Supporting Accessible Data Visualization Through Audio Data Narratives

Alexa F. Siu
Stanford University and Adobe Research
asiu@adobe.com

Gene S-H Kim
Stanford University
gene.sh.kim@stanford.edu

Sile O'Modhrain
University of Michigan
sileo@umich.edu

Sean Follmer
Stanford University
sfollmer@stanford.edu

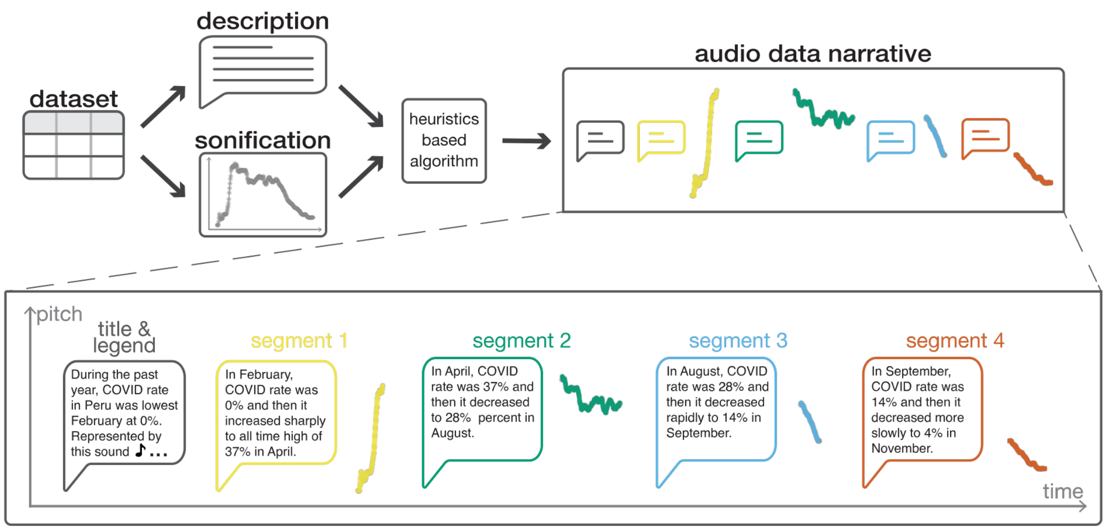

Figure 1: We introduce audio data narratives, which combine textual data representations and data sonification. Based on identified design principles and relevant auditory processing characteristics, we propose a heuristics-based approach to automatically generating a narrative given a time-series dataset.

ABSTRACT

Online data visualizations play an important role in informing public opinion but are often inaccessible to screen reader users. To address the need for accessible data representations on the web that provide direct, multimodal, and up-to-date access to the data, we investigate audio data narratives –which combine textual descriptions and sonification (the mapping of data to non-speech sounds). We conduct two co-design workshops with screen reader users to define design principles that guide the structure, content, and duration of a data narrative. Based on these principles and relevant auditory processing characteristics, we propose a dynamic programming approach to automatically generate an audio data narrative from a given dataset. We evaluate our approach with 16 screen reader users. Findings show with audio narratives, users gain significantly more insights from the data. Users describe data narratives help them better extract and comprehend the information in both the sonification and description.

CCS CONCEPTS

• **Human-centered computing**; • **Accessibility**; • **Accessibility systems and tools.**; • **Visualization**; • **Visualization systems and tools.**;

KEYWORDS

accessibility, data visualization, sonification, data narratives

ACM Reference Format:
Alexa F. Siu, Gene S-H Kim, Sile O'Modhrain, and Sean Follmer. 2022. Supporting Accessible Data Visualization Through Audio Data Narratives. In *CHI Conference on Human Factors in Computing Systems (CHI '22), April 29–May 05, 2022, New Orleans, LA, USA.* ACM, New York, NY, USA, 19 pages. https://doi.org/10.1145/3491102.3517678

CHI '22, April 29–May 05, 2022, New Orleans, LA, USA

ACM ISBN 978-1-4503-9157-3/22/04...$15.00
https://doi.org/10.1145/3491102.3517678

1 INTRODUCTION

Online data visualizations are increasingly used, by both domain experts and general audiences, to communicate important insights from complex data [44]. Insights gained from data visualization can help people make important decisions concerning health (e.g., COVID-19) and finances (e.g., stock trends), guide policy makers and scientists in understanding natural phenomena, and support communication in journalism, etc.

Data visualizations are effective in amplifying cognition for data exploration [14]. Much of the guidelines and tools used for creating data visualizations have been investigated for producing effective visual graphics, which poses a significant disadvantage to those that cannot benefit from visual consumption [26, 65]. Recent work has highlighted some of the challenges faced by people who are Blind and/or Visually Impaired (BVI), as primarily screen reader users, when accessing data visualizations on the web [35, 48, 65, 66]. Screen reader users access data visualizations in alternative ways using additional and/or different modalities (e.g., primarily speech and audio). Towards improving screen reader users' access to non-visual data representations, in this work, we investigate audio data narratives –which combine textual data representations and data sonification. Our work is driven by the need for accessible data representations that provide rich and direct access to the data, are updatable, and are robust to access on the web by screen reader users [48].

Image descriptions, presented as alternative text or alt text, and tabular data are among the most common alternative representations that are accessible to screen reader users on the web, as recommended by web accessibility guidelines such as WCAG [74]. A limitation of descriptions, or more generally textual representations, is that they do not provide direct access to the data [65] and instead only capture the author's interpretation of the data, rather than supporting the reader in making their own interpretation [66]. The level and quality of description can also vary widely [35]. Compounded with these issues is the fact that descriptions often require human authoring for them to be done well. With data charts that are updated automatically, descriptions can become outdated, creating a mismatch with the information presented visually [35, 66]. Lastly, textual representations rely on speech modalities which can pose a high cognitive load to interpret compared to a direct perceptual interface, especially when communicating graphical and highly spatial information [28].

Data sonification is an alternative representation that addresses some of the challenges with textual representations by providing more direct access to the data. Sonification is the mapping of relationships in data into perceived relations in an acoustic signal, taking advantage of human's auditory perceptual capabilities to make the relationships comprehensible [34]. Much of the work on sonification has focused on the low-level data inquiries, by investigating auditory parameters that make the information more perceivable [11, 49, 53, 71]. Considerably less work has focused on additional information that contributes to higher level communication goals (e.g., trend identification, predictions, decision-making) [5]. Communication is the primary purpose of visualizations presented for casual consumption such as in the news and articles on the web. To improve data communication with audio graphs using real-world datasets, more focus is needed on understanding how to support the user in interpreting the sounds in the graph and making connections between perceptual and conceptual levels to gain meaningful insights from the data.

With visual graphics, several studies have assessed the benefits of tight integration between text and visualization through different spatial layouts, data annotations, and interactions [36, 73, 81]. Data narratives are increasingly used to aid narrative communication of findings to non-domain experts by helping to clearly highlight and emphasize one or more intended messages in the data [7, 60]. Segel et al. describe narrative visualization as "*tours through visualized data*" which can be organized in a linear or non-linear sequence or "*they can also be interactive, inviting verification, new questions, and alternative explanations*" [64]. Data narratives have mostly been explored with visual graphics. We posit that effective narrative techniques can be extended and applied to also improve consumption of audio graphics. A data representation that more tightly integrates descriptions as a narrative to guide the reader, with data sonification to provide direct access to the data, could improve data communication through accessible modalities for screen reader users.

In this work, we introduce audio data narratives and explore their benefits and tradeoffs for data communication purposes. We conducted two virtual workshops with BVI co-designers to define design principles that guide the structure, content, and duration of an audio data narrative. As a starting point, our design investigation focuses on improving communication of time-series data. Temporal data is among the most common data types, typically presented visually through a line chart [8], and much prior work investigating auditory mappings for sonification have set forth guidelines for time-series data [12]. We apply these design principles driven by findings from the co-design workshops and prior work in auditory perception, to develop a heuristics-based algorithm for automatically generating an audio data narrative given a time-series dataset. Figure 1 provides an overview of our approach.

To evaluate how our approach with audio data narratives supports data communication of real-world datasets, we conduct an evaluation with sixteen BVI screen reader users. We find the audio data narrative representation, which interleaves both description and sonification, helps users gain a more complete gist of the data when compared to a standard sonification representation (control). The control representation has the same description but presented first rather than interleaved with the sonification. Users draw more insights from the sonification when consuming the information in narrative form. The audio data narratives are especially helpful for communicating complex real-world datasets that have more than two trend reversals [54], but their benefits are lesser with more simple datasets. Users describe their preference for audio data narratives in helping them better extract and comprehend the information in the sonification when gaining a gist of the data. Based on the evaluation findings, we believe, audio data narratives are a promising approach to provide automatic and up-to-date access to data visualizations on the web.

2 RELATED WORK

First, we review important considerations on auditory perception to inform effective information encodings. Additionally, we review related work on data representations accessible to screen reader users focusing on the use of speech and non-speech modalities.

2.1 Auditory Perception & Encoding

There are three perceptual tasks central to auditory cognitive processing: segmentation, localization, and categorization [46]. Given a sensory stream, the auditory system segments multiple potential sound sources into distinct sources to form a coherent spatial scene of the environment. Bregman termed the task of analyzing a mixture of sounds *auditory scene analysis* [9]. The auditory system makes sense of an auditory stream by 1) making use of primitive processes of auditory grouping, and 2) leveraging known schemas incorporating our knowledge of familiar sounds. Both processing mechanisms involve bottom-up and top-down processes. This task has similarities with visual processing where given a visual scene, the visual system must partition the scene into one or more objects and foreground and background [40]. However, given the temporal nature of hearing and the fact that sounds are transient, supporting memory and minimizing workload is particularly important for audition [46].

In perceiving and categorizing complex patterns as a whole, the auditory system, like haptics and vision, has the ability of auditory gestalt formation [33]. Organization of sound components into a meaningful element is referred to as an auditory object [70]. A number of gestalt principles have been shown to apply in auditory perception, these include: grouping by timbre, frequency proximity of sound events, good continuation of sound events, common fate, and closure [20, 33]. Effective data representations can take advantage of the auditory perceptual system capabilities to make the information more comprehensible and reduce workload.

2.2 Natural Language Descriptions of Data

Image descriptions or alternative text, often referred to as alt text, is the most common way in which BVI users encounter graphical representations. Alt text provides a textual alternative to graphical content. When accessing an image with alt text, if it is available, it is presented to the users' assistive technology. With screen readers, alt text would be read aloud or, if not present, an image may just be announced as "image" with no description of its content. The Web Content Accessibility Guidelines (WCAG) provide general guidelines for the creation of alt text [74], while the National Center for Accessible Media (NCAM) provides more specific guidelines for describing STEM images including data charts [30]. Using the NCAM guidelines, Morash et al. developed and evaluated a template-based description generator for data charts which lead to more standardized word usage and structure [49]. To address challenges with human authoring of image descriptions and provide access to up-to-date information, other approaches have investigated automating image descriptions combining computer vision and natural language processing. These approaches have been investigated for general-purpose images (e.g. in social media, web search) [28, 75] as well as for describing specific types of data visualizations and data features [17, 23, 38, 51].

An important consideration with natural language descriptions, is that meaningful information may be strongly reader-specific [47]. BVI users' preferences on image descriptions can vary depending on the media source and information-seeking goals [67]. Prior studies investigating BVI users' experiences with descriptions of data graphics have shown that while descriptions support the user in gaining a general overview, they are not comprehensive enough in supporting a rich and detailed understanding [65, 77] and that there is a gap in supporting the user in confidently generating their own insights [66]. Lungard et al. investigate a model of semantic content to better guide the content of image descriptions and position natural language as a data interface coequal with visualization [47].

A last consideration with descriptions is that while textual representations can accurately describe information, such presentation tends to be more verbose, error prone to interpret, and require more cognitive load than a perceptual interface that directly renders the same information through touch or vision [29]. In this work, we investigate how shortcomings from natural language descriptions can be addressed with complementary information provided through data sonification (and vice versa).

2.3 Tabular Data Representations

For data-driven content, in addition to image descriptions, guidelines also recommend including the source data in tabular form. While tabular representations provide direct access to the data, there are several limitations including: overloaded speech feedback [59] and working memory [37 ,68], and lack of an overall picture of the data structure [59, 65, 66]. Speech and non-speech sounds have been used to improve navigation and comprehension of 2D tabular representations by significantly reducing workload and providing a better overview of the information [37, 59, 68]. To address the challenges with textual representations while providing more direct access to the data, in this work, we explore audio data narratives and assess the benefits of more tightly integrating both textual representations and non-speech sounds.

2.4 Data Sonification

Sonification is another method that exploits sound to make data graphics more accessible by transforming data relations into perceived relations in an acoustic signal [34]. Sonification has been investigated for communicating a variety of data such as time-series data [12, 21, 42], georeferenced data [80], and mathematical functions [2, 56, 77]. Recently we have seen a number of these technologies used in practice such as in the Desmos graphing calculator [18] used in education to support both visual and audio representations of mathematical functions, Apple's audio graph accessibility API [6] that allows specification of audio graphs and sonified data, and the SAS Graphics Accelerator [62] which allows importing tabular data and exploration through a variety of sonified graphs. Auditory displays can deliver high amounts of detail but there are multiple mapping possibilities and few standards in place [63]. In considering mapping possibilities for data communication, Sawe et al. recommend striking a balance between four key but interrelated elements: fidelity to the data, level of complexity, aesthetics, and accessibility. Prior works have also investigated the effectiveness of specific data to audio mappings for sonification displays; providing recommendations on pitch and time mapping, choosing distinct timbres [12], polarity mapping [71], the use of rhythmic clicks for context [52], reference tones and white noise [53], and duration [53] (See a more general review and guidelines in [34]). Much of this work on sonification has focused on the low-level data tasks, by investigating auditory parameters that make the information more perceivable.

To better support high-level task interactivity, Zhao et al. proposed a framework for Auditory Information Seeking Actions (AISAs) [79]. The framework includes Gist, Navigate, Filter, and Details-on-Demand. Gist is the first step in obtaining an overview of the data, to guide further exploration. In Zhao et al.'s work, a common sonification exploration strategy used by BVI users to obtain a gist of georeferenced data was to break down the data into smaller ranges and then sweep each range in a consistent order to systematically build an overview [79]. This strategy was more effective than repeated sweeps of the entire map and could be explained by the limited working memory capacity with auditory stimuli, especially when investigating complex data. Brown et al. also reported sighted users' systematic isolation of specific regions was helpful in understanding sonified line graphs [11]. Similar to other kinds of graphics, with sonification, good strategies also need to be learned [33, 79]. Working with BVI novices unfamiliar with sonification, Zhao recommended training on specific auditory sweep and pattern recognition strategies [79]. For casual data consumption on the web, providing such one-on-one training on successful exploratory strategies might not be as feasible. In this work, we investigate audio data narratives to improve communication and interpretation of auditory graphs through sonification. A narrative can help contextualize the information and support the user in their interpretation.

3 CO-DESIGN WORKSHOPS

To identify important considerations when creating an audio data narrative for data communication using sonification and descriptions, we conducted two virtual workshops with BVI co-designers who were primarily screen reader users. Our goal was to work with co-designers to lead discussions and generate ideas on how we might make auditory graphs easier to navigate and interpret through a narrative. All workshops were conducted online through the Zoom video conferencing platform and lasted between 90 and 120 minutes.

3.1 Participants

Four co-designers were recruited through snowball sampling to participate in two recurrent group design workshops. All co-designers were working professionals residing in the United States, with an interest in data accessibility. All co-designers identified as blind and/or visually impaired and used screen readers as their primary assistive technology. The median age was 26.5 ($SD = 17.9$, $range = 40$). Two co-designers had a strong preference for tactile graphics when consuming data graphics while the remaining two co-designers had a stronger preference for audio-based methods such as sonification. Additionally, two members of the research team participated as facilitators during both workshops. One facilitator was sighted, and the second facilitator identifies as blind and uses a screen reader as their primary assistive technology.

3.2 Materials & Methods

Before the workshops, the research team met with each co-designer individually to explain the goal of the co-design workshops, answer any questions, and understand any accessibility needs. Physical and digital materials that were used to support different workshop activities were shared with all co-designers one week in advance. Physical materials were mailed to co-designers; these included tactile graphics, tactile prototyping materials (e.g., wikki stix, pipe cleaners). Digital materials included a detailed agenda for each workshop, background information, and sample datasets. For each dataset, we provided three different data representations: a textual description, a tabular representation and a sonified representation.

The first workshop focused on the brainstorming and ideation stages of the design process. The goal of the activities for this workshop were to encourage conversation about different available data representations (tactile, speech, and audio), discuss preferences and tradeoffs between representation types, and to formulate a list of guidelines for when each representation was useful or preferred. The workshop began with co-designers individually familiarizing and exploring each of five different datasets provided through both physical (tactile graphic) and digital representations (tabular data and graph sonification). Co-designers were encouraged to think about the story behind the data and how they might share that story to their peers, including non-experts. After individual exploration participants discussed their insights as a group. Facilitators prompted questions for users to reflect on information available with each of the different data representations and co-designers' preferences based on what they were interested in learning from the data. The workshop concluded with a brainstorming activity where facilitators prompted co-designers to propose prototype ideas to improve the sonification representation. In addition to the materials provided, some co-designers made use of additional software tools: Audacity (audio-editing tool), the SAS Graphics Accelerator [62], Desmos graphing calculator [18] to explore prototypes of the proposed ideas. Throughout the discussions between participants, the facilitator took notes and at the end shared a collective summary of prototype design suggestions based on the co-designers' feedback. We include sample prototypes from the workshops in Supplementary Materials.

Between the first and second workshops, the research team generated prototype alternatives that were discussed in the first workshop. Continuing the conversation over email, facilitators provided a summary of what was learned from the workshop and answers to any of co-designers' questions that could not be answered at the time of the workshop. The second workshop focused on the prototyping stage of the design process. The workshop focused on critiquing the different prototype probes that were generated from the first workshop and prompting co-designers to suggest any improvements. The workshop again concluded with a summary of the main insights learned and discussion of remaining questions. After the workshop, facilitators transcribed observation notes and meeting recordings from both workshops. Open coding was used to organize the data and identify common findings.

3.3 Observations & Findings

Prototypes generated from the workshops primarily explored the use of data sonification and speech narration. Observations from the workshops point to important design considerations. When exploring a complex dataset, co-designers found it most helpful to break down exploration of the graph sonification by segments (**F1**). When sharing dataset insights with the group, co-designers would provide contextual information verbally (e.g., axis values, trend shape) and play smaller relevant segments of the sonification. This exploration strategy, similarly, discussed in prior work with both BVI and sighted users [11, 79], was described by co-designers as being helpful for "*keeping* track of how far along the time-series I am at any given moment". With simpler datasets that conveyed fewer trend reversals or that were cyclical "with no dramatic changes", co-designers suggested "just play the whole series without a break... You can keep track of where you are in the series". The number of trend reversals as well as number of data points has been reported in prior sonification studies as increasing complexity and impacting global integration [15, 54].

Co-designers cautioned on maintaining a balance between too many and too few segments (**F2**); with co-designers suggesting, "determining what length of time a person can retain in memory". Regarding the content of each segment, co-designers suggested the segment should either help identify relevant patterns in the audio or provide a systematic breakdown by time periods depending on the purpose of the visualization (**F3**).

In combining sonification with descriptions, co-designers emphasized the need for descriptions always preceding any presentation of the audio (**F4**), "I feel strongly that the audio [sonification] should never precede the description... otherwise, it's like looking at a graph with no markers on it." Co-designers described the narration as helpful for creating an expectation before listening to the audio [33], "for explaining, identifying specific values that are significant, or making a comment about the general pattern and the significance of it". Co-designers also strongly suggested avoiding any overlap between the speech and sonification (**F5**). Overlapping of the narration with sonification was described as increasing the mental demand for what to focus on, but perhaps at the cost of losing context. One co-designer suggested, overlap should only be used "if there isn't anything interesting [that] you're trying to call out within the series".

As investigated in prior work, co-designers also found it helpful to include rhythmic clicks or beats to mark passing of time along the x-axis [33] (**F6**). This was helpful for providing a more granular marking of time, in addition to the narrative segments. Co-designers had different preferences on sound characteristics (e.g., pitch range, timbre, tempo, etc.), suggesting that these characteristics would be better personalized by the user (**F7**). Though as discussed from prior work some ideal ranges have been reported for parameters such as tempo, pitch, and duration [12, 24] which can be helpful in providing a starting point for novice users.

4 DESIGN PRINCIPLES FOR AUDIO DATA NARRATIVES

We summarize takeaways from the co-design and their connection to prior work to define audio data narratives for time-series data. An audio data narrative is composed of sonification segments and verbal descriptions. Given a time-series dataset, a data narrative should include the following:

D1. The description segment precedes (and does not overlap) the data segment to provide context, structure and set the expectation for the upcoming data sequence. With audio, users' interpretation is influenced by the expectation created by contextual cues [1, 49]. Co-designers described the preceding speech as important to help guide their attention by setting the context for how to interpret the sounds. The description should avoid overlapping with the sonification to mitigate mental load.

D2. The structure of the description segment is consistent across the narrative and describes at minimum the start and end point for the upcoming data sonification segment. Additionally, the description can provide external context to explain the data or highlight key points in the upcoming segment.

D3. The sonification segments maintain consistent trends, maintaining the rhythmic pattern. Co-designers grouped sonification segments based on minimal trend changes. In prior work, the number of trend reversals has also been suggested as the fundamental psychological unit in line graphs [15] as a higher number of trend reversals in a sequence impacts global integration [54]. Temporal resolution changes at the sub-milliseconds level are perceived as pitch changes, while temporal changes at the sub-seconds level are perceived as rhythm [19]. The impact of trend reversals on graph comprehension has been attributed to rhythmic theory that patterned sequences of notes are comprehended more easily than less structured and random combinations [19].

D4. The narrative maintains a moderate number of overall segments. Users can focus on a limited number of items in the overall picture [1]. Too long of an auditory signal and too many trend reversals, increases the number of items to remember [15]. Based on our co-design findings, providing too many segments could also impact workload by requiring the user to constantly switch attention between speech and non-speech sounds.

D5. The sonification segments maintain a moderate duration. Sonification segments should be neither too short nor too long. Target identification in an auditory stream is mediated by top-down processes rather than bottom-up (pre-attentively) [22]. This means that buildup of trends from a stream of tones requires more time, compared to pre-attentive pitch changes in the order of milliseconds. Users need a few seconds (in order of 3 to 5 seconds) to perceive overall changes in rhythm [22]. With an auditory stream that is too short, it might be difficult to extract or understand the pattern in the data. Similarly, an auditory stream that is too long might exceed how much information can be held in working memory. Prior work has suggested keeping the duration of an auditory graph up to 10-12 seconds, with a duration per note of 50-10 msec [53].

5 GENERATING DATA NARRATIVES

Findings from the co-design workshop suggested that data narratives could help BVI users better comprehend auditory graphs and lead to meaningful insights. To support our goal of providing up-to-date and accurate access to the data, we investigated how the process of generating the data narratives could be automated.

We pose the problem as a temporal segmentation problem, where our goal is, given a set of data points, find a narrative composed of speech and sonification segments. Driving the algorithm outcomes are heuristics based on the identified design principles (Section 4). A heuristic approach to generating narratives, as opposed to a more data-driven approach, was more suitable since discussions in the co-design did not necessarily identify a ground truth. More importantly, a heuristic approach allows more customization as well as the possibility of easily defining and including additional heuristics depending on the application context being used. However, a disadvantage is that different heuristics might need to be investigated for different data visualization types whereas a data-driven approach could provide a more general solution. Here we limit our scope to single timeseries datasets. With the defined heuristics, we solve this optimization problem using dynamic programming. In the next subsections, we describe each step and how we define the problem.

5.1 Identifying Boundary Points

To define the narrative segments, we first need to identify where a segment could begin and end. To identify candidate boundary points, we define units ($u_1, \ldots, u_n$), and use the Perceptually Important Points Algorithm (PIP), which provides a reduced set of points that most contribute to the overall shape of the time-series [25, 72]. Units are a set of points in the data that cannot be broken down further ($u_i = [p_i, p_{i+1}]$). Since we want to keep together points that form a consistent pattern, these points are the most likely to be the points where a segment begins or ends.

5.2 Data Segmentation with Dynamic Programming

Next, we want to combine units into segments ($s_1, \ldots, s_m$) that compose an optimal narrative (N_j). Given n units ($u_1, \ldots, u_n$), there are 2^n possible subsets. To solve this problem, we define cost functions based on principles identified in Section 4 and solve this problem using dynamic programming [32]. An optimal narrative will be composed of a collection of segments which maximizes the cost function.

For a given narrative composed of segments ($N_j = s_1, \ldots, s_m$), our algorithm identifies the optimal segment boundaries. Units are processed in order, and the cost of a segment ($s_i = \{u_q, \ldots u_r, \ldots, u_s\}$) is evaluated against the cost of splitting the segment into smaller segments ($s_i = \{u_q, \ldots u_r\},\ s_{i+1} = \{u_{r+1}, \ldots u_s\}$). The algorithm proceeds recursively, keeping track of the best cost as well as the optimal partial narrative. At the end, the optimal narrative will have the lowest cost. We define three cost functions:

C1. Consistency cost. To maintain segments with minimal trend reversals that maintain the rhythmic pattern in the sonification (**D3**), we define a consistency cost proportional to the difference in angle between adjacent units [15].

$$C_{consistency}\left(\mathrm{N}_j\right) = \sum_{i=1}^{m} f(s_i)$$

$$f\left(s_i\right) = f\left(\{u_q, \ldots, u_s\}\right) = \sum_{k=1}^{s-1} g(u_k,\ u_{k+1})$$

$$g\left(u_k,\ u_{k+1}\right) = \begin{cases} 0, & if\ \angle\left(u_k\right) - \angle\left(u_{k+1}\right) < \beta_1 \\ \angle\left(u_k\right) - \angle\left(u_{k+1}\right), & otherwise \end{cases}$$

where $\beta_1 = \pi/9$.

C2. Duration cost. A narrative maintains a moderate duration for each segment (**D5**), not too short (at least 3 seconds) and not too long (ideally within 12 seconds). We calculate the duration of a given segment ($dur(s_i)$) based on the number of data points and sonification tempo. Extremely short segments (below T_{lower}) are strongly penalized since we do not want a segment close to zero. For segments above T_{upper}, we include a linear increase in cost since we do not have a strict cut-off. For segments in the ideal range between T_{lower} and T_{upper}, we define a quadratic loss function.

$$C_{duration}\left(N_j\right) = \sum_{i=1}^{m} C_{duration}\left(s_i\right)$$

$$C_{duration}\left(s_i\right) = \begin{cases} 11^{-dur(s_i) + T_{lower}}, & if \quad dur\left(s_i\right) \leq T_{\mathrm{lower}} \\ \beta_2 \times \left(dur\left(s_i\right) - 0.5T_{upper} + 1.5T_{lower}\right)^2, & \\ if \quad T_{lower} < dur\left(s_i\right) < T_{\mathrm{upper}} & \\ 4 \times dur\left(s_i\right) - 10), & if \quad dur\left(s_i\right) \geq T_{\mathrm{upper}} \end{cases}$$

where $T_{lower} = 3$, $T_{upper} = 12$, and $\beta_2 = 0.075$.

C3. Number of segments cost. To prioritize a moderate number of segments ($num(N)$) in a narrative (N) between $n_{lower} = 2$, and $n_{upper} = 4$ (**D4**), we use a quadratic loss function with a minimum at 3. At minimum we want one segment, so anything below 1 is strongly penalized.

$$C_{number}\left(N_j\right) = \begin{cases} \left[num\left(N_j\right) - 3\right]^2, & if \quad num\left(N_j\right) \geq 1 \\ \infty, & otherwise \end{cases}$$

Final cost. The final cost of a narrative (N_j) is the weighted as the sum of the individual cost functions (C1, C2, C3). For the datasets shown in Figure 2, we use the following hyperparameter values: $\alpha_{consistency} = 0.05$, $\alpha_{number} = 2.0$, and $\alpha_{duration} = 1.0$. We chose these parameters through experimentation.

$$\begin{aligned} C_{total}\left(N_j\right) = {} & \alpha_{consistency} \times C_{consistency}\left(N_j\right) \\ & + \alpha_{duration} \times C_{duration}\left(N_j\right) \\ & + \alpha_{number} \times C_{number}\left(N_j\right) \end{aligned}$$

Applying this segmentation process, Figure 2 shows the results for four different datasets. The simple datasets have 115 and 120 data points while the complex datasets have 270 and 281 data points. Appendix Figure A.1 provides additional samples generated with our approach.

5.3 Sonification Parameters and Generation

With the identified optimal segments, the next step is generating the sonification. For the sonification, we use a frequency mapping paradigm [53]. A piano sound font is used for the timbre, the pitch range is set between MIDI notes 30 to 127, and the tempo is set to 400 beats per minute [12]. We use a positive polarity mapping, such that the dataset maximum value corresponds to the maximum pitch frequency.

5.4 Speech Generation

With identified sonification data segments, the last step is determining the descriptions which precede each data segment (s_i) in the narrative (**D1**). The first description segment always contains the

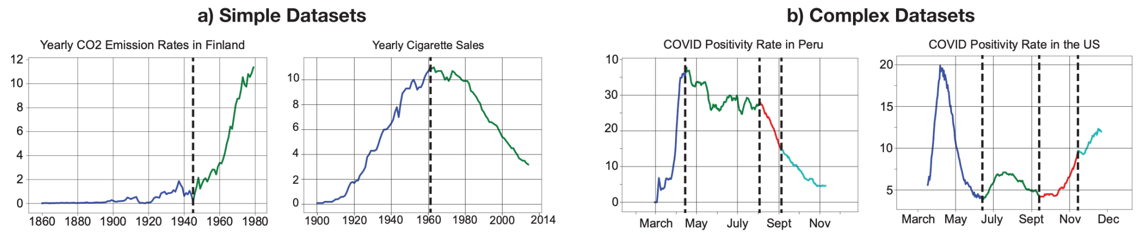

Figure 2: Sample datasets and the segments identified by the algorithm (marked by dashed lines). These datasets were used in the evaluation and included a) Simple datasets with two segments, and b) Complex datasets with four segments.

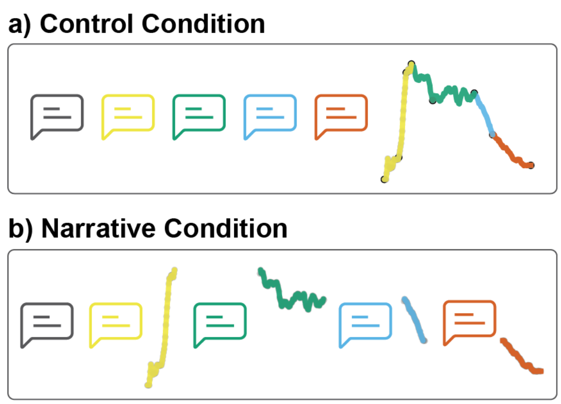

Figure 3: Our evaluation included two conditions. a) In the Control condition, the verbal description was presented all together in one segment followed by one complete segment of the corresponding data sonification. b) In the Narrative condition, the verbal description was interleaved between segments of the corresponding data sonification.

dataset title and a legend that provides context to the maximum and minimum pitch and its corresponding value in the dataset (Figure 1). To compose the description preceding each segment, we use sentence templates. The speech descriptions provide at minimum the value of the start and end points (**D2**). Additionally, if there is a maxima or minima point in the upcoming segment, we include it in the description. The following are two representative speech templates:

- In [time-period-start], [dimension] was [value-start] and then it [slope (increased / decreased)] to [value-end] in [time-period-end].
- In [time-period-start], [dimension] was [value-start] and then it [slope (increased / decreased)] to all-time [high / low] of [value-end] in [time-period-end].

Finally, the sonification and description segments are stitched together in sequential order. Figure 1 shows the narrative output for a dataset on COVID positivity rate. Additional samples are provided in Appendix 1.

6 EVALUATION

We conducted a study to understand the benefits and limitations of data sonification narratives in contrast to the standard approach of presenting auditory graphs through sonification. We were interested on whether data narratives help users better understand the data tones and thus influence the kind of insights users gain.

6.1 Experimental Conditions

The study was conducted as a repeated measures $2x2$ within-subjects study design. A balanced Latin square design was used to reduce order effects. Participants were presented with an auditory graphic in one of two ways (**Factor 1**). In the Control condition (Figure 3a), the verbal description was presented all together in one segment followed by one complete segment of the corresponding data sonification. In the Narrative condition (Figure 3b), the verbal description was interleaved between segments of the corresponding data sonification. This condition evaluated the output from the generation pipeline described in Section 5.

In addition to varying *how* the information was presented, we included datasets from two levels of varying complexity (**Factor 2**). The datasets used for the study are shown in Figure 2. In the Narrative condition, the simple datasets included two segments (115 and 120 data points), while the complex case included four segments (270 and 281 data points). The datasets varied in the number of points and trend reversals present, which prior sonification studies have reported as factors that impact global integration [15, 54].

6.2 Measures

6.2.1 Data insights (quantity, type & quality). To assess what users understood from a gist of the data representation, we asked users to provide a description of the overall trend(s) or pattern(s) in the data and what insights they gained from the data [55]. Prior insight-based evaluations have defined a data insight as "*an individual observation about the data by the participant, a unit of discovery*" [61]. For an initial gist, we asked users to provide an answer after listening to the representation no more than two times.

Additionally, we were interested in making the distinction whether users were simply recalling and repeating information given in the description or generating new insights based on what they understood from both the description and the sonification. Thus, we also coded the quality of an insight as Exact (information provided in the description) or Inferred (not provided in the description). Our assessment protocol is similar to that used by Carswell et al. [15] where participant descriptions are used to assess local and global integration of sonified features in a line graph.

6.2.2 Comprehension. To assess the effectiveness of each representation in specific tasks, we used comprehension questions based on common tasks with time-oriented data [4]. In total we asked four comprehension questions per dataset, and we distinguished between elementary tasks (direct lookup, indirect lookup) and synoptic tasks (pattern lookup, pattern comparison) [3]. The elementary task questions related to individual data values and can be answered from an understanding of just the descriptions. The synoptic tasks require consideration of sets of values of data, where descriptions are not sufficient for answering. For each question, we recorded accuracy and time.

6.2.3 Self-reported ratings & qualitative comments. To assess cognitive load in understanding the information, we measured users' self-perceived mental effort, on a 9-point Paas Likert scale [57]. Additionally, after conclusion of all trials, we asked users' open-ended questions on their strategies, experience, and preferences.

6.3 Procedure

The study was conducted entirely online using the Zoom videoconferencing platform and the Qualtrics survey platform was used for data collection. Studies were scheduled to last up to 90 minutes. On average the study lasted 63.5 minutes ($SD = 20.2$).

After obtaining participation consent, participants were introduced to the study procedures and were provided a brief background explaining what data sonification is. Participants were told they would be asked to listen to different data representations in which the information might be presented differently, and then asked to answer a number of questions. To progress through the study, participants opened a Qualtrics survey which contained all the instructions, trial materials and questions. Participants accessed the survey using their preferred browser. An experimenter was present during the study to remotely guided participants through the survey.

Participants completed one practice trial followed by four experimental trials, each with varying Condition (Control, Narrative) and dataset Complexity (Simple, Complex). For all trials, the data representations were accessible as audio clips using participants' native browser media player. The practice trial was used to familiarize participants with the study procedure, the survey mechanics, and the data representations, and to answer any questions that might arise.

A trial began by asking participants to entirely listen to the data representation to gain a general gist of the information. Participants could listen to the audio up to two times. After this initial listen, based on their recollection, participants provided a few sentences describing the representation and insights or takeaways learned, and completed a set of Likert ratings assessing task mental effort. Participants also rated the accuracy and completeness of their responses. In the second half of the trial participants answered four specific comprehension questions. For these questions, participants could re-visit the representation as many times as they wanted. Participants were told they would be assessed based on both the accuracy and timeliness of their responses. This general procedure was repeated for each of the four trials plus the practice trial. After completion of the practice and experimental trials, participants answered open ended questions.

6.4 Hypotheses

In connection with our study goals and motivated by prior work, we formulated the following study hypotheses:

H1. The narrative representation helps users gain a more complete gist of the data integrating both description and sonification. In the narrative condition, descriptions are provided closer to the relevant sonification segment. Co-designers described this facilitated identifying important patterns in the sonification and grounding them with values provided in the description. For the Narrative condition, we expect this will result in overall more insights gained from the data. Furthermore, we expect insights will rely more on the information contained in both the description and sonification. While for the non-narrative (Control) condition, we expect users will have less insights and rely more on the descriptions alone resulting from a less comprehensive understanding of the sonification. We expect this to also reflect in the comprehension questions. In the Narrative condition, participants will be able to more efficiently recall and answer questions that require understanding of the sonification (synoptic tasks). Regarding dataset Complexity, we expect Complex datasets will result in a higher overall number of insights since there is naturally more information compared to the Simple datasets. However, we expect Simple datasets to have a higher proportion of inferred insights since they are easier to comprehend.

H2. Identifying relevant segments in the sonification will reduce mental effort. Prior work investigating BVI users' exploration strategies of data sonification displays have reported one strategy to gain a gist of the data entails breaking down the segment to better re-investigate and identify relevant patterns [79]. We observed this to be especially helpful to co-designers when the dataset was complex with several trend reversals. In the Narrative condition, relevant patterns are identified and presented to the user as individual segments, thus we expect this will reduce user's mental effort in having to do the work in identifying these segments.

6.5 Participants

Participants were recruited through announcements sent to local and national blindness organization mailing lists. Participant eligibility included being at least 18 years old, residing in the United States, identifying as blind and/or visually impaired, and being a primarily screen reader user. In total, 16 participants took part in this study. Ten participants identified as woman and six participants identified as man. The median age was 29 ($SD = 15.3$, $range = 50$). The primary screen reader used for access was JAWS (9/16), followed by VoiceOver (5/16), and NVDA (2/16). All participants rated their screen reader expertise highly ($\bar{x} = 5$, $SD = 0.97$) on a scale from 1 (*Not familiar at all*) to 6 (*Expert*). Participants rated their expertise interpreting data through tactile charts ($\bar{x} = 3.1$, $SD = 1.5$)

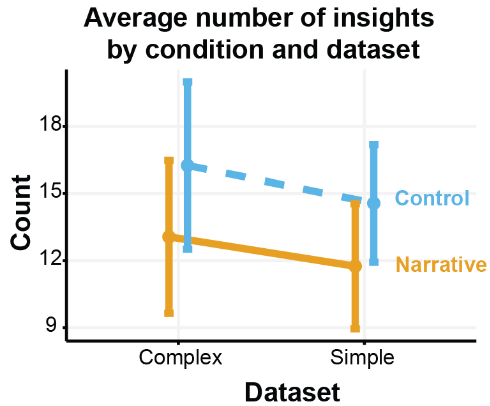

Figure 4: The Narrative condition had a significantly higher number of insights compared to the Control condition. There were no interaction effects with dataset Complexity. Error bars show 95% bootstrap confidence intervals.

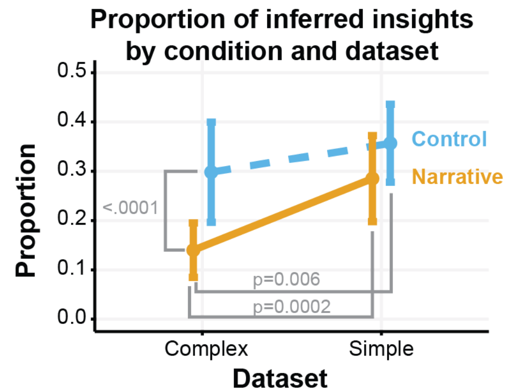

Figure 5: Condition had a significant effect on the proportion of inferred insights with a higher likelihood of an insight being inferred in the Narrative condition compared to the Control. Error bars show 95% bootstrap confidence intervals.

slightly higher than their expertise with auditory graphs ($\bar{x} = 2.2$, $SD = 0.94$) on a scale from 1 (*Not familiar at all*) to 6 (*Expert*).

6.6 Data Analysis

To code participants' insights and descriptions, we developed a codebook using the fact taxonomy described by Law et al. [43] and additional categories described in Vande Moere et al.'s coding process [76]. Our codebook categorized an insight according to the following: Fact type (value, trend, range, extreme, compound fact), Emotional, Rational, and Sound characteristics (pitch, rate). This codebook was initially defined based on five pilot studies and then applied to coding the evaluation data from the sixteen participants. To verify the codebook applicability, two members of the research team first coded a portion of the data (21.7%, 208/957 statements). Before resolving disagreements, we calculated inter-rater reliability (IRR) as the percentage of agreement between raters as 87.5% (182/208), which indicated reasonable agreement [31]. Raters jointly discussed and resolved disagreements and then coded the remaining data independently. Appendix 2.1 defines our codebook and provides examples from the data for each category.

6.7 Results

We present results from our quantitative analysis followed by findings from our qualitative analysis.

6.7.1 Higher number and quality of data insights in Narrative condition. Figure 4 shows the number of insights by Condition and Complexity. Likelihood ratio tests were used to test for the effects of Condition comparing a full model to a restrained model [58]. We fit a generalized linear mixed effects model, using a Poisson distribution, predicting the number of insights (Count). We included fixed effects for Condition (Narrative, Control), Complexity (Simplex, Complex), and their two-way interaction (Condition: Complexity), and a random intercept for each participant.

We find a main effect of Condition on the number of insights provided by participants ($\beta = 0.22$, $SE = 0.067$, $\chi^2(1) = 10.4$, $p = 0.0013$). The average number of insights is higher in the Narrative condition ($\bar{x} = 15.4$, $SD = 6.5$) compared to the Control condition ($\bar{x} = 12.4$, $SD = 6.3$). In the Narrative condition compared to the Control condition, while holding all other variables constant, we expect to have a rate 1.24 times greater in the number of insights provided (Table A.1). There was no significant interaction between Condition and dataset Complexity ($\beta = -0.004$, $SE = 0.13$, $\chi^2(1) = 0.0008$, $p = 0.98$).

Figure 5 shows the proportion of Inferred insights by Condition and Complexity. Using the same analysis procedure as before, we fit a mixed effects logistic regression, predicting the proportion of inferred insights. We find a main effect of Condition on the proportion of inferred insights provided ($\beta = 0.78$, $SE = 0.23$, $\chi^2(1) = 8.15$, $p = 0.004$). The proportion of inferred insights is significantly higher in the Narrative condition ($\bar{x} = 0.33$, $SD = 0.18$) compared to the Control ($\bar{x} = 0.21$, $SD = 0.16$). In the Narrative condition, while holding all other variables constant, the odds are 1.56 times higher that an insight is inferred compared to the Control (Table A.3).

There was also a marginally significant interaction between Condition and dataset Complexity ($\beta = -0.60$, $SE = 0.32$, $\chi^2(1) = 3.65$, $p = 0.05$). Simultaneous pairwise comparisons, adjusting for p-values using Tukey's HSD test indicated the proportion of insights in the Control condition was significantly different between the Simple and Complex datasets ($Z = -1.03$, $p = 0.0002$) but not between the Narrative condition Simple and Complex datasets ($Z = -0.174$, $p = 0.13$). There were also significant differences between the Control Complex and Narrative Complex ($Z = -0.77$, $p = 0.0060$) and the Control Complex and Narrative Simple ($Z = -1.21$, $p < 0.0001$). Table A.4 lists all the contrasts.

Figure 6 shows the distribution of insights by categories. The main distinctions between the Narrative and Control conditions are the higher count of Compound facts and facts describing Trends and Values (X and Y).

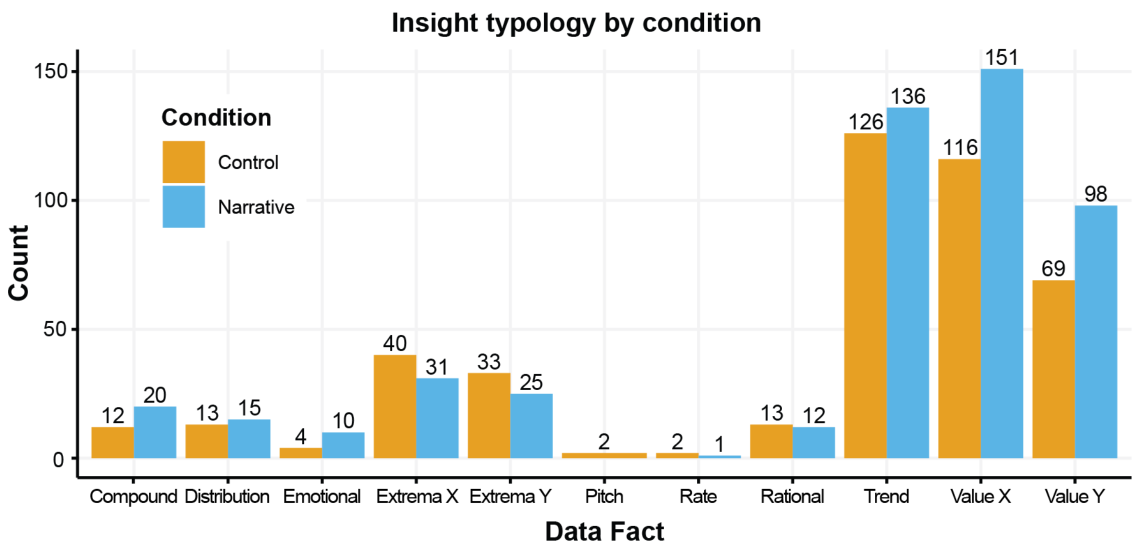

Figure 6: The number of facts that are compound and that describe Trends, and X-Y Values is higher in the Narrative condition compared to the Control.

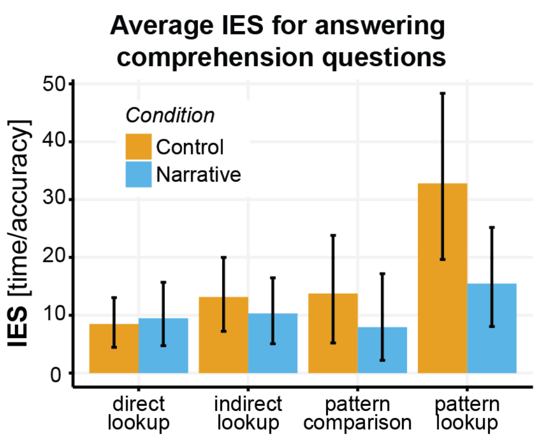

Figure 7: On average, participants were more *inefficient* answering questions in the Control condition compared to the Narrative. A higher IES indicates completion is more inefficient. The pattern lookup questions, which require understanding of both the description and sonification, were most inefficient. Error bars show 95% confidence intervals.

6.7.2 Higher efficiency in the Narrative condition. To assess the effect of Condition on performance in the comprehension questions, we computed the Inverse Efficiency Score (IES) as the time spent answering a comprehension question divided by the accuracy [81]. Since participants may focus on generating an accurate response while sacrificing time, the IES provides us a combined score to assess performance. A higher IES means completion is more inefficient. Following the same analysis procedure as before, we fit a generalized mixed effects model using a Gamma distribution and identity link function, predicting IES (continuous). We included fixed effects for Condition, Complexity and Question type (direct lookup, indirect lookup, pattern comparison, pattern lookup), as well as a random intercept for each participant. We use a Gamma distribution [45], since inspection of our data revealed a non-normal distribution.

Figure 7 shows the average IES score and 95% bootstrap confidence intervals. Controlling for dataset Complexity and Question type, we find a significant interaction between Condition and Question type ($\beta_{indirect\_lookup} = -3.51$, $SE_{indirect\_lookup} = 1.54$, $\beta_{pattern\_lookup} = -11.94$, $SE_{pttern\_lookup} = 4.3$, $\beta_{pattern\_comparison} = -0.61$, $SE_{pattern\_comparison} = 4.3$, $\chi^2(3) = 12.90$, $p = 0.004865$). On average, participants were more inefficient at responding comprehension questions in the Control condition ($\bar{x} = 17.89$, $SD = 28.8$) compared to the Narrative condition ($\bar{x} = 10.81$, $SD = 19.73$). Pairwise contrasts on Condition with Holm-Šidák correction indicated these differences were significantly different depending on the Question type. For the pattern lookup questions ($Z = 11.66, p = 0.0066$), performance was more efficient in the Narrative condition ($\bar{x} = 15.45$, $SD = 24.03$) compared to the Control ($\bar{x} = 32.77$, $SD = 41.38$). An example of this question asked, "*When was the rate of change the fastest?*". Correctly answering these synoptic questions requires integrating an understanding of both the description and sonification. Smaller differences were also found for the indirect lookup questions ($Z = 3.23$, $p = 0.0269$), where performance was slightly better in the Narrative condition ($\bar{x} = 10.28$, $SD = 16.27$) compared to the Control ($\bar{x} = 13.13$, $SD = 18.13$). Table A.6 lists all the contrasts.

6.7.3 High mental effort across conditions. Participants' self-reported rating for the mental effort required to understand each representation and complete the tasks was high across both conditions. Figure 8 shows Likert responses on a 9-point Paas scale (*very very high mental effort* to *very very low mental effort*). In the Control condition the average self-reported mental effort ($\bar{x} = 6.1$, $M = 6$

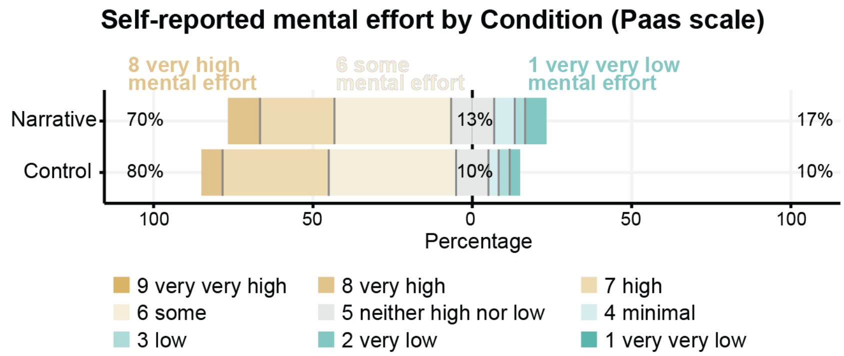

Figure 8: Participants reported high mental effort in both the Narrative and Control conditions.

$SD = 1.31$) was higher than in the Narrative condition ($\bar{x} = 5.8$, $M = 6$, $SD = 1.56$). 80% of participants reported mental effort in the *very high* to *neither high nor low* range compared to 70% in the Narrative condition. A mixed effects model analysis predicting mental effort and controlling for Participant, did not reveal significant differences between these conditions ($\beta = -0.25$, $SE = 0.28$, $\chi^2(1) = 0.81$, $p = 0.368$). However, some qualitative comments from participants suggested the Narrative condition required less effort by the making the information easier to digest (P4, P7, P8, P12, P13) and recall (P6, P9). We discuss users' qualitative feedback in the next sections.

6.7.4 Benefits of integrating description and sonification. Overall, regardless of study condition, all participants were strongly enthusiastic about the complementing benefits of using structured descriptions and sonification. Participants described both "*re-enforced [each] other*" (P6) and provided complementing details (P1, P5, P7, P9, P10, P11, P12, P13). Participants described the sonification provided an easier and better understanding of the trends, verifying what is usually verbally described (P8, P12, P15, P16). Other participants described the sonification as more memorable than just hearing a description (P6, P11, P12). P5 described the benefits of both together saying "*I think the sound and description were really good... It's a verification because with the data you can only focus on so much information but the change in pitch clarifies that. And I think it's really helpful both together.*"

Participants also compared hearing the data narratives to their typical experience encountering descriptions. P11 for example, described how "*listening to the tones is easier to grasp and hold in memory*" compared to descriptions they typically encounter. This could be indicative of the higher cognitive load imposed by text-based representations, as opposed to rendering the information through a direct perceptual interface, when accessing spatial graphics [29]. P7 described how usually when accessing NPR, they read "*the captions and data but it doesn't really give me a whole picture*" compared to having access to both the description and sonification. One participant (P2), however, did not see value in using the sonification and asked, "*if the audio was describing the rates and the percentage and all that stuff, what is the purpose of having the sound?*".

6.7.5 Different understanding between the Narrative and Control conditions. For the Narrative condition, participants described being able to better comprehend the information in more detail, especially the individual trends when compared to the Control condition (P4, P8, P9, P10, P12, P13, P16). P12 emphasized this saying, "*slice it up and then give it to me so I can understand the exact shape of the data.*" These observations explain some of the quantitative findings showing how participants in the Narrative condition performed better in the synoptic tasks (Figure 7) and provided more facts related to trends (Figure 6). Participants described the Narrative allowed them to better keep track of where events were happening in the graph which may also be indicative of the higher number of data facts involving value-x (Figure 6). Several participants qualified their preference for the Narrative representation depending on the dataset and context where they might be consuming the information (P1, P3, P4, P7, P10, P12, P13, P16). Participants described the Narrative as being most helpful for complex datasets with "*many data points and differing trends*" (P16) but less necessary for the simpler datasets. These are also reflected in the quantitative results (Figure 5) and support prior work investigating parameters, including data complexity, which may impact interpretation of an auditory graphic [54].

While the Narrative allowed participants to better contextualize the information and make more inferences based on the data sonification, participants described how it provided a lesser appreciation of the overall sonification (P4, P10). Despite representations used in both the Narrative and Control condition having the same duration, some participants perceived the Narrative as longer in duration. P4 described it as, "*the graph of the overall would be much slower but I could comprehend it more*". Whereas for the Control condition, participants described their understanding as more general, "*more of a big trend understanding... so I can say it was high or low but there was no accurate way to gauge*" (P4).

6.7.6 Adjusting the narrative based on the task & context. After a general gist of the data, several participants were interested in accessing specific details in the data or making adjustments to the narrative (P10, P11, P12, P14, P16). For example, participants wanted to know more exact dates for events they had picked up from the audio or wanted to reduce the narrative to a specific segment to better appreciate the changes (P13, P16). Several participants also wanted to adjust some of the audio parameters such as the speed, timbre, and interval between rhythmic clicks (P10, P11, P12, P14, P16). Additionally, some participants noted the benefits of the Narrative versus Control for use in different contexts (P1, P3, P4, P10, P13, P16). For example, P4 described the sonification would be useful for "*general discussion*" while the narrative might be more useful for "*data analysis or political science class*". These were similar to discussions with co-designers on adjusting the narrative segments and number based on different information-seeking goals. A few participants emphasized that while they had specific preferences, it was important to let the person choose (P3, P9, P10).

7 DISCUSSION & FUTURE WORK

7.1 For complex datasets, the audio narrative representation helps users gain a more complete gist

The evaluation results provide evidence supporting our first hypothesis that the narrative representation helps users gain a more complete gist of the data integrating both the description and sonification. Results show that with the narrative, participants draw more insights from the sonification forming their own interpretation of the data (inferred insights) compared to repeating information explicitly provided in the description. This is particularly the case when the dataset is complex. Results (Section 6.7.1) indicate that for a complex dataset, the Narrative condition resulted in a significantly higher proportion of inferred insights ($\bar{x} = 0.30$, $SD = 0.20$) compared to the Control condition ($\bar{x} = 0.14$, $SD = 0.11$). Whereas when the dataset was simple, the difference between the Narrative ($\bar{x} = 0.36$, $SD = 0.16$) and Control ($\bar{x} = 0.30$, $SD = 0.18$) was higher but not significant. Thus, we see that, especially with complex datasets, audio data narratives can support users in drawing their own insights and gaining a more complete understanding of the data.

Participants described in the Narrative condition they could gain a more detailed understanding, whereas in the Control, they described their understanding as more general. Our narrative generation approach includes a heuristic that aims to minimize the number of trend reversals contained in a sonification segment which might explain some of participants' responses. In a study assessing global integration of sonified line graphs, Carswell et al. reported that interpreting more complex graphs (higher data density and trend reversals) resulted in more global insights at the expense of local detail [15]. These differences indicate the Narrative could be adjusted depending on the dataset complexity. In this evaluation, we considered complexity based on the number of trend reversals and data points. Investigating additional factors that might impact complexity in interpreting an auditory graphics such as noise, symmetry, and variance might also be important to consider [15, 54].

7.2 Reducing the high cognitive load

One downside of our investigated approach is that across conditions, participants self-reported high mental effort (M = 6 on a scale from 1 to 9). We initially hypothesized the Narrative condition would result in lower cognitive load since the narrative helps identify and segment relevant patterns in the sonification. However, participants' self-reported ratings for mental effort were consistently high and no significant differences were observed between conditions. One participant attributed this to the novelty of using sonification. P12 explained that listening to verbal information is "*very common... so it doesn't necessarily require high levels of concentration*" whereas the "*sonification is such a new way of representing information*" it requires conscious attention to "*to combine all the pieces*". Exploring other complementing modalities such as haptics, might help lessen the auditory load and reduce users' cognitive load [78].

Future studies could also investigate whether repeated exposure and greater familiarity from users could reduce the high mental load. Though not a requirement for the study, all participants reported their expertise in using audio graphs as relatively low ($\bar{x} = 2.2$, on a scale from *Not familiar at all* [1] to *Expert* [6]). Nonetheless, with a short introduction and one practice trial, we found most participants were able to gain a comprehensive gist of the data. Providing more ways for users to directly interact with the audio graph could also help reduce the high effort required. In our study, we focused on understanding the benefits of the representation itself and thus offered limited interaction techniques. We discuss further interactions in the next section.

7.3 Exploring additional interactions

In our study, we used real world datasets of relevance and several participants were enthusiastic about being able to understand the data at greater depth when compared to access through typical news channels with just image descriptions or tabular data. All participants were generally enthusiastic about the use of data sonification to complement typical descriptions available with data visualizations. Participants described their potential benefits in providing more comprehensive access to data representations, being able to quickly understand trends, and being able to verify information provided through descriptions. In efforts to address the data accessibility gap [65, 66], data sonification could be more widely integrated with existing image descriptions for data-driven content on the web. The work on data narratives with visual graphics is comparatively extensive. We believe this work demonstrates there are ample research opportunities in similarly extending data narrative patterns and techniques to the auditory domain.

Most evaluation participants also had interest in gaining a greater understanding of specific events in the data, as well as having external context that might explain the data. Co-designers had similar interests, emphasizing opportunities for the data narratives to be more engaging by including relevant external context. In our approach, no external context was included in the descriptions provided. Instead at minimum the description just provided the start and end point values of a segment. Prior works with visual narratives have investigated methods to automatically include relevant annotations that tie external context to a data visualization [36, 69]. These methods could also be applied to audio data narratives to improve the descriptions provided and enhance the narrative by better explaining the data.

Participants accessed the information through their native browser media player and had minimal interaction as they were asked to passively listen. Moreover, the interactions available with the narrative were time-based manipulations (e.g., play/pause, timeline scrolling). Exploring dynamic data visualizations and interactions that support semantic navigation of the narrative might better support data-driven tasks and could address some of participants' feedback discussed in the qualitative results (Section 6.7.6). Prior work has also suggested access to meaningful data insights is strongly reader-specific [47]. Thus, dynamic visualizations that provide users the flexibility to adjust the narrative, both the sound characteristics as well as the details provided in the description, are promising areas of future work. The approach we investigated to generate the data narratives would be able to support these interactions. Depending on different contexts or tasks, it would be possible to add or change heuristics to accommodate each need. Future work could investigate an authoring tool for data narratives and design parameters that might be adjustable by the user for use in different scenarios.

8 LIMITATIONS

There are limitations to consider in the work presented. On the algorithm side, we proposed a heuristics-based approach to generate audio data narratives. We discussed three cost functions and experimentally tuned parameters weighing each function. These parameters may vary depending on the dataset and may need to be adjusted accordingly. Other factors might also impact the weights used (e.g., data density). Furthermore, we implemented a small set of heuristics, but our approach could be extended to include other considerations discussed in the co-design workshops. For example, highlighting additional significant points such as outliers in the data.

We investigated audio data narratives specifically for communicating time-series datasets with only one variable of interest. For multidimensional time-series datasets, the principles investigated could be extended. Investigating interaction methods, as discussed in Future Work (Section 7.3), would likely be critical when presenting multiple variables to mitigate further auditory load on the user. However, some of the principles we applied in our approach might not directly extend to communication of other types of commonly available data visualizations (e.g., scatter plot) and will instead require further investigation.

In compliance with COVID-19 health guidance, we conducted all tasks remotely through a videoconferencing platform for both the co-design workshops and evaluation. Users accessed the audio representations with their own personal device; thus, we were not able to control for the audio quality or environmental noise in participants' location. Participants also had different prior background and experiences with data graphics and the order in which they were presented the different conditions might lead to some practice effects. We aimed to mitigate these effects in our study design and analysis by counterbalancing conditions as well as accounting for participant as a random effect in our statistical analysis modelling. We also used a qualitative insight analysis process to evaluate participants findings from the data. North et al. discusses some of the difficulties with these methods for assessing visualizations, including the greater variance in results compared to more controlled benchmark tasks [55].

In proposing audio data narratives to increase access to data, we focused on addressing the needs of BVI screen reader users. Our investigation was largely motivated by insights gained from formative work with four BVI co-designers. With a limited participant group, we might not capture the diverse accessibility needs in the broad continuum of visual conditions and abilities in this population [27]. Furthermore, while the approach we investigated may improve access for some users, it may be entirely inaccessible to others. Marriot et al. review the current state of access to visualization and discuss challenges across three disability groups (visual, cognitive, and motor impairments) that affect access to visualization [48]. Focusing on visual disabilities and relying on primarily auditory perception, our approach to data narratives may be inaccessible to others such as users with hearing loss or even make data interpretation more difficult for these users. Thus, in aiming for greater access and equity for all users, it is important to consider access in a more holistic context.

9 CONCLUSION

We have discussed one approach to improve consumption of auditory graphs through audio data narratives which interleave segments of speech description and data sonification. We have focused specifically on communicating time-series datasets. Informed by prior work and a series of co-design workshops with BVI users, we summarized design principles for audio data narratives. We applied these in the development of a heuristics-based algorithm for generating data narratives given a time-series dataset. To validate our approach, we conducted a user evaluation with sixteen BVI screen reader users exploring the benefits of data narratives in helping users gain insights from the data. Our evaluation compared different metrics between a Narrative condition with segments to a Control condition without segments. Our findings show that consuming the information in narrative form helps BVI screen reader users gain more insights that integrate both description and sonification. Like consumption of visual data graphics, consumption of audio data graphics can also benefit from effective narrative techniques that help guide the reader. Our work shows that audio data narratives can support screen reader users in forming their own interpretation of the data, promoting independent and equitable access to data.

ACKNOWLEDGMENTS

This material is based upon work supported by the National Science Foundation under Grants DGE-2016789 and 2016363. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation. We thank Dae Hyun Kim for feedback and Peter Saathoff-Harshfield and Frank Welte for their advice and support on conducting the remote workshops and in creating accessible study materials.

A APPENDICES

A GENERATING DATA NARRATIVES

We provide additional sample outputs created from our proposed generation algorithm described in Section 5. Datasets were collected from Our World in Data and the Federal Reserve Economic Data (FRED).

A.1 Description segments for the datasets shown in Figure 2

2 shows samples of the data segmentation through dynamic programming. Below are the accompanying descriptions for each segment, including the title and legend.

Yearly CO_2 Emission Rates in Finland. During the past 119 years between 1860 and 1979, the emission rate was lowest in January 1860 at 0%. Represented by this sound [*notes*]. And was the highest in January 1979 at 11.4% [*notes*]. Tick sounds mark 10-year intervals. In January 1860, emission rate was 0% and then it increased slowly to 0.2% in January 1945 [*notes*]. In January 1945, emission rate was 0.2% and then it increased sharply to all time high of 11.4% in January 1979 [*notes*].

Yearly Cigarette Sales. During the past 114 years between 1900 and 2014, the sales was lowest in January 1900 at 0.1%. Represented by this sound [*notes*]. And was the highest in January 1961 at 11% [*notes*]. Tick sounds mark 10-year intervals. In January 1900, sales was 0.1% and then it increased slowly to all time high of 11.0% in January 1961 [*notes*]. In January 1961, sales was 11% and then it decreased slowly to 3.2% in January 2014 [*notes*].

COVID Positivity Rate in Peru. During the past year between February and November, the COVID rate was lowest in February 2020 at 0%. Represented by this sound [*notes*]. And was the highest in April 2020 at 37.7% [*notes*]. Tick sounds mark monthly intervals. In February 2020, COVID rate was 0% and then it increased sharply to all time high of 37.7% in April [*notes*]. In April 2020, COVID rate was 37.7% and then it decreased more slowly to 28.2% in August [*notes*]. In August 2020, COVID rate was 28.2% and then it decreased rapidly to 14.6% in September [*notes*]. In September 2020, COVID rate was 14.6% and then it decreased more slowly to 4.6% in November [*notes*].

COVID Positivity Rate in the US. During the past year between March and December, the COVID rate was highest in April 2020 at 19.9%. Represented by this sound [*notes*]. And was the lowest in June 2020 at 3.8% [*notes*]. Tick sounds mark monthly intervals. In March 2020, COVID rate was at 5.6%, when it sharply increased to all time high of 19.9% in April then decreased back to 3.8% in June [*notes*]. In June 2020, COVID rate was at 3.8%, then it increased slowly then decreased to 4.2% in September [*notes*]. In September 2020, COVID rate was 4.2% and then it increased slowly to 9.7% in November [*notes*]. In November 2020, COVID rate was 9.7% and then it increased slowly to 12.0% in December 2020 [*notes*].

A.2 Results from the Segmentation Algorithm

Figure A.1 shows additional datasets and results from the segmentation algorithm. Segments are visually marked by both different colors and dashed lines.

B EVALUATION

B.1 Codebook

A codebook was developed to code the insights participants provided during the evaluation. The codebook uses the fact taxonomy described by Law et al. [43] and additional categories described in Vande Moere et al.'s insight coding process [76]. There are three possible categories: data fact, rational, and emotional. These are defined as follows:

- Rational: An observation that contains some reasoning, such as 'why' an event in the data might have occurred [76].
- Emotional: An observation that contains a subjective interpretation [76].
- Data Fact: one of seven possible data descriptors (value [x, y], extrema [x, y], trend, range, outlier, compound fact, sonification fact) [43]. Sonification quality was a category specifically added for our study and describes references to the sonification rate and/or pitch qualities.

For each category, below provide examples collected from participants during the evaluation:

1. Rational:

- "In 1860 maybe they didn't have many cars or the stuff that would put carbon because there was none."
- "Despite the fact that people know that there is a link between cancer and cigarette smoking, people still continue to do it, but the trends have really gone down."
- "It kind of did what I would expect that in the 50s and 60s people were smoking a lot, so it made sense for it to go up. But then it started going down because we learned about the dangers of nicotine and all that, so sales went back down."
- "I think that goes to show that was the time period we were under mandatory shelter in place and that seems to have been effective."
- "Contrary to popular belief COVID rates were in fact much lower during the shutdown."

2. Emotional:

- "What surprised me was how long it took of the number of cases to decline."
- "I'm so glad that it has gotten lower because I hate cigarettes."

3. Value (X, Y):

- "In November, it was 4.2%."
- "In December it went back to 12%."

4. Extrema (X, Y):

- "It had some spikes with a high in 1961."
- "Then it drops to the all-time low in June at 3.4%."

5. Trend:

- "For the first few decades, I would say 3 or 4 decades, it was a very slow increase, or I guess barely any increase."
- "1945 onwards, for a short amount of period on the x-axis there is a sharp rise in the y coordinate."
- "There was a gradual increase starting from 0% in 1860 to 0.2%."

6. Range:

- "It was the COVID rate in the United States between April and December."
- "The data spanned from January of 1860 to January of 1979."

7. Compound fact:

- "It decreased more rapidly between April and August of 2020 compared to like between August and November."
- "Early on in the beginning it was a lot higher and decreased more slowly over time."

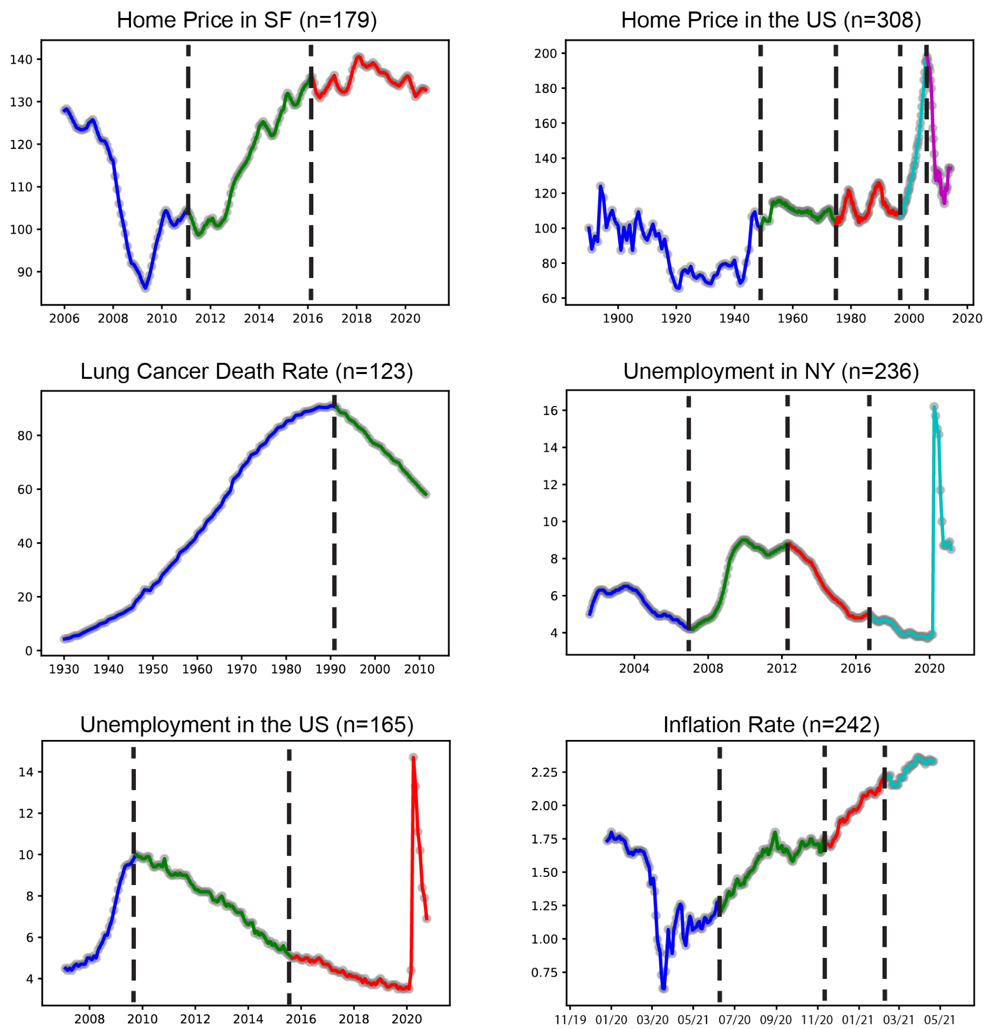

Figure 9: Results from the segmentation through dynamic programming.

- "How much faster it went up than it was able to come back down again overall."
- "I think it's interesting that between 1945 and 1979, the rate jumps so quickly and in that 24-year span, as opposed to the 115 years before that, and I would wonder why that was."

8. Sonification fact:

- "The data consisted of sounds, piano sounds that were rising in pitch as the CO_2 level increased."
- "The kick sounds represented every 10-year intervals."

B.2 Results: Poisson Mixed Effects Model Predicting the Number of Insights

Likelihood ratio tests were used to test for the effects of Condition comparing a full model to a restrained model [58]. We fit a generalized linear mixed effects model, using a Poisson distribution, predicting the number of insights (Count). We included fixed effects for Condition (Narrative, Control), Complexity (Simplex, Complex), and their two-way interaction (Condition:Complexity), and a random intercept for each participant. Table A.1 shows the regression model and estimates. Table A.2 lists the sample mean and SD.

B.3 Results: Logistic Mixed Effects Model Predicting the Proportion of Inferred Insights

Likelihood ratio tests were used to test for the effects of Condition comparing a full model to a restrained model [58]. We fit a logistic mixed effects regression, predicting the proportion of inferred insights. We included fixed effects for Condition (Narrative, Control), Complexity (Simplex, Complex), and their two-way interaction (Condition: Complexity), and a random intercept for each participant. Table A.3 shows the regression model estimates. Table shows pairwise contrasts between Condition*Complexity calculated using Tukey's HSD. Table A.5 lists the sample mean and SD.

Table 1: Mixed effects model predicting the *count of insights*

	Restrained Model			Full Model		
Predictors	*Incidence Rate*	*CI*	*p*	*Incidence Rate*	*CI*	*p*
(Intercept)	13.54	10.78–17.0	<0.001	12.08	9.51–15.34	<0.001
Complexity [Simple]	0.90	0.79–1.02	0.11	0.90	0.79–1.02	0.11
Condition [Narrative]				**1.24**	**1.09–1.42**	**0.001**

Table 2: Mean and SD for the number of insights by condition and dataset complexity

Condition	Complexity	Mean	SD
Control	Complex	13.1	6.9
	Simple	11.8	5.7
Narrative	Complex	16.2	7.6
	Simple	14.6	5.4

Table 3: Logistic mixed effects model predicting the *proportion of inferred insights*

	Restrained Model			Full Model			Full Model + Interaction		
Predictors	*Odds Ratio*	*CI*	*p*	*Odds Ratio*	*CI*	*P*	*Odds Ratio*	*CI*	*p*
(Intercept)	0.27	0.20–0.37	<0.001	0.21	0.14–0.30	<0.001	0.17	0.11–0.26	<0.001
Complexity	1.94	1.43–2.61	<0.001	1.94	1.44–2.63	<0.001	2.80	1.72–4.57	<0.001
Condition				**1.56**	**1.15–2.12**	**0.005**	2.17	1.36–3.47	0.001
Interaction							**0.55**	**0.29–1.02**	**0.057**

Table 4: Pairwise Contrast using Tukeys HSD for *Condition: Complexity*

Contrast	*Estimate*	*SE*	*z-ratio*	*P*
Control Complex – Narrative Complex	-0.777	0.238	-3.266	0.0060
Control Complex – Control Simple	-1.031	0.249	-4.145	0.0002
Control Complex – Narrative Simple	-1.205	0.237	-5.084	<0.0001
Narrative Complex – Control Simple	-0.254	0.212	-1.199	0.6273
Narrative Complex – Control Simple	-0.428	0.197	-2.176	0.1298
Control Simple – Narrative Simple	-0.174	0.210	-0.827	0.8416

B.4 Results: Linear Mixed Effects Model Predicting Inverse Efficiency Score

Likelihood ratio tests were used to test for the effects of Condition comparing a full model to a restrained model [58]. We fit a generalized mixed effects model using a Gamma distribution and identity link function, predicting IES (continuous). We included fixed effects for Condition, Complexity and Question type (direct lookup, indirect lookup, pattern comparison, pattern lookup), as well as a random intercept for each participant. We use a Gamma distribution [45], since inspection of our data revealed a non-normal

Table 5: Mean and SD for the proportion of inferred insights by condition and dataset complexity

Condition	Complexity	Mean	SD
Control	Complex	0.14	0.11
	Simple	0.28	0.18
Narrative	Complex	0.30	0.21
	Simple	0.36	0.16

Table 6: Pairwise Contrasts with Holm-Šidák Correction for *Condition: Question*

Question	*Condition*	*Mean*	*SE*	*P*
Direct lookup	Control	9.56	2.71	0.6805
	Narrative	9.84	2.71	
Indirect lookup	Control	12.37	2.98	0.0269
	Narrative	9.14	2.69	
Pattern Comparison	Control	9.88	2.71	0.6149
	Narrative	9.56	2.71	
Pattern Lookup	Control	25.53	4.70	0.0066
	Narrative	13.87	3.03	

Table 7: Mean and SD for IES by condition and dataset question

Condition	Question	Mean	SD
Control	Direct lookup	8.5	12.4
	Indirect lookup	13.1	18.1
	Pattern comparison	13.8	27.4
	Pattern lookup	32.8	41.4
Narrative	Direct lookup	9.5	15.3
	Indirect lookup	10.3	16.3
	Pattern comparison	7.9	22.1
	Pattern lookup	15.4	24.0

distribution. Pairwise contrasts on Condition with Holm-Šidák correction indicated differences were significantly different depending on the Question type. Table A.6 lists the means and contrasts.

Table A.7 lists the sample mean and SD.

REFERENCES

[1] James L Alty and Dimitrios Rigas. 2005. Exploring the use of structured musical stimuli to communicate simple diagrams: the role of context. International journal of human-computer studies, 62(1):21–40.

[2] Dragan Ahmetovic, Cristian Bernareggi, João Guerreiro, Sergio Mascetti, and Anna Capietto. 2019. Audiofunctions. web: Multimodal exploration of mathematical function graphs." In Proceedings of the 16th International Web for All Conference, pp. 1-10.

[3] Wolfgang Aigner, Silvia Miksch, Heidrun Schumann, and Christian Tominski. Visualization of time-oriented data. Springer Science & Business Media, 2011.

[4] Natalia Andrienko and Gennady Andrienko. 2006. Exploratory analysis of spatial and temporal data: a systematic approach. Springer Science & Business Media.

[5] Robert Amar, James Eagan, and John Stasko. 2005. Low-level components of analytic activity in information visualization. In IEEE Symposium on Information Visualization. INFOVIS 2005., pages 111–117. IEEE.

[6] Apple Audio Graphs API. 2021. Retrieved August 20, 2021 from https://developer.apple.com/documentation/accessibility/audio_graphs

[7] Benjamin Bach, Moritz Stefaner, Jeremy Boy, Steven Drucker, Lyn Bartram, Jo Wood, Paolo Ciuccarelli, Yuri Engelhardt, Ulrike Koeppen, and Barbara Tversky. 2018. Narrative design patterns for data-driven storytelling. In Data-driven storytelling, pp. 107-133. AK Peters/CRC Press.

[8] Michael Behrisch, Michael Blumenschein, Nam Wook Kim, Lin Shao, Mennatallah El-Assady, Johannes Fuchs, Daniel Seebacher, Alexandra Diehl, Ulrik Brandes, Hanspeter Pfister. 2018. Quality metrics for information visualization. In Computer Graphics Forum, volume 37, pages 625–662. Wiley Online Library.

[9] Albert S Bregman. 1994. Auditory scene analysis: The perceptual organization of sound. MIT press.

[10] Matthew Brehmer, and Tamara Munzner. 2013. A multi-level typology of abstract visualization tasks. IEEE transactions on visualization and computer graphics, 19(12), pp.2376-2385.

[11] Lorna M. Brown, Stephen Brewster, Ramesh Ramloll, Wai Yu, and Beate Riedel. 2002. Browsing modes for exploring sonified line graphs. In Proceedings of the British Human-Computer Interface Conference. 6–9.

[12] Lorna M. Brown, Stephen A. Brewster, S. A. Ramloll, R. Burton, and Beate Riedel. 2003. Design guidelines for audio presentation of graphs and tables. International Conference on Auditory Display (ICAD).

[13] Matthew Butler, Leona M Holloway, Samuel Reinders, Cagatay Goncu, and Kim Marriott. 2021. Technology developments in touch-based accessible graphics: A systematic review of research 2010-2020. In Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems, pages 1–15.

[14] Stuart T. Card, Jock D. Mackinlay, and Ben Scheiderman. 1999. Readings in Information Visualization, using vision to think. San Francisco: Morgan Kaufmann.

[15] C Melody Carswell, Cathy Emery, and Andrea M Lonon. 1993. Stimulus complexity and information integration in the spontaneous interpretations of line graphs. Applied Cognitive Psychology, 7(4):341–357.

[16] Chartability. Retrieved August 20, 2021 from https://chartability.fizz.studio

[17] Jinho Choi, Sanghun Jung, Deok Gun Park, Jaegul Choo, and Niklas Elmqvist. 2019. Visualizing for the non-visual: Enabling the visually impaired to use visualization. In Computer Graphics Forum, volume 38, pages 249–260. Wiley Online Library.

[18] Desmos Graphing Calculator. Retrieved August 20, 2021 from https://www.desmos.com/calculator

[19] Diana Deutsch. 1980. The processing of structured and unstructured tonal sequences. Perception & psychophysics, 28(5):381–389.

[20] Diana Deutsch. 1999. Grouping mechanisms in music. In The psychology of music, pages 299–348. Elsevier.

[21] Wanda L. Diaz-Merced, Robert M. Candey, Nancy Brickhouse, Matthew Schneps, John C. Mannone, Stephen Brewster, and Katrien Kolenberg. 2011. Sonification of astronomical data. Proceedings of the International Astronomical Union, 7(S285), 133-136.

[22] Mounya Elhilali, Juanjuan Xiang, Shihab A. Shamma, and Jonathan Z. Simon. 2009. Interaction between attention and bottom-up saliency mediates the representation of foreground and back- ground in an auditory scene. PLoS biology, 7(6):e1000129.

[23] Leo Ferres, Avi Parush, Shelley Roberts, and Gitte Lindgaard. 2006. Helping people with visual impairments gain access to graphical information through natural language: The i-Graph system. In International Conference on Computers for Handicapped Persons, pages 1122–1130. Springer.

[24] John H. Flowers, Laura E. Whitwer, Douglas C. Grafel, and Cheryl A. Kotan. 2001. Sonification of daily weather records: Issues of perception, attention and memory in design choices. Faculty Publications, Department of Psychology, page 432.

[25] Tak-chung Fu, Fu-lai Chung, Robert Luk, and Chak-man Ng. 2008. Representing financial time series based on data point importance. Engineering Applications

of Artificial Intelligence, 21(2):277–300.

[26] Steven L. Franconeri, Lace M. Padilla, Priti Shah, Jeffrey M. Zacks, and Jessica Hullman. The science of visual data communication: What works. Psychological Science in the Public Interest 22, no. 3 (2021): 110-161.

[27] Nicholas A. Giudice. 2018. Navigating without vision: Principles of blind spatial cognition. In Handbook of behavioral and cognitive geography. Edward Elgar Publishing.

[28] Cole Gleason, Amy Pavel, Emma McCamey, Christina Low, Patrick Carrington, Kris M Kitani, and Jeffrey P Bigham. 2020. Twitter a11y: A browser extension to make twitter images accessible. In Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems, pages 1–12.

[29] Jenna L. Gorlewicz, Jennifer L. Tennison, Hari P. Palani, and Nicholas A. Giudice. 2018. The graphical access challenge for people with visual impairments: Positions and pathways forward. In Interactive Multimedia-Multimedia Production and Digital Storytelling. IntechOpen.

[30] Bryan Gould, Trisha O'Connell, and Geoff Freed. 2008. Effective practices for description of science content within digital talking books.

[31] Donald P Hartmann. 1977. Considerations in the choice of interobserver reliability estimates. Journal of applied behavior analysis 10, no. 1: 103-116.

[32] Michael Held, and Richard M. Karp. 1962. A dynamic programming approach to sequencing problems. Journal of the Society for Industrial and Applied mathematics 10, no. 1: 196-210.

[33] Thomas Hermann. 2002. Sonification for exploratory data analysis. PhD thesis, Bielefeld University.

[34] Thomas Hermann, Andy Hunt, and John G. Neuhoff. 2011. The sonification handbook. Logos Verlag Berlin.

[35] Leona Holloway, Matthew Butler, Samuel Reinders, and Kim Marriott. 2020. Non-visual access to graphical information on COVID-19. In The 22nd International ACM SIGACCESS Conference on Computers and Accessibility. 1–3.

[36] Jessica Hullman, Nicholas Diakopoulos, and Eytan Adar. 2013. Contextifier: automatic generation of annotated stock visualizations. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '13). Association for Computing Machinery, New York, NY, USA, 2707–2716. DOI:https://doi.org/10.1145/2470654.2481374

[37] Johan Kildal, and Stephen A. Brewster. 2006. Providing a size-independent overview of non-visual tables. In Proceedings of the International Conference on Auditory Display (ICAD).

[38] Edward Kim, and Kathleen F. McCoy. 2018. Multimodal deep learning using images and text for information graphic classification. In Proceedings of the 20th International ACM SIGACCESS Conference on Computers and Accessibility (pp. 143-148).

[39] N. W. Kim, S. C. Joyner, A. Riegelhuth, and Y Kim. 2021. Accessible visualization: Design space, opportunities, and challenges. In Computer Graphics Forum, volume 40(3), pages 173–188. Wiley Online Library.

[40] Hirohito M Kondo, Anouk M van Loon, Jun-Ichiro Kawahara, and Brian CJ Moore. 2017. Auditory and visual scene analysis: an overview. Philosophical Transactions of the Royal Society B: Biological Sciences, 372(1714):20160099.

[41] Heidi Lam, Enrico Bertini, Petra Isenberg, Catherine Plaisant, and Sheelagh Carpendale. 2011. Empirical studies in information visualization: Seven scenarios. IEEE transactions on visual- ization and computer graphics, 18(9):1520–1536.

[42] Mark Last, and Anna Gorelik. 2008. Using sonification for mining time series data. In Proceedings of the 9th International Workshop on Multimedia Data Mining: held in conjunction with the ACM SIGKDD (August 2008), pp. 63-72. 2008.

[43] Po-Ming Law, Alex Endert, and John Stasko. 2020. Characterizing automated data insights. In 2020 IEEE Visualization Conference (VIS), pages 171–175. IEEE.

[44] Po-Ming Law, Alex Endert, and John Stasko. 2020. What are Data Insights to Professional Visualization Users?. In 2020 IEEE Visualization Conference (VIS), pp. 181-185. IEEE Computer Society.

[45] Steson Lo, and Sally Andrews. 2015. To transform or not to transform: Using generalized linear mixed models to analyse reaction time data. Frontiers in psychology 6: 1171.

[46] Andrew Lotto and Lori Holt. 2011. Psychology of auditory perception. Wiley Interdisciplinary Reviews: Cognitive Science, 2(5):479–489.

[47] Alan Lundgard, and Arvind Satyanarayan. 2021. Accessible Visualization via Natural Language Descriptions: A Four-Level Model of Semantic Content. IEEE transactions on visualization and computer graphics.

[48] Kim Marriott, Bongshin Lee, Matthew Butler, Ed Cutrell, Kirsten Ellis, Cagatay Goncu, Marti Hearst, Kathleen McCoy, and Danielle Albers Szafir. 2021. Inclusive data visualization for people with disabilities: a call to action. Interactions 28, no. 3: 47-51.

[49] George A. Miller, George A. Heise, and William Lichten. 1951. The intelligibility of speech as a function of the context of the test materials. Journal of experimental psychology 41, no. 5: 329.

[50] Valerie S Morash, Yue-Ting Siu, Joshua A Miele, Lucia Hasty, and Steven Landau. 2015. Guiding novice web workers in making image descriptions using templates. ACM Transactions on Accessible Computing (TACCESS), 7(4):1–21.

[51] Tomas Murillo-Morales and Klaus Miesenberger. 2020. Audial: A natural language interface to make statistical charts accessible to blind persons. In International Conference on Computers Helping People with Special Needs, pages 373–384. Springer.

[52] Michael A. Nees and Bruce N. Walker. 2006. Relative intensity of auditory context for auditory graph design. In Proceedings of the International Conference on Auditory Display (ICAD).

[53] Michael A. Nees and Bruce N. Walker. 2007. Listener, task, and auditory graph: Toward a conceptual model of auditory graph comprehension. In Proceedings of the International Conference on Auditory Display (ICAD).

[54] Michael A. Nees and Bruce N Walker. 2008. Data density and trend reversals in auditory graphs: Effects on point-estimation and trend-identification tasks. ACM Transactions on Applied Perception (TAP), 5(3):1–24.

[55] Chris North. 2006. Toward measuring visualization insight. IEEE computer graphics and applications 26, no. 3: 6-9.

[56] Keita Ohshiro, Amy Hurst, and Luke DuBois. 2021. Making Math Graphs More Accessible in Remote Learning: Using Sonification to Introduce Discontinuity in Calculus. In The 23rd International ACM SIGACCESS Conference on Computers and Accessibility (pp. 1-4).

[57] Fred G. W. C. Paas. 1992. Training strategies for attaining transfer of problem-solving skill in statistics: a cognitive-load approach. Journal of educational psychology 84, no. 4 (1992): 429.

[58] Nicolaas Prins. 2018. Applying the model-comparison approach to test specific research hypotheses in psychophysical research using the Palamedes toolbox. Frontiers in psychology 9: 1250.

[59] Rameshsharma Ramloll, Stephen Brewster, Wai Yu, and Beate Riedel. 2001. Using non-speech sounds to improve access to 2D tabular numerical information for visually impaired users. In People and Computers XV—Interaction without Frontiers, pp. 515-529. Springer, London.

[60] Nathalie Henry Riche, Christophe Hurter, Nicholas Diakopoulos, and Sheelagh Carpendale, eds. Data-driven storytelling. CRC Press, 2018.

[61] Purvi Saraiya, Chris North, and Karen Duca. 2005. An insight-based methodology for evaluating bioinformatics visualizations. IEEE transactions on visualization and computer graphics, 11(4), 443-456.

[62] SAS Graphics Accelerator. Retrieved August 20, 2021 from http://support.sas.com/software/products/graphics-accelerator/index.html

[63] Nik Sawe, Chris Chafe, and Jeffrey Treviño. 2020. Using Data Sonification to Overcome Science Literacy, Numeracy, and Visualization Barriers in Science Communication. Frontiers in Communication 5 (2020): 46.

[64] Edward Segel, and Jeffrey Heer. 2010. Narrative visualization: Telling stories with data. IEEE transactions on visualization and computer graphics 16, no. 6: 1139-1148.

[65] Ather Sharif, Sanjana Chintalapati, Jacob O. Wobbrock, and Katharina Reinecke. 2021. Understanding screen-reader users' experiences with online data visualizations. Proceedings of the ACM Conference on Computers and Accessibility (ASSETS '21). Virtual Event (October 18-22, 2021). New York: ACM Press. To appear.

[66] Alexa F. Siu, Danyang Fan, Gene S-H Kim, Hrishikesh V. Rao, Xavier Vazquez, Sile O'Modhrain, and Sean Follmer. 2021. COVID-19 highlights the issues facing blind and visually impaired people in accessing data on the web. In Proceedings of the 18th International Web for All Conference, pp. 1-15. 2021.

[67] Abigale Stangl, Nitin Verma, Kenneth R. Fleischman, Meredith Ringel Morris, and Danna Gurari. 2021. Going Beyond One-Size-Fits-All Image Descriptions to Satisfy the Information Wants of People Who are Blind or Have Low Vision. To appear.

[68] Tony Stockman, Christopher Frauenberger, and Greg Hind. 2005. Interactive sonification of spreadsheets. In Proceedings of the International Conference on Auditory Display (ICAD).

[69] Charles D. Stolper, Bongshin Lee, Nathalie Henry Riche, and John Stasko. 2016. Emerging and recurring data-driven storytelling techniques: Analysis of a curated collection of recent stories.

[70] Sarinah Sutojo, Joachim Thiemann, Armin Kohlrausch, and Steven van de Par. 2020. Auditory gestalt rules and their application. The Technology of Binaural Understanding, pages 33–59.

[71] Bruce N. Walker and Lisa M. Mauney. 2010. Universal design of auditory graphs: A comparison of sonification mappings for visually impaired and sighted listeners. ACM Transactions on Accessible Computing (TACCESS), 2(3):1–16.

[72] Yuqing Wan, Xueyuan Gong, and Yain-Whar Si. 2016. Effect of segmentation on financial time series pattern matching. Applied Soft Computing, 38:346–359.

[73] Zezhong Wang, Shunming Wang, Matteo Farinella, Dave Murray-Rust, Nathalie Henry Riche, and Benjamin Bach. 2019. Comparing effectiveness and engagement of data comics and infographics. In Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems, pp. 1-12.

[74] Web Content Accessibility Guidelines (WCAG) 2.1. Retrieved January 14, 2021 from https://www.w3.org/TR/WCAG21/

[75] Shaomei Wu, Jeffrey Wieland, Omid Farivar, and Julie Schiller. 2017. Automatic alt-text: Computer-generated image descriptions for blind users on a social network service. In Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing, pages 1180–1192.

[76] Andrew Vande Moere, Martin Tomitsch, Christoph Wimmer, Boesch Christoph, and Thomas Grechenig. 2012. Evaluating the effect of style in information visualization. IEEE transactions on visualization and computer graphics, 18(12):2739–2748.

[77] Karen Vines, Chris Hughes, Laura Alexander, Carol Calvert, Chetz Colwell, Hilary Holmes, Claire Kotecki, Kaela Park, and Victoria Pearson. 2019. Sonification of numerical data for education, Open Learning: The Journal of Open, Distance and e-Learning, 34:1, 19-39, DOI: 10.1080/02680513.2018.1553707

[78] Wai Yu, and Stephen Brewster. 2002. Comparing two haptic interfaces for multimodal graph rendering. In Proceedings 10th Symposium on Haptic Interfaces for Virtual Environment and Teleoperator Systems. HAPTICS 2002, pp. 3-9. IEEE.

[79] Haixia Zhao. 2006. Interactive Sonification of Abstract Data-Framework, Design Space, Evaluation, and User Tool. PhD thesis, University of Maryland (College Park, Md.).

[80] Haixia Zhao, Catherine Plaisant, Ben Schneiderman, and Jonathan Lazar. 2008. Data sonification for users with visual impairment: a case study with georeferenced data. ACM Transactions on Computer-Human Interaction (TOCHI), 15(1), 1-28.

[81] Qiyu Zhi, Alvitta Ottley, and Ronald Metoyer. 2019. Linking and layout: Exploring the integration of text and visualization in storytelling. In Computer Graphics Forum, volume 38(3), pages 675–685. Wiley Online Library.