J Vis (2023) 26:937–956
https://doi.org/10.1007/s12650-023-00909-3

REGULAR PAPER

Sadia Rubab · Lingyun Yu · Junxiu Tang · Yingcai Wu

Exploring Effective Relationships Between Visual-Audio Channels in Data Visualization

Received: 28 April 2022 / Revised: 4 October 2022 / Accepted: 16 January 2023 / Published online: 10 April 2023

Abstract In recent years, there has been a growing trend towards taking advantage of audio–visual representations. Previous research has aimed at improving users' performance and engagement with these representations. The attainment of these benefits primarily depends on the effectiveness of audio–visual relationships used to represent the data. However, the visualization field yet lacks an empirical study that guides the effective relationships. Given the compatibility effect between visual and auditory channels, this research presents the effectiveness of four audio channels (timbre, pitch, loudness, and tempo) with six visual channels (spatial position, color, position, length, angle, and area). In six experiments, one per visual channel, we observed how each audio channel, when used with a visual channel, impacted users' ability to perform the differentiation or similarity task accurately. Each experiment provided the ranking of audio channels along a visual channel. Central to our experiments was the evaluation at two stages, and accordingly, we identified the effectiveness. Our results showed that timbre, with spatial position and color, aided in more accurate target identification than the three other audio channels. With position and length, pitch allowed a more accurate judgment of the magnitude of data than loudness and tempo but was less accurate than the other two channels along angle and area. Overall, our experiments showed that the choice of representation methods and tasks had impacted the effectiveness of audio channels.

Keywords Audio–visual correspondence · Categories identification · Magnitude estimation · Data visualization

1 Introduction

Krygier (Krygier 1994), decades ago, emphasized the importance of using audio channels (abstract sounds) with the visual channels to represent nominal and ordinal data in visualizations. Redundant channels *(audio+visual)* significantly improve users' performance, understanding, and engagement (Sawe et al. 2020; McCormack et al. 2018; Kim et al. 2022; Metatla et al. 2016; Ziemer and Schultheis 2018). Visualization research introduces systems with redundant channels. These systems could offer varied benefits

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s12650-023-00909-3.

S. Rubab · J. Tang · Y. Wu (✉)
State Key Lab of CAD &CG, Zhejiang University, Hangzhou, China
E-mail: ycwu@zju.edu.cn

L. Yu
Department of Computing, Xi'an Jiaotong-Liverpool University, Suzhou, China

including a unique identification in an occluded view (Roodaki et al. 2017), magnitude estimation (Du et al. 2018), ease and certainty in searching in a small area (Roodaki et al. 2017), reduced task errors (Rönnberg 2019; Kim et al. 2022; Janata and Childs 2004), and improved engagement (Su et al. 2021; Metatla et al. 2016). In the systems, audio–visual relationships such as loudness-position (Du et al. 2018), pitch-categories (Roodaki et al. 2017), and pitch-angle (Roodaki et al. 2017) were used.

Based on the following observations, we argue that the attainment of the aforementioned benefits primarily depend on the effectiveness of chosen audio–visual relationships. First, studies in cross-modal correspondence [e.g., (Evans and Treisman 2010; Adeli et al. 2014)] had reported the difference in users' perception of different relationships. For instance, users had considered pitch strongly compatible with position but less with size. Second, Hogan discussed the intrinsic nature of users relating a representation of audio data to a representation of visual data. They can relate changes in pitch with changes in a scatter plot or a bar chart (Hogan et al. 2017). Third, there is evidence [e.g., (Du et al. 2018; Ren et al. 2013; Brewster 2018; Schito and Fabrikant 2018)] that users have had experienced problems identifying patterns and categories due to inappropriate relationships.

Previous research shows that designers of audio–visual systems had focused on the strengths of auditory channels and the compatibility of audio representation with the visual (Roodaki et al. 2017; Du et al. 2018; Lee et al. 2021). They clearly stated the relationship between individual visual features such as angle (Roodaki et al. 2017), color, and depth (Lee et al. 2021) of the visual representation, and the audio channels (Roodaki et al. 2017; Du et al. 2018; Lee et al. 2021). However, as discussed earlier, the chosen relationship may not provide the required benefits. The research which observed parameter mapping in the visualization field suggests a clear relationship between data and sound (e.g., timbre for nominal and pitch/loudness/tempo for ordinal (Krygier 1994; Schito and Fabrikant 2018; Sawe et al. 2020)). Nevertheless, it provides very little guidance on effective relationships between visual and audio channels and does not present a comparison between various relationships.

The literature reports empirical studies in which a significant difference in various audio–visual relationships was found. The studies explored musical notations for the corpus of sounds based on various audio dimensions (Tsiros 2014) and musical parameters for generating animated visuals (Lipscomb and Kim 2004). The researchers also surveyed the mapping of the auditory dimensions to the physical dimensions that give the impression of movement (Dubus and Bresin 2013). Inspired by these works, we aimed to contribute to the knowledge about which and when mappings are successful or unsuccessful in data visualization. The guidance could be provided by a review of mappings suggested in the literature like timbre-color (Sun et al. 2018), pitch-position (Khulusi et al. 2020), loudness-position (Du et al. 2018), pitch-length (Hansen et al. 2019), and tempo-area (Khulusi et al. 2020). However, it cannot imply which mappings are more effective. Therefore, we conducted an empirical study to answer the question: Which audio–visual relationships are effective in data visualization?

The study provided accuracy measures, based on which we presented the effectiveness of relationships between visual and audio channels used for data representation. To that end, we selected the most effective visual channels used for presenting categories - spatial position and color (identity channels) and magnitude - position, length, angle, and area (magnitude channels) of data (Munzner 2014). We selected the audio channels: pitch, loudness, tempo, and timbre, representing the data magnitude or categories (Flowers 2005). We ran six experiments, one per visual channel. In each experiment, we tested which audio channel(s) with the visual channel results in higher accuracy of the performed task. In the first two experiments, we asked participants to differentiate between audio–visual pairs to find the effective audio channels with identity channels. The differentiation task can estimate how accurately targets are identified with a channel (Hermann et al. 2011; Wang et al. 2019b). In the last four experiments, to find effective audio channels with magnitude channels, for each of three audio channels (timbre excluded), we asked participants to judge the similarity between visual and audio representations of data. The similarity task can provide the accuracy with which users observe the similarity level between the patterns of two data representations (Hermann et al. 2011; Gogolou et al. 2019). In each experiment, we compared the results from different audio channels to assess their effectiveness along a visual channel.

The results from six experiments show that pitch was noticeable with all visual channels, however, with different significance levels. In the conditions where more visual attention was required, users perceived changes in pitch significantly better than the other audio channels. However, the greater size of a visual negatively affected the effectiveness of pitch. Collectively, in all experiments, we observed a significant difference between different audio–visual relationships. Furthermore, our study design helped us identifying the factors that could impact mappings.

Our contribution is threefold:

- A study that systematically explores the difference in relationships between six visual and four audio channels
- Identification of the factors that impact the effectiveness of audio–visual relationships, and the prospects of further refining the mapping options
- Accuracy-based ranking of audio channels with visual channels

2 Related work

2.1 Visual channels for data representation

Research on how to map the data attributes to perceived visual channels was widely studied several decades ago (e.g., (Cleveland and McGILL 1984; Mackinlay 1986)). Prior studies (e.g., (Mackinlay 1986; Cleveland and McGILL 1984; Heer and Bostock 2010)) reported various rankings of visual channels used for representing categorical, ordered, and quantitative attributes. Among these works, Heer and Bostock (Heer and Bostock 2010) presented a significant study in which a ranking based on how accurately people perceive the quantities behind different ways of encoding the data was proposed. Based on the findings of previous works, Munzner (Munzner 2014) proposed a comprehensive taxonomy, in which visual channels were grouped into two categories, i.e., identity channels for categorical attributes and magnitude channels for ordered/quantitative attributes. Overall, there are four identity and eight magnitude channels. This paper considered two identity channels, namely, spatial position and color, and four magnitude channels, namely, position, length, angle, and area. We chose these channels based on observations (e.g., (Spence 2011; Evans and Treisman 2010)) that more effective visual dimensions can better match the auditory dimensions.

2.2 Audio channels for data representation

Sound is a valuable source for presenting information, and both realistic (Kong et al. 2019; Jin et al. 2023; Wang et al. 2022b; Ghosh et al. 2018) and abstract sounds have been used in visualization. However, realistic sounds (speech or earcons) are unsuitable for data representation (Krygier 1994). Additionally, previous studies [e.g., (Rouben and Terveen 2007; Wersényi et al. 2015)] observed people preferred audio channels over speech. They feel understanding speech requires more concentration and negatively affects focus on visual display. Even novice users who did not have the familiarity with mapping between visual and audio significantly performed better with non-speech audio than with speech (Harada et al. 2011).

Prior studies present data mapped to abstract sounds (also called audio variables or audio channels), mainly pitch, tempo, loudness, and timbre (Ferguson and Brewster 2018; Zhao et al. 2022; Hansen et al. 2019). The significance of the audio channels can be traced back to the 1980 s when Willison used them to present iris data (Wilson 1982). Yeung suggested that the four channels may present at least nine or more dimensions of data (Yeung 1980). This data transformation to audio channels follows a pipeline similar to the information visualization pipeline (Daudé and Nigay 2003). The authors (Daudé and Nigay 2003) proposed that first transform each data attribute to an audio channel and then merge them to present multidimensional data.

In these channels, pitch, tempo, and loudness are suitable to estimate the data magnitude (trends/patterns) and timbre to present the categories (Flowers 2005). A timbre can have multiple harmonic, envelope, or roughness. However, in this work, we had not considered the dimensions within the timbre as explicit audio channels as they depend on the intramodal association. The envelope depends on the loudness, and roughness requires changes in pitch (Hermann et al. 2011; Brewster 2018). The study of the influence of audio channels' association on mappings was beyond the scope of this work for two reasons. First, consideration of all possible combinations of audio channels would have resulted in many combinations, which is not manageable in one study. Second, the simultaneous use of multiple audio channels while testing a mapping would make it difficult to justify which of the two or three audio channels had played a major role in the results achieved. This work investigated the mapping of visual and audio channels on a one-to-one basis.

2.3 Correspondence between sensory modalities

Previous research aimed to investigate audio–visual correspondence (Spence 2011, 2020) provided us sufficient motivation and guidance for studying the effective relationships in data visualization. For instance, empirical studies found that users had not become accustomed to every mapping even after sufficient training (Knoeferle et al. 2016; Metatla et al. 2016). Their performance was significantly different with different mappings. To know the effective mappings, developers can take guidance from empirical studies in experimental psychology (e.g., (Evans and Treisman 2010; Adeli et al. 2014; Knoeferle et al. 2016)) or music (e.g., (Tsiros 2014)). However, researchers emphasized studying mapping based on the visual stimuli from the target field (Sanabria et al. 2004; Spence 2007). The visual design significantly influences the combination of auditory and visual dimensions (Sanabria et al. 2004). The design and its complexity should be iteratively investigated to refine the mapping options (Spence 2007).

Furthermore, in empirical studies exploring correspondence, users performed tasks on audio–visual pairs that were congruent or incongruent (e.g., (Parise and Spence 2013; Knoeferle et al. 2016; Evans and Treisman 2010; Adeli et al. 2014)). The two types of pairs were used because correct discrimination between them reflects a strong crossmodal integration effect (Spence 2007). Apart from the design and stimuli type, Metatla et al (Metatla et al. 2016) demonstrated the impact of task selection on evaluating a mapping. In the study (Metatla et al. 2016), users were asked to take their time to learn five shape-pitch and size-pitch pairs. They then differentiated between the pairs. The result shows that the task accuracy was according to the correspondence of each mapping. Users performed better with shape-pitch than size-pitch. Two inferences based on this result are: First, pitch and size are suitable for presenting the magnitude of data and correspond to each other (Spence 2020); the correspondence had not worked when they were simultaneously used to identify categories. Thus, the task influenced the tested effectiveness of the mapping. Second, the audio pattern was the same in both tested mapping, yet the performance differs. Thus, rigorous memorization had not made the mapping suitable for identification (Metatla et al. 2016) or searching task (Knoeferle et al. 2016). Basically, identifying mapping between audio and visual channels requires an analytical-rule-based approach, which means that those mappings are proven effective that users perceive similar (Shenkar et al. 2005). Users' experience with effective mappings provides the transfer of knowledge. They can successfully apply mapping experience to a new set of systems (Shenkar et al. 2005).

2.4 Audio–visual channels for data representation

Prior research on audio–visual data representations used various mappings between audio and identity channels. However, in most cases the effectiveness is unclear. For instance, pitch was used with spatial position (Nesbitt and Barrass 2002; Franklin and Roberts 2003) and loudness with texture (Roodaki et al. 2017) to identify categories. The effectiveness of pitch-spatial position mapping is doubtful for several reasons. First, along with pitch, loudness and panning had also been altered (Nesbitt and Barrass 2002). Thus, merely users can relate pitch with the two identity channels, is unsure. Second, researchers had not investigated how users performed with pitch (Franklin and Roberts 2003). Third, a prior studies reported a weak correspondence between the pitch and identity channels (Spence 2011; Adeli et al. 2014). The effectiveness of loudness-texture mapping is questionable as loudness was based on the acceleration sound effect (Roodaki et al. 2017). Studies proposed correspondence of timbre with colors, shapes (Adeli et al. 2014; Sun et al. 2018), and spatial position (Papachristodoulou et al. 2015). Thus, we can regard the mapping between timbre and identity channels as putative; all others need evidence.

The mappings between audio and magnitude channels used in previous systems also need evidence of effectiveness. For instance, both pitch (Roodaki et al. 2017) and tempo (Franklin and Roberts 2003) had been used for estimating angles. In pitch-angle mapping, three states (less, exact, and more) of the angle paired with three significantly different pitches. The estimation of angles with a reasonable difference in pitches is doubtful, whereas, in tempo-angle mapping, tempo with a dash-dot structure was applied. Smith and Walker (Smith and Walker 2005) proposed that features like dash-dot or context cues, irrespective of the audio channel, help users in the exact magnitude estimation. Thus, they make it hard to estimate the accuracy with which users perceive changes in an audio channel. Both pitch (Smith and Walker 2005) and loudness (Du et al. 2018) had also been used for estimating position. In the case of pitch-position, mapping was based on context cues. While, for loudness-position, researchers had not found the mapping helpful. However, their conclusion was not based on empirical evidence.

3 Experiments

This paper extends the line of research in audio–visual correspondence by exploring the effective relationships between audio and visual channels in data visualization from two perspectives: i) identification of categories and ii) estimation of data magnitude. To that end, we conducted an empirical study comprising six independent experiments. Exp **1** and **2** were based on spatial position and color, respectively. Each experiment used the differentiation task to evaluate the effectiveness of each of the four audio channels with the visual channel in representing categories. Exp **3**, **4**, **5**, and **6** were based on position, length, angle, and area, respectively. The experiments used the similarity task to evaluate the effectiveness of the three audio channels (timbre excluded) with the visual channel in estimating the changes in the magnitude of data.

3.1 Participants and setup

We recruited 81 participants (male = 37, female = 44) through an online survey announcement for all experiments. All the participants were from the same university. They were either undergraduate or graduate students of different departments. They all had an idea of basic charts. Out of them, 20 participated in Exp **1**, **2**, **3**, and **4**, 26 in Exp **5**, and 23 in **6**. We set 20 as a target for the number of participants for an experiment (Demiralp et al. 2014a; Ren et al. 2013). The same group of participants completed an entire experiment, and some of them participated in more than one experiment. The participants' ages ranged from 20-38. All participants reported normal vision and hearing. They were compensated on an hourly basis.

All experiments were run separately, on different days and times, and in a lab environment. Lab environment eliminates the chances of factors that can create noise in the audio (Smith and Walker 2005; Turnage et al. 1996). We completed an entire experiment in one sitting to avoid the risk of change in participants' strategy (Hermann et al. 2011). Visual representations and audio icons were shown on a projection screen; audio stimuli were played on speakers.

The effectiveness of the visual channels in representing categories or magnitude is well known. So, in all experiments, we considered the audio channels as the primary experimental factor. We used Audacity[1] to develop the audio stimuli. It is an audio editing tool helpful in processing the features of audio. We synthesized pure sine wave tones for the stimuli of pitch, loudness, and tempo. In the tempo stimuli, we applied the music length concept, i.e., the variation in sound and silence duration (Walker 2007). Pitch having frequency 300Hz-10000Hz (i.e., within the human hearing range (20Hz-20,000Hz) (Werner et al. 2011)), loudness with amplitude change.03 to 1 (Range: 0 to 1, used in Audacity to set the loudness), and tempo with the duration of.3 s to 2 s were used to create the audio stimuli of the three audio channels. The stimuli of a set were based on one channel. For instance, in a set of pitch based stimuli, each stimulus had a different pitch but the same loudness, tempo, and timbre. Consideration of intramodal association, for instance, between timbre and pitch (Adeli et al. 2014), was beyond the scope of this work (cf. Sect. 2.2). Further description of audio channels is covered in Sects. 3.2.1 and 3.3.1.

3.2 Experiments 1 and 2

In both experiments, we iteratively investigated each audio–visual relationship (Tsiros 2014). Our investigations, were based on the two usages of the audio–visual relationships. First, the literature suggests that users can identify categories by differentiating stimuli based on visual and audio channels (e.g., (Ren et al. 2013)). Therefore, in stage X, we presented audio and visual stimuli simultaneously. Second, users can take audio assistance to differentiate between the visual categories (e.g., (Papachristodoulou et al. 2015)) or focus on a visual after listening to the audio (e.g., Ghosh et al. (2018)). Therefore, in stage Y, we presented only audios.

3.2.1 Stimuli and trial generation

Visual Stimuli: In Exp **1** and **2**, we developed five rectangular areas to present spatial position and color, respectively (Fig. 1A). The rectangular areas had been commonly used in perceptual studies with identity channels (e.g., Giovannangeli et al. 2022; Demiralp et al. (2014a, 2014b)). In the pilot study, we observed

[1] Par96 https://www.audacityteam.org/.

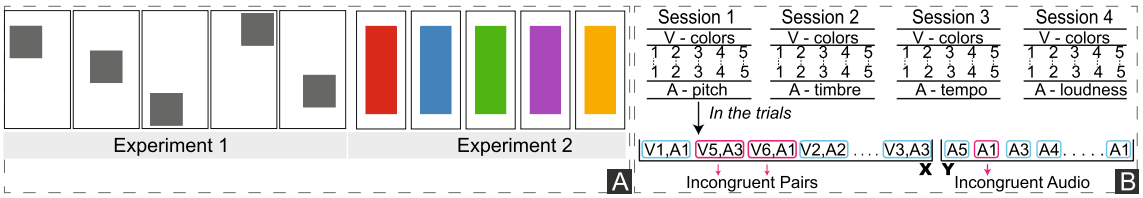

Fig. 1 This figure presents **A** Visual stimuli from experiments 1-Spatial position and 2-Color; **B** The pairing between visual (V) and audio (**A**) stimuli in the four sessions of experiment 2. Pairings in experiment 1 are the same as in experiment 2. A session's trials present the difference between the congruent and incongruent stimuli. Supplementary figure 1 and 2 also illustrates the pairing

that participants focused on both the color and position of the rectangles. Therefore, to avoid the intramodal association effect, in Exp **1**, we kept color constant and in Exp **2** spatial position.

Audio Stimuli: For both experiments, we developed four sets of auditory stimuli representing the four audio channels. In each set, we developed five considerably distinct stimuli. The difference in the stimuli of a set was based on one channel (cf. Sect. 3.1). The stimuli for pitch, loudness, and timbre were 1 s in length, and the tempo stimuli were 0.5 s to 2.5 s. In Exp **1**, we selected timbre based on musical instruments (Adeli et al. 2014). While in Exp **2**, timbres matched with the hue (like bird sound paired with green color, or water and thunder sound), as humans relate timbres with color (Adeli et al. 2014). We took only five stimuli due to the limitations of human auditory memory (Papachristodoulou et al. 2015; Harding et al. 2002).

Trial Generation: In both experiments, the five visual stimuli were paired on a one-to-one basis with the five audio stimuli in each of the four sets; to generate four audio–visual relationships (Fig. 1B). For each relationship, the testing trials were formed from those five pairs. One pair presented in a trial, and the pairs were repeated. The approach had been followed in the literature (Tsiros 2014; Evans and Treisman 2010; Papachristodoulou et al. 2015). In some trials, we applied incongruence. For the incongruence, in X, we either interchanged the audio stimulus or presented an unseen visual (V6 in Fig. 1B). In Y, we presented an audio other than those used in the five pairs.

3.2.2 Task

Participants performed the differentiation task, which can suggest the audio–visual relationship that represents easily separable categories (Ernst 2007; Ren et al. 2013; Giovannangeli et al. 2022; Wang et al. 2019b). Figure 2A presents the task. For each tested audio–visual relationship, participants first learned its five audio–visual pairs. Then, in X, they viewed randomly presented pairs. After each pair, their task was to check whether the audio and visual stimulus are from the same pair. Participants' answered yes and no for the congruent and incongruent pairs, respectively. In Y, they were required to mention the visual after listening to the audio. The task was based on objective assessment and provided the difference in relationships based on the accuracy of the results.

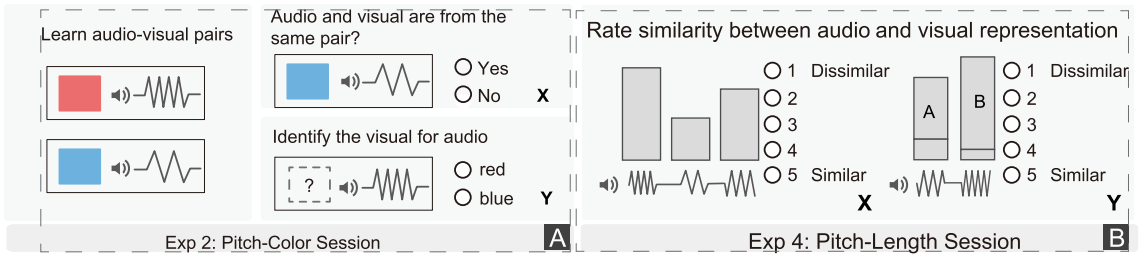

Fig. 2 This figure presents **A** The task used in experiment 2-Color: Participants differentiated between the audio–visual pairs, which they had learned. Here, we present only two pairs of the pitch-color relationship. We used the same task in experiment 1; **B** The task used in experiment 4-Length: Participants observed how well the audio representation matched the visual. In both stages of the session, participants rated the similarity. The same task was used in experiments 3-Position, 5-Angle, and 6-Area. **Note**. *Waveforms presented in this and all other figures are just an imaginary representation of audio stimuli. In all experiments, participants viewed visuals and listened to the audio. They did not see a waveform of audio stimuli. Additionally, the pitches, loudness, or tempo presented in all figures are not the actual ones*

3.2.3 Experimental design

In both experiments, we had run four sessions for the four tested relationships. We compared the results of sessions to find the difference in the effectiveness of the four audio channels with the visual channel. All sessions were comprised of two stages and shared the same procedure. Supplementary figure 3 illustrates the procedure.

1. Practice Phase. At the start of each session, participants went through the practice phase. The procedure for the practice phase was similar to the training & testing phase. Participants completed three practice trials in both X and Y.

2. Training. In every session, participants first learned the five pairs. We played the audio stimulus of each pair four times (suitable to memorize five crossmodal pairs (Knoeferle et al. 2016; Metatla et al. 2016)). Training is customarily performed before the differentiation task (Papachristodoulou et al. 2015; Ren et al. 2013; Ernst 2007; Metatla et al. 2016).

3. Testing Phase. During the testing phase, participants first finished all testing trials of X and then moved to Y. In each trial, participants first viewed/listened to the stimulus, and then with their handheld devices, supplied the response on an online form.

For the testing trails, each of the five pairs was repeated five times in X and three times in Y. We randomly presented the pairs. Thus, in total, 40 testing trials were completed by a participant. Approximately half of the trials of X and one-third of Y were based on incongruence. From 20 participants, in one session, 800 responses were collected.

4. Break. On completion of all testing trials, participants submitted the form. They had a 5-minute break before moving to the next session.

3.2.4 Hypothesis

In Exp **1** and **2**, we had considered the audio channels and the deviation (congruence/incongruence) in their stimuli as the independent variables and the accuracy in identifying stimuli as the dependent variable. The hypotheses for each experiment were:

H1 We expected a significant difference in the differentiation task accuracy with the different audio channels. We based our assumption on prior research, which found that humans better differentiated in tempo than pitch (Harding et al. 2002), and timbre was a suitable option for representing categories (Papachristodoulou et al. 2015).

H2 We expected the accuracy would remain consistent within an audio channel regardless of the deviation in the stimuli. A previous study had identified the impact of incongruence on performance (Papachristodoulou et al. 2015). Besides, Ernst reported a high differentiation accuracy for the incongruent pairs with well-perceived relationships (Ernst 2007).

3.3 Experiments 3 to 6

We based the two stages (X and Y) of each experiment on the following perspective of data exploration. The exploration of data magnitude involves an overview of the entire data (e.g., Tang et al. (2019); Du et al. (2018)) and the extraction of specific points (e.g., Roodaki et al. (2017); Heer and Bostock (2010); Hermann

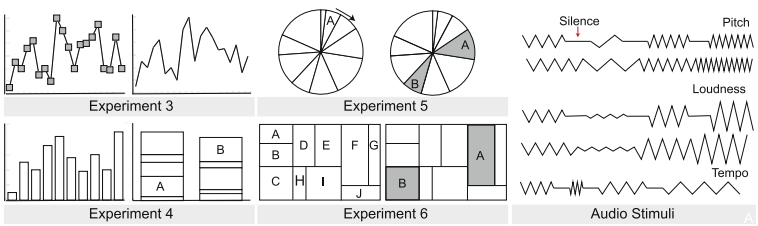

Fig. 3 This figure presents visual stimuli from the experiment 3-Position, 4-Length, 5-Angle, and 6-Area. From the two visual stimuli of each experiment, the left one is a representation from stage X and the right from stage Y. The figure also shows the waveforms based on the three audio channels. In the waveforms based on pitch and loudness, the second ones shows continuous audio streams (no silence), which we used with stage Y line charts (experiment 3)

et al. (2011)). The exploration entailing these tasks is achievable with the audio–visual representations (e.g., Du et al. (2018); Smith and Walker (2005); Roodaki et al. (2017)). These representations can represent datasets as a series of points (Smith and Walker 2005; Roodaki et al. 2017; Janata and Childs 2004) or continuous streams (Tang et al. 2019; Smith and Walker 2005; Du et al. 2018). Previous research (e.g., Hermann et al. (2011); Smith and Walker (2005)) showed the effect of data exploration and representation methods on the estimation of the magnitude. Therefore, we assumed they would also affect the similarity task. Thus, we considered them in designing the two stages of Exp **3**-**6**. The following section provides the details.

3.3.1 Stimuli and trial generation

Visual Stimuli: In Exp **3**-**6**, we developed graphical representations, like a bar or line chart (Fig. 3). Previous works (e.g., (Cleveland and McGILL 1984; Saket et al. 2019; Skau and Kosara 2016)) had used them in perceptual studies. Because they kept the identity of magnitude channels, and the evidence drawn from them could be reproduced with a different visualization.

In Exp **3**, following Smith and Walker's (Smith and Walker 2005) approach, we created line charts for the position channel. Line charts in X comprised of a series of points, and in Y, a continuous line. We took ten sets of random numbers to create five line-charts for X and five for Y. Each set was of 20 numbers, from 3 to 100 (Cleveland and McGILL 1984). Researchers proposed small datasets can provide sufficient evidence in support of the results (Harrison et al. 2014; Wang et al. 2022a; Ondov et al. 2019).

In Exp **4**, we created bar/stacked bar, in **5**, pie charts, and in **6**, treemaps, for length, angle, and area, respectively (Heer and Bostock 2010; Ondov et al. 2019; Harrison et al. 2014; Skau and Kosara 2016). In each experiment, X contained representations of the entire dataset, and in Y, representations with two selected points were used. In the Y of Exp **4**, we used stacked bars because the aligned scale in bar charts provides additional cues that make the comparison between lengths easy (Heer and Bostock 2010). Finding the similarity between a series of points is challenging with the bar charts (Ondov et al. 2019), so we created them for X. For Exp **6**, treemaps with the aspect ratio of 1:3/2 (Heer and Bostock 2010) were developed with the squarified treemap algorithm (Bruls et al. 2000). In each experiment, twenty sets - each of ten values (from 3 to 100), were used to create ten visual representations for X and ten for Y.

Audio Stimuli: In the last four experiments, timbre was excluded as considered categorical (Hermann et al. 2011)). In every experiment, based on each of the three audio channels (pitch, loudness, and tempo), we created three audio stimuli for each visual representation. In X of **3**, the stimuli based on pitch and loudness comprised.5 s sound for each of 20 data values, separated by.2 s silence. For both channels, the total duration for a stimulus was 13.8 s. In the tempo stimuli,.2 s silence was applied between notes. In Y of **3**, pitch and loudness stimuli were continuous streams of 20 s length, i.e., 1 s sound for every data value and no silence (Fig. 3). In X of Exp **4**-**6**, the duration of pitch and loudness stimuli was 6.8 s,.5 s sound for each of the 10 data values, separated by.2 s silence. In Y, they were 1.7 s (.75 s×2+.2 s silence).

Trial Generation: In the trials, we presented a visual representation and audio representation based on the same dataset. In each trial, the audio representation could match the visual at one of the three different similarity levels (SL) (Fig. 4A). A single visual representation can pair with the audio of different similarity levels in multiple trials (Fig. 4B). Supplementary figure 4 and 5 also illustrates the pairing. Multiple similarity levels provide a fair judgment of the similarity task (Turnage et al. 1996; Gogolou et al. 2019).

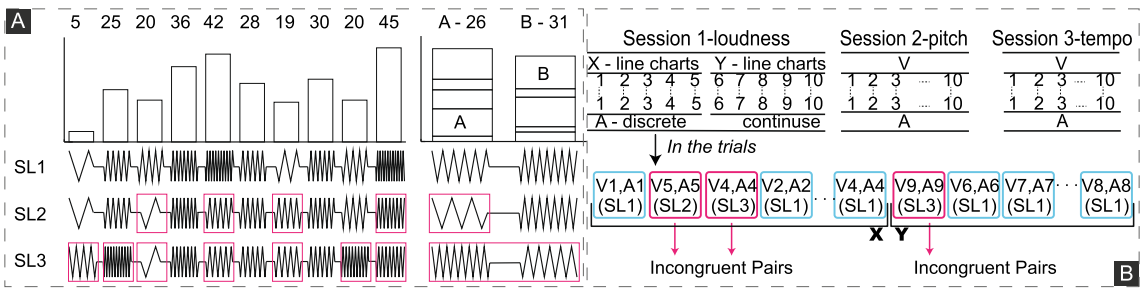

Fig. 4 This figure presents **A** The three similarity levels in pitch-based audio stimuli for both stage X (left) and Y (right). In a trial, the visual representation was paired with any of the three audio stimuli; **B** The pairing between visual and audio stimuli in the three sessions of experiment 3-Position. A session's trials present the difference between the congruent and incongruent stimuli. Pairings in experiments 4-Length, 5-Angle, and 6-Area were the same as experiment 3

SL1 Audio representation exactly matched the visual.
SL2 For the entire dataset representation, less than half of the audio representation was mismatched with the visual. The difference between the two audio values was varied by a maximum of 100% from the exact difference for marked slots.
SL3 For the entire dataset representation, more than half of the audio representation was mismatched with the visual. For marked slots, the two audio values were varied considerably from the correct difference or interchanged.

3.3.2 Task

Participants had performed the similarity task. The task suggests the audio–visual relationship, which can provide a correct unified perception of the data magnitude (Gogolou et al. 2019; Evans and Treisman 2010). Figure 2B presents the task. In every trial, participants had first viewed the visual and listened to its paired audio. They had then assigned a rating on a 5-point scale. Our participants made a point-to-point comparison—whether the visual and audio representation of each data value is similar or not. So, we expected a significant difference in the rating of different similarity levels. In these experiments, the accuracy that provided the difference between the audio channels was analogous to consistency in ratings with the similarity levels.

3.3.3 Experimental design

In each experiment, we used three sessions - one per audio channel. Each session had two stages. All sessions were comprised of practice, testing, and break. In the practice phase, participants completed three trials for X and three for Y.

In every session of Exp **3**, during the testing phase, each of the five visual representations of X and Y was repeated five times to generate the testing trials. In both stages, 10 pairs (randomly selected) out of 25 were with SL1, 5 with SL2, and 10 with SL3. In total, from 20 participants, 1000 responses were collected per session.

In each session of Exp **4**-**6**, each of the ten visual representations of X and Y was repeated three times. Out of the three pairs per visual representation, two had the same similarity level, and in total, each participant had completed 60 testing trials. In Exp **4**, from 20 participants, per session, 1200 responses were collected. In Exp **5**, 1560 responses were collected per session, and in Exp **6**, per session, 1380 responses were collected.

3.3.4 Hypothesis

In Exp **3**-**6**, we had considered the audio channels and the similarity levels as the independent variables and the variation in the similarity ratings in an audio channel as the dependent variable. The hypotheses for each experiment were.

H3 We expected a significant difference in the perception of similarity with the different audio channels. Mainly, users follow the changes in an audio stream according to the visual representation (Hogan et al. 2017; Du et al. 2018). Previous studies show the varied technique used in the creation of stimuli influences the similarity perception between two different stimuli (Evans and Treisman 2010; Gogolou et al. 2019).
H4 We expected that in the case of well-perceived mappings, users' will rate similarity between visual and audio representation according to the similarity levels. The audio channel which would lead to a good judgment of similar audio–visual pairs will be suitable for observing deviations.

4 Results

Given the two perspectives (cf. Sect. 3), in this section, we present our results corresponding to the suitable choices in audio–visual relationships for presenting the categories (Sect. 4.1) and the magnitude of data (Sect. 4.2).

4.1 Effective relationships for categories

In Exp **1** and **2**, we had first encoded the results as 0 s and 1 s for the correct and incorrect responses, respectively. We then considered the two metrics based on the hypotheses **H1** and **H2** to observe the difference in the effectiveness of audio channels.

First, we measured variation in the differentiation task accuracy. For this, we compared the percentage of correct responses collected from the four audio channels. Wilcoxon signed-rank sum test was applied to each audio channel to identify how significantly its results differ from other audio channels. We could not expect a normal distribution in our responses, but we had considered data as ordinal. Our data thus met the assumptions of the non-parametric Wilcoxon signed rank-sum test.

Second, to assess the impact of incongruence on a mapping, we measured the consistency in differentiation task accuracy. For this, we compared how accuracy differs between congruent and incongruent pairs in the case of each audio channel. We applied the Chi-square goodness of fit test for the comparison.

4.1.1 Exp 1. audio channels with spatial position

Overall results (X+Y, Fig. 5A) show that participants observed the difference in the audio–visual pairs with high accuracy with the timbre. Their mean percentage of correct answers was 94.1% (SD=6.2), which drops by 24.1%±2.95 ([78.62% (SD=5.98)]) with the loudness (Fig. 5B). Pitch and tempo were in the middle, with the accuracy of tempo [81.75% (SD=7.35)] more than the pitch [78.62% (SD=5.98)]. In **H1**, we expected a significant difference in accuracy with the four audio channels. The Friedman test showed statistically significant difference in the results from four sessions. The pairwise comparisons between overall responses gave a significant difference in all pairs except tempo and pitch ($p = 0.093$ Fig. 5A, tempo-pitch 3.13%±3.19 Fig. 5B).

In either stage, we found the same sequence in the difference between the audio channels. In X, except for the difference between tempo and pitch ($p = .847$), all others were significant. In Y, we observed a high difference of timbre from pitch=26.7%±5.2, loudness=39.6%±4.9, and tempo=18.3%±2.8, and all differences, including tempo-pitch ($p = .025$), were significant.

The results of Y were impacted by participants' performance with the incongruent stimuli (Table 1). The high chi-squared values for pitch (51.253), loudness (83.737), and tempo (119.29) revealed how inaccurately participants observed the incongruent stimuli with the three audio channels. In the case of timbre, the difference was insignificant ($\chi^2 = 1.57, p = .209$). In **H2**, we expected consistent accuracy. Further, we considered the high accuracy for incongruent stimuli as a criterion for effectiveness. In Y, merely timbre had met the requirements.

In X for all audio channels, the difference between the incongruent and congruent stimuli remained insignificant (Table 1). The consistency in the results can be due to the way we applied incongruence. In pitch, loudness, and tempo, we did not replace the audio stimulus with the next lower or higher value stimulus of the same set. We used this to minimize the chances of results being biased towards timbre. However, the difference between the results of X and Y and the incongruent and congruent stimuli of Y shows the ineffectiveness of these three channels when more sensory attention is required.

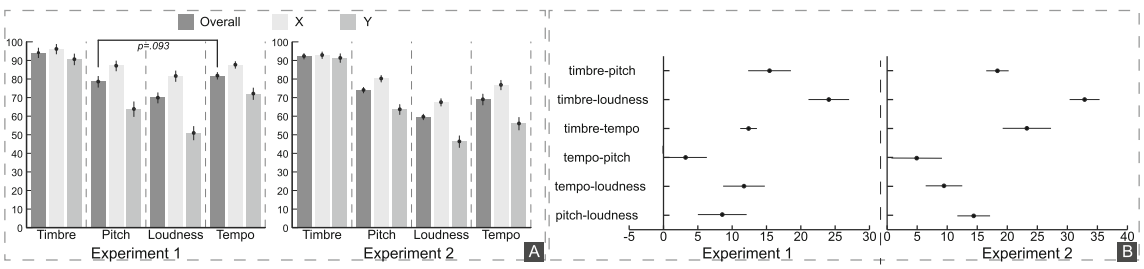

Fig. 5 This figure presents **A** The mean percentage of correct responses to show the difference between four audio channels along with spatial position (experiment 1) and color (experiment 2). For the difference, we present the overall (X+Y) results and results in X and Y. In experiment 1, the line over bars indicates two audio channels that are not significantly different (Wilcoxon signed rank-sum test). Error bars are 95% Cls; **B** The mean percentage difference with 95% Cls in six pairwise comparisons in experiments 1 and 2

Table 1 Experiment 1-Spatial position: Mean percentage of correct responses in both stages for congruent and incongruent pairs when we used the four audio channels along with the spatial position. The results are 95% Cls

		Timbre	Pitch	Loudness	Tempo
X	Congruent	96.5%±1.30	85.0%±3.84	77.3%±4.26	88.5%±2.64
	Incongruent	95.0%±2.38	89.3%±3.99	82.1%±4.50	83.6%±3.15
Y	Congruent	92.3%±2.49	75±5.13	62.3%±5.35	87.7%±3.04
	Incongruent	89%±3.39	43%± 4.17	22%±4.07	46%±4.12

Table 2 Experiment 2-Color: Mean percentage of correct responses in both stages for congruent and incongruent pairs when we used the four audio channels along with the color. The results are 95% Cls

		Timbre	Pitch	Loudness	Tempo
X	Congruent	93.5%±1.39	82.3%±2.44	64.6%±2.70	73.8%±4.55
	Incongruent	92.1%±2.84	68.5%±3.52	57.1%±4.26	74.3%±4.35
Y	Congruent	91.5%±2.23	74.5±2.94	58.5%±2.74	68.5%±4.31
	Incongruent	93%±2.62	48%± 4.45	36%±3.25	38%±5.71

4.1.2 Exp 2. audio channels with color

Overall results (Fig. 5A) show that participants had the best differentiation performance with timbre [92.% (SD=4.01)] and worst with the loudness [59.5% (SD=9.78)]. Pitch and tempo remained in the middle. However, the mean percentage of accuracy was better with pitch [74% (SD=5.27)] than tempo [69% (SD=7.8)]. The difference from Exp **1** is probably due to the audio used in tempo stimuli. In Exp **1**, we altered a piano sound for the tempo stimuli, while in the **2**, we used a pure sine wave. The Friedman test showed statistically significant difference in the results from four sessions. Furthermore, pairwise comparisons show, though the difference between pitch and tempo was small (Fig. 5B pitch-tempo = 5±4.16), all differences were significant with $p < .05$ as expected in **H1**. In either stage, results were in the same sequence. However, in X, the mean percentage difference between pitch and tempo was insignificant ($p = .194$).

The overall less accuracy for pitch, loudness, and tempo was not merely due to Y. In X, the mean percentage of accuracy for the three channels was less [80.2% (SD = 6.7), 67.4% (SD = 7.2), 76.8% (SD = 7.6)]. Additionally, the accuracy for the incongruent pairs with the three channels remained less. In the case of Y, participants' performance for the incongruent stimuli significantly differed from congruent with the three channels (Table 2). With timbre, the difference between X and Y ($\chi^2 = 1.28$ with $p = .258$) and congruent and incongruent stimuli was insignificant.

4.2 Effective relationships for data magnitude

In Exp **3**-**6**, to analyze X, Y, and overall results of each audio channel, we calculated the percentage for every point of the Likert scale. In each result, we did this for the three similarity levels. Further, we took the sum of the percentages for 5 and 4. We call it a high rating. Likewise, we took the sum of percentages of 1 and 2 as a low rating. This encoding criterion was suggested for the analysis of Likert-scale results (Xia et al. 2016). We used it because, in SL3, we applied the difference from 60% to 100% between the visual and audio stimuli. Stringent criteria could affect the analysis. We had then used the two metrics based on the hypotheses **H3** and **H4**.

First, to measure the variation in the similarity ratings, we compared how, with the different audio channels, the percentages shift from high to low rating with the shift in the similarity levels from SL1 to SL3. The diverging bar charts (Fig. 6), with precisely similar on the right, dissimilar on the left, and centered at the zero, provide a detailed comparison (Borkin et al. 2011). Wilcoxon signed-rank sum test was applied to identify how significantly an audio channel's ratings differ from the others. This test was used where the difference between the two channels seemed minimal.

Second, to measure the consistency in the similarity rating, we applied a one-way ANOVA test to each audio channel and identified how significantly its rating differs between its three similarity levels. Though the Shapiro-Wilk test, a formal normality procedure, showed our collected data violates the condition of normal distribution, our data was suitable for the One-way ANOVA test. We based our argument on the literature. Researchers (Kim 2013; Blanca Mena et al. 2017) had found that the data with skewness between

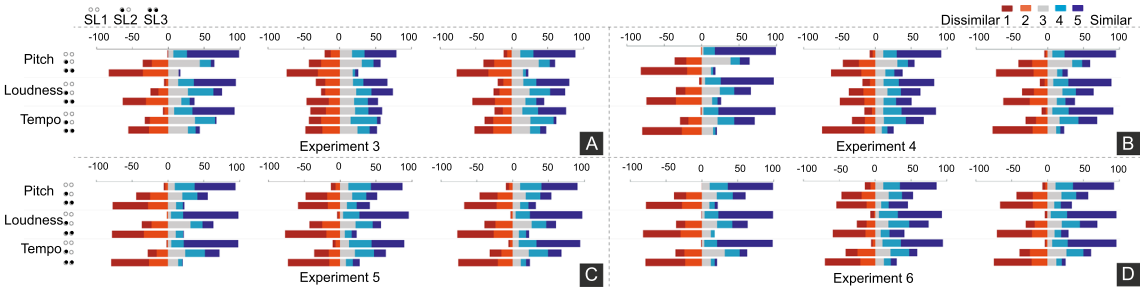

Fig. 6 This figure presents similarity ratings with the three audio channels paired with the four visual channels; 3-Position **A**, 4-Length **B**, 5-Angle **C**, and 6-Area **D**. Each part is broken down into three diverging bar charts: left presents X, middle - Y, and right - the overall results. For an audio channel, each bar represents a different similarity level

− 1 and 1 and the z-score value less than 2 do not show a substantial departure from the normality. Our data met this requirement.

The One-way ANOVA test helps observe the difference in the similarity ratings (Hermann et al. 2011). We had also computed the effect sizes. The results with significant differences and large effect sizes show consistency in the ratings. Table 3 presents the results. In this table, in the results of the pitch-position relationship, the first row presents the data collected at stage X. It shows how significantly the ratings for three similarity levels differ. The second shows results of stage Y, and the third shows overall results. The table presents the results of all other audio–visual relationships in the same pattern.

4.2.1 Exp 3. audio channels with position

Figure 6A shows that participants had better judged the difference in the three (SLs) with pitch compared to loudness and tempo. With pitch, we found a strong negative correlation between ratings and (SLs) (-0.71). However, it was moderately weak with loudness (-0.41) and tempo (-0.36). The small effect size for loudness ($\eta_p^2 = .16$) and tempo ($\eta_p^2 = .13$) ([cf. Table 3 column 1]), and the mean difference MD between three SLs with loudness (.46±.27, 1.32±.23,.66±.28) and tempo (.74±.26, 1.12±.22,.37±.27) had made the two channels less significant than pitch (MD: 1.43±.17, 2.29±.19,.86±.17). Figure 7A illustrates MD. More specifically, participants' observation of changes in the loudness ($\eta_p^2 = .05$) and tempo ($p = .108$) [cf. Table 3 column 1]) when used with the continuous line charts (Y) was very poor.

For all audio channels, the continuous lines made it difficult to precisely judge the position of data points, notably when the slope between points was zero. However, the observation of similarity with pitch (MD: 1.01±.34, 1.82±.29,.81±.36) was much better than loudness (MD: -.145[−.56, .27],.64±.34,.79±.42) and tempo (MD =.07±.39,.27±.38,.20±.4). MD results are further supported by pairwise comparisons with the Games-Howell test. It has provided a non-significant difference in the case of loudness (in SL1,SL2 $p = .682$) and tempo (in SL1,SL2 $p = .903$, in SL1,SL3 $p = .098$, and in SL2,SL3 $p = .42$). Additionally, the percentage of high ratings with loudness (SL1 = 48% [SD = 8.2%], SL2 = 51% [SD = 6%], SL3 = 34% [SD = 6.5%]) and tempo (SL1 = 38% [SD = 6.4%], SL2 = 42% [SD = 4.8%], SL3 = 28% [SD = 8.4%])

Table 3 Experiments 3-6 (Position/Length/Angle/Area): One-way ANOVA test shows how significantly the ratings of the three similarity levels differed in X (first row), Y (second row), and overall (third row) results of each audio channel. p and η_p^2 (in brackets) values present the significance. η_p^2 is based on Cohen's d classification (Sawilowsky 2009). p-value is set at 0.05 with 95% confidence interval. *→non-significant, **→small effect size

	Position	Length	Angle	Area
Pitch	.000 (0.65)	.000 (0.63)	.000 (0.44)	.000 (0.48)
	.000 (0.31)	.000 (0.32)	.012 (0.17)	.001 (0.15)
	.000 (0.46)	.000 (0.46)	.001 (0.28)	.001 (0.28)
Loudness	.000 (0.38)	.000 (0.46)	.000 (0.57)	.000 (0.59)
	.003 (0.05)	.022 (0.09)	.000 (0.47)	.002 (0.24)
	.001 (0.16)	.000 (0.24)	.000 (0.51)	.000 (0.39)
Tempo	.000 (0.42)	.000 (0.58)	.000 (0.58)	.000 (0.60)
	.108	.004 (0.27)	.002 (0.30)	.000 (0.38)
	.031 (0.13)	.001 (0.41)	.000 (0.43)	.000 (0.48)

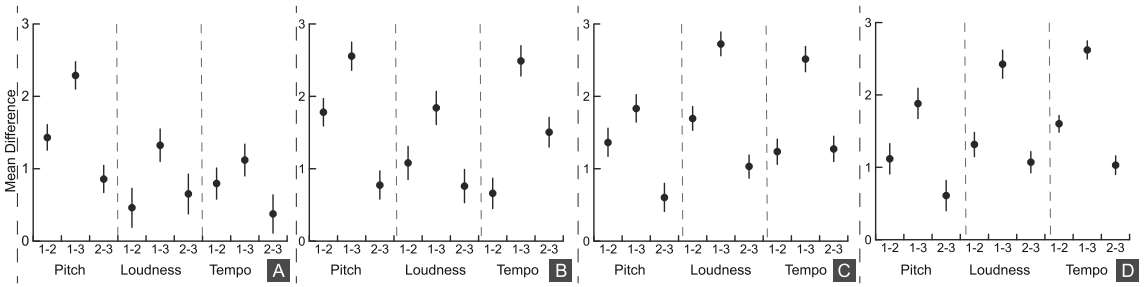

Fig. 7 This figure presents, based on overall results, the mean difference in users' rating of the three similarity levels for each audio channel in experiments 3-Position **A**, 4-Length **B**, 5-Angle **C**, and 6-Area **D**. Error bars are 95% Cls. Along the x-axis, 1 presents SL1, 2 - SL2, and 3 - SL3

further strengthens our analysis. In X, though the pitch was better than others, within-group significant difference serves as evidence for the effectiveness of loudness and tempo along with a series of points.

4.2.2 Exp 4. audio channels with length

Figure 6B shows that participants' perception of similarity was better with the pitch than tempo and loudness. With pitch, we found a strong negative correlation between ratings and SLs (-0.69). However, it was moderate with loudness (-0.49) and tempo (-0.63). Contrary to Exp **3**, participants better judged the (SLs with tempo compared to the loudness. The MD (Fig. 7B) and the Wilcoxon signed-rank sum test results that show similarity rating for SL2 and SL3 was better with tempo than loudness, highlight the difference between the two channels. Across all three audio channels, we observed the difference in the results of X and Y.

On average, in Y, the high rating of SL1 dropped by 20%, and SL3 increased by 18.6%. However, with loudness and tempo, we observed a considerable chance of incorrect judgment of the magnitude of selected points. With loudness, in SL2, 46% [SD = 10.3%] and in SL3, 39% [SD = 12%] of the responses were high ratings. While with tempo, 56% [SD = 6.5%] of the responses in SL2 were high ratings. Additionally, though the difference in the ratings was significant, the small effect sizes significantly lower the chance of selecting loudness (.09) and tempo (.27) (cf. Table 3 column 2) along the selected lengths.

4.2.3 Exp 5. audio channels with angle

Figure 6C shows that pitch was less effective in the perception of changes in angle than loudness and tempo. In loudness and tempo, the ratings were consistent with the SLs (η_p^2=.51,.43 cf. Table 3 column 3) and had a strong negative correlation (loudness -0.72 and tempo -0.66). However, with pitch, the effect size was small ($\eta_p^2 = .28$). The MD (Fig. 7C) was better for loudness (1.69±.15, 2.73±.16, 1.03±.16) and tempo (1.24±.17, 2.58±.16, 1.28±.17) than for pitch (1.32±.19, 1.97±.19,.60±.19). Further, participants' performance was better with loudness than tempo. With loudness, approximately 92% [SD = 7.6%] had judged SL1, and 77% [SD = 9.2%] finely observed SL3. With the tempo, it was 86% [SD = 6.6%] and 76% [SD = 5.7%]. The Wilcoxon signed-rank sum test shows in SL1-loudness and SL2-loudness, the similarity rating was significantly better ($p = .001$ and $p = .005$) than SL1-tempo and SL2-tempo. The main difference between the results of X and Y was observed with pitch. For pitch, in Y, the Games-Howell test had provided a non-significant difference between SL2 and SL3 ($p = .194$). The result shows how weakly the changes in 2d visual marks were perceived with pitch.

4.2.4 Exp 6. audio channels with area

Figure 6D shows that the results of Exp **6** were similar to those of **5**. The ratings were well consistent with the similarity levels in loudness (η_p^2=.39) and tempo (η_p^2=.48) but less in pitch (η_p^2=.28) (cf. Table 3 column 4). The MD (Fig. 7D) was better for loudness and tempo than for pitch. Contrary to Exp **5**, participants had better performed with tempo (in SL2-34% [SD = 9.3%] and in SL3-13% [SD = 7.8%] rated high) than loudness (in SL2-50% [11.3%] and in SL3-20% [SD = 8.2%]). In SL2-tempo and SL3-tempo at $p = .009$ and $p = .017$, respectively, the similarity rating was significantly better than SL2-loudness and SL3-

loudness. In the case of pitch, in stage Y, the effect size was small (η_p^2 =.15). It is due to a non-significant difference in pairwise comparisons. The Games Howell test has provided the non-significant difference between SL2 and SL3 ($p = .820$).

5 Discussion and conclusion

The work contributes to enhancing users' experience of audio–visual representations. To that end, it had provided empirical evidence about the effectiveness of four audio channels along six visual channels. The results suggested that all four audio channels can be used in audio–visual representations. However, the insights listed in the paper are essential for their practical use. This section discusses the analysis of results, the implications, applications, and limitations of user studies.

5.1 Differentiation accuracy

The Scope of the Effectiveness of a Mapping. The first two experiments show that the found mappings were inconsistent across the two stages of the experiment. In X, besides timbre, with pitch and tempo, participants were able to identify stimuli with an accuracy of more than 80%. However, in Y, except for timbre with all other channels, participants had identified visuals with low accuracy. The results suggest that the studies for effective mappings need to be expanded for all possible tasks. Furthermore, we can also assess the dependency of the effectiveness on the amount of information presented. The literature proposed that five distinct audios are easy to memorize (Metatla et al. 2016; Zhao et al. 2004). Thus, it might have influenced the results X. A large set may help to demonstrate a more reliable compatibility effect. It would provide further guidance on the factors that affect compatibility between the visual and audio channels.

In contrast to the two stages, the mappings were consistent across the two experiments. We can assume that an audio channel effective along an identity channel will also be effective for other identity channels. However, this assumption is doubtful as previous studies found an association between pitch and shapes but not between pitch and colors (Adeli et al. 2014; Metatla et al. 2016). Additionally, our study results show a significant difference between pitch/spatial position and pitch/color mapping ($p = 0.024$) and between tempo/spatial position and tempo/color mapping ($p = .000$).

Design Considerations. Overall, users' performance and the variation in the results show that a strong interaction between different senses occurs. It suggests the designer should ensure consistency in the audio across the whole visualization. An inconsistency might not be noticed but can confuse users and make a task difficult. Specifically, if a visual representation is very complex and users identify the categories with audio help. In stage Y of either experiment, a significant difference in users' performance suggests that when audio directs users' attention to the visual, there should not be any compromise on selecting audio channels. In addition, if the potential users are familiar with the visual form, provide them with the audio they like to hear with the image. We made this assertion based on the results of experiment 2. In the experiment, the relevance of timbre with color had resulted in a considerable difference between timbre and other audio channels (Table 2).

5.2 Magnitude judgment

Consistency in the Ranking of Audio Channels The found mappings in Exp **3** and **4** show that along with the continuous line representations and stacked bar charts where the selected targets were without color (*idea mimicked the perceptual visualization studies* (Cleveland and McGILL 1984)), the accuracy of similarity judgment with pitch (SL1 = 62% in Exp **3** and 81% in Exp **4**) was well above the loudness (48%, 70%) and tempo (37%, 72%). Du et al. (Du et al. 2018) had used loudness alongside position. However, it was proven not very useful, and the low significance of loudness is evident in our results. The results of Exp **5** and **6** show that users could observe loudness and tempo better than the pitch along with representations having additional visual cues *such as colors in the selected target*. Lipscomb and Kim (Lipscomb and Kim 2004) had suggested loudness and pitch for the area. We observed a considerable performance difference between pitch, loudness, and tempo. More-importantly, our study is significant to develop solid foundations as it is based on accuracy.

Regarding consistency in the mappings across the two stages, our study found that an audio channel that was more effective in a challenging task or display conditions (Y) had a higher ranking than others when

the conditions were simple (X). It shows that a paired visual channel has a strong influence on the perception of an audio channel. Although the audio channels' ranking remained the same in either stage of each experiment, users' performance remained stable with pitch. Varied conditions in the display or task did not have a drastic impact on the pitch effectiveness.

Design Considerations Here based on result we suggest few design considerations for usage of audio channels. First, in our results, we found that variation in visual representations has affected the perception of audio channels. The extreme imprecision with loudness and tempo in Y of Exp **3** had made them an unsuitable option with the representations like storylines (Tang et al. 2019). It further suggests that users can observe the changes in pitch along the conditions requiring more visual attention. Loudness and tempo are effective along with the simple visual representations. Taking inspiration from the results, the influence, of visual representations from other visual channels (e.g., circular area chart), on the audio channels could also be investigated.

Second, the audio pattern of pitch and loudness also differed in X (sound + silence) and Y (no silence) of Exp **3**. The pattern for tempo was the same. Nevertheless, with the tempo, we observed a significant performance difference between the two display types. Thus, results reveal that an audio pattern had not influenced the users' performance. However, a comprehensive follow-up study can confirm these findings.

Third, our results also guides task-based effectiveness of audio channels. We could consider pitch *(effective along with line and bar charts)* is a suitable choice of representing trends, distribution, or anomalies in the data (Zacks and Tversky 1999). Taking into account the effectiveness of loudness and tempo along with the pie charts and treemaps, they could be considered useful for representing clusters, or proportional relationships (Itoh et al. 2023). Here, we cannot advocate the selection of audio channels according to the task, because the assertions are based on task effectiveness of the visual channels (Saket et al. 2019). However, it inspires for further investigations. We could examine the effectiveness of audio channels for different tasks. It would be interesting to determine the tasks which though we can accomplish with visuals, but audio cannot be used for them.

Fourth, as users' visual abilities are addressed (such as color blindness, low vision), designers of audio–visual representations can consider hearing capabilities. However, it is vital to consider the similarity rating results, which show how closely users had observed the differences in visual and audio representations of data. Designers should avoid a mismatch between the visual and audio representation, down to the marginal differences.

Impact of Visual Channel. Our results show that even with the simple visual stimuli compared to continuous line charts (Exp **3**- Y) and marked-slots without visual cues (Exp **4**- Y), in Exp **5** and **6**, on average, 40% of users had wrongly perceived the deviation in audio stimuli. Thus, as suggested in the literature (Spence 2011; Evans and Treisman 2010), we can argue for the impact of visual channels' effectiveness on their compatibility with the audio channels.

5.3 Implications of user study

First, this research presents the first attempt to understand the mappings between visual and audio channels used in data visualization through comprehensive, two-stage experiments. The two stages provide the first list of observations: a) The visual representation has a strong influence on the perception of the audio channels (e.g., loudness and tempo in Y of Exp **3**). b) Though, with an audio channel, users can perceive a small magnitude difference, the location and size of targets impact the effectiveness of an audio channel (e.g., pitch in Y of Exp **5** and **6**). The observations suggest the judgment of accuracy should be further expanded into a sequence of stages.

Second, based on the results, this research also suggests the task-based effectiveness of audio channels. Mainly, no prior study has explored relationships between tasks and audio channels, except timbre for identifying categories (Flowers 2005). We had considered this limitation in the selection of the tasks for our six experiments. We selected the tasks that provide generalizable results and are suitable for studies with visual and audio channels. These tasks also did not have any specific visual display requirements. However, our results suggest the suitability of audio channels with the tasks like representing trends, distribution, anomalies in the data, or representing clusters.

Third, based on the study results, we proposed a ranking of four audio channels along six visual channels. Exp **5** and **6** results help us assume that perception of loudness and tempo along 2D visual elements remains similar. The similarity could be due to the correspondence between angle and area. However, previous studies (Adeli et al. 2014; Blazhenkova and Kumar 2018) show the difference in user

preferences for audio dimensions along with related visual dimensions. They also report that an association between visual channels does not impact the relationship between a visual and audio channel. For example, the use of color in shapes had not impacted users' decisions about the compatibility between shape and timbre (Adeli et al. 2014). Position and size were when used simultaneously; they had not influenced each other's mapping with pitch (Evans and Treisman 2010). Thus, the assumption may not be generalizable to all 2D shapes. However, it requires further investigations. Here it is essential to know that ranking does not provide complete guidance on making use of mappings. It gives guidelines on effective mappings that must be considered while designing audio–visual representations.

Fourth, users' perception of minor changes in Exp **3**-**6** suggests audio could be used to encode the data attributes to keep the visualizations simple, e.g., in the case of small display space (Mansoor et al. 2023) or occluded views (Limberger et al. 2023; Jin et al. 2021; Lan et al. 2022).

Fifth, the previous research (e.g., (Wan et al. 2019a; Tang et al. 2020; Wen et al. 2020; Batch et al. 2023;Ning et al. 2021)) proposed generating images from audio and audio from images. These works motivate the development of systems that uses machine learning models based on effective audio–visual mappings and generate an audio representation for visualization or vice versa.

Sixth, a previous study investigated the mapping between audio and the two interactions, pan and zoom and fisheye lens (Bouchara et al. 2010). The authors observed that for both pan &zoom and fisheye lens, audio resulted in a decrease in completion time. However, in pan &zoom, users found audio distracting. But in the fisheye lens, where they adjusted loudness with the size of the focused option, users preferred audio–visual over visual. Taking inspiration from this study, we can investigate the role of audio channels in other interaction techniques (Rubab et al. 2021).

5.4 Significance of audio–visual modalities in data visualization

Simultaneous use of audio and visual modalities provides several benefits. Like, they enrich users' immersion in AR and VR (Han and Surve 2019; Kwok et al. 2019; Su et al. 2021; McCormack et al. 2018). Audio augmented narratives provide users with an innovative and dynamic experience (Kwok et al. 2019). Well-suited audio eases the stress on the visual content, improves comprehension and retention of narrative, and leads to better task performance than visual-only conditions (Kim et al. 2018; Brewster and Clarke 2005). Recently, visualization researchers have proposed audio–visual experience in data exploration and interactions (Rind et al. 2018; Enge et al. 2022; Yang and Hermann 2018). Encoding data attributes with channels of different modalities is better than increasing the number of channels of a single modality (Nees and Walker 2011) or using varied interactions (Enge et al. 2022) while analyzing multidimensional data (Zhou et al. 2022). However, to achieve the benefits of redundant (cross-modal) modalities, we need to provide an audio channel that users accurately perceive with visuals (Yang and Hermann 2018; Rogińska et al. 2013; Zhao et al. 2004).

Our user studies provide a solid foundation for effective audio–visual relationships. We identified effectiveness and guidelines for 20 audio–visual relationships. Additionally, researchers (Tsuchiya et al. 2016) developed tools that take data as input and generate audio and visual representation as output. However, these tools can only encode attributes to pitch and timbre along with position and color, respectively. The tools can be extended based on our findings. Enge et al. (Enge et al. 2022) also proposed a tool for audio–visual representation. The tool interface allows users to select pitch and position for the same or different attributes. Visualization research (Munzner 2014) suggests suitable visual channels for attributes. For instance, the area can represent the size and the angle for a diverging attribute. However, the sonification field does not comprehensively guide the relevance between audio and data attributes. The tool (Enge et al. 2022) can be extended to other channels based on our findings. Our findings can help overcome the limitations of visual channels like area cannot be length or shape coded (Munzner 2014). The additional attributes can be audio encoded. Our results can define the constraints on the interactive adjustments of the audio channels (Ness et al. 2010; Enge et al. 2022), which otherwise may lead to the wrong interpretation of data representation.

5.5 Limitations and future work

First, we had limited our study to basic representations. Charles Spence (Spence 2007) also pointed at difference between the experimental setup and the complexity of actual usage. Simultaneously, he suggested that simple displays and tasks can provide significant evidence of binding between the channels (Spence

2007). Our results show the separable estimates with different audio channels and two stages. Thus, we believe they are plausible representative of effective audio–visual mappings strategies. We expect that our proposed rules will undergo testing and revisions with the accumulation of new information. **Second**, our experiments were based on representations with small datasets, inspired by existing studies (Heer and Bostock 2010; Smith and Walker 2005). If we had used a large dataset, participants might have felt difficulty analyzing the similarity level that relies on memory over a sequence of a long pairing of points. The effect of long audio/visual pairing on results should be investigated in future studies. **Third**, we had not considered intramodal association (e.g., between pitch and loudness (Neuhoff et al. 2002)). However, there is scope for studying the influence of association between the visual or audio channels on audio–visual mapping.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s12650-023-00909-3.

Acknowledgements The work was supported by NSFC (61761136020), NSFC-Zhejiang Joint Fund for the Integration of Industrialization and Information (U1609217), Zhejiang Provincial Natural Science Foundation (LR18F020001) and the 100 Talents Program of Zhejiang University. This project was also partially funded by Microsoft Research Asia.

References

Adeli M, Rouat J, Molotchnikoff S (2014) Audiovisual correspondence between musical timbre and visual shapes. Front Hum Neurosci 8:352

Batch A, Ji Y, Fan M, Zhao J, Elmqvist N (2023) uxSense: Supporting user experience analysis with visualization and computer vision. IEEE Trans Vis Comput Graph, To appear

Blanca Mena MJ, Alarcón Postigo R, Arnau Gras J, Bono Cabré R, Bendayan R (2017) Non-normal data: is anova still a valid option? Psicothema 29(4):552–557

Blazhenkova O, Kumar MM (2018) Angular versus curved shapes: correspondences and emotional processing. Perception 47(1):67–89

Borkin M, Gajos K, Peters A, Mitsouras D, Melchionna S, Rybicki F, Feldman C, Pfister H (2011) Evaluation of artery visualizations for heart disease diagnosis. IEEE Trans Vis Comput Graph 17(12):2479–2488

Bouchara T, Katz BF, Jacquemin C, Guastavino C (2010) Audio-visual renderings for multimedia navigation. In: Proc. of International Conference on Auditory Display, pp 245–252

Brewster SA, Clarke CV (2005) The design and evaluation of a sonically enhanced tool palette. ACM Trans Appl Percept 2(4):455–461

Bruls M, Huizing K, Wijk JJV (2000) Squarified treemaps. In: Proc. of Eurographics Conference on Visualization, pp 33–42

Cleveland WS, McGILL R (1984) Graphical perception: Theory, experimentation, and application to the development of graphical methods. J Am Stat Assoc 79(387):531–554

Daudé S, Nigay L (2003) Design process for auditory interfaces. In: Proc. of International Conference on Auditory Display, pp 176–179

Demiralp Ç, Bernstein MS, Heer J (2014) Learning perceptual kernels for visualization design. IEEE Trans Vis Comput Graph 20(12):1933–1942

Demiralp Ç, Scheidegger CE, Kindlmann GL, Laidlaw DH, Heer J (2014) Visual embedding: a model for visualization. IEEE Comput Graph Appl 34(1):10–15

Du M, Chou JK, Ma C, Chandrasegaran S, Ma KL (2018) Exploring the role of sound in augmenting visualization to enhance user engagement. In: Proc. of IEEE Pacific Visualization Symposium, pp 225–229

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

Enge K, Rind A, Iber M, Ḧoldrich R, Aigner W (2022) Towards multimodal exploratory data analysis: Soniscope as a prototypical implementation. In: Proc. of Eurographics Conference on Visualization-Short Papers, pp 67–71

Ernst MO (2007) Learning to integrate arbitrary signals from vision and touch. J Vis 7(5):1–14

Evans KK, Treisman A (2010) Natural cross-modal mappings between visual and auditory features. J Vis 10(1):6

Ferguson J, Brewster SA (2018) Investigating perceptual congruence between data and display dimensions in sonification. In: Proc. of ACM CHI Conference on Human Factors in Computing Systems, pp 1–9

Flowers JH (2005) Thirteen years of reflection on auditory graphing: Promises, pitfalls, and potential new directions. In:Proc. of International Conference on Auditory Display, pp 406–409

Franklin KM, Roberts JC (2003) Pie chart sonification. In: Proc. of International Conference on Information Visualisation, pp 4–9

Ghosh S, Winston L, Panchal N, Kimura-Thollander P, Hotnog J, Cheong D, Reyes G, Abowd GD (2018) Notifivr: exploring interruptions and notifications in virtual reality. IEEE Trans Vis Comput Graph 24(4):1447–1456

Giovannangeli L, Bourqui R, Giot R, Auber D (2022) Color and shape efficiency for outlier detection from automated to user evaluation. Vis Inform 6(2):25–40

Gogolou A, Tsandilas T, Bezerianos P, Bezerianos A (2019) Comparing similarity perception in time series visualizations. IEEE Trans Vis Comput Graph 25(1):523–533

Han YC, Surve P (2019) Eyes: Iris sonification and interactive biometric art. In: Proc. of ACM CHI Conference on Human Factors in Computing Systems Extended Abstracts, pp 1–4

Hansen B, Baltaxe-Admony LB, Kurniawan S, Forbes AG (2019) Exploring sonic parameter mapping for network data structures. In: Proc. of International Conference on Auditory Display, pp 67–74
Harada S, Wobbrock JO, Landay JA (2011) Voice games: investigation into the use of non-speech voice input for making computer games more accessible. In: Proc. of IFIP International Conference on Human Computer Interaction, pp 11–29
Harding C, Kakadiaris IA, Casey JF, Loftin RB (2002) A multi-sensory system for the investigation of geoscientific data. Elsevier Comput Graph 26(2):259–269
Harrison L, Yang F, Franconeri S, Chang R (2014) Ranking visualizations of correlation using weber's law. IEEE Trans Vis Comput Graph 20(12):1943–1952
Heer J, Bostock M (2010) Crowdsourcing graphical perception: Using mechanical turk to assess visualization design. In: Proc. of ACM CHI Conference on Human Factors in Computing Systems Conference on Human Factors in Computing Systems, pp 203–212
Hermann T, Hunt A, Neuhoff JG (2011) The sonification handbook. Logos Verlag Berlin, Germany
Hogan T, Hinrichs U, Hornecker E (2017) The visual and beyond: Characterizing experiences with auditory, haptic and visual data representations. In: Proc. of ACM Conference on Designing Interactive Systems, pp 797–809
Itoh T, Nakabayashi A, Hagita M (2023) Multidimensional data visualization applying a variety-oriented scatterplot selection technique. J Vis 26(1):199–210
Janata P, Childs E (2004) Marketbuzz: Sonification of real-time financial data. In: Proc. of International Conference on Auditory Display
Jin Z, Cao N, Shi Y, Wu W, Wu Y (2021) EcoLens: visual analysis of ecological regions in urban contexts using traffic data. J Vis 24(2):349–364
Jin Z, Wang X, Cheng F, Sun C, Liu Q, Qu H (2023) ShortcutLens: A visual analytics approach for exploring shortcuts in natural language understanding dataset. IEEE Trans Vis Comput Graph, To appear
Khulusi R, Kusnick J, Meinecke C, Gillmann C, Focht J, Jänicke S (2020) A survey on visualizations for musical data. Comput Graph Forum 39:82–110
Kim HY (2013) Statistical notes for clinical researchers: assessing normal distribution (2) using skewness and kurtosis. Restor Dent Endod 38(1):52–54
Kim YJ, Kumaran R, Sayyad E, Milner A, Bullock T, Giesbrecht B, Ḧollerer T (2022) Investigating search among physical and virtual objects under different lighting conditions. IEEE Trans. Vis. Comput. Graph 28(11):3788–3798
Kim K, Billinghurst M, Bruder G, Duh HBL, Welch GF (2018) Revisiting trends in augmented reality research: a review of the 2nd decade of ISMAR (2008–2017). IEEE Trans. Vis. Comput. Graph 24(11):2947–2962
Knoeferle KM, Knoeferle P, Velasco C, Spence C (2016) Multisensory brand search: how the meaning of sounds guides consumers' visual attention. J. Exp. Psychol 22(2):196
Kong HK, Zhu W, Liu Z, Karahalios K (2019) Understanding visual cues in visualizations accompanied by audio narrations. In: Proc. of ACM CHI Conference on Human Factors in Computing Systems, pp 1–13
Krygier JB (1994) Sound and geographic visualization. Modern Cartography Series 2:149–166
Kwok TC, Kiefer P, Schinazi VR, Adams B, Raubal M (2019) Gaze-guided narratives: adapting audio guide content to gaze in virtual and real environments. In: Proc. of ACM CHI Conference on Human Factors in Computing Systems, pp 1–12
Lan J, Wang J, Shu X, Zhou Z, Zhang H, Wu Y (2022) RallyComparator: visual comparison of the multivariate and spatial stroke sequence in table tennis rally. J Vis 25(1):1–16
Lee Y, Lee CH, Cho JD (2021) 3d sound coding color for the visually impaired. Electronics 10(9):1037
Limberger D, Scheibel W, Ḋollner J, Trapp M (2023) Visual variables and configuration of software maps. J Vis 26(1):249–274
Lipscomb SD, Kim EM (2004) Perceived match between visual parameters and auditory correlates. In: Proc. of International Conference on Music Perception and Cognition, pp 72–75
Mackinlay J (1986) Automating the design of graphical presentations of relational information. ACM Trans Graph 5(2):110–141
Mansoor H, Gerych W, Alajaji A, Buquicchio L, Chandrasekaran K, Agu E, Rundensteiner E, Rodriguez AI (2023) INPHOVIS: Interactive visual analytics for smartphone-based digital phenotyping. Vis Inform, To appear
McCormack J, Roberts JC, Bach B, Freitas CDS, Itoh T, Hurter C, Marriott K (2018) Multisensory immersive analytics. In: Immersive analytics, Springer, pp 57–94
Metatla O, Correia NN, Martin F, Bryan-Kinns N, Stockman T (2016) Tap the ShapeTones: Exploring the effects of crossmodal congruence in an audio-visual interface. In: Proc. of ACM CHI Conference on Human Factors in Computing Systems, pp 1055–1066
Munzner T (2014) Visualization analysis and design. CRC Press, Boca Raton, FL
Ness RS, Reimer P, Krell N, Odowichuck G, Schloss WA, Tzanetakis G (2010) Sonophenology: a tangible interface for sonification of geo-spatial phenological data at multiple time-scales. In: Proc. of International Conference on Auditory Display, pp 335–341
Nees MA, Walker BN (2011) Auditory displays for in-vehicle technologies. Rev Hum Factors Ergon 7(1):58–99
Nesbitt KV, Barrass S (2002) Evaluation of a multimodal sonification and visualization of depth of market stock data. In: Proc. of International Conference on Auditory Display, pp 1–6
Neuhoff JG, Wayand J, Kramer G (2002) Pitch and loudness interact in auditory displays: Can the data get lost in the map? J Exp Psychol Appl 8(1):17–25
Ning H, Zheng X, Yuan Y, Lu X (2021) Audio description from image by modal translation network. Neurocomputing 423:124–134
Ondov B, Jardine N, Elmqvist N, Franconeri S (2019) Face to face: evaluating visual comparison. IEEE Trans Vis Comput Graph 25(1):861–871
Papachristodoulou P, Betella A, Manzolli J (2015) Augmenting the navigation of complex data sets using sonification: A case study with brainx 3. In: Proc. of IEEE VR Workshop: Sonic Interaction in Virtual Environments, pp 1–6

Parise C, Spence C (2013) Audiovisual cross-modal correspondences in the general population. The Oxford handbook of synaesthesia 790:815
Ren Z, Yeh H, Klatzky R, Lin MC (2013) Auditory perception of geometry-invariant material properties. IEEE Trans Vis Comput Graph 19(4):557–566
Rind A, Iber M, Aigner W (2018) Bridging the gap between sonification and visualization. In: Proc. of AVI Workshop on Multimodal Interaction for Data Visualization
Rogińska A, Friedman K, Mohanraj H (2013) Exploring sonification for augmenting brain scan data. In: Proc. of International Conference on Auditory Display, pp 95–105
Rönnberg N (2019) Musical sonification supports visual discrimination of color intensity. Behav Inform Technol 38(10):1028–1037
Roodaki H, Navab N, Eslami A, Stapleton C, Navab N (2017) Sonifeye: Sonification of visual information using physical modeling sound synthesis. IEEE Trans Vis Comput Graph 23(11):2366–2371
Rouben A, Terveen L (2007) Speech and non-speech audio: Navigational information and cognitive load. In: Proc. of International Conference on Auditory Display, pp 468–475
Rubab S, Tang J, Wu Y (2021) Examining interaction techniques in data visualization authoring tools from the perspective of goals and human cognition: a survey. J Vis 24(2):397–418
Saket B, Endert A, Demiralp C (2019) Task-based effectiveness of basic visualizations. IEEE Trans Vis Comput Graph 25(7):2505–2512
Sanabria D, Soto-Faraco S, Spence C (2004) Exploring the role of visual perceptual grouping on the audiovisual integration of motion. Neuroreport 15(18):2745–2749
Sawe N, Chafe C, Treviño J (2020) Using data sonification to overcome science literacy, numeracy, and visualization barriers in science communication. Front comm 5:46
Sawilowsky SS (2009) New effect size rules of thumb. J Mod Appl Stat Methods 8(2):597–599
Schito J, Fabrikant SI (2018) Exploring maps by sounds: using parameter mapping sonification to make digital elevation models audible. Int J Geogr Inf Sci 32(5):874–906
Shenkar O, Weiss PL, Algom D (2005) Auditory representation of visual stimuli: Mapping versus association. In: Proc. of International Conference on Auditory Display, pp 273–275
Skau D, Kosara R (2016) Arcs, angles, or areas: individual data encodings in pie and donut charts. Comput Graph Forum 35(3):121–130
Smith DR, Walker BN (2005) Effects of auditory context cues and training on performance of a point estimation sonification task. Appl Cogn Psychol 19(8):1065–1087
Spence C (2007) Audiovisual multisensory integration. Acoust Sci Technol 28(2):61–70
Spence C (2011) Crossmodal correspondences: a tutorial review. Atten Percept Psychophys 73(4):971–995
Spence C (2020) Simple and complex crossmodal correspondences involving audition. Acoust Sci Technol 41(1):6–12
Su C, Yang C, Chen Y, Wang F, Wang F, Wu Y, Zhang X (2021) Natural multimodal interaction in immersive flow visualization. Vis Inform 5(4):56–66
Sun X, Li X, Ji L, Han F, Wang H, Liu Y, Chen Y, Lou Z, Li Z (2018) An extended research of crossmodal correspondence between color and sound in psychology and cognitive ergonomics. PeerJ 6:e4443
Tang T, Rubab S, Lai J, Cui W, Yu L, Wu Y (2019) iStoryline: effective convergence to hand-drawn storylines. IEEE Trans Vis Comput Graph 25(1):769–778
Tang Z, Bryan NJ, Li D, Langlois TR, Manocha D (2020) Scene-aware audio rendering via deep acoustic analysis. IEEE Trans Vis Comput Graph 26(5):1991–2001
Tsiros A (2014) Evaluating the perceived similarity between audio-visual features using corpus-based concatenative synthesis. In: Proc. of International Conference on New Interfaces for Musical Expression, pp 421–426
Tsuchiya T, Freeman J, Lerner LW (2016) Data-driven live coding with datatomusic api
Turnage KD, Bonebright TL, Buhman DC, Flowers JH (1996) The effects of task demands on the equivalence of visual and auditory representations of periodic numerical data. Behav res meth instrum comput 28(2):270–274
Wang J, Cai X, Su J, Liao Y, Wu Y (2022a) What makes a scatterplot hard to comprehend: data size and pattern salience matter. J Vis 25(1):59–75
Wang L, Sun G, Wang Y, Ma J, Zhao X, Liang R (2022b) AFExplorer: Visual analysis and interactive selection of audio features. Vis Inform 6(1):47–55
Walker BN (2007) Consistency of magnitude estimations with conceptual data dimensions used for sonification. Appl Cogn Psychol 21(5):579–599
Wan CH, Chuang SP, Lee HY (2019) Towards audio to scene image synthesis using generative adversarial network. In: Proc. of IEEE International Conference on Acoustics, Speech and Signal Processing, pp 496–500
Wang Y, Chen X, Ge T, Bao C, Sedlmair M, Fu CW, Deussen O, Chen B (2019) Optimizing color assignment for perception of class separability in multiclass scatterplots. IEEE Trans Vis Comput Graph 25(1):820–829
Wen X, Wang M, Richardt C, Chen ZY, Hu SM (2020) Photorealistic audio-driven video portraits. IEEE Trans Vis Comput Graph 26(12):3457–3466
Werner L, Fay RR, Popper AN (2011) Human auditory development, vol 42. Springer, Newyork
Wersényi G, Nagy H, Csapó A (2015) Evaluation of reaction times to sound stimuli on mobile devices. In: Proc. of International Conference on Auditory Display, pp 268–272
Wilson SR (1982) Sound and exploratory data analysis. In: COMPSTAT symposium, Springer, pp 447–450
Xia H, Araujo B, Grossman T, Wigdor D (2016) Object-oriented drawing. In: Proc. of ACM CHI Conference on Human Factors in Computing Systems, pp 4610–4621
Yang J, Hermann T (2018) Interactive mode explorer sonification enhances exploratory cluster analysis. AES: J Audio Eng Soc 66(9):703–711
Yeung ES (1980) Pattern recognition by audio representation of multivariate analytical data. Anal Chem 52(7):1120–1123
Zacks J, Tversky B (1999) Bars and lines: a study of graphic communication. Memory Cogn 27(6):1073–1079

Zhou Y, Meng X, Wu Y, Tang T, Wang Y, Wu Y (2022) An intelligent approach to automatically discovering visual insights. J Vis, To appear

Zhao Y, Jiang J, Chen Y, Liu R, Yang Y, Xue X, Chen S (2022) Metaverse: Perspectives from graphics, interactions and visualization. Vis Inform 6(1):56–67

Ziemer T, Schultheis H (2018) A psychoacoustic auditory display for navigation. In: Proc. of International Conference on Auditory Display, pp 136–144

Zhao H, Plaisant C, Shneiderman B, Duraiswami R (2004) Sonification of geo-referenced data for auditory information seeking: Design principle and pilot study. In: Proc. of International Conference on Auditory Display, pp 1–8