DOI: 10.1111/cgf.14523
Eurographics Conference on Visualization (EuroVis) 2022
R. Borgo, G. E. Marai, and T. Schreck
(Guest Editors)

Volume 41 (2022), Number 3

Seeing Through Sounds: Mapping Auditory Dimensions to Data and Charts for People with Visual Impairments

R. Wang[1], C. Jung[1] and Y. Kim[1]

[1]University of Wisconsin-Madison, United States

Abstract

Sonification can be an effective medium for people with visual impairments to understand data in visualizations. However, there are no universal design principles that apply to various charts that encode different data types. Towards generalizable principles, we conducted an exploratory experiment to assess how different auditory channels (e.g., pitch, volume) impact the data and visualization perception among people with visual impairments. In our experiment, participants evaluated the intuitiveness and accuracy of the mapping of auditory channels on different data and chart types. We found that participants rated pitch to be the most intuitive, while the number of tappings and the length of sounds yielded the most accurate perception in decoding data. We study how audio channels can intuitively represent different charts and demonstrate that data-level perception might not directly transfer to chart-level perception as participants reflect on visual aspects of the charts while listening to audio. We conclude by how future experiments can be designed to establish a robust ranking for creating audio charts.

CCS Concepts

• ***Human-centered computing*** → ***Empirical studies in visualization; Empirical studies in accessibility;***

1. Introduction

Visualizations have become a major medium to communicate data. However, visually impaired people (VIP) who cannot fully leverage their vision have been excluded from getting the benefit that visualizations can offer. This population includes people who are blind and have low vision, which is around 285 million people in the world according to the World Health Organization [Wor10]. Given the widespread use of visualizations, providing equal access to their contents regardless of the user's ability is imperative. Serving a broader audience has drawn the attention of the visualization community (e.g., [LCI*20]), promoting research efforts to design alternative modalities to communicate visualization. Sonification is an especially appealing alternative, given its practical nature. Unlike other modalities, such as tactile graphs, sonification can be implemented on most devices without specialized hardware.

To this end, extensive research has been carried out on auditory perception, including a few works from the visualization community (e.g., [RHEJ16]). For example, prior work has studied the optimal range of pitch for sonification [FM33] and the optimal slope of pitch increments to perceptually match one unit increment (i.e., power-law exponent) [WM10]. Based on these empirical studies, various software has been developed to create audio charts (e.g., [Pie19, WC03]). However, the penetration rate of audio charts in the wild remains low due to several reasons, including the low accuracy and individual variances in perceiving audio signals and biases of creators who design audio charts [Neu19]. Moreover, one prominent reason is the lack of a consolidated framework that can prescribe designers how to create an audio chart given data or convert a visualization into an audio chart. In contrast, when creating visual charts, designers and automated systems can consult with established frameworks, such as Mackinlay's effective rankings [Mac86], to get guidance on the appropriate mappings between visual channels and given data, enabling an effective and efficient creation process.

The guiding objective of this work is to create a ranking of audio channels to help create an audio chart, specifically for VIP. Toward this vision, we investigated how VIP perceive various auditory channels to represent different data and chart types. In designing visualizations, effectiveness (approximated by accuracy) is often the primary criteria to determine which encodings should be used. While effectiveness is an essential component, we believe that lowering the burden of constructing a mental model of the data and the visual structure of charts is equally important to serving people with visual impairments. Therefore, we investigated the perceived intuitiveness of each mapping. Through a controlled online study conducted via Zoom, we evaluated the intuitiveness and accuracy of five auditory channels (pitch, volume, length, tapping, timbre) to represent three data types (quantitative, ordinal, and nominal). We also studied whether these intuitiveness mappings transfer to chart-level perception. We derive suggestions for future studies to establish a robust auditory channel ranking.

Our study revealed that participants considered pitch as the most

Dimensions / Channels			Note
Temporal Dimension			Often used to convey the axis. Equivalent to position encoding.
Auditory Dimension	Speech Channel		Often used to convey ticks, labels and legend.
	Non-speech Auditory Channel	Pitch	The way in which a frequency of a sound wave is perceived.
		Volume	The strength or intensity of a sound.
		Panning	Positioning of sounds in the left or right spectrum of a stereo.
		Length	Representing playing a sound for a certain amount of time.
		Tapping	Representing repeating a short sound a certain number of times.
		Timbre	The quality or "texture" of the sounds.
		Modulation Index	Being composed of 2-seconds sounds using a modulation frequency, which uses a different frequency from the frequency used in the 2 secs,and a modulation index, which is the number of harmonics.
Display Dimension	Continuity	Continuous	Continuous sounds are used to represent data points across the axis.
		Discrete	Discrete sounds include beats of silence between sounds that represent data points across the axis.
	Tempo		The speed at which a sound is played.
Mapping Dimension	Polarity	Positive	The lower value of an auditory channel is mapped to the lower data value, and the higher value of an auditory channel is mapped to the higher data values.
		Negative	The lower value of an auditory channel is mapped to the higher data value, and the higher value of an auditory channel is mapped to the lower data values.

Table 1: *The design space of auditory charts.*

intuitive channel to encode the data, regardless of the data type. However, other channels such as tappings (the number of taps to encode data) and length (the length of sounds to encode data) were more accurate in decoding the underlying data. The intuitiveness rankings of quantitative and ordinal data were similar but slightly different from nominal data. We observed that visual metaphors such as line charts are "continuous" and scatter plots are "scattered," impacting perceptions when evaluating audio charts' intuitiveness. For example, participants find continuous sounds more intuitive in representing line charts, whereas discrete sound encodings were preferred when representing scatter plots. We find that even when charts encode the same sets of variables, participants voted for different auditory channels to be more intuitive to represent the charts due to their visual characteristics.

We make several contributions. First, we present the design space of audio charts by aggregating prior work to inform our study and future designs. We demonstrate the use of the design space in our experiment. Second, we report evidence that VIP find different auditory channels intuitive to represent different data types. In contrast to prior work, which mostly focuses on quantitative variables, our work extends the assessment to all three data types, providing more generalizable guidance in designing audio charts. We also report findings from what combinations of auditory channels best represent different chart types. We observed that the most intuitive mapping at the data level might not be aligned with the most intuitive mapping at the chart level due to their "visual" characteristics, implying that additional consideration is needed to design audio charts. Lastly, informed by our experiments, we present directions for future studies to establish a robust effectiveness ranking.

2. Related Work & Background

2.1. Visualization accessibility for VIP

Several approaches have been investigated to communicate visualizations using alternative modalities beyond vision. The approaches include sonification (e.g., [BBR*02, Bre02, SFH05]), tactile visualization [WM18, EW17a, EW18, EW17b, PR10, YMB*20, FM15, Hu15, EMW19], olfactory strategies [PBE19], and summarization through text [JMK*22, ESC*07, MSMC14, GTPG13]. Kim et al. conducted a survey regarding accessible visualizations and provided an exhaustive list of literature of the last decade [KJRK21].

2.2. Sonification

Sonification is the use of audio channels to convey data [KWB-fAD99]. While our focus is to support VIP, not every study investigated sonification to serve VIP. Indeed, there is evidence demonstrating the difference in perception between sighted people and VIP [CW10, WM10]. However, we include them in the following section as they provide full design space of sonification and potentially relevant empirical findings.

While it is not common, some prior work explored sonification in the context of visualizations, as opposed to the context of data. Various chart types have been explored with sonification, such as bar charts, line charts, and maps [Bre02, BBR*02, CJP*19, CW10, SJJJ19, YB02, ZPS05, CMS07, KWKH19, Tom16]. An early audio access method prototyped by Bulatov and Gardner emphasized that visually impaired users must be capable of building a mental image of where objects and data points are located [BG98]. Moreover, research conducted by Sakhardande et al. explored audio to represent bar charts and found that visually impaired users were faster using audio than speech on point estimation and point comparison tasks [SJJJ19]. This finding is similarly reflected in work by Brewster, which investigates line charts through various visualization tasks [Bre02]. Audio graphs resulted in better interaction than graphs with only speech, as well as a reduction in cognitive demand. These works highlight how the sonification of visualizations can contribute to the effective interpretation of visualizations and completion of visualization tasks. However, there is a lack of comprehensive exploration of a variety of auditory channels, the different polarity of mapping (i.e., positive polarity maps lower values in a channel to lower data values and higher values in a channel to higher data values; negative polarity maps the opposite), and different data types that may be used in visualizations.

The prior studies can be classified by auditory channels that they target to investigate. Pitch has been one of the most popular channels that have been extensively studied [WL01, DW07, Bre02, BBR*02, BB03, CMS07, FWGK01, PD08, KWKH19, NKW02a, SBS*02, Tom16, WK05, Wal02, Wal07, ZPSL08, NH05, NKW02b, PL05]. Pitch is how a frequency of a sound wave is perceived [Cha21a]. Previous studies have explored using positive and negative polarities with pitch [DW07, SBS*02, WK05, Wal02, Wal07]. Walker and Kramer [WK05] found that the sighted participants' preferred polarity depends on the mapping in consideration, as well as that it is difficult to predict the preferred polarity but can be determined empirically. Most of the studies explored only quantitative data types to examine the auditory perception [Bre02, WK05, Wal02, Wal07] with only a few exceptions [HHN11].

Volume is the strength or intensity of a sound [Sci21]. Volume has been explored to represent data, often used with pitch [FR03, NKW02a, FCBS06, NH05, PL05]. Neuhoff's work demonstrated that changes in volume could influence pitch change and vice versa [NKW02a]. Tapping represents repeating a short sound a certain number of times. Length represents playing a sound for a certain amount of time. Although there is a lack of work directly looking into tapping and length as auditory channels to encode data, tick marks were involved in several studies to measure axes or length [SW02, BNCM01, WC03, FR03, CW10, DW07, NW06].

Work by Smith and Walker showed that including tick marks in the form of clicks to provide x-axis information resulted in a greater proportion of correct answers than the condition that did not provide tick marks [SW02]. Panning is the positioning of sounds in the left or right spectrum of a stereo [pan21]. Panning has been investigated in various studies [DW07, WMG04, BB03, CW10]. For example, panning was used to distinguish between different sequences of data points [DW07, BB03] and to represent decimals [WMG04]. Timbre is a specific quality or texture of a particular sound [Cha21b]. For example, different instruments have different timbres. There have been a few works using the timbre in their charts [WC03, SB07, FR03, NH05, BB03]. For example, work by Brown and Brewster demonstrated a study using same-instrument and different-instrument timbres to represent one data set (a math function) in each ear. Participants were measured on their accuracy in drawing the auditory graph. Results showed high accuracy in participants' drawings with no effect between same-instrument and different-instrument timbres to represent data series [BB03].

Neuhoff et al. found that prior musical experience plays an essential role in skills such as mapping, scaling, and conceptual relationships in sonified visualizations. Moreover, perceiving the direction of pitch change is difficult for participants with little musical experience [NKW02a]. Sandor and Lane also reported that people with more musical training perform better in tasks related to pitch [SL03]. However, some experiments have shown there is no absolute relationship between musical experience and the ability to interpret encoded data messages [WK94].

Our work is different in various aspects. First, our study investigates auditory perception by different *data types* in the context of *visualization design*. All prior studies focus on the quantitative data type [Bre02, WK05, Wal02, Wal07]), which limits the contribution to designing real-world audio charts. Numerous works only look at the data aspect (e.g., [WK05, Wal02]) without charts involved, not providing insights on audio chart design. Second, our study extends the number of auditory channels to a broader variety: pitch, volume, tapping, length, and timbre. Previous studies have looked at a limited set of auditory channels at a time, hindering deriving some comparative conclusion among the channels (e.g., [CMS07, YB02]). Third, our study investigates the intuitiveness and effectiveness of mapping for VIP. As visually impaired individuals may not have any visual information, the intuitiveness of mappings can lower the cognitive burden of constructing a mental model for the visualization. Only a handful of works measure intuitiveness and specifically target users with visual impairments [WK05].

2.3. Design Space for Audio Chart

Based on the prior work, we derive the design space for our experiment and future audio chart design (Table 1). First, we created a Google Sheet where each column consisted of the metadata of the paper (e.g., type of paper, title, authors), the type of auditory signals (e.g., auditory channels), the characteristics of the signals (e.g., parallel, serial) and the experiment set-up (e.g., conditions, tasks) if the paper is an experimental paper. As we reviewed papers, we iterated on the columns, resulting in 19 different items. Three researchers clustered the audio signal-related columns iteratively, resulting in 14 items. We used the term "dimension" for higher-level clusters and "channel" for lower-level elements.

Some dimensions/channels are not possible to combine with others. For example, timbre can not be combined with polarity. Also, some auditory channels, such as tapping and length, can not be combined with consistency elements. In addition to the auditory channels and polarity we illustrated above from prior work, we considered a few more dimensions to cover the entire space. Temporal dimensions can be considered equivalent to the position encoding in the visual channel. While there is a lack of explicit characterization, prior work has considered temporal dimension when creating audio charts [SJJJ19]. Also, most of the available software maps one of the quantitative variables to temporal dimensions. For example, to represent a line chart, Sonification Sandbox would map one quantitative variable to the temporal dimension and one with an audio channel such as pitch [WC03]. Speech channel has also been used in a few works (e.g., [Bre02]). For example, in representing a bar chart, the system would read out the x-axis categories when the x-axis variable is mapped to the speech channel. We used this design space to build our experimental conditions.

3. Experiment: Intuitiveness & Accuracy of Mapping

We designed a within-subject experiment to evaluate how participants perceive different auditory channels (e.g., pitch, volume). The experiment consists of three parts. Part 1 investigates how intuitively different auditory channels can represent different data types (e.g., nominal, ordinal) and how accurately participants perceive the underlying data. Part 2 investigates whether the intuitiveness of the mapping in data-level perception transfers to chart-level perception. In other words, we study whether the channels that participants perceived as intuitive to encode a specific data type are also intuitive when they represent the data type in a chart. Part 3 examines their prior experience with audio charts and open-ended feedback for our experiments.

In this experiment, we consider intuitiveness and accuracy (i.e., effectiveness). While it is apparent for sighted people what types of data are presented in visualizations, intuitiveness in the mapping between a data type and auditory stimuli is crucial to communicate the data effectively for VIP. It is notoriously hard to define and measure "intuitiveness", but as a first step, we prompted participants with a simple question, "how intuitive the mapping is," using a five-point Likert scale. We adopted this methodology following a study that investigates the intuitiveness of different visual channels (e.g., fuzziness, color saturation) representing uncertainty [MRO*12]. Before the beginning of the session, we ensured that participants were in a quiet environment. We presented demo audio to test whether participants could hear the audio clearly from their devices.

3.1. Participants

Participants were recruited from mailing lists of organizations serving VIP. Our recruitment criteria were participants who are 1) at least 18-year-old, 2) legally blind, and 3) not experiencing hearing loss. We recruited a total of 20 participants. Participants' ages ranged from 20 to 38 (M=30.4, SD=5/6). Among the 20 participants, 16 participants were blind, and 4 had functional visions (all of those 4 participants have a visual acuity of 20/200 or less). The detailed demographic information is in the Supplemental Material. All study sessions were conducted via Zoom. The average length of a study session was around 1 hour (SD=5 mins). Participants were compensated with a $30 Amazon gift card.

3.2. Part 1: Evaluating Intuitiveness and Accuracy on Mapping Auditory Channels to Data

We investigated intuitive audio mappings for different data types. Since data abstraction is the starting point in the visualization design, we sought to understand how participants' perception differs for different data types. We expected to see the differences since quantitative and ordinal data have a notion of magnitude (i.e., small or large) while nominal variables do not. Barrass [Bar05] also conjecture that timbre might be the best auditory channel for mapping nominal variables. We formulated the study conditions by conjugating three factors: data type, auditory channel, and polarity of mapping.

3.2.1. Study Stimuli & Conditions

We formulated the study conditions based on three factors.

Data type: We varied the data types encoded by auditory channels.

- **Quantitative (Q):** We presented data as test scores for a class of students. The lowest score is 1 and the highest score is 100.
- **Ordinal (O)**: We presented data as five educational degrees: middle school, high school, bachelor's degree, master's degree,and PhD.
- **Nominal (N):** We presented data as five different fruits: apple, banana, peach, grape, and watermelon.

Auditory channel: We experimented with five auditory channels (pitch, volume, tapping, length, and timbre) to map to each data described above. We studied these five channels as they are commonly used in sonification software.

- **Pitch:** The frequency used for the minimum value was Ab #56 (207.65Hz), and the frequency used for the maximum value was F #113 (5587.65Hz). Humans tend to hear higher-frequency sounds louder than lower frequencies [FM33]. Therefore, we limited the pitch range to higher frequencies to control the perception differences.
- **Volume:** The minimum volume was 70%, and the maximum was 100%. We anticipate that participants will not always wear earphones. Thus, the minimum volume of 70% was the lowest sound that allowed participants to hear clearly, as tested in our informal pilot study.
- **Tapping:** Each data point was represented by a variable-length sound featuring multiple taps, where each tap represents a unit of the data. The length of each tap was set after the pilot study to 0.5 seconds.
- **Length:** Each data point was encoded through the length of a continuous sound. We set the length proportional to the data value with a proportionality constant of 0.5 seconds for consistency with the tapping condition.
- **Timbre:** Timbre is the quality or texture of sounds [Cha21b]. We used three distinctive tone qualities (string ensembles, violins, and pianos) to encode different data values.

While manipulating one channel, we kept all remaining channels constant. The default setup used 100% volume, C3 pitch, and a string ensemble instrument sound. These are either the defaults of an established sonification software [WC03] used to generate our study stimuli or were determined via the informal pilot.

Polarity of Mapping: We varied the polarity of auditory channels except for timbre since it does not have the notion of polarity.

- **Positive:** Positive polarity involves using lower values of a channel for lower data values and higher channel values for higher data values. For example, pitch-positive mapped lower pitch to lower values and higher pitch to higher values.
- **Negative:** Negative polarity involves using lower channel values for higher data values and higher channel values for lower data values.

We combined the five audio channels and the two polarities (except timbre), resulting in 9 auditory conditions (pitch-positive, pitch-negative, volume-positive, volume-negative, tapping-positive, tapping-negative, length-positive, length-negative, and timbre). These were then conjugated with the three data types (Quantitative, Ordinal, and Nominal), yielding 27 conditions (e.g., Q-pitch-positive, O-pitch-positive, N-pitch-positive...). All participants examined all conditions. Participants analyzed one data type at a time, where the order was randomized across participants. Within each data type block, the order of the nine auditory condition blocks was also randomized. To generate the stimuli, we used the Sonification Sandbox [WC03] developed by Walker and Cothran, which provides a mapping between data and various audio channels such as pitch, volume, and timbre. The Sonification Sandbox provides reliable default values which have been empirically validated as a result of research on people's auditory perception [WL01, SW02, Wal02]. Since the tapping condition was not supported, we repeated the default sound produced by the Sonification Sandbox with a gap in between.

3.2.2. Procedure

We first asked participants their demographic information, including gender, education level, occupation, and vision condition (e.g., functional vision, light sensitivity, onset age). We also asked about their musical experience to analyze potential correlations with their performance [NKW02a].

Part 1 aims at assessing the intuitiveness and accuracy of different audio encodings for different data types. To measure **intuitiveness**, participants first heard the minimum and maximum values represented by the assigned condition and were asked how intuitive the mapping was on a 5-point Likert scale. For example, in Q-pitch-positive condition, participants were told that the data ranges from 1 to 100, followed by the lowest pitch sound representing the minimum value, and the highest pitch sound representing the maximum value. To measure **accuracy**, the participant heard the minimum and maximum values first (e.g., for the Q-pitch-positive condition: "1 [the sound of Ab #56;], 100 [the sound of F #113].")), followed by three audio clips representing three data values randomly generated from the data set within the minimum and maximum range. After each audio clip, we asked participants to estimate the encoded value by comparing the minimum and maximum sounds. The participant was asked to estimate between 1 to 100 for quantitative data and from middle school to Ph.D. for ordinal data. Since nominal variables cannot be inferred from an audio clip without a legend, we did not measure accuracy for nominal data.

3.2.3. Data Preliminary

To analyze intuitiveness by the factor (data type, auditory channel, and polarity) and the condition, we built ordinal Logistic Regression models known to be suitable for analyzing Likert scale data [WK16]. To analyze the accuracy, we built mixed-effect models [SG91] to accommodate the fixed random effects in our study setup. To measure the accuracy, we calculated the absolute difference between the participants' three responses and the respective ground truth data, which were encoded in the audio clips using the Sonification Sandbox. The three absolute errors were averaged to compute the final metric. Our analysis did not deviate from the original plan based on the study design (e.g., factor level analysis, condition level analysis) and did not conduct any opportunistic analysis.

3.2.4. Results

Factor Level Analysis on Intuitiveness We used the self-rated intuitiveness as a dependent variable, and the data type, polarity, auditory channels and the interaction between polarity and auditory channels as independent variables. The result shows that the average values of self-rated intuitiveness were significantly varied by the polarity (χ^2=22.41, p<0.001) and auditory channel (χ^2=29.83, p<0.001). There was an interaction effect between auditory channels and the polarity (χ^2=12.74, p<0.05). The data type does not affect the average values of intuitiveness (χ^2=2.58, p=0.27, Mean of quantitative=3.2, SD of quantitative=1.2, Mean of ordinal=3.0, SD of ordinal=1.2, Mean of nominal=3.0, SD of nominal=1.2).

Effect of Polarity on Intuitiveness For the polarity analysis, we excluded the timbre as it does not have a notion of magnitude (i.e., no notion of higher/lower timbre). We also excluded the intuitiveness of audio encodings evaluated on nominal data for the same reason. We observed that polarity impacts encoding intuitiveness. Participants' self-rated intuitiveness was higher for auditory channels with positive polarity (i.e., larger data values map to higher channel values). The average rating of the positive polarity was 3.39 (SD=1.15), and that of the negative polarity was 2.76 (SD=1.09). The difference was statistically reliable (t=6.02, p<0.001).

The effect of Auditory channels on Intuitiveness The self-rated intuitiveness ratings by the auditory channel are shown in Fig. 1. Participants found the pitch to be the most intuitive to encode data, regardless of data type and polarity. Tapping had the second-highest self-rated intuitiveness, followed by length, timbre and volume. In our post hoc analysis, the intuitiveness rating of the pitch was reliably higher than tapping (t=-2.05, p<0.05) and all following channels. Tapping conditions were rated higher than length conditions (t=-2.82, p< 0.01) and all following channels. Length conditions and timbre conditions were (t=-0.79, p=0.4) not reliably different as well as with volume conditions (t=-1.65, p=0.98). The timbre conditions and the volume conditions were not significantly different (t=-0.50, p=0.62).

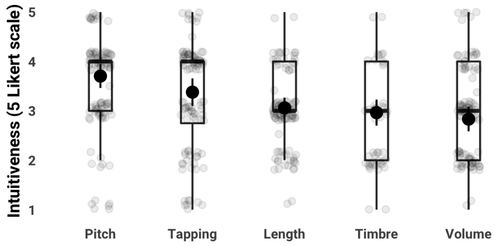


Figure 1: *Evaluating intuitiveness by auditory channels. Participants find pitch most intuitive to encode data, followed by tapping. The error bar indicates the standard error.*

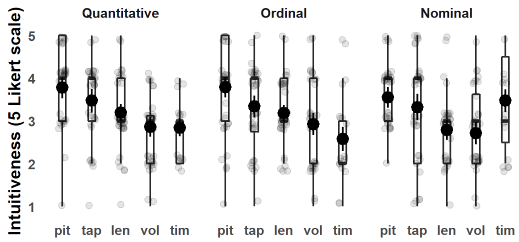


Figure 2: *Evaluating intuitiveness by auditory channel and data type. The aggregated participant's rankings of intuitiveness were similar when auditory channels were mapped to the quantitative and ordinal data but slightly different when those were mapped to the nominal data. The error bar indicates the standard error.*

The effect of Auditory Channel by Data Type on Intuitiveness As shown in Figure 2, the intuitiveness ranking was similar across data types with a few exceptions. Especially when auditory channels are mapped to quantitative and ordinal data, the trends are very similar. The main exception was timbre, which ranked second to last for quantitative and ordinal data types but second (third in terms of median value) for the nominal data type. We measured the correlation of rankings between the three data types using Spearman's ranking correlation coefficient. The intuitiveness rankings of the auditory channels of quantitative and ordinal were highly correlated (S=0, p<0.001), but the rankings of quantitative and nominal were not correlated (S=20.68, p=0.13).

We further break down the result by polarity (Fig. 3). The rankings between the two data types were similar (S=4, p<0.001). The intuitiveness ratings of pitch-positive conditions for both quantitative and ordinal data were reliably different from tapping-positive (Q: t=-3.71, p<0.001, O: t=-4.68, p<0.001) and all the following conditions. While tapping-positive conditions were more intuitive than tapping-negative conditions (t=3.21, p<0.01) and all the following conditions for encoding the quantitative data, tapping-positive conditions were similar to length-positive conditions (t=-1.35, p=0.18). All the combinatorial analysis is in the Supplemental Material.

For mapping the nominal data, the equivalent result is presented in Figure 2 (nominal) since polarity does not apply to nominal data.

Factor Level Analysis on Accuracy We now analyze what fac-

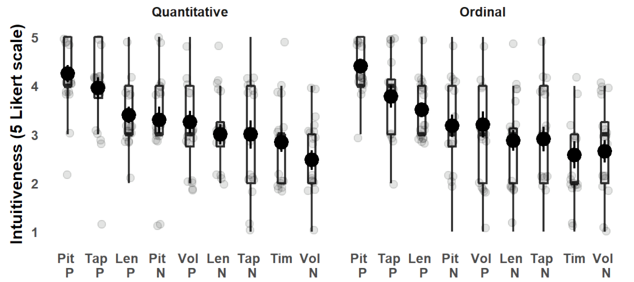
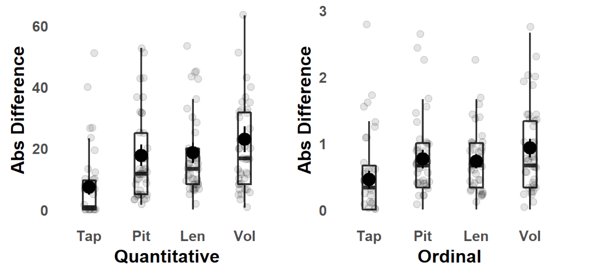
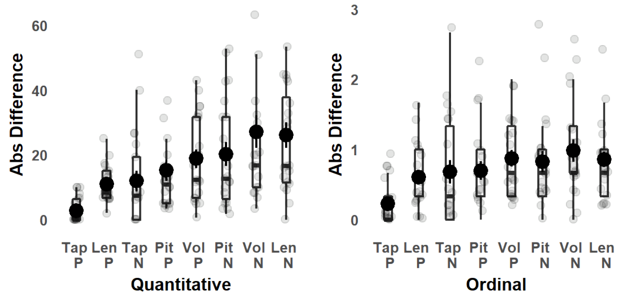


Figure 3: *Evaluating intuitiveness by auditory channel and polarity. The rankings of the quantitative data and ordinal were similar. Refer to figure 2 for the ranking of nominal data. The error bar indicates the standard error. P=Positive, N=negative*



(a) (b)

Figure 4: *Evaluating participants' accuracy of perceiving the data by auditory channel. The error bar indicates the standard error.*



(a) (b)

Figure 5: *Evaluating participants' accuracy of perceiving the data by auditory channels and polarity. The error bar indicates the standard error. P=Positive, N=negative*

tors impact the accuracy of participants' perceptions. The accuracy is computed as the average accuracy of three data points that we prompted participants to decode. The accuracy was calculated by taking absolute values of the difference between participants' responses and the ground truth. Ground-truth quantitative data ranged from 1 to 100, and ordinal data ranged from middle schools to Ph.D., which was mapped into a 1 to 5 range. We did not evaluate perception accuracy on nominal data since these cannot be predicted from audio signals.

For the analysis, we created a mixed-effect model using the lme4 package in R [BMBW15] for the quantitative and ordinal data conditions, respectively. We used the average absolute difference as a dependent variable, polarity, auditory channel, musical experience and the interaction between polarity and the auditory channel as fixed effects, and participants as a random effect. The result shows that the accuracy varied by polarity (χ^2=5.93, p<0.05) and auditory channel (χ^2=10.64, p<0.05). There was no effect of musical experience (χ^2=0.52, p=0.97) nor the interaction between auditory channels and the polarity (χ^2=1.43, p=0.84).

The Effect of Polarity on Accuracy We found that polarity impacts the accuracy of decoding quantitative data (t=3.33, p<0.001). Negative polarity conditions yielded an average error of 22.74 (SD=20.01) and positive polarities yielded an average error of 14.07 (SD=13.40) when the data is quantitative. Similarly, positive polarities (M=0.68, SD=0.60) yielded reliably lower decoding errors than negative polarities (M=0.91, SD=0.69) when perceiving ordinal data (t=2.54, p<0.05).

The effect of Auditory Channels on Accuracy Figure 4 shows the effect of the different auditory channels on the perception errors of quantitative and ordinal data. Tapping was the most accurate auditory channel to represent quantitative and ordinal data. In perceiving quantitative data, tapping conditions yielded significantly lower errors than pitch conditions (t=4.75, p<0.001) and all other channels. Similarly, for ordinal data, tapping conditions yielded reliably lower errors than pitch conditions (t=3.40, p<0.01) and all other channels. We found no difference between the pitch and length conditions for both quantitative (t=0.63, p=0.53) and ordinal data (t=0.17, p=0.86). We also found no reliable difference between length and volume (t=0.97, p=0.33) when encoding quantitative data. In the case of ordinal data, the length encoding was not reliably different from volume (t=1.21, p=0.23).

The Effect of Auditory Channel and Polarity on Accuracy Figure 5 shows the effect of the auditory channel and polarity for quantitative and ordinal data. The trends under quantitative and ordinal data conditions were similar (S=8, p<0.001). Tapping-positive conditions appeared to be the most accurate in perceiving quantitative and ordinal data. The tapping-positive encoding was reliably superior to the second-best encoding length-positive (Q: t=3.19, p<0.01, O: t=3.19, p<0.01) and all following conditions. In representing quantitative data, we found no differences between length-positive and tapping-negative (t=-1.33, p=0.18), pitch-positive (t=0.30, p=0.76), pitch-negative (t=1.25, p=0.21) or volume-positive (t=1.38, p=0.17). For ordinal data, the length-positive encoding was not reliably superior to tapping-negative (t=-0.49, p=0.63) and all other encodings. The full analysis is in Supplemental Material.

Difference Between Early-onset vs. Late-onset, Blind vs. Low-vision Individuals We did not observe differences in perceived intuitiveness or accuracy between participants who had early-onset and who had late-onset (we use age 16 as a threshold to define early-onset [LLQ*12]. Similarly, we did not observe differences in perceived intuitiveness or accuracy between blind or low vision participants. The analysis is in the Supplemental Material.

3.2.5. Summary

Participants considered pitch to be the most intuitive channel to encode the quantitative, ordinal, and nominal data. However, other channels such as tappings (i.e., the number of tappings to encode data) and length (the length of sounds to encode data) were more accurate in decoding the underlying data. The overall intuitiveness rankings of quantitative and ordinal data were structured similarly but slightly different from nominal data.

Dimension			Bar & Pie Chart (1N-1Q)				Line & Scatter Plot (1Q-1Q)			
Temporal Dimension			N	N	N	N	Q	Q	Q	Q
Auditory Dimension	Speech Channel									
	Non-speech Auditory Channel	Pitch	Q				Q		Q	
		Volume		Q				Q		Q
		Length			Q					
		Tapping				Q				
		Timbre								
Display Dimension	Continuity	Continuous					x	x		
		Discrete	x	x	x	x			x	x
Mapping Dimension	Polarity	Positive	x	x	x	x	x	x	x	x
		Negative								

Table 2: *Study condition we formulated with the aggregated design space (Q=quantitative data, N=nominal data, x=check mark).*

3.3. Part 2: Does Data-level Intuitiveness Transfer to Visualization Communication?

When designing visualizations, it is assumed that the data-level efficacy of visual channels can transfer to the chart-level. For example, according to Mackinlay ranking, the position is the most effective visual channel in visualizing quantitative data (data-level); therefore, scatter plots (chart-level) are a good choice to visualize quantitative data. In the sonification world, however, we hypothesize that the direct transfer of channel perceptions from data to visualization might not be possible, as listeners are interested in capturing both the underlying data and essential visual components of the chart. As a first step in validating our hypothesis, participants examined a sonified chart and were asked to assess the intuitiveness of audio mappings to represent various types of charts.

3.3.1. Pre-survey

Prior work demonstrates that VIP are generally familiar with many types of charts since they have experienced them via tactile display or embossed materials in school [JMK*22] They are most familiar with simple bar charts, pie charts, and line charts, and have significantly lower familiarity with stacked bar charts, scatter plots and area charts [EW17b,EW18]. To ensure participants' familiarity when evaluating data and chart mappings, we evaluated the familiarity of seven charts. We surveyed participants before the study to investigate their familiarity with the following chart types: area chart, bar chart, line chart, pie chart, scatter plot, donut chart, and violin plot. We asked two questions to measure their familiarity. First, we asked them to rate their familiarity on a 5-point Likert scale (e.g., How familiar are you with area charts? "Not familiar at all" to 'Extremely familiar"). Also, we asked whether they had seen or touched each chart (e.g., Have you seen or touched area charts through a tactile display or an embossed paper before? "Yes", "No", "I don't know what an area chart is"). We gave them a short description of each chart before asking these questions. We sent out a Qualtrics survey link through participants' emails, and responses were collected via Qualtrics before the study. With the result from the survey (refer to Sec. 3.3.4), we set out to study bar charts, pie charts, line charts, and scatter plots, as these were identified as the most familiar chart types in the preliminary survey.

3.3.2. Study Stimuli & Conditions

Table 2 summarizes all combinations of auditory channels used to represent each chart type. The design space helped us think through all the combinations available to create an audio chart for each chart type. Since the combinatorial design space is large, we limited the study conditions based on the feasibility (e.g., timbre cannot convey quantitative information without legend) and our design intuition (e.g., when presenting nominal data, sounds should be discrete).

Chart Type: We used four chart types based on the familiarity.

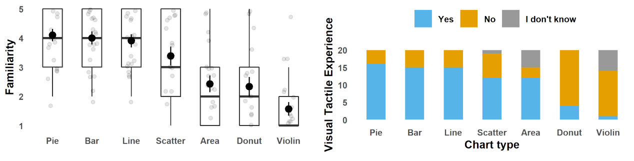
- **Bar Chart & Pie Chart (x-axis: nominal variable, y-axis: quantitative variable):** The sonification of bar charts and pie charts was created using either volume, pitch, length, and tapping to encode the y-axis and the temporal dimension with speech to encode the x-axis. For example, the audio chart will read out each category on the x-axis (e.g., watermelon) and play the corresponding quantitative data value with the assigned auditory channels (e.g., pitch). We only test discrete conditions for those two chart types based on the intuition that discrete sounds represent nominal variables.
- **Line Chart & Scatter Plot (x-axis: temporal dimension (without speech), y-axis: quantitative variable):** Line charts and scatter plots were represented using volume or pitch to encode the y-axis. They played continuously or discretely to evaluate whether the visual properties would influence their intuitiveness. Length and tapping were excluded as encodings for the y-axis based on our design intuition, where they would give listeners the impression that each data point was represented by a lengthy object, such as a bar.

Data Distribution (as a random variable): We varied the data distribution encoded in a chart. To create bar and pie chart datasets, we re-used the five nominal data values used in Part 1 for the x-axis and created corresponding quantitative data (y-axis) by considering two trends (Monotonic Trend, Bell Curve with Noise) as a starting point. For the line chart and scatter plot, we sampled 10 data points for the x-axis ranging from 0 to 100, similar to Part 1, and created the corresponding y-value by considering the same two trends.

- **Monotonic Trend:** The data points increase. We generated the data by adding random noise to the series [2, 4, 6, 8, 10] for bar and pie charts, and [1, 2, 3, 4, 5, 6, 7, 8, 9, 10] for line charts and scatter plots.
- **Bell Curve:** The data points follow a Bell curve with noise.

Bar charts and pie charts were created using one of the four audio channels (pitch, volume, length, or tapping) to represent the y-axis and the temporal dimension with speech to represent the x-axis, resulting in four study conditions each. Line charts and scatter plots were created using one of two audio channels (pitch or volume) played as continuous or discrete sounds to represent the y-axis and temporal dimension without speech to encode the x-axis, resulting in 4 conditions. All participants examined all conditions.

To generate the audio charts, we used the Sonification Sandbox [WC03]. Since this tool does not support nominal data, we used Kukarella [ttv] to convert text-to-speech to read out nominal categories in our data and used Apple GarageBand to concatenate the audio clips to create the stimuli for bar and pie charts. The visual-



(a) (b)

Figure 6: *(a) Results from preliminary survey measuring familiarity with different chart types using a 5-point Likert scale. (b) Results from preliminary survey gauging their visual and tactile experience with different chart types.*

ization of the data and chart types are shown in the Supplementary Material.

3.3.3. Procedure

We started by explaining what type of data was encoded for each chart. Participants listened to three audio charts representing the chart and were asked which audio chart best represented the chart type. Participants were allowed to say "none of them" to "all of them". To assess all conditions without overloading participants, we split the stimuli for each chart type into groups to contain at most three audio clips. The order of chart types was randomized. The order of conditions in each chart was randomized within each chart block. We also randomized the data distribution per chart type (i.e., either a data with monotonic trend or bell curve trend would be played, and participants receive the same dataset within the same distribution). During the entire session, we encouraged participants to share the rationales for their responses.

3.3.4. Results

Familiarity with Charts on 5-Point Likert Scale Figure 6 (a) shows the result. Participants' self-ratings reveal that they are most familiar with pie charts (M=4.1, SD=0.9), followed by bar charts (M=4.0, SD=1.0), line charts (M=3.9, SD=0.9), and scatter plots (M=3.4, SD=1.5). Participants were relatively unfamiliar with violin charts (M=1.6, SD=1.0), donut charts (M=2.3, SD=1.5), as well as area charts (M=2.4, SD=1.3).

Visual or Tactile Experience of Charts Figure 6 (b) shows the result. The majority of participants have seen or touched pie charts (16/20), bar charts (15/20), and line charts (15/20). About half of the participants experienced scatter plots (12/20) and area charts (12/20) visually or through tactile displays. Only a few participants had experience with donut charts (4/20) and violin charts (1/20).

Overall, the self-rated familiarity and the visual or tactile experiences were aligned with each other. Participants were primarily familiar with basic charts, especially pie charts, bar charts, line charts, and scatter plots. Based on the collection aggregated by Borkin et al. [BVB*13], circle charts, including pie charts, comprised only 3.2% of all visualizations. We conjecture that pie charts are more prevalent in education settings than the overall usage in the wild; therefore, they had more experience with pie charts.

Self-rated Intuitiveness Since participants were asked to choose the most intuitive audio chart encoding out of multiple choices, we aggregated participants' votes for each chart. We filtered the results

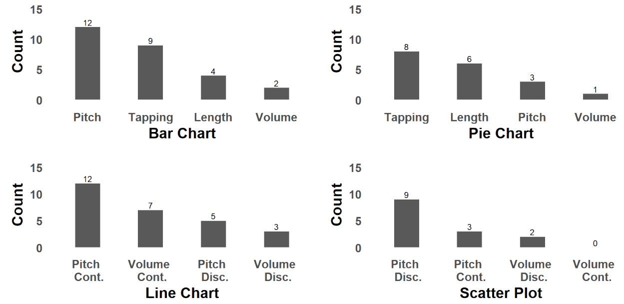


Figure 7: *Evaluating intuitiveness of auditory channels to represent different types of charts. The response from participants who haven't seen or touched the chart were excluded. Cont.=continuous, Disc.=discrete*

from participants who indicated that they had not seen or touched the specific charts (138 out of 424 data points were removed).

Figure 7 shows the votes that each condition received. The pitch obtained the most votes to represent bar charts, followed by tapping. Tapping had the highest number of votes, followed by length, representing pie charts. The pitch-continuous condition earned the most votes to represent line charts, followed by volume-continuous. Pitch-discrete encoding obtained the most votes when representing scatter plots, followed by pitch-continuous. Participants voted more often for continuous encodings to represent line charts and discrete encodings to represent scatter plots.

3.3.5. Summary

As we hypothesized, participants' responses demonstrated that they consider factors other than data type intuitiveness. While the pitch was the most intuitive channel to represent all data types, participants voted tapping and length the most to represent pie charts, which contain a quantitative variable. We suspect that participants may take into account the charts' "visual" look in judging the intuitiveness of audio channels to represent charts. The patterns became apparent when we compared line charts and scatter plots. Participants considered discrete sounds more intuitive when representing scatter plots, while continuous sounds were more intuitive when representing line charts.

3.4. Part 3: Post-Task Interview

After the main experiment, we followed up with post-task questions. We asked participants to share their prior experience with audio charts and experience with the experiments.

Prior Experience with Audio Charts 13 out of 20 participants had limited experience with audio charts. Prior experience included the use of the Texas Instrument graphing calculator, VoiceOver, Apple Health App, etc. P3 stated, "The only experience I had was when I was in college, I had the software called audio graphing calculator on my computer. So I used that a little bit when I was in college, but that's my only experience." P8 also had a limited experience with calculators for blind people: "I used ones that are built into like a Texas Instrument talking calculator. Okay, I don't know if you've seen it, and they made it for blind people, and it has like tones and stuff (for) the charts and graphs."

Designing Audio Charts beyond Mapping Participants liked the idea of playing audio for the minimum and maximum in the beginning to set the scale of the data range. P4 further suggested playing the entire scale to provide a better sense of it: "Go from a lower scale to a higher scale, to give a representation of that data before everything." P3 emphasized the importance of providing the overviews of the mapping: "Make it very clear what sounds are being played and what sounds are assigned to what kind of data point." Several participants recommended speeding up the audio chart. The audio should "move fairly quick" and should be limited "between 10-15 seconds," as suggested by P15. P13 asserted that "We are pretty accustomed to not having things really slow, so speed the sounds up would probably help to get the accuracy without going crazy."

4. Discussion

Our results showed that participants found auditory channels mapped with positive polarity to be more intuitive and enable more accurate data decoding. Pitch is the most intuitive to represent any data type in terms of individual channels. Given the continuous nature of pitch encodings, we found it surprising that pitch was deemed intuitive for nominal variables. We conjecture that participants' preference for pitch encodings stems from their familiarity with this channel and that pitch enables easy distinction between two data points when compared side by side.

While rated as the most intuitive for encoding data, the pitch is not always accurate in conveying data compared to more discrete channels, such as tapping. Since effectiveness and intuitiveness do not go together in this case, further studies are needed to understand the trade-offs of using pitch and tapping channels to represent quantitative and ordinal data in audio charts.

4.1. Alignment/misalignment with Prior Work

Our finding regarding people's perception of the polarity of data encodings is aligned with prior work. Walker and Mauney [WM10] reported that more people find positive polarities to be better when size, temperature, pressure, and velocity data are encoded. Similarly, most participants found positive polarities more intuitive in our experiments for data types that implied the notion of magnitude (i.e., quantitative, ordinal). Data decoding accuracies were also reliably higher for encodings with positive polarity. Participants found the pitch to be the most intuitive regardless of data types. This result is partially aligned with the findings of Walker and Kramer [WK05]. They found that pitch is better for representing quantitative variables than other channels such as tempo. We partially confirmed the conjecture made by Barrass, who speculates that timbre is a good auditory channel for mapping nominal variables [Bar05], by observing that timbre was the second most intuitive channel to represent nominal data but one of the least intuitive for quantitative and ordinal data. While a prior work claims that musical experience can impact people's perception of pitch [NKW02a], the sheer amount of evidence indicates the opposite (e.g., [HHN11]). Our result also shows no pattern that people with musical experience performed better at accuracy tasks, including pitch perception. Our findings reveal that participants associate the tapping mapping with pie charts. Interestingly, prior work done by Franklin and Roberts [FR03] demonstrates that a "Morse code" style (similar to tappings) encoding has the highest accuracy in perceiving values in the pie chart out of the five designs experimented with.

4.2. Expressiveness Criteria

While effectiveness criteria are important to consider in designing data representations, expressiveness criteria also play a role in the mapping process [Mac86]. We would like to start an initial discussion around the expressiveness criteria when mapping audio signals to data and charts. Since the discussions are based on the reflection of our results and intuition gained from conducting the study, further empirical validations are required. We can apply expressiveness criteria on two levels: 1) can an audio signal convey all the facts but only the facts that the data carries? 2) can an audio signal convey all the facts but only the facts that a chart visually carries?

Data-level Expressiveness: All the channels that have the notion of magnitude (e.g., pitch, volume) may express the quantitative and ordinal variables without violating the expressiveness criteria, as both carry the notion of "larger and smaller". However, the timbre, which does not intuitively map to values, may confuse listeners when it encodes quantitative or ordinal variables.

Chart-level Expressiveness: Discrete audio signals, regardless of the type of audio channels, can be mapped to visual marks that are not continuous (i.e., points). Chart types using line and area as marks may be better expressed by continuous signals, as the result of Part 2 alludes. In conjunction with data-level expressiveness, some visual encodings that often represent categorical variables (e.g., color, shape) may be more expressive when using an audio channel that doesn not have the notion of magnitude (e.g., timbre). Beyond element-level (e.g., marks and visual channels) consideration, expressiveness criteria can be applied to charts' overall "look". Such as what we showed in Study 2, the audio signals can convey visual appearances to some extent. For example, 3D audios may express the spatiality that pie charts carry better than other channels.

4.3. Suggested Directions for Future Experiments

While researchers and developers have worked on sonification extensively and created many libraries, the penetration rate of audio charts in the wild remains low. One potential reason is the lack of generalizable design guidelines that work with all data types and chart types. If a perceptual ranking of auditory channels is formalized and compatible with data and images models that visualization frameworks use, the creation and conversion space of audio charts can be drastically smaller, helping designers choose effective auditory encodings. Toward that vision, we list a few directions for future experiments that derived from our exploratory investigation.

Study Chart-Level Perception: One major takeaway of our study is that participants consider extra information beyond the data type when evaluating audio encodings to represent charts, unlike visual chart perception. While a bar chart and a pie chart may encode the same set of data (1N-1Q), participants found the pitch condition (followed by tapping) to be more intuitive to represent a bar chart, whereas the two most voted encoding channels for pie charts were tapping and length. Some participants alluded that they try to map acoustic characteristics to visual characteristics when assessing the intuitiveness of an audio mapping to represent a visual chart. For example, many participants explicitly highlighted the appearance of a pie chart and tried to map what they listened to, to the shape of the pie. P14 mentioned, “It’s kind of actively going around the circle, and maybe having another constant note as a guide that kind of takes you around it would be good." P3 also stated that “a bar chart is almost like blocks of lines and the pie chart is more of a circle segmented into pieces. They should be different.” Another example from our study is that participants find that pitch-continuous encodings provide the most intuitive audio chart to represent line charts. In contrast, pitch-discrete encodings are the most intuitive for scatter plots. P4 stated that “some progression of changes in the pitch makes me imagine a continuous line.” P1 also shared that “you want to know how scattered the points are, and where they’re located. I really like the idea of different notes because that really makes it seem like it is scattered."

Unlike visual charts, our study suggested that audio chart designs should consider effective and intuitive mappings for the chart type and take “visual” metaphors into account to convey the chart better. Future studies should explore chart-level effectiveness and intuitiveness to construct a universal ranking based on chart type instead of data type.

Narrow Down Condition Space: Using our design space (Table 1) and approach to formulate conditions (Table 2), the future study can expand the evaluation to other chart types and combinations of the channels. However, the condition spaces of possible encodings may be too large. Since it is evident that participants consider the visual look of the charts when they evaluate auditory mappings, the future experiment can narrow down the conditioned space by leveraging this fact. For example, researchers can only evaluate discrete dimensions if a chart uses a discrete mark (e.g., point), or only evaluate continuous dimensions if a chart uses a continuous mark (e.g., area, line).

In designing conditions for more complex chart types such as stacked bar charts or grouped bar charts, researchers can extract the visual metaphors from it (e.g., “stack” for a stacked bar chart, “juxtaposition” for grouped bar charts) and prioritize the auditory channels that align with these metaphors.

Using Tactile Stimuli or Extensive Description for Unfamiliar Chart Types: We observed that many participants were not familiar with some chart types, such as donut charts and violin charts. In future studies, researchers can use tactile stimuli to convey the visuals first to ensure participants’ familiarity with a particular chart type. Researchers can also prepare an extensive description based on the charts they are familiar with (e.g., Area charts are similar to line charts, but the area under the line is filled to encode data).

Study Speed as Design Factor It is known that people with visual impairments can perceive verbal information spoken in fast speed [MHD*08]. Our study set up the speed of tapping and length conditions relatively low (i.e., 0.5 seconds for one data unit) to ensure clear understanding. However, P13 asserted that the speed can be increased “We are pretty accustomed to not having things slow, so speed the sounds up would probably help get the accuracy without going crazy.” A future study can vary the speed to learn the maximum threshold speed to warrant accurate perception.

4.4. Limitations

We did not require people to wear headphones or earphones in our setting. As a result, 7 out of 20 participants wore headphones with stereo sound support, and one participant used a laptop with stereo sound outputs. While it can be minimal, this may impact participants’ responses during our study.

We tested the effect of data types with one dataset per type as an initial experiment. We envision that this set-up can be scaled up in various directions: 1) varying ranges of the datasets (e.g., for quantitative variable datasets, does the change of order of magnitude change the perception?), 2) varying topic of datasets [WM10], 3) varying the number of points in datasets (e.g., does complexity change the perception?), and 4) varying the combination of encodings (e.g., Are there interactions effects?). We believe that the different setup may not modulate the results of Study 1 (intuitiveness) since participants mostly relied on the notion of each audio channel to judge intuitiveness. However, the accuracy results may be changed as the dataset space changes.

We did not measure the accuracy of nominal data since audio cannot provide any signal to predict the ground truth. However, an open question for future research is how each audio channel influences the recall of nominal values. A legend must be provided for a user to decode nominal data in an audio chart. Future work could examine the effectiveness of audio channels by playing the legend first, then evaluate how well participants can recall the value later to gauge how effectively different audio channels support nominal data comprehension.

We evaluated the perceived intuitiveness, inspired by a method suggested by prior work [MRO*12, YCS*21]. However, other methods might provide informative validation to our findings [MJUO17, RBP*15].

5. Conclusion

We aggregated prior work on sonification to formulate a design space of audio charts to inform our study and future designs. We conducted an experiment with 20 visually impaired people to investigate the intuitiveness and effectiveness of auditory channels to represent data and their intuitiveness to represent charts. We found evidence that the data type can impact the intuitiveness toward auditory channels through the experiment. The study findings also suggested that the visual aspect of charts impacts participants’ preferences towards different auditory channels. We concluded by how future experiments can be conducted to construct generalizable guidelines.

References

[Bar05] BARRASS S.: A Perceptual Framework for the Auditory Display of Scientific Data. TAP 2 (10 2005), 389–402. doi:10.1145/1101530.1101532. iv, ix

[BB03] Brown L. M., Brewster S. A.: Drawing by Ear: Interpreting Sonified Line Graphs. In Proceedings of the 9th International Conference on Auditory Display (Boston, MA, USA, 2003). ii, iii

[BBR*02] Brown L. M., Brewster S. A., Ramloll R., Yu W., Riedel B.: Browsing Modes For Exploring Sonified Line Graphs. In Proceedings of 16th British HCI Conference (London, UK, 2002), vol. 2, pp. 6–9. ii

[BG98] Bulatov V., Gardner J.: Visualization by People without Vision. In Content Visualization and Intermedia Representations (CVIR'98) (1998). ii

[BMBW15] Bates D., Mächler M., Bolker B., Walker S.: Fitting Linear Mixed-Effects Models Using lme4. Journal of Statistical Software 67, 1 (2015), 1–48. URL: https://www.jstatsoft.org/index.php/jss/article/view/v067i01, doi:10.18637/jss.v067.i01. vi

[BNCM01] Bonebright T. L., Nees M. A., Connerley T. T., McCain G. R.: Testing the Effectiveness of Sonified Graphs for Education: A Programmatic Research Project. In Proceedings of the 7th International Conference on Auditory Display (Espoo, Finland, 2001), Georgia Institute of Technology. ii

[Bre02] Brewster S. A.: Visualization Tools for Blind People Using Multiple Modalities. Disability and Rehabilitation 24, 11-12 (Jan. 2002), 613–621. doi:10.1080/09638280110111388. ii, iii

[BVB*13] Borkin M. A., Vo A. A., Bylinskii Z., Isola P., Sunkavalli S., Oliva A., Pfister H.: What Makes a Visualization Memorable? IEEE Transactions on Visualization and Computer Graphics 19, 12 (2013), 2306–2315. doi:10.1109/TVCG.2013.234. viii

[Cha21a] Chase S.: What Is Pitch In Music? |HelloMusicTheory, 2021. URL: https://hellomusictheory.com/learn/pitch/. ii

[Cha21b] Chase S.: What Is Timbre In Music? |HelloMusicTheory, 2021. URL: https://hellomusictheory.com/learn/timbre/. iii, iv

[CJP*19] Choi J., Jung S., Park D. G., Choo J., Elmqvist N.: Visualizing for the Non-Visual: Enabling the Visually Impaired to Use Visualization. Computer Graphics Forum 38, 3 (June 2019), 249–260. doi:10.1111/cgf.13686. ii

[CMS07] Cohen R. F., Meacham A., Skaff J.: Teaching Graphs to Visually Impaired Students Using an Active Auditory Interface. In Proceedings of the Thirty-Seventh SIGCSE Technical Symposium on Computer Science Education (2007), pp. 279–282. doi:10.1145/1121341.1121428. ii, iii

[CW10] Choi S. H., Walker B. N.: Digitizer Auditory Graph: Making Graphs Accessible to the Visually Impaired. In CHI '10 Extended Abstracts on Human Factors in Computing Systems (New York, NY, USA, 2010), CHI EA '10, Association for Computing Machinery, p. 3445–3450. URL: https://doi.org/10.1145/1753846.1753999, doi:10.1145/1753846.1753999. ii, iii

[DW07] Davison B. K., Walker B. N.: Sonification Sandbox Reconstruction: Software Standard for Auditory Graphs. In Proceedings of the 13th International Conference on Auditory Display (Montréal, Canada, 2007). ii, iii

[EMW19] Engel C., Müller E. F., Weber G.: SVGPlott: An Accessible Tool to Generate Highly Adaptable, Accessible Audio-tactile Charts for and from Blind and Visually Impaired People. In Proceedings of the 12th ACM International Conference on PErvasive Technologies Related to Assistive Environments (2019), pp. 186–195. URL: http://doi.acm.org/10.1145/3316782.3316793, doi:10.1145/3316782.3316793. ii

[ESC*07] Elzer S., Schwartz E., Carberry S., Chester D., Demir S., Wu P.: A Browser Extension for Providing Visually Impaired Users Access to the Content of Bar Charts on the Web. In Proceedings of the Third International Conference on Web Information Systems and Technologies - Volume 1: WEBIST (2007), INSTICC, SciTePress, pp. 59–66. doi:10.5220/0001274600590066. ii

[EW17a] Engel C., Weber G.: Analysis of Tactile Chart Design. In Proceedings of the 10th International Conference on PErvasive Technologies Related to Assistive Environments (New York, NY, USA, 2017), PETRA '17, Association for Computing Machinery, p. 197–200. URL: https://doi.org/10.1145/3056540.3064955, doi:10.1145/3056540.3064955. ii

[EW17b] Engel C., Weber G.: Improve the Accessibility of Tactile Charts. In 16th IFIP Conference on Human-Computer Interaction (INTERACT) (Bombay, India, Sept. 2017), Bernhaupt R., Dalvi G., Joshi A., Balkrishan D. K., O'Neill J., Winckler M., (Eds.), vol. LNCS-10513 of Human-Computer Interaction - INTERACT 2017, Springer International Publishing, pp. 187–195. Part 3: Assistive Technology for Blind Users. URL: https://hal.inria.fr/hal-01676157, doi:10.1007/978-3-319-67744-6_12. ii, vii

[EW18] Engel C., Weber G.: A User Study to Evaluate Tactile Charts with Blind and Visually Impaired People. In Computers Helping People with Special Needs (06 2018), Springer International Publishing, pp. 177–184. doi:10.1007/978-3-319-94274-2_24. ii, vii

[FCBS06] Ferguson S., Cabrera D., Beilharz K., Song H.: Using Psychoacoustical Models for Information Sonification. In Proceedings of the 12th International Conference on Auditory Display (London, UK, 2006), pp. 113–120. ii

[FM33] Fletcher H., Munson W. A.: Loudness, Its Definition, Measurement and Calculation. Bell System Technical Journal 12, 4 (1933), 377–430. URL: https://onlinelibrary.wiley.com/doi/abs/10.1002/j.1538-7305.1933.tb00403.x, arXiv:https://onlinelibrary.wiley.com/doi/pdf/10.1002/j.1538-7305.1933.tb00403.x, doi:10.1002/j.1538-7305.1933.tb00403.x. i, iv

[FM15] Fusco G., Morash V. S.: The Tactile Graphics Helper: Providing Audio Clarification for Tactile Graphics Using Machine Vision. In Proceedings of the 17th International ACM SIGACCESS Conference on Computers & Accessibility (New York, NY, USA, 2015), ASSETS '15, Association for Computing Machinery, p. 97–106. URL: https://doi.org/10.1145/2700648.2809868, doi:10.1145/2700648.2809868. ii

[FR03] Franklin K. M., Roberts J. C.: Pie Chart Sonification. In Proceedings on 7th International Conference on Information Visualization (2003), pp. 4–9. doi:10.1109/IV.2003.1217949. ii, iii, ix

[FWGK01] Flowers J. H., Whitwer L. E., Grafel D. C., Kotan C. A.: Sonification of Daily Weather Records: Issues of Perception, Attention and Memory in Design Choices. In Proceedings of the 7th International Conference on Auditory Display (Espoo, Finland, 2001), pp. 222–226. ii

[GTPG13] Granz H., Tuccar M., Purushe S., Grinstein G.: Implementing Disability Accommodations in a Widely Distributed Web Based Visualization and Analysis Platform - Weave. In International Conference on Universal Access in Human-Computer Interaction (2013), Springer Berlin Heidelberg, pp. 31–39. ii

[HHN11] Hermann T., Hunt A., Neuhoff J. G. (Eds.): The Sonification Handbook. Logos Publishing House, Berlin, Germany, 2011. URL: http://sonification.de/handbook. ii, ix

[Hu15] Hu M.: Exploring New Paradigms for Accessible 3D Printed Graphs. In Proceedings of the 17th International ACM SIGACCESS Conference on Computers & Accessibility (New York, NY, USA, 2015), ASSETS '15, Association for Computing Machinery, p. 365–366. URL: https://doi.org/10.1145/2700648.2811330, doi:10.1145/2700648.2811330. ii

[JMK*22] Jung C., Mehta S., Kulkarni A., Zhao Y., Kim Y.-S.: Communicating Visualizations without Visuals: Investigation of Visualization Alternative Text for People with Visual Impairments. IEEE Transactions on Visualization and Computer Graphics 28, 01 (2022), 1095–1105. doi:10.1109/TVCG.2021.3114846. ii, vii

[KJRK21] KIM N., JOYNER S., RIEGELHUTH A., KIM Y.: Accessible Visualization: Design Space, Opportunities, and Challenges. In Computer Graphics Forum (2021), vol. 40, Wiley Online Library, pp. 173–188. doi:10.1111/cgf.14298. ii

[KWBfAD99] KRAMER G., WALKER B., BARGAR R., FOR AUDITORY DISPLAY I. C.: Sonification Report: Status of the Field and Research Agenda. International Community for Auditory Display, 1999. URL: https://books.google.com/books?id=Kv7yNwAACAAJ. ii

[KWKH19] KIM Y.-S., WALLS L., KRAFFT P., HULLMAN J.: A Bayesian Cognition Approach to Improve Data Visualization. In Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (New York, NY, USA, May 2019), CHI '19, Association for Computing Machinery, p. 1–14. URL: https://doi.org/10.1145/3290605.3300912, doi:10.1145/3290605.3300912. ii

[LCI*20] LEE B., CHOE E. K., ISENBERG P., MARRIOTT K., STASKO J.: Reaching Broader Audiences With Data Visualization. IEEE Computer Graphics and Applications 40, 2 (2020), 82–90. doi:10.1109/MCG.2020.2968244. i

[LLQ*12] LI J., LIU Y., QIN W., JIANG J., QIU Z., XU J., YU C., JIANG T.: Age of Onset of Blindness Affects Brain Anatomical Networks Constructed Using Diffusion Tensor Tractography. Cerebral Cortex 23, 3 (02 2012), 542–551. URL: https://doi.org/10.1093/cercor/bhs034, arXiv:https://academic.oup.com/cercor/article-pdf/23/3/542/17306926/bhs034.pdf, doi:10.1093/cercor/bhs034. vi

[Mac86] MACKINLAY J.: Automating the Design of Graphical Presentations of Relational Information. ACM Transactions on Graphics 5, 2 (1986), 110–141. URL: https://doi.org/10.1145/22949.22950, doi:10.1145/22949.22950. i, ix

[MHD*08] MOOS A., HERTRICH I., DIETRICH S., TROUVAIN J., ACKERMANN H.: Perception of Ultra-Fast Speech by a Blind Listener-Does He Use His Visual System. In Proceedings of the International Speech Production Seminar, Strasbourg (2008), pp. 297–300. x

[MJUO17] MATSUZOE S., JIANG S., UEKI M., OKABAYASHI K.: Intuitive Visualization Method for Locating Off-Screen Objects Inspired by Motion Perception in Peripheral Vision. In Proceedings of the 8th Augmented Human International Conference (New York, NY, USA, 2017), AH '17, Association for Computing Machinery. URL: https://doi.org/10.1145/3041164.3041198, doi:10.1145/3041164.3041198. x

[MRO*12] MACEACHREN A. M., ROTH R. E., O'BRIEN J., LI B., SWINGLEY D., GAHEGAN M.: Visual Semiotics & Uncertainty Visualization: An Empirical Study. IEEE transactions on visualization and computer graphics 18, 12 (2012), 2496–2505. doi:10.1109/TVCG.2012.279. iii, x

[MSMC14] MORAES P., SINA G., MCCOY K., CARBERRY S.: Evaluating the Accessibility of Line Graphs Through Textual Summaries for Visually Impaired Users. In Proceedings of the 16th International ACM SIGACCESS Conference on Computers & Accessibility (New York, NY, USA, 2014), ASSETS '14, Association for Computing Machinery, p. 83–90. URL: https://doi.org/10.1145/2661334.2661368, doi:10.1145/2661334.2661368. ii

[Neu19] NEUHOFF J. G.: Is Sonification Doomed to Fail? In Proceedings of the 25th International Conference on Auditory Display (Newcastle-upon-Tyne, UK, 06 2019), pp. 327–330. doi:10.21785/icad2019.069. i

[NH05] NEUHOFF J. G., HELLER L. M.: One Small Step: Sound Sources and Events as the Basis for Auditory Graphs. In Proceedings of the 11th International Conference on Auditory Display (Limerick, Ireland, 2005), pp. 1–3. ii, iii

[NKW02a] NEUHOFF J. G., KNIGHT R., WAYAND J.: Pitch Change, Sonification, and Musical Expertise: Which Way is Up? In Proceedings of the 8th International Conference on Auditory Display (Kyoto, Japan, 2002), pp. 1–6. ii, iii, iv, ix

[NKW02b] NEUHOFF J. G., KRAMER G., WAYAND J.: Pitch and Loudness Interact in Auditory Displays: Can the Data Get Lost in the Map? Journal of Experimental Psychology: Applied 8, 1 (2002), 17–25. doi:10.1037//1076-898X.8.1.17. ii

[NW06] NEES M. A., WALKER B. N.: Relative Intensity of Auditory Context for Auditory Graph Design. In Proceedings of the 12th International Conference on Auditory Display (London, UK, 2006), pp. 95–98. ii

[pan21] Audio Panning for Music Producers - Skills & Techniques, 2021. URL: https://www.renegadeproducer.com/audio-panning.html. iii

[PBE19] PATNAIK B., BATCH A., ELMQVIST N.: Information Olfactation: Harnessing Scent to Convey Data. IEEE Transactions on Visualization and Computer Graphics 25, 1 (2019), 726–736. doi:10.1109/TVCG.2018.2865237. ii

[PD08] PEDERSEN T. H., DELTA: The Semantic Space of Sounds: Lexicon of Sound-Describing Words. DELTA, Hørsholm, 01 2008. ii

[Pie19] PIETRUCHA M.: Sonification of Spectroscopy Data. PhD thesis, WORCESTER POLYTECHNIC INSTITUTE, 2019. i

[PL05] PERES S. C., LANE D. M.: Auditory Graphs: The Effects of Redundant Dimensions and Divided Attention. In Proceedings of the 11th International Conference on Auditory Display (Limerick, Ireland, 2005), pp. 169–174. ii

[PR10] PANEELS S., ROBERTS J. C.: Review of Designs for Haptic Data Visualization. IEEE Transactions on Haptics 3, 2 (2010), 119–137. doi:10.1109/TOH.2009.44. ii

[RBP*15] RUGINSKI I. T., BOONE A. P., PADILLA L. M. K., HEGARTY M., THOMPSON W. B., HOUSE D. H., CREEM-REGEHR S. H.: Non-expert Interpretations of Hurricane Forecast Uncertainty Visualizations. Spatial Cognition & Computation 16 (2015), 154–172. doi:10.1080/13875868.2015.1137577. x

[RHEJ16] RÖNNBERG N., HALLSTRÖM G., ERLANDSSON T., JOHANSSON J.: Sonification Support for Information Visualization Dense Data Displays. Proceedings of IEEE VIS Infovis Posters (VIS2016). IEEE VIS, Baltimore, Maryland (2016), 1–2. i

[SB07] SONG H. J., BEILHARZ K.: Spatialization and Timbre for Effective Auditory Graphing. In Proceedings of the 8th WSEAS International Conference on Acoustics & Music: Theory & Applications (Stevens Point, Wisconsin, USA, 2007), AMTA'07, World Scientific and Engineering Academy and Society (WSEAS), p. 18–26. iii

[SBS*02] SMITH L. D., BEST L. A., STUBBS D. A., ARCHIBALD A. B., ROBERSON-NAY R.: Constructing Knowledge: The Role of Graphs and Tables in Hard and Soft Psychology. American Psychologist 57, 10 (2002). doi:10.1037/0003-066X.57.10.749. ii

[Sci21] SCIENCE WORLD: Exploring Pitch and Volume, 2021. URL: https://www.scienceworld.ca/resource/exploring-pitch-and-volume/. ii

[SFH05] STOCKMAN T., FRAUENBERGER C., HIND G.: Interactive Sonification of Spreadsheets. In Proceedings of the 11th International Conference on Auditory Display (Limerick, Ireland, 2005), pp. 134–139. ii

[SG91] SHEINER L. B., GRASELA T. H.: An introduction to mixed effect modeling: concepts, definitions, and justification. Journal of pharmacokinetics and biopharmaceutics 19, 3 (1991), S11–S24. v

[SJJJ19] SAKHARDANDE P., JOSHI A., JADHAV C., JOSHI M.: Comparing User Performance on Parallel-Tone, Parallel-Speech, Serial-Tone and Serial-Speech Auditory Graphs. In Human-Computer Interaction – INTERACT 2019, Lamas D., Loizides F., Nacke L., Petrie H., Winckler M., Zaphiris P., (Eds.), vol. 11746. Springer International Publishing, Cham, 2019, pp. 247–266. doi:10.1007/978-3-030-29381-9_16. ii, iii

[SL03] SANDOR A., LANE D.: Sonification of Absolute Values with Single and Multiple Dimensions. In Proceedings of the 9th International

Conference on Auditory Display (Boston, MA, USA, 08 2003), pp. 243–246. iii

[SW02] SMITH D. R., WALKER B. N.: Tick-marks, Axes, and Kabels: The Effects of Adding Context to Auditory Graphs. In Proceedings of the 8th International Conference on Auditory Display (Kyoto, Japan, 2002), pp. 1–6. ii, iii, iv

[Tom16] TOMLINSON S. M.: Perceptions of Accessibility and Usability by Blind or Visually Impaired Persons: A Pilot Study. In Proceedings of the Association for Information Science and Technology (2016), vol. 53, pp. 1–4. doi:10.1002/pra2.2016.14505301120. ii

[ttv] Text to Voice. https://www.kukarella.com/. Accessed: 2021-06-29. vii

[Wal02] WALKER B. N.: Magnitude Estimation of Conceptual Data Dimensions for Use in Sonification. Journal of Experimental Psychology: Applied 8, 4 (2002), 211–221. doi:10.1037/1076-898X.8.4.211. ii, iii, iv

[Wal07] WALKER B. N.: Consistency of Magnitude Estimations with Conceptual Data Dimensions Used for Sonification. Applied Cognitive Psychology 21, 5 (July 2007), 579–599. URL: http://doi.wiley.com/10.1002/acp.1291, doi:10.1002/acp.1291. ii, iii

[WC03] WALKER B. N., COTHRAN J. T.: Sonification Sandbox: A Graphical Toolkit for Auditory Graphs. In Proceedings of the 9th International Conference on Auditory Display (Boston, MA, USA, 2003), pp. 161–163. i, ii, iii, iv, vii

[WK94] WATSON C. S., KIDD G. R.: Factors in the Design of Effective Auditory Displays. In Proceedings of the 2nd International Conference on Auditory Display (Santa Fe, New Mexico, 11 1994), pp. 293–303. iii

[WK05] WALKER B. N., KRAMER G.: Mappings and Metaphors in Auditory Displays: An Experimental Assessment. ACM Trans. Appl. Percept. 2, 4 (oct 2005), 407–412. URL: https://doi.org/10.1145/1101530.1101534, doi:10.1145/1101530.1101534. ii, iii, ix

[WK16] WOBBROCK J. O., KAY M.: Nonparametric Statistics in Human–computer Interaction. In Modern statistical methods for HCI. Springer, 03 2016, pp. 135–170. doi:10.1007/978-3-319-26633-6_7. v

[WL01] WALKER B. N., LANE D. M.: Psychophysical Scaling of Sonification Mappings: A Comparision of Visually Impaired and Sighted Listeners. In Proceedings of the 7th International Conference on Auditory Display (Espoo, Finland, 2001), pp. 90–94. ii, iv

[WM10] WALKER B. N., MAUNEY L. M.: Universal Design of Auditory Graphs: A Comparison of Sonification Mappings for Visually Impaired and Sighted Listeners. ACM Transactions on Accessible Computing 2, 3 (mar 2010). URL: https://doi.org/10.1145/1714458.1714459, doi:10.1145/1714458.1714459. i, ii, ix, x

[WM18] WATANABE T., MIZUKAMI H.: Effectiveness of Tactile Scatter Plots: Comparison of Non-visual Data Representations. In International Conference on Computers Helping People with Special Needs (2018), Springer, pp. 628–635. ii

[WMG04] WALKER B. N., MAUNEY L. M., GODFREY J.: The Audio Abacus: Representing a Wide Range of Values with Accuracy and Precision. In Proceedings of the 10th International Conference on Auditory Display (Sydney, Australia, 2004). iii

[Wor10] WORLD HEALTH ORGANIZATION: Global Data on Visual Impairment. World Health Organization, 2010. URL: https://www.who.int/blindness/publications/globaldata/en/. i

[YB02] YU W., BREWSTER S.: Multimodal Virtual Reality versus Printed Medium in Visualization for Blind People. In Proceedings of the Fifth International ACM Conference on Assistive Technologies (New York, NY, USA, July 2002), Assets '02, Association for Computing Machinery, pp. 57–64. doi:10.1145/638249.638261. ii, iii

[YCS*21] YANG Y., CHEN Y., SONG H., TANG S., YANG J.: Depth Perception Enhancement in 2D/3D Vascular Image Fusion. In 2021 5th International Conference on Digital Signal Processing. Association for Computing Machinery, New York, NY, USA, 2021, p. 46–52. URL: https://doi.org/10.1145/3458380.3458389. x

[YMB*20] YANG Y., MARRIOTT K., BUTLER M., GONCU C., HOLLOWAY L.: Tactile Presentation of Network Data: Text, Matrix or Diagram? In Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems (New York, NY, USA, 2020), Association for Computing Machinery, p. 1–12. URL: https://doi.org/10.1145/3313831.3376367. ii

[ZPS05] ZHAO H., PLAISANT C., SHNEIDERMAN B.: "I Hear the Pattern": Interactive Sonification of Geographical Data Patterns. In CHI '05 Extended Abstracts on Human Factors in Computing Systems (New York, NY, USA, 2005), CHI EA '05, Association for Computing Machinery, p. 1905–1908. URL: https://doi.org/10.1145/1056808.1057052, doi:10.1145/1056808.1057052. ii

[ZPSL08] ZHAO H., PLAISANT C., SHNEIDERMAN B., LAZAR J.: Data Sonification for Users with Visual Impairment: A Case Study with Georeferenced Data. ACM Transactions on Computer-Human Interaction 15, 1 (May 2008), 4:1–4:28. doi:10.1145/1352782.1352786. ii