Accessible Data Representation with Natural Sound

Md Naimul Hoque
nhoque@umd.edu
University of Maryland, College Park
College Park, MD, USA

Md Ehtesham-Ul-Haque
mfe5232@psu.edu
Pennsylvania State University
University Park, PA, USA

Niklas Elmqvist
elm@umd.edu
University of Maryland, College Park
College Park, MD, USA

Syed Masum Billah
sbillah@psu.edu
Pennsylvania State University
University Park, PA, USA

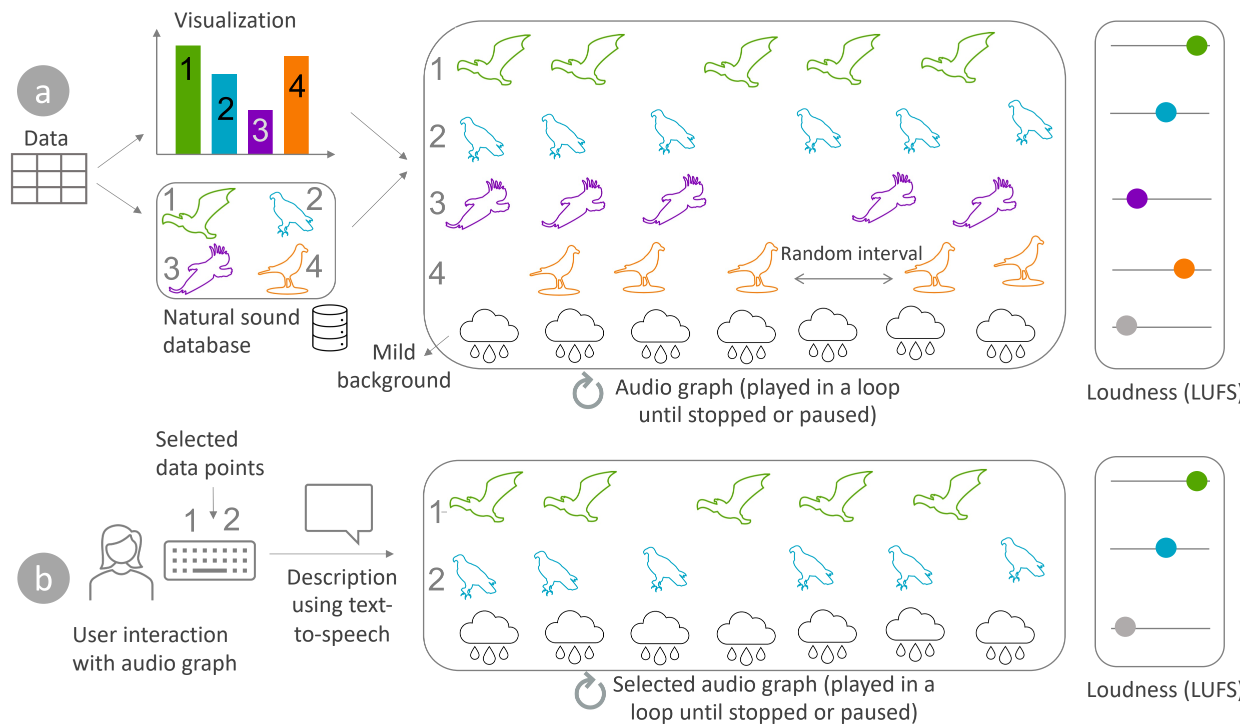


Figure 1: Sonification of a bar chart with Susurrus. (a) Our technique maps each bar to a natural sound drawn from an ambient theme (e.g., forest and birds). Using Loudness Levels Relative to Full Scale (LUFS), we convey the data values by setting the loudness of the sounds in decibels proportionately (i.e., height) to the bars. In this instance, we have mapped four bars in a bar chart to four bird sounds (e.g., robin, woodpecker, raven, and dove). We play the sounds together (i.e., in parallel) in a loop with random intervals and use a calm forest ambiance as background, thus making sonification of the bar chart similar to listening to bird sounds in the forest. (b) A user can interact with the audio graph using specific keys. For example, the user can select the first two bars using 1 and 2 number keys and listen to the corresponding sounds. With the selection, the user can also listen to the description of the selected data values using Text-to-Speech (TTS). The audio for this example is provided in the supplement.

ABSTRACT

Sonification translates data into non-speech audio. Such auditory representations can make data visualization accessible to people who are blind or have low vision (BLV). This paper presents a sonification method for translating common data visualization into a blend of natural sounds. We hypothesize that people's familiarity with sounds drawn from nature, such as birds singing in a forest,

CHI '23, April 23–28, 2023, Hamburg, Germany

ACM ISBN 978-1-4503-9421-5/23/04.
https://doi.org/10.1145/3544548.3581087

and their ability to listen to these sounds in parallel, will enable BLV users to perceive multiple data points being sonified at the same time. Informed by an extensive literature review and a preliminary study with 5 BLV participants, we designed an accessible data representation tool, Susurrus, that combines our sonification method with other accessibility features, such as keyboard interaction and text-to-speech feedback. Finally, we conducted a user study with 12 BLV participants and report the potential and application of natural sounds for sonification compared to existing sonification tools.

CCS CONCEPTS

• **Human-centered computing → Accessibility technologies**.

KEYWORDS

Sonification, accessibility, accessible data, natural sounds.

ACM Reference Format:

Md Naimul Hoque, Md Ehtesham-Ul-Haque, Niklas Elmqvist, and Syed Masum Billah. 2023. Accessible Data Representation with Natural Sound. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems (CHI '23), April 23–28, 2023, Hamburg, Germany*. ACM, New York, NY, USA, 19 pages. https://doi.org/10.1145/3544548.3581087

1 INTRODUCTION

Data visualization is becoming commonplace in many domains—from data science and machine learning to online newspapers, business intelligence, medical science, and elementary school education. This proliferation makes it imperative that visualizations are accessible to people from diverse backgrounds. However, due to their visual nature, people who are blind and have low vision (BLV)[1] often cannot access them [17, 53, 57]. In response, the accessibility and visualization field have recently embarked on a significant enterprise of increasing the accessibility of data visualizations for BLV users [39, 55, 58, 70, 71]. *Sonification* [23], the representation of data using non-verbal audio, has seen success in this regard. Sonification can be used as a standalone solution [66] or combined with other accessibility features such as text-to-speech (TTS) and alternative text (alt-text) generation [55, 70, 71]. However, several challenges persist, including perceiving multiple sounds, the role of attention in the process, and dynamic sound perception [34]. Some of these unsolved challenges have contributed to sonification enjoying only limited deployment in the field and none in commercial data visualization products, such as Tableau or Spotfire, despite the area being more than 30 years old.

In this paper, we address some of these perceptual aspects by proposing a novel sonification technique that uses ambient sounds drawn from nature—such as birds chirping, rain dripping, or waves crashing—for representing data. We believe such natural sounds are well suited for sonification since humans are familiar with them from birth. We perceive them on a daily basis while walking, running, or just existing in nature. In fact, blind individuals routinely use natural sounds for navigation [9], suggesting that natural sounds are central to sensing the surroundings. Moreover, they have a positive impact on health and wellbeing [7], including calming the mind [51], reducing stress [1, 48], increasing attention [4, 48], and enhancing mood [30, 31]. They also evoke a similar response in the human brain's early visual cortex [65], regardless of a person's visual ability.

Our work draws on the rich literature in audiology, psychoacoustics, digital signal processing, and accessible data representation. Informed by the literature and a pilot study with 5 BLV participants (§6), we designed a natural sound-based sonification prototype, **Susurrus**. The key features of Susurrus include (1) using *Loudness Levels Relative to Full Scale* (LUFS) [49] for conveying data values; (2) sonification of multiple data values together (i.e., *in parallel*), instead of the current norm of playing the data values *serially*, one after one (e.g., [2, 8, 55]); and (3) supporting common auditory information-seeking actions (AISA) [70] with a keyboard, the primary mode of interaction for BLV screen reader users.

We then used Susurrus to conduct a user study with 12 BLV participants where we compared natural sounds to a traditional sonification approach using artificial sounds. Our findings suggest that natural sound-based sonification (1) can support commonly used data operations in bar, line, and scatter plots; (2) is better suited to represent charts representing multiple categories (e.g., bar charts); (3) is most useful to users who do not have musical training and have difficulties in differentiating pitch; and (4) has hedonic value for enjoyment, relaxation, emotional connection, and personalizing. Based on these findings, we conclude that natural sound-based sonification is a promising approach for making data more accessible. In sum, our contributions are as follows:

(1) We show that natural sounds that are generally considered as *ambient noise* have distinct expressive patterns that can be harnessed to represent data;
(2) We propose Susurrus, a sonification tool that implements our technique for representing common charts (e.g., bar, scatter, line) using natural sounds; and
(3) We report on a user study with 12 blind participants to compare Susurrus with existing sonification tools.

The remainder of the paper is organized as follows. In Section 2, we describe sound and sonification concepts relevant to our method. Section 3 discusses accessible data representation literature. Drawing on the literature review, in Section 4, we present the design space for accessible data representation with natural sounds. Section 5 presents the initial design of the Susurrus tool. Sections 6 and 7 present our pilot study and the revised prototype. Section 8 presents the evaluation of Susurrus with 12 BLV participants. Finally, we discuss and conclude the broader impact and design implications of Susurrus in Sections 9 and 10.

2 BACKGROUND: SOUND AND SONIFICATION

Here we briefly describe relevant theories, concepts, and terminologies to position our work in the literature and contextualize our contribution.

2.1 Periodic and Aperiodic Sounds

A sound is generated by a vibrating object (e.g., vocal cords, string, or wind) that displaces air molecules, resulting in local regions of compression and rarefaction, which travels through the air as a

[1]While we generally use people-first language when referring to people with disabilities—e.g., "person who is blind"—we note that the Blind community often does not make this distinction; i.e., the term "blind person" is acceptable, and often preferable, to many people who are blind or have low vision [21].

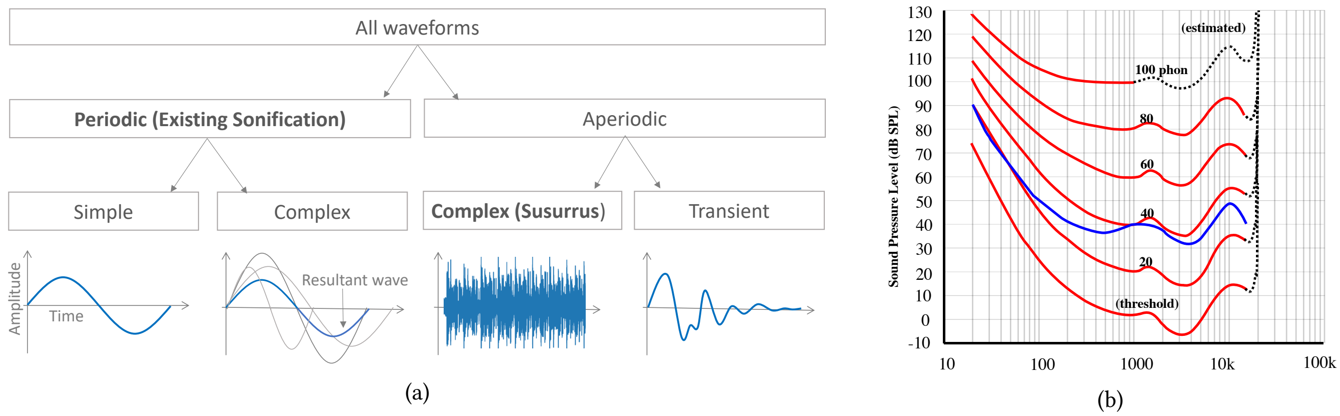


Figure 2: Periodic and aperiodic sounds. (a) Different types of periodic and aperiodic sounds. Existing sonification tools for accessible data representation predominantly use both simple and complex periodic sounds. In contrast, Susurrus uses natural sounds, which falls under complex aperiodic sounds. (b) Equal loudness contours or Fletcher-Munson curves (red) from ISO 226:2003 revision. The original ISO standard is shown in blue for 40-phons [36]. Here, the X-axis represents sound frequency (logarithmically spaced); the Y-axis represents the sound pressure level (in dB SPL); and each curve or contour line indicates a fixed loudness (measured in phons).

wave. This change in pressure with time can be visualized as a waveform, where the x-axis represents the time, and the y-axis represents the air pressure (Figure 2a). A waveform is *periodic* if the alternating air pressure repeats in a regular fashion; *aperiodic* otherwise.

The *simplest* type of periodic waveform is a single sine wave (*sinusoid*), as shown in Figure 2a (bottom-left). Most sinusoids are generated artificially (e.g., using computers) and can be used to prototype the acoustic realization of a musical note [42]. Often, a sinusoid is referred to as a *pure tone* and is the basis of data sonification in prior work.

A *complex periodic* waveform (e.g., a musical note) is a superposition of multiple sinusoids, each with its own frequency, amplitude, and phase. Figure 2a (second from the left) shows a complex periodic sound waveform made up of 3 sinusoids: all have equal amplitudes and zero phases but varying frequencies (e.g., 2 Hz, 4 Hz, and 8 Hz). The lowest frequency among the constituent sinusoids is called the *fundamental frequency* (2 Hz in the example). Like simple periodic sounds, complex periodic sounds (e.g., musical notes) are also extensively used in data sonification. Our technique is a departure from prior work, as we investigate the data sonification with complex aperiodic sounds.

Complex aperiodic sound waves do not exhibit patterns (e.g., repetition, well-defined frequencies) in air pressure, as shown in Figure 2a (third from the left). Most environmental sounds (e.g., traffic, coffee shop, forest, and rain) fall under this category. In sonification research, such sound is often referred to as audio icons [23], a distinct auditory category. Our technique synthesizes aperiodic, ambient sound for a certain duration and plays it in a loop to add periodicity. Users can adjust the duration dynamically and pause (or play) the loop at any time.

Some aperiodic sound waves are *transient* (e.g., a pulse, a popping sound, and a clicking sound), as shown in Figure 2a (bottom-right) [20]. To our knowledge, transient sounds are not used in data representation but for notifications. We used a short-beep transient sound to indicate the end of a loop.

2.2 Frequency vs. Pitch

Frequency is an objective property of a sound (i.e., the number of cycles per second). The audible frequency range for humans is between 20 Hz to 20 kHz. *Pitch*, on the other hand, is the perception of frequency in humans and is measured in *Mels*; humans perceive two frequencies similarly if they differ by a power of two [41, 72]. For example, musical notes with fundamental frequencies (i.e., the lowest frequency in a complex periodic sound) of 440 Hz and 880 Hz belong to the same note *A*. These notes are said to be one *octave* apart, where the former (440 Hz) has a low pitch (e.g., A4), and the latter (880 Hz) has a high pitch (e.g., A5). Most prior work on sonification (e.g., iSonic [70]) mapped data to pitch.

Prior work on psychoacoustics reports that an octave can be divided into 1,200 units or *cents*, and most individuals can recognize pitch differences of 25 cents [42]. Trained musicians, in contrast, can recognize pitch differences of as small as 10 cents: a 60% higher recognition resolution compared to untrained individuals. We observed this phenomenon in our study: participants with a musical background performed better on a pitch-based sonification system, compared to those without a musical background.

2.3 Factors Affecting the Perceived Loudness in Analog Sounds

Loudness is a subjective measure of human perception of sound that correlates with sound pressure and frequency. Sound *pressure* is an objective measure related to the physical property of an analog sound wave. It indicates how much the local air pressure deviates from the atmospheric pressure caused by a sound wave (unit: pascal). Humans can perceive the minimum sound pressure of $2x10^{-5}$ pascal, which serves as the reference sound pressure. Psychoacoustics studies [41, 72] report that humans perceive sound pressure level (SPL) on a decibel scale (dB SPL) as the log-ratio of a given pressure to the reference pressure.

Loudness also correlates with sound frequencies, following Fletcher-Munson curves (Figure 2b). In this figure, each curve indicates an equal perception of the loudness of a sound in relation to different frequencies (X-axis) and sound pressure levels (Y-axis). Since we map data to perceived loudness, these curves are particularly interesting. However, we cannot directly use these curves because the sound pressure level (Y-axis) is a physical property of analog sounds that is lost when we convert an analog sound into digital audio. The next section describes this process.

2.4 Factors Affecting the Perceived Loudness in Digital Sounds

Decibels relative to full scale (dbFS). An analog sound is first converted into digital sounds by drawing samples at least twice the rate of the frequency [64]. In digital sound, decibels relative to full scale (dB FS) is a unit of measurement of a signal level relative to full scale (FS). It is equivalent to the sound pressure of analog sound. The full scale depends on the number of bits used to represent individual audio samples. For example, the maximum possible level of a sample in a 16-bit digital audio system is 32,767 (i.e., $2^{16-1} - 1$). Thus, the full scale or FS of this system is 32,767. A sound of 0 dBFS indicates that it peaks at the FS. Similarly, another sound of -18 dBFS indicates that it peaks 18 dB below the FS.

Loudness Levels Relative to Full Scale (LUFS) is the unit for subjective loudness levels relative to full scale (FS), calculated by the ITU-R BS.1770 algorithm [49]. This algorithm approximates Fletcher-Munson curves (Figure 2b) for digital audio. LUFS is now being used by streaming companies to normalize the loudness of audio on their platforms. We use the loudness range between -11 LUFS (loud) to -23 LUFS (quiet), following the range used by Spotify [60].

Distance and Duration. Loudness is also a function of the distance between the sound source and the listener, as well as the duration of a sound [42, 68]. These two factors are easily controllable by wearing headphones (which keeps the distance between the sound source and the ears to *3* centimeters) and keeping the duration approximately equal.

2.5 Timbre of a Sound

The timbre of a sound is another perceptual property [52] similar to pitch and loudness. It allows listeners to distinguish the musical note of different instruments in the same category (e.g., a violin, an oboe, or a trumpet) even if the same note is played at the same pitch (e.g., A4) and with the same loudness [42]. Prior work in sonification has used timbre to represent two independent categories (e.g., piano notes for category 1 and acoustics guitar for category 2) [67]. Although timbre is associated with musical instruments, we observe this timbre-like property in natural sounds within the same category (e.g., birdsong). For example, humans can easily distinguish the sound of a robin from that of a chickadee. This observation also inspires our design.

3 RELATED WORK

Our work is related to prior research on accessible and multimodal data representation, especially with sonification. We briefly describe research in these areas below.

3.1 Accessible Data Representation

Recent research have outlined accessibility issues with data visualization, such as the lack of informative alternate texts and support for perceiving trend and overview from data [9, 16, 17, 27, 28, 38, 53, 57]. Several efforts are starting to address these concerns using alternate sensory mediums such as touch, smell, and audio.

Examples of tactile interfaces include physical bar charts [18, 19, 62, 63], tactile feedback based interpretation of 2D information [69], 3D printed tangible maps [24, 29], and wheeled robots for physical data visualizations [35]. In contrast, the field of smell-based (olfactory) representations is much smaller. Patnaik et al. [45] explored the olfactory system by conveying data through scents. A follow-up paper based on this work provided a ranking of the sensory channels through a perception study [3].

While effective, both tactile and smell-based systems require external hardware and are only applicable to specific data types (e.g., bar charts), thus making their applicability limited to a lab or a data domain. In contrast, the majority of data visualizations today appear on the web. Thus, researchers have called for accessibility solutions that are flexible, can work in web browsers, are compatible with screen-readers, and do not require any external hardware [53, 57].

A promising research direction is improving alternate texts that BLV users can access using screen-readers for gathering insights about the data. However, the utility of alt-texts depend on their quality, which is directly related to the effort that chart creators are willing to invest [28]. Recent research suggests alt-texts on the web are largely non-existent, and even if they exist, they vary in terms of structure, quality, and content [28, 39]. Alt-texts also do not allow BLV users to explore and interact with a visualization [39, 55, 57]. Thus, there is significant research on improving the quality of alternate texts [71] as well as novel tools that allow users to interact with a visualization non-visually [2, 8, 55]. One of the promising non-visual methods is Sonification, which we discuss next.

3.2 Sonification for Accessible Data Representation

Sonification creates data representations using non-speech audio [34]. While sonification has many uses for art, design, and ambient data, it is particularly powerful for representing data to BLV users because of its non-visual property. When applied to represent data, sonification is often referred to as audio graphs [55]. The Sonification Sandbox [66] and iSonic [70] are early examples of audio graphs. Highcharts [8] extends Sonification Sandbox and integrates sonification in a Javascript-based commercial data visualization library. Apple Audio Graphs [2] uses a similar sonification method, with the function to interact with the sonification using touch in Apple products.

In recent times, several emerging accessible data representation solutions have featured sonification. For example, Siu et al. [58] proposed Audio Data Narratives, a method that automatically partitions a timeline into important segments and then interleaves

System	Design Dimensions				
	Audio Type	Sonic Mapping	Play Order	Auditory Information Seeking Actions (AISA)	Application Domain
Sonification Sandbox [66]	Simple + complex periodic	Pitch, Volume, Timbre, Pan	Serial	Gist	Desktop
iSonic [70]	Simple + complex periodic, TTS	Pitch	Serial	Gist, navigation, situate, details on demand, select, brush	Desktop
Apple Audio Graph [2]	Simple + complex periodic, TTS	Pitch	Serial	Gist, navigation, situate, details on demand, select	Mobile app
Highcharts [8]	Simple + complex periodic, TTS	Pitch	Serial	Gist, navigation, situate, details on demand, select	Proprietary, online
VoxLens [55]	Simple + complex periodic, TTS	Pitch	Serial	Gist, navigation, situate, details on demand, select	Online plugin, open-sourced
Audio data narratives [58]	Simple + complex periodic, TTS	Pitch	Serial	Gist	—
Rich Screen Reader [71]	TTS	—	—	Navigation, situate, details on demand, select	Online

Table 1: Design space for accessible data visualization across five design dimensions. Here TTS refers to Text-to-Speech. We do not include Sonifier [54] as a separate system as it was originally proposed as a module in VoxLens.

description with sonification rather than presenting the description separately. VoxLens [55] uses a mix of textual description of a visualization (equivalent to alternate texts), sonification produced from a dedicated tool named Sonifier, and voice commands for supporting accessible visualization on the web. Using Sonifier as a probe, Sharif et al. [54] further highlighted the need to improve the usefulness as well as usability of sonified responses. They conducted a study to measure the usability of different periodic signals (e.g., a synthesizer with a square waveform) in sonification across four Likert scales: *Pleasantness, Clarity, Confidence,* and *Overall Score.*

While these lines of work are inspiring, at their core, they follow a similar sonification method: using the *pitch* of *artificial and musical sounds (i.e., simple or complex periodic waveforms, as described in §2.1)* to *serially* sonify data. Prior work has highlighted the need to improve such sonification techniques [9, 25, 54, 67]. This paper pushes the status quo of sonification, drawing on the rich literature on audiology, data sonification, and psychoacoustics studies. Indeed, for the first time, we show that natural sounds—typically considered ambient noises and comprising complex aperiodic waveforms—can represent common data visualizations. Further, our method introduces several new concepts for sonification research, such as using LUFS as sonic mapping and parallel sonification.

3.3 Sonification with Natural Sounds

To our knowledge, we are not aware of any prior work that uses natural sounds for accessible data representation. However, we note several uses of natural sounds in other applications. For example, traffic data from computer networks are often beyond human capabilities to process due to their large scale and speed. Debashi et al. [13] used different bird sounds to raise situational awareness of users monitoring such large-scale network traffic data. Similarly, Lockton et al. [37] proposed Bird-Wattching, a hardware device that uses bird sounds to alert users about electricity usage.

Another direction relevant to our research is *soundscapes* [26, 33, 46]: the use of sounds to create an acoustic environment, where natural sounds are used as a matter of course. The concept is often applied in game design and virtual and augmented reality. We differ from soundscapes in the sense that we do not use natural sounds to *enhance* the realism of an environment or *raise* spatial awareness, but instead *create* an environment to *represent* abstract data. Put differently, a raven cawing would indicate the presence of the bird in the soundscape, whereas we utilize the raven's call as a vehicle to convey other data. We, however, acknowledge that our work can be interpreted as falling under the umbrella of "soundscape" research because of our use of natural sounds as our representational medium.

In summary, while natural sounds have previously been used in sonification and soundscape research, the process of mapping natural sounds to data visualization such as bar, line, or scatter plot, is still unknown. In this paper, we aim to address this gap.

4 DESIGN SPACE FOR ACCESSIBLE DATA REPRESENTATIONS WITH NATURAL SOUNDS

Here we discuss current accessible data representation solutions across five design dimensions. These design dimensions are the key parameters for designing appropriate solutions, and we identified them using our literature review (§2 and §3). Together, these dimensions define the design space for developing our natural sound-based sonification technique.

Table 1 summarizes this design space. We include only systems that feature an audio interface for accessing visualizations and exclude tactile and smell-based systems. We further discuss the design dimensions below in detail and outline design challenges for natural sound-based sonification.

4.1 D1 – Audio Type

Most existing solutions use simple (i.e., a single sine wave) and complex periodic sounds (i.e., musical notes) for sonification and synthesized speech (e.g., text-to-speech or TTS) for providing details of selected data points. However, we intend to use natural sounds (complex aperiodic sounds) for sonification and synthesized speech for details on demand. The simple and complex periodic sounds common to current tools are created programmatically and are easy to manipulate. However, there is no straightforward way to create natural sounds programmatically. Thus, we need to collect natural sounds that we can manipulate for sonification (**design challenge 1 or C1**).

4.2 D2 – Sonic Mapping

Sonic mapping captures what sonic features are used to convey data. Based on Table 1, there are four options for auditory mapping: (1) pitch, (2) loudness, (3) timbre, and (4) pan. Existing solutions typically use the pitch of a simple or complex periodic sound to represent data values. This is consistent with long practice in sonification confirmed as recently as by Wang et al. [67] in 2022 and Chundury et al. [9] in 2021.

Using pitch to convey data makes sense for periodic sounds since they either contain a single sine waveform (i.e., simple periodic sounds) or a fundamental frequency (i.e., complex periodic sounds), which can be changed easily. However, it is not applicable for natural sounds (i.e., complex aperiodic sounds) as they do not exhibit any well-defined frequency. Moreover, changing the frequency distribution of natural sounds can change their signature and make them unrecognizable. This poses a **unique challenge** (**C2**) for our work as we need to investigate which sonic property is appropriate for mapping data to natural sounds.

4.3 D3 – Play Order

The play order captures the sequencing of the sonification of data points, serially or in parallel. All current systems play data points serially, one after one. However, this is contrary to how data visualizations are perceived visually, where sighted users consume data points together rather than in isolation [43]. One key motivation behind using natural sounds is that they are typically perceived by humans **in parallel**, all blended together. Remarkably, the human auditory system can simultaneously consume parallel sound streams while selectively focusing on streams of interest [5, 11]. We thus hypothesize that natural sounds can be used for sonifying multiple data points in parallel. However, there exists limited knowledge of how parallel mechanisms can be facilitated in sonification. We anticipate that parallel playback will yield its own challenges (**C3**).

4.4 D4 – Supported Auditory Information Seeking Actions

The fourth dimension is the **Auditory Information Seeking Actions (AISA)** proposed by Zhao et al. [70]. There are eight actions in this set: *gist, navigation, situate, select, details on demand, brush, filter, and search.* Zhao et al. showed that these actions are analogous to how sighted users use data visualization and should be supported in audio graphs. Gist provides an overview of data with small audio.

ID	Description
C1	Collecting a natural sound database
C2	Mapping data to an auditory channel of natural sounds
C3	Supporting parallel sonification with natural sounds
C4	Support AISA: *gist, navigation, situate, select*, and *details on demand*
C5	Supporting visualization in a web browser and using a screen reader

Table 2: Design challenges for accessible data representation with natural sounds. The challenges are identified from the five design dimensions discussed in Section 4.

The other actions depend on user interaction. For example, a user can select a data point and listen to a verbal description of the data point from a text-to-speech converter.

With the exception of iSonic and Susurrus (proposed in this paper), most systems in Table 1 were not explicitly designed to support AISA. We reviewed interaction support in these systems and identified which AISA are supported. A few patterns emerged from Table 1: (1) all systems provide a gist or overview of the data by combining sonification and Text-to-Speech (TTS) description, except for the Rich Screen Reader [71], which focuses more on allowing users to drill down a visualization using a screen reader; (2) navigation, situate, select, and details on demand are commonly supported using keyboard interaction; and (3) brush, filter, and search are not commonly supported. Based on this prior art, our approach should support commonly used actions: *gist, navigation, situate, select,* and *details on demand* (**C4**).

4.5 D5 – Application Domain

The most recent work in this space, such as VoxLens [55], focus on web-based data visualization in a web browser. This is not surprising since most data visualization today appear on the web. Thus, our future prototype should be compatible with most browsers, data visualizations, and screen readers (**C5**).

5 INITIAL PROTOTYPE: SUSURRUS

Susurrus is our sonification tool based on natural sounds. To address the identified design challenges, **C1-C5** (summarized in Table 2), we adopted an iterative approach—we first designed a prototype, conducted a pilot study (§6), and refined the prototype (§7) based on the study findings. This section describes our initial prototype.

5.1 Natural Sounds Database

We collected a wide variety of royalty-freely natural sounds from the web. Our database includes sounds of rain pattering, distant thunder, dry leaves rustling, the wind howling, waves crashing, soft wind whispering, wood burning, chimes chiming, as well as many animal sounds from crickets, frogs, owls, woodpeckers, robin, nightingale, cuckoo, grasshopper, and seagull (addressing **C1**). Our goal was to include a wide array of sounds to test how they complement each other when played together. We selected sounds with a single tap of the source (e.g., a bird chirping only once) without any other background sounds. This is so we can repeat the sounds

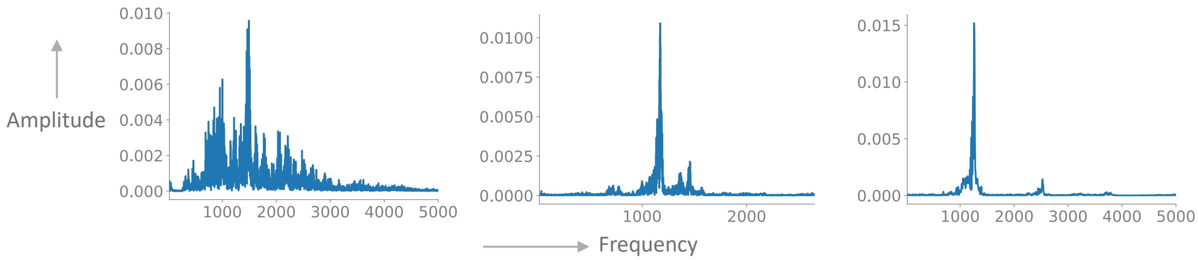


Figure 3: Frequency distribution for three soundtracks (woodpecker, seagull, and raven) in our database. Notice that unlike single frequency wave form, our soundtracks are comprised many frequencies.

Data (d_i)	Normalized Data ($\hat{d}_i$)	Sound (s_i)	Target Loudness ($\hat{l}_i$)
8	0	Seagull	-23.00
25	0.46	Woodpecker	-17.49
45	1	Raven	-11.00
23.5	0.42	Robin	-17.97

Table 3: Sonic mapping using loudness. We show how this approach is applied to four example natural sounds.

with variable intervals. Sounds in our database were between 1 and 2 seconds in duration.

5.2 Sonic Mapping

To compare different audio channels, we selected perceived loudness, sound or tapping interval, and sound duration for mapping data values in Susurrus (**C2**). According to the discussion in Section 4.2, pitch is not well-defined for natural sounds. Indeed, we plotted the frequency distribution for a few sounds (Figure 3) and found that each has a unique distribution. We noticed that there is typically a frequency with high amplitude. We experimented by changing the frequency with high amplitude. However, this tended to distort the quality of the natural sound. Similarly, we did not change timbre (§2.5) as we did not want to change the distinctiveness of a sound.

Thus, we selected perceived loudness, sound interval, and duration for sonic mapping. We use the following notations to demonstrate our mapping functions. Suppose $D(d_1, \ldots, d_n)$ is our dataset containing n data points; $S(s_1, \ldots, s_n)$ is a list of sounds available in our sound database, where a data point d_i is assigned a natural sound s_i.

5.2.1 Perceived Loudness Mapping. This mapping conveys the relative differences between data points through their differences in loudness. For a dataset $D(d_1, \ldots, d_n)$ containing n data points, this mapping produces a list of loudness levels $L(l_1^s, \ldots, l_n^s)$ of n natural sounds, where the perceived loudness of a natural sound s_i, assigned to sonify an individual data point d_i, is adjusted to the target loudness level l_i^s. More specifically, l_i^s is a mapping function, defined as $l_i^s = Susurrus^L(d_i, s_i)$. The decomposition of $Susurrus^L(d_i, s_i)$ function is given below:

$$\hat{d}_i = \frac{d_i - D_{min}}{D_{max} - D_{min}}$$
$$\hat{l}_i = L_{min} + \hat{d}_i * (L_{max} - L_{min})$$
$$l_i^s = LUFS_{Implementation}(s_i, \hat{l}_i) \tag{1}$$

Here, D_{max} and D_{min} are the maximum and minimum values in the dataset D; L_{max} and L_{min} are the maximum and minimum loudness possible in our system; $\hat{d}_i$ is the min-max normalization of d_i; $\hat{l}_i$ is target loudness for normalized data point $\hat{d}_i$; and $LUFS_{Implementation}(.)$ is an audio operation that mimics the equal loudness curves (Figure 2b) for digital sound s_i to achieve the target loudness of $\hat{l}_i$. We set $L_{min} = -23$ LUFS (quietest) for the minimum value in the dataset (D_{min}) and $L_{max} = -11$ LUFS (loudest) for the maximum value in the dataset (D_{max}).

Recall that the perceived loudness is a function of the full scale (FS), frequency distribution, distance, and duration of a digital sound (§2.4). The usefulness of LUFS is that it can achieve the desired loudness level regardless of the underlying frequency distribution of the sound [49]. For example, the frequency distributions of woodpecker, seagull, and raven sounds (Figure 3) are different. Therefore, to achieve the same (target) perceived loudness for these sounds, they need to be normalized differently, mimicking the equal loudness curves (Figure 2b) for individual, constituent frequencies. We used an open-source implementation of LUFS [61] for this purpose.

Additionally, we kept the duration of different sounds in our database similar in order to minimize the effect of sound length on perceived loudness. Furthermore, we require listeners to wear headphones to control the distance between the sound source and the ears.

Table 3 shows an example of converting a bar chart containing four values to the desired loudness levels.

5.2.2 Sound Interval. We repeated the sounds multiple times while controlling the interval between the sounds to map the data values (Figure 4a). Suppose $T(t_1^s, \ldots, t_n^s)$ is the list of the sound interval for n data points. We need to define a mapping function, $t_i^s = Susurrus^T(d_i, s_i)$ that maps a data point d_i to a sound s_i having the target interval t_i^s in order to convey the data values. The decomposition of $Susurrus^T(d_i, s_i)$ function is given below:

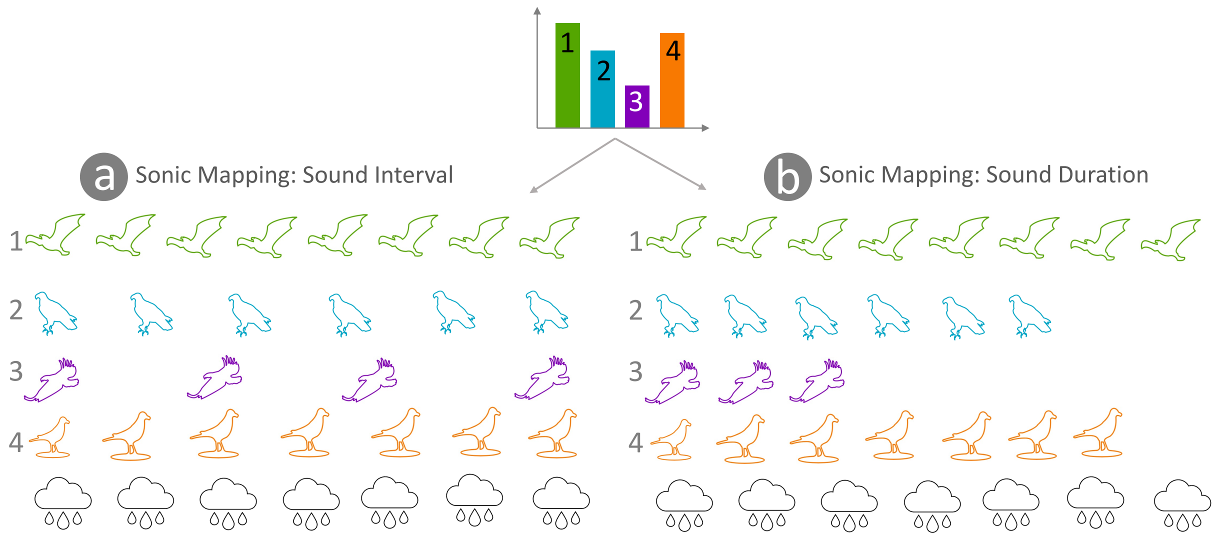


Figure 4: Sonic mapping using sound interval and duration. (a) Variable intervals between the sounds with constant total duration. Notice that the interval between two sounds for the raven bird (1st bar) is the shortest, representing the highest value. (b) Variable duration of the sounds with a constant interval between repetitions. Notice that all four birds have the same interval. However, the total duration of the raven sound (1st bar) is the longest, representing the highest value.

$$\hat{d_i} = \frac{d_i - D_{min}}{D_{max} - D_{min}}$$
$$t_i^s = T_{max} - (T_{min} + \hat{d_i} * (T_{max} - T_{min})) \quad (2)$$

The interval between the sounds are set to be inversely proportional to the data values. That is, a high value will mapped to a birds having a small interval. For example, in Figure 4a, the first bar has the highest value and the sound representing it (raven) has the lowest interval. We set $T_{min} = 0.01s$ and $T_{max} = 5s$. The total duration of the sounds were 20 seconds.

5.2.3 Sound Duration. We mapped each data value to a sound with variable duration (Figure 4b). The minimum and maximum sound duration was set to $SD_{min} = 0.5$ and $SD_{max} = 20$ seconds. Similar to Equation 2, the mapping function $Susurrus^{SD}(d_i, s_i)$ maps the duration of a sound s_i, assigned to represent the data point d_i, to the target sound duration sd_i^s. We repeated the sounds multiple times to achieve the target duration. For example, in Figure 4b, the target duration for the raven bird is 20 seconds (SD_{max}), representing the first bar with the highest value. The intervals between the repetition for all sounds were constant (0.1 seconds).

5.3 Supporting Gist with Parallel Sonification

Gist provides an overview of all data points so that users can infer patterns and trends. It is the most common application of sonification and a core action in AISA. We wanted to use parallel sonification for providing gist (**C3, C4**). We implemented gist (by default a 20 seconds long sonification) for three types of data visualization in this initial version: (1) bar chart, (2) line chart, and (3) multi-line chart. We determined the default duration (20 seconds) by experimenting with several time durations. We also provide keyboard shortcuts to customize the duration of sonification (Table 4). The 20 seconds gist is played in a loop until the user stops or pauses the sonification.

5.3.1 Bar chart. In a bar chart, a user should be able to identify the number of bars and their relative values for comparison. To support that, we mapped each bar to a unique natural sound and then played them together (in parallel). The selected audio channel (perceived loudness, sound interval, or duration) was set in proportion to the bar's height.

5.3.2 Line chart. In a line chart, the height of the line represents its value at a point. Since it represents continuous data, we needed an auditory channel that we can change continuously. Sound interval and duration are unsuitable for that. Thus, we used varying loudness to represent the change in a line chart.

5.3.3 Multi Line chart. Like a bar chart, we assigned a unique natural sound to each line in a multi-line chart. We changed the loudness of each natural sound, similar to a single-line chart.

5.4 Supporting Navigation, Situate, Select, and Details on Demand with Interaction

Interaction helps users navigate, situate, select, and obtain details on demand from an audio graph (**C4**). A user can press the Space button to play the gist (Table 4). A user can also increase or decrease the length of the sonification by pressing ↑ and ↓ arrow keyboard shortcuts.

We used ← and → arrow keys to allow users to move (navigate) from one data point to another. During navigation, users may get lost in the virtual auditory space. To situate themselves, similar to iSonic [70], users can press I to get verbal feedback from the system about their current status during navigation.

Shortcut	AISA Support	Details
Space	Gist	Play or stop the sonification
↑	Gist	Increase playback speed of the audio
↓	Gist	Decrease playback speed of the audio
←	Navigate	Move to the the previous data point
→	Navigate	Move to the the next data point
I	Situate	Play the current status
1 to 9	Select	Select a datapoint(s) of interest
I	Details on demand	Description of the selected datapoint(s) using speech-to-text
Esc	Control	Reset data selection

Table 4: Keyboard shortcuts in Susurrus. A user needs to press a control button (default Ctrl) with the keystrokes.

A user can press 1 to 9 to select data points. For example, the user can press 2 to select the second data point and listen to the sonification of that data point. In this selection mode, pressing the I key will play the details of the selected data point using text-to-speech software. A subsequent press on 2 will deactivate the data point in the sonification. This way, a user can listen to any combination of the data points, similar to how a user would interact with a visualization. Pressing the Esc button will reset any selection.

5.5 Web Application Design

To address **C5**, we implemented Susurrus as a web application. We used Python in the backend for sound processing and modern web technologies with JavaScript in the frontend for supporting user interaction. For performing the sonic mapping (e.g., LUFS) between data and sounds, we used `pyloudnorm`[2]. `Howler.js`[3] was used to implement audio playback features, such as play, pause, stop, and fast forward. We used Web speech API to verbally describe the data points. Finally, the accompanying visualizations were produced using D3.

6 PILOT STUDY

We conducted a pilot study with 5 BLV participants to assess the feasibility Susurrus and identify possible design issues.

6.1 Participants

We recruited 5 participants (male = 2, female = 3) from university mailing lists and word of mouth. All participants were legally blind—three had some light perception, and two were fully blind. All of our participants had familiarity with basic data statistics (e.g., *min*, *max*, *mean*). Two participants reported having prior musical training, either as a self-taught musician or through formal training. One participant had experience with sonification. Participant ages ranged from 19 to 35. Participation was voluntary, with no compensation.

[2]https://github.com/csteinmetz1/pyloudnorm
[3]https://howlerjs.com

6.2 Study Materials and Procedure

We created ten sonifications in total for the study: two each for bar, line, and multi-line charts using *loudness*, two for bar charts using *sound duration*, and two for bar charts using *sound interval* as sonic mapping. We deployed Susurrus along with the pre-loaded sonification on a public web server. Participants used their web browsers to listen to and interact with the sonification.

All study sessions were conducted remotely using teleconferencing software (e.g., Zoom). Each participant attended the session separately. We recorded and transcribed each session. Participants wore headphones during the study.

After giving consent, participants listened to the sonification, one at a time. We provided verbal instructions to the participants to interpret the sounds. We also guided participants to interact with the charts using keyboard shortcuts. To facilitate discussion and ideation, we asked participants to identify *minimum* and *maximum* and *compare* two data points from the sonification of the bar charts. We also asked participants to describe the *trends* from the sonification of line and multi-line charts. We followed a think-aloud protocol. Throughout the sessions, participants shared their feedback about the sonification with the research team.

6.3 Data Analysis

Two authors independently open-coded the anonymized transcripts of the participants' feedback and then conducted a thematic analysis. The authors met regularly to discuss and refine the codes and themes. During this process, we also discussed the codes and themes with the entire research team.

6.4 Findings and Recommendations

Participants were enthusiastic about the tool and found the sonification easy to understand. We noticed that participants quickly identified the natural sounds from the sonification and suggested that the sounds were easy to distinguish, even when played together. Participants further provided several recommendations to improve Susurrus.

6.4.1 Recommendation for Sonic Mapping. All participants found loudness to be the easiest for comparing data points. Participants easily identified the maximum and minimum values from the *gist*. For values with small differences, participants used *select* and *details on demand*.

P1, P2, and P4 reported that *sound interval* was confusing as a sonic mapping. This is because the sounds differed slightly in terms of duration. Even though we selected sounds with approximately the same duration, they are not exactly the same due to their ambient nature. This is an unavoidable problem with natural sounds as each sound is different from others. For example, a single chirp of a seagull is slightly shorter than a single chirp of a robin. While this did not affect the perception of loudness, it was problematic when we controlled the interval between the repetition of the sounds. Additionally, P1 and P5 mentioned that the repetition was annoying in some cases, especially when they were played with small intervals.

Participants found *sound duration* to be an easy mapping. However, we noticed that participants were slow to infer the order of the data since they had to wait till the end of the sounds.

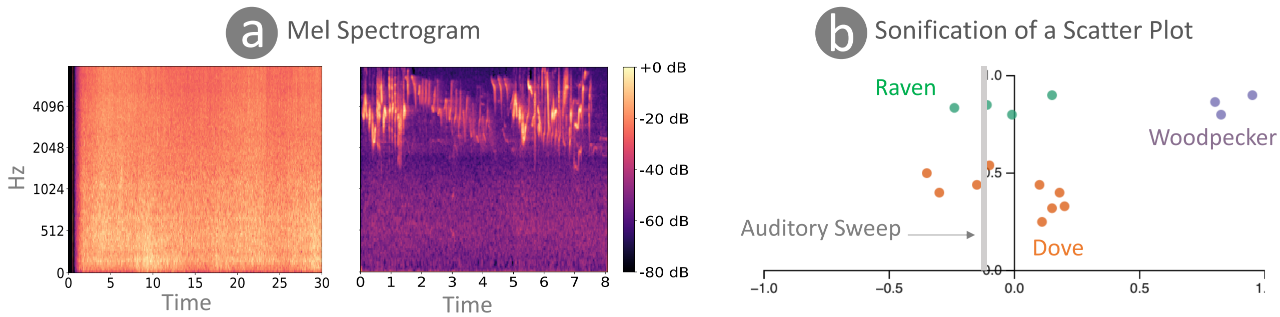


Figure 5: Revising sounds and supported charts in Susurrus. (a) We excluded sounds that have a wide range frequency spectrum (e.g., rain sound, on the left) for sonification and include sounds that have distinct frequency spectrum (e.g., a bird sound, on the right). The x axis represents time (seconds) while the y axis represents frequency (Hz). The color scale represents amplitude in decibel (dBFS). (b) We extended supported charts to scatter plots. Each class is represented with a distinct natural sound. An auditory sweep (the gray line) moves from left to right and plays the corresponding sounds whenever it intersects with a point or a group of points.

6.4.2 Recommendation for Selecting Sound Sources. Participants suggested selecting sounds with similar **semantic meaning** (P1, P3, P5) so they are easy to compare. They also suggested removing sounds that had no distinct pattern and consumed other sounds (e.g., rain). For example, participants found rain and bird sound difficult to compare. In contrast, multiple birds were easy to compare. P1 said:

> *"Bird and rain sounds were difficult to compare. It did not feel natural to compare two sounds that are so different in terms of what they represent physically. The birds were easy to compare."* (P1)

6.4.3 Recommendation for Number of Data Points. We used parallel sonification for bar and multi-line charts. While participants understood the presence of different sounds easily, they suggested to limit the number of parallel sounds to five so that the sonification remains manageable (P1, P2). This feedback was expected since we anticipated that we will need to account for the limited *bandwidth* of our auditory system.

7 SUSURRUS: FINAL IMPLEMENTATION

Based on the findings from the pilot study, we refined several components in Susurrus.

7.1 Loudness as Sonic Mapping

We decided to use perceived loudness as the sonic mapping for Susurrus because (1) our method is able to account for variable frequency spectrum of the sounds for mapping loudness; (2) participants in our pilot study found loudness to be an easy mapping; (2) sound interval was confusing to participants; and (3) sound duration is only applicable to bar charts and requires significantly more time for inference.

7.2 Ambient Theme

We separated sounds with a wide frequency range (white noises) from our sound database. White noise is a sound that consumes all frequencies with similar intensity [32]. For example, Figure 5a (on the left) presents a spectrogram for a rain sound, a white noise used in our pilot study. Its frequency distribution is expanded to the full spectrum with high intensities. Such white noises created confusion among participants in the pilot study as the white noise consumed frequencies of other sounds, making the other sounds hard to distinguish. In contrast, a bird only takes a fixed portion of the frequency spectrum (Figure 5a right). Therefore, white noises should not be used for mapping data. However, we noticed that white noises, with small intensity, are useful for creating ambience, such as a rainy day or a forest.

According to the pilot study, the sound sources need to be comparable. To facilitate this, we created a collection of different bird sounds that are freely available on the web. Susurrus uses the white noises as background sounds while the bird sounds are used for representing data. Thus, they create an ambient theme (e.g., bird sounds in the forest). It is worth noting that although we created one theme for demonstration, different types of sounds may create different themes (e.g., a coffee house, walking in a city, or sounds from the seaside), and they will be supported the same way.

Finally, to make the sonification consistent with the ambient theme, we decided to use variable intervals between the repetition of the sounds. This is because constant repetition can sound mechanical and reduce the pleasantness of the sounds. We play the sounds in repetition till a user pauses or stops the sounds by pressing the Space key.

7.3 Extending Chart Types

We extended the supported chart types to scatter plots. To sonify a scatter plot, we first assign each class of points to a unique natural sound. For example, in Figure 5b, the three sets of points are represented by Raven, Dove, and Woodpecker. We then convert the x-axis value of the points to a $[-1, 1]$ range so that we can represent the location of the points using stereo panning. We also convert the y-axis value of the points to the range of loudness levels.

The sonification starts with a mild white noise (i.e., auditory sweep) playing from the left ear to the right ear of a user using stereo panning. The continuous auditory sweep helps users situate

ID	Gender	Age	Vision Level	Education	Musical Experience
P1	Female	36	Blind	Doctorate	Music teacher
P2	Female	25	Blind	Bachelors	Pitch-based sonification
P3	Female	21	Blind	Associates	Pitch-based sonification
P4	Male	38	Low Vision	Bachelors	Pitch-based sonification
P5	Male	20	Blind	Bachelors	Sound engineer for media
P6	Female	43	Blind	Masters	Professional singer
P7	Female	20	Blind	High school	None
P8	Male	34	Blind	Doctorate	None
P9	Female	32	Blind	Masters	None
P10	Female	48	Low Vision	Associates	None
P11	Male	49	Blind	College	None
P12	Male	30	Blind	High school	None

Table 5: Participants demographics. Most participants were blind; two were low-vision. All were full-time screen reader users.

themselves in the virtual auditory space. In the case of an intersection between the data points and sweep, we play the points with their corresponding natural sounds, stereo locations, and loudness levels. For example, Figure 5b shows a case where two points belonging to Raven and Dove intersect with the auditory sweep. Both Raven and Dove sounds will be played together in this case where stereo panning and loudness will represent their coordinates. We repeat the natural sounds to indicate the number of points. However, the repetition can overlap and create annoyance if the points are very nearby. Thus, we only repeat a sound if it is not already playing. This way our method can represent the relative number of points (i.e., density) for each class and where they are in the 2D Cartesian space.

7.4 Implementation Notes

Susurrus is currently a research prototype. However, we have made our source code open-source, allowing others to build upon it or modify it to their needs. For example, with appropriate modification to our codebase, Susurrus can be deployed as a *plug-in* for browsers or screen readers. It only uses JavaScript libraries and runs in popular browsers (e.g., Chrome). Moreover, it is compatible with screen readers and supports keyboard shortcuts. Since many visualization tools (e.g., D3, Tableau) preserve raw data, if Susurrus plug-in can access the raw data, it can sonify them in response to users' keyboard commands, without needing a separate graphical interface. Prior work has shown that plug-ins are very effective in combating accessibility issues for blind users [40]. Similarly, Susurrus can provide API services (e.g., RESTful APIs), which other researchers and practitioners can integrate into their web interfaces. Finally, Susurrus can be integrated into existing efforts for accessible data visualization (e.g., Sonifier[4]), to enable multiple sonic profiles (e.g., natural sound, musical notes, artificial sounds such as sinusoidal and square waves) and conversational ability [9, 55].

8 EVALUATION

We conducted a user study with 12 BLV participants to understand the novel aspects of the final Susurrus prototype in the context of current sonification techniques. The study was reviewed and approved by our university's IRB office and the Research Advisory Council of the U.S. National Foundation of the Blind (NFB) in Baltimore, MD, USA. We intended to answer the following research questions with this study:

RQ1: How *effective* is Susurrus in representing *bar* charts, compared to existing sonification tools?
RQ2: How *effective* is Susurrus in representing *line* charts, compared to existing sonification tools?
RQ3: How does *musical background* of the participants affect their *performance* with Susurrus, compared to existing sonification tools?
RQ4: What are the benefits of using sonification proposed in Susurrus in terms of *user experience*?

RQ1 and RQ2 are motivated from the supported charts in Susurrus. RQ3 is based on prior works that showed that users with musical background can discern changes in pitch 60% better than untrained users. We anticipated this phenomenon could impact our study results and hence needs to be studied. Finally, RQ4 seeks to find hedonic values of natural sounds for sonification.

8.1 Participants

We recruited 12 BLV participants through the National Foundation of the Blind (NFB). Our inclusion criteria included participants owning stereo headphones, having no hearing impairments, and familiarity with web browsing using assistive technology (e.g., screen readers) of choice. Although prior experience with music or sound was not required, most respondents from the initial recruitment pool had music careers (e.g., professional singers or music teachers). We believe our recruitment flyer may have attracted participants with musical background. Some participants had also used musical note-based sonification techniques before, but none had used them regularly. Our final selection included 7 women and 5 men, with an average age of 33 (SD: 10.3). Half (6) of them had musical backgrounds or familiarity with pitch-based data sonification techniques. Table 5 presents participants' demographics.

[4]https://github.com/athersharif/sonifier

Task	Category	Chart	Purpose (RQ)	Example Question
T1	Point Estimation	Bar	RQ1, RQ3, RQ4	Which student has the highest (or lowest) score?
T2	Point Comparison	Bar	RQ1, RQ3, RQ4	Which one has a higher value between the scores sonified by the raven and woodpecker?
T3	Trend Identification	Line	RQ2, RQ3, RQ4	What is the overall trend for the stock price?
T4	Trend Forecasting	Line	RQ2, RQ3, RQ4	Based on the audio, what will be your forecast for the stock price in the near future?
T5	Trend Comparison	Line	RQ2, RQ3, RQ4	Which of the following is true about the prices of the two stocks in recent years?
T6	Point Estimation	Scatter	RQ4	Which side of your ear the crickets are buzzing?
T7	Point Comparison	Scatter	RQ4	Which pair of points have the shorter average distance between them?

Table 6: Task list for the user study. The tasks are designed for specific research questions.

8.2 Study Condition

We conducted a repeated-measures within-subject experiment with the following two conditions:

C1 Highcharts: Sonifications created using Highcharts [8].
C2 Susurrus: Sonifications created using our tool.

The Highcharts condition featured bar and line charts. We used the default setting from Highcharts: a *triangleMajor (Major 7th chord)* instrument with a frequency range from 520 Hz (C5 note) to 1,050 Hz (C6 note). Highcharts played the bar and line charts serially, where the data values were represented with the pitch in the frequency range. We include a video in the supplement with the sonification from Highcharts.

8.3 Study Tasks and Questions

We designed the study questions based on the RQs (Table 6). To answer **RQ1**, we presented the scenario of inferring test scores for *n* students from a bar chart. We designed two tasks (**T1** and **T2**). T1 and T2 are motivated by two common sonification tasks [23]: (1) **Point Estimation** for estimating the magnitude of a data point; and (2) **Point Comparison** for estimating multiple data points and comparing their magnitudes. Similar tasks have also appeared in previous sonification research [70]. Based on the two tasks, we generated six questions for each condition.

To answer **RQ2**, we designed three tasks (**T3**, **T4**, and **T5**) based on trend identification and comparison from a sonification of one or two stock prices. These tasks were also designed from common sonification tasks [23] and prior studies [70]. Using these three tasks, we generated six questions for each condition.

Lastly, we designed **T6** and **T7** for scatterplots, only to perform with Susurrus (C2). This is because Highcharts (C1) does not support finding the location of a data point in a scatterplot, as well as the sonification of multiple classes of points in scatterplots. Along with T1-T5, these two tasks are designed only to answer **RQ4**.

We provided multiple choices for each question. The supplemental materials include a detailed list of questions and multiple choices. The underlying task datasets in each condition had similar statistical properties and complexity.

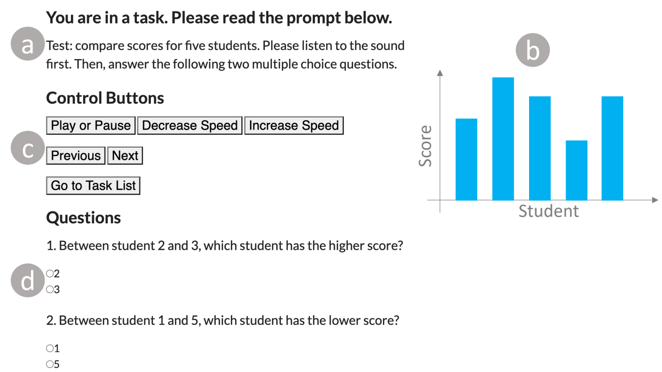


Figure 6: Web interface for the user study. (a) A prompt for the current task; (b) The corresponding visualization (not shown on the study web page); (c) Control buttons to play/pause sounds, select data points, and move between study tasks, mirroring the keyboard shortcuts in Table 4; (d) Questions relevant to the current task. The interface was fully accessible to screen reader users.

8.4 Study Procedure

Similar to our pilot study, we deployed Susurrus on a public server and conducted each session using video-conferencing software (e.g., Zoom). Our web interface (shown in Figure 6) was fully accessible to screen readers.

Each session started with participants signing the consent form and providing demographic information. We then shared the URL for the study. To minimize learning effects, we counterbalanced the order of study conditions and tasks. Note that the sonification for scatterplots always appeared last in Susurrus.

We provided verbal and written instructions to interpret the sounds and interact with the sounds with keyboard shortcuts. Both conditions supported interactions listed in Table 4; however, we disabled the details on demand (i.e., text-to-speech) mode of both tools to measure the sonification technique's effectiveness. Additionally, we provided training tasks and encouraged participants to ask questions. Once participants were comfortable with the condition and different sonification, we presented the study tasks.

Each session lasted for 1.5 hours and ended with an exit interview. We include the semi-structured questionnaire for the exit interviews as a supplement. To gauge the usability, we administered a post-completion System Usability Scale (SUS) questionnaire [6], which consists of 10 Likert scale statements, where the participant rated each statement on a scale of 1—strongly disagree—to 5—strongly agree. We also verbally administered NASA-TLX [22] to measure an individual's perceived workload.

The experimenter took notes during the session. All sessions were video-recorded and transcribed for post-analysis. The participant received a $25 Amazon gift card to compensate for their time.

8.5 Data Collection and Analysis

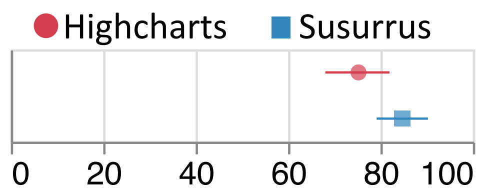
Similar to VoxLens [55], we used Accuracy of Extracted Information (AEI) to measure performance. AEI is a binary variable for a single question (i.e., "inaccurate" or 0 if the user was unable to answer the question correctly, and "accurate" or 1 otherwise). The overall accuracy was calculated by taking the ratio of the number of correct answers to the total number of questions and then converting it to a percentage.

Following guidelines for statistical analysis in HCI [14], we intentionally avoided traditional null-hypothesis-based statistical testing in favor of estimation methods to derive 95% confidence intervals (CIs) for all measures. We employed non-parametric bootstrapping [15] with $R = 1,000$ iterations. We also report the mean difference as a sample effect size and Cohen's d as a standardized measure of effect size [10].

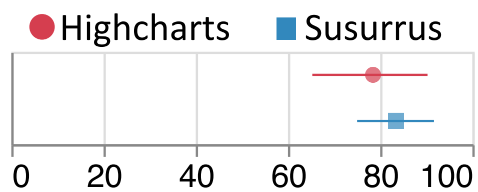
Similar to our pilot study, two authors of this paper independently open-coded the anonymized post-study interview transcripts and then conducted a thematic analysis. The coders met regularly to discuss and refine the codes and themes. The coders also discussed the codes and themes with the entire research team. All study data along with the study instructions and questions are available in our OSF repository.

8.6 Study Design Rationale

Our primary goal with this study was to understand the benefits and limitations of using the new sonification technique proposed in Susurrus. To achieve that, we decided participants should have experience with Susurrus and an existing solution that uses artificial or musical notes. We decided Highcharts is a suitable baseline because (1) it is a widely used visualization library that includes a state-of-the-art sonification API; (2) it is built upon over 20 years of empirical research (starting from Sonification Sandbox [66]); and (3) it is open sourced and, similarly to Susurrus, it can easily be integrated with a web browser. This was a requirement since we wanted to conduct the study online because of the COVID-19 pandemic. However, as a first-ever sonification tool based on natural sounds, our goal was not to beat a tested tool such as Highcharts quantitatively but to position our work in the sonification design space with evidence.



(a) Bar chart (T1, T2).



(b) Line chart (T3, T4, T5).

Figure 7: Accuracy of Extracted Information (AEI). Performance for bar and line charts with 95% confidence intervals (CI); the higher the better. (a) AEI for task T1 and T2. (b) AEI for task T3, T4, and T5.

A challenge in our study is that it contains multiple confounding factors such as audio type (natural vs. musical), sonic mapping (loudness vs. pitch), and play order (serial vs. parallel). Due to the fundamental difference in the waveform of natural (complex aperiodic) and musical sound (complex periodic), it is not possible to study these factors in isolation. Thus, we decided to compare the systems (Susurrus vs. Highcharts), not the individual features of the systems. The RQs reflect this decision.

8.7 Results

Here we report on our findings from the user study. We organize the findings around the research questions.

8.7.1 RQ1 – Bar Charts. For task T1 and T2 (Figure 7a), participants were 84.7% (CI = [79.2, 90.3]) accurate with Susurrus. In contrast, participants were 75.2% (CI = [68.05, 81.9]) accurate with Highcharts. On average, participants were 9.7% (CI = [8.6, 13.7]) more accurate with Susurrus for bar charts. The effect size (Cohen's $d = 0.45$) indicates a medium effect of the study condition.

8.7.2 RQ2 – Line Charts. For task T3, T4, and T5 (Figure 7b), participants were 83.5% (CI = [75.0, 91.7]) accurate with Susurrus. In contrast, participants were 78.4% (CI = [65.2, 90.3]) accurate with Highcharts. On average, participants were 4.3% (CI = [2.3, 8.1]) more accurate with Susurrus for line charts. The standardized effect size (Cohen's $d = 0.20$) indicates a small effect of the study condition.

8.7.3 RQ3 – Effect of Musical Background. To answer RQ3, we divided participants into two groups: `Non-experts`: participants who had no prior experience in music and pitch-based sonification (P7-12); and `Experts`: participants who had those experiences (P1-6). We then compared their performance for bar and line charts.

On average, for task T1 and T2 (Figure 8a-top), `Non-experts` were 86.6% (CI = [76.7, 96.7]) accurate while answering the questions using Susurrus. In contrast, `Non-experts` were only 66.8% (CI =

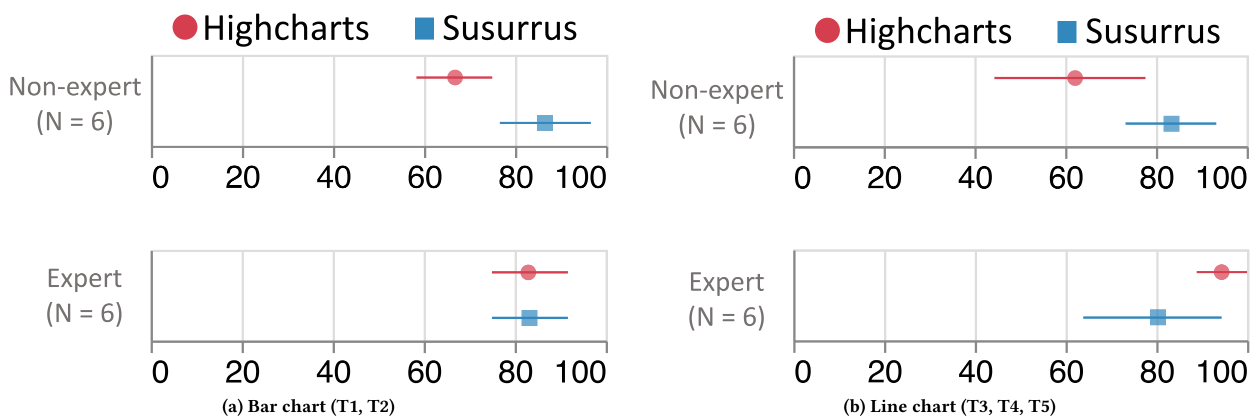


Figure 8: Accuracy of Extracted Information (AEI) for expert vs. non-experts. Experts are musically trained, and non-expert are not. The higher the better. (a) AEI for task T1 and T2. (b) AEI for task T3, T4, and T5.

[58.3, 75.0]) accurate while using Highcharts. The difference, 19.5% (CI = [12.8, 26.5]), with an effect size of 0.86 (Cohen's *d*) indicates a large effect of the study condition.

We noticed that `Experts` achieved similar accuracy for both sonification tools while performing T1 and T2. On average, for task T1 and T2 (Figure 8a-bottom), `experts` were 83.3% (CI = [75.0, 91.7]) accurate while using Susurrus. `Experts` achieved almost exactly the same accuracy, 83.0% (CI = [75.0, 91.7]), while using Highcharts. The difference, 0.27% (CI = [0.0, 0.33]), with an effect size of 0.01 (Cohen's *d*) indicates no practical difference.

We observed a similar trend for line charts. `Non-experts` performed better while using Susurrus for T3, T4, and T5. On average, for task T3, T4 and T5 (Figure 8b-top), `Non-experts` were 83.4% (CI = [73.3, 93.3]) accurate while using Susurrus. In contrast, `Non-experts` were only 62.2% (CI = [44.4, 77.7]) accurate while using Highcharts. The large difference in accuracy—20.5% (CI = [13.4, 28.4])—with an effect size of 0.78 (Cohen's *d*) indicates a large effect of the study condition.

Finally, we noticed `Experts` performed better for task T3, T4, and T5 while using Highcharts (Figure 8b-bottom). On average, `Experts` were 80.4% (CI = [63.9, 94.4]) accurate while using Susurrus. However, `Experts` were 94.4% (CI = [88.9, 100]) accurate while using Highcharts. The difference in accuracy, 13.9% (CI = [9.3, 19.5]), with an effect size of 0.41 (Cohen's *d*) indicates a medium effect of the study condition.

8.7.4 RQ4 – User Experience. We discuss the relevant themes emerged from the post-study interviews below.

Parallel vs. Serial Sonification. In the post-study interview, several participants mentioned that natural sounds were easy to distinguish and helped them to separate multiple categories of a bar chart (P1-12). Further, they found it easy to compare data values since they did not have to memorize the sounds. Using our parallel sonification technique, they could listen to the sounds in reference to each other. In contrast, when using Highcharts, participants listened to the serial sonification multiple times to answer the questions for bar charts. P1 and P4 said,

> *"It was easy to compare because I knew the [natural] sounds and could identify them easily."* (P1)

> *"Natural sounds helped me separate the students [bars]. I was able to differentiate the data I was trying to identify. With beep sounds [Highcharts], it did not feel like I am comparing 4 or 5 students."* (P4)

This overall positive feedback was reflected in the quantitative results for bar charts, as separating and estimating multiple data points were easier using Susurrus.

Loudness vs. Pitch. In the post-study interviews, participants suggested that natural sounds and loudness can provide an alternative to people who are not familiar with musical notes and pitches. Several participants mentioned that pitch is difficult for them to decode (P7-9, P12). P7 said,

> *"Pitch is confusing to me. I can understand high and low frequencies, but the middle ones [mid-range frequencies] really confuse me."* (P7)

P1, who is a music teacher, said,

> *"I have worked with plenty of students at university who cannot recognize differences in pitch. For them, natural sounds will be a better choice."* (P1)

Half of our participants were well-attuned to pitch and musical notes from prior experience. They mentioned that natural sound-based sonification was also easy for them. On the flip side, participants suggested that it is possible that loudness can be confusing too, especially for people with hearing loss (P1, P4).

Localizing Data Points in Scatterplots. Participants understood the different classes of data points easily when listening to scatterplots (P1-12). P6's comment below summarizes how participants interpreted the sonification:

> *"It is really amazing! If I hear only one Dove in my right ear, I know there is only one type of point on the right side. If I hear* ***several*** *Doves and a* ***single*** *Woodpecker on my left ear* ***together****, I know there are several Dove points on the left side with one Woodpecker*

> *point overlapping with them. If I hear the Woodpecker sound a little bit later and in the middle of my ears, I know it is a bit far away from the Dove points."* (P6)

Thus, our sonification method not only enabled participants to localize points but also helped them measure the relative distance between the points. P12 said that they have been trying to interpret "XY axis-based data" all their life and our sonification is the most intuitive they have found so far.

Overview and Zoom with Susurrus. Susurrus provided an auditory equivalent of visualization mantra by Ben Shneiderman [56]: *Overview first, zoom and filter, details on demand.* While participants did not use the details on demand feature as it was disabled for study purpose, we noticed participants used gist for obtaining an overview of data and then selected specific data points using our keyboard shortcuts for comparison. This helped them understand data using a top-down approach (P1-4, P8-9) and concentrate on data of interest (P3-4, P6-9).

Natural Sound helps in Concentration and Focus. Participants appreciated the acoustic nature of Susurrus and suggested it changed their mood positively. P9 mentioned that they felt relaxed when listening to natural sounds.

> *"I tend to concentrate better when I am relaxed. The natural sounds made me relaxed so it was easier for me to answer the questions. It did not feel like a test."* (P9)

Enjoyment and Relaxation with Natural Sounds. Sonification with natural sounds provided a sense of realism to participants. P11 compared natural sound-based sonification to taking a walk in a park. P8 appreciated the background white noise in each sonification, which created the ambience and provided an immersive experience. Even participants who performed better using musical notes preferred natural sounds for long-term use. P4 said,

> *"Listening to the same artificial monotones [single note tone] over a long period of time can be monotonous and tiring. I can imagine natural sounds to be more versatile and give me more options."* (P4)

Personalizing Natural Sounds. Natural sounds are better suited for personalization than musical or artificial sounds. Participants enthusiastically asked us to expand the database of natural sounds to other kinds of sounds, such as `farm animals (P1)`, `trains and cars (P8)`, `ocean sounds (P3)`, `water drops (P11)`, and `cats (P5)`. Participants believed that doing so will help them adjust sonification to their liking that existing solutions do not offer. For example, P10 said,

> *"It will be great if I can set a natural sound to a stock, the one that I track all the time, I can get used to the sound and find important information quickly over the time. I can imagine it will also help me remember and refer to data. It is difficult to remember what happened to a musical note, but I can easily remember a train, or a car."* (P10)

Potential Use Cases for Natural Sounds. Participants mentioned several use cases for naturals sounds beyond sonification. They

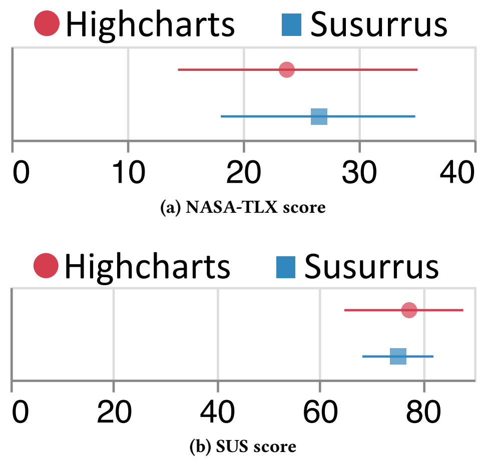


Figure 9: NASA-TLX and SUS scores. (a) NASA-TLX score calculated from on six workload measures. (b) SUS score calculated from ten usability measures.

were particularly interested in using our technique for navigation (P1, P5, P7-11).

> *"The louder the sound is, the closer I am to an object!"* (P6)

P6 thought musical notes were not well suited for such applications; presumably, natural sounds would be. P3 mentioned that natural sounds could help them navigate their computers since screen reader descriptions can sometimes be overwhelming. P3 suggested assigning different natural sounds to different tabs so that they can identify the tabs quickly. Finally, P12, who teaches BLV high school students, thought scatterplot sonification could be used with Braille to create a multi-modal interface. P12 said,

> *"We use Braille graphs or papers where there are multiple objects, and we ask students to pick out a certain object. Using audio stimuli with touch would be much more engaging not only for students, but also for teachers. I would prefer natural sounds with spatial positioning for that over musical notes because it is more real and engaging."* (P12)

8.7.5 Perceived Workload. Figure 9a presents NASA-TLX ratings for BLV participants in two conditions. On average, participants reported a similar workload for both sonification tools. For Susurrus, the average NASA-TLX score was 26.6 (CI = [18.1, 34.9]). For Highcharts, the average NASA-TLX score was 23.8 (CI = [14.4, 35.1]). The small effect size 0.1 (Cohen's *d*) indicates very small practical difference.

8.7.6 Usability. On average, participants rated both sonification tools equally in terms of usability (Figure 9b). For Susurrus, the average SUS score was 75.3 (CI = [68.3, 82.1]). For Highcharts, the score was 77.4 (CI = [64.8, 87.9]). The small effect size 0.08 (Cohen's *d*) indicates very small practical difference.

9 DISCUSSION

In this section, we first summarize and generalize our study results. We then discuss design implications, limitations, and future plans for Susurrus.

9.1 Summary of Study Results

To answer **RQ1**, we found that Susurrus is better suited than Highcharts to represent bar charts with five or less categories. The standardized effect size (d = 0.45) indicated a medium effect of Susurrus on participants' performance. Further, in the post-study interviews, participants suggested that Susurrus made it easier to compare values in a bar chart since (1) they could compare values without the need to memorize them, and (2) the sounds were easier to differentiate. We discuss ways to extend Susurrus for more than five data points in §9.2.2.

We found that Susurrus was equally effective, in comparison to Highcharts, to represent line charts (**RQ2**). Based on the post-study interviews, loudness is an effective mapping for showing trends in a line chart.

To answer **RQ3**, we found that participants' prior musical background affected their performance. Both experts and non-experts performed equally when using Susurrus. However, non-experts struggled with answering questions from pitch-based artificial sounds. In contrast, experts were highly accurate in discerning changes in pitch from artificial sounds. This result matched our intuition since, based on prior research, we anticipated that experts would be better at detecting changes in pitch (§2.2). This disparity in performance suggests Susurrus is most useful to participants with no formal musical background.

Finally, we found that Susurrus improved user experience in terms of enjoyment, concentration, and relaxation (**RQ4**). This result validated our initial motivation to use natural sounds for sonification. We discuss future work to improve pleasantness with Susurrus in §9.2.4.

9.2 Design Implications and Future Work

Based on our experience with this work, we outline the following design implications for future research.

9.2.1 Extending Chart Types. A limitation of Susurrus is that it currently only supports bar, line, and scatterplots. Although these are widely used charts, the web features a much wider variety of visualization (e.g., node-link-based graphs, tree maps, and heat maps). We believe adopting our technique for a new chart can be straightforward, or require additional design depending on the chart type, or may not be feasible. For example, a pie chart can be considered a radial counterpart of a bar chart and can be sonified in parallel with little change. However, a node-link-based graph is more complex and requires a specific task-driven design. In contrast, a heat map that encodes 2D spatial data (e.g., correlations of multiple variables) may not be an easy feat. Therefore, it is important to investigate which aspects of a chart can be sonified easily and which cannot to raise awareness among visualization researchers and practitioners.

9.2.2 Parallel Sonification vs. Bandwidth. A key insight of our findings is that natural sounds can sonify categorical data simultaneously (i.e., in parallel), but only when the number of categories is small (e.g., five or fewer). Surprisingly, this limit of five or fewer is in line with the seminal work on pitch-based auditory displays, where Pollack reported that the average listener could reliably distinguish only six pitches [47]. Therefore, the low bandwidth of the human auditory sense is a common challenge for data sonification, not unique to Susurrus.

We envision several ways to mitigate these limitations of parallel sonification. The first is to apply the information-seeking mantra, *"overview first, zoom and filter, then details-on-demand"* [56] for larger datasets. For example, we can first organize data into a virtual hierarchy with a maximum of five items at any level. Consider 20 bars being hierarchically clustered, with the top level containing 5 items and the secondary level containing 4 items each ($5 \times 4 = 20$ items, in total). Susurrus can first play the top-level items in parallel to provide an overview of the 20 items. Users can then select a top-level item and listen to the corresponding secondary-level items (4 at a time) in parallel. Such a method will open new research directions, such as finding an effective hierarchical clustering criterion and supporting fluid interaction to navigate the hierarchy. The fluid interaction can be achieved by a multi-wheel-based input device (e.g., [59]), where different wheels are dedicated to navigating different levels.

Another potential solution is data filtering. Users can filter or manually choose a small subset of data points and use Susurrus to sonify them in parallel. The filtering method can be supported by integrating features such as natural language queries [44] and semi-automated insight generation engines [12]. In the future, we aim to empirically evaluate these potential solutions to scale Susurrus.

9.2.3 Prior Experience as a Design Dimension. Prior work on accessible data representation has rarely explored participants' backgrounds or experiences during design. However, our findings indicate that individuals' backgrounds (e.g., musically trained or not) can significantly affect their ability to extract information from sonification. We hope our work will motivate future research to consider this novel design dimension.

9.2.4 Synthesizing Sounds. Blending natural sounds into a single mix is a significant challenge in this work. To increase the quality of blending, we experimented with several neural networks. For example, we experimented with *variational autoencoders* to learn the underlying latent representation of natural sounds, allowing us to produce these sounds without external noise [50]. However, the results were not convincing. Future work may explore more advanced models, such as a hierarchical recurrent variational autoencoder for learning the latent spaces.

Finally, the sounds used in this paper are collected from freely available web content. However, these sounds were not originally designed for data sonification. We believe a dataset of natural sound sources specifically designed for data sonification will produce a better mix and increase overall aesthetics. Future work should curate such a sound library.

9.2.5 New Opportunities: Personalized Data Sonification, Discovering Patterns, and Immersive Data Representation. A secondary benefit of our technique is that natural sounds can evoke emotions and give listeners a sense of realism. This can be leveraged to make personalized data sonification. For instance, participants enthusiastically suggested expanding supported themes by adding sounds, such as car sounds and farm animals, so that they could choose a theme matching their mood. In contrast, Artificial sounds are less suitable for customization, except for changing musical instruments (e.g., changing a guitar sound to a violin). However, this will likely benefit those with a musical background. Listening to artificial sounds for a long time can also be monotonous [33]. Thus, natural sounds can offer a less monotonous, more personalized experience to long-term continuous data consumption, such as surveillance or network traffic data. We believe such personalized experiences will motivate future research. We lay down this path in Figure 10.

Yet another opportunity is to integrate natural sounds in immersive and multi-modal data representation. Natural sounds commonly represent acoustic environments in virtual reality (VR) [26]. Since such environmental sounds can represent abstract data, as we show in this work, future research can investigate how visualization and sonification can be integrated into immersive, accessible data representation in VR.

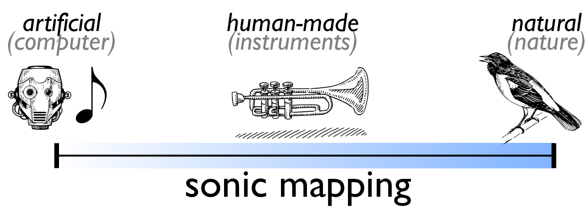


Figure 10: Sonic mapping spectrum for sonification. Many sonifications use computer-generated artificial sounds (left). Others use human-made artifacts such as musical instruments (middle). Our work proposes the use of natural sounds for sonification (right).

10 CONCLUSION

We have presented an approach for using natural sounds for sonification, which could be seen as existing on the far end of a spectrum of sonic mappings whose other end is characterized by artificial, single-frequency tones (Figure 10). While obviously not based on a single frequency, natural sounds can represent data. Our work is grounded in the observation that natural sounds are integrated into our day-to-day life, are easily distinguishable, and have hedonic values for meditation and well-being. We designed Susurrus, a sonification tool that uses several intelligent audio processing functionalities to blend multiple natural sounds into a single coherent sound. To evaluate the tool, we conducted a summative user study. Our findings indicate that Susurrus improved an individual's performance in understanding categorical data and it is most useful to people who do not have musical training.

POSTSCRIPT

We close this paper with the eternal question: *What does the fox say?* A raspy bark, it turns out, or sometimes an eerie scream. Both would be useful—albeit potentially unsettling—natural sounds for Susurrus.

ACKNOWLEDGMENTS

This work was partially supported by the U.S. National Science Foundation grant 2211628. Any opinions, findings, and conclusions, or recommendations expressed here are those of the authors and do not necessarily reflect the views of the funding agencies.

REFERENCES

[1] Jesper J. Alvarsson, Stefan Wiens, and Mats E. Nilsson. 2010. Stress recovery during exposure to nature sound and environmental noise. *International Journal of Environmental Research and Public Health* 7, 3 (2010), 1036–1046. https://doi.org/10.3390/ijerph7031036
[2] Apple, Inc. 2021. Audio Graphs. Available at https://developer.apple.com/documentation/accessibility/audio_graphs.
[3] Andrea Batch, Biswaksen Patnaik, Moses Akazue, and Niklas Elmqvist. 2020. Scents and Sensibility: Evaluating Information Olfactation. In *Proceedings of the ACM Conference on Human Factors in Computing Systems*. ACM, New York, NY, USA, 1–14. https://doi.org/10.1145/3313831.3376733
[4] Marc G. Berman, John Jonides, and Stephen Kaplan. 2008. The cognitive benefits of interacting with nature. *Psychological Science* 19, 12 (2008), 1207–1212. https://doi.org/10.1111/j.1467-9280.2008.02225.x
[5] Albert S. Bregman. 1994. *Auditory Scene Analysis: The Perceptual Organization of Sound*. MIT Press, MA, USA. https://doi.org/10.7551/mitpress/1486.001.0001
[6] John Brooke. 1996. SUS – A quick and dirty usability scale. In *Usability Evaluation in Industry*. CRC Press, Boca Raton, FL, USA, 194–200.
[7] Rachel T. Buxton, Amber L. Pearson, Claudia Allou, Kurt Fristrup, and George Wittemyer. 2021. A synthesis of health benefits of natural sounds and their distribution in national parks. *Proceedings of the National Academy of Sciences* 118, 14 (2021), e2013097118. https://doi.org/10.1073/pnas.2013097118
[8] Stanley J. Cantrell, Bruce N. Walker, and Øystein Moseng. 2021. Highcharts Sonification Studio: An Online, Open-Source, Extensible, And Accessible Data Sonification Tool. In *Proceedings of the International Conference on Auditory Display*. International Community For Auditory Display, Atlanta, GA, USA, 7. https://doi.org/10.21785/icad2021.005
[9] Pramod Chundury, Biswaksen Patnaik, Yasmin Reyazuddin, Christine Tang, Jonathan Lazar, and Niklas Elmqvist. 2022. Towards Understanding Sensory Substitution for Accessible Visualization: An Interview Study. *IEEE Transactions on Visualization and Computer Graphics* 28, 1 (2022), 1084–1094. https://doi.org/10.1109/TVCG.2021.3114829
[10] Jacob Cohen. 1988. *Statistical Power Analysis for the Social Sciences*. Routledge, New York, NY, USA. https://doi.org/10.4324/9780203771587
[11] Martin Cooke and Daniel P. W. Ellis. 2001. The auditory organization of speech and other sources in listeners and computational models. *Speech Communication* 35, 3-4 (2001), 141–177. https://doi.org/10.1016/S0167-6393(00)00078-9
[12] Zhe Cui, Sriram Karthik Badam, Mehmet Adil Yalçin, and Niklas Elmqvist. 2019. DataSite: Proactive visual data exploration with computation of insight-based recommendations. *Information Visualization* 18, 2 (2019), 1–14. https://doi.org/10.1177/1473871618806555
[13] Mohamed Debashi and Paul Vickers. 2018. Sonification of network traffic flow for monitoring and situational awareness. *PLoS One* 13, 4 (2018), e0195948.
[14] Pierre Dragicevic. 2016. Fair statistical communication in HCI. In *Modern Statistical Methods for HCI*, Judy Robertson and Maurits Kaptein (Eds.). Springer, Cham, Germany, 291–330. https://doi.org/10.1007/978-3-319-26633-6
[15] Bradley Efron. 1992. Bootstrap methods: another look at the jackknife. In *Breakthroughs in Statistics*. Springer, New York, NY, USA, 569–593. https://doi.org/10.1007/978-1-4612-4380-9_41
[16] Frank Elavsky, Cynthia L. Bennett, and Dominik Moritz. 2022. How accessible is my visualization? Evaluating visualization accessibility with Chartability. *Computer Graphics Forum* 41, 3 (2022), 57–70. https://doi.org/10.1111/cgf.14522
[17] Niklas Elmqvist. 2023. Visualization for the Blind. *Interactions* 30, 1 (Jan/Feb 2023). To appear.
[18] Stephanie Elzer, Edward Schwartz, Sandra Carberry, Daniel Chester, Seniz Demir, and Peng Wu. 2008. Accessible bar charts for visually impaired users. In *Proceedings of the IASTED International Conference on Telehealth and Assistive Technologies*. International Association of Science and Technology for Development, Calgary, AB, Canada, 55–60.
[19] Sean Follmer, Daniel Leithinger, Alex Olwal, Akimitsu Hogge, and Hiroshi Ishii. 2013. inFORM: dynamic physical affordances and constraints through shape and object actuation. In *Proceedings of the ACM Symposium on User Interface Software and Technology*. ACM, New York, NY, USA, 417–426. https://doi.org/10.1145/2501988.2502032

[20] Bill Gibson. 2020. *The Ultimate Live Sound Operator's Handbook*. Rowman & Littlefield Publishers, Lanham, USA.

[21] Vicki L Hanson, Anna Cavender, and Shari Trewin. 2015. Writing about accessibility. *Interactions* 22, 6 (2015), 62–65.

[22] Sandra G. Hart. 2006. NASA-task load index (NASA-TLX); 20 years later. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, Vol. 50. Sage Publications, Los Angeles, CA, 904–908.

[23] Thomas Hermann, Andy Hunt, and John Neuhoff. 2011. *The Sonification Handbook*. Logos Publishing House, Berlin, Germany.

[24] Leona Holloway, Kim Marriott, and Matthew Butler. 2018. Accessible maps for the blind: Comparing 3D printed models with tactile graphics. In *Proceedings of the ACM Conference on Human Factors in Computing Systems*. ACM, New York, NY, USA, 1–13. https://doi.org/10.1145/3173574.3173772

[25] Leona M. Holloway, Cagatay Goncu, Alon Ilsar, Matthew Butler, and Kim Marriott. 2022. Infosonics: Accessible Infographics for People who are Blind using Sonification and Voice. In *Proceedings of the ACM Conference on Human Factors in Computing Systems*. ACM, New York, NY, USA, 480:1–480:13. https://doi.org/10.1145/3491102.3517465

[26] Stine S. Johansen, Niels van Berkel, and Jonas Fritsch. 2022. Characterising Soundscape Research in Human-Computer Interaction. In *Proceedings of the ACM Conference on Designing Interactive Systems Conference*. ACM, New York, NY, USA, 1394–1417. https://doi.org/10.1145/3532106.3533458

[27] Shakila Cherise S. Joyner, Amalia Riegelhuth, Kathleen Garrity, Yea-Seul Kim, and Nam Wook Kim. 2022. Visualization Accessibility in the Wild: Challenges Faced by Visualization Designers. In *Proceedings of the ACM Conference on Human Factors in Computing Systems*. ACM, New York, NY, USA, 83:1–83:19. https://doi.org/10.1145/3491102.3517630

[28] Crescentia Jung, Shubham Mehta, Atharva Kulkarni, Yuhang Zhao, and Yea-Seul Kim. 2022. Communicating Visualizations without Visuals: Investigation of Visualization Alternative Text for People with Visual Impairments. *IEEE Transactions on Visualization and Computer Graphics* 28, 1 (2022), 1095–1105. https://doi.org/10.1109/TVCG.2021.3114846

[29] Shaun K. Kane and Jeffrey P. Bigham. 2014. Tracking @stemxcomet: teaching programming to blind students via 3D printing, crisis management, and Twitter. In *Proceedings of the ACM Symposium on Computer Science Education*. ACM, New York, NY, USA, 247–252. https://doi.org/10.1145/2538862.2538975

[30] Rachel Kaplan. 2001. The nature of the view from home: Psychological benefits. *Environment and Behavior* 33, 4 (2001), 507–542. https://doi.org/10.1177/00139160121973115

[31] Stephen Kaplan. 1995. The restorative benefits of nature: Toward an integrative framework. *Journal of Environmental Psychology* 15, 3 (1995), 169–182. https://doi.org/10.1016/0272-4944(95)90001-2

[32] Steven M. Kay and Stanley Lawrence Marple. 1981. Spectrum Analysis—A Modern Perspective. *Proc. IEEE* 69, 11 (1981), 1380–1419.

[33] Fredrik Kilander and Peter Lönnqvist. 2002. A whisper in the woods—an ambient soundscape for peripheral awareness of remote processes. In *Proceedings of the International Conference on Auditory Display*. International Community For Auditory Display, Atlanta, GA, USA, 5.

[34] Gregory Kramer, Bruce Walker, Terri Bonebright, Perry Cook, John H. Flowers, Nadine Miner, John Neuhoff, Robin Bargar, Stephen Barrass, Jonathan Berger, Grigori Evreinov, W. Tecumseh Fitch, Matti Gröhn, Steve Handel, Hans Kaper, Haim Levkowitz, Suresh Lodha, Barbara Shinn-Cunningham, Mary Simoni, and Sever Tipei. 1997. *Sonification Report: Status of the Field and Research Agenda*. Technical Report. National Science Foundation.

[35] Mathieu Le Goc, Charles Perin, Sean Follmer, Jean-Daniel Fekete, and Pierre Dragicevic. 2018. Dynamic composite data physicalization using wheeled microrobots. *IEEE Transactions on Visualization and Computer Graphics* 25, 1 (2018), 737–747. https://doi.org/10.1109/TVCG.2018.2865159

[36] Lindosland. 2005. Equal-loudness contours created by lindosland using OpenOffice Draw by reference to ISO 226:2003 standard. http://en.wikipedia.org/wiki/Image:Lindos1.svg

[37] Dan Lockton, Flora Bowden, Clare Brass, and Rama Gheerawo. 2014. Birdwattching: exploring sonification of home electricity use with birdsong. In *Proceedings of the Conference on Sonification of Health and Environmental Data*. University of York, York, UK, 6. https://www.york.ac.uk/media/c2d2/media/sonihedconference/Lockton_etal_SoniHED_2014.pdf

[38] Alan Lundgard, Crystal Lee, and Arvind Satyanarayan. 2019. Sociotechnical Considerations for Accessible Visualization Design. In *Proceedings of the IEEE Conference on Visualization*. IEEE, Piscataway, NJ, USA, 16–20. https://doi.org/10.1109/VISUAL.2019.8933762

[39] Alan Lundgard and Arvind Satyanarayan. 2022. Accessible Visualization via Natural Language Descriptions: A Four-Level Model of Semantic Content. *IEEE Transactions on Visualization and Computer Graphics* 28, 1 (2022), 1073–1083. https://doi.org/10.1109/TVCG.2021.3114770

[40] Farhani Momotaz, Md Touhidul Islam, Md Ehtesham-Ul-Haque, and Syed Masum Billah. 2021. Understanding Screen Readers' Plugins. In *Proceedings of the ACM Conference on Computers and Accessibility*. ACM, New York, NY, USA, 15:1–15:10. https://doi.org/10.1145/3441852.3471205

[41] Brian C. J. Moore. 2012. *An Introduction to the Psychology of Hearing* (6th ed.). Emerald Group Publishing, Bingley, UK.

[42] Meinard Müller. 2015. *Fundamentals of Music Processing: Audio, Analysis, Algorithms, Applications*. Springer, Cham, Germany. https://doi.org/10.1007/978-3-319-21945-5

[43] Tamara Munzner. 2014. *Visualization Analysis and Design*. CRC Press, Boca Raton, FL, USA.

[44] Arpit Narechania, Arjun Srinivasan, and John T. Stasko. 2021. NL4DV: A Toolkit for Generating Analytic Specifications for Data Visualization from Natural Language Queries. *IEEE Transactions on Visualization and Computer Graphics* 27, 2 (2021), 369–379. https://doi.org/10.1109/TVCG.2020.3030378

[45] Biswaksen Patnaik, Andrea Batch, and Niklas Elmqvist. 2018. Information olfactation: Harnessing scent to convey data. *IEEE Transactions on Visualization and Computer Graphics* 25, 1 (2018), 726–736. https://doi.org/10.1109/TVCG.2018.2865237

[46] Bryan C. Pijanowski, Luis J. Villanueva-Rivera, Sarah L. Dumyahn, Almo Farina, Bernie L. Krause, Brian M. Napoletano, Stuart H. Gage, and Nadia Pieretti. 2011. Soundscape ecology: the science of sound in the landscape. *BioScience* 61, 3 (2011), 203–216. https://doi.org/10.1525/bio.2011.61.3.6

[47] Irwin Pollack. 1952. The information of elementary auditory displays. *The Journal of the Acoustical Society of America* 24, 6 (1952), 745–749. https://doi.org/10.1121/1.1906969

[48] Eleanor Ratcliffe, Birgitta Gatersleben, and Paul T. Sowden. 2013. Bird sounds and their contributions to perceived attention restoration and stress recovery. *Journal of Environmental Psychology* 36 (2013), 221–228. https://doi.org/10.1016/j.jenvp.2013.08.004

[49] Recommendation ITU-R BS.1770. 2022. Algorithms to measure audio programme loudness and true-peak audio level. https://tech.ebu.ch/docs/r/r128.pdf

[50] Adam Roberts, Jesse Engel, Colin Raffel, Curtis Hawthorne, and Douglas Eck. 2018. A hierarchical latent vector model for learning long-term structure in music. In *Proceedings of the International Conference on Machine Learning*. PMLR, Stockholm, Sweden, 4364–4373.

[51] Kavous Salehzadeh Niksirat, Chaklam Silpasuwanchai, Mahmoud Mohamed Hussien Ahmed, Peng Cheng, and Xiangshi Ren. 2017. A framework for interactive mindfulness meditation using attention-regulation process. In *Proceedings of the ACM Conference on Human Factors in Computing Systems*. ACM, New York, NY, USA, 2672–2684. https://doi.org/10.1145/3025453.3025914

[52] William A. Sethares. 2005. *Tuning, Timbre, Spectrum, Scale*. Springer Science & Business Media, London, UK.

[53] Ather Sharif, Sanjana Shivani Chintalapati, Jacob O. Wobbrock, and Katharina Reinecke. 2021. Understanding Screen-Reader Users' Experiences with Online Data Visualizations. In *Proceedings of the ACM Conference on Computers and Accessibility*. ACM, New York, NY, USA, 14:1–14:16. https://doi.org/10.1145/3441852.3471202

[54] Ather Sharif, Olivia H. Wang, and Alida T. Muongchan. 2022. "What Makes Sonification User-Friendly?" Exploring Usability and User-Friendliness of Sonified Responses. In *Proceedings of the ACM Conference on Computers and Accessibility*. ACM, New York, NY, USA, 45:1–45:5. https://doi.org/10.1145/3517428.3550360

[55] Ather Sharif, Olivia H. Wang, Alida T. Muongchan, Katharina Reinecke, and Jacob O. Wobbrock. 2022. VoxLens: Making Online Data Visualizations Accessible with an Interactive JavaScript Plug-In. In *Proceedings of the ACM Conference on Human Factors in Computing Systems*. ACM, New York, NY, USA, 478:1–478:19. https://doi.org/10.1145/3491102.3517431

[56] Ben Shneiderman. 1996. The eyes have it: A task by data type taxonomy for information visualizations. In *Proceedings of the IEEE Symposium on Visual Languages*. IEEE, Los Alamitos, CA, USA, 336–343.

[57] Alexa F. Siu, Danyang Fan, Gene S.-H. Kim, Hrishikesh V. Rao, Xavier Vazquez, Sile O'Modhrain, and Sean Follmer. 2021. COVID-19 highlights the issues facing blind and visually impaired people in accessing data on the web. In *Proceedings of the Web for All Conference*. ACM, New York, NY, USA, 11:1–11:15. https://doi.org/10.1145/3430263.3452432

[58] Alexa F. Siu, Gene S.-H. Kim, Sile O'Modhrain, and Sean Follmer. 2022. Supporting Accessible Data Visualization Through Audio Data Narratives. In *Proceedings of the ACM Conference on Human Factors in Computing Systems*. ACM, New York, NY, USA, 476:1–476:19. https://doi.org/10.1145/3491102.3517678

[59] Noushad Sojib, Syed Masum Billah, and Mohammad Ruhul Amin. 2020. Towards Multi-Wheel Input Device for Non-Visual Interaction. In *Adjunct Proceedings of the ACM Symposium on User Interface Software and Technology*. ACM, New York, NY, USA, 7–9. https://doi.org/10.1145/3379350.3416168

[60] Spotify for Artists. 2022. Loudness normalization. https://artists.spotify.com/help/article/loudness-normalization

[61] Christian J. Steinmetz and Joshua D. Reiss. 2021. pyloudnorm: A simple yet flexible loudness meter in Python. *Journal of the Audio Engineering Society* 150, Article 10483 (2021), 8 pages.

[62] Faisal Taher, John Hardy, Abhijit Karnik, Christian Weichel, Yvonne Jansen, Kasper Hornbæk, and Jason Alexander. 2015. Exploring interactions with physically dynamic bar charts. In *Proceedings of the ACM Conference on Human Factors in Computing Systems*. ACM, New York, NY, USA, 3237–3246. https://doi.org/10.1145/2702123.2702604

[63] Faisal Taher, Yvonne Jansen, Jonathan Woodruff, John Hardy, Kasper Hornbæk, and Jason Alexander. 2016. Investigating the use of a dynamic physical bar chart for data exploration and presentation. *IEEE Transactions on Visualization and Computer Graphics* 23, 1 (2016), 451–460. https://doi.org/10.1109/TVCG.2016.2598498

[64] PP Vaidyanathan. 2001. Generalizations of the sampling theorem: Seven decades after Nyquist. *IEEE Transactions on Circuits and Systems I: Fundamental Theory and Applications* 48, 9 (2001), 1094–1109.

[65] Petra Vetter, Łukasz Bola, Lior Reich, Matthew Bennett, Lars Muckli, and Amir Amedi. 2020. Decoding natural sounds in early "visual" cortex of congenitally blind individuals. *Current Biology* 30, 15 (2020), 3039–3044. https://doi.org/10.1016/j.cub.2020.05.071

[66] Bruce Walker and Joshua Cothran. 2003. Sonification Sandbox: A graphical toolkit for auditory graphs. In *Proceedings of the International Conference on Auditory Display*. International Community For Auditory Display, Atlanta, GA, USA, 3. http://hdl.handle.net/1853/50490

[67] Ruobin Wang, Crescentia Jung, and Yea-Seul Kim. 2022. Seeing Through Sounds: Mapping Auditory Dimensions to Data and Charts for People with Visual Impairments. *Computer Graphics Forum* 41, 3 (2022), 71–83. https://doi.org/10.1111/cgf.14523

[68] Bosun Xie. 2013. *Head-related Transfer Function and Virtual Auditory Display*. J. Ross Publishing, Ft. Lauderdale, FL, USA.

[69] Cheng Xu, Ali Israr, Ivan Poupyrev, Olivier Bau, and Chris Harrison. 2011. Tactile display for the visually impaired using TeslaTouch. In *Extended Abstracts on Human Factors in Computing Systems*. ACM, New York, NY, USA, 317–322. https://doi.org/10.1145/1979742.1979705

[70] Haixia Zhao, Catherine Plaisant, Ben Shneiderman, and Jonathan Lazar. 2008. Data Sonification for Users with Visual Impairment: A Case Study with Georeferenced Data. *ACM Transactions on Computer Human Interaction* 15, 1 (2008), 4:1–4:28. https://doi.org/10.1145/1352782.1352786

[71] Jonathan Zong, Crystal Lee, Alan Lundgard, JiWoong Jang, Daniel Hajas, and Arvind Satyanarayan. 2022. Rich Screen Reader Experiences for Accessible Data Visualization. *Computer Graphics Forum* 41, 3 (2022), 15–27. https://doi.org/10.1111/cgf.14519

[72] Eberhard Zwicker and Hugo Fastl. 2013. *Psychoacoustics: Facts and Models*. Springer, Berlin, Germany.