AUDITORY GRAPHS NEED ERROR BARS: VALIDATING ERROR-TO-SOUND MAPPINGS AND SCALINGS

Jared M. Batterman
Georgia Institute of Technology
School of Psychology
Atlanta, GA, USA
jmbatterman@gatech.edu

Bruce N. Walker
Georgia Institute of Technology
School of Psychology
Atlanta, GA, USA
bruce.walker@psych.gatech.edu

ABSTRACT

Regardless of presentation modality (visual, auditory, etc.), graphs often need to convey uncertainty about the data values. Visual graphs may use error bars or confidence intervals. How to convey uncertainty about the data in an *auditory* graph remains an issue. Fundamental questions need to be answered first, such as how much change in, say, pitch, represents a given level of uncertainty. Here we present the results of a study that validates the uncertainty-to-sound-parameter mappings that had been determined in a previous phase of the research. That earlier research utilized conceptual magnitude estimation to identify preferred auditory mappings and scalings for error and uncertainty [1]. The present study sought to evaluate and validate those mappings and scalings in an applied context. Participants listened to pairs of auditory stimuli, and reported which they felt more closely represented a given range of error. Results showed that participants selected the error scalings that had been identified in the prior research, over other higher or lower slope values. This supports the validity of the mappings and scalings, and also provides more support for the utility of the conceptual magnitude estimation procedure developed by Walker [2].

1. INTRODUCTION

Data comprehension and statistical knowledge is becoming increasingly important in today's data-rich world, and it is simply crucial if one wishes to pursue any sort of scientific career. Presenting data in a non-numerical summary format, such as a visual graph, is a staple of modern scientific communication. For students or scientists with visual impairment, or indeed anyone who is unable to see a visual graph, *auditory* graphs have been developed to help make data more accessible [3] [4]. In any graph, regardless of modality (visual, auditory, tactile, etc.), scientific rigor often requires there to be a representation of any uncertainty about the data values. Visual graphs may use error bars or confidence intervals, for example. How to convey uncertainty about the data in an auditory graph remains an under-studied problem. Unfortunately, this has meant that auditory error bars (and similar representations of uncertainty) have largely been omitted from sonifications and auditory graphs (but see [5]). Fundamental issues abound, such as how much change in, say, pitch, represents a given level of uncertainty. Here we present the results of a study that validates the uncertainty-to-sound-parameter mappings that were determined in a previous phase of the research. This is foundational to being able to design effective auditory error bars or confidence intervals, as part of effective auditory graphs.

To begin to understand such auditory representations, our earlier phase of work utilized conceptual magnitude estimation in an attempt to develop optimal auditory mappings and scalings for error and uncertainty [1]. Whereas other conceptual magnitude estimation studies [e.g., 2] have used musical notes or pure tones, our recent work has utilized stimuli made from bandpass-filtered resonant noise [1]. This was because we anticipate that these mappings may one day be used in a wide range of tools and devices, many of which utilize notes or pure tones for other purposes. For example, one of the most common data-to-sound mapping frameworks for auditory graphs varies pitch as the y-axis data changes, and uses time to represent the x-axis. Existing software such as the Sonification Sandbox (which makes use of this mapping framework) would be able to easily incorporate these noise-based error-bar mappings without making any major changes to its existing structure [6].

Results of our previous study showed that participants had strong positive tempo mappings for both error and uncertainty. That is, for the majority of participants, as the tempo of a sound set increased, so too did their perception of how much error or uncertainty they felt it represented. However, the results for the frequency dimension were more mixed (in order to give white noise a specific frequency, band pass filters were built around specific central frequencies with a 6 dB decay per octave). For the error dimension, 9 participants utilized a "positive" mapping (i.e. higher frequencies represented more error) and 5 utilized a "negative" mapping (i.e. lower frequencies represented more error) (see [2] for more on these definitions). This demonstrated a slight preference for a positive mapping, with both mappings being strong fits and possessing r^2 values of over 0.90. For the uncertainty dimension, the preferences were reversed, with 7 participants preferring the negative mapping and 4 preferring the positive one. Furthermore, though the negative mapping was preferred, it produced a less precise mapping than its positive counterpart, yielding an r^2 of 0.88, lower than the 0.94 yielded by the positive mapping, despite drawing its data from a larger number of participants.

In order to determine the validity and usefulness of these mappings, in the research we report on here we adapted the slope validation procedure utilized by Walker [2]. All mappings (both positive and negative) were tested for *frequency* in this new experiment; however, for *tempo* only the positive mappings warranted testing, due to their preference by a clear majority of participants.

2. METHODS

2.1 PARTICIPANTS

Twenty-six Georgia Tech undergraduates participated for extra credit in introductory psychology courses.

2.2 SOUND STIMULI

There were two sets of stimuli, a frequency set and a tempo set. The three stimuli in the frequency set each consisted of a series of five 1 second long segments of bandpass-filtered resonant noise of set ascending central frequencies (each with a 6 dB decay per octave) separated by 0.25 seconds of silence. The central frequency for the first sound of each frequency stimulus was 200 Hz, increasing over five equally spaced steps to final central frequencies of 400, 600, and 1000 Hz (central frequencies adapted from the pure tone frequencies utilized by Walker in his initial studies) [2]. Thus, the three stimuli in the frequency set were as follows: F1 consisted of central frequency steps of 200, 250, 300, 350, and 400 Hz, F2 consisted of steps of 200, 300, 400, 500, and 600 Hz, and F3 consisted of steps of 200, 400, 600, 800, and 1000 Hz.

The three stimuli in the tempo set were constructed in a similar manner with each stimulus having a starting tempo of 60 bpm and taking equally spaced steps to final tempos of 120, 180, and 300 bpm. Each step consisted of white noise pulsing in an on/off pattern for 2 seconds per step and then proceeding on to the next step in the set. The tempo set stimuli were as follows: T1 consisted of steps of 60, 75, 90, 105, and 120 bpm, T2 consisted of steps of 60, 90, 120, 150, and 180 bpm, and T3 consisted of steps of 60, 120, 180, 240, and 300 bpm.

All sound stimuli were generated by Audacity v.1.3.14, with filters created by the LS Filter plugin. Each step in each stimulus set was equated for equal loudness using the ReplayGain plugin.

2.3 VALUE RANGES

The purpose of Walker's (2002) procedure is to demonstrate the validity of the mappings generated from conceptual magnitude estimation by taking value ranges generated from those mappings, and comparing them to value ranges that would not be ideal were the mappings accurate. To do this, ideal final values (initial values were the same for each trial) for both error and uncertainty need to be generated for each of the six stimuli (F1, F2, F3, T1, T2, and T3). To do this, the following equation was used:

$$FV = IV \cdot (S_5/S_1)^m, \quad (1)$$

In Equation (1), FV represents the 'Final Value' being calculated and IV represents the 'Initial Value' for each of the value ranges. IV stays the same for every pair, and in this study IV was always equal to 100. S_5 was the final step for a given stimulus and S_1 was the initial step for a given stimulus. So, for the tempo set, S_1 was always 60 (representing the starting point of each tempo stimulus at 60 bpm) and for the frequency set S_1 was always 200 (representing the starting point of each frequency stimulus at 200 Hz). 'm' is the slope of the ideal mapping equation for a given data dimension generated based on data collected in our past work [2].

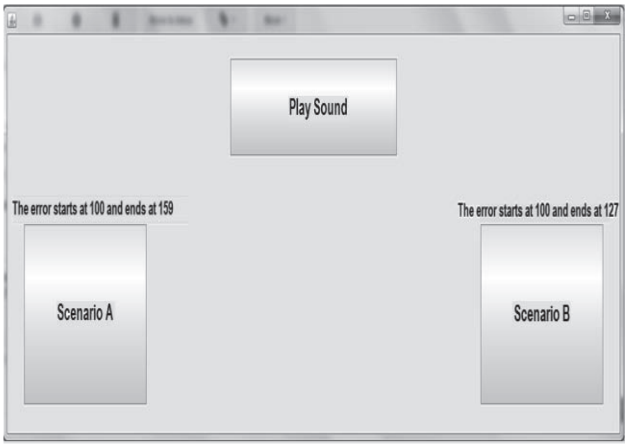

Figure 1: This is an image of the interface utilized by participants for a given trial. Clicking the 'Play Sound' button would play the sound stimulus for a given a trial, while clicking either the 'Scenario A' or 'Scenario B' buttons would record the participants' choice of that value range as the best fit for the given sound stimulus. This figure is a screenshot taken of a single trial in the error tempo block. The value range located above the 'Scenario A' button is the ideal, having been generated by Equation (1). The value range above the 'Scenario B' button goes from 100 to 127, which is 0.8 times the ideal value of 159.

For example, the equation to calculate the ideal FV for the positive frequency mapping for uncertainty for stimuli F1 would be as follows:

$$FV = 100 \cdot (400/200)^{0.5343}, \quad (2)$$
$$FV = 145$$

So, if our mappings are accurate, the correct value range that listeners perceive for uncertainty for F1 should go from 100 to 145 (all FVs were rounded off to whole numbers). In order to generate the incorrect value ranges that this ideal range would be compared to, FVs were multiplied by 0.8 and 1.2. Staying with the current example, this would generate incorrect FVs of 116 and 174. If our mappings are correct, participants should listen to F1, be told it represents uncertainty, and select a range of 100 to 145 significantly more often than any of the 'incorrect' values.

2.4 DESIGN

The study was conducted entirely within-subjects with each participant receiving 6 trial blocks; 3 blocks utilized uncertainty as the conceptual dimension and 3 utilized error. The blocks were presented in random order and each block contained every possible pairing of F1, F2, and F3 or T1, T2, and T3 (depending on whether it was a tempo or frequency block) with each possible coupling of their ideal value range and the two incorrect ranges. This yielded a total of 18 trials per block, with 12 having a correct answer (meaning the ideal value range was present as an option) and 6 being 'foil' trials where participants only had the choice of either of the two incorrect value ranges. Within each block, the order of trials was also randomized.

Frequency Set	All Data	Error Pos	Error Neg	Uncertain Pos	Uncertain Neg
Proportion Correct	0.601	0.660	0.612	0.596	0.5353
Variance	0.015	0.017	0.011	0.014	0.013
Participant N	104	26	26	26	26
Value of *t*	8.316*	6.338*	5.434*	4.186*	1.551
Value of *p*	0.0001	0.0001	0.0001	0.0001	0.133

Table 1: This table displays the sum of the data utilizing frequency as a display dimension. Asterisks indicate statistical significance at the $p<0.01$ level.

Frequency Set	All Data	Error Pos	Uncertain Pos
Proportion Correct	0.601	0.619	0.5833
Variance	0.017	0.013	0.021
Participant N	52	26	26
Value of *t*	5.587*	5.217*	2.964*
Value of *p*	0.0001	0.0001	0.007

Table 2: This table displays the sum of the data utilizing tempo as a display dimension. Asterisks indicate statistical significance at the $p<0.01$ level.

2.5 TRIAL STRUCTURE & TASK

Upon their arrival, participants were given oral instructions as to the purpose of the study and the use of the interface (see Figure 1). They were also given the following written instructions (adapted from Walker [2]):

You are going to be listening to several different sounds meant to represent either error or uncertainty. Your task is to indicate which of the two data descriptions the sound best represents. For each trial, you can listen to the sound by clicking on the 'Play Sound' button located in the top center region of the display. To select which data description you feel fits the sound best, please click on the button below description labeled either 'Scenario A' or 'Scenario B'. At the end of a block of trials, a message will appear asking you if you would like to continue to the next block. At this point, feel free to ask the experimenter any additional questions you have, and/or take a short break. When you are ready, please click 'Yes' to go on to the next block of trials. After you have finished all 6 trial blocks, the program will exit. At this point, please see the experimenter for further instructions.

After all trial blocks were completed, participants were given a short demographics survey which also inquired about any past musical training or professional musical experience they may have.

3. RESULTS

For each participant, performance scores were calculated for every trial block. These scores consisted of the number of times they chose the ideal value range divided by 12(the number of times it was presented in a block). The grand mean was then calculated for all performance scores within each display dimension and was compared to what would be expected by chance (i.e. 50% or 0.5). For the 4 blocks where frequency was the display dimension, the participants selected the ideal or 'correct' mapping significantly more than chance [score=0.601, SD=0.1238; $t(103)=8.316$, $p<0.0001$]. For the 2 blocks where tempo was the display dimension, this was also the case [score=0.601, SD=0.1303; $t(51)=5.587$, $p<0.0001$]. Following this, each individual block was tested against chance, with all of their means being greater than chance, and with all of them being statistically significant, with the exception of the negative frequency mapping for uncertainty (results summarized in Tables 1 and 2). Furthermore, there were also no effects of age, gender, or musical training.

Performance on the 'foil' trials was also analyzed. These were the trials that contained only 'incorrect' value ranges. There was no reason to think that either of these values would perform any different from chance and subsequent analyses showed this to be the case.

4. DISCUSSION

These results not only demonstrate that our mappings for error and uncertainty hold up in an applied context, but they also provide further support for the methodology proposed by Walker to both develop and test optimal sonifications [2]. This is also the first time this methodology has been utilized using stimuli composed of anything other than pure tones. It is not entirely clear why the negative frequency mappings for uncertainty were not significant, but it is possible that this was due to it having the lowest r^2 value of any of the other mappings tested (being the only mapping with an r^2 below 0.90). However, regardless of the reasoning, it does suggest that when we apply these mappings in the future, we may want to avoid utilizing a negative frequency mapping for uncertainty in favor of either a positive tempo or frequency mapping.

5. CONCLUSION

Now that these mappings have been validated, the next step is to take them and apply them directly to an auditory graphs context in the form of an auditory equivalent to standard error bars and confidence intervals. In addition, these mappings could also benefit current work focusing on the sonification of future climate data and spatial locations by providing them with a scaling function that directly relates user perception of uncertainty to variations in frequency and tempo [7] [8].

It is often the case (as in [7] and [8]) that decisions regarding the creation and application of sonifications are made arbitrarily [9] [10] [11]. Unfortunately, this limits their utility outside of the specific parameters of a given study [11]. In order to ensure that sonifications are designed in a way that makes them both more intuitive and more generalizable, it makes sense to use a standardized procedure. Walker's [2] method has been shown to develop reliable scaling functions that hold up in various applied contexts (including the current study) and would thus be a prime candidate for a 'standard' way to create optimal sonifications [1] [2] [3].

6. REFERENCES

[1] Batterman, J. M., & Walker, B. N. (2012). Displaying Error & Uncertainty in Auditory Graphs. *Proceedings of the 14th International ACM SIGACCESS Conference on Computers and Accessibility (ASSETS2012)*, Boulder, CO (22-24 October, 2012).

[2] Walker, B. N. (2002). Magnitude estimation of conceptual data dimensions for use in sonification. Journal of Experimental Psychology: Applied, 8, 211-221. DOI: 10.1037/1076-898X.8.4.211

[3] Walker, B. N. & Mauney, L. (2010). Universal Design of Auditory Graphs: A Comparison of Sonification Mappings for Visually Impaired and Sighted Listeners. *ACM Transactions on Accessible Computing 2, 3*. Article 12 (March 2010), 16 pages. DOI = 10.1145/1714458.1714459. http://doi.acm.org/10.1145/1714458.1714459

[4] Flowers, J. H. (2005). Thirteen Years of Reflection on Auditory Graphing: Promises, Pitfalls, and Potential New Directions. Proceedings of ICAD 05-Eleventh Meeting of the International Conference on Auditory Display, Limerick, Ireland, July 6-9, 2005, pp. 406-409.

[5] Peres, C., & Lane, D. (2003). Sonification of Statistical graphs, *Proceedings of the Ninth International Conference on Auditory Display*, *9*, 157-160.

[6] Walker, B. N., & Cothran, J. T. (2003). Sonification Sandbox: A graphical toolkit for auditory graphs. Proceedings of the Ninth International Conference on Auditory Display ICAD 2003, Boston, MA (6-9 July) pp. 161-163.

[7] Brown, E. & Bearman, N. (2012). Listening to uncertainty: Information that sings, Significance, 9(5) p. 14-17. DOI:10.1111/j.1740-9713.2012.00601.x *Runner-up in Young Statisticians Group/Significance writing competition*

[8] Bearman, N., Jones, P. & Lovett, A. (2011). Using Sound to Represent Uncertainty in UK Climate Predictions 2009 Data: Communication of Uncertainty, From Climate Research to Climate Services workshop, Presented 29/06/2011 at Royal Meteorological Society Conference, Exeter, UK.

[9] Edworthy, J. (1998). Does sound help us to work better with machines? A commentary on Rautenberg's paper 'About the importance of auditory alarms during the operation of a plant simulator', Interacting with Computers, vol. 10, pp. 401-409.

[10] A. Pirhonen, E. Murphy, G. McAllister, and W. Yu, "Non-speech sounds as elements of a use scenario: A semiotic perspective," presented at 12th International Conference on Auditory Display (ICAD06), London, UK, 2006

[11] Nees, M. A., & Walker, B. N. (2007). Listener, task, and auditory graph: Toward a conceptual model of auditory graph comprehension. *Proceedings of the International Conference on Auditory Display (ICAD 2007)*, Montreal, Canada (26-29 June). pp. 266-273.