Cartography and Geographic Information Science

ISSN: 1523-0406 (Print) 1545-0465 (Online) Journal homepage: http://www.tandfonline.com/loi/tcag20

Sonifying data uncertainty with sound dimensions

Andrea Ballatore, David Gordon & Alexander P. Boone

To cite this article: Andrea Ballatore, David Gordon & Alexander P. Boone (2018): Sonifying data uncertainty with sound dimensions, Cartography and Geographic Information Science

To link to this article: https://doi.org/10.1080/15230406.2018.1495103

CARTOGRAPHY AND GEOGRAPHIC INFORMATION SCIENCE
https://doi.org/10.1080/15230406.2018.1495103

Taylor & Francis
Taylor & Francis Group

ARTICLE

Sonifying data uncertainty with sound dimensions

Andrea Ballatore [a], David Gordon [b] and Alexander P. Boone [c]

[a]Department of Geography, Birkbeck, University of London, UK; [b]Department of Music, University of California, Santa Barbara, CA, USA; [c]Department of Psychological and Brain Sciences, University of California, Santa Barbara, CA, USA

ABSTRACT
The communication of data uncertainty is a crucial problem in data science, information visualization, and geographic information science (GIScience). Effective ways to communicate the uncertainty of data enables data consumers to interpret the data as intended by the producer, reducing the possibilities of misinterpretation. In this article, we report on an empirical investigation of how sound can be used to convey information about data uncertainty in an intuitive way. To answer the research question *How intuitive are sound dimensions to communicate uncertainty?* we carry out a cognitive experiment, where participants were asked to interpret the certainty/uncertainty level in two sounds A and B (N = 33). We produce sound stimuli by varying sound dimensions, including loudness, duration, location, pitch, register, attack, decay, rate of change, noise, timbre, clarity, order, and harmony. In the stimuli, both synthetic and natural sounds are used to allow comparison. The experiment results identify three sound dimensions (loudness, order, and clarity) as significantly more intuitive to communicate uncertainty, providing guidelines for sonification and information visualization practitioners.

ARTICLE HISTORY
Received 11 February 2018
Accepted 27 June 2018

KEYWORDS
Sonification; uncertainty; aural cognition; sound variables; communication

Introduction

The understanding and representation of uncertainty is a persistent challenge in geographic information science (GIScience) and its cognate fields (Çöltekin, Bleisch, Andrienko, & Dykes, 2017). As all spatio-temporal data is constrained by spatial, temporal, and thematic precision and accuracy (Ballatore & Zipf, 2015), it is paramount to communicate uncertainty effectively, both to scientists and to non-specialist audiences. For example, different weather forecasts have a varying probability to occur; environmental models are devised to simulate sea level rise scenarios, ranging from likely to unlikely; remotely sensed night lights indicate economic activities to varying degrees of certainty in different geographical areas. Depending on the context, data present different types of uncertainty, starting from the conceptualization to the measurement and analysis of data. Uncertainty also includes conceptual and semantic dimensions, as some geographical notions are imbued with vagueness and have several, only partially compatible definitions. This broadly applies to land use, land cover, and urban/rural classifications that are commonly used in the environmental and social sciences.

Ideally, every piece of geospatial information should be presented with its associated uncertainty, helping the user understand it correctly. In addition to text-based metadata (e.g. "50 m satellite imagery," "the classification precision is 80%," "measurement accuracy ±5%"), visualization methods have been employed to represent data uncertainty in an intuitive way, showing the spatial variation in uncertainty. In a cartographic context, visual metaphors such as transparency, fog, and blur have been shown as relatively effective in communicating this aspect of the data (Kinkeldey, MacEachren, & Schiewe, 2014).

Despite the advances in this field, in practice, representing uncertainty has proven difficult in many contexts, and misunderstandings are common. While it is increasingly easy and inexpensive to produce and disseminate data representations, cognizing the data uncertainty correctly is challenging, particularly for non-expert data consumers (Brown et al., 2013). For this reason, cognitive psychological inquiries can help uncover the cognitive patterns (and errors), leading to more cognitively adequate representations. Notably, showing different designs of hurricane forecasts to citizens tends to result in different risk evaluations (Ruginski et al., 2016). Along similar lines, alternative visual designs of the "blue dot" that signals the user location in Google Maps significantly influence the user perception and

CONTACT Andrea Ballatore a.ballatore@bbk.ac.uk
The supplemental data of this article can be accessed here.

reasoning about locational uncertainty (Hegarty, Friedman, Boone, & Barrett, 2016; McKenzie, Hegarty, Barrett, & Goodchild, 2016). Importantly, this work indicates that some people have a difficulty in understanding the uncertainty inherent to the display in the visual channel, highlighting how the type of display and task influence the cognition of uncertainty.

The majority of efforts in the communication of uncertainty have focused on the visual channel, leaving the other senses unexplored, including the aural channel. Sound-based communication offers several advantages that can complement and—in some contexts—replace visual media. For more than two decades, the discipline of sonification has explored the communication of information through non-speech sounds, devising techniques to translate data into audible sound waves (Hermann, Hunt, & Neuhoff, 2011). For instance, seismic waves generated by earthquakes have been used to generate sounds and even music (http://www.seismicsoundlab.org, accessed on January 2018). The core driver in this area is the possibility of tapping into the unused aural channel to offload semantic content from the overcrowded visual channel (Dubus & Bresin, 2013). Sonification is used in several fields for system monitoring, for sports analytics, for science outreach, for vision-impaired users, and for exploratory data analysis, exploiting the ability of the human auditory system to discriminate the amplitude and frequency of sounds.

Sonification has been deployed in several scientific domains but, to date, little research has tackled the communication of uncertainty using auditory techniques (Bearman, 2013). In this article, we investigate whether sound can be used to convey information about data uncertainty intuitively, identifying the most effective basic sound metaphors for this purpose. Despite the limited applicability observed to date, we consider communicating uncertainty with sound an avenue worth exploring for several reasons. In cartography-based visualizations, the visual channel is often overloaded with several pieces of information, making aural augmentation desirable. Changes in sound dimensions, such as pitch and volume, can be detected by most people and are cognitively general and simple. Hence, we hypothesize that intelligible aural representations can be devised. Rather than focusing on specific applications, we aim at uncovering foundational aspects. General findings in this area can inform a number of fields that deal with varying levels of uncertainty in spatial data.

Grounding this interdisciplinary work in cognitive psychology and semiotics (MacEachren et al., 2012) and sonification research (Hermann et al., 2011), we designed a cognitive experiment to investigate how sound dimensions can be used to signify uncertainty to non-expert users. In this experiment, we focus on the research question: *How intuitive are sound dimensions to communicate uncertainty?* Participants are asked to compare 26 pairs of sounds, interpreting the meaning of the stimuli intuitively as more or less certain. In each pair, a different sound dimension varies, e.g. duration, loudness, pitch, etc. All sound stimuli are available online as supplementary material, as well as in our repository (https://github.com/andrea-ballatore/SonificationUncertainty).

The remainder of this article is organized as follows. The next section discusses prior work in this area, including GIScience, cognitive psychology, and sonification, which inform our study. Subsequently, we outline the experiments, and we report on the tasks and stimuli design. For each experiment, we discuss the results and the guidelines that can be derived for scientists and sonification practitioners. Finally, we draw conclusions from these experiments and we indicate possible directions for future research.

Related work

In our effort to sonify uncertainty, we draw on concepts and findings from GIScience, cognitive psychology, and sonification. This section reviews these research areas, highlighting the lack of systematic assessments.

Data quality and uncertainty in giscience

Assessing and expressing the quality of geospatial information is a core part of GIScience (Zhang & Goodchild, 2002). Several terms are used in GIScience interchangeably to refer to data quality issues, including "uncertainty," "certainty," and "errors." This terminology is problematic and includes implicit biases. Notably, the terms "uncertainty" and "error" have a negative connotation, while "certainty" and "quality" suggest positive qualities. All terms refer to the discrepancy between some knowable state of affairs in the real world and the knowledge acquired through scientific observations, which is necessarily limited and imperfect (Couclelis, 2003). The concept of quality (and uncertainty) has been dissected into orthogonal dimensions, such as accuracy, precision/resolution, completeness, lineage, and currency (Ballatore & Zipf, 2015).

While no clear terminological consensus exists, spatial data and models should embed notions of quality, guiding data consumers to the correct assessment of fitnessfor-purpose, particularly for critical decision-making (Zhang & Goodchild, 2002). Uncertainty is epistemically present in all phases of the information life cycle, starting from the conception, representation, and collection of the data, to its processing and analysis. In this sense, data quality issues are propagated cumulatively from one step to the next in each analysis (Longley, Goodchild, Maguire, & Rhind, 2015). In recent years, the emergence of crowdsourcing and the big data paradigm has prompted new research on uncertainty, focusing on contexts of abundant but non-representative information samples (Kitchin, 2014).

The semiotics of uncertainty

As understanding and describing uncertainty is paramount in scientific work, the communication of uncertainty poses an important challenge, both for scientists, decision-makers, and journalists. When presenting datasets, models, and forecasts, several ways exist to describe their associated uncertainty, minimizing the semantic gap between the information producers and consumers. To meet this challenge, studies in GIScience and cartography analyze how uncertainty can be represented visually. Adopting a semiotic framework, uncertainty is a "signified" that must be associated with effective "signifiers," for which no clear conventions and standards exist. Uncertainty can be displayed either together with the data, or in a separate channel. Different types of uncertainty, such as lineage, positional accuracy, and completeness, are best represented with different techniques, varying Bertin's visual variables, e.g. color, size, and shape (MacEachren et al., 2005). Visualizations can be static or dynamic, including a temporal dimension. Effective communication of uncertainty can benefit information users in diverse contexts, such as hurricane forecast and user self-location (McKenzie et al., 2016; Ruginski et al., 2016).

More than 40 empirical evaluations have been carried out on the symbolization of uncertainty, exploring different communication techniques and tasks (Kinkeldey et al., 2014), confirming overall that uncertainty is not just another variable, but that it deserves special treatment. In particular, MacEachren et al. (2012) investigated alternative visualizations of uncertainty, asking participants how intuitive each representation was with respect to different dimensions of spatial uncertainty. Different iconic and abstract symbols obtained different scores of intuitiveness, showing that fuzziness is the most effective visual metaphor of uncertainty, hence providing guidelines for more effective data visualizations. We adopt this methodology for our study as well.

Sonification research

The translation of data into intelligible sounds has been investigated systematically for more than two decades (Hermann et al., 2011; Kramer, 1994; Kramer, Walker, Bonebright, Cook, & Flowers, 1999). Sonification techniques mainly consist of producing synthetic sound waves to represent some input non-sonic data, e.g. seismic waves, crime rates, and air pollution levels. This approach differs from "earcons," fixed sonic motifs that are used to signify actions and events in user interfaces (e.g. emptying the trash can on a computer). Foundational research in sonification aims at identifying effective mappings between ordinal or nominal variables and sound dimensions, in order to support the design of usable aural representations. Empirical investigations explored the effects of varying pitch (low/high), volume (quiet/loud), spatialization (left/right, front/back), timbre (violin/guitar), and tempo (slow/fast), reducing the huge sonic combinatorial space to manageable samples (Dubus & Bresin, 2013). It is reasonable to assume that, as for visual metaphors, some sonic metaphors are more intuitive than others: for example, high pitch can be employed to represent high values, and vice versa.

Sonification can be used for both complementing and replacing the visual channel through aural communication. Listening to data has proven useful not only in the arts and music composition, but also in seismology and astrophysics, for example to interpret data streams from the Voyager 2 space probe (The Economist, 2016). While evidence of major scientific discoveries made through sonification remains scant, auditory displays are common in several domains, including environmental monitoring (e.g. Geiger counters), complex system monitoring (e.g. telecom networks and stock market trends), and, perhaps more obviously, in interfaces for vision-impaired users (Grond & Hermann, 2014; Loomis, Golledge, & Klatzky, 1998; Nesbitt & Barrass, 2004).

In the context of GIScience, abstract sound variables have been discussed as a way to represent spatial information. Sounds can be shaped by varying location, loudness, pitch, register, timbre, duration, rate of change, order, and attack/decay (Krygier, 1994). Bearman (2013) also compiled an extensive review of the use of sonification to tackle geographic problems. In his thorough appraisal of the field, he notes that some pioneering authors have suggested ways of sonifying uncertain data, without providing empirical evidence (e.g. Fisher, 1994; Pang, Wittenbrink, & Lodha, 1997). In his own work, Bearman evaluated sonification to communicate positional uncertainty through pitch, using piano notes as signifiers (Bearman & Lovett, 2010). As limited research on the sonification of uncertainty has been carried out, particularly using different sound dimensions, the remainder of this article outlines our empirical investigation to fill this knowledge gap.

Sound dimensions and design

To support the sonification of uncertainty, in this article, we aim at identifying the most effective sound variables that can be intuitively associated with different levels of uncertainty. As suggested by MacEachren et al. (2012) in the context of visual metaphors, our study can be framed as a semiotic problem: What sound variables can signify uncertainty? To what extent are these variables intuitive to an untrained participant?

Hence, our experimental design required the development of (1) a set of tasks, and (2) a set of sound stimuli. The purpose of this design is to include all major sound dimensions that can be used in sonification and observe how they perform in the sonification of uncertainty with participants. For this purpose, we adopted the tasks from MacEachren et al. (2012) and the sound dimensions by Krygier (1994) as starting points. While the tasks are relatively simple and described in each experiment's section, the sound design has proven to be more complex, and deserves detailed treatment.

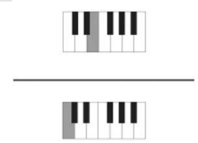
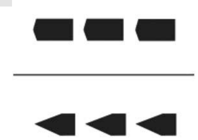
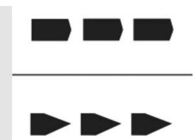
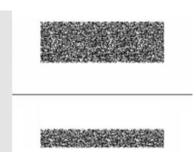
An important distinction in our sound design is that between abstract, machine-generated sounds and natural, real-world sounds. Natural sounds are associated to a known referent (human speech, birds chirping, wheel screeching, musical instruments, etc.), while abstract sounds are not, although they can occasionally appear similar to natural sounds (e.g. continuous noises). This distinction is analogous to abstract and iconic visualizations by MacEachren et al. (2012). For our experiments, we designed a set of sounds to capture these two categories, distinguishing between natural (*acoustic*) and abstract (*synthetic*) sounds.

Sound dimensions were more complex to select, as the range of possibility is extremely broad. Hence, we selected 13 dimensions, making sure to include all the basic dimensions (see Table 1 for a summary). While the human ear can discriminate between more than two levels in a single dimension, we limited our design to two levels to control experimental complexity. For each sound dimension, we designed two clearly distinct sound stimuli, described as *low* and *high* sounds. Three of the most basic sound dimensions are *loudness*, a sound's perceived amplitude; *duration*, the length of a sound; and *location*, where a sound appears in space, e.g. to the left, right, in front of, or behind the listener. We hypothesized that loudness and duration might be effective in the sonification of uncertainty: as loudness is correlated with the distance of the source to the listener, louder sounds could indicate greater certainty, and vice versa. The sonic metaphor for duration seems less clear, and it is possible to hypothesize that short duration could either indicate high certainty, or vice versa. Location, by contrast, was included in the experiment for completeness, assuming a sound from the left and right can be associated with uncertainty arbitrarily.

Pitch and *register* refer, respectively, to the perceived frequency of a sound and the octave in which a note or melody is played. Changing a note's pitch moves it to a different frequency (e.g. C to D), while changing the register of a note causes the same note (e.g. C) to be played at a different octave. During the sound design phase, we had no clear intuition as to how pitch or register might convey the certainty or uncertainty of data. Unlike pitch and register, which refer to perceived frequency, *attack* and *decay* refer to the "envelope" of a sound, that is, how its amplitude changes over time. More specifically, attack describes how long a sound increases in volume before reaching its highest point (sustain level), while decay describes how long it takes for a sound to decrease from the sustain level to zero amplitude. We hypothesized that shorter attack/decay times might indicate higher certainty, as a faster volume change could appear as a more certain statement than a slow one. Another parameter, *rate of change*, referring to how rapidly a sound changes, can be interpreted in a number of ways. To keep the sound design reasonably simple, we interpreted rate of change in terms of slower and faster tempos.

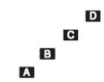
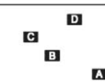
An important dimension that we added to the original set is *noise*. Commonly described as an "unwanted disturbance" in a signal, noise can be acoustically defined as "sound in which the amplitude over time changes with a degree of randomness" (Roads, 2015, p. 97). Intuition suggests that noisier sounds might convey greater data uncertainty, while less noisy sounds could convey greater certainty. We included three kinds of noise: white noise, pink noise, and red noise. *White noise*, containing equal energy for all bandwidths, has the highest degree of randomness. In contrast, for *pink noise*, amplitude is determined by $1/f$ (1/frequency), attenuating the higher frequencies; while for *red noise*, also known as brown or Brownian noise, amplitude is given by $1/f^2$, representing a sharper attenuation of the upper frequencies. Since pink and red noise have lower amplitudes at higher frequencies,

Table 1. Core sound dimensions, based on Krygier (1994).

Sound dimension	Description	Sample values	Visual depiction
loudness	A sound's perceived amplitude	Quiet, loud	
duration	Length of a sound	Short, long	
location	Where a sound appears in space	Left, right	
pitch	Perceived frequency of a sound	C, D	
register	Octave in which a note or melody is played	C4, C5	
attack	How a sound increases in volume before reaching its highest point (sustain level)	Short, long	
decay	How a sound decreases in volume from the sustain level before reaching zero amplitude	Short, long	
rate of change	How rapidly the sound changes	Slow tempo, fast tempo	
noise	Unwanted signal or distortion; mathematically, the degree of randomness in a complex sound spectrum	Noisy, clear	
clarity	Amount of original sound heard, rather than distorted or hidden in noise	Distorted, undistorted	

(*Continued*)

Table 1. (Continued).

Sound dimension	Description	Sample values	Visual depiction
timbre	Hard to define. Parameter that encompasses all aspects not captured by the other parameters. Brightness, tone quality	Flute, trumpet	 
order	Order of execution of different sounds	Regular, irregular	 
harmony	Consonance or dissonance between notes played at the same time	Consonant, dissonant	 

they better approximate acoustic sounds, such as wind or crashing waves (Oxenham, 2013).

When designing noisy stimuli, potential unpleasantness for participants was a concern. We anticipated that pink and red noise would appear less unpleasant than white noise since they are closer to natural sounds; however, even pink and red noise, if unmodified, can seem harsh to the listener. Since unmodified noise can be unpleasant, filtered noise is often used in electronic music to create sounds that appear less "synthetic" to the listener than pure synthesized sounds. Applying a bandpass filter to a pure noise signal creates a noise band of a given width. If the band is narrow enough, the result can be described as "pitched" noise, mixing unpitched white noise and a sine wave (Roads, 2015). In the sounds used in the experiment, a bandpass filter was applied to sounds in each of the noise parameters to attempt to reduce unpleasantness. Within each sound pair, a wider band of noise for the high condition sound was compared to a narrower noise band for the low condition. Since a narrower noise band is quieter than a wider one, in each case the amplitude of the low condition sound was boosted to match the volume of the high condition sound.

In contrast to noise, the sound parameter of *clarity* represents the degree to which an original signal is present without being distorted by or covered in noise (Alten, 2013, p. 4). We separate noise and clarity into two categories to reflect the fact that noise can both be a sound source and a signal added to another sound. Whereas the noise category compares bands of pure noise of different widths, the clarity category adds distortion to an existing sound. *Timbre* is often defined negatively, as the "perceptual attribute that enables us to distinguish among orchestral instruments that are playing the same pitch and are equally loud" (Risset & Wessel, 1999, p. 26). As it typically conveys the identity of a sound source, it seems timbre can be useful for the sonification of uncertainty, although we had no clear intuition on what different timbres may suggest. It is important to note that the parameters of clarity and timbre are linked, since a reduction in sound clarity will likely affect timbral characteristics as well. Like timbre, we hypothesized that the parameter of clarity may potentially be effective to signify uncertainty, as processes such as distorting or muffling a sound could intuitively convey that data has been degraded or lost.

Two of the more complex sound dimensions, order and harmony, refer to relationships between sounds, rather than qualities of individual sounds. The *order* dimension might hold potential for the sonification of uncertainty, as sounds that are out of order could convey greater uncertainty, suggesting less structured data. At its most basic level, *harmony* refers to the consonance or dissonance between notes played at the same time. A "harmonic interval" is the ratio between the frequencies of two simultaneously sounding notes. Harmonic intervals with low whole number ratios, such as 3/2, 4/3, 5/4, and 6/5 sound consonant, while other intervals sound dissonant (Plomp & Levelt, 1965). A musical chord contains three or four notes, whose combinations in harmonic intervals make the chord sound harmonious or inharmonious. We hypothesized that harmony could be a potentially useful parameter for the experiment, since an inharmonious chord could suggest greater uncertainty.

It is worth noting that many sound dimensions perceptually correspond to physical quantities in non-linear ways. For instance, pitch is perceived logarithmically with respect to sound frequency: an increase of one octave corresponds to a doubling in frequency (Moore, 2012). Since, in our study, participants compare discrete sounds with high and low values clearly distinguishable from each other, these considerations are mainly applicable to future research investigating representations of continuous changes in uncertainty through sound.

From a technical viewpoint, acoustic sounds are based on violin recordings available in the public domain, modified with the audio editing software Audacity (https://www.audacityteam.org). Synthetic sounds were either generated algorithmically by the Minim library for Processing 2.1 (http://code.compartmental.net/minim), created manually in Audacity, or a combination of both. All sounds are available in our online repository as open data (https://github.com/andrea-ballatore/SonificationUncertainty).

The experiment is described below. Informed consent was obtained for each participant, and all experiment procedures were approved by the University of California, Santa Barbara, Human Subjects Committee, before the trials. Participants were always free to leave at any time during all trials.

Experiment: comparison of two sounds

In this experiment, we investigate the sonification of uncertainty through a set of sound dimensions. Participants were asked to compare two sounds (A and B), indicating which one evokes "better" or "worse" quality. The core research question targeted by this experiment is

How intuitive are different sound dimensions to communicate uncertainty?

The experiment was first tested on a pilot run with 10 participants, identifying and correcting several design issues. The final design used in the main experiment is described below.

Stimuli design

For this experiment, we designed a set of sound stimuli, containing alternative sonic representations of the dimensions in Table 1. We designed 26 separate stimuli, including 11 acoustic stimuli and 15 synthetic stimuli, varying 16 sound dimensions: *attack/decay, clarity, duration, harmony, location 1, loudness, noise pink, noise red, noise white, order, pitch, rate of change, register, timbre, timbre 1*, and *timbre 2*. To produce the 52 sounds, many technical choices had to be made, reducing the huge space of possibilities to a small, manageable, intelligible, and yet representative set, summarized in Table 2. Each stimulus is generated in a high-low pair, for a total of 52 sound waves. In order to provide sufficient technical clarity and ensure replicability, we detail below the rationale, techniques, and tools used in this sound design. The sound files and full result tables are accessible in the online repository.

In this experiment, most sounds were 4.0-second long, with a 0.5-second fade out, a suitable length that we determined empirically, balancing the intelligibility and pleasantness of sounds, as well as task duration. The exceptions were *attack/decay, duration*, and *rate of change*, since these dimensions require different sound lengths between high and low sounds. For the *duration* parameter in the synth category, a sine tone at C4 was used for both the low and high conditions. In the acoustic category, a violin note at C4 played legato non-vibrato was used for both the low and high duration sounds. For both categories, the notes are spaced 1.5 seconds apart, and for the high condition, the notes are 1-second long, while for the low condition the notes have a shorter duration of 0.5 seconds.

In the *harmony* dimension in the synth category, the high condition sound consists of three simultaneous sine tones forming a major third chord, or E-flat, G and B-flat. The low condition is the same chord, but instead of the sine tone G being played, a sine tone lower than G by ½ semitone was used. In the acoustic category, the sounds for the harmony dimension were created using a similar method, except instead of sine tones, the sound sources were violin notes played legato non-vibrato. To produce the inharmonious chord, the violin G note was detuned using the Audacity Change Pitch feature. In the low condition for both categories, since the harmonic interval between E-flat and the detuned G is not the major third, but an interval with a non-whole number ratio, the low condition should sound less consonant to the listener.

The high clarity synth sound consists of an unmodified triangle waveform played at C4 (261.4 Hz), using the Minim library. Since we wanted the low clarity synth to be noticeably distorted, yet not become unpleasant to the listener, we followed a two-step process. First, using the

Table 2. The 16 sound dimensions used in the experiment, resulting in 26 sound pairs. The sounds files are available online (https://github.com/andrea-ballatore/SonificationUncertainty).

Dimension	Dimension variation	Sound type	Sound design description
attack_decay	Fast vs. slow	Synth	Sine tone at C4 (261.6 Hz)
-		Acoustic	Violin note C4 played legato non-vibrato
clarity	Clear vs. distorted	Synth	High: triangle waveform played at C4 (261.4 Hz) Low: distorted triangle waveform at C4 (261.4 Hz)
-		Acoustic	High: violin note G2 played legato non-vibrato Low: violin note G2 note digitally distorted
duration	Short vs. long	Synth	High: sine tones at C4, 1 second durations Low: sine tones at C4, 0.5 second durations
-		Acoustic	High: legato violin note C4 1 second notes Low: legato violin note C4 0.5 second notes
harmony	Harmonious vs. inharmonious	Synth	High: sine tone major chord, E-flat, G and B-flat Low: sine tone major chord with detuned G
-		Acoustic	High: violin major chord, E-flat, G and B-flat Low: violin major chord with detuned G
location	Left vs. right	Synth	High: sine tone in right channel Low: sine tone in left channel
-		Acoustic	High: violin C4 legato non-vibrato in right channel Low: violin C4 legato non-vibrato in left channel
loudness	Soft vs. loud	Synth	High: sine tone C4 normal volume Low: sine tone C4 lower volume
-		Acoustic	High: violin note C4 normal volume Low: violin note C4 lower volume
noise_white	Small vs. wide band	Synth	High: 400 Hz white noise band Low: 40 Hz white noise band
noise_pink	Small vs. wide band	Synth	High: 400 Hz white noise band Low: 40 Hz white noise band
noise_red	Small vs. wide band	Synth	High: 400 Hz white noise band Low: 40 Hz white noise band
order	Low vs. high order	Synth	High: rising chromatic scale from C4 to F4 Low: unpredictable sequence
-		Acoustic	High: rising chromatic scale from C4 to F4 Low: unpredictable sequence
pitch	Low vs. high	Acoustic	High: violin note B-flat 4 Low: violin note at C4
-		Synth	Low: sine tone at C4 High: sine tone at B-flat 4
rate_of_change	Slower vs. faster tempo	Acoustic	High: violin staccato C4 notes: 3/second Low: violin staccato C4 notes: 1/second
-		Synth	High: sine tone pulses on C4: 3/second Low: sine tone pulses on C4: 1/second
register	Low vs. high octave	Acoustic	High: violin note G4 Low: violin note G3
-		Synth	High: sine tone G4 Low: sine tone G3
timbre	Tremolo vs. legato	Acoustic	High: violin note G3 with tremolo Low: violin note G3 legato non-vibrato
timbre_1	Simple tone vs. Complex tone	Synth	High: square wave approximation with 5 sine waves Low: sine tone at C4
timbre_2	Simple tone vs. Complex tone	Synth	High: sawtooth wave approximation with 5 sine waves Low: sine tone at C4

Wavetable.warp method in the Minim library, a slightly distorted triangle waveform at 261.4 Hz was recorded. Next, in Audacity, the Noise Removal plugin was applied to the whole sound to reduce the distortion noise to a more tolerable level. Noise Removal requires a noise profile, for which a small portion from the middle of the sound of about 0.2 seconds was used. Finally, the amplitude was adjusted slightly to bring it back to the same level as the high condition sound.

To design the *loudness* dimension in the synth category, we chose a sine tone at C4 as the high sound, while the low sound was created by using Audacity to lower the volume of the sound by −12 dB. In the acoustic category, the sound for the high condition is the note C4 played legato non-vibrato on the violin, while the low condition sound was produced by lowering the volume by −6.3 dB in Audacity. A smaller decrease in amplitude is needed for the acoustic sound, to account for the fact that in complex tones, where sonic energy is present in several critical bands, "[t]otal loudness is greater than when the same amount of energy is concentrated within one critical band" (Oxenham, 2013, p. 12).

For the *clarity* dimension in the acoustic category, the high clarity sound is the note G2 played legato non-vibrato on the violin. For the low clarity sound, the G2 note was digitally distorted using Audacity. Specifically, a short portion of audio, around 0.4 seconds, was loaded as a noise profile and the Noise Isolation feature was used, boosting the amplitude of the portions of the harmonic spectrum contained in the noise profile, but maintaining the same fundamental frequency. As the noise isolation process reduces the volume, the overall amplitude was later adjusted to match the original sound.

For the low condition sound for each *noise* parameter, the raw noise was filtered using a bandpass filter with a width of 40 Hz. This bandwidth was used since it would be noticeably smaller than the high condition, yet wide enough that it would avoid a clear sense of pitch. In the high condition sound, the noise signal was filtered using a bandpass filter with a width of 400 Hz. The 400 Hz bandwidth filter was chosen to produce a noise band clearly wider than the low condition, yet still avoid some of the unpleasantness of unfiltered noise. As pink noise is quieter than white noise, due to containing fewer high frequencies, and red noise contains fewer high frequencies than either white or pink noise, volume levels of the sounds needed to be adjusted to match.

The sounds for *order* were tested by comparing a smooth chromatic scale with a disconnected sequence of notes. As the chromatic scale is a regular pattern, we used it to indicate a high degree of order. In both the synth and acoustic categories, the high condition is a rising chromatic scale from C4 to F4. For the low condition in the synth category, an unpredictable sequence of notes was used: A3, Bb4, G5, C#4, E5, G#3. For the low condition in the acoustic category, a different unpredictable sequence of notes was used: D4, A4, E4, G3, Eb4, B4.

For the *rate of change* dimension in the synth category, the sound source was a sine tone, while in the acoustic category the sound source is a violin note played staccato. In both the acoustic and synth categories, the high condition sound consisted of short C4 notes, about 0.25-seconds long, with a short gap separating the notes, so the pulses repeat at a rate of three per second. For the low condition, silence was added in between notes so that the pulses sound at one per second.

Timbre is a complex meta-dimension that mixes several other sound dimensions (Risset & Wessel, 1999). For timbre, the acoustic sounds used were a violin note G3 played legato non-vibrato for the low condition and the same note played with a tremolo effect for the high condition. There are two timbre parameters in the synth category, which we called timbre #1 and timbre #2. In the synth category for timbre #1, the low condition sound is a sine tone at C4, while the high sound was generated with a waveform approximating a square wave through the addition of five harmonically related sine waves. The Minim wavetable function used was *Waves.squareh(5)*. Timbre #2 compared a sine tone at C4 for the low condition with five sine harmonics approximating a sawtooth wave for the high condition. The Minim wavetable functions used was *Waves.sawh(5)*. We avoided using pure square and sawtooth waves in Timbre #1–2, since these sounds might be unpleasant for the listener.

Task design

Based on cognitive psychological work in the area of communication of uncertainty (MacEachren et al., 2012), we designed a task for non-expert participants. The task consists of listening to two sounds (high and low) for a set of dimensions, to then express a judgment about them. After preliminary consultation with participants, the terms "certainty" and "uncertainty" were deemed to be too confusing, and the term "data quality" was considered clearer. To reduce the positive/negative bias, a positive and negative phrasing of the task were alternated between subjects, using either "better" or "worse" in each trial. The core question was phrased as followed:

Which sound, A or B, seemed as if it was coming from a source with BETTER/WORSE data quality?

To present participants with a meaningful and interesting context, a scenario was formulated, where a scientist needs vision for data collection, and a computer system is being designed to communicate about the data quality through sound. Hence, the participant is asked to help train the computer to learn how humans interpret different sounds (see Supplementary Material for the complete experiment protocol). The instructions state explicitly that task has no right or wrong answer. Each participant carried out the tasks individually, and not in groups of participants.

To ensure that all sounds are clearly audible, not unpleasant nor too loud, a volume calibration phase was introduced at the beginning of the task, testing both the loudest and faintest sounds. As stated in the ethical clearance, participants were asked to fill in an informed consent form, and were free to interrupt the experiment and leave at any time. Participants were instructed to listen to each of two sounds presented through consumer over-the-ear headphones for approximately four seconds each.

Each sound was presented with a green bounding box that indicated which sound was currently playing, sound A or B. After both sounds concluded, both boxes were bound with green lines along with the question above, asking participants to make a choice to this question.

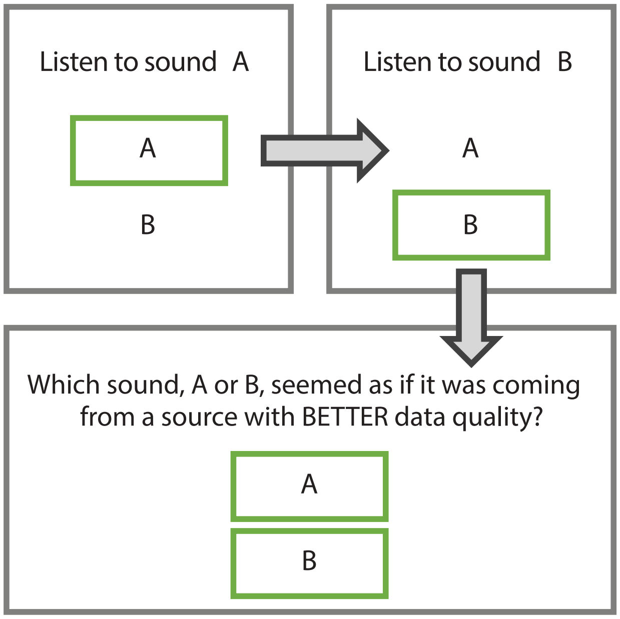


Figure 1. Task interface, starting from the top-left screen, and then moving to the other ones.

Table 3. Results for sound dimensions, aggregating acoustic and synthetic stimuli (33 participants, 1848 responses). For each of the 16 sound dimensions, it shows the median response time (RT), the better/worse divergence index (BWD), and the preference index (PI), and its corresponding binomial probability at 95% confidence, (.) $p < .1$, (*) $p < .05$, (**) $p < .01$, (***) $p < .001$.

Sound variable	RT median (ms)	Better/worse div. (BWD)	Pref. index (PI)	Binomial prob.
loudness	518	0.12	0.58	0.79 ***
timbre_2	578	0.15	0.39	0.70 **
order	741	0.15	0.33	0.67 ***
clarity	506	0.03	0.29	0.64 **
attack_decay	569	0.05	0.23	0.61 *
noise_red	718	0.03	0.21	0.61
timbre	872	0.03	0.21	0.61
harmony	608	0.03	0.18	0.59 *
location	683	0.03	0.15	0.58.
duration	690	0.06	0.15	0.58.
register	553	0.20	0.11	0.55
pitch	630	0.06	0.09	0.55
rate_of_change	783	0.05	0.08	0.54
timbre_1	512	0.03	−0.24	0.62.
noise_pink	602	0.06	−0.36	0.68 **
noise_white	694	0.06	−0.36	0.68 **

Each participant then chooses between sound A and sound B to answer the question by pressing the number pad key marked with stickers indicating their association. The interface flow is summarized in Figure 1.

After the calibration phase, participants were given four practice trials, selected from the set of all stimuli pairs. In order to minimize possible effects on the results and reduce confusion, no new sound was introduced and the selection was randomized across participants. Subsequently, each participant was asked to evaluate 26 high-low pairs of stimuli, shown in random order to reduce ordering bias. To increase intra-subject reliability, each of the 26 pairs was played twice, once in the high-low order, and once in low-high, for a total of 52 sound pairs. At the end of the task, participants were asked to provide feedback in a separate online form on the Qualtrics platform.

Results

The experiment was run on 36 participants at University of California, Santa Barbara, in April and May 2015. The total duration of the task ranged from 9:03 to 16:01 minutes, with 12:37 as the average. Each participant evaluated four test pairs, 52 sound pairs (26 sound designs in two playing orders), for a total 56 selections. Three participants were removed from the dataset as they stated in the feedback that they had made mistakes in the task. As a result, 1848 responses from 33 participants were analyzed, including 20 female (60%) and 13 male participants (40%). Participants were mostly undergraduate students, with age ranging from 18 to 39, with a median of 19. The responses were verified in terms of completeness and collated and are illustrated in Tables 3 and 4.

Table 4. Results for sound designs (33 participants, 1848 responses). For each of the 26 sound designs, it shows the median response time (RT), the better/worse divergence index (BWD), the preference index (PI), and its corresponding binomial probability at 95% confidence, (.) $p < .1$, (*) $p < .05$, (**) $p < .01$, (***) $p < .001$.

Sound type	sound dimension	RT median (ms)	Better/worse div. (BWD)	Pref. index (PI)	Binomial prob.
Synth	Loudness	482	0.03	0.73	0.86 ***
Acoustic	Order	652	0.15	0.58	0.79 ***
Acoustic	Clarity	496	0.06	0.42	0.71 ***
Acoustic	Loudness	555	0.24	0.42	0.71 ***
Synth	timbre_2	578	0.15	0.39	0.70 **
Synth	Register	574	0.21	0.33	0.67 **
Synth	Harmony	626	0.03	0.27	0.64 *
Synth	attack_decay	634	0.06	0.24	0.62.
Acoustic	attack_decay	514	0.03	0.21	0.61
Acoustic	Timbre	872	0.03	0.21	0.61
Synth	noise_red	718	0.03	0.21	0.61
Synth	Location	628	0.06	0.18	0.59
Synth	Pitch	650	0.06	0.18	0.59
Acoustic	Duration	658	0.15	0.15	0.58
Synth	Clarity	510	0.09	0.15	0.58
Synth	duration	712	0.03	0.15	0.58
Acoustic	location	696	0.06	0.12	0.56
Acoustic	harmony	590	0.03	0.09	0.55
Synth	order	772	0.15	0.09	0.55
Synth	rate_of_change	747	0.09	0.09	0.55
Acoustic	rate_of_change	884	0.18	0.06	0.53
Acoustic	pitch	615	0.06	0.00	0.50
Acoustic	register	518	0.18	−0.12	0.56
Synth	timbre_1	512	0.03	−0.24	0.62
Synth	noise_pink	602	0.06	−0.36	0.68 **
Synth	noise_white	694	0.06	−0.36	0.68 **

Response times

The median response time (RT) in milliseconds was collected for each judgment, indicating how much cognitive processing is necessary to give an answer. The values fall in a broad range, from 1 ms to 25 s. As the median is relatively low (613 ms), values above 5 seconds can be considered as a small tail. For this reason, we consider medians as robust indicators of cognitive load, and not means. The median per participant indicates a broad range of RTs, from about 250 ms to 1.5 s. These values are coherent with the task, and we did not find values indicating unusually fast responses. In terms of medians, male and female participants do not show significant differences ($t = -0.63$, $df = 23.7$, $p = 0.5$). Similarly, the type of stimulus (acoustic or synthetic) does not show significant RT differences ($t = 0.64$, $df = 1493.6$, $p = 0.5$). Even the positivity of the question ("better" or "worse") does not seem to influence the RT at all ($t = -0.41$, $df = 1710.8$, $p = 0.7$).

Different RTs can also be observed in the sound dimensions (see Table 3). The median RT of each sound dimension ranges from about 500 ms to 870 ms, with a median of 620. This suggests that some sound dimensions (and therefore stimuli) are easier to cognize, such as *clarity, loudness*, and *register* (RT < 600 ms), while a significantly higher load is needed to process *order* and *rate of change* (RT > 700 ms). Interestingly, the complexity of *timbre* is visible in the wide range of RTs for the three alternative designs. The easiest to process is the square wave (timbre 1, 512 ms), followed by the sawtooth wave (timbre 2, 578 ms). The violin stimulus is, in fact, the slowest dimension (timbre, 872 ms). As the parametric assumptions of ANOVA are not met, we use the Kruskal–Wallis test as a non-parametric alternative to test the significance of these differences for each pair of dimensions. While the inter-dimension differences appear significant as a whole (Kruskal–Wallis chi-squared = 45.6, $df = 15$, $p < .001$), most of the comparisons are not statistically significant when observed pair-wise ($p > .05$). By observing the median per stimulus pair (see Table 4), the values span from about 480 ms (*synth loudness, acoustic clarity*, and *synth clarity*) to more than 800 ms (*acoustic timbre, acoustic rate of change, synth order*). Similarly, the data shows significant differences (Kruskal–Wallis chi-squared = 49.9, $df = 25$, $p < .01$), but most pair-wise differences are not significant ($p > .05$), limiting the conclusions that can be drawn from RTs alone. Furthermore, it must be noted that the protocol did not include rewards for fast answers, limiting the informativeness of RTs.

Better and worse divergence

The answer to the comparison question ("sound A" or "sound B") is the core piece of information to consider. As the task poses two questions with terms "better" and "worse," it is important to first evaluate the impact of this difference on the responses. For this purpose, we devised a *better/worse divergence* (BWD) index, ranging from 0 (no difference between the two terms) to 1 (all responses are skewed to either better or worse). BWD is calculated as follows:

$$tot = better\_low + better\_high + worse\_low + worse\_high$$
$$BWD = (|better\_high - worse\_low| + |worse\_high - better\_low|)/tot$$

where *better_high* is the number of participants who selected the high sound while being asked the question with the "better" option, and *tot* is the total number of responses. A low value of BWD can be interpreted as better/worse having little or no effect, hence indicating a positive outcome for the experimental design. Relatively high values might indicate, by contrast, a possible bias, as the two terms ("better" or "worse") resulted in different responses. In Table 3, relatively high values (BWD > .1) are visible for *loudness, timbre 2, order*, and *register*, while the other sound designs obtained very low values in the range [.03,.06].

As this is a potential concern for the experimental design, the differences between the two modes were analyzed with Fisher's exact test. For each of the 26 designs, we built a contingency table with low/high as rows, better/worse as columns, and the number of user selections as values. For the vast majority of designs, the odd ratios were in the range ~.8 and ~1.2, indicating a balanced choice between the "better" and "worse" options. All tests were non-statistically significant ($p > .5$). Only in three cases (*acoustic loudness, synth register*, and *acoustic order*), while still non-significant, odd ratios were more divergent, respectively, .49, .62, and 1.42. As even these three cases result in non-significant, moderate divergence, we consider the task design robust.

High-low preference

The preference assigned by participants to either sounds in each trial marks the intuitiveness of those sound stimuli to be used for the signification of uncertainty. To quantify the extent to which participants prefer either "high" or "low" version of a sound pair, we devised the preference index (PI), calculated as follows:

$tot\_pref\_high = better\_high + worse\_low$
$tot\_pref\_low = better\_low + worse\_high$
$PI = (tot\_pref\_high - tot\_pref\_low)/tot$

where *tot* is the total number of preferences expressed. PI ranges from −1 (all participants prefer the low sound as indicating "high quality") to 1 (all participants prefer the high sound as indicating "high quality"). Values around 0 indicate that half participants selected the low sound, while the other half selected the high sound. Hence, values close to 1 and −1 are an indicator of effective designs, because users tend to express a preference, showing that the sound design matches the participants' intuition. By contrast, values near 0 suggest that those sound designs are ineffective.

To test the significance of the PI, we observe the divergence of the user selection from a random selection ($p(low) = p(high) = .5$) using an exact binomial test. This test is designed to either confirm or reject the null hypothesis by calculating the probability of success in a Bernoulli experiment with successive trials (low/high sound selections in this context), with the corresponding p-value. For example, when exposed to the *loudness* stimuli, participants selected the same sound ("high") 104 times out of 132 trials, resulting in a binomial probability of .79 at 95% confidence, with $p < .01$. As is possible to observe in Table 3, 8 out of 16 dimensions resulted in significant preferences ($p < .05$), indicating a positive result—namely, that the selection shows a clear non-random pattern, corresponding to high PIs. The other eight dimensions show a binomial probability too near .5 (i.e. random selection), indicating them as ineffective sound dimensions.

As each participant expressed their opinion about a sound pair twice, we also analyze the consistency in this choice by counting for how many pairs the participants expressed the same preference (selection of high *or* low sound twice), and for how many pairs elicited contradictory responses (selection of high *and* low sounds). A consistency index ranging from 0 (all selections were contradictory) to 1 (all selections were consistent) was calculated for each sound pair. This index ranged from .4 to .88, with a median .63. When comparing this index with the PI, high variability in the infra-subject consistency emerges for the sounds with low PI ([.2,−.2]). However, for high PI (significant binomial probability >.65), high values for the consistency index are obtained (>.7), indicating that most participants selected the same sound twice. As this corroborates our main findings, and for the sake of brevity, we do not include the full values for this index.

Acoustic and synthetic sounds

Differences between acoustic and synthetic sound designs are important to observe. As shown in Figure 2, the values of PI for the different sound designs vary from −.36 to .73, with a median of .17. BWD and PI show no correlation ($r = .1$, $p > .4$), indicating that they capture orthogonal aspects of the results. *Synthetic loudness, acoustic order*, and *acoustic clarity* emerged as stimuli for which participants had a strong preference (PI > .4). Out of 26 designs, 10 obtained a binominal probability higher than .64 ($p < .05$), indicating a significant non-random preference. The top four designs (*synth loudness, acoustic order, acoustic clarity*, and *acoustic loudness*) obtain probability higher than .7 ($p < .001$).

Discussion

This experiment shows a large variability in the judgments expressed by participants. As shown in Table 4, overall, 23% of the 26 sound designs obtained PI > .3, with binomial probability >.65 ($p < .01$). We consider this to be reasonably strong evidence for usable sound designs in aural interfaces to signify uncertainty. Interestingly, *noise pink* and *noise white* also show high preference, but opposite to the expected one (PI < −.3). The cases with PI < .3 do not exhibit a sufficient preference to recommend usage, suggesting unsuitable designs (and sound dimensions). *Synthetic loudness* is without doubt the most effective and robust sound design (PI = .73, BWD = .03, bin. prob. = .86, $p < .001$). Among the selected sound dimensions, *loudness* definitely features as the most promising for practical use to signify uncertainty (see Table 3). *Order* provides a promising dimension (PI = .33), contrasting highly structured and random-sounding sequences. *Clarity* in its acoustic version (PI = .42) appears usable, as well as *timbre 2* (sawtooth wave, PI = .39).

We had anticipated that since loudness carries an association with sound source distance, a louder sound would appear closer, which would be seen to represent greater data certainty. The strong results in the loudness dimension support this hypothesis, though further research is needed to determine the precise intuition behind this preference. For example, other sonic metaphors are possible besides that of a sound source moving closer or farther away, including that of a speaker putting emphasis on certain words. The positive results for *order* provide support for the claim that an ordered sequence of notes, compared to an unordered, random-sounding sequence carries an intuition of greater data quality. Interestingly, the preference for order is

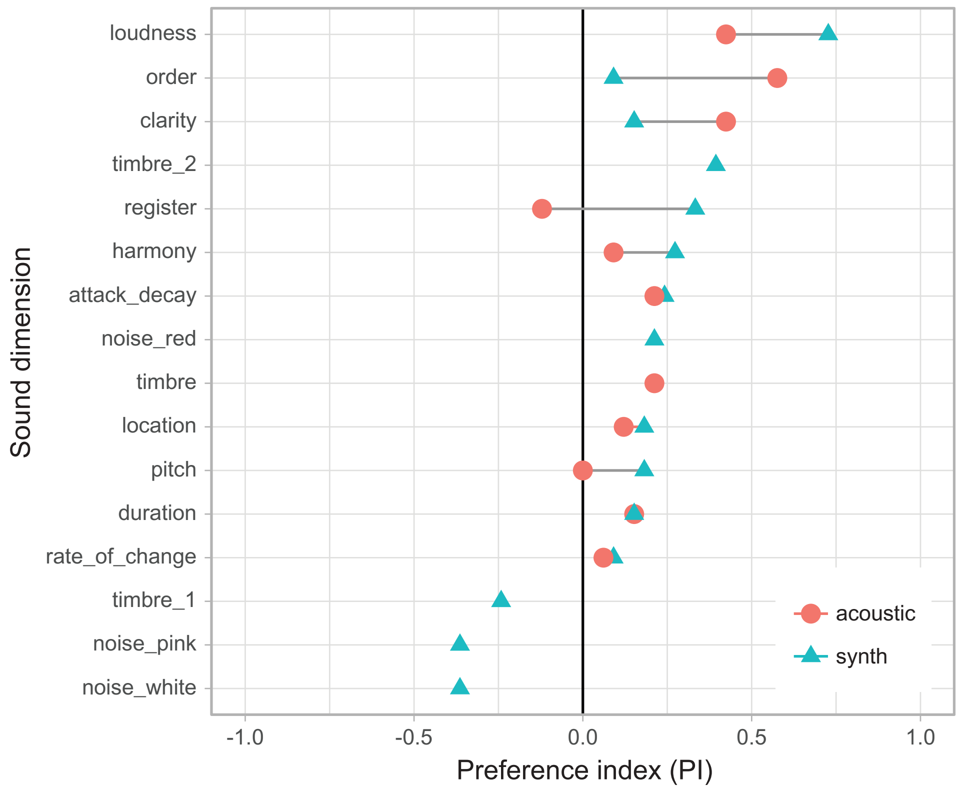


Figure 2. Comparison of acoustic and synthetic sound designs for each sound dimension (33 participants, 1848 responses). For each of the 26 sound designs, it shows the preference index (PI). The sound dimensions are ordered by maximum PI in descending order.

stronger for the acoustic sounds than for the synthetic sounds. Random notes on a violin seem to convey a greater sense of unpredictability, perhaps due to a cultural expectation that violin sounds will appear more melodious than synthetic tones, or that disordered notes become less unpredictable or unpleasant when heard as simple sine waves.

The *timbre 2* dimension, comparing a sine tone (low) with an approximation of a sawtooth wave, revealed a preference for the high condition, contrasting with a slight preference for the low condition for *timbre 1*, which compares a square wave (high) with a sine wave (low). Since we had little intuition how timbre might convey data quality, the high and low conditions for timbre were designated largely arbitrarily, so we were not surprised that results for *timbre 1* and *timbre 2* diverged. Since timbre is a complex dimension, many possible factors could contribute to this result. One possible explanation is that a richer or "brighter" harmonic spectrum makes a sound appear more "present," creating a positive association with data certainty. Since a sawtooth wave has "brighter," richer timbre than a sine wave, this theory could explain the preference for the high condition in *timbre 2*. Since a square wave, on the other hand, is missing the even harmonics (2nd, 4th, 6th, etc.), it tends to be described as "hollow," which perhaps can explain the preference for the low condition in *timbre 1*. In absence of further research, these considerations remain rather speculative.

Interestingly, two of the noise-based sound designs (*pink* and *white* noise) obtained significant preferences (PI = −.36, bin. prob. = .68, $p < .01$), but inverted with respect to our intuition during the design phase. The preference for narrower noise bands in the *pink* and *white noise* dimensions was surprising, since we had anticipated that wider noise bands would seem noisier and therefore be associated with greater uncertainty. Only the *red noise* parameter seems to weakly support our hypothesis. As *red noise* contains the fewest high frequencies, a feature shared with natural sounds such as wind, it is the least noisy of the three. One explanation of the divergent results for *red noise* could be that only in this category was the low condition sound "clear" enough for the noise metaphor to take precedence. For *pink* and *white noise*, both conditions may have appeared "noisy" enough, so that a sonic metaphor more related to richness of spectrum or "presence" took precedence, similar to the preference shown for the high condition in the *timbre 2* dimension, a sawtooth wave. If this is the case, future research might investigate different configurations of red noise, or devise a different way of testing the pink and white noise parameters.

A rather unexpected result lies in the variability between synthetic and acoustic versions of the same dimensions, compared in Figure 2. As is possible to notice, sound dimensions with a strong temporal component obtained tightly clustered results (*attack/decay, location, duration*, and *rate of change*). Other dimensions, including those that obtained the best results, show instead high variability between the designs (*loudness, order, clarity, register*, and *pitch*). A possible explanation is that acoustic sounds suggest different sonic metaphors than synthetic sounds, perhaps due to their association with a musical context. The results might also reflect the fact that acoustic sounds, even solo violin notes, are more complex and contain a richer harmonic spectrum, so it is harder to isolate their effects.

While we consider the aforementioned results to be robust and usable for uncertainty sonification, the experimental design shows limitations that should be borne in mind. First, the experiment has a very abstract focus and does not address the communication of spatial information, which is of more direct relevance to GIScience and geo-visualization research and applications. Furthermore, all participants came from a homogenous, young population, and a more diverse sample in terms of age group and ethnicity would be desirable. For instance, older participants might perceive the stimuli differently, particularly those involving higher frequencies.

The experimental protocol included instructions aimed at reducing the ambiguity of terms such as "uncertainty" and "quality," explicitly referring to a scenario involving a probe that captures data (see Supplementary Material). When asked how well they understood the task on a 1–7 Likert scale (1 not at all, 7 very well), participants rated their understanding of the task itself fairly high in general (M = 5.27, SD = 1.61). Some participants offered feedback indicating that they felt confident while performing the task, but found it difficult not be influenced by their own aesthetic preferences ("I wasn't sure if I was supposed to choose the sound I liked better, or the one where the quality is more clear"; "The sounds can be pleasing when it comes to music but I become conflicted because the abrupt sounds may give you a better understanding when it comes to data."). Although some participants were able to articulate this issue, it is likely that others were not.

User feedback (both informal and formal) indicated that the semantics of terms "certain," "uncertain," "high," and "low quality" was deemed to be problematic by several participants. Of the 12 participants offering feedback regarding the nature of the task, many stated some degree of confusion regarding the interpretation of "data quality." Examples of participant statements include: "I was unsure on what was considered a better data quality sound" and "'Quality' is somewhat subjective." However, this problem has already been encountered by previous cognitive psychological work (MacEachren et al., 2005), and seems to be avoidable only through experimental designs that do not mention any of these terms. Thus, these results highlight a need for indirect methods of assessing behavior without reliance on each person's intuition of the construct in question.

Conclusions

In this article, we have explored empirically the challenge of representing uncertainty through sound. This approach can be useful to complement and even replace visual representations of spatial data, tapping into an under-utilized communication channel. After reviewing the existing work in the interdisciplinary field of uncertainty semiotics and sonification, we selected a number of sound dimensions to investigate how appropriate they are to represent uncertainty, such as loudness, duration, pitch, and register. In order to explore empirically the intuitiveness of these dimensions, we designed a set of pairs of aural stimuli, including acoustic and synthetic sounds, in which only one of these dimensions was altered. These stimuli were then used in an experiment with human participants (N = 33), which collected their intuitive judgments about the sound stimuli meaning with respect to perceived data quality.

This data, freely available in our online repository, allows ascertaining to what extent different dimensions and sound stimuli elicit a preference for a specific sound, as opposed to a random choice. The analysis of results revealed a number of findings, useful to provide guidelines for practitioners and sonification designers. The findings can be summarized as follows:

(1) The most effective dimensions, which provide statistically significant preferences, are loudness, order, clarity, timbre based on a sawtooth wave, as well as pink and white noise (see Tables 3 and 4 for details). The other dimensions generated judgments that do not diverge from random in a significant way, indicating their unsuitability for the representation of uncertainty.
(2) Pitch, the most popular dimension used in sonification to represent quantities (Dubus & Bresin,

2013), obtains near-random results. This indicates the need for specific metaphors for the sonic representation of uncertainty, as occurred in the visual domain (MacEachren et al., 2012).

(3) For dimensions focused on temporal changes (attack-decay, location, duration, and rate of change), acoustic and synthetic sound designs strongly converged to similar results. By contrast, non-temporal dimensions show more divergent results, indicating a higher impact of properties specific to acoustic and synthetic designs (see Figure 2), possibly related to expectations about known musical instruments. In this sense, synthetic sounds should be preferred as less culturally loaded.

(4) In our experiment, some participants reported semantic confusion around terms "certainty," "uncertainty," and "quality." While this might prove hard in practice, the experimental design in further work should favor more indirect tasks that do not mention these ambiguous terms.

These guidelines provide the first step toward effective sonification techniques of data uncertainty. However, more work is needed to develop effective sonic and multisensorial representations of uncertainty in order to support data exploration and interpretation in data science and analytics. As the work in this article focuses on sound dimensions in an abstract space, further work is needed to explore the spatial dimension that is central to cartographic representations, benefitting GIScience more directly. In future experiments, it will be worth tackling spatially structured information, arranging the sound stimuli in two-by-two grids, asking participants to listen to, interpret, and compare sound sequences. Based on the guidelines reported in this article, this approach will enable the design of experiments with right and wrong answers, getting closer to usable sound designs.

From a methodological standpoint, it will be important to develop tasks that allow cognitive measurement without eliciting an explicit semantic interpretation by participants. For example, as uncertainty is an inherently difficult concept to understand, other research has asked participants to use uncertainty in the visual channel to perform a secondary task rather than to estimate uncertainty directly (Hegarty et al., 2016; Ruginski et al., 2016). Knowing which sound dimensions individuals find intuitive can support this type of experimental design.

More research is also needed to investigate the properties of complex sound dimensions, such as timbre and noise, as well as the impact of volume on the results, going beyond the sound design presented in this article. In parallel, sonification approaches should be tested in more concrete scenarios, in different domains and contexts, such as risk assessment, decision-making in emergencies, and data quality assessment. Such an approach will allow researchers to further verify our findings, and would suggest novel designs, hopefully enabling real applications of sonification in GIScience and geovisualization, which have, to date, failed to appear. Communicating uncertainty effectively in sonic representations would be greatly beneficial to support complex scientific, political, and administrative tasks, beyond vision-impaired users that constitute the most obvious beneficiaries of the sonification of uncertainty.

Acknowledgments

We thank the Center for Spatial Studies (UC Santa Barbara), Prof. Mary Hegarty for her constructive input, and the experiment participants for patiently listening to our sounds.

ORCID

Andrea Ballatore http://orcid.org/0000-0003-3477-7654
David Gordon http://orcid.org/0000-0001-8334-1895
Alexander P. Boone http://orcid.org/0000-0002-0367-2887

References

Alten, S. R. (2013). *Audio in media* (10th ed.) Boston, MA: Cengage Learning.

Ballatore, A., & Zipf, A. (2015). A conceptual quality framework for volunteered geographic information. In S. I. Fabrikant & M. Raubal (Eds.), *Spatial information theory* (Vol. 9368, pp. 89–107). Berlin: Springer. http://doi.org/10.1007/978-3-319-23374-1

Bearman, N. (2013). *Using sound to represent uncertainty in spatial data* (PhD thesis). University of East Anglia. Norwich, UK.

Bearman, N., & Lovett, A. (2010). Using sound to represent positional accuracy of address locations. *The Cartographic Journal*, 47(4), 308–314. http://doi.org/10.1179/000870410X12911302296833

Brown, M., Sharples, S., Harding, J., Parker, C. J., Bearman, N., Maguire, M., … Jackson, M. (2013). Usability of geographic information: Current challenges and future directions. *Applied Ergonomics*, 44(6), 855–865. http://doi.org/10.1016/j.apergo.2012.10.013

Çöltekin, A., Bleisch, S., Andrienko, G., & Dykes, J. (2017). Persistent challenges in geovisualization – A community perspective. *International Journal of Cartography*, 9333 (April), 1–25. http://doi.org/10.1080/23729333.2017.1302910

Couclelis, H. (2003). The certainty of uncertainty: GIS and the limits of geographic knowledge. *Transactions in GIS*, 7 (2), 165–175. http://doi.org/10.1111/1467-9671.00138

Dubus, G., & Bresin, R. (2013). A systematic review of mapping strategies for the sonification of physical quantities. *PloS One*, 8(12), e82491.

The Economist (2016, March). Now hear this. *The Economist*.

Fisher, P. F. (1994). Hearing the reliability in classified remotely sensed images. *Cartography and Geographic Information Systems*, 21(1), 31–36. http://doi.org/10.1559/152304094782563975

Grond, F., & Hermann, T. (2014). Interactive sonification for data exploration: How listening modes and display purposes define design guidelines. *Organised Sound*, 19(1), 41–51. http://doi.org/10.1017/S1355771813000393

Hegarty, M., Friedman, A., Boone, A. P., & Barrett, T. J. (2016). Where are you? The effect of uncertainty and its visual representation on location judgments in GPS-like displays. *Journal of Experimental Psychology: Applied*, 22 (4), 381–392. http://doi.org/10.1037/xap0000103

Hermann, T., Hunt, A., & Neuhoff, J. G. (Eds.). (2011). *The sonification handbook*. Berlin: Logos.

Kinkeldey, C., MacEachren, A. M., & Schiewe, J. (2014). How to assess visual communication of uncertainty? A systematic review of geospatial uncertainty visualisation user studies. *The Cartographic Journal*, 51(4), 372–386. http://doi.org/10.1179/1743277414Y.0000000099

Kitchin, R. (2014). Big Data, new epistemologies and paradigm shifts. *Big Data & Society*, 1, 1.

Kramer, G. (Ed.). (1994). *Auditory display: Sonification, audification, and auditory interfaces*. Boca Raton, FL: CRC Press.

Kramer, G., Walker, B., Bonebright, T., Cook, P., & Flowers, J. H. (1999). *Sonification report: Status of the field and research agenda*. International Community for Auditory Display. http://digitalcommons.unl.edu/psychfacpub/444

Krygier, J. B. (1994). Sound and geographic visualization. In A. M. MacEachren & D. R. F. Taylor (Eds.), *Visualization in Modern Cartography* (pp. 149–166). New York: Pergamon. http://doi.org/10.1016/B978-0-08-042415-6.50015-6

Longley, P., Goodchild, M., Maguire, D., & Rhind, D. (2015). *Geographic information science and systems* (4th ed.) Hoboken, NJ: John Wiley & Sons.

Loomis, J. M., Golledge, R. G., & Klatzky, R. L. (1998). Navigation system for the blind: Auditory display modes and guidance. *Presence: Teleoperators and Virtual Environments*, 7(2), 193–203. http://doi.org/10.1162/105474698565677

MacEachren, A. M., Robinson, A., Hopper, S., Gardner, S., Murray, R., Gahegan, M., & Hetzler, E. (2005). Visualizing geospatial information uncertainty: What we know and what we need to know. *Cartography and Geographic Information Science*, 32(3), 139–160. http://doi.org/10.1559/1523040054738936

MacEachren, A. M., Roth, R. E., Brien, J. O., Li, B., Swingley, D., & Gahegan, M. (2012). Visual semiotics & uncertainty visualization: An empirical study. *IEEE Transactions on Visualization and Computer Graphics*, 18(12), 2496–2505. http://doi.org/10.1109/TVCG.2012.279

McKenzie, G., Hegarty, M., Barrett, T., & Goodchild, M. (2016). Assessing the effectiveness of different visualizations for judgments of positional uncertainty. *International Journal of Geographical Information Science*, 30(2), 221–239.

Moore, B. C. J. (2012). *An introduction to the psychology of hearing* (6th ed.) Bingley, UK: Emerald.

Nesbitt, K. V., & Barrass, S. (2004). Finding trading patterns in stock market data. *IEEE Computer Graphics and Applications*, 24(5), 45–55. http://doi.org/10.1109/MCG.2004.28

Oxenham, A. J. (2013). The perception of musical tones. In D. Deutsch (Ed.), *The psychology of music* (pp. 1–33). Cambridge, MA: Academic Press. doi:10.1016/B978-0-12-381460-9.00001-8

Pang, A. T., Wittenbrink, C. M., & Lodha, S. K. (1997). Approaches to uncertainty visualization. *The Visual Computer*, 13(8), 370–390. http://doi.org/10.1007/s003710050111

Plomp, R., & Levelt, W. J. M. (1965). Tonal consonance and critical bandwidth. *The Journal of the Acoustical Society of America*, 38(4), 548–560.

Risset, J.-C., & Wessel, D. L. (1999). Exploration of timbre by analysis and synthesis. In D. Deutsch (Ed.), *The psychology of music* (pp. 113–169). Cambridge, MA: Academic Press. doi:10.1016/B978-012213564-4/50006-8

Roads, C. (2015). *Composing electronic music: A new aesthetic*. Oxford, UK: Oxford University Press.

Ruginski, I. T., Boone, A. P., Padilla, L. M., Liu, L., Heydari, N., Kramer, H. S., … Creem-Regehr, S. H. (2016). Non-expert interpretations of hurricane forecast uncertainty visualizations. *Spatial Cognition & Computation*, 16(2), 154–172.

Zhang, J., & Goodchild, M. (2002). *Uncertainty in geographical information. research monographs in geographic information systems*. New York: Taylor & Francis. doi:10.1111/j.1467-8306.2003.09304014_8.x