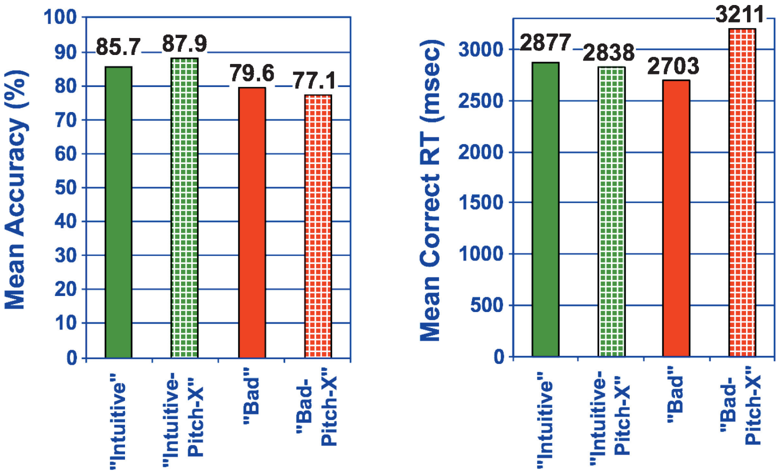

Fig. A2. Accuracy and reaction time (RT) for the "Intuitive" and "Bad" ensembles, as well as for the ensembles created by simply flipping the polarity of one data-to-display mapping. Note that the "Intuitive" mapping was robust to this change, whereas the "Bad" ensemble was relatively fragile (i.e., performance with the "Bad-Pitch-X" ensemble was much poorer than it was with the "Bad" ensemble).

Table A2. Summary of Mapping Effectiveness

	Loudness	Pitch	Tempo	Onset
Temperature	**good**	okay	okay	poor
Pressure	okay	okay	poor	poor
Size	okay	okay	poor	**good**
Rate	okay	**good**	okay	poor

A.3 Discussion

The first main conclusion from these results is that data-to-sound mappings that seem intuitive to a sound designer may actually result in less effective performance. This demonstrates how crucial it is to empirically test an auditory display with listeners representative of the final users. Next, it became clear from this experiment that the exact pairing of a data dimension to a sound attribute could affect the way listeners interpreted the sound's meaning. That is, sound design and mapping matter! There may be particular sound dimensions that are best for representing a given data type. Overall, the mapping recommendations made at the ICAD 1996 conference are summarized in Table A2. Clearly only a few mappings were considered to be "good" or effective, while many were considered "okay," and some "poor." In terms of the effectiveness of a given display attribute, this table also shows that tempo and onset were generally not as effective as pitch and loudness. This is not surprising when RT is one of the metrics of effectiveness, because both tempo and onset changes are relatively slow to manifest. On the other hand, when representing "size" (and only size, in this experiment), onset can be an effective display dimension. Post hoc explanations may be able to explain that result in terms of slower changes being representative of larger objects (i.e., due to inertia), but such an effective mapping was never predicted a priori. This again points to the need for iterative prototyping and verification of interface designs. The third main conclusion is that the polarity of a mapping is a crucial element of sound design. This point had never really be made before, since it was generally assumed that increasing data values should be represented by increases in a sound attribute (e.g., rising temperature mapped to increasing pitch). However, this experiment, along with other work in our lab at the time, demonstrated that at least some of the listeners brought expectancies about how certain data "ought to" sound. In some cases, these mental models reflected a preference for a negative polarity. For example, we now know that increasing mass is generally best represented by decreasing pitch. Of course, there are considerable individual differences in mental models as they relate to auditory displays and this is an area undergoing some investigation now—a decade later. The implications of this research for auditory displays are that guidelines may emerge, based on this kind of study, so that designers will not need to guess about effective mappings, and will be able to deploy more truly “intuitive” mapping ensembles that actually allow a listener to understand the message the display is meant to convey.

REFERENCES

Fitch, T. and Kramer, G. 1994. Sonifying the body electric: Superiority of an auditory display over a visual display in a complex, multivariate system. In *Auditory Display: Sonification, Audification, and Auditory Interfaces*, G. Kramer, Ed. Santa Fe Institute Studies in the Sciences of Complexity. Addison-Wesley, Reading, MA. 307–325.

Gaver, W., Smith, R. B., and O'Shea, T. 1991. Effective sounds in complex systems: The ARKola simulation. In *Proceedings of CHI1991*, held April 28–May 2, New Orleans, LA. ACM Press/Addison-Wesley, Reading, MA.

Kramer, G. 1994a. An introduction to auditory display. In *Auditory Display: Sonification, Audification, and Auditory Interfaces*, G. Kramer, Ed. Santa Fe Institute Studies in the Sciences of Complexity. Addison-Wesley, Reading, MA. 1–77.

Kramer, G. 1994b. Some organizing principles for representing data with sound. In *Auditory Display: Sonification, Audification, and Auditory Interfaces*, G. Kramer, Ed. Santa Fe Institute Studies in the Sciences of Complexity. Addison-Wesley, Reading, MA. 185–221.

Lakoff, G. and Johnson, M. 1980. *Metaphors We Live By*. University of Chicago Press, Chicago, IL.

Melara, R. D. and O'Brien, T. P. 1990. Interaction between synesthetically corresponding dimensions. *Journal of Experimental Psychology: General 116*, 4, 323–336.

Patterson, R. D. 1982. *Guidelines for auditory warning systems on civil aircraft*. Paper No. 82017, Civil Aviation Authority, London.

Pollack, I. and Ficks, L. 1954. Information of elementary multidimensional auditory displays. *Journal of the Acoustical Society of America 26*, 155–158.

Proctor, R. W. and Reeve, T. G., Eds. 1990. *Stimulus-Response Compatibility: An Integrated Perspective*. North Holland, Amsterdam.

Révész, G. 1954. *Introduction to the Psychology of Music*. University of Oklahoma Press, Norman, OK.

Shneiderman, B. 1992. Designing the User Interface: Strategies for Effective Human-Computer Interaction (2e.). Addison-Wesley, Reading, MA.

Tufte, E. 1990. *Envisioning Information*. Graphics Press.

Walker, B. N. and Ehrenstein, A. 1996. Cross-dimensional compatibility effects with dynamic auditory stimuli. Paper presented at the Southwest Psychological Association 42nd Annual Convention. Houston, April 4–6, 1996.

Received February 2005; accpeted June 2005