

Consistency of Magnitude Estimations with Conceptual Data Dimensions Used for Sonification

BRUCE N. WALKER*

Georgia Institute of Technology, USA

SUMMARY

In two experiments listeners assessed how sounds represented data, as might be used in sonification. In Experiment 1, 209 undergraduates used magnitude estimation to define the relationship between three sound attributes (frequency, tempo, modulation index) and 10 data dimensions (size, temperature, pressure, velocity, number of dollars, urgency, proximity, attractiveness, danger, mass). Polarities and slopes (i.e. power function exponents) are reported and compared to predictions from the literature. In Experiment 2, 226 new participants demonstrated polarities and slopes are stable across a direct replication. Results show that listener expectations depend on both sound and data dimensions in use. While there are some unanimous expectations across listeners, there are also differences due to different mental models formed by participants, which may relate to listening experience. Copyright © 2006 John Wiley & Sons, Ltd.

Designers of data displays often apply the results of perception and cognition research to improve effectiveness. For example, the design of dials, gauges, charts, and graphs can be informed by studies of visual acuity thresholds, colour discrimination, perceptual grouping and motion perception (e.g. Sanders & McCormick, 1993). Not all displays are visual, of course; when a person cannot look at or cannot see a visual display, a well-designed auditory display may be very effective (Kramer et al., 1999; Sanders & McCormick, 1993; Wickens, Lee, Liu, & Gordon Becker, 2004). In fact, the speed and pattern recognition capabilities of the human auditory system mean that in some data analysis tasks auditory displays of data can even be more effective than visual displays (e.g. Fitch & Kramer, 1994; Kramer et al., 1999). However, there is still only a relatively thin empirical foundation upon which to base auditory display design decisions; fundamental research is needed to bring auditory displays up to the same level of effectiveness as the more ubiquitous and more well-studied visual displays.

A *sonification* is a particular type of auditory display, in which relationships in a data set are translated into, or represented by, sounds for the purpose of understanding or discovering patterns in the data set (Kramer et al., 1999; Walker, 2002; Walker & Kramer, 2004). While it is clearly an oversimplification, it is nevertheless helpful to think of sonifications as sophisticated auditory graphs in which data values are used to change the parameters of a synthesized tone. For example, changes in temperature might be indicated

*Correspondence to: Bruce N. Walker, School of Psychology, Georgia Institute of Technology, 654 Cherry Street, Atlanta, GA 30332-0170, USA. E-mail: bruce.walker@psych.gatech.edu

by changing the pitch of a sound. The resulting sonification can be used for a broad array of data analysis and pattern recognition tasks, ranging from the discovery of patterns in weather data (e.g. Flowers & Grafel, 2002), to assisting in the detection of tumours (e.g. Martins & Rangayyan, 1997), examining structural data from concrete highway bridges (e.g. Valenzuela, Sansalone, Krumhansl, & Streett, 1997), and analysing stock market data (e.g. Nesbitt & Barrass, 2002). Sonifications and auditory graphs in particular, can also be useful in providing access to data by students and scientists with visual impairments (e.g. Gardner, 1999; Gardner, Lundquist, & Sahyun, 1996; Meijer, 2004; Walker & Cothran, 2003; Walker & Lowey, 2004).

A sonification's success depends heavily on the details of the mapping of data dimensions (e.g. temperature, stock price) onto the sound dimensions that are used to represent them (e.g. pitch, tempo). This dependence on effective mappings is true in the design of any kind of data display, be it visual or auditory, and extends to the design of auditory warnings and trend monitoring sounds (e.g. Edworthy, Hellier, Aldrich, & Loxley, 2004). The research presented here continues a line of investigation into the best ways to map data specifically onto sound parameters, and discusses initial investigations of the stability of mapping preferences.

As Walker has pointed out (Walker, 2002; Walker & Kramer, 2005), the mapping issue for sonification is at the centre of at least three design choices. First, it matters which specific sound dimension is chosen to represent a given data dimension. This is partly because there seems to be some agreement among listeners about what sound attributes are good (or poor) at representing particular data dimensions. For example, pitch is generally good for representing temperature, whereas tempo is not as effective (Walker, 2002). It is also partly because some sound dimensions (e.g. loudness) are simply not very effective in auditory displays for practical design reasons (Walker & Kramer, 2004). Second, designers must choose an appropriate polarity for the data-to-display mappings. For example, listeners might agree that pitch should increase in order to represent increasing temperature (a positive mapping polarity, Walker, 2002), while at the same time feel that pitch should decrease in order to represent increasing size (a *negative* polarity). The issue of polarity is not typically an issue for visual displays, but it can be very important in auditory representations ranging from helicopter warning sounds (Edworthy et al., 2004) to interfaces for the visually impaired (Walker & Lane, 2001). The third design choice is how much change in, say, the pitch of a sound is used to convey a given change in, for example, temperature. Matching the data-to-display scaling function to the listener's internal conceptual scaling function between pitch and temperature is critical if the sonification is to be used to make accurate comparisons and absolute or exact judgements of data values, as opposed to simple trend estimations. This is a key distinction between sonifications and warnings or trend monitoring sounds. All three of the display design choices discussed here depend on knowing more about the preferred mappings, polarities, and scaling functions listeners have for a range of data and sound dimensions that are likely to be used in auditory displays and sonifications. A core underlying assumption is that such preferences exist and that they can be determined empirically. This assumption has been supported in a variety of instances, often using the psychophysical method of magnitude estimation.

Magnitude estimation has become a standard tool for studying the relationship between an acoustic variable and its perceptual correlate, such as between sound frequency and pitch (for several examples, see Stevens, 1975). In addition, magnitude estimation can be used to study the relationship between an acoustic variable and a listener's concept of a data variable that is not perceptual, but rather conceptual. For example, magnitude estimation

can identify the scaling function between sound frequency and conceptual data scales ranging from the level of racism attributed to certain acts (Dawson & Brinkler, 1971), to urgency (Hellier, Edworthy, & Dennis, 1995), to conceptual adjectives such as 'dangerous', 'jerky', 'safe' and 'heavy' for use in cockpit warnings (Edworthy, Hellier, & Hards, 1995). Walker (2002, Experiments 2 and 3) extended the use of magnitude estimation even further to determine listener expectancies for not only scaling functions, but also for mappings and polarities that may be useful in all sorts of data sonifications. That paper reported on the relationship between the auditory display dimensions of frequency and tempo, and the data dimensions of temperature, pressure, velocity, size and number of dollars, which are common and generic enough to be included in many sonifications, certainly including the sonification examples already mentioned. It turns out that the preferred mapping, polarity, and scaling slopes all depend on the data dimension that listeners are instructed to scale, but within the categories there is still considerable agreement among listeners.

The key conclusions from these lines of research are: (1) the polarity and scaling function both matter for auditory display designs; (2) the preferred polarity and optimal scaling function depend on the data-to-display mapping; and (3) it is difficult to predict with certainty either the preferred polarity or the scaling function for a given group of people, but these can be determined empirically. There are many data dimensions, both generic like *temperature* (atmospheric, nuclear, chemical, etc.) and specific like *number of dollars*, that are of interest to display designers but for which there remain little or no empirically-obtained guidance. Understanding how and why a listener interprets sounds in a particular manner is also of general interest to cognitive and psychophysical scientists. This, then, points to the need to assess the data-to-display mapping relationship for many more data categories and sound attributes.

The results of such expanded studies are of practical interest only if there is also an accompanying idea of both the applicability and stability of the results reported. In terms of applicability, Edworthy and her colleagues (e.g. Edworthy, Loxley, & Dennis, 1991; Hellier, Edworthy, & Dennis, 1993; Hellier, Edworthy, Weedon, Walters, & Adams, 2002) have repeatedly shown that designing warning sounds based on the results of magnitude estimation studies leads to predictable orderings of 'urgency', and therefore to better auditory warning systems. Guillaume, Pellieux, Chastres and Drake (2003) have recently argued that, at least in the case of warnings, in addition to acoustic variables, such as frequency or repetition rate, more cognitive factors need to be considered. Both of these assertions (i.e. the effectiveness of magnitude estimation as a tool, as well as the need to consider mental models) is supported by the results from Walker (2002) that showed that a set of magnitude estimation values obtained for one set of listeners was generally effective for conveying information in an auditory graph to other listeners, but that there were differences in the mental models of some listeners in relation to conceptual data dimensions like temperature and number of dollars.

In terms of the question of stability, there has been considerable investigation over the years looking at both intra- and interindividual stability of magnitude estimation slopes for perceptual dimensions. At the level of the individual, for example, M. and R. Teghtsoonian (1971, 1983) have looked at how stable magnitude estimation slopes are for a single participant, across multiple replications, with different inter-session intervals. Interestingly, they have found only weak intra-individual stability from

session to session unless there is essentially no break between the sessions (see M. Teghtsoonian & Teghtsoonian, 1983 for a review). This is in contrast to the situation for magnitude estimation slopes determined from the data of groups of participants, where intergroup stability does seem to be evident. The values obtained in the various studies are not in exact agreement, but the differences have often been ascribed to slight differences in procedure or random variation. That is, most of the comparisons of this sort have been across experiments where stimuli (e.g. R. Teghtsoonian & Teghtsoonian, 1997), methods (e.g. Hellier et al., 1995), and participant pools (e.g. Hellier, Wright, Edworthy, & Newstead, 2000; M. Teghtsoonian, 1980) were slightly different. The fact that the slope values have nevertheless shown some stability across experiments has led to the assumption of a 'real' underlying or consensus slope value for many dimension pairs (e.g. ~ 0.66 between intensity and loudness, Stevens, 1966, 1969). A simple averaging of the exponents of the power functions that describe the magnitude estimation curves across studies is often used to estimate the supposed real scaling value (see Stevens, 1975 for several such summary values). In the realm of more conceptual (also known as, 'second-order') data dimensions like urgency, Hellier, Edworthy, and their colleagues have, as well, shown some stability of group exponents from study to study, and even from the US to the UK to some degree (e.g. Hellier et al., 2000). There is, nevertheless, still some variability between magnitude estimation studies with conceptual dimensions, which has been attributed to different stimulus ranges, response ranges, analysis methods, and instructions to participants (e.g. Hellier et al., 1995). It is important to note that magnitude estimation is not the only method used to explore the relationship between sounds and perceived data values, and that there can be variability between results, depending on the experimental methods employed (Edworthy et al., 2004). However, the present paper is focused mostly on the application of magnitude estimation in the examination of data-to-sound mappings.

Considering the discussion to this point, it may be problematic for a display designer to assume there is a 'correct' or 'best' value for a magnitude estimation slope (i.e. a population parameter, or shared scaling factor) that should be used for a display, when the actual exact value to use may not be clear, given that different researchers have arrived at slightly different values via slightly different stimuli and procedures. It is not yet clear how much of the variations in slopes from experiment to experiment are the result of the variabilities in procedure and stimuli, and how much is simply due to inevitable, random variability from group to group. An important first step, then, is to determine how stable the magnitude estimation slopes are under an identical replication of experimental procedures within a population. It does not appear that there has been a study involving a direct replication of a complete magnitude estimation study with virtually identical conditions and certainly not for estimations of conceptual data dimensions. Further, the notion of mapping polarity, which is considerably more important for conceptual data dimensions than for traditional psychophysical dimensions like pitch (Walker, 2002), has never been directly assessed for its stability (but see Edworthy et al., 2004). Thus, the research presented here reports on the polarities and magnitude estimation slopes for 30 conceptual data-to-auditory display pairs, and further investigates the stability of both slopes and polarities across replications of the data collection process. Predictions are made in advance for some of the key dimension pairings, based on related results in the literature.

EXPERIMENT 1

Method

Participants

A total of 209 undergraduates (118 males, 91 females; mean age 20.7 years) from the Georgia Institute of Technology participated in Experiment 1 for partial course credit. All reported normal or corrected-to-normal vision and hearing, and provided informed consent.

Stimuli and apparatus

The experiments used an Apple Macintosh G4 computer and 17 inch (43.2 cm) Studio Display to present instructions; the auditory stimuli were played through Sony MDR-7506 headphones. The experiment ran via JavaScript in Netscape Navigator v.4.7 on Mac OS 9.2.

There were three sets of auditory stimuli: the Frequency Set, Tempo Set and Modulation Index Set. These sets of stimuli varied along acoustic dimensions that are often easily manipulated in sonification and auditory display systems. This is certainly only a subset of the possible acoustic dimensions that could be varied, both in sonification systems and in a listening study, however it does represent some of the major dimensions in common usage. As discussed earlier, loudness is also often manipulable in such systems, but is problematic (and therefore not recommended) in practical settings, so it was not included in this study. The 10 stimuli in the Frequency Set were each 2-seconds long pure sine wave tones synthesized with frequencies of 90, 205, 320, 415, 790, 1000, 1350, 1750, 2410 and 3200 Hz. The stimuli were normalised for perceived loudness, matching the 1000 Hz tone at 60 dBA SPL. Each stimulus had a 10-ms amplitude ramp at onset and offset to avoid sudden loudness changes. The frequency values were chosen to be within the frequency range of typical sonifications, and similar to the values used previously by Walker (2002).

For the Tempo Set, each auditory stimulus was made up of a series of one-beat long pure sine wave tones (1000 Hz), each followed by a half-beat of silence. These sound and silence elements were looped to create continuous on-off stimuli of different tempos (i.e. beep beep beep vs. beeeep . . . beeeep . . . beeeep . . .). Note that the terminology used to describe tempo here is based on the computer music foundations of many sonification systems (see, e.g. Kramer et al., 1999). When using such systems to represent data, the relative lengths of different sounds (and silences) are often specified in terms of beats, and then the whole 'score' (i.e. the sounds representing the whole data set) can be performed or rendered at any tempo, indicated in units of beats per minute (bpm). This music-based approach means that both sounds and silences become shorter or longer as the tempo increases or decreases. By contrast, a different concept of tempo (e.g. 'speed', Edworthy et al., 1995) might keep the sound portion of a signal constant in duration, and then adjust tempo by changing the interval between sounds, often called the interpulse interval. Both approaches are valid; the musical approach was taken here to preserve the applicability to music-based auditory display and sonification software (e.g. The Sonification Sandbox, Walker & Cothran, 2003). The 10 stimuli in this Tempo Set were each 6 seconds long, but as described above, the length of the component tones and silences in the stimulus depended on the performance tempos of 41, 60, 107, 167, 203, 270, 415, 505, 572 and 685 bpm. For example, at 60 bpm one beat lasts one second, so the exact length of the tone portion of the stimulus can be determined simply by dividing 60 by the tempo value,

resulting in lengths of 1.46, 1.00, 0.56, 0.34, 0.30, 0.22, 0.14, 0.12, 0.10 and 0.09 seconds, respectively. As was the case for the Frequency Set, a 10-ms onset and offset amplitude ramp was applied to the tone portion of each stimulus. Also as with frequency, the tempo values were chosen to be in the range typically used in sonification, and similar to the tempo values used in Walker (2002). All of the sounds in the Tempo Set were presented at 60 dBA SPL.

The 10 sounds in the Modulation Index Set were composed of 2-seconds long FM-synthesized sounds each with a carrier frequency of 100 Hz, a modulation frequency of 300 Hz, and a modulation index (i.e. number of harmonics) of 1, 2, 3, 4, 5, 6, 7, 8, 9 or 10. Increasing the modulation index has the effect of increasing the perceived 'brightness' or spectral centroid of the sound, while maintaining the fundamental frequency (and perceived pitch) as constant. The Modulation Index Set was a new set of stimuli varying along a dimension that had not previously been studied in this paradigm, but one that can be (and often is) directly controlled by sonification software employing FM-synthesis to drive sound creation. Again, through pre-testing, all sounds within the set were equated for apparent loudness and matched to the 1000 Hz tone in the Frequency Set.

Design and block structure

These three physical display dimensions (frequency, tempo and modulation index) were paired with their corresponding perceptual dimension (pitch, perceived tempo and perceived tonal brightness), and with 10 other conceptual data dimensions for the purposes of magnitude estimation. The conceptual data dimensions were size, temperature, pressure, velocity, number of dollars, urgency, proximity, attractiveness, danger and mass. These data dimensions are commonly used in data sonification and other auditory displays; some have been studied previously using magnitude estimation, whereas the rest have not.

Each data dimension (e.g. temperature) was paired with one of the display dimensions (e.g. frequency) for an entire block of trials, which will be referred to in *data:display* notation (e.g. a temperature:frequency block). The perceptual dimension of pitch was only paired with frequency; the perceptual dimension of perceived tempo was only paired with tempo; and the perceptual dimension of perceived brightness was only paired with modulation index, since those perceptual dimensions were calibrations of their respective acoustic dimensions. All of the other data dimensions were paired with each of the display dimensions, so there were all possible blocks such as size:frequency, size:tempo, size:modulation index, urgency:frequency, urgency:tempo, urgency:modulation index and so on.

Each participant completed a total of three blocks of trials, including one block with the Frequency Set, one block with the Tempo Set, and a third with the Modulation Index Set of sounds. Each block used a different data dimension, so a sample set of blocks for a participant would be: urgency:tempo, danger:modulation index and temperature:frequency. The type and order of blocks was counterbalanced pseudorandomly, subject to the constraint that there be approximately equal participants in each block type.

Trial structure and task

The task of the listener was to estimate the magnitude of the data dimension that the sounds were meant to represent. Before each block of trials, the experimenter read aloud instructions like the following, while the participants followed along on the computer screen:

You will hear a series of sounds, one at a time, in random order. Your task is to indicate what *temperature* the sounds seem to represent, by assigning numbers to them. For the first sound, assign it any number of your choosing that represents a *temperature*. Then, for each of the remaining sounds, estimate its '*temperature*', relative to your subjective impression of the first sound. For example, if the second sound seems to represent a *temperature* that is 10 times as hot as the first, then assign it a number that is 10 times bigger than the first number. If the sound seems to represent a *temperature* that is one-fifth as hot, assign it a number that is one-fifth as large as the first number, and so on. You can use any range of numbers, fractions, or decimals that seem appropriate, so long as they are greater than zero. [*italics added here, only*]

These instructions were modelled closely after the instructions often used in modulus-free magnitude estimation tasks (e.g. Engen, 1971; Stevens, 1975; Walker, 2002), and were designed to encourage the use of a ratio scale, while avoiding the imposition of any particular range for that scale. For example, the listener was free to think in terms of daily temperatures, thermonuclear temperatures, or any other concept of temperature that he or she chose. The requirement that responses be greater than zero is a mathematical necessity of the ratio scale, and was included in the instructions merely as a practical reminder to the participant, rather than as any experimental constraint, since if their responses were multiples or fractions of a starting number, they should always be greater than zero.

On each trial the participant heard one stimulus from the set being used for that block (e.g. Frequency Set) and entered a number for the subjective value (e.g. the temperature) of that stimulus. In a block of 20 trials, each one of the 10 stimuli was randomly presented twice, with the constraint that the largest or smallest stimulus in that set could not occur first (see R. Teghtsoonian & Teghtsoonian, 1978; Walker, 2000). Following a brief rest, the participant began the next block with new instructions that introduced different data and display dimensions. Another brief rest then preceded the third block of trials.

Predicted results

For many of the data dimensions in this study there are no directly comparable results in the literature, making predictions about preferred polarities and magnitude estimation slopes difficult. Of course, that is also one of the reasons for completing this study. On the other hand, for the data-to-display pairs where prior results do exist, it is possible to make predictions about polarities, and in some cases, the actual magnitude estimation exponents. It should be noted that in psychophysics studies that do not involve conceptual data dimensions, if participants do not respond in the polarity direction that the experimenter presumes to be correct, or with a sufficiently linear set of responses, those data are often removed from subsequent analyses (e.g. M. Teghtsoonian, 1980). Thus, the very idea of polarities in magnitude estimation is relatively new, and generally only applies to conceptual data dimensions (see Edworthy et al., 2004; Edworthy et al., 1995; Walker, 2002). Figure 1 provides a brief overview of the data that lead to the predictions detailed below, along with a summary of the results from the two experiments reported here. Full details of the present study's results are contained in the data tables discussed in the Results section.

Walker (2002) scaled frequency and tempo to some of the data dimensions used here. There were several non-unanimous polarities reported, which complicates the predictions. A simplified summary of the results from Walker (2002, Experiment 2) is contained within

Sound Dimension	Concept/Data Dimension	Previous Polarity & Exponent		Refs	Experiment 1		Experiment 2	
Frequency	Size	↓↑	-0.76	1	↓↑	-0.87	↓↑	-0.64
	Temperature	↑	0.95	1	↑↓	0.65	↑↓	0.56
	Pressure	↑↓	0.78	1	↑↓	0.77	↑↓	0.63
	Velocity	↑	1.06	1	↑	0.77	↑↓	0.82
	Num. Of Dollars	↑↓	0.77	1	↑↓	1.36	↑↓	0.84
	Urgency	↑		2,3	↑	0.70	↑	0.68
	Proximity				↑↓	0.78	↑↓	0.54
	Attractiveness				↓↑	-0.58	↓	-0.81
	Danger	↑		2	↑	0.69	↑	0.49
	Mass				↓	-0.91	↓↑	-0.96
Tempo	Size	↓	-0.94	1	=		↓↑	-0.58
	Temperature	↑↓	0.43	1	↑↓	0.63	↑↓	0.45
	Pressure	↑↓	0.68	1	↑	0.84	↑↓	0.75
	Velocity	↑	1.04	1	↑	0.90	↑	0.83
	Num. Of Dollars	↑↓	0.66	1	↑↓	1.07	↑↓	0.92
	Urgency	↑	0.61	2,3,4	↑	0.64	↑↓	0.66
	Proximity				↑↓	0.51	↑↓	0.60
	Attractiveness				↑↓	0.80	↑↓	0.60
	Danger	↑		2	↑	0.61	↑↓	0.51
	Mass				↓↑	-0.71	↓↑	-0.99
Modulation Index	Size				=		↑↓	0.52
	Temperature				↑↓	0.62	↑	0.74
	Pressure				↑	0.69	↑↓	0.68
	Velocity				↑	0.67	↑	0.82
	Num. Of Dollars				↑↓	0.69	↑↓	0.87
	Urgency				↑	0.94	↑	0.88
	Proximity				↑↓	0.66	↑↓	0.61
	Attractiveness				↓↑	-0.69	↓	-0.90
	Danger				↑	1.00	↑	0.72
	Mass				↓↑	-0.39	↓↑	-0.64
References indicated by numbers above:								
1. Walker (2002)				3. Edworthy et al. (1995)				
2. Edworthy et al. (2004)				4. Hellier & Edworthy (1989)				

Figure 1. Brief depiction of previous polarities and representative (or majority) exponents, plus polarities and majority exponents for the two experiments reported here. A single arrow pointing up or down indicates a unanimous or near-unanimous polarity. A pair of up and down arrows indicates a split polarity, with the arrow appearing first indicating the majority polarity. Thus, ↓↑ represents a mixed polarity, with the negative polarity being more common or majority (e.g. for size:frequency). An equal sign (=) indicates that the polarities were equal or very nearly so. The references cited provide pointers to previous studies with sufficiently similar methods to allow comparisons here. If more than one study is cited, the polarities and exponents depicted here may be an estimated average or representative value, for simplicity of display

Figure 1. It is at least possible to predict that the polarity preferred by the majority of participants in the present study (in comparable groups) would be the same as in Walker (2002, Experiment 2). In addition, there should be a significant correlation between the exponents obtained in the two studies. Since there are some small differences between that

study and the current one (participant pool, stimuli etc.), it is unlikely that there will be perfect agreement in the exponents. Most of the data-to-display mappings in Walker (2002) had not been explicitly studied elsewhere in the literature. A notable exception seems to be the concept of number of dollars, which Walker had found to produce split polarities for frequency and tempo, with many no-polarity responses for each. Those results suggested that there are different mental models for how sound should represent money. As an interesting comparison, in several cross-modal matching studies Galanter investigated perceptions of money (e.g. Galanter, 1962; all discussed in Galanter, 1974; Galanter & Pliner, 1974; Stevens & Galanter, 1957), leading to the conclusion that perceptions of *loss* and *gain* of money may differ. This subtlety could be reflected in different exponents, but not likely in different polarities between money (value) and acoustic dimensions.

In terms of the concept of urgency, one can predict a positive polarity for both frequency and tempo (e.g. Edworthy et al., 2004; Edworthy et al., 1995), although there does not seem to emerge a clear expected value for the magnitude estimation exponents. If tempo as defined here is considered analogous to the pulse repetition rate used by Edworthy and colleagues, then exponent values of about 0.61 from Hellier and Edworthy (1989; as cited by Hellier et al., 1995, p. 665) would probably be the most likely candidate value, given similarities to the methods and analysis used here.

The findings of Edworthy and colleagues (2004, table 4, 1995, in particular table 5 and figures 2–6) would suggest a positive polarity for danger:tempo and danger:frequency (note that they also found a corresponding negative polarity for safe:frequency). The concepts of mass and velocity here might have been predicted from Edworthy et al. (1995), though unfortunately neither the specific exponents nor regression coefficients were reported in that paper.

As for the data dimensions first appearing in the present study (e.g. attractiveness), no firm predictions can be made. The display dimension added in the current experiment, namely modulation index, or number of harmonics, at first seemed similar to the ‘inharmonicity’ manipulation that has been used by Edworthy and her colleagues. However, after informally listening to stimuli created in both manners, it was clear that the two sound dimensions were dissimilar enough to prevent predictions of one based on the other.

Analysis

Auditory stimuli: individual analyses of polarity

Previous research with conceptual data dimensions, mapped to both visual and auditory stimuli (e.g. Edworthy et al., 2004; Walker, 2000, 2002), has shown that there is often some variance within a group in terms of the *polarity* of the preferred mapping. Within a block, most individual participants apply a consistent mapping polarity (be it positive or negative), and make fairly monotonic responses, so that, for example, low frequencies are given lower numbers and higher frequencies are given higher numbers. In order to separate the positive polarity responses from the negative polarity responses in an algorithmic manner, three polarity categories have been defined: ‘positive’, ‘negative’, and ‘no’ polarity, as follows (see Walker, 2000, 2002 for more details). For each listener in each block the Pearson correlation coefficient is computed between the log of the responses (e.g. estimated temperature values) and the log of the actual stimulus values (e.g. frequencies). Data from a specific participant in a given block is considered to have ‘no’ polarity, and is not used in subsequent slope analyses, if the absolute value of the correlation coefficient in

that block does not reach conventional levels of statistical significance ($r_{\text{critical}} = 0.444$, $df = 18$, $\alpha = 0.05$). That is, there was no significant linear component to the relationship between the sound parameter and the data it was supposed to represent.

Auditory stimuli: aggregate analyses of slope

For each polarity, geometric means were calculated for all judgements of a given stimulus, collapsing across participants in a given data and display pair. These mean estimation values were plotted against the actual stimulus values in log–log coordinates, and fitted with a power function of the form $y = bx^m$. The exponent m , which is also the slope of the fit line, indicates how much the perceived, or estimated value changes as the actual stimulus parameter changes. For each data-to-display mapping a slope was determined (including both positive and negative slopes, where obtained), as well as the standard error of the slope estimate, the r^2 measure for each slope, and the number of listeners whose data contributed to each slope. If both polarities were obtained, the *majority polarity* was defined as greater than 50% of *all* participants in that block, including the ‘no’ polarity responses (see Walker, 2002). For example, if 11 participants responded with a positive polarity, and 10 responded with a negative polarity, then the positive polarity would be considered as the majority (11 out of 21). However, if there were also two ‘no’ polarities that would mean there was no majority for that block. The reasoning behind this approach is to attempt to determine a ‘preferred’ polarity among the group of participants whenever possible, while at the same time acknowledging that there may not be a unanimous polarity for a given data-to-display pairing, especially if the responses from many participants represented ‘no’ polarity. In the practical case of sonification design, an ambiguous or no-majority polarity situation would warrant careful consideration of another mapping, or might indicate the need to explicitly train listeners on how to interpret the display (should such training be effective).

Results and discussion

Perceptual dimensions: pitch, perceived tempo, and perceived brightness

The magnitude estimation method effectively yielded exponents relating three physical sound dimensions to 10 conceptual data dimensions. The results for each block are summarized in Table 1. As indicated in the three sections across the top of Table 1a, the majority polarity was positive for each of the perceptual dimensions, as would be expected. The slope of the regression line for estimations of pitch versus actual frequency for the majority was $m = 0.89$ ($SE_m = 0.08$; $r^2 = 0.95$). Previous pitch:frequency estimates that have yielded exponents slightly higher than the Mel scale, in the range of about 0.73–0.80 (Beck & Shaw, 1961, 1962, 1963; Walker, 2002). See, also, Stevens (1975) for details of the Mel scale. A confidence interval is a simple way to compare the single value obtained here to the variety of values that have previously been obtained through similar means. The 95% confidence interval for the regression slope is $m \pm t_{(\alpha/2, df)}SE$ where t is the tabled ratio at, in this case, $\alpha/2 = 0.025$ and $df = N - 2 = 8$, hence $t = 2.306$. SE is the standard error of the slope estimate (Pedhazur, 1997, p. 30). Note that the confidence interval depends on the number of points (and therefore the number of stimuli) used to estimate the regression slope in magnitude estimation. For pitch:frequency here, the interval is $0.71 = m - 1.07$, which indicates that the present results are in line with previous findings.

The slope for perceived tempo versus actual tempo for the majority positive polarity was $m = 0.85$ ($SE_m = 0.06$; $r^2 = 0.96$). For tempo:tempo, Eisler (1976) suggested an approximate value of 0.9, and Walker (2002) obtained an exponent of 0.95. Thus, the

Table 1. Magnitude estimation results for Experiment 1

	Frequency set				Tempo set				Modulation index set			
	<i>n</i>	<i>m</i>	<i>SE(m)</i>	<i>R</i> ²	<i>n</i>	<i>m</i>	<i>SE(m)</i>	<i>R</i> ²	<i>n</i>	<i>m</i>	<i>SE(m)</i>	<i>R</i> ²
a. Perceptual dimensions (pitch, perceived tempo and brightness for frequency, tempo, and modulation index, respectively)												
Perceptual dimension												
Positive polarity	16	0.89	0.08	0.94	15	0.85	0.06	0.96	18	0.83	0.07	0.95
'No' polarity	1				1				1			
Negative polarity	2	-0.74	0.14	0.79	3	-0.94	0.07	0.95	0	—	—	—
Total	19				19				19			
b. Conceptual dimensions												
Size												
Positive polarity	9	0.66	0.04	0.98	9	0.87	0.03	0.99	9	0.69	0.06	0.95
'No' polarity	0				1				1			
Negative polarity	10	-0.87	0.05	0.97	9	-0.77	0.07	0.94	9	-0.65	0.09	0.86
Total	19				19				19			
Temperature												
Positive polarity	11	0.65	0.06	0.93	12	0.63	0.04	0.96	14	0.62	0.03	0.98
'No' polarity	2				2				0			
Negative polarity	6	-0.59	0.05	0.94	5	-0.76	0.11	0.85	6	-0.48	0.07	0.84
Total	19				19				20			
Pressure												
Positive polarity	13	0.77	0.05	0.97	10	0.84	0.06	0.97	17	0.69	0.06	0.94
'No' polarity	3				3				0			
Negative polarity	3	-0.96	0.05	0.98	6	-0.74	0.17	0.71	2	-0.20	0.03	0.82
Total	19				19				19			
Velocity												
Positive polarity	18	0.77	0.06	0.96	17	0.90	0.03	0.99	17	0.67	0.04	0.98
'No' polarity	1				1				0			
Negative polarity	0	—	—	—	1	-1.29	0.13	0.92	2	-0.34	0.07	0.74
Total	19				19				19			
Num. Of Dollars												
Positive polarity	9	1.36	0.14	0.92	14	1.07	0.06	0.97	12	0.69	0.04	0.98
'No' polarity	2				2				1			
Negative polarity	8	-1.09	0.05	0.99	5	-1.12	0.03	0.99	6	-1.03	0.11	0.92
Total	19				21				19			
Urgency												
Positive polarity	14	0.70	0.03	0.98	17	0.64	0.02	0.99	19	0.94	0.05	0.98
'No' polarity	3				1				0			
Negative polarity	2	-0.61	0.10	0.82	1	-0.70	0.11	0.84	0	—	—	—
Total	19				19				19			
Proximity												
Positive polarity	10	0.78	0.10	0.89	8	0.51	0.03	0.98	12	0.66	0.05	0.96
'No' polarity	6				5				0			
Negative polarity	3	-0.61	0.11	0.79	6	-0.71	0.08	0.90	7	-0.36	0.07	0.78
Total	19				19				19			
Attractiveness												
Positive polarity	3	0.61	0.06	0.94	9	0.80	0.06	0.95	4	0.70	0.17	0.69
'No' polarity	8				5				2			
Negative polarity	8	-0.58	0.09	0.83	5	-0.66	0.06	0.93	13	-0.69	0.12	0.80
Total	19				19				19			

(Continues)

Table 1. (Continued)

	Frequency set				Tempo set				Modulation index set			
	<i>n</i>	<i>m</i>	<i>SE(m)</i>	<i>R</i> ²	<i>n</i>	<i>m</i>	<i>SE(m)</i>	<i>R</i> ²	<i>n</i>	<i>m</i>	<i>SE(m)</i>	<i>R</i> ²
Danger												
Positive polarity	10	0.69	0.02	0.99	14	0.61	0.02	0.99	17	1.00	0.07	0.96
'No' polarity	8				5				1			
Negative polarity	1	-0.47	0.07	0.86	0	—	—	—	1	-0.34	0.12	0.49
Total	19				19				19			
Mass												
Positive polarity	2	0.85	0.10	0.90	4	0.72	0.05	0.97	5	0.60	0.08	0.87
'No' polarity	0				5				4			
Negative polarity	17	-0.91	0.04	0.99	9	-0.71	0.05	0.96	10	-0.39	0.06	0.84
Total	19				18				19			

value obtained in the present study, with a confidence interval of $0.71 = m = 0.99$, is in line with the expected range.

For perceived brightness versus modulation index, the unanimously positive polarity slope was $m = 0.83$ ($SE_m = 0.07$; $r^2 = 0.95$). Since there is no referent to predict the slope for brightness:modulation index, the value of 0.83 can only be discussed in terms of it being less than 1.0, as is typical for many acoustic dimensions.

The presence of negative polarities for more than one participant in the pitch:frequency and tempo:tempo categories is unusual, and was not observed in previous studies. However, as mentioned earlier, psychophysical studies have often excluded such data as erroneous. That there was more than one participant responding in such an unusual manner warranted more scrutiny of the data. The listeners who responded with negative polarities in these groups did so with consistency throughout the block, as demonstrated by their significant (albeit negative) correlation coefficients. This also refutes any suggestions that participants are simply not able to discriminate between the sounds. They also did not exhibit any particular pattern of polarities in their subsequent blocks that were contrary to the group majorities. In sum, it remains mysterious as to why these participants responded so consistently and linearly, yet with negative polarities for the perceptual dimensions. It is interesting, although somewhat anecdotal, to note that at least one other researcher has occasionally encountered similar 'backward' responding for at least pitch:frequency, among musically naive listeners (J. Neuhoff, personal communication, 2003).

Conceptual dimensions

As indicated by the data in Table 1b, within nearly all blocks with conceptual data dimensions there were some participants who responded with positive, and some with negative polarities. The first comparison is between the present polarities and previously reported polarities. In all but two cases, for the frequency and tempo stimulus sets the majority polarity here corresponded to the majority polarity reported in Walker (2002). One of the differences was for the number of dollars:frequency group. In the present study, despite the slim nine to eight advantage for the positive polarity, the presence of two 'no' polarity responses meant that there was no majority polarity, by the established definition. Walker (2002) found six positive, five negative, and no 'no' polarity responses for the same dimensions, very similar in character to the present finding. In both cases there was a nearly even split among listeners, which has been explained in terms of the mental models the listeners invoke when interpreting the sounds. The other difference between the polarities

in the present study and the previous study is for the size:tempo mapping. In contrast to the previous unanimously positive polarity, the current results include an even split of nine positive and nine negative polarities, with one 'no' polarity.

To assess the overall stability of the exponent values for the data dimensions that appeared in both studies, the Pearson correlation coefficient was computed between the present study and the results of Walker (2002). Exponent values were included only if there were at least two participants responding with a given polarity, to ensure that the value was at least to some degree the result of a consensus, and not just one person's response. The exponents were highly correlated between the two studies, $r = 0.94$, $n = 19$, $p < 0.01$.

The next comparison is between the data dimensions that are 'new' to this particular study, but which have been studied in similar ways in the past. First, the polarity for urgency in relation to both frequency and tempo was positive, in agreement with the repeated findings by Edworthy and colleagues. For urgency:frequency the majority exponent of 0.70 is similar to the value of 0.61 from Hellier and Edworthy (1989; as cited by Hellier et al., 1995). Since the standard error of the exponent from the earlier study was not available, a confidence interval can suffice in comparing the two studies. The previous value of 0.61 falls just outside the 95% confidence interval of $0.63 \leq m \leq 0.77$ for urgency:frequency obtained in the present study. For urgency:modulation index, increases in modulation were unanimously reported here as indicating an increase in urgency, which may make that display dimension effective for both sonification and warning designs. The majority positive polarities for danger:frequency and danger:tempo were also in agreement with predictions. It is important to note that both perceived danger and perceived urgency increase as the frequency or tempo of a sound increases. If these dimensions had been in opposition (i.e. urgency increases but danger decreases when tempo increases) the dimensions may have been problematic for not only warning sound design, but also data sonification, given the obvious and frequent relation between the empirical danger of a situation and the urgency of a required response (for more on this, see Edworthy et al., 2004). The strongly positive polarity for danger:modulation index would point to modulation index as a potential warning sound dimension, and an effective sonification dimension. In fact, modulation index yielded results that were more nearly unanimous than either the Frequency Set or Tempo Set, suggesting that by this metric modulation index is even more effective in communicating levels or degrees of danger.

While most of the majority polarities were positive, there were some mappings where the majority response pattern was in the negative polarity. For example, 17 out of 19 participants in the mass:frequency block responded that increasing sound frequency corresponds to decreasing mass. This is, however, consistent with previous findings, and in line with predictions. For both the Tempo Set and Modulation Index Set there was, as well, a predominant negative polarity for mass, but the positive and 'no' polarities received more responses than there were for the Frequency Set. In terms of sonification interpretation, this makes frequency more reliable than tempo or modulation index for communicating the concept of mass.

The concept of size is typically closely related to the concept of mass, so one may have predicted results for size that are very similar to those for mass. However, for each audio dimension, the responses for size were nearly evenly split between positive and negative polarities. This suggests two (or more) equally valent ways of interpreting the sounds that are meant to represent size. As discussed by Walker (2002), some participants seem to consider a 'default' model of representing things, as in a typical visual graph: increases in the display dimension (i.e. height above the x-axis for graphs, or frequency for sounds)

mean increases in whatever it is that is being represented. Traditional visual graphs tend to follow this model, so some listeners assume the same must apply to auditory graphs. Other listeners tend to assume a more physical mental model, in which, in the case of size, bigger things make lower pitched sounds, leading to a negative polarity mapping. It makes sense, then, that other dimensions that could be construed to represent more 'of something', such as number of dollars, could be interpreted with this physical-model inspired negative polarity as well. Some concepts, such as mass, may be so valent that a greater number of participants automatically think in terms of the physical mental model, and relatively few fall back on the default positive polarity when it comes to sound. Similarly, Edworthy et al. (2004; see also Solomon, 1958; Solomon, 1959a,b) discuss that the preferred polarity may depend on whether the listener attributes changes in the sound to real changes in an object (a direct or 'denotative' meaning), as compared to the application of an intermediary mental model of how things work (a more 'connotative' meaning of the sound).

None of the foregoing seems to apply to the concept of attractiveness; thus it is difficult to explain the split, or even majority negative polarities for attractiveness in terms of any physical mental model. However, it is possible that the term attractiveness is strongly associated with aesthetic judgements and rarely considered as a data dimension, so it may be that some listeners simply made subjective assessments of the sounds themselves, despite the instructions to consider the sounds in terms of the data values they would represent. Similar issues would arise if one wanted to map the concept of attractiveness onto a visual display dimension such as line thickness, shape or colour.

EXPERIMENT 2

Many of the conceptual data dimensions investigated in Experiment 1, for example attractiveness, have not been studied before in this manner. In addition, there were some split and ambiguous polarities. Thus, as discussed in the opening section, it is important to determine how stable both the polarities and the exponents are under direct replication.

Method

Experiment 2 was conducted in the academic semester immediately following the previous study, for a separation of approximately 2 months. There were a total of 226 participants (122 males, 104 females, mean age = 20.9 years) from the same pool of undergraduates as Experiment 1. None of these participants had been in the earlier study. In Experiment 2, there were approximately 21 participants in each data-to-display pairing.

All stimuli, materials, scripts and apparatus were identical to Experiment 1. As a reminder, participants listened to sounds that varied along the acoustic dimensions of frequency, tempo or modulation index, and made magnitude estimations about data dimensions that the sounds were meant to represent. The data dimensions were the same as those in Experiment 1: size, temperature, pressure, velocity, number of dollars, urgency, proximity, attractiveness, danger and mass. The data were sorted and analysed as before, and Table 2 presents the results for slope, standard error and r^2 for each data-to-display mapping.

Table 2. Magnitude estimation results for Experiment 2

	Frequency set				Tempo set				Modulation index set			
	<i>n</i>	<i>m</i>	<i>SE(m)</i>	<i>R</i> ²	<i>n</i>	<i>m</i>	<i>SE(m)</i>	<i>R</i> ²	<i>n</i>	<i>m</i>	<i>SE(m)</i>	<i>R</i> ²
a. Perceptual dimensions (pitch, perceived tempo, and brightness for frequency, tempo, and modulation index, respectively)												
Perceptual dimension												
Positive polarity	20	0.79	0.07	0.94	21	0.83	0.05	0.97	18	0.61	0.02	0.99
'No' polarity	1				0				0			
Negative polarity	0	—	—	—	0	—	—	—	2	-0.58	0.09	0.85
Total	21				21				20			
b. Conceptual Dimensions												
Size												
Positive polarity	7	0.66	0.06	0.94	7	0.80	0.05	0.97	16	0.52	0.04	0.96
'No' polarity	3				4				1			
Negative polarity	11	-0.64	0.03	0.99	9	-0.58	0.08	0.88	3	-0.38	0.04	0.92
Total	21				20				20			
Temperature												
Positive polarity	17	0.56	0.04	0.96	13	0.45	0.02	0.99	18	0.74	0.03	0.99
'No' polarity	0				2				0			
Negative polarity	4	-0.73	0.07	0.94	5	-0.52	0.03	0.98	2	-2.03	0.55	0.63
Total	21				20				20			
Pressure												
Positive polarity	11	0.63	0.03	0.98	12	0.75	0.04	0.98	16	0.68	0.05	0.96
'No' polarity	2				2				1			
Negative polarity	7	-0.62	0.04	0.96	6	-0.48	0.04	0.95	3	-0.84	0.18	0.74
Total	20				20				20			
Velocity												
Positive polarity	17	0.82	0.06	0.96	17	0.83	0.04	0.98	21	0.82	0.05	0.97
'No' polarity	0				2				0			
Negative polarity	3	-0.99	0.15	0.84	1	-0.49	0.07	0.86	0	—	—	—
Total	20				20				21			
Num. Of Dollars												
Positive polarity	11	0.84	0.04	0.98	15	0.92	0.05	0.97	13	0.87	0.08	0.93
'No' polarity	1				3				2			
Negative polarity	8	-0.72	0.11	0.84	3	-0.98	0.06	0.97	6	-0.78	0.14	0.80
Total	20				21				21			
Urgency												
Positive polarity	14	0.68	0.03	0.99	16	0.66	0.02	0.99	18	0.88	0.07	0.96
'No' polarity	6				1				1			
Negative polarity	0	—	—	—	3	-0.40	0.05	0.91	1	-1.69	0.30	0.80
Total	20				20				20			
Proximity												
Positive polarity	10	0.54	0.05	0.94	11	0.60	0.03	0.98	13	0.61	0.05	0.96
'No' polarity	2				5				4			
Negative polarity	8	-0.65	0.04	0.97	5	-0.76	0.08	0.92	4	-0.51	0.05	0.92
Total	20				21				21			
Attractiveness												
Positive polarity	1	0.24	0.02	0.93	9	0.60	0.06	0.92	2	1.01	0.13	0.89
'No' polarity	2				7				1			
Negative polarity	17	-0.81	0.13	0.84	4	-0.79	0.16	0.76	17	-0.90	0.11	0.89
Total	20				20				20			

(Continues)

Table 2. (Continued)

	Frequency set				Tempo set				Modulation index set			
	<i>n</i>	<i>m</i>	<i>SE(m)</i>	<i>R</i> ²	<i>n</i>	<i>m</i>	<i>SE(m)</i>	<i>R</i> ²	<i>n</i>	<i>m</i>	<i>SE(m)</i>	<i>R</i> ²
Danger												
Positive polarity	12	0.49	0.02	0.99	16	0.51	0.04	0.96	17	0.72	0.05	0.97
'No' polarity	7				3				3			
Negative polarity	1	-0.94	0.15	0.83	3	-0.39	0.04	0.91	0	—	—	—
Total	20				22				20			
Mass												
Positive polarity	5	0.80	0.13	0.83	5	0.87	0.06	0.96	4	0.49	0.02	0.99
'No' polarity	1				3				0			
Negative polarity	14	-0.96	0.08	0.95	12	-0.99	0.06	0.97	17	-0.64	0.05	0.95
Total	20				20				21			

Results and discussion

Perceptual dimensions: pitch, perceived tempo, and perceived brightness

In this experiment, there were no negative polarity responses for either pitch:frequency or perceived tempo:tempo, but there were two participants who responded with a negative polarity for brightness:modulation index (see Table 2a). As before, it is not clear why this might be the case, though the brightness continuum is likely to be less familiar to listeners than pitch and perceived tempo. The majority slopes are of most interest, again. The slope of the regression line for estimations of pitch versus actual frequency was $m = 0.79$ ($SE_m = 0.07$; $r^2 = 0.94$), which is again in line with the expected range of 0.73–0.80, based on previous studies, including Experiment 1. The slope for perceived tempo versus actual tempo was $m = 0.83$ ($SE_m = 0.05$; $r^2 = 0.97$), which is very close to the slope value from Experiment 1 (0.85). The slope for the positive polarity responses for perceived brightness versus modulation index was $m = 0.61$ ($SE_m = 0.02$; $r^2 = 0.99$). Among the three perceptual dimensions, this brightness:modulation index exponent was the only one that was statistically different from the results of Experiment 1, $t(32) = (m_1 - m_2) / \sqrt{(SE_1^2 + SE_2^2)} = 3.022$, $p < 0.01$. These results suggest that whether a magnitude estimation exponent is consistent from one experiment to another may depend on the dimension itself, even for perceptual dimensions.

Conceptual dimensions

As indicated by the data in Table 2b, there were, again, both positive and negative polarity responses in nearly all blocks. The majority polarity was the same in all but a few cases, and those cases represent only minor differences. For example, the concept of number of dollars had a split polarity with a slight preference for the positive polarity with each of the stimulus sets in Experiment 1. However, the Frequency Set did not yield a majority polarity by the definition used here (nine positive, two 'no', and eight negative polarity responses). In Experiment 2, though, the split was 11, 1, and 8, which does yield a technical majority for the positive polarity.

Dimensions for which there were previous predictions of polarity, such as temperature and urgency, all yielded majority polarities in line with Experiment 1 and with the previous studies. Even the mappings that were split or somewhat ambiguous in Experiment 1, such as number of dollars and attractiveness, exhibited similar split results in Experiment 2. This

underlines the idea that there may be different stereotypes or mental models that cause opposite interpretations of those sounds. As a result, those kinds of data-to-display pairings should either be avoided, or, as discussed earlier, may necessitate some training of the listeners in order to 'align' their perceptions of the dimensions to the intended polarity held by the designer.

To assess the stability of the actual exponent values from Experiment 1 to Experiment 2, the overall Pearson correlation coefficient was calculated as before. The results from Experiments 1 and 2 were highly correlated, $r = 0.93$, $n = 57$, $p < 0.01$. Looking at the three display dimensions separately, the correlations between Experiments 1 and 2 for frequency, tempo and modulation index were $r = 0.97$, 0.98 and 0.88 , respectively. In the Modulation Index Set there was one value pair that seemed to contain an outlier. For temperature:modulation index in Experiment 2 the exponent was $m = -2.02$, $SE(m) = 0.55$, $n = 2$, and it arose because one participant responded with an extreme exponent of -3.5 (though still with a significant r , so the data were included in the analysis). In Experiment 1 the corresponding exponent was $m = -0.48$, $SE(m) = 0.07$, $n = 6$, so given the small sample sizes and large SE for the group in Experiment 2, the difference does not actually reach statistical significance, $t(4) = 0.99$, $p > 0.05$. Nevertheless, this points out how divergent results can be obtained when there are only a few respondents with a given polarity. Removing that one pair of exponents results in a correlation coefficient for modulation index of $r = 0.94$, and an overall correlation of $r = 0.96$, $n = 56$, $p < 0.01$ across all three stimulus sets. The exponents for Experiment 2 were also highly correlated with the findings from Walker (2002), $r = 0.93$, $n = 20$, $p < 0.01$, overall. Thus, the actual exponent values are very stable across all three studies.

GENERAL DISCUSSION

Magnitude estimation has again proven useful in determining the relationship between physical sound parameters such as frequency, tempo and modulation index, and a variety of conceptual data dimensions. Both the preferred mapping polarity and the function relating physical sound changes to conceptual data changes (i.e. slope or exponent) emerge from this approach. The determination of majority polarities for a given data-to-display mapping is an important component in display design, and can also provide insight into the kinds of mental models that listeners employ when interpreting what they hear. As pointed out by Walker (2002), mappings that result in unanimous polarities should be more effective in displays and warnings, since there is no ambiguity over the data being represented. For example, if all users of an environmental monitoring system interpret rising pitch as rising temperature, then the correct actions required when the pitch of the auditory display increases is not going to be in doubt. Several of the mappings studied here, including, for example, velocity:frequency and urgency:modulation index, elicited unanimous or near-unanimous polarities, and so they should be consistently interpreted by listeners in terms of changes in a represented data set.

For most mappings the positive polarity was dominant, but some mappings elicited negative polarities from the majority of listeners. The fact that negative polarities can dominate a data-to-display mapping in an auditory display is interesting in itself. This generally does not happen in visual displays, due to the prevalence of a Cartesian model for mapping data onto visual displays, such that 'more means more' and 'more means farther from the axis'. Clearly, then, auditory display designers have another parameter (polarity)

to consider when constructing sonifications and auditory graphs. Adding to the challenge, in some cases the majority polarity could not be predicted in advance. This echoes Walker and Kramer's (2005) results in which even expert sound designers were not able to predict perfectly which sound mappings would lead to better speed or accuracy in a crystal factory monitoring sonification. Taken together, these results demonstrate the need for empirical studies to determine preferred polarities.

In addition to the possibility of either a positive or negative majority polarity, a mapping can lead to a split polarity, presumably because different listeners use different internal analogies or mental models to determine what a particular sound change 'means'. Such mappings should be avoided in displays, to avoid a mismatch between the designer's intention and the listeners' interpretations. Consider a sonification intended for stock price monitoring: there would be serious problems if half the trading floor interpreted rising pitch as rising stock price (buy!) and the other half interpreted the same sounds as declining stock price (sell!). Whether such mental models are flexible and can be modified by training or instruction remains to be determined (but see Smith & Walker, 2005; Walker & Nees, 2005a,b for initial research in that area).

The magnitude estimation procedure also can determine the distinct exponents relating changes in the sounds to perceived changes in data values. This is, of course, crucial in sonification, and can be useful in other display design, such as in developing warnings that communicate increasing levels of urgency. As has been found in the past, the specific pairing of both data and display dimensions needs to be considered, since the exponents can differ across mappings.

The replication from Experiment 1 to 2 showed that within a mapping, polarities were very consistent and stable between studies, even down to the level of ambiguities and split polarities. This suggests that once a mapping has been assessed for its dominant polarity, that information can be reliably utilized by display designers. In addition to the polarities being stable, the exponents obtained in the two studies here were highly correlated ($r = 0.96$), which indicates that the relative ordering of the exponents also remains stable. However the exact values can and do differ between the studies for some dimensions. Thus, additional replications may be useful in further establishing the stability and predictability of these exponents, and in narrowing in to a 'suggested' value, as Stevens did for several dimensions only after many years of investigations (Stevens, 1975). It is important not to over-interpret the exact exponent values reported in a single study (including here), to the point where they are considered *the* correct values for use in all cases. As with any performance data that are used to drive interface guidelines, care must always be taken to avoid the treating the numbers as components of a design recipe. Rather, they should be treated as guidance, at least until repeated measurements converge towards a clear value or range.

The studies reported here also make an important contribution in terms of methodology. Stevens (1975) suggests that about 10 participants are sufficient to obtain reliable magnitude estimation exponents. However, it is clear that when conceptual data dimensions are under investigation, more like twice that number seems to be appropriate. That way, even with a split polarity in the responses, a sufficient number of responses will be available to determine a group magnitude estimation exponent.

In summary, the magnitude estimation results presented here, as well as the methods used to obtain them, have important theoretical and applied implications. In particular, perception and cognition researchers may now have a greater understanding of how listeners interpret what they hear, and sonification designers will be able to make more

effective and comprehensible displays. This represents an excellent example of cognitive psychology being applied to make dramatic improvements in the design of interfaces that have real everyday utility, and that can change the lives of many researchers, scientists and others for whom visual displays are not available, not accessible or not sufficient.

REFERENCES

- Beck, J., & Shaw, W. A. (1961). The scaling of pitch by the method of magnitude-estimation. *American Journal of Psychology*, *74*, 242–251.
- Beck, J., & Shaw, W. A. (1962). Magnitude estimations of pitch. *Journal of the Acoustical Society of America*, *34*(1), 92–98.
- Beck, J., & Shaw, W. A. (1963). Single estimates of pitch magnitudes. *Journal of the Acoustical Society of America*, *35*(11), 1722–1724.
- Dawson, W. E., & Brinkler, R. P. (1971). Validation of ratio scales of opinion by multimodality matching. *Perception & Psychophysics*, *9*(5), 413–417.
- Edworthy, J., Hellier, E. J., Aldrich, K., & Loxley, S. (2004). Designing trend-monitoring sounds for helicopters: Methodological issues and an application. *Journal of Experimental Psychology: Applied*, *10*(4), 203–218.
- Edworthy, J., Hellier, E. J., & Hards, R. (1995). The semantic associations of acoustic parameters commonly used in the design of auditory information and warning signals. *Ergonomics*, *38*(11), 2341–2361.
- Edworthy, J., Loxley, S., & Dennis, I. (1991). Improving auditory warning design: Relationship between warning sound parameters and perceived urgency. *Human Factors*, *33*(2), 205–231.
- Eisler, H. (1976). Experiments on subjective duration 1868–1975: A collection of power function experiments. *Psychological Bulletin*, *83*, 1154–1171.
- Engen, T. (1971). Scaling methods. In J. Kling, & L. Riggs (Eds.), *Experimental psychology* (pp. 45–86). London: Methuen.
- Fitch, W. T., & Kramer, G. (1994). Sonifying the body electric: Superiority of an auditory over a visual display in a complex, multi-variate system. In G. Kramer (Ed.), *Auditory display: Sonification, audification, and auditory interfaces* (pp. 307–326). Reading, MA: Addison-Wesley.
- Flowers, J. H., & Grafel, D. C. (2002). Perception of sonified daily weather records. *Proceedings of the Annual Meeting of the Human Factors and Ergonomics Society*, Baltimore, Maryland.
- Galanter, E. (1962). The direct measurement of utility and subjective probability. *American Journal of Psychology*, *75*, 208–220.
- Galanter, E. (1974). Psychological decision mechanisms and perception. In E. Carterette, & M. P. Friedman (Eds.), *Handbook of perception II: Psychophysical judgement and measurement* (pp. 85–126). New York: Academic Press.
- Galanter, E., & Pliner, P. (1974). Cross-modality matching of money against other continua. In H. R. Moskowitz, B. Sharf, & J. C. Stevens (Eds.), *Sensation and measurement: Papers in honor of S. S. Stevens*. Dordrecht, Netherlands: Reidel.
- Gardner, J. A. (1999). The Accessible Graphing Calculator: A Self-voicing Graphing Scientific Calculator for Windows. Retrieved November, 2004, from <http://dots.physics.orst.edu/calculator/>
- Gardner, J. A., Lundquist, R., & Sahyun, S. (1996). TRIANGLE: A practical application of non-speech audio for imparting information. *Proceedings of the International Conference on Auditory Display*, San Francisco, CA; pp. 59–60.
- Guillaume, A., Pellioux, L., Chastres, V., & Drake, C. (2003). Judging the urgency of nonvocal auditory warning signals: Perceptual and cognitive processes. *Journal of Experimental Psychology: Applied*, *9*(3), 196–212.
- Hellier, E. J., & Edworthy, J. (1989). Quantifying the perceived urgency of auditory warnings. *Canadian Acoustics*, *17*(4), 3–11.
- Hellier, E. J., Edworthy, J., & Dennis, I. (1993). Improving auditory warning design: Quantifying and predicting the effects of different warning parameters on perceived urgency. *Human Factors*, *35*, 693–706.

- Hellier, E. J., Edworthy, J., & Dennis, I. (1995). A comparison of different techniques for scaling perceived urgency. *Ergonomics*, *38*, 659–670.
- Hellier, E. J., Edworthy, J., Weedon, B., Walters, K., & Adams, A. (2002). The perceived urgency of speech warnings: Semantics versus acoustics. *Human Factors*, *44*(1), 1–17.
- Hellier, E. J., Wright, D. B., Edworthy, J., & Newstead, S. (2000). On the stability of the arousal strength of warning signal words. *Applied Cognitive Psychology*, *14*, 577–592.
- Kramer, G., Walker, B. N., Bonebright, T., Cook, P., Flowers, J., Miner, N., et al. (1999). The Sonification Report: Status of the Field and Research Agenda. Report prepared for the National Science Foundation by members of the International Community for Auditory Display. Santa Fe, NM: International Community for Auditory Display (ICAD).
- Martins, A. C. G., & Rangayyan, R. M. (1997). Experimental evaluation of auditory display and sonification of textured images. *Proceedings of the Fourth International Conference on Auditory Display (ICAD97)*, Palo Alto, CA; pp. 129–134.
- Meijer, P. (2004). The vOICe Accessible Graphing Calculator. Retrieved November, 2004, from <http://www.seeingwithsound.com/winmath.htm>
- Nesbitt, K. V., & Barrass, S. (2002). Evaluation of a multimodal sonification and visualization of depth of market stock data. *Proceedings of the International Conference on Auditory Display (ICAD2002)*, Kyoto, Japan; pp. 233–238.
- Pedhazur, E. J. (1997). *Multiple regression in behavioral research: Explanation and prediction* (3rd ed.). New York: Wadsworth Thompson Learning.
- Sanders, M. S., & McCormick, E. J. (1993). *Human factors in engineering and design* (7th ed.). New York: McGraw-Hill.
- Smith, D. R., & Walker, B. N. (2005). Effects of auditory context cues and training on performance of a point estimation sonification task. *Applied Cognitive Psychology*, *19*(8), 1065–1087.
- Solomon, L. N. (1958). Semantic approach to the perception of complex sounds. *Journal of the Acoustical Society of America*, *30*, 421–425.
- Solomon, L. N. (1959a). Search for physical correlates to psychological dimensions of sounds. *Journal of the Acoustical Society of America*, *31*, 492–497.
- Solomon, L. N. (1959b). Semantic reactions to systematically varied sounds. *Journal of the Acoustical Society of America*, *31*, 986–990.
- Stevens, S. S. (1966). A metric for the social consensus. *Science*, *151*, 530–541.
- Stevens, S. S. (1969). On predicting exponents for cross-modal matches. *Perception & Psychophysics*, *6*, 251–256.
- Stevens, S. S. (1975). *Psychophysics: Introduction to its perceptual, neural, and social prospects*. New York: Wiley.
- Stevens, S. S., & Galanter, E. (1957). Ratio scales and category scales for a dozen perceptual continua. *Journal of Experimental Psychology*, *54*, 377–411.
- Teghtsoonian, M. (1980). Children's scales of length and ludness: A developmental application of cross-modal matching. *Journal of Experimental Child Psychology*, *30*, 290–307.
- Teghtsoonian, M., & Teghtsoonian, R. (1971). How repeatable are Stevens's power law exponents for individual subjects? *Perception & Psychophysics*, *10*(3), 147–149.
- Teghtsoonian, M., & Teghtsoonian, R. (1983). Consistency of individual exponents in cross-modal matching. *Perception & Psychophysics*, *33*(3), 203–214.
- Teghtsoonian, R., & Teghtsoonian, M. (1978). Range and regression effects in magnitude scaling. *Perception & Psychophysics*, *24*(4), 305–314.
- Teghtsoonian, R., & Teghtsoonian, M. (1997). Range of acceptable stimulus intensities: An estimator of dynamic range for intensive perceptual continua. *Perception & Psychophysics*, *59*(5), 721–728.
- Valenzuela, M. L., Sansalone, M. J., Krumhansl, C. L., & Streett, W. B. (1997). Use of sound for the interpretation of impact-echo signals. *Proceedings of the Fourth International Conference on Auditory Display (ICAD97)*, Palo Alto, CA; pp. 47–56.
- Walker, B. N. (2000). Magnitude estimation of conceptual data dimensions for use in sonification. Unpublished Ph.D. Dissertation, Rice University, Houston, TX.
- Walker, B. N. (2002). Magnitude estimation of conceptual data dimensions for use in sonification. *Journal of Experimental Psychology: Applied*, *8*(4), 211–221.
- Walker, B. N., & Cothran, J. T. (2003). Sonification Sandbox: A graphical toolkit for auditory graphs. *Proceedings of the International Conference on Auditory Display (ICAD2003)*, Boston, MA; pp. 161–163.

- Walker, B. N., & Kramer, G. (2004). Ecological psychoacoustics and auditory displays: Hearing, grouping, and meaning making. In J. G. Neuhoff (Ed.), *Ecological psychoacoustics* (pp. 150–175). New York: Academic Press.
- Walker, B. N., & Kramer, G. (2005). Mappings and metaphors in auditory displays: An experimental assessment. *ACM Transactions on Applied Perception*, 2(4), 407–412.
- Walker, B. N., & Lane, D. M. (2001). Psychophysical scaling of sonification mappings: A comparison of visually impaired and sighted listeners. *Proceedings of the International Conference on Auditory Display (ICAD2001)*, Espoo, Finland; pp. 90–94.
- Walker, B. N., & Lowey, M. (2004). Sonification Sandbox: A graphical toolkit for auditory graphs. *Proceedings of the Rehabilitation Engineering & Assistive Technology Society of America (RESNA) 27th International Conference*, Orlando, FL.
- Walker, B. N., & Nees, M. A. (2005a). Brief training for performance of a point estimation sonification task. *Proceedings of the International Conference on Auditory Display (ICAD2005)*, Limerick, Ireland.
- Walker, B. N., & Nees, M. A. (2005b). Conceptual versus perceptual training for auditory graphs. *Proceedings of the Annual Meeting of the Human Factors and Ergonomics Society (HFES2005)*, Orlando, FL; pp. 1598–1601.
- Wickens, C. D., Lee, J. D., Liu, Y., & Gordon Becker, S. E. (2004). *An introduction to human factors engineering* (2nd ed.). Upper Saddle River, NJ: Pearson Prentice Hall.