International Journal of Geographical Information Science

ISSN: 1365-8816 (Print) 1362-3087 (Online) Journal homepage: https://www.tandfonline.com/loi/tgis20

Exploring maps by sounds: using parameter mapping sonification to make digital elevation models audible

Joram Schito & Sara Irina Fabrikant

To cite this article: Joram Schito & Sara Irina Fabrikant (2018) Exploring maps by sounds: using parameter mapping sonification to make digital elevation models audible, International Journal of Geographical Information Science, 32:5, 874-906, DOI: 10.1080/13658816.2017.1420192

To link to this article: https://doi.org/10.1080/13658816.2017.1420192

INTERNATIONAL JOURNAL OF GEOGRAPHICAL INFORMATION SCIENCE, 2018
VOL. 32, NO. 5, 874–906
https://doi.org/10.1080/13658816.2017.1420192

RESEARCH ARTICLE

Exploring maps by sounds: using parameter mapping sonification to make digital elevation models audible

Joram Schito [a] and Sara Irina Fabrikant [b]

[a]Institute of Cartography and Geoinformation, ETH Zurich, Zurich, Switzerland; [b]Department of Geography, University of Zurich, Zurich, Switzerland

ABSTRACT

This study empirically investigates the potential of auditory displays for spatial data exploration, as an additional means to broaden the accessibility and dissemination of geographic information for a diverse body of users. In a mixed factorial experiment, three parameter mapping sonification methods are empirically evaluated to interactively explore discrete and continuous digital elevation models by auditory means. Contrasting prior sonification research, this study's unique empirical evidence suggests that participants can indeed successfully interpret sonified displays containing continuous spatial data. Specifically, the auditory variable pitch leads to significantly better response accuracy, compared to the sound variable duration. Background and training has a weak effect on data interpretation performance with the auditory display. The more immersive the experienced soundscape, the better participants can interpret the sonified terrain. These encouraging empirical results indeed suggest that interactive auditory displays might offer additional means to disseminate spatial information, and to increase the accessibility to spatial data, beyond the currently dominant visual paradigm.

ARTICLE HISTORY
Received 1 September 2016
Accepted 18 December 2017

KEYWORDS
Auditory display; parameter mapping sonification; spatial data analysis; digital elevation model; GIS

Introduction

Since the rise of multimedia desktop computing in the early 1990s, where digital authoring systems allowed even nonexpert developers to creatively assemble a broad range of digital data sources such as, text, charts, graphs, video and sound, GIScientists have been engaged in exploiting the potential of multimodal map displays to access and disseminate spatial data (Cartwright *et al*. 1999). In most geographic multimedia systems (i.e. digital atlases etc.), the assembled information was, and still is, accessed predominantly through visual means (Andrienko *et al*. 2013). The visual primacy of communication has been argued successfully with the dominance of neurons in the human brain associated with vision (McCormick *et al*. 1987). Whereas the amount of spatiotemporal information available today increases exponentially, the already limited human visual system's capacity to process this information has remained largely the same (Ware 2008). Jacobson (2010) argues that the invention of sound carriers, that is

CONTACT Sara Irina Fabrikant sara.fabrikant@geo.uzh.ch

This paper is based on the first author's unpublished Master's thesis, completed under the direction of the second author. The thesis received a distinction from the Faculty of Science of the University of Zurich.

Supplemental data for this article can be accessed here.

the telephone, or the radio, help to inform, guide and shape human understanding about places. Sound may therefore complement or even replace visual data encodings, as to offload cognitive effort especially in situations where the visual perception system might already be overloaded, that is rapidly changing situations or in virtual reality or augmented reality scenarios. One thus might wonder why the nonvisual human perceptual modalities have been underutilized for geographic information so far. This might become even more critical today, as current spatiotemporal data displays are increasingly used in mobile contexts. One's sight might already be occupied with alerting humans of appearing obstacles, and potentially suddenly emerging harmful contextual situations, for example, while wayfinders are trying to navigate to a destination in a rapidly changing environment. Not surprisingly then, successful mobile navigation systems provide not only visual displays but are accompanied by textual turn-by-turn instructions, coupled with real-time voice-overs. Also large-screen, 3D virtual display systems increasingly employ sound and spatialized audio sources to enhance the immersive experience of the projected 3D information (Théberge 2005, Ruotolo *et al*. 2013, Paterson and Conway 2014). Spatiotemporal data access with sound also supports disadvantaged user groups such as the visually impaired or congenitally blind population to benefit from the spatial data revolution (Simonnet *et al*. 2009). In the following related work sections, we highlight relevant research in the context of auditory data representation that motivated our empirical study using parameter mapping sonification (PMS) to explore and make sense of discrete and continuous spatial data.

Data sonification with auditory displays

The idea of making the inaudible hearable can be traced back to two inventions (Schwartz 2003). The first, since 1816, René Théophile Hyacinthe Laënnec's stethoscope amplified existing sound sources that would otherwise be inaudible for humans. The second, the Geiger–Mueller counter, invented in 1908 by Hans Johann Wilhelm Geiger, turns alpha particles that cannot be directly perceived by the human perceptual system through electrical pulses into a synthetic sound source (Brazil and Fernström 2011). Arguably, the Geiger counter systems could be called one of the earliest known auditory displays.

Pollack and Ficks (1954) empirically investigated the potential of high dimensional information transmission through sound and in doing so first introduced the idea of auditory displays. They asked participants to distinguish sounds manipulated through eight sound variables (i.e. noise frequency and loudness, tone frequency and loudness, the rate of alternation between noise and tone, the 'on'-time fraction of the tone, the total duration of presentation and the sound direction) and found that respondents' error rates vary significantly across sound variables. Early empirical sonification research since the digital revolution by Yeung (1980), Bly (1982) and Williams *et al*. (1990) further investigated the effect of different sound parameters to human perception. For example, Yeung (1980) sonified seven data dimensions of chemical compounds and tested participants achieving high classification accuracies (up to 98%). To examine the potential of sound to convey more information beyond visual means in exploratory data analysis, Bly (1982) systematically compared the effect of sound parameters in temporal, multivariate and logarithmic datasets. He discovered that the combination of graphics and sound yielded the highest interpretation accuracy. Similarly, Williams *et al.* (1990) evaluated compound sound textures, so-called sound icons (i.e. by combining pitch, attack rate, decay rate, volume and depth of frequency modulation), with multivariate glyphs. They found that participants' accuracy increases significantly when graphic symbols are explored jointly with sound icons. In this early epoch of empirical sonification research, researchers focused primarily on testing the mappings of various dataset dimensions to varying sound variables, including both discrete (e.g. timbre, harmonics, damping and waveform) and continuous sound parameters (e.g. pitch, loudness, panning, duration and attack).

Since the first International Conference on Auditory Display convened in 1992, sonification established itself as an interdisciplinary research field in its own right (Kramer 1992), drawing upon knowledge from cognate disciplines including cognitive science and human factors research and considering visually impaired or blind people.

Sonification in GIScience

Only a few GIScientists have experimented with using sound to explore and communicate spatial data, as early as the 1990s. For example, Krygier (1994) proposed nine sound variables including location, loudness, pitch, register, timbre, duration, rate of change, order and envelope to sonify spatial data. His theoretical framework was inspired by Bertin's fundamental visual variable system, already employed by cartographers for a long time to appropriately assign graphic symbols to depicted data characteristics. Krygier (1994) contends that sound variables should only be used to represent nominal or ordinal levels of information in spatial data. In contrast, Fisher (1994) shows how sound variables can be used to represent uncertainty in land-use classification using the continuous sound variables pitch and loudness and in doing so is able to provide compelling evidence for the use of sound variables for at least interval level data. Lodha *et al.* (1999) sonified a discrete raster map by means of parametric auditory icons and reported improved understanding of the presented information in Geographic Information System (GIS) displays in a user study, combining geographic visualization with sonification for various spatial tasks, including estimates of raw data values, local averaging and global comparisons. Using similar sonification methods, but not supported with any user study, Brauen (2006) explored the use of sound to represent election results in online web maps, whilst MacVeigh and Jacobson (2007) sonified land usage classes with four simultaneously rendered auditory icons.

Pushing the sonification boundary further, Heuten *et al.* (2006) introduced the head-related transfer functions (HRTF) with auditory icon streams and hierarchical earcons to represent geographic features including landmarks as sound areas within a sound room. More recently, Geronazzo *et al.* (2016) demonstrated in two experiments how customized HRTF can be combined effectively with a tactile mouse to help users orient themselves acoustically and haptically by means of virtual anchor sounds located in specific points of a sonified map. Zhao (2006) developed iSonic, a speech-supported PMS tool, to sonify categorized statistical data mainly through the variables pitch and balance to facilitate interactive spatial data exploration for visually impaired and blind data users. Her system allows interactive exploration of map content by means of a touchpad and keyboard, supplemented with so-called Auditory Information Seeking

Actions (Zhao 2006). Extending this work, Delogu *et al*. (2010) empirically compared a sonified map with either tablet or keyboard interaction with four nonauditory tactile map versions. Sighted and non-sighted participants were asked to compare cross-modally which of the nonauditory maps corresponded best with the sonified map they had explored first. Participants performed well with the sonified map in all of the four tactile map discrimination tasks. No significant performance differences were found between participants who used keyboard or tablet interaction modes, or between blind or sighted participants.

More recently, sonification researchers explored various techniques to further advance the sonification frontiers. For example, by using the raster scanning approach as a framework for both image sonification and sound visualization, Yeo and Berger (2008) proposed an algorithm to linearly map brightness value ranges in continuous gray scale raster images to ranges of audio samples. Furthermore, Adhitya and Kuuskankare (2012) developed a PMS library to sonify images along a vector path defined by a user. These sonification methods were found to provide a rapid overview of a raster image.

Grond and Berger (2011) suggest the use of the sound variable duration for the sonification of continuous spatial data models. Bearman and Lovett (2010) implemented Earcons based on triad notes to represent positional information, but without clear evidence for improved data understanding. In a subsequent study, using piano notes and bird calls to sonify distances, Bearman and Lovett (2010) found a significant effect in favor of sonification, compared to visualization alone. Two similar approaches are also noteworthy: On the one hand, Schiewe and Weninger (2012) evaluated the usability of gamut-based and pitch-modified piano earcons for the extraction of qualitative information in maps, but with no clear efficiency gains when combining sonification with visualization. On the other hand, Josselin (2011) built a theoretical framework that proposes the use of soundscapes or compound earcons. By coupling interactive map displays with a musical synthesizer, a user is able to query an image using sonified summary statistics at cursor location. Furthermore, sonification research has successfully moved on to mobile map display use. For example, Laakso and Sarjakoski (2010) propose an interactive mobile hiking map with sonified soundscapes that include land-use classes. Finally, Josselin *et al*. (2016) implemented a QuantumGIS plug-in to first isolate geographic features depicted on an official French topographic map by the proportion of their dominant colors and then sonified them by means of soundscapes. In subsequent usability studies, they found that blind and visually impaired participants were more successful at interpreting more complex sonification patterns, compared to sighted participants, perhaps due to their compensatory capacities or due to previous experience using sound to encode this information. This contradicts the findings of Afonso *et al*. (2005) and Bujacz *et al*. (2011) who could not find any significant performance differences between blind and sighted participants using auditory displays.

A key for successful interpretation of spatial representations is spatial metaphors, which are based on image schema that are at the core of human cognition, such as MORE-IS-UP (Lakoff and Johnson 1980). The MORE-IS-UP image schema is also frequently employed in film to synch an action on screen to the movie's soundtrack. This film technique is also called Mickey Mousing (Kalinak 1992). For example, a sequence of rising notes is heard in synch when an actor walks up the stairs. Kuhn (1996) and others have shown that spatial metaphors in GIS user interfaces help users make sense of spatial and nonspatial data.

From this brief review, one can distill that auditory displays have already been successfully implemented in various ways, mostly through mapping categorical data. However, empirical evidence on its success for data exploration and communication, especially for continuous spatial data, is scarce and inconclusive at best. We thus wondered whether one could push the sonification envelope further and provide additional empirical evidence for the successful use of advanced parameter sonification methods, specifically to explore continuous geographical datasets presented interactively in auditory displays.

PMS for continuous spatial data

Similarly, to information visualization or spatialization, where nonspatial data dimensions are mapped to a set of visual variables to construct abstract visuospatial data displays, PMS assigns sound variables to multivariate data characteristics. As PMS offers great flexibility in data transformation, input data of almost any kind can be mapped for sonification, including specifically, continuous data models (Grond and Berger 2011). Of course, such types of mapping flexibility can also bear potential dangers. In order to achieve usable solutions, not only technical or methodological aspects must be considered (e.g. signal processing etc.), but also informed decisions must be made related to human sound perception, including human information processing modes such as thinking, tuning and listening (Grond and Berger 2011). As with information visualization displays, the parameterization logic of a PMS must be learnt by a user for a particular display, and thus the display is ideally complemented with an auditory legend to increase usability (Grond and Berger 2011). Since the choice of appropriate sound variables is essential for intuitive understanding of auditory displays, knowledge about human acoustics including psychophysics is necessary in sonification research.

Acoustic principles for continuous auditory displays

By means of Fourier analysis, a sound can be decomposed into the continuous wave components amplitude, frequency and phase (Campenhausen 1993, Goldstein 2002). These three physical wave components are perceived as sound, from which different discrete and continuous sound variables used in music and sound engineering can be derived (see Table 1). First, a wave's amplitude corresponds to the sound variable loudness, including the continuous musical variables panning, which is the distribution of a sound between different sources. Second, a wave's frequency corresponds to the sound variable pitch. Third, a wave's phase incorporates two different sound characteristics: On the one hand, it reflects the sound variable duration and consequently its pace. For example, consider a beeping sound with a sequence and duration of the beep and a period of silence in between. On the other hand, a sound wave's shape affects the sound's timbre, which is typically characterized on a nominal level of information. A wave's envelope components translated into the sound variables attack, decay, sustain and release are continuous. Lastly, the combination of amplitude, frequency, and phase allows for the analysis of the more complex sound variables sound texture and spatial

Table 1. Relationship between wave components, sound variables and musical variables.

Wave component	Sound variable	Variables used in music and sound engineering
Amplitude	Loudness	Pan
		Dynamics
Frequency	Pitch	Tone within a gamut
		Cardinality within a gamut
		Register
		Order
		Harmony
Phase	Duration	Pace
	Timbre	Rhythm
		Meter
		Clarity
		Damping
		Reverberation
		Envelope (attack, decay, sustain, release)
Amplitude, frequency and phase combined	Texture	Instrumentation
	Location	Distance
		Angle

location, e.g. by decomposing a sound into its sources in space, or by locating a sound source in the space. Thus, the continuous sound variables loudness, pitch and duration can directly be mapped to continuous data, whereas the qualitative variable timbre must be decomposed into its quantitative envelope components, to be useful for continuous data sonification. While sound variables can be physically modeled and measured, identical sound wave parameters might not be perceived the same way by a human, due to environmental effects (i.e. room acoustics etc.), and varying individual differences in human sound perception.

In the following, we detail our own empirical study aimed at evaluating a novel, interactive PMS by means of sound variables (Table 1), to explore digital elevation models (DEMs) of various types. Through this study, we wish to provide additional quantitative empirical evidence on the utility and usability of sonification of spatial data for exploratory data analysis.

Experiment

The aims of this exploratory, empirical sonification study reported below are manifold. On the one hand, we wish to replicate results with previously suggested sound variables (e.g. those of Krygier 1994), specifically in the context of auditory displays for continuous spatial data analysis (i.e. DEMs). We also aim at extending prior work by investigating how different sonification methods might influence spatial data interpretation. As very little is known on how user characteristics (i.e. musicality, geographic data analysis etc.) might influence performance with auditory displays, we were also interested in evaluating the role of user background knowledge and training for decision-making with sonified data.

Experimental design

We chose a *mixed factorial design* for our study, consisting of the within-subject factor terrain model type (three levels: abstract discrete, abstract continuous and realistic continuous; Figure 4) discussed in the 'Materials' section, and the between-subject factor sonification method (three levels: A, B, C; Table 3). We further assessed participants' individual domain relevant background knowledge and training. The set of controlled variables is summarized in Table 2.

The dependent variable *response accuracy* measured during the auditory display portion of the experiment contains three components, as shown in Figure 1. For relation *a*, a participants' sound interpretation at response location in the auditory display is compared to the modeled sound parameters at that location. Similarly, for relation *c*, a participants' sound interpretation is compared to the modeled sound parameters at the correct target location. Finally, in relation *b*, the distance between correct target location and a participants' response location is assessed. Response accuracy is additionally recorded in a series of closed-ended questions, related to the structure of the explored, sonified space, as well as in pre- and posttest questionnaires related to user background and training.

Table 2. Controlled and independent variables of the experiment.

Controlled variables	Within factors		Between factors
Background knowledge	Terrain model type	Complexity level	Sonification method
Technical ability	Abstract discrete	1	A
Musicality	Abstract continuous	2	B
Terrain interpretation	Realistic continuous	3	C

Table 3. Mapped sound variables to the rendered dimensions of each group.

Sonification method	*X*-coordinate	*Y*-coordinate	Altitude
A	Increasing pitch on left ear	Increasing pitch on right ear	Increasing pace
B	Moving balance from left to right ear	Moving pan from full square wave to full sine wave	Increasing pitch
C	Moving balance from left to right ear	Moving pan from full square wave to full sine wave	Increasing pitch and acoustic elevation profile

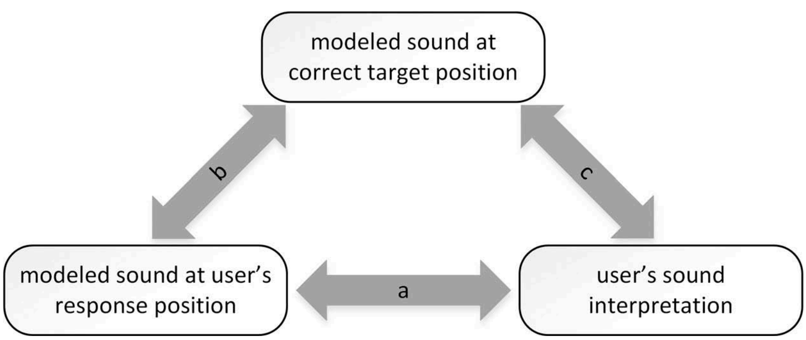


Figure 1. Components of the dependent variable response accuracy.

Participants

In total, 61 participants were invited (26 females and 35 males) who were willing to spend 90 min for a test session. The study was aimed at recruiting a participant pool balanced by gender and familiar with the interpretation of spatial data. The participants recruited also specifically spanned a broad range of ages, as hearing capacity is known to deteriorate with age. Tested participants ranged from 18 to 56 years. Their participation was voluntary. All participants are sighted but were blindfolded during the display exploration portion of the test.

Sound spatialization: the mapping of sound variables

The continuous acoustic variables *pitch, duration* and *pan* were used in this study. *Pan was mapped to loudness* and *to waveform* (Table 1). Pitch is continuous, and based on reviewed literature, easily distinguishable. The *octave* was chosen as the spectral unit, ranging from 0 to 1. This is because it represents as simplest harmonic oscillation, the first and loudest overtone of a harmonic series (Amon 2005). The octave is used across various music cultures as a basis for building a scale. To minimize potentially perceived isophonic loudness fluctuations (Genuit 2010), the fundamental frequency was set to 440 Hz. As loudness is perceived very subjectively due to varying hearing sensitivity, it was purposefully not used to map elevation data. Instead, the *duration* of a sound, and the pace, was chosen to increase with data magnitude (i.e. elevation). To avoid further lengthening of the long test session, the fastest pace that the used technology allowed was chosen. The difference in loudness was mapped to pan, that is to move within two orthogonal directions in the sonified space. East–west movement in the terrain was implemented by taking advantage of the binaural perception between the left ear (i.e. western edge of terrain) and the right ear (eastern edge of terrain). The north–south direction changes are represented by loudness panning between two types of sound waves: sine waves (towards the northern edge of the terrain) and square waves (by going south). As a square wave has a larger integral and thus is louder than a sine wave of equal amplitude, the square wave's amplitude was reduced by 38% to get wave types of equal loudness (Equation 1).

$$r = \int_0^{\pi} 1 - \sin x \; dx \tag{1}$$

Sonification methods

Based on these considerations, two fundamentally different sound parameterizations (A and B) were implemented. To reduce overall cognitive load during the already quite long and demanding experiment, only two continuous variables were selected simultaneously. For *method A*, the origin of the coordinate system was set to the lower left corner of the auditory display; that is, the *x*-axis was rendered to the left ear and the *y*-axis to the right ear. With increasing distance from the origin, the pitch rises on each ear individually up to an octave when reaching the maximum. Elevation was sonified by using the duration of a sound: The higher the elevation at cursor position, the shorter the sound. The minimal duration t between sound and silence was set to 200 ms and mapped to elevation as shown in Equation 2, where t is the duration, h_{max} is the maximum DEM height and h_{cursor} is the elevation at the cursor position.

$$t = \frac{200[ms] \cdot h_{max}}{h_{cursor}} \tag{2}$$

The mapping of the sound variables in *method B* was designed to be more immersive. First, the elevation was mapped to pitch as it is commonly done. Second, the location was mapped on the *xy*-plane to pan. As mentioned above, participants could take advantage of binaural hearing to locate a sound through pan on the *x*-axis. As a novelty, pan was employed a second time on the *y*-axis, using two distinct waveforms. In doing so, participants could orient themselves by comparing nuances of sound changes when moving away from the origin in the *xy*-plane in any direction.

Furthermore, the sound variables mapping of *method C* was almost identical to those of method B, with one slight difference in that participants could additionally choose to use an acoustically rendered elevation profile between two independently chosen points whenever they desired to do so. After selecting a start and an end point by pressing a given key, an audio generator successively played the respective sound for the elevation of every pixel between chosen start/end points with a pace of 15 t/s. Table 3 summarizes the mapped sound variables to the rendered dimensions. The developed sound parameterizations for each method are visualized in Figure 2.

Method A has its sound mapping origin in the lower left corner of the terrain. In method B and method C, the origin of the sound source is located in the center of the terrain, to give the user a stronger feeling of immersion (Figure 2).

Stimuli implementation and data

The auditory stimuli and testing environment were developed using the Java-based *Processing* software including the Java Sound library *Ess*. Three DEMs were constructed for sonification with decreasing levels of abstraction, thus increasing realism and complexity (see Table 2 and Figure 3), but at an identical spatial resolution of 1280 × 800 pixels, and equal color depth of 8 bits (i.e. 256 shades of gray). The most abstract terrain, presented to participants first in the experiment, and thus the least complex sonified

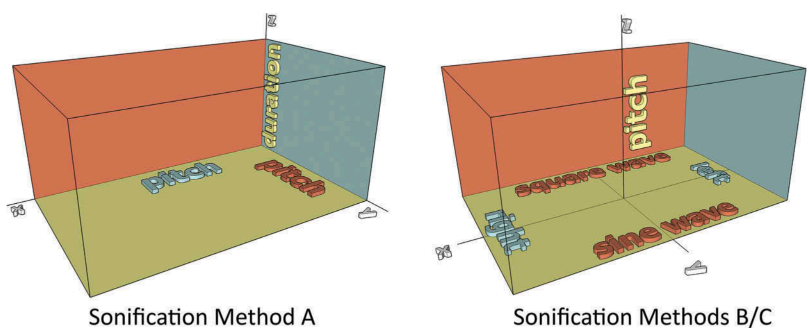


Figure 2. Sonification methods employed in the study.

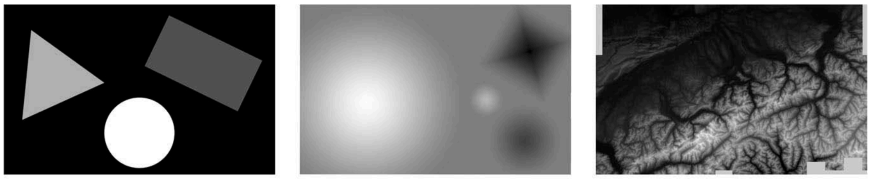


Figure 3. Three tested digital elevation models (DEMs) with decreasing levels of abstraction, increasing degrees of realism and increasing complexity levels from left to right.

terrain, consisted of a discrete elevation model including three simple geometric shapes. An artificial, continuous elevation model containing two protrusions and two depressions served as the second complexity level. Finally, a DEM from Swisstopo served as the third and highest complexity level in our empirical study. The employed Swisstopo DEM extends 256 km in east–west direction and 160 km in north–south direction. The DEM covers about 35,000 km^2 of Swiss territory (84%) and about 5100 km^2 of the adjacent countries including France, Italy, Germany, Austria and Liechtenstein. The mapped elevation ranges approx. between 400 and 4200 m above sea level. It also contains about 1100 km^2 of pixels without any elevation information.

Materials and equipment

The technical equipment employed to run the study should optimally support immersive and focused auditory data exploration, while reducing as many potentially distracting noise sources and sources for visual interference as possible. For various reasons, a *Wacom Bamboo CTH-460* tablet with a 14.7 × 9.1 cm active area was chosen. This was operated with a stylus as an input device for data exploration and to record participant responses. This is because most importantly, currently available tablets with gesture interfaces (i.e. touch pads etc.) did not provide the level of spatial resolution we needed to accurately record user interactions and respective responses on a given area of 1280 × 800 pixels. As the previous test experiences with this kind of tablet and input modality (Heuten *et al*. 2006, Zhao 2006, Delogu *et al*. 2010, Laakso and Sarjakoski 2010) were deemed good, this choice seemed reasonable. The surface area on the tablet beyond the DEM display was covered with adhesive tape so that participants without vision deficiencies could feel the borders of the display while solving the test blindfolded. Participants wore closed, over ear, studio quality earphones (i.e. Beyerdynamic DT770) during the test session with the auditory displays. To increase immersion and focus on the sound source, participants also wore sun glasses (i.e. an S. Tacchini model) with a lens size that covered a large field of view and respective vision angle. The lenses were additionally covered with light-proof, black foil to increase opaqueness. In contrast to prior reviewed studies, an interactive sound legend providing acoustic information at cursor location could be activated during the experiment on request and heard through the earphones, pressing the cursor control keys on the notebook of the experimenter. This sound legend played the highest/lowest tone for comparison between the current pitch and the absolute minimum/maximum. Furthermore, the sound on the current

Figure 4. Test setup with experimenter (left) and test participant (right).

position could be split into the left/right ear and into its proportions of sine/square wave to further help determine the cursor's absolute position on the tablet display.

The experimenter monitored the progress of the experiment using a visual console and by auditory means. Figure 4 depicts the test environment. The experimenter (first author) seated at the notebook monitors progress on the left-hand side of the photograph. A test participant using the stylus on the tablet to solve a task is shown on the right-hand side of the image.

Procedure

A test session lasted approximately 90 min. While the experimental procedure was identical for all tested participants, testing locations, times and lighting conditions varied randomly. The locations typically included a quiet office without acoustic disturbances. First, participants were welcomed to the experiment and seated. After signing the consent form, information about the testing procedure was provided and participants' personal information was gathered, including a question about their hearing. Following that, relevant domain knowledge and abilities were recorded, using a mixture of self-assessments, questionnaires and ability tests. The response categories related to technical ability (i.e. the handling of digital devices and respective input mechanisms), musicality (i.e. their understanding of music) and the ability to read/interpret surface/terrain features (i.e. topographic map reading). For each assessment type, a sample question is given in a summary table below (Table 4). The entire questionnaire and respective scoring of the questions are available in the Appendix A.

Next, participants performed the auditory display portion of the experiment. Participants wore darkened glasses and headphones. They were randomly assigned to one of the three experiment conditions, including one of the three tested sonification methods (A, B and C), as discussed earlier (see Table 3 for details). Participants were

Table 4. Sample questions, grouped by assessment category.

Assessment categories [Appendix A]	Question	Assessed ability
Background questionnaire/ability tests		
Technical ability [1b]	'Without looking at the keyboard. Show me, where the "D" key is'	Ability to find a key while being blindfolded
Musicality [2d]	'If the basic tone were 0% and its octave 100%, estimate the percentage of the pitch relative to the basic tone you hear'. [Experimenter plays an interval]	Ability to assess a pitch within a given range
Terrain interpretation ability [3b]	[Experimenter shows a map] 'Look at this map. What is the orientation of the main valley? Draw it into the map'	Ability to interpret large-scale topographic structures
Auditory display question types		
Abstract-discrete terrain model [4b]	'Describe the shape of the geometric object you are hovering over right now'	Ability to identify simple shapes in an abstract sonified terrain
Abstract-continuous terrain model [5c]	'Find the highest point in the space and set a marker. Then, estimate its x-coordinate relative to the left display edge'	Ability to identify a location by attribute in an abstract sonified terrain
Realistic-continuous terrain model [6e]	'Find a lake and set a marker. Estimate its relative elevation using the minimum and maximum elevation values provided in the sound legend'	Ability to identify geographic features in a realistic sonified terrain and estimate their elevation on a relative scale

asked to interpret location and elevation information (including surface patterns), as accurately as possible, based on the sound heard across the entire sonified display and for various tasks with varying complexity. All participants were tested in three DEM conditions, with increasing complexity levels; always presented in the same sequence. While one could argue that an identical sequence would potentially yield learning effects and introduce response biases, the decision was still taken to apply such a sequence purposefully. Therefore, we counteracted potential learning effects by increasing the questions' difficulty depending on the complexity level. The first complexity level aimed at correctly interpreting number, location, shape, orientation and height of different geometric shapes on a discrete DEM. The second complexity level aimed at additionally finding the correct location and accurately interpreting surface characteristics such as slope of local and absolute peaks and sinks on a continuous DEM. Finally, the third complexity level additionally aimed at correctly interpreting absolute elevation information and further morphological patterns.

We began each test sequence by orally explaining the general structure of the DEM and by introducing the main goals of the tasks. Each DEM condition included the invitation to first freely explore the sonified terrain. This was followed by a series of trials with tasks designed to find particular locations in the DEM that matched given conditions, or to interpret a surface feature at a given location. Participants were closely monitored while solving the tasks. If participants found difficulty with a particular task, this task was repeated. Participants could request their scores after completion of the experimental session.

Time, cursor position and key button presses were recorded digitally and saved into a log file, including participant responses to test questions. This allowed us to later compare self-reports on performance with actual performance to further provide insights into the three response accuracy relations as shown in Figure 2. Depending on the task, responses were recorded at different levels of detail. Whenever possible, quantitative responses were measured on a continuous measurement level which was automatically recorded during the test session by the testing software. This allowed us to compute a ratio score for the three components of the response accuracy (as shown in Figure 2). Other responses were manually coded when necessary, using an ordinal scoring system, as follows: zero for non-satisfactory responses, 0.5 for fair responses and 1 for a satisfactory response. The experimenter graded participant's answers based on the valuation key listed in Appendix A while the requirements for obtaining the specified points depended on the task. Tasks within the same experimental condition were weighted equally. Participant responses were averaged and normalized to scores ranging from 0 to 1, and the participant performance was statistically assessed for each portion of the experiment. A complete overview over all tasks is provided in Appendix A, including the scoring scheme, the weighting factors and the formula for calculating the response scores.

Results

Below, the collected data across sonification methods and terrain complexity levels are detailed. This is followed by correlation analyses to further evaluate users' background knowledge on response accuracy.

How are sonified terrains interpreted by users?

As the sonification of discrete data has been typically investigated in prior work, we specifically wondered, whether and how sonification might also work in comparison with continuous data, such as a continuous DEM. The box plots in Figure 5 show participants' response accuracies organized by increasing complexity of the sonified terrain and grouped by sonification method. The black (trend) lines connect mean accuracy across sonification methods and terrain complexity. There are only small differences across sonification methods, but rather high standard deviations (12.8–13.4%) for all groups in this condition.

Surprising to us, and in direct contrast to what prior research suggests (e.g. Krygier 1994), higher response accuracies (and lower response variability) were observed with the continuous terrain models, compared to the seemingly easier discrete terrain. For the discrete DEM, participants' average response accuracy overall ranges from 69.1% to 72.7%. Figure 5 also shows that participants' average accuracy increases steeply from the discrete to the simple continuous terrain model (to approx. 86.7%), when using sonification methods B and C, but then levels off at about ±1% accuracy for the most complex terrain model. In contrast, with sonification method A, participants' average accuracy decreases from 76.7% to 73.3% across complexity levels. It is interesting to see that lower response averages (i.e. with the discrete terrain) tend to be accompanied with noticeably larger response variability. A repeated measures ANOVA suggests a significant main effect of terrain complexities (F [1, 59] = 14.82, $p < 0.001$, $\eta^2 = 27.3\%$, $f = 61.3\%$). A *post-hoc* analysis including a Bonferroni correction yields significant differences in response accuracy between terrain complexity levels 1 and 2

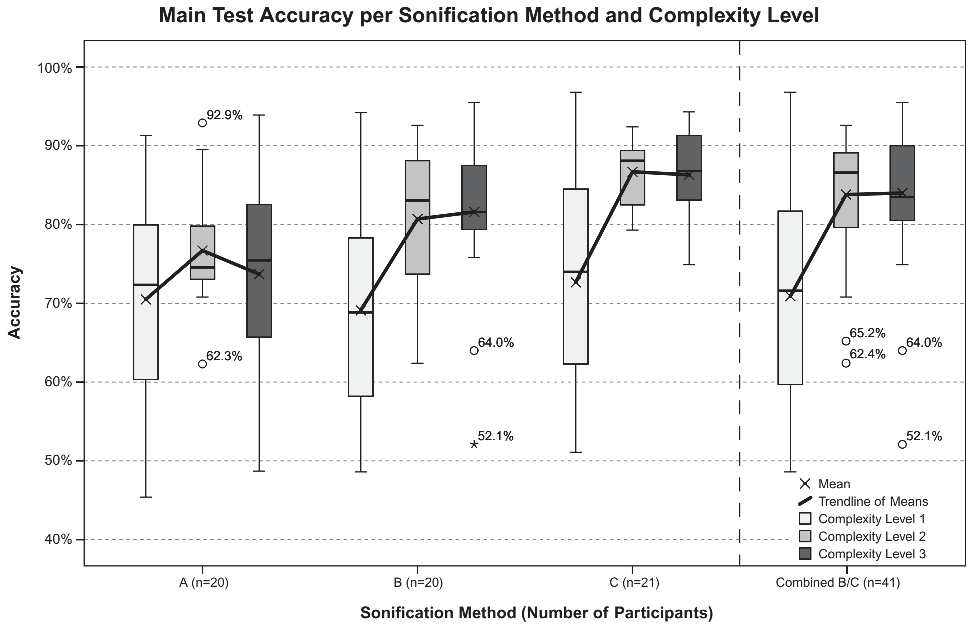


Figure 5. Response accuracies across sonification methods and terrain complexity levels.

($p < 0.001$), and between terrain complexity levels 1 and 3 ($p < 0.01$), suggesting that participants' response accuracy was significantly worse with the discrete terrain model. In other words, quite in contrast to prior research suggesting that sonification might be best used for ordinal data (e.g. Krygier 1994), our participants perform even better with the more complex continuous terrains, compared to the discrete terrain model.

An interaction effect was also found across sonification methods A and B/C with complexity levels (F [1, 59] = 3.541, $p < 0.05$, $\eta^2 = 6.5\%$, $f = 26.4\%$). This suggests that not only terrain complexity is relevant for participants' performance with an auditory display but also the employed sonification method, as hypothesized. Thereby, attention is now drawn to the assessment of the developed sound mapping methods.

What role does the sound mapping method play for terrain interpretation?

One aim of this research was to propose various continuous sound mapping methods to assess their strength and weaknesses for the sonification of DEMs. As shown for the results on terrain complexity above, there seems to be an interaction effect across sonification methods. Indeed, the right panel of Figure 6 reports lowest response accuracy with method A (M = 74%, SD = 7%), followed by method B (M = 77%, SD = 6%) and lastly highest elevation interpretation accuracy with method C (M = 82%, SD = 5%). A repeated measures ANOVA suggests these differences to be significant (F [1, 59] = 8.990, $p < 0.001$, $\eta^2 = 23.7\%$, $f = 7\%$, $d = 79.7\%$). A *post-hoc* analysis using a Bonferroni correction indicates that participants' response accuracy differs significantly across sonification methods A and C ($p < 0.001$). The magnitude of the effect expressed by Cohen's d (1.331) is also large. When using Tukey's HSD, participants'

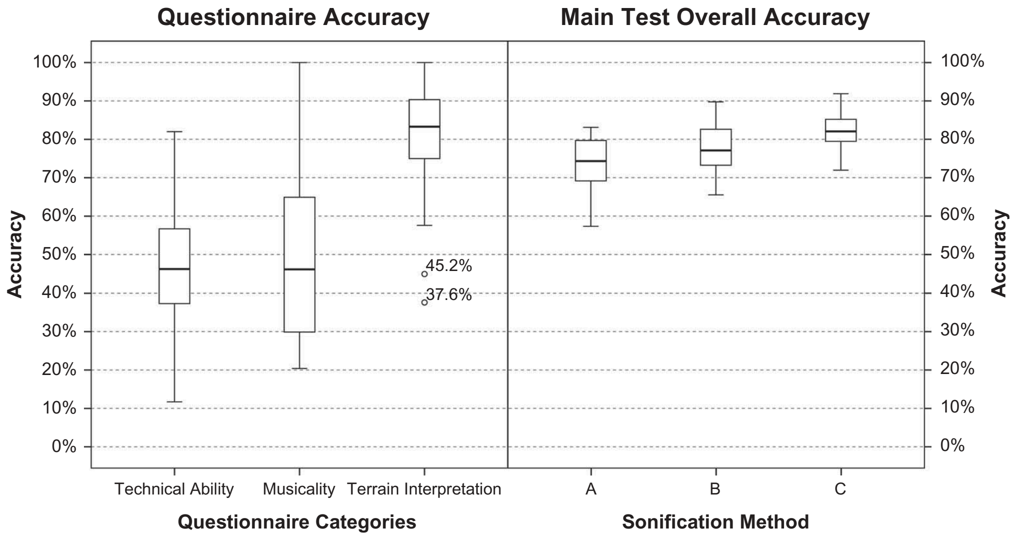


Figure 6. Participant response accuracy based on expertise (left) and on sonification method.

response accuracy not only differs significantly between sonification methods A and C ($p < 0.001$) but also between methods B and C ($p < 0.05$). The size of the effect is again large ($d = 0.797$). In other words, a significant effect on terrain interpretation could be found when participants additionally had the opportunity to peruse a sonified elevation profile (the only difference between sonification methods B and C).

As the difference between sonification methods B and C is only related to the addition of the elevation profile, the respective results were merged. Below, we thus report aggregated results for a new, combined B/C method. Again, combined methods B/C yield the highest response accuracy (M = 80%, SD = 6%) compared to method A (M = 74%, SD = 7%). This difference is significant (F [1, 59] = 11.226, $p < 0.01$, η^2 = 16.0%, f = 43.6%, t [59] = −3.353) showing a large effect ($d = 0.915$), with observed power of 0.959. As previously demonstrated, participants perform significantly better with either method B or C compared to A. In other words, participants are more accurate in their terrain interpretation when the auditory variable pan is employed for localization and pitch is used for elevation (methods B/C), compared to method A, with the auditory variable pitch assigned to localization, and the sound variable duration mapped to represent elevation. This difference seems particularly evident when participants interpret continuous DEMs, again, somewhat surprising based on prior research.

Furthermore, the participants' accuracy was analyzed when having to estimate absolute elevation, comparing sonification methods A (M = 79%, SD = 11%) with collapsed methods B/C (M = 87%, SD = 7%). A t-test yields significant differences across both sonification methods (t [59] = −3.187, $p < 0.01$, $|\rho|$ = 39.2%). Again, our participants are significantly better at estimating elevation when the sound variable pitch is employed to represent terrain, compared to the sound variable duration. Aside from investigating the influence of the data characteristics, and the sonification methods applied to discrete and continuous elevation data in sonification, assessment of the human factor was also of interest.

Domain knowledge and training

Below, the results relating to users' background and training are discussed, as hypothesized that participants' performance with auditory displays might also depend on domain expertise differences as prior works contends (i.e. Bearman and Lovett 2010), but which has not been systematically investigated until now in sonification research in GIScience. The detailed catalog of assessed background and training questions is listed in Appendix A (supplementary).

Boxplots presented in the left panel in Figure 6 show participants' average accuracy overall, assessed with the background knowledge and ability questionnaire (normalized across question types). The box plots are organized by questions related to technical ability, musicality and the ability to identify terrain features. The boxplots in the right panel in Figure 6 depict average response accuracy across sonification methods for direct comparisons.

As shown in Figure 6, users' background knowledge in reading terrain features tends to be better and less variable than their technical ability or musicality. Participants' accuracy is especially high for the tasks related to the identification of terrain features (>75%). Musicality and technical ability are somewhat low with approximately 50% of answers being correct. Overall, participants' auditory response accuracy is higher and shows less variability in the main experiment compared to the background and ability questionnaire.

As one can see in Figure 6, overall, participants' variability in performance due to domain expertise seems to be larger, than the accuracy of response variability when working with the sonified displays.

One of the compelling factors mentioned for the potential success of data sonification is typically users' background and level of training in music. Thus, further investigation was carried out into whether and how domain expertise and relevant abilities (i.e. technical ability, musicality and topographic map reading) might explain response accuracy with sonified terrains. The results of this analysis are presented in the next section below.

How does background knowledge and training influence terrain data interpretation with sonified displays?

Correlation analyses were conducted including Pearson's r and Spearman's rho for each participant (applying a Fisher's r-to-z transformation and converting back to r) to assess potential associations of individual domain knowledge and abilities scores (assessed with the background questionnaire) with sonified data interpretation accuracy across DEM complexity levels (recorded in the main experiment). Results are shown in Table 5. No significant correlations could be found for the lowest (discrete) DEM complexity level. However, technical ability, musicality and topographic feature interpretation ability seem to be associated with sonified terrain interpretation for terrain complexity levels 2 and 3. As shown in Table 4, average correlations are significant and moderately strong

Table 5. Relationships of participant performance using the sonified display with domain expertise and ability.

Terrain complexity level		Domain knowledge and abilities Pearson's correlation			Domain knowledge and abilities Spearman's rho		
		Technical	Musical	Topographical	Technical	Musical	Topographical
Discrete	Correlation	0.114	−0.055	0.183	0.098	−0.061	0.181
	Power	0.221	0.110	0.412	0.186	0.119	0.405
Continuous (abstract)	Correlation	0.342**	0.432***	0.359**	0.377**	0.411***	0.334**
	Power	0.863	0.972	0.893	0.920	0.956	0.848
Continuous (realistic)	Correlation	0.233*	0.351**	0.296*	0.287**	0.387**	0.431**
	Power	0.571	0.880	0.759	0.734	0.932	0.971

Pearson's correlations (r) and Spearman's rho (ρ) including statistical power.
*$p < 0.05$; **$p < 0.01$; ***$p < 0.001$.

for the abstract continuous model (level 2) with r coefficients between 0.34 and 0.43. For the most complex, that is realistic sonified terrain (level 3), r coefficients range only from 0.23 to 0.35 and thus are somewhat weaker and more variable than for terrain complexity level 2. These results suggest that indeed relevant domain knowledge (especially topographic map reading) and domain-related abilities (i.e. musicality and technical ability) can influence the readability of a sonified display.

Discussion

We set out to investigate the utility of PMS for spatial data exploration and wished to empirically assess how people might interpret auditory displays showing DEMs of varying complexity. We were also interested in assessing how an auditory elevation profile might support users in interpreting auditory terrain displays. Furthermore, we wished to identify to what extent domain expertise and respective relevant abilities (i.e. technical ability, musicality and topographic map reading) might explain response accuracy with auditory displays.

How people might interpret auditory displays showing DEMs

First, the results show that continuous spatial elevation models *can* indeed be successfully interpreted by (blindfolded) users, overall yielding a mean accuracy up to 86.7% (method C) using PMS. What is more, participants' mean accuracy was even higher when interpreting continuous DEMs compared to the seemingly easier discrete DEMs. This finding contradicts prior research, suggesting that sound variables would not be useful for the interpretation of continuous data (Krygier 1994). These results show that participants were indeed able to more accurately interpret continuous terrain sonifications compared to discrete versions. Specifically, participants' accuracy significantly improved by 6% with method A and 13% with sonification method B/C. In contrast to method A, participants' accuracy did not significantly decline with increasing complexity, thus moving from abstract-continuous to realistic-continuous models using method B/C. In other words, the sonification method itself is also an important factor to consider, as it may add perceptual and cognitive load to the interpretation of terrain models, irrespective of the complexity of the terrain.

Given that the stimuli sequence was identical for all tested participants, the improvement in response accuracy from the first to the second terrain complexity level might have been influenced by a potential learning effect. If learning effects might have confounded results (albeit differently across sonification methods), then this begs the question why response accuracy did not further increase for the third, most complex level. One could explain this by a potential ceiling effect. One the one hand, method A is perhaps not appropriate enough to explore a natural terrain model that features complex elevation variations, over varying terrain distances. On the other hand, method B/C might develop its full potential only at the most complex level. As explained earlier, additional (more complex) questions were also included for the more complex terrain trials. These questions were specifically adapted to the changing affordances at each level. We did this to further distinguish performances across complexity levels. Perhaps, the rather long duration of the entire experiment might have negatively affected participants' performance for the most complex level always presented last due to fatigue. This might have especially affected participants with method A, as this method also might have been more difficult. A potential learning effect in our study might mean for future experiments to randomize the stimuli presentation order, as to better disentangle this potential confound. It might also mean that real-world uses of sonified displays might indeed require a training phase to unleash its fully potential. At this point in time, however, it is not clear whether a learning effect might have affected results, and how, and if so, how much it matters.

Nevertheless, these results still shed new light on sonification of spatial data, considering that most sonification projects in GIScience have been conducted with categorical spatial data (Fisher 1994, Lodha *et al*. 1999, Brauen 2006, Zhao 2006, MacVeigh and Jacobson 2007, Bearman and Lovett 2010, Delogu *et al*. 2010, Laakso and Sarjakoski 2010), and there are very few studies using continuous data, but mostly outside of GIScience (Heuten *et al*. 2006, Yeo and Berger 2008, Geronazzo *et al*. 2016).

The results also indicate that the specific combination of sound variables and respective mapping to data dimensions needs to be carefully considered in sonified displays, as it can improve or hinder data interpretation. Specifically, for our study, mapping the sound variables *pitch* to elevation and *pan* for orientation along cardinal directions (N–S–E–W), successfully supported participants in more accurately interpreting sonified displays (methods B/C), compared to using sound variables *pitch* for orientation and *duration* for elevation (method A). Variable *pitch* was employed for all tested sonification methods, as prior work suggests, and this found that it indeed serves as a key sound variable. Even though many of the participants could successfully orient themselves in the sonified terrain by distinguishing two pitches, it was not suggested that *pitch* is intuitive for spatial orientation, as its interpretation is based on a musical convention that might not be universally understood (i.e. full screen height/width was equal to an octave in our study). Moreover, the comparison of two pitches on different ears requires training.

The proposed method B/C is based on the same sound variables as proposed by Fisher (1994), but it was decided to map them differently. As loudness is perceived very subjectively (Campenhausen 1993) and depends on pitch (Genuit 2010), loudness was not used to represent a magnitude change of a mapped attribute. Instead, it was chosen to be mapped to the location change (i.e. pan) to take advantage of binaural localization in a soundscape and thus to increase the immersive experience of sound spatialization. Although spatialized sounds were applied in B and C (Nasir and Roberts 2007), immersion might have been increased by implementing HRTF, as suggested by various authors (Heuten *et al*. 2006, Zhao 2006, Geronazzo *et al*. 2016). Still, these results confirm that the appropriate spatialization of sound variables can indeed support listeners to orient themselves in a soundscape. For terrain interpretation, the continuous data variable *elevation* is more accurately interpreted using *pitch* compared to *duration*.

The empirical results for the assessment of two different sound parameterization techniques also suggest that the appropriate use of spatial metaphors, in this case auditory mappings, is also key for the success of sonified displays (Kuhn 1996, Fabrikant 2017). Especially, the consideration of the sound variable pitch for mapping elevation or the use of pan for orientation in the soundscape transferred very well into sound spatialization. It was thus suggested to use increasing pitch as a fundamental metaphorical mapping of the MORE-IS-UP image schema in sonification, as it serves as the basis for understanding any type of linguistic, auditory or graphical metaphor to communicate the magnitude or quantity of an item of interest (Fabrikant 2017).

How auditory elevation profiles might support users in interpreting a DEM

Evidence was found that an auditory elevation profile (method C) could facilitate sonified DEM interpretation. However, a significant difference could only be detected when compared directly to a similar sound parametrization. Here again, it is suggested to more carefully consider the employed spatialization technique. The SOURCE-PATH-DESTINATION image schema and the MORE-IS-UP image schema might be more effectively combined in future studies. Similarly to Yeo and Berger (2008), one could consider to map the sound variable *pitch* to elevation with a sonification method that plays a series of sounds within a short amount of time. This is a solution to create a holistic hearing sensation that might also reduce cognitive load. Such an approach could be interesting for further studies, for example, to be able to distinguish between important and unimportant features and to get a general impression of the sonified space. The visual analog here would be the information visualization mantra ‘overview-first, then details-on-demand’ proposed by Shneiderman (1996), as many participants had difficulty to remember the already retrieved information, especially while interpreting complex, realistic elevation models.

Does domain expertise explain response accuracy?

As many prior empirical visualization studies show, we also find that background and training can influence user’s effectiveness and efficiency in decision-making with a sonified display design. Specifically, the general ability of handling technical devices (i.e. blind typing), musicality and terrain feature interpretation ability facilitated the interpretation of the auditory display in our study, even if this effect turned out not to be a strong predictor for each of the tested complexity levels. Even if the pattern is not straightforward, user background and training could have interacted in predictable ways with potential learning effects, as participants with better abilities generally achieved significantly higher accuracy, at least at complexity levels 2 and 3. A base capacity due to background and training might be necessary to achieve high interpretation accuracy in sonification. Bearman and Lovett (2010, p. 310) suggest that knowledge of geographical data 'resulted in an increased likelihood of a correct answer' with sonified displays, however, without being statistically significant. In contrast, Deutsch (2013) asserts that musical understanding indeed plays an important role in understanding of complex, multivariate auditory displays. These results thus also replicate many visualization studies in suggesting that background knowledge or expertise, for example, in handling digital devices blindfolded, map reading and spatial data interpretation will indeed improve performance with other kinds of visuospatial displays.

Summary and outlook

This study empirically investigated the potential of auditory displays for exploring DEMs by means of three PMS methods. The empirical results suggest, contrary to own expectations and the results of prior research, that the sound variable pitch can be successfully interpreted by users of sonified displays when it is applied to represent continuous spatial elevation data. These encouraging results further suggest that sonification, yet another spatialization method, indeed has untapped potential for spatiotemporal data exploration in GIScience and beyond – even on complex terrain models. Perhaps given some time and practice, it could become as popular as (geographic) information visualization, to support (or even replace?) visual data encodings, when and where needed. Particularly, PMS might be an additional attractive multivariate data dimensionality technique for sound spatialization, as PMS can flexibly map various parameters representing different data dimensions to a set of sound variables. Sound variables can be further fine-tuned by adaptive parameterizations, dependent on the data exploration task at hand.

Contrary to prior work suggesting only nominal or ordinal data for sonification, unique empirical evidence was found indicating that continuous spatial data models are not interpreted significantly worse when sonified, compared to sonified displays showing discrete spatial data. As a potential learning effect might have occurred due to the purposefully chosen test setup, at this point, it is not clear how much training might be needed, to achieve significant accuracy improvements. Novel empirical evidence was also found to suggest that the sound parameter mapping methods themselves offer new and exciting opportunities for sound spatialization research in GIScience. Future work could, for example, empirically explore successful sound variable combinations that could effectively and efficiently support users in exploring geographic data in sonified landscapes. It remains to be further empirically investigated how sound variables might best capture essential spatial data characteristics, for example, when mapping various data source domains (e.g. continuous terrains of different kinds) into sonified target domains, represented as spatialized soundscapes. Our own successful sonification method uses binaural pan for west–east orientation and pan between square and sine waves for north–south orientation to localize elevation data represented with the pitch sound variable in a sonified terrain. As sound information must be decoded and processed sequentially, the amount of perceived sound details also depends on the temporal level of detail or on the auditory scale. Auditory generalization across scales in sonified terrains could be another line of exciting future research.

Furthermore, as we already know this for visuospatial displays, domain knowledge and expertise are also important factors to consider to predict data exploration success with sonified displays. For soundscapes that contain terrain data, musicality and the ability to read topographic maps are certainly relevant. Furthermore, a general ability to handle technical devices (without sight) seems to play an important role as well.

Future studies will concentrate on extending this PMS approach to three or more sound variables, which has not been investigated yet. These results further encourage empirical sonification research concerning the communication of continuous geographical datasets. For example, census data or other types of data collected on the ratio scale from surveys could be sonified and evaluated, without having to transform the data to the ordinal level. As the purposefully chosen experimental setup gradually increased terrain complexity, future studies might further investigate if and how training might influence performance with sonified displays. Finally, the sighted participants were blindfolded for the experiment. It remains to be seen whether performance differences using auditory displays might also be explained by the level of visual impairment.

Note

Appendix A (supplementary) contains the administered questions, including level of measurement of collected data, and weighting schemes for scoring. Appendix B (published online) shows the questionnaire used for the study. The developed software including use instructions, the spatial dataset for the highest complexity level and the applied methods B/C as well as some audio streams and short films are available as a JAVA applet on the Web at http://www.geo.uzh.ch/~jschito. The source code and the calculation of the variable operationalization are available on request from the first author.

Acknowledgments

This project was inspired by and is in honor of Prof. Pete Fisher's trailblazing research on sonification in GIScience. We wish to thank all participants who provided valuable data by serving as research participants in the experiment. We also thank the anonymous reviewers who provided feedback to improve the communication of our research results.

Disclosure statement

No potential conflict of interest was reported by the authors.

ORCID

Joram Schito http://orcid.org/0000-0003-1138-8496
Sara Irina Fabrikant http://orcid.org/0000-0003-1263-8792

References

Adhitya, S. and Kuuskankare, M., 2012. Sum: from image-based sonification to computeraided composition. *In*: *Proceedings of the 9th international symposium on computer music modelling and retrieval. Presented at the international symposium on computer music modelling and retrieval (CMMR)*. London: Queen Mary University of London.

Afonso, A., *et al*., 2005. A study of spatial cognition in an immersive virtual audio environment: comparing blind and blindfolded individuals. *In*: *Proceedings of the 17th meeting of the international conference on auditory display* (icad2011), *20-23 June, 2011*. 6–9.

Amon, R., 2005. *Lexikon der Harmonielehre*. Wien: Doblinger.

Andrienko, G., *et al*., 2013. *Visual analytics of movement*. Berlin, Germany: Springer Science & Business Media.

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

Bly, S., 1982. *Sound and computer information presentation*. CA, USA: Lawrence Livermore National Lab. California Univ., Davis (USA).

Brauen, G., 2006. Designing interactive sound maps using scalable vector graphics. *Cartographica: International Journal for Geographic Information and Geovisualization*, 41 (1), 59–72. doi:10.3138/5512-628G-2H57-H675

Brazil, E. and Fernström, M., 2011. Navigation of data. *In*: T. Hermann, J. Neuhoff, and A. Hunt, eds. *The sonification handbook*. Berlin: Logos, 509–523.

Bujacz, M., Skulimowski, P., and Strumillo, P., 2011. Sonification of 3d scenes using personalized spatial audio to aid visually impaired persons. *In*: *Proceedings of the 17th international conference on auditory display (icad2011),* 20-23 June, 2011. Budapest, Hungary. International Community for Auditory Display

Campenhausen, C., 1993. *Die Sinne des Menschen. Einführung in die Psychophysik der Wahrnehmung*. 2nd ed. Stuttgart: Georg Thieme Verlag.

Cartwright, W., Peterson, M.P., and Gartner, G., eds. 1999. *Multimedia cartography*. Berlin Heidelberg: Springer.

Delogu, F., *et al*., 2010. Non-visual exploration of geographic maps: does sonification help? *Disability & Rehabilitation: Assistive Technology*, 5 (3), 164–174. doi:10.3109/17483100903100277

Deutsch, D., 2013. *The psychology of music*. 3rd ed. London: Elsevier.

Fabrikant, S.I. (2017). Spatialization. In: Castree, N., Goodchild, M.F., Kobayashi, A., Liu, W, Marston, R., Richardson, D. (eds). AAG International Encyclopedia of Geography: People, the Earth, Environment, and Technology. Wiley. doi:10.1002/9781118786352.wbieg0812

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

Genuit, K., ed., 2010. *Sound-Engineering im Automobilbereich: Methoden zur Messung und Auswertung von Geräuschen und Schwingungen*. 1st ed. Berlin: Springer.

Geronazzo, M., *et al*., 2016. Interactive spatial sonification for non-visual exploration of virtual maps. *International Journal of Human-Computer Studies*, 85, 4–15. doi:10.1016/j.ijhcs.2015.08.004

Goldstein, E.B., 2002. *Wahrnehmungspsychologie*. 2nd ed. Berlin: Spektrum Akademischer Verlag.

Grond, F. and Berger, J., 2011. Parameter mapping sonification. *In*: T. Hermann, J. Neuhoff, and A. Hunt, eds. *The sonification handbook*. Berlin: Logos, 363–397.

Heuten, W., Wichmann, D., and Boll, S., 2006. Interactive 3d sonification for the exploration of city maps. *In:* A. Morch, et al., eds. Proceedings of the 4th Nordic Conference on Human-computer interaction: changing roles, 14–18 October 2006, Oslo, Norway. New York, NY: ACM.

Jacobson, R.D., 2010. Non-visual geographies. *In*: B. Warf, ed. *Encyclopedia of geography*. London: Sage Publications.

Josselin, D., 2011. Spatial analysis using sonorous cartography. Some propositions. *In*: A. Ruas, ed. *Proceedings of the 25th international cartographic conference*. 3-8 July 2011, Paris, France.

Josselin, D., *et al*., 2016. Sonorous cartography for sighted and blind people. *In*: *AGILE'2016-19th AGILE International conference on Geographic Information Science,* June 2016, Helsinki, Finland.

Kalinak, K., 1992. *Settling the score: music and the classical Hollywood film*. Madison: University of Wisconsin Press.

Kramer, G., 1992. *Auditory display: sonification, audification, and auditory interfaces*. Santa Fe: Addison-Wesley.

Krygier, J.B., 1994. Sound and geographic visualization. *In*: *Visualization in modern cartography*. New York: Pergamon, 149–166.

Kuhn, W., 1996. Handling data spatially: spatializating user interfaces. *In:* M.-J kraak and M. molenaar, eds. Proceeding of *7th International Symposium on spatial Data handling, SDH'96, Advances in GIS research II, in Delft,* 12–16 August 1996, The Netherlands: IGU

Laakso, M. and Sarjakoski, T., 2010. Sonic maps for hiking – use of sound in enhancing the map use experience. *The Cartographic Journal*, 47 (4), 300–307. doi:10.1179/000870410X12911298276237

Lakoff, G. and Johnson, M., 1980. The metaphorical structure of the human conceptual system. *Cognitive Science*, 4 (2), 195–208. doi:10.1207/s15516709cog0402_4

Lodha, S.K., Joseph, A.J., and Renteria, J.C., 1999. Audio-visual data mapping for GIS-based data: an experimental evaluation. *In*: *Proceedings of the 1999 workshop on new paradigms in information visualization and manipulation in conjunction with the eighth ACM international conference on Information and knowledge management*. 41–48. Kansas City, Missouri: ACM.

MacVeigh, R., and Jacobson, R.D., 2007. Implementing Auditory Icons in Raster GIS. *In: Proceedings of the 13th International Conference on Auditory Display,* 26–29 June 2007, Montréal, Québec, Canada. 530–535

McCormick, B.H., DeFanti, T.A., and Brown, M.D., 1987. Visualization in scientific computing. *ACM SIGGRAPH Computer Graphics*, 21, 6.

Nasir, T., and Roberts, J.C., 2007. Sonification of spatial data. *In: Proceedings of the 13th International Conference on Auditory Display,* 26–29, 2007 June, Montréal, Québec, Canada. 32–37

Paterson, N. and Conway, F., 2014. Engagement, Immersion and Presence. *In:* K. Collins, B. Kapralos, and H. Tessler, eds. *The Oxford Handbook of Interactive Audio*. Oxford, UK: Oxford University Press.

Pollack, I. and Ficks, L., 1954. Information of elementary multidimensional auditory displays. *The Journal of the Acoustical Society of America*, 26, 155–158. doi:10.1121/1.1907300

Ruotolo, F., *et al*., 2013. Immersive virtual reality and environmental noise assessment: an innovative audio–visual approach. *Environmental Impact Assessment Review*, 41, 10–20. doi:10.1016/j.eiar.2013.01.007

Schiewe, J. and Weninger, B., 2012. Acoustic encoding of quantitative information in maps – results of a study comparing with classical forms of representation. *Kartographische Nachrichten*, 3, 126–135.

Schwartz, H., 2003. The indefensible ear: a history. *In*: M. Bull and L. Back, eds. *The auditory culture reader*. New York: Berg, 487–501.

Shneiderman, B., 1996. The eyes have it: a task by data type taxonomy for information visualizations. *In: VL '96 Proceedings of the 1996 IEEE Symposium on Visual Languages*, 03–06, 1996 September. Washington, DC: IEEE Computer Society

Simonnet, M., *et al*., 2009. SeaTouch: a haptic and auditory maritime environment for non visual cognitive mapping of blind sailors. *In*: G. Ligozat, *et al*., eds. *Spatial information theory. Presented at the COSIT 2009*. Aber Wrac'h, France: Springer, 212–226.

Théberge, P., 2005. Sound maps: music and sound in cybercartography. *Modern Cartography Series*, 4, 389–410.

Ware, C., 2008. *Visual thinking for design*. San Francisco, CA: Morgan Kaufmann.

Williams, M.G., Smith, S., and Pecelli, G., 1990. Computer-human interface issues in the design of an intelligent workstation for scientific visualization. *ACM SIGCHI Bulletin*, 21 (4), 44–49. doi:10.1145/379106

Yeo, W.S., and Berger, J., 2008. Raster scanning: a new approach to image sonification, sound visualization, sound analysis, and synthesis. *In: International Computer Music Conference*. http://hdl.handle.net/2027/spo.bbp23722006.008

Yeung, E.S., 1980. Pattern recognition by audio representation of multivariate analytical data. *Analytical Chemistry*, 52 (7), 1120–1123. doi:10.1021/ac50057a028

Zhao, H., 2006. *Interactive Sonification of Abstract Data - Framework, Design Space, Evaluation, and User Tool*. Dissertation. University of Maryland, College Park.

Appendix A (for publication in the printed edition)

Administered questions translated from German into English, including level of detail of collected data and weighting schemes for scoring.

Test category	Question type	Measurement level	Scoring range [0,1]	Test type	Weight
Technical ability	**Video game experience** *Score ranges from 0.0 to 1.0 and are calculated as follows:* $a = \frac{\text{number of game console generations the participant was or is used to play with}}{5}$ **The number of game console generations is limited to 5. For further information, see: Landsman, V., & Stremersch, S. (2011). Multihoming in two-sided markets: an empirical inquiry in the video game console industry. *Journal of Marketing, 75*(6), 39–54**	Discrete	$\left\{x = \frac{k}{5} \mid 0 \leq x \leq 1 \wedge k \in \mathbb{N}_0\right\}$	Experience	$\frac{1}{4}$
	Blind-typing self-report *Task 1 – Score ranges from 0.0 to 1.0 and are calculated as follows:* *1.0 = Participant could find the marks on the 'F' and 'J' keys without looking onto the keyboard* *0.0 = Participant could not find the marks on the 'F' and 'J' keys without looking onto the keyboard* *Task 2 – Score ranges from 0.0 to 1.0 and are calculated as follows:* *1.0 = Participant could find 3 arbitrarily chosen keys without looking onto the keyboard* *0.0 = Participant could not find 3 arbitrarily chosen keys without looking onto the keyboard* *Task 3 – Score ranges from 0.0 to 1.0 and are calculated as follows:* *1.0 = Participant is experienced in blind typing* *0.0 = Participant is not experienced in blind typing* **This answer was further used in the next aptitude assessment**	3× Dichotomous	$\{0, 1\}$	Experience aptitude	$\frac{1}{4}$
	Blind-typing ability test *Score ranges from 0.0 to 1.0 and are calculated as follows:* $a = \frac{precision_{min}\langle experience \cdot time\ [s] \cdot (1 + 0.25 \cdot mistakes\ [])\rangle}{precision_{subject}\langle experience \cdot time\ [s] \cdot (1 + 0.25 \cdot mistakes\ [])\rangle}$ *experience = 1 if the participant was experienced in blind-typing or* *experience = 1.25 if the participant was not experienced in blind-typing* **After the assessment of the whole test battery, the participant with the best score (equal to the fewest errors) has been chosen as reference in the numerator. This value is divided by the participant's score to obtain a relative score between 0% and 100%**	Continuous	$\{x \in \mathbb{R} \mid 0 \leq x \leq 1\}$	Aptitude	$\frac{1}{4}$

(*Continued*)

(Continued).

Test category	Question type	Measurement level	Scoring range [0,1]	Test type	Weight
	Precision of handling a pen tablet *Score ranges from 0.0 to 1.0 and are calculated as follows:* $a = \frac{precision_{subject}\langle q\rangle}{precision_{max}\langle q\rangle}$ *with* $q = \frac{1000 \cdot registeredPoints\ []}{time\ [ms] \cdot SD^{0.0625}\ []}$ **The participant gazed at a screen visualizing straight lines he/she should trace as precisely as possible. Simultaneously, a software registered the time needed, the number of points and the standard deviation from the traced to the given lines. After the assessment of the whole test battery, the exponent 0.0625 has been empirically determined to assign a score of 1.0 for best performance**	Continuous	$\{x \in \mathbb{R} \mid 0 \leq x \leq 1\}$	Aptitude	$\frac{1}{4}$
Musicality	**Musical experience** *Score ranges from 0.0 to 1.0 and are calculated as follows:* *1.0 = Professional musician* *0.67 = Serious amateur musician* *0.33 = Self-educated occasional musician* *0.0 = Developed musical preferences* **Participants had to assess their musical experience by using following classification: Vogt, S. (2005). *Clubräume-Freiräume: musikalische Lebensentwürfe in den Jugendkulturen Berlins* (Vol. 6). Bärenreiter, 44**	Discrete	$\left\{x = \frac{k}{3} \mid 0 \leq x \leq 1 \wedge k \in \mathbb{N}_0\right\}$	Experience	$\frac{1}{6}$
	Precision of singing five western solfeggio *Score ranges from 0.0 to 1.0 and are calculated as follows:* *1.0 = No note of a western scale varied more than one major diesis (62.6 cents)* *0.5 = Maximally one note of a western scale deviated less than a minor second (100.0 cents)* *0.0 = Worse than described above*	5× Discrete	$\left\{x = \frac{k}{2} \mid 0 \leq x \leq 1 \wedge k \in \mathbb{N}_0\right\}$	Aptitude	$\frac{1}{6}$
	Precision of recognizing five western intervals within a perfect octave *Score ranges from 0.0 to 1.0 and are calculated as follows:* *1.0 = Intervals were correctly recognized* *0.5 = Answer deviated less than a minor second* *0.0 = Worse than described above*	5× Discrete	$\left\{x = \frac{k}{2} \mid 0 \leq x \leq 1 \wedge k \in \mathbb{N}_0\right\}$	Aptitude	$\frac{1}{6}$

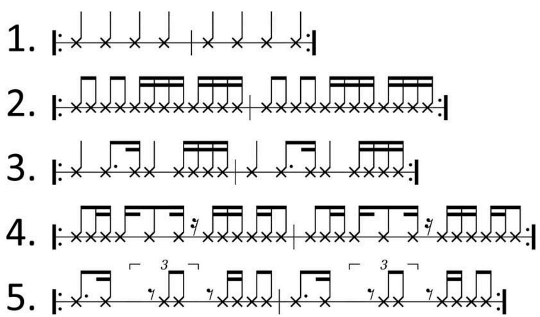
(*Continued*)

(Continued).

Test category	Question type	Measurement level	Scoring range [0,1]	Test type	Weight
	Precision of quantifying five western intervals within one perfect octave by using a percentage scale *Score ranges from 0.0 to 1.0 and are calculated as follows*: $a = \left(1.08 \cdot exp\left(\sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - E(X))^2}\right)\right)^{-1}$ **After the assessment of the whole test battery, the constant of 1.08 has been empirically determined to assign the best participant a score of 1.0 points. $E(X)$ is the true value, whereas x_i is the given answer within a range between 0% and 100%. A perfect octave with equal temperament consists of 12 equal minor seconds of 100 cents each. Thus, a perfect octave is equal to 1200 cents = 100%, whereas a minor second is equal to 100 cents = 8.33%**	Continuous	$\{x \in \mathbb{R} \vert 0 \leq x \leq 1\}$	Aptitude	$\frac{1}{6}$
	Precision of recognizing five wave shapes *Score ranges from 0.0 to 1.0 and are calculated as follows*: *1.0 = The wave shape was identified correctly* *0.0 = The wave shape was not identified correctly*	5× Dichotomous	$\left\{x = \frac{k}{5} \vert 0 \leq x \leq 1 \wedge k \in \mathbb{N}_0\right\}$	Aptitude	$\frac{1}{6}$

(*Continued*)

(Continued).

Test category	Question type	Measurement level	Scoring range [0,1]	Test type	Weight
	Precision of reproducing five rhythms of increasing complexity *Score ranges from 0.0 to 1.0 and are calculated as follows*: $a = c \cdot (x_1 + 2 \cdot x_2 + 3 \cdot x_3 + 4 \cdot x_4 + 5 \cdot x_5)$ *1.0 = Rhythm tapped without altering its tempo and its character* *0.5 = Maximally one mistake without decaying the rhythm* *0.5 = Tenuous tapping with a total delay of 0.5 to 1.0 s* *0.0 = Worse than described above*  **Constant factor *c* ranges from 0.8 to 1.0 and was calculated follows**: *1.0 = Given rhythm was continued seamlessly* *0.8 = Given rhythm was not continued seamlessly* **Complexity was increased from a regular pattern (x_1, single weight) to a mixed, doubling rhythm (x_2, double weight) to a mixed, dotted rhythm (x_3, triple weight) to a syncopated rhythm with breaks (x_4, fourfold weight) and finally, to a mixed rhythm with syncopated triplets (x_5, fivefold weight). The sum of this result was multiplied by the factor *c*, which diminished the result by 20% (factor *c* = 0.8) if the participant continued did not continue the given rhythm seamlessly**	5× Discrete	$\left\{x = \frac{k}{2} \vert 0 \leq x \leq 1 \wedge k \in \mathbb{N}_0\right\}$	Aptitude	$\frac{1}{6}$
Terrain interpretation	**Ability of recognizing significant geomorphological structures** *Score ranges from 0.0 to 1.0 and are calculated as follows*: *1.0 = Answer was clear and correct* *0.5 = Answer was satisfying* *0.0 = Worse than described above*	4× Discrete	$\left\{x = \frac{k}{2} \vert 0 \leq x \leq 1 \wedge k \in \mathbb{N}_0\right\}$	Aptitude	$\frac{1}{4}$

(*Continued*)

(Continued).

Test category	Question type	Measurement level	Scoring range [0,1]	Test type	Weight
	Ability of estimating effects of fluvial structures *Score ranges from 0.0 to 1.0 and are calculated as follows:* *1.0 = Answer was clear and correct* *0.5 = Answer was satisfying* *0.0 = Worse than described above*	4× Discrete	$\left\{x = \frac{k}{2} \mid 0 \leq x \leq 1 \wedge k \in \mathbb{N}_0\right\}$	Aptitude	$\frac{1}{4}$
	Ability of recognizing topographical structures *Score ranges from 0.0 to 1.0 and are calculated as follows:* *1.0 = Answer was clear and correct* *0.5 = Answer was satisfying* *0.0 = Worse than described above*	4× Discrete	$\left\{x = \frac{k}{2} \mid 0 \leq x \leq 1 \wedge k \in \mathbb{N}_0\right\}$	Aptitude	$\frac{1}{4}$
	Ability of recognizing which area the map extension might represent *Score ranges from 0.0 to 1.0 and are calculated as follows:* *1.0 = Answer was correct* *0.5 = Answer was reasonable because of topographic similarity* *0.0 = Worse than described above*	3× Discrete	$\left\{x = \frac{k}{2} \mid 0 \leq x \leq 1 \wedge k \in \mathbb{N}_0\right\}$	Aptitude	$\frac{1}{4}$
Complexity level 1 (abstract discrete)	**Recognizing the correct number of shapes on a homogenous background** *Participants obtained no points (0.0) for this introductory question*	Dichotomous	$\{0, 1\}$	Auditory display response	0
	Identifying three geometrical shapes on a homogenous background *Score ranges from 0.0 to 1.0 and are calculated as follows:* *1.0 = The shape has been identified correctly* *0.0 = The shape has not been identified correctly*	3× Dichotomous	$\{0, 1\}$	Auditory display response	$\frac{3}{14}$
	Identifying the correct location of a geometrical shape's centroid *Score ranges from 0.0 to 1.0 and are calculated as follows:* $a = 1 - \frac{\textit{subject's distance to centroid } [px]}{\max \textit{yield distance } [px] + 50[px]}$ **Participants could get a full score within 50 pixels of tolerance to the centroid**	3× Continuous	$\{x \in \mathbb{R} \mid 0 \leq x \leq 1\}$	Auditory display response	$\frac{3}{14}$

(*Continued*)

(Continued).

Test category	Question type	Measurement level	Scoring range [0,1]	Test type	Weight
	Accuracy of identifying the orientation of a regular triangle and of a rectangle *Score ranges from 0.0 to 1.0 and are calculated as follows*: *1.0 = 0° ≤ φ < 5°* *0.8 = 5° ≤ φ < 10°* *0.6 = 10° ≤ φ < 15°* *0.4 = 15° ≤ φ < 20°* *0.2 = 20° ≤ φ < 25°* *0.0 = φ ≥ 25°* **φ describes the angular deviation between a participant's answer and the given solution. An upper error margin of 25° has been chosen because it is smaller than half the angle of a regular triangle** ($25^\circ < \frac{60^\circ}{2}$) **or of a rectangle** ($25^\circ < \frac{90^\circ}{2}$)	2× Discrete	$\left\{x = \frac{k}{5} \mid 0 \leq x \leq 1 \land k \in \mathbb{N}_0\right\}$	Auditory display response	$\frac{2}{14}$
	Accuracy of identifying a shape's size *Score ranges from 0.0 to 1.0 and are calculated as follows:*$a = 1 -$ *deviating percents of the real size* []	Continuous	$\{x \in \mathbb{R} \mid 0 \leq x \leq 1\}$	Auditory display response	$\frac{3}{14}$
	Accuracy of identifying a shape's elevation *Score ranges from 0.0 to 1.0 and are calculated as follows:*$a = 1 -$ *deviating percents of the relative height* []	Continuous	$\{x \in \mathbb{R} \mid 0 \leq x \leq 1\}$	Auditory display response	$\frac{3}{14}$

(*Continued*)

(Continued).

Test category	Question type	Measurement level	Scoring range [0,1]	Test type	Weight
Complexity level 2 (abstract continuous)	**Accuracy of identifying a target's location** *Score ranges from 0.0 to 1.0 and are calculated as follows*: $a = \frac{3}{2} \cdot \left(1 - \frac{d_{cursorTarget}[px]}{d_{tol}[px]}\right) + \frac{1}{2} \cdot \left(1 - \frac{d_{cursorTarget}[px]}{d_{max}[px]}\right) + \left(1 - \left\lvert \widehat{x} \square - \frac{x_{target}[px]}{res_x[px]} \right\rvert \right) + \left(1 - \left\lvert \widehat{y} \square - \frac{y_{target}[px]}{res_y[px]} \right\rvert \right)$ *a = Modeled accuracy for this task* $d_{cursorTarget}$ = *Measured raster distance from the cursor's position to the target point in pixel* d_{tol} = *Tolerance that must not be exceeded to get a score (approx. a third of a structure's extension)* d_{max} = *Largest distance from the recorded position to an edge of the pen tablet* $\widehat{x}$ = *Position the participant estimated on the x-axis whereas* $0 \leq \widehat{x} \leq 1$ x_{target} = *Target's effective position on the x-axis* res_x = *Auditory display's resolution on the x-axis* $\widehat{y}$ = *Position the participant estimated on the y-axis whereas* $0 \leq \widehat{y} \leq 1$ y_{target} = *Target's effective position on the y-axis* res_y = *Auditory display's resolution on the y-axis* **The formula's first two brackets refer to the question, how accurately a participant could effectively locate the *xy* target location of a geomorphological structure (mountain, dip, sink, valley). The formula's last two brackets refer to the question, how accurately a participant could identify the target's absolute position** **Please note that the cursor's position has been recorded at the moment when the participant answered a question regarding the absolute position of a geomorphological structure. Since the target point was known for each question, the distance d_{ras} was calculated. It has been associated with the two denominators in the first two brackets: d_{tol} and d_{max}. In the first bracket, d_{tol} has been chosen to model the accuracy regarding to a given tolerance distance; if this tolerance distance was exceeded, the whole first bracket resulted in zero points. In the second bracket, d_{max} has been chosen to model the distance over the whole pen tablet extent. In this way, the sense of perceiving distances over the whole pen tablet could be taken into account. The first bracket has been weighted with 1.5 whereas the second bracket has been weighted with 0.5** **The last two brackets refer to the answer the participant gave in order to estimate the location of a geomorphological structure. Since the target point was known for each question (x_{target} and y_{target}), its relative position in accordance to a fixed coordinate system could be determined by dividing this value by the *x* or *y* resolution (res_x and res_y), leading to a value between 0% and 100% for each axis. This relative value has been subtracted from the participant's estimation of the imagined location $\widehat{x}$ and $\widehat{y}$ in order to model the deviation between the mental model to the effective target point**	Continuous	$\{x \in \mathbb{R} \mid 0 \leq x \leq 1\}$	Auditory display response	$\frac{4}{13}$

(*Continued*)

(Continued).

Test category	Question type	Measurement level	Scoring range [0,1]	Test type	Weight
	Accuracy of identifying the effective cursor position *Score ranges from 0.0 to 1.0 and are calculated as follows*: $a = \frac{11}{32} \cdot \left(1 - \left\|\hat{x}\,\square - \frac{x_{cursor}}{res_x}\right\|\right) + \frac{11}{32} \cdot \left(1 - \left\|\hat{y}\,\square - \frac{y_{cursor}}{res_y}\right\|\right) + \frac{10}{32} \cdot \left(1 - \frac{d_{mindCursor}\,[px]}{d_{tol}\,[px]}\right)$ *a = Modeled accuracy for this task* $d_{mindCursor}$ *= Modeled distance from the imagined position to the cursor's position in pixel* d_{tol} *= Tolerance that must not be exceeded to get a score (approx. a third of a structure's extension)* $\hat{x}$ *= Position the participant estimated on the x-axis whereas* $0 \leq \hat{x} \leq 1$ x_{cursor} *= Cursor's position registered on the x-axis* res_x *= Auditory display's resolution on the x-axis* $\hat{y}$ *= Position the participant estimated on the y-axis whereas* $0 \leq \hat{y} \leq 1$ y_{cursor} *= Cursor's position registered on the y-axis* res_y *= Auditory display's resolution on the y-axis* **This task compared a participant's perceived location with the effective location at the cursor's position. The first two brackets associate the imagined location ($\hat{x}$ and $\hat{y}$) with the cursor's position (x_{cursor} and y_{cursor}). In the third bracket, d_{max} has been chosen to model the distance over the whole pen tablet extent. In this way, the sense of perceiving distances over the whole pen tablet could be taken into account. The first two brackets have been weighted with $\frac{11}{32}$ each whereas the third bracket has been weighted with $\frac{10}{32}$**	Continuous	$\{x \in \mathbb{R} \| 0 \leq x \leq 1\}$	Auditory display response	$\frac{4}{13}$
	Accuracy of estimating a target's absolute elevation *Score ranges from 0.0 to 1.0 and are calculated as follows*: $a = \hat{z}\,\square - \frac{z_{target}}{res_z}$ *a = Modeled accuracy for this task* $\hat{z}$ *= Height the participant estimated on the z-axis whereas* $0 \leq \hat{z} \leq 1$ z_{target} *= Target's effective height on the z-axis* res_z *= Auditory display's resolution on the z-axis* **A geomorphological structure (mountain, hill, valley, sink). Since the target point was known for each question (z_{target}), its relative height in accordance to a fixed coordinate system could be determined by standardizing this value by the z resolution (res_z), leading to a value between 0% and 100%. This standardized value has been subtracted from the participant's estimation of the imagined height $\hat{z}$ to model the deviation between the mental model to the effective target point**	Continuous	$\{x \in \mathbb{R} \| 0 \leq x \leq 1\}$	Auditory display response	$\frac{2}{13}$

(*Continued*)

(Continued).

Test category	Question type	Measurement level	Scoring range [0,1]	Test type	Weight
	Accuracy of estimating the effective height at a cursor's position *Score ranges from 0.0 to 1.0 and are calculated as follows*: $a = \widehat{z}\,[] - \frac{z_{cursor}}{res_z}$ *a = Modeled accuracy for this task* $\widehat{z}$ = *Height the participant estimated on the z-axis whereas* $0 \le \widehat{z} \le 1$ z_{cursor} = *Height registered at the cursor's position* res_z = *Auditory display's resolution on the z-axis* **This task compared a participant's perceived height with the effective height at the cursor's position. The formula associates the imagined height ($\widehat{z}$) with the height at the cursor's position (z_{cursor}). The division by the height resolution (res_z) returns a relative error, according to the relative estimated height ($\widehat{z}$)**	Continuous	$\{x \in \mathbb{R} \mid 0 \le x \le 1\}$	Auditory display response	$\frac{2}{13}$
	Identifying if a geomorphological structure (mountain, dip, sink, valley) is steep or flat *Score ranges from 0.0 to 1.0 and are calculated as follows*: 1.0 = The answer was correct 0.0 = The was not correct	2× Dichotomous	$\{0, 1\}$	Auditory display response	$\frac{1}{13}$
Complexity level 3 (real continuous)	**Accuracy of identifying the effective cursor position** *Score ranges from 0.0 to 1.0 and are calculated as follows*: $a = \frac{1}{4} \cdot \left(1 - \left\lvert \widehat{x}\,[\,] - \frac{x_{cursor}}{res_x} \right\rvert \right) + \frac{1}{4} \cdot \left(1 - \left\lvert \widehat{y}\,[\,] - \frac{y_{cursor}}{res_y} \right\rvert \right) + \frac{1}{2} \cdot \left(1 - \frac{d_{mindCursor}\,[px]}{d_{max}\,[px]}\right)$ *a = Modeled accuracy for this task* $d_{mindCursor}$ = *Modeled distance from the imagined position to the cursor's position in pixel* d_{max} = *Maximum model extent, which corresponds to the auditory display's width* $\widehat{x}$ = *Position the participant estimated on the x-axis whereas* $0 \le \widehat{x} \le 1$ x_{cursor} = *Cursor's position registered on the x-axis* res_x = *Auditory display's resolution on the x-axis* $\widehat{y}$ = *Position the participant estimated on the y-axis whereas* $0 \le \widehat{y} \le 1$ y_{cursor} = *Cursor's position registered on the y-axis* res_y = *Auditory display's resolution on the y-axis*	Continuous	$\{x \in \mathbb{R} \mid 0 \le x \le 1\}$	Auditory display response	$\frac{6}{15}$

(*Continued*)

(Continued).

Test category	Question type	Measurement level	Scoring range [0,1]	Test type	Weight
	This task compared a participant's perceived location with the effective location at the cursor's position. The first two brackets associate the imagined location ($\widehat{x}$ and $\widehat{y}$) with the cursor's position (x_{cursor} and y_{cursor}). In the third bracket, d_{max} models the distance over the whole auditory display's extent. In this way, the maximal estimation error could be reduced to the width, which is the largest distance on the auditory display's extentAccuracy of identifying the effective elevation at a cursor's position *Score ranges from 0.0 to 1.0 and are calculated as follows*: $a = 1 - \left\| \widehat{z}\,[] - \frac{z_{cursor}\,[px]}{max_z[px] - min_z[px]} \right\|$ *a = Modeled accuracy for this task* $\widehat{z}$ *= Height the participant estimated on the z-axis whereas* $0 \leq \widehat{z} \leq 1$ z_{cursor} *= Height registered at the cursor's position* max_z *= Auditory display's maximum elevation (on the z-axis)* min_z *= Auditory display's minimum elevation (on the z-axis)* **This task compared a participant's perceived height with the effective height at the cursor's position. The formula associates the imagined height ($\widehat{z}$) with the height at the cursor's position (z_{cursor}). The division by the height range ($max_z - min_z$) returns a relative error, according to the relative estimated height ($\widehat{z}$)**	Continuous	$\{x \in \mathbb{R} \| 0 \leq x \leq 1\}$	Auditory display response	$\frac{4}{15}$
	Ability of finding specific geomorphological structures (mountain, dip, lake, valley) $a = \frac{points_{earned}}{points_{max}}$ *Score ranges from 0.0 to 1.0 and are calculated as follows*: *1.0 = The target area was found within 5 pixels of tolerance* *0.0 = The target area was not found within 5 pixels of tolerance*	4× Dichotomous	$\{x \in \mathbb{R} \| 0 \leq x \leq 1\}$	Auditory display response	$\frac{4}{15}$
	Accuracy of identifying the azimuth of a mountain chain *Score ranges from 0.0 to 1.0 and are calculated as follows*: *1.0 =* $0° \leq \varphi < 5°$ *0.8 =* $5° \leq \varphi < 10°$ *0.6 =* $10° \leq \varphi < 15°$ *0.4 =* $15° \leq \varphi < 20°$ *0.2 =* $20° \leq \varphi < 25°$ *0.0 =* $\varphi \geq 25°$ **φ describes the angular deviation between a participant's answer and the given solution. An upper error margin of 25° has been chosen because it is smaller than half the angle of a regular triangle ($25° < \frac{60°}{2}$) or of a rectangle ($25° < \frac{90°}{2}$)**	Discrete	$\left\{x \in \mathbb{N}_0 \| 0 \leq \frac{x}{5} \leq 1\right\}$	Auditory display response	$\frac{1}{15}$