Audio-Visual Data Analytics

Towards a Design Theory

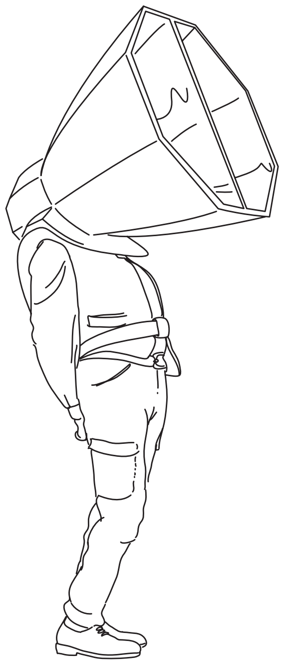
PhD Thesis

Kajetan Enge
Matriculation Number: 1230641

University of Music and Performing Arts Graz
PhD program: Sound and Music Computing

Supervisors:
O.Univ.Prof. Mag.art. DI Dr.techn. Robert Höldrich
Univ.Prof. Dipl.-Ing. Dr.techn. Alois Sontacchi
FH-Prof. Priv.-Doz. Dipl.-Ing. Dr. Wolfgang Aigner, MSc

December 2024



“Hearing is a way of touching at a distance.”

– R. Murray Schafer (1933 – 2021)

The illustration was designed and kindly provided by the La Strada Graz Festival and depicts Michel Risse listening through one of his "Kaleidophones." The quote originates from R. Murray Schafer’s book: "The Soundscape: Our Sonic Environment and the Tuning of the World."

…………………………………… (Name in Blockbuchstaben)

…………………….. (Matrikelnummer)

Ehrenwörtliche Erklärung

Mit meiner Unterschrift bestätige ich, dass mir der *Leitfaden für schriftliche Arbeiten an der KUG* bekannt ist und ich die darin enthaltenen Bestimmungen eingehalten habe. Ich erkläre ehrenwörtlich, dass ich die vorliegende Arbeit selbständig verfasst, andere als die angegebenen Quellen nicht verwendet und die wörtlich oder inhaltlich entnommenen Stellen als solche kenntlich gemacht habe. Sofern fremde Hilfe (z.B. Korrekturlesen durch Native Speaker) und/oder KI-Dienste verwendet wurden (z.B. internetbasierte Übersetzungstools und/oder KI-basierte Sprachmodelle) habe ich dies ausgewiesen.

Graz, den ……………………………………….

…………………………………………………….
Unterschrift der Verfasserin*des Verfassers

Statutory Declaration

I declare that I have authored this thesis independently, that I have not used other than the declared sources/resources, and that I have explicitly marked all material that has been quoted either literally or by content from the used sources. I utilized AI support for grammar and spell-checking.

Thank you,

Robert Höldrich, for convincing me to start a PhD in the first place and for guiding me through the process. I am grateful for the many inspiring conversations and for you effortlessly providing a week's worth of work in just one-hour meetings.
Wolfgang Aigner, for providing an environment that enabled deeply focused and unobstructed work, for your guidance, and for supporting my research stay in Sweden.
Michael Iber, for your critical spirit and appreciation for detail, which pushed me to do better, and for trusting me to co-chair the Audio Mostly 2022 conference with you.
Alexander Rind, for the many inspiring conversations and for always being available, even late at night when a deadline was close.
Niklas Rönnberg for hosting me in your research group, the *SoniBoyz*, during spring 2022, shaping my research for the better.
Elias Elmquist for our extensive discussions and reflections, but even more for warmly welcoming me to both of your hometowns, Norrköping and Varberg, and for becoming a dear friend.
Alois Sontacchi and **Silvia Miksch** for reviewing this PhD thesis.
to my **colleagues** at the St. Pölten UAS and KUG, as well as to **all of the co-authors** that I was lucky to publish with over the last four years.

—

Mom for your love and for always being close, regardless of where your home is.
Dad for your love and for ... everything.
Grandma for your love and for being a crazy cool granny.
Lolo, for never being more than one phone call away.
Valentin, for bringing well-curated randomness to my life.
to **all of my dear friends**.

—

In my research, I was embedded into the team of the FWF-funded project *SoniVis*[1] with grant number P33531, led by Wolfgang Aigner at the St. Pölten University of Applied Sciences in Austria. My research was also funded by the Gesellschaft für Forschungsförderung Niederösterreich (GFF) with grant number SC20-006. I want to express my gratitude to both agencies for funding this research.

[1] fwf.ac.at/en/research-radar/10.55776/P33531 (accessed Nov. 9th, 2024)

Abstract | Kurzfassung

Information visualization and sonification are two techniques that share the objective of making abstract data interpretable to humans. To do so, visualization employs visual representations of data, and sonification employs auditory representations. Over the recent decades, the two respective research communities have developed mainly in parallel, and only a few integrated audio-visual designs have been proposed, even if the integration of sonification and visualization potentially offers to be "more than the sum of the two." This thesis contributes to establishing a design theory for audio-visual data analytics. It explores the potential of analysis tools that explicitly combine sonification and visualization. The first of three articles composing this thesis suggests an integrated theoretical framework for both types of display to support the formal description and analysis of audio-visual data analytics idioms. The second article suggests a classification system for such idioms, presents the field's state of the art, and identifies research gaps. The third article presents the design and evaluation of an idiom called "Parallel Chords." Methodologically, this thesis is rooted in design science research and utilizes an eight-part theoretical framework to present the contributions of the articles included.

Informationsvisualisierung und Sonifikation sind zwei Techniken, die das gemeinsame Ziel verfolgen, abstrakte Daten für Menschen interpretierbar zu machen. Dabei verwendet die Visualisierung visuelle und die Sonifikation auditive Datenrepräsentationen. Während der letzten Jahrzehnte haben sich die entsprechenden Forschungsfelder weitgehend parallel entwickelt, und nur wenige integrierte audiovisuelle Designs wurden vorgeschlagen, obwohl die Kombination der Darstellungsformen „mehr also die Summe ihrer Teile“ sein könnte. Die vorliegende Dissertation trägt zur Etablierung einer Designtheorie für audiovisuelle Datenanalytik bei und untersucht das Potential von Kombinationen aus Sonifikation und Visualisierung. Der erste der drei enthaltenen Artikel stellt einen kombinierten theoretischen Rahmen für beide Darstellungsformen vor, um eine formale Beschreibung und Analyse von audiovisuellen Idiomen der Datenanalytik zu unterstützen. Der zweite Artikel stellt ein Klassifikationssystem für ebendiese Idiome vor, beschäftigt sich mit dem aktuellen Stand des Forschungsfelds und identifiziert Forschungslücken. Der dritte Artikel präsentiert das Idiom „Parallel Chords“ und bespricht dessen Design und Evaluierung. Methodisch orientiert sich diese Dissertation an der Designwissenschaft und verwendet einen achtteiligen theoretischen Rahmen, um die Beiträge der enthaltenen Artikel zu präsentieren.

Contents

List of Figures

1. Introduction

We connect to the physical world around us through our human senses. We see, hear, touch, smell, and taste what our surroundings offer. Especially our capabilities for multimodal sensation allow us to interact with and interpret our environment. While we are great at navigating through an analog world, our sensory system is not designed to effectively interpret digital data, at least not in its raw form, often arranged in numbers and tables. Nevertheless, the keyword "data" accompanies many people's everyday personal and professional lives in a multi-faceted way. Data is an abstract term defined in the Cambridge Dictionary[1] as "information, especially facts or numbers, collected to be examined and considered and used to help decision-making, or information in an electronic form that can be stored and used by a computer." There are various reasons to be interested in the analysis and interpretation of data, in whatever specific context that may be. On a personal level, one might be interested in better understanding and acting on personal health issues, economic situations, or social media behavior. On a professional level, people might be interested in production data, text analysis of historical books, or all kinds of data related to the Sustainable Development Goals [UN 24]. Data in its various shapes and forms significantly impacts our lives. Yet, our human senses do not support their direct interpretation. In short: Our senses are sensational, but not for the analysis of data.

This gap between the necessity to work with digital information and the inappropriateness of our basic human sensing capabilities led people to build tools for support. The visualization of data is defined as "the use of computer-supported, interactive, visual representations of data to amplify cognition" [CMS99, p. 6] and has been a broadly researched topic for several decades [Ber83, CM97, Mun15]. The practice of visualizing abstract data is so established that children learn the basic forms of visualization early in school [BSK*24] and an international research community has evolved, meeting yearly at conferences such as IEEE VIS[2], EG EuroVis[3], or IEEE PacificVis[4].

[1] dictionary.cambridge.org/dictionary/english/data (accessed Nov. 9th, 2024)
[2] ieeevis.org (accessed Nov. 9th, 2024)
[3] eg.org/wp/eg-events/visualization-eurovis (accessed Nov. 9th, 2024)
[4] pacificvis.github.io (accessed Nov. 9th, 2024)

While visualization is the predominant technique to transform data into interpretable structures, alternatives do exist. One of them is sonification, defined in [KWB*99, p. 3] as "the use of nonspeech audio to convey information." The sonification research community studies the potential of sound to translate data into informative audible structures. Research in sonification is arguably younger than in visualization, and conferences such as the International Conference on Auditory Display (ICAD)[5], the Interactive Sonification Workshop (ISon)[6], or the ACM Audio Mostly[7] are smaller in number of participants. Nevertheless, sonification is a promising alternative to visualization, and the core goals of both fields are identical: Communicating (abstract) data via a perceptualization so that the representation supports a user with their interpretation of data. While both fields and their respective research communities have co-existed for several decades in parallel, the integration of sonification and visualization is promising as well, potentially offering to be "more than the sum of the two." This idea of the two modalities effectively supporting each other is rooted in phenomena such as the different temporal and spatial resolutions offered by the two senses. While the visual sense offers great spatial resolution, the auditory sense offers great temporal resolution for incoming stimuli [Bla96], building a fruitful soil for combinations.

In an integrative endeavor, this thesis explores the potential of analysis tools that explicitly combine the two forms of display: visualization and sonification. The thesis presents the current state of the art of the relatively young field of audio-visual data analytics [EEC*24], provides a theoretical foundation for the development and discussion of combined analysis tools [ERI*23], and introduces two integrated designs intended for exploratory data analysis, called *Parallel Chords* [EER*24] and *SoniScope* [ERI*22]. For the remainder of this thesis, the phrase "audio-visual data analytics" will mostly be abbreviated as "audio-visual analytics" or simply as "AVA."

This chapter will motivate studying audio-visual analytics, specify the scope of this research, and introduce the research questions. The employed methodology is introduced in Chapter 2. Chapter 3 provides an overview of related and prior work, embedding this thesis in the two fields of visualization and sonification. Chapter 4 will discuss three journal articles that jointly compose the core of this work. The section also introduces conference contributions and community-building activities, which were integral to the research. Chapter 5 discusses the research questions and contributions through the lens of an envisioned design theory of audio-visual analytics. Finally, Chapter 6 will describe a vision for the future of audio-visual analytics research.

[5]icad.org (accessed Nov. 9th, 2024)
[6]interactive-sonification.org (accessed Nov. 9th, 2024)
[7]audiomostly.com (accessed Nov. 9th, 2024)

1.1 Motivation & Inspiration

How our senses work together to support us in navigating our daily environments [KvKM17] is a fundamental inspiration to study the potential of integrated audio-visual analysis. Moving through an unknown area may require both visual and auditory attention to navigate safely. Two thought experiments make the point: (1) Walking in complete darkness requires focusing on other senses, such as touch, proprioception, and the auditory sense, to navigate safely. (2) Walking without hearing, such as when using headphones in public spaces, requires us to pay more attention to the visual sense to navigate safely (for example, by visually checking for cars more rigorously when crossing the street). In general, the multi-sensory way we perceive and explore our surroundings suggests that multi-sensory displays can support data analysis as well [RC15]. The multi-sensory (here, audio-visual) representation of data seems especially promising in the context of exploratory data analysis, where a user has no prior knowledge about what or where exactly to search for phenomena in the data [Tuk77, Mun15].

Multi-sensory integration, also on a theoretical level, is not a new idea and has previously been called for [RW10]. Originating from the field of visualization, seminal work by Wilkinson [Wil05] or Spence [Spe07] made clear their understanding of visualization also embraces other sensory modalities such as the auditory. Early on, in 1995, related ideas were discussed when Minghim and Forrest [MF95] suggested that sonification could support visualization by adding complementary or redundant dimensions or by providing a natural mapping for time-oriented data. These arguments align with the idea that one sensory modality could help overcome challenges the other might have [TC05]. For example, the topic of overview is a classic challenge discussed in the visualization literature [Shn03, EF10] that could be tackled using sonification. In our daily lives, our auditory perception gives a direct overview of our 360-degree surroundings. At the same time, the eyes are responsible for a detailed analysis of the area where we take our next steps. Similarly, a sonification designed as an abstract soundscape could provide an overview and context, while an interactive visualization system lets users explore detailed views of their data. A visualization such as a regular scatter plot, for example, is not able to display an unlimited number of data attributes at the same time, and some attributes could be displayed instead by an accompanying sonification [ERI*22]. In a more general sense, two fundamental challenges are related to the two display techniques: (1) Visualization is limited by space, and (2) sonification is limited by time. A data visualization cannot use an endless amount of space, and a sonification cannot use an endless amount of time. This "balance of fundamental limitations" can be considered an inspiration to study the potential of combining sonification and visualization.

1.2 Scope

The combination of sonification and visualization is a topic that can be studied from various perspectives. This section provides clarification on the scope of this thesis and what it explicitly is *not* covering.

Audio-visual Analytics Idioms
An "Audio-visual Analytics Idiom" is "a distinct approach to creating and manipulating audio-visual representations of data" [EEC*24, p. 1], which is a definition informed by Munzner's definition [Mun15, p. 10] of the "visualization idiom" as "a distinct approach to creating and manipulating visual representations." The rest of this thesis will utilize the phrasing audio-visual analytics idiom or AVA idiom instead of similar phrasings such as AVA-design or AVA-tool. To be considered an AVA idiom, the data representation needs to communicate the data employing both visual and auditory modalities. For example, an idiom that uses a visual representation just to provide an interface for interaction is not considered an audio-visual analytics idiom. Figure 1 describes the process of mapping data to an audio-visual analytics idiom that is used for exploration.

Exploratory Data Analysis
This thesis focuses on the context of exploratory data analysis [Tuk77] of multivariate data. Exploration is considered interactively searching for structures, patterns, or insights without knowing their form or location within the data [Mun15]. An analogy of exploration is the search for an object in a haystack without knowing (1) that there is something to be found and (2) that it will take the shape of a needle after all. For exploratory analysis, the interactivity of an idiom is crucial, as it should offer a user flexible visual and auditory perspectives on their data. Staying in our analogy, a needle will not be found by just watching the haystack without using one's hands.

Adjacent Topics
Regarding the integration of sonification and visualization, two adjacent topics come to mind: accessibility and monitoring. While the potential of sonification to enhance the accessibility of visual displays is a most timely field of research, it is not the focus of this thesis. This thesis is meant to contribute to the knowledge around idioms requiring a user to see and hear the display without restrictions. Nevertheless, some of this research's findings may be relevant to the development of idioms meant to raise the inclusiveness of analysis environments. The field of auditory monitoring has been studied extensively and is distinct from the task of data exploration. As such, it is not covered in this thesis.

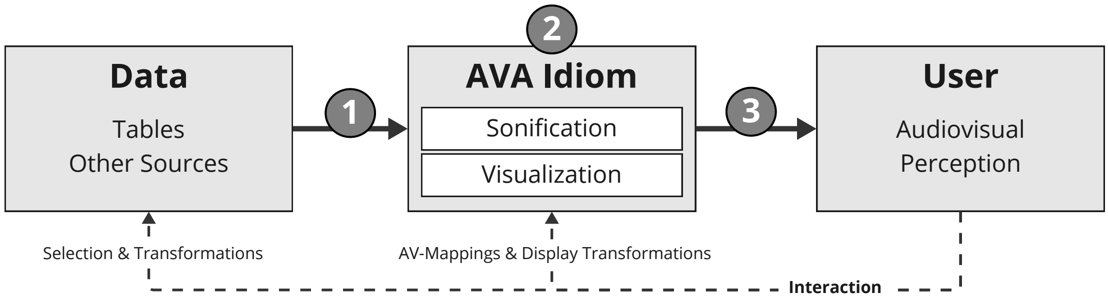


Figure 1: A simplified audio-visual analytics model, inspired by the reference model for visualization by Card et al. [CMS99]. Data is mapped to an interactive audio-visual analytics idiom, allowing users to explore the data. The three numbers (1), (2), and (3) indicate where in the model to locate the three journal contributions composing this thesis. (1) relates to the theoretical constructs article describing the mapping from data to an AVA idiom [ERI*23], (2) relates to the state-of-the-art report on AVA idioms [EEC*24], and (3) relates to the Parallel Coordinates article and the psychophysical evaluation of the idiom [EER*24].

1.3 Research Questions

This thesis is written from a sound and music computing student's perspective. It aims for a balanced proportion of inspiration from the sonification and visualization fields. Nevertheless, visualization is the far more established technique in the context of exploratory data analysis of multivariate data. This is a fact that is mirrored in the phrasing of the third research question (RQ 3) studied in this thesis:

- **RQ 1:** Can we identify basic theoretical constructs that build a bridge between sonification and visualization? How do we define such constructs to offer terminology for the formal analysis and description of AVA idiom design?
- **RQ 2:** What is the current state of the art of audio-visual analytics, both with respect to its idiom design space and its research community? Can we identify apparent research gaps from the current state of the art?
- **RQ 3:** In the context of exploratory data analysis, can sonification support visualization challenges and offer solutions? Can we describe examples and what advantages emerge from combined designs?

The present thesis is built around three articles published between 2023 and 2024. While all of them relate to all of the mentioned research questions, they each focus on one of the research questions primarily (Article # relates to RQ #).

- **Article 1:** Enge K., Rind A., Iber M., Höldrich R., Aigner W.: Towards a unified terminology for sonification and visualization. Personal and Ubiquitous Computing 27, 5 (2023), 1949–1963. doi: 10.1007/s00779-023-01720-5.
- **Article 2:** Enge K., Elmquist E., Caiola V., Rönnberg N., Rind A., Iber M., Lenzi S., Lan F., Höldrich R., Aigner W.: Open Your Ears and Take a Look: A State-of-the-Art Report on the Integration of Sonification and Visualization. Computer Graphics Forum (EuroVis '24) 43, 3 (2024), e15114. doi: 10.1111/cgf.15114.
- **Article 3:** Elmquist E., Enge K., Rind A., Navarra C., Höldrich R., Iber M., Bock A., Ynnerman A., Aigner W., Rönnberg N.: Parallel Chords: An Audio-Visual Analytics Design for Parallel Coordinates. Personal and Ubiquitous Computing 28 (2024). doi: 10.1007/s00779-024-01795-8.

All three articles are written with methodological considerations based on design science research, a field briefly introduced in the next chapter.

2. Design Science Research & Design Theory

Methodologically, this thesis is rooted in design science research [Sim08, VK15, JP14, GJ07, Cro01]. The design of (software) artifacts is common practice both in sonification as well as in visualization research. Both fields have a vast corpus of designs available, forming their current state of the art. Also, audio-visual analytics research and this thesis are characterized by the implementation of audio-visual analytics idioms, hence artifacts.

The conventional understanding of science relates to natural or social science. How do the terms "design," "science," and "research" fit into our image of science? While, in a simplified manner, natural sciences or social sciences are concerned with knowledge about phenomena in nature or society, the science of the artificial generates knowledge about human-made systems and artifacts designed with a purpose in mind [VK15, Sim08]:

> The natural sciences are concerned with how things are. [...] Design, on the other hand, is concerned with how things ought to be, with devising artifacts to attain goals. [Sim08, p. 114]

What Simon called the "science of the artificial" is widely known as design science today. More specifically, in the context of this thesis, we are speaking of design science research in the field of information systems (as opposed to other disciplines such as architecture or industrial design) [GJ07, VK15]. A follow-up question will be: What makes the process of design qualify to be "research" or "science?" While designers are free to be interested primarily in solutions relevant to single actors or local interests, design science researchers are interested in solutions of general interest, producing and communicating knowledge about artifacts to a (research) community [VK15, JP14]. Johannesson and Perjons [JP14, p. 7] describe design science as "the scientific study and creation of artefacts as they are developed and used by people with the goal of solving practical problems of general interest." Scholars have also described the difference between design science and design as related to the difference between a discipline and a craft. What makes design science research "more" than a craft is the fact that researchers theorize about

Figure 2: A simplified flowchart of the general design science methodology, as described in detail in Figure 2.5 in [VK15].

their artifacts and that they follow a guided process, including the evaluation of the design [VK15, JP14, Cro01].
Different guidelines for such a structured design process have been suggested. Early on, the core element of process guidelines was the loop between design knowledge and the designed artifacts [TVY90, Owe98]. Vaishnavi and Kuechler, for example, adopted the structure of Takeda et al. [TVY90] by adding dedicated outputs to the five suggested process steps (compare Figure 2). In the specific context of visualization research, Sedmair et al. presented a seminal paper on design study methodology, suggesting a framework of nine stages, with each step potentially looping back into all previous stages [SMM12]. Since its publication in 2012, many papers published at visualization conferences have followed this methodology (at least in parts). The following will briefly introduce the methodology, to be discussed later in the context of audio-visual analytics in Section 5.2.

Design Study Methodology

Sedlmair et al.'s paper "Design Study Methodology: Reflections from the Trenches and the Stacks" [SMM12] describes a process that has been especially relevant for the visualization research community. The authors introduce guidelines on effectively conducting design studies, an increasingly popular form of visualization research. A design study is "a project in which visualization researchers analyze a specific real-world problem faced by domain experts, design a visualization system that supports solving this problem, validate the design, and reflect about lessons learned in order to refine visualization design guidelines" [SMM12, p. 2]. In many cases, researchers will deeply immerse themselves in the field of the domain expert they are working with, trying to understand the visualization challenges within the domain context. Sedlmair et al. consider this collaborative process with a domain expert mandatory and critical [SMM12]. Ten years after its publication, the paper received a "10-year test of time award" at IEEE VIS 2022, also because "[d]esign study methods are now the 'gold standard' in approaching visualization solutions to real-world problems[1]." Over the last decade, many design studies have been published, and design science methodology has evolved into an established practice in the field. It is only plausible to consider design studies

[1]ieeevis.org/year/2022/info/awards/test-of-time-awards (accessed Nov. 9th, 2024)

as the appropriate methodological approach in audio-visual analytics as well. Nevertheless, Section 5.2 will argue that the field of audio-visual analysis lacks (or recently lacked) a fundamental knowledge base that is considered a prerequisite for a rigorous contribution through design studies.

Eight Components of a Design Theory

Design study research generally aims to develop what is called a design theory. Design theories foster the communication, justification, and development of design knowledge [GJ07]. Theories on visualization and sonification can be described as design theories, and a theory of audio-visual analytics will be such a design theory as well. Gregor and Jones identified eight components to be relevant for the establishment of design theories in information systems [GJ07]. They do so by extending the work of Walls et al. [WWES92]. This thesis contributes to the establishment of a design theory for audio-visual analytics in the context of exploratory data analysis. Such a theory is supposed to describe the integration of information visualization and sonification into one form of audio-visual display. The list below introduces the eight components and relates to Table 2 in the publication of Gregor and Jones [GJ07].

1. The **purpose and scope** of a design theory describes what a system is designed for. They specify a type of artifact and set its scope and boundaries.

2. **Constructs** describe or represent the "entities of interest" in a design theory. They are on the most basic level of a theory and may be abstract theoretical terms or physical phenomena.

3. **Principles of form and function** are described as the "abstract blueprint" of an information systems artifact. These principles provide a designer with insights into the "structure, organization, and functioning" of a design product, in our case, an AVA idiom.

4. **Artifact mutability** is the component relating to the ability of a design theory to anticipate the mutable nature of information systems artifacts. A design theory should allow for the description of an artifact that has evolved into a new form or shape over time. When an AVA idiom is equipped with new features, for example, the respective design theory should not fail to represent such a mutation.

5. **Testable propositions** are statements or hypotheses about the artifact (and its functionality) that can be tested by studying an instance of the described artifact. Gregor and Jones also describe them as "truth statements."

6. **Justificatory knowledge**, also described as "kernel theories," is the knowledge that is foundational to the design space. In other words, it builds the theoretical foundation that an artifact is built on. For audio-visual analytics, one such informing field is psychoacoustics.

7. **Principles of implementation** describe the processes that are employed when implementing a design product or method that is covered by the design theory. They guide a designer when building an artifact.

8. **Expository instantiations** are actual implementations of an artifact described by the theory. These instantiations can also be used as representations and test environments of the theory.

Section 5.1 will discuss the contributions of this thesis through the theoretical lens of these eight components. Before discussing the articles composing the present thesis, the following chapter will provide context by introducing selected related work.

3. Related Work

One of the articles presented later in this thesis is a state-of-the-art report (STAR) on the field of audio-visual analytics itself. While the STAR discusses AVA idioms presented between 2011 and 2023, this section introduces related work in the field on a wider level. Selected theoretical contributions are followed by selected practical examples. Three topics relevant to this section are covered in detail in the publications composing this thesis and would be of considerable redundancy if discussed here again: For an introduction to the most established sonification techniques such as parameter mapping sonification or model-based sonification, and for an introduction to the major milestones in the recent history of sonification research, the reader is referred to section 1.1 "Sonification Background" in the STAR [EEC*24]. For an introduction to the different techniques of visualization of multivariate data, the reader is referred to section 2.1, "Visualization of multivariate data," in the article on the Parallel Chords AVA idiom [EER*24].
As mentioned earlier, Minghim and Forrest suggested that sonification and visualization could be integrated beneficially early on in the 6th IEEE Visualization conference in 1995 [MF95]. The authors called for integration just one year after Kramer published the seminal book "Auditory Display: Sonification, Audification and Auditory Interfaces," covering the proceedings of the first ICAD conference [Kra94]. In their paper, Minghim and Forest identified several topics where sonification might support visualization, such as providing additional data dimensions through sound. Other early theoretical contributions to audio-visual analytics are routed in work by Nesbitt, who suggested two possibilities for the design space of multi-modal data displays, also including haptic displays [Nes01, Nes04, Nes06]. One of them is inspired by the reference model for visualization by Card, Mackinlay, and Shneiderman [CMS99]. This first model uses space as the necessary foundation for both visual and auditory displays [Nes01]. Nesbitt's second model of the multi-modal design space employs three kinds of metaphors: spatial metaphors, temporal metaphors, and direct metaphors [Nes06]. With this second model, Nesbitt acknowledges the relevance of time for auditory displays more clearly, and therefore, it is better in line with the ideas presented later in the present thesis [ERI*23]. In a more recent endeavor, Caiola et al. [CLR22] analyzed a large number of audio-visual idioms stemming

from the Data Sonification Archive (DSA)[1] and a Google (not Google Scholar) search. The authors present an "audiovisual design map" that informs a reader about sensory correspondences commonly used in AVA idioms (the authors use the wording "audiovisual sonification"). Such sensory correspondences describe designs where a data attribute is mapped redundantly to both a visual and an auditory channel. They found, for example, that the visual mapping of vertical position often comes with an auditory mapping of pitch. The mentioned Data Sonification Archive is a curated website presenting sonification work not only from academic backgrounds but also from fields such as the arts or data journalism. Recently, Lenzi and Ciuccarelli, two of the curators of the DSA, presented the Data Sonification Canvas [LC24], a tool meant to support designers by providing helpful questions to think of when designing a sonification. The canvas also includes a section explicitly asking about multimodality and whether the sonification is coupled with other sensory modalities. The above-mentioned work can be considered inspirational in investigating the integration of theories of sonification and visualization. What was missing from the field before this thesis was such a theoretical integration employing a balanced level of inspiration from both fields.

On a practical level, many idioms combining sonification and visualization have been presented over the years. Some of the earlier examples are the following: In 1990, Rabenhorst et al. [RFJ*90] used sonification to augment a vector field visualization, allowing a user to visually focus on one field while listening to another. Nesbitt and Barrass [NB02] presented an evaluation of a sonification, a visualization, and an audio-visual combination for the analysis of market stock data. Their results show that both the sonification and the multimodal approach supported participants in better predicting price movements. Franchin et al. [FdLM09] proposed a Java web tool for audio-visual data analysis using a scatterplot and a shock wave metaphor for its sonification. Chang et al. used an audio-visual approach to explore the activity of neurons in the brain [CWB10] and Hildebrandt et al. [HAR16] used a combination of visualization and sonification to analyze business process execution data.

While these and other examples are described in the respective sections of the presented articles, this section focuses on introducing more recent articles published in 2024 that were not mentioned yet elsewhere: Schütz et al. recently presented a "Framework for Multimodal Medical Image Interaction" [SMS*24], that is designed to support physicians with the audio-visual exploration of human tissue. The authors conducted two studies. The first one shows that 34 participants were able to effectively learn the audio-visual correspondences between visuals and sounds. In the second study, the participants were significantly better at

[1] sonification.design (accessed on Oct. 8th, 2024)

localizing a brain tumor with the audio-visual modality compared to the visual-only modality. The study relates well to one of the conclusions we describe in our STAR [EEC*24]: The potential of combining scientific visualization and sonification for exploratory data analysis. In this particular case, the authors even used model-based-sonification, while most other AVA idioms utilize parameter mapping sonification. Linke and Ziemer [LZ24] recently presented "SOMson," a sonification of Kohonen Maps, also called self-organizing maps [Koh95]. The authors use a sonification designed to display several data dimensions via sound simultaneously. With SOMson, a user is presented with a two-dimensional grid map that visually displays high-dimensional data but is necessarily affected by the loss of details. The authors present several examples and even an interactive website[2], making it plausible that the design could be useful. Nevertheless, one might oppose the following statement from their paper: "We think SOMson's benefit is so obvious that a formal evaluation is obsolete, especially since most of the dimensions have been implemented and evaluated before" [LZ24]. Later in this thesis, the relevance of evaluations in our field will be discussed in more detail. Two other examples that both explore the artistic perspectives onto audio-visual data representation (and that both already cite our STAR) are the work by Pile et al. [PLP*24], an installation presented at the Ars Electronica Festival in 2022[3], and Armitage et al.'s work [ACJ24] on the combination of two python libraries for efficient and interactive audio-visual display. The above-mentioned work, especially the artistic work, demonstrates the timeliness of fundamental research in audio-visual analytics.

When abstracting from the level of data displays, a lot of related work exists in psychophysics, a field that studies our perception as such. It is out of scope for this thesis to discuss in detail the vast amount of justificatory knowledge of the field, such as psychovisual studies [FG03, Gre15], psychoacoustic studies [Neu04, Bre90], gestalt theory [Wer23, RC15], or theories of perceptual objecthood [Kub81, KVV01]. Nevertheless, two more perspectives should be mentioned: (1) Audio-visual scene analysis, hence the way we perceive our environment, depends not only on physical sensations but also on factors such as attention and prior experience [vdH03, KvKM17]; (2) Research shows phenomena where a visual stimulus can modify our auditory sensations [MM76, AB04], and vice versa [SKS02]. Phenomena such as these should be considered when designing audio-visual analytics idioms.

In general, the mentioned related work shows the relevance of both theoretical and practical contributions toward an audio-visual analytics design theory. Abstracted from individual research gaps, described in more detail in our STAR [EEC*24], it

[2]simon-linke.github.io/SOMson/simple (accessed October 9th, 2024)

[3]ars.electronica.art/planetb/en/spin-wave-voices (accessed Oct. 10th, 2024)

is the employment of a "balanced mindset" between sonification and visualization that characterizes the present thesis and its contributions to the development of the field.

4. Contributions to the Field of Audio-Visual Data Analytics

This section first introduces this thesis's main contributions and results in the form of three journal articles [ERI*23, EER*24, EEC*24]. The section will discuss the individual articles themselves and their current or potential future impact on the field. Section 4.4 will introduce selected other contributions in the form of conference papers, public outreach activities, and the founding of the Audio-Visual Analytics Community. In Appendix B, the *CRediT* author role taxonomy will be employed to make contributions to the individual articles transparent.

4.1 Towards a Unified Terminology

This section discusses the following article that was published in the Journal on Personal and Ubiquitous Computing in 2023:

[ERI*23] Enge K., Rind A., Iber M., Höldrich R., Aigner W.: Towards a unified terminology for sonification and visualization. Personal and Ubiquitous Computing 27, 5 (Oct. 2023), 1949–1963. doi: 10.1007/s00779-023-01720-5.

When building a bridge between two heavily related but largely separated research communities, it is beneficial to have shared theoretical constructs available. Such constructs help members from both communities to communicate their ideas to each other. As described earlier, both the sonification and the visualization theories can be considered design theories. To integrate the two approaches of sonification and visualization, basic theoretical concepts, hence "constructs," are needed to support the discussion, the design, and potentially even the semi-automatic design of audio-visual analytics idioms in the future. With sonification theory being less developed, it seemed favorable to review the visualization literature for theoretical constructs that lend themselves for adaption for sonification. We identified three fundamental theoretical constructs established within the visualization research community. They are the *spatial substrate*, the *visual mark*, and the *channel* [CMS99, Mun15]. A visual mark is a geometric entity that is

placed in space and can take one of four fundamental forms: (0D) points, (1D) lines, (2D) areas, and (3D) volumes [CMS99]. A wide range of visualization idioms are designed based on the concept of the mark as an essential building block of visualizations [CMS99, Mun15, Wil05, Mac86, CM97]. A mark is informative to its viewers through the channels that alter its appearance [Mun15]. Frequently occurring channels are a visual mark's shape, color, position, or length. The conceptual geometric space, such as a piece of paper or a desktop display, necessary to "hold" the marks is called spatial substrate [CMS99] and is the third of the widely established constructs found in the visualization literature.
The publication "Towards a Unified Terminology for Sonification and Visualization" [ERI*23] discusses the definitions of these three theoretical constructs in more detail, then abstracts them to their modality-independent meaning, translates them to the field of sonification, and provides examples of how to use those to describe selected design examples from the literature. In short, the article identifies time as the best-fitting substrate of sonification while arguing against space or frequency as two alternatives. The article defines 0D and 1D auditory marks to expand in time, providing information through the encoded channels. Typical auditory channels are the pitch and timbre of a sound.

Existing and expected impact of the publication:

Since the publication of the article, other researchers have taken inspiration from the theoretical constructs. The following paragraphs offer a reflection on selected examples from the recent literature.
One challenge within the sonification research community is the establishment of design software. Such software should support not only experts but also beginners with the streamlined drafting of sonification ideas. The visualization community used constructs such as marks and channels not only to describe their designs but also to develop visualization design software. To some extent, they have successfully made the necessary step towards a lay audience with toolboxes such as matplotlib [Hun07], vega-lite [SMWH17], or Tableau[1]. In a very similar manner, and inspired by visualization toolkits such as matplotlib, Reinsch and Hermann presented three layers of software that might finally show the potential democratize sonification design [HR21, RH22, RH23]. The fundamental layer of their framework is called *sc3nb*, a "python package for audio coding and interactive control of the SuperCollider programming environment" [HR21, p.208]. The second layer is called *mesonic*, described as a "novel framework for crossplatform, back-end independent, pythonic sonification framework" [RH22, p. 1]. Finally, the top layer is called *sonecules*, described as "a flexible, extensible, enduser friendly and open-source Python sonification toolkit to bring 'sonification to the masses'" [RH23,

[1] tableau.com (accessed Nov. 9th, 2024)

p. 62]. These contributions are worth mentioning in such detail as they show a substantial link to the theoretical constructs suggested in the present thesis. Especially in their publication on the mesonic-framework [RH22], the authors contextualize their implementation of "mutable" synths and "immutable" synths to the differentiation between 0D auditory marks and 1D auditory marks defined in [ERI*21, ERI*23]:

> The distinction of *mutable* and *immutable* Synths was found by ourselves to be practical in common use cases, and is also suggested by the research of Enge, Rind, Iber, *et al.* [44][2] who suggest the adaption of theoretical constructs from visualization to sonification. [RH22, p. 4]

Note that Reinsch and Hermann refer to the initial publication of the same ideas, presented virtually at Audio Mostly 2021 [ERI*21] and extended to the journal version in 2023 [ERI*23]. Reinsch and Hermann do not only use the theoretical constructs to describe their design, but they even suggest a theoretical extension by drawing another parallel to the visualization theory:

> Using the terminology of Enge, Rind, Iber, *et al.* [44], that the time is the substrate of the sonification like the space is the substrate of the visualization, this introduces the Context as counterpart of the Figure or Scene and allows building a further bridge between sonification and visualization concepts. [RH22, p. 4]

A second example is Elmquist's recent publication on the sonification of bird species [EEB*24]. He refers to the construct of auditory magnitude and identity channels, as they are discussed in [ERI*23]:

> Therefore, the bird songs used for the auditory icons should be chosen not only by how well they represent each bird order, but also by how distinguishable they are to each other as one set according to their auditory characteristics. Enge *et al.* [8][3] makes the connection of an auditory icon having an identity channel through its timbre, similar to how a user can differentiate the identity of data points in visualizations by their color [19][4]. [EEB*24, p. 2]

In the case of Zaho et al. [ZLN24] o,ur proposed terminology of auditory marks even made it into the title of their paper: "Speech-based Mark for Data Sonification." Time will show the applicability of the three constructs to bridge the gap between the communities and to support the design of audio-visual analytics idioms. Subjectively speaking, the found constructs were helpful to our research group

[2]Reference [44] is part of this thesis' bibliography as [ERI*21]
[3]Reference [8] is part of this thesis' bibliography as [ERI*23]
[4]Reference [19] is part of this thesis' bibliography as [Mun15]

at the FWF SoniVis project[5]. They allowed us to communicate effectively within our team of sonification and visualization experts, to share ideas, and to discuss potential prototypical designs. Furthermore, the constructs were helpful with the classification of designs reported in the STAR on audio-visual analytics idioms [EEC*24], discussed next.

4.2 Open Your Ears and Take a Look

Presented at the EG EuroVis conference in Odense, Denmark, in 2024, this State-of-the-Art report (1) identifies 57 papers from 2011 until 2023 as relevant to the field, (2) introduces a classification system for audio-visual analytics idioms, and (3) discusses the current state of the field.

[EEC*24] Enge K., Elmquist E., Caiola V., Rönnberg N., Rind A., Iber M., Lenzi S., Lan F., Höldrich R., Aigner W.: Open Your Ears and Take a Look: A State-of-the-Art Report on the Integration of Sonification and Visualization. Computer Graphics Forum (EuroVis '24) 43, 3 (2024), e15114. doi: 10.1111/cgf.15114.

For a new field to evolve, knowing the foundation upon which it is built is most relevant. On the one hand, this can be a theoretical foundation, as discussed in detail in [ERI*23]; on the other hand, a practical perspective on the field helps understand its potential and challenges for future work. Together with a team of nine international researchers from the fields of sonification and visualization, an extensive literature search was conducted, initially resulting in 1498 potentially relevant articles from all major publishers in the respective fields. These 1498 articles were scanned and filtered down to 57 articles relevant to our research interest through an elaborate system of checks and rechecks of two to three of the co-authors. The collected data was made available in the supplemental material of the article and is a contribution of its own. We classified the corpus using a variety of tags such as their purpose [Mun15], visualization idiom [Mun15], sonification technique [HHN11], reading level [Ber83], search level [Mun15], dataset type [Mun15], level of redundancy, evaluation system [IIJ*13], target display platform, and more. The STAR offers a thematic corpus overview identifying a variety of themes such as astronomy, earth sciences, medicine and health, or domain agnostic displays. It also offers insights into the details of selected and representative articles, describing them from specific thematic perspectives, such as their reading level [Mun15] or their level of redundancy.

[5]Descibed in more detail in the Acknowledgment Section of this thesis. This research was funded in part by the Austrian Science Fund (FWF): 10.55776/P33531

Existing and expected impact of the publication:

The STAR will help both visualization and sonification researchers identify research gaps in their own field, as well as in one that is closely related. The STAR is also intended to help researchers find interesting collaborators from other fields. In an endeavor similar to Reinsch and Hermann's contributions, Armitage et al. recently presented their work in integrating two Python libraries to support the design of interactive audio-visual displays [ACJ24]. They motivate their article with a quote from our State-of-the-Art report, discussing the abilities of researchers and domain experts to design their audio-visual analytics idioms. An AVA idiom not described in the STAR due to late publishing is the Parallel Chords idiom, introduced below.

4.3 Parallel Chords

The following article was published in the Journal on Personal and Ubiquitous Computing in 2024.

[EER*24] Elmquist E., Enge K., Rind A., Navarra C., Höldrich R., Iber M., Bock A., Ynnerman A., Aigner W., Rönnberg N.: Parallel Chords: An Audio-Visual Analytics Design for Parallel Coordinates. Personal and Ubiquitous Computing (May 2024). doi: 10.1007/s00779-024-01795-8.

Parallel Chords is an audio-visual analytics idiom fusing two widely established techniques for data analysis: the parallel chords plot as the visualization [Ins85, ID90, Weg90] and auditory graphs as the sonification [Nee18, Flo05]. The parallel coordinates plot is a visualization idiom explicitly designed to display multivariate data. Its main advantage is the flexibility in the number of parallel axes displayed simultaneously. A dataset with N dimensions can be displayed along N parallel axes. Of course, the number of dimensions must be reasonable regarding the available space. Figure 3 illustrates the Parallel Chords idiom with a four-dimensional dataset. One polyline in a Parallel Chords plot represents one data item along its N attributes.
The article contributes a design of an AVA idiom, a qualitative result inspection [IIJ*13] in the form of a fictive user scenario and the presentation of prototypical audio-visual data patterns, as well as a formal evaluation. In this formal evaluation, we conducted a user experiment with 35 participants and a staircase test design [Lev71] to study the thresholds of discernibility between very similar datasets. The user evaluation did not show perceptual advantages for the audio-visual integration. The participants scored well when using the visualization or the combination, and they scored lower when using the sonification alone. While

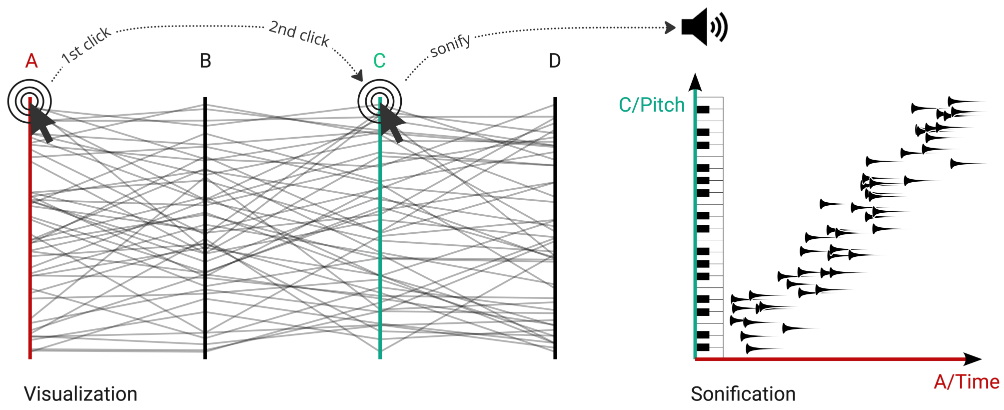


Figure 3: An illustration showing the interactivity of the parallel chords design. A user clicks on two of the axes and hears a sonification of the relationship between these axes. In this example, the positive correlation between axes A and C is not visible due to the two axes not being adjacent to each other. A rising sound sequence will inform a user about the correlation (original figure CC BY in [EER*24]).

the combination didn't show advantages regarding the "hard metric" of perceptual sensitivity, our results suggest the combination to be more enjoyable and to support participants' confidence in their answers.

Existing and expected impact of the publication:

While the article on Parallel Chords has not yet been picked up by the respective research communities (also due to its recent publication), its evaluation results suggest a discussion on the core motivation to combine sonification and visualization. Why should we combine sonification and visualization? What advantages do combinations provide to a user? Should we focus on finding perceptual advantages such as improved just noticeable differences (JND)? Or should we focus on less "hard" metrics, such as user engagement and entering a flow state during analysis, ultimately leading to more (quality) time spent with the analysis? In chapter 5, user engagement will be presented as one alternative objective that should be studied more rigorously next to user performance. Furthermore, the potential future impact of the Parallel Chords study should be assessed in the context of evaluation culture for audio-visual analytics idioms. Our article introduces the staircase method for evaluating AVA idioms, which will potentially be picked up by the community.

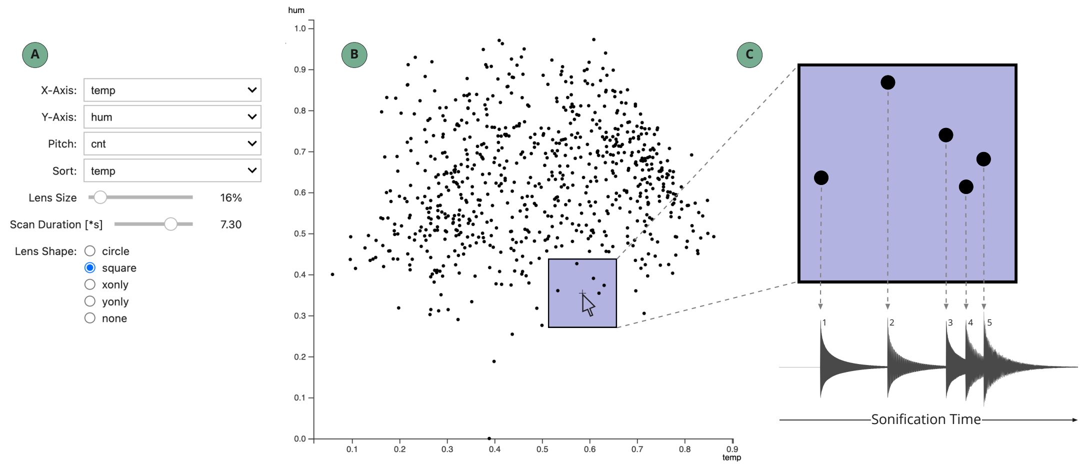


Figure 4: Clicking into the scatterplot will generate a sound sequence representing the selected items. The pitch of the sounds can be mapped to any of the data attributes, also to those that are not visible (original figure CC BY in [ERI*22]).

4.4 Further Contributions

While the contributions described in the following do not formally qualify to compose this thesis, they are relevant to contextualize the conducted research. The following pages describe (1) selected conference contributions, (2) public outreach activities, and (3) the founding of the Audio-Visual Analytics Community (AVAC).

Conference Contributions

Several conference contributions were published as part of this research, out of which two are worth mentioning here briefly. The SoniScope [ERI*22], inspired by the stethoscope and displayed in Figure 4, is a prototypical implementation combining a visual and an auditory scatterplot. While a stethoscope is used to listen into a human's body, the SoniScope is used to listen into a visualization. A user can interact by using a lens to select an area of interest in the scatterplot. When clicked, the SoniScope sonifies additional, not visible dimensions of the selected data points by playing them back as short individual musical sounds. It does so by using the pitch and the onset times of these musical sounds (marimba tones), with lower-pitched and earlier sounds representing lower data values. With the SoniScope, we touch on several topics that are discussed widely within the visualization community, such as visual clutter, multiple views, or the visual information seeking mantra [Shn03] "Overview first, zoom and filter, then details-on-demand." The scatterplot is a technique widely used to display two-dimensional data, but it can only appropriately display so many additional dimensions through non-spatial channels such as color, shape, or size. The objective of SoniScope is to support the exploration of multivariate data by distributing the data dimensions

to the auditory and the visual sense, hence increasing the number of dimensions a user could explore without changing the view itself. On a technical level, the paper contributes an open source design[6], combining a D3.js[7] visualization in a Jupyter Notebook[8] and a SuperCollider[9] sonification controlled via sc3nb [HR21]. The short paper offers a qualitative result inspection [IIJ*13] describing potential usage scenarios.
A second conference contribution worth mentioning is a preliminary replication study of the influential paper by Flowers et al. from 1997: "Cross-Modal Equivalence of Visual and Auditory Scatterplots for Exploring Bivariate Data Samples" [FBT97]. The data collection and software design for the replication study [EFH24] was done by Fabry Liam. The results were presented as a poster at ICAD 2024 in Troy, New York. The original study compared the human capability to estimate the strength of correlations between two data variables when presented through visual or auditory scatterplots. The present extended abstract replicated the original study and strengthened the findings thereof. Participants, again, seemed to have a similar ability to estimate Pearson-r values between two correlated variables regardless of the modality of the display. Also, in line with the original results, the audio-only condition resulted in a higher standard deviation of the responses, signifying that participants were less certain about their response with only the sonification available. With our study, we also wanted to bring back a discussion about replicability to ICAD 2024. Chapter 5 will discuss establishing a culture of evaluation and documentation within the sonification research community to enable replication studies in our collective future work.

Public Outreach

After the EG EuroVis conference in 2022, SoniScope was developed further so that the interaction could be done with a physical stethoscope as an interface. To do so, a user would place the stethoscope on a touch-sensitive display, selecting a region of data points to be sonified. This version of the SoniScope was presented to lay audiences twice in public outreach events. One was the European Researcher's Night 2023 at Universalmuseum Joanneum in Graz, Austria, and one was the Long Night of Research 2024 at the St. Pölten University of Applied Sciences, Austria. On both occasions, visitors were free to explore data using SoniScope after getting a brief introduction to the ideas behind sonification and visualization. While subjective observations are far from an evaluation, two are worth mentioning: (1) Most audience members seemed to understand the concept behind SoniScope

[6] phaidra.fhstp.ac.at/o:4776
[7] d3js.org (accessed Nov. 9th, 2024)
[8] jupyter.org (accessed Nov. 9th, 2024)
[9] supercollider.github.io (accessed Nov. 9th, 2024)

quickly and could draw correct conclusions from the sounds to the underlying data. This observation holds true also for children of various ages. (2) For many children, interacting with the SoniScope seemed fun enough to spend up to ten minutes exploring the data (or rather the sounds themselves). These subjective impressions are inspirational for studying not only user performance metrics but also metrics such as user engagement in future work in the field.

The Audio-Visual Analytics Community

Early in the SoniVis project and in this PhD project, it became clear that the long-term integration of sonification and visualization cannot happen without closing the gap between their communities. Such a gap is best filled when members of both fields exchange ideas and collaborate. Two factors are indispensable for a fruitful exchange of ideas between two communities: A shared language and an opportunity to meet. Supporting meeting opportunities, our research group organized and hosted several workshops and panel discussions at visualization and sonification conferences. Between 2021 and 2024, workshops and panel discussions were hosted at the 2021 ACM Audio Mostly Conference (virtual), the 2021 IEEE VIS Conference (virtual), the 2022 ACM AVI Conference (on-site), the 2022 IEEE VIS Conference (hybrid), at ICAD 2023 (on-site), and at the 2024 EG EuroVis conference (hybrid). In 2024, the yearly All Around Audio Symposium at the St. Pölten UAS offered a special session about audio-visual analytics. With the so-called "AVAC Meet-Ups," three online meet-up sessions brought together the communities, inviting researchers and artists from both fields to present their interdisciplinary work. These meet-up presentations were scheduled for 20 and the discussions for 40 minutes. This allowed the audience and the presenters to exchange ideas and discuss the presented work in more detail than is typically possible during a conference Q&A session. A list of all AVAC events can be found in the Appendix. An upcoming AVAC event in 2025 will be a five-day Dagstuhl seminar[10]. All event descriptions can be found in the archive of the AVAC website[11].

We started AVAC at the Audio Mostly conference in 2021, inviting the sonification and visualization communities to meet online to identify research gaps for integrating the fields. The circumstance that the COVID-19 pandemic forced many conferences to relocate to the virtual space turned out to be an advantage for initiating what we had planned. Multiple conferences waived conference fees for online participation, enabling visualization researchers to attend audio conferences such as Audio Mostly 2021 and vice versa. When asked for their main discipline, most participants attending our first-ever AVAC workshop, virtually

[10]dagstuhl.de/25072 (accessed Nov. 9th, 2024)

[11]audio-visual-analytics.github.io (accessed Oct. 10th, 2024)

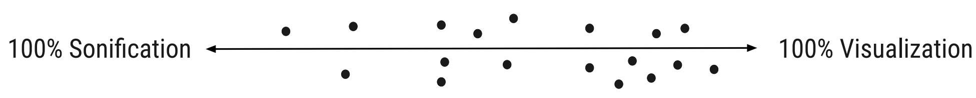


Figure 5: The visualization shows the self-assessment of the participant's core research field. Each dot represents a researcher and where they see themselves on a continuum between sonification and visualization. Overall, 18 researchers participated in this poll at the virtual Audio Mostly 2021 conference, and 11 of them tended toward the visualization side. The vertical positions of the marks are not informative.

at Audio Mostly 2021, placed themselves towards the visualization end of a continuum between sonification and visualization. Figure 5 displays the self-assessed placement of AM21 participants between the two disciplines. Three years later, hence post-pandemic, our team invited sonification experts to provide their expertise in an interdisciplinary panel discussion at EuroVis 2024. Unfortunately, this time, the conference fees for the participation of the sonification experts in our panel were not waived, causing us to have one of the sonification researchers attend remotely instead of in person. In the spirit of interdisciplinary research, this is a call for an inclusive approach at our conferences. Such an inclusive approach will strengthen our communities, especially when experts from other fields offer their expertise in a panel or a hands-on workshop without presenting a paper of their own.

Retrospectively speaking, organizing and hosting the described events was a notable contribution to our research community. The mentioned events supported the development of a community, and in some cases, organizing these events even resulted in publications [AEI*22, EE22, SAR22, RF22, EK22, REI*24]. Most of the publications are made available via a Zenodo community page[12]. On a personal level, hosting AVAC events supported a development towards becoming a "dual citizen," placing one foot in the sonification and the other in the visualization community. Being such a dual citizen is essentially characterized by the ability to effectively communicate with members of both parties. Lastly, initiating AVAC led to a four-month research stay at Linköping University in Sweden in 2022, resulting in a broadened research network early in the project and various collaborations and exchange activities.

[12]zenodo.org/communities/audio-visual-analytics-community/ (accessed Oct. 10th, 2024)

5. Results in the Light of Design Science Research

This section reflects on the results emerging from this thesis in a two-fold way. On the one hand, using the lens of design science research to take a meta-perspective on the contributions to the field and its current state. On the other hand, discussing the research questions posed in Section 1.3. The present section takes a step back from the results presented in the individual articles and offers a discussion of this thesis' contributions through the lens of design science research.

5.1 Contributing to the Eight Components of a Design Theory

Chapter 2 introduced the eight components of a design theory suggested by Gregor and Jones [GJ07]. The following will discuss if and how the presented articles contribute to each of the eight components. Each of the paragraphs will begin with the definitions initially provided by the authors in Table 2 of [GJ07].

Purpose and Scope

"What the system is for," the set of meta-requirements or goals that specifies the type of artifact to which the theory applies and in conjunction also defines the scope, or boundaries, of the theory. [GJ07, p. 322]

In general, the theory discussed in this thesis is defined by the context of exploratory data analysis. Both monitoring and accessibility are adjacent fields with designs in partly different design spaces than exploratory analysis tools. In some cases, contributions, such as the definition of auditory marks and channels, could also be used for these adjacent fields.

Constructs

Representations of the entities of interest in the theory. [GJ07, p. 322]

According to the authors, the constructs could be "physical phenomena or abstract theoretical terms." Our theory publication[ERI*23] introduced auditory marks, auditory channels, and the substrate of sonification. The definition of these terms contributes to the basic linguistic constructs of an AVA design theory. They do not only affect the "constructs" component. The definition of these terms also affects other components of a design theory, such as principles of form and function, artifact mutability, and principles of implementation. This also shows the relevance of such basic "entities of interest" in a theory [GJ07].

Principles of Form and Function

The abstract "blueprint" or architecture that describes an IS artifact, either product or method/intervention. [GJ07, p. 322]

Gregor and Jones describe this component as referring to the principles that define "the structure, organization, and functioning of the design product or design method" [GJ07]. With the introduction of a versatile classification system, the STAR on audio-visual analytics idioms [EEC*24] contributes to the component of principles of form and function. Whether an idiom's purpose is the exploration of data, the presentation of data, or both, is a principle of form and function. Whether an idiom is designed in a complementary manner, hence mapping attributes of a dataset exclusively to only one of the senses, or it is designed redundantly is also such a principle of form and function. The classification system was initially developed to enable the analysis of AVA idioms. This same categorization system, however, also provides an overview of the design space and its available options to inform design decisions in the first place.

Artifact Mutability

The changes in state of the artifact anticipated in the theory, that is, what degree of artifact change is encompassed by the theory. [GJ07, p. 322]

A theory should account for the gradual evolution of technology and not fail whenever an artifact changes its state, such as when a feature is added. On the one hand, this relates to the evolution of individual artifacts that are updated during their development and potentially even after their deployment. On the other hand, and more importantly in the context of this thesis, artifact mutability means a theory should cover a design space that allows for different instances of artifacts. Hence, a design theory of audio-visual analytics should open a design space that offers more opinions than combining one exact type of visualization with one exact type of sonification. Our theoretical considerations, such as marks, channels, the substrate, communicative redundancy, reading level, or the purpose of an idiom, are at a basic level of design and will be relevant whether or not the underlying technology evolves.

Testable Propositions

Truth statements about the design theory. [GJ07, p. 322]

A design theory should allow the formulation of hypotheses testable by using instances of the design. In the case of audio-visual analytics idioms, this could be a statement such as "When the visualization X and a sonification Y are used in combination, the just noticeable difference (JND) between two data items will be smaller than when using only one of the two display techniques." This component of a design theory is strongly related to evaluation practices and techniques. The concrete proposition that we tested with the Parallel Chords idiom [EER*24] was that a user would be able to differentiate between two very similar datasets better when they have both a visualization and a sonification available. To test this hypothesis, we employed a staircase test method, which is established in the audio research community but not in the visualization community. Hence, in terms of assessing perceptual phenomena, we have also made methodological contributions to the field.
Another proposition is that the combination of sonification and visualization will affect the engagement of a user of such a system. While capturing user engagement is generally more complex than measuring metrics such as task completion times or error rates, it is another testable proposition to be studied in the future work of the community.

Justificatory Knowledge

The underlying knowledge or theory from the natural or social or design sciences that gives a basis and explanation for the design (kernel theories). [GJ07, p. 322]

Speaking of audio-visual analytics, justificatory knowledge is provided by psychophysics, computer graphics, cognitive science, electroacoustics, and other fields, which provide basic knowledge applied in the design of visualization and sonification. Both sonification and visualization are based on a solid foundation of justificatory knowledge individually, and we employed this knowledge when designing SoniScope and Parallel Chords and developing other theoretical contributions. While much of the justificatory knowledge will stem from the individual fields of visualization and sonification, some can also be based on studies of inherently multimodal perception. A popular example is the McGurk effect [MM76], where a visual stimulus modifies "what we hear." This thesis does not contribute to the psychophysical justificatory knowledge of sensual interdependencies and their impact on the design of AVA idioms. This component will be relevant for future work within the field.

Principles of Implementation

A description of processes for implementing the theory (either product or method) in specific contexts. [GJ07, p. 322]

Principles of implementation support a designer with the process of implementing an information system. Hence, this component is less concerned with the outcome of the design process than with guiding the process itself. The mathematical description of 0D and 1D auditory marks can be regarded as such a principle of implementation. The differentiation of marks that hold time-dependent channels (1D) and ones that hold time-independent channels (0D) is essential for the guided implementation of an appropriate synthesizer. The mathematical definition of different mark types essentially is a “mathematical description of the process for implementing the theory.” Put differently, when designers ask themselves how to implement a sonification required to display a data series that changes over time, they will need to use a synthesizing technique that allows for modifications while the sound is already playing. This relationship between the theoretical constructs and the implementation of sonification has been described in detail by Dennis Reinsch and Thomas Hermann in their publication on the mesonic framework [GJ07].

Expository Instantations

A physical implementation of the artifact that can assist in representing the theory both as an expository device and for purposes of testing. [GJ07, p. 322]

Expository instantiations of an audio-visual analytics design theory will be idioms that integrate sonification and visualization into one form of display. It is worth mentioning that Gregor and Jones also consider software to be such an instantiation, even if it is not "physical." The two AVA idioms designed during this project are examples of expository instantiations. They show how integrated designs can be designed with justificatory knowledge in mind. The SoniScope combines visual and auditory scatter plots and allows users to interactively explore their data. The SoniScope is an instantiation of the constructs of marks, channels, and substrates. Hence, it follows two purposes at the same time: It is an explanatory instance of the theory at hand, and it can be used to further develop the same theory. The same is true for the Parallel Chords idiom. With this design, we can explain theoretical constructs, and in the article, we provided an evaluation of testable propositions.

5.2 Is the Field Ready for Design Studies?

This section offers thoughts on the maturity of the field with respect to the applicability of the design study methodology suggested by Sedlmair et al. [SMM12]. Sedlmaier et al. define a design study as "a project in which visualization researchers analyze a specific real-world problem faced by domain experts, design a visualization system that supports solving this problem, validate the design, and reflect about lessons learned in order to refine visualization design guidelines" [SMM12, p. 2]. In recent years, several meta-studies have been published, discussing design studies and their methodological implications [Sed16, MD18, MD20, ALC*23]. Sedlmair himself, for example, characterizes seven different possible contributions that can result from design studies [Sed16]. His seven scenarios for design study contributions are the following: (1) propose a novel technique, (2) reflect on methods, (3) illustrate design guidelines, (4) transfer to other problems, (5) improve understanding of a VIS sub-area, (6) address a problem that your readers care about, and (7) strong and convincing evaluation. Regarding, for example, the "illustration of design guidelines," Sedlmair identifies the following challenge:

> Often, it can be hard to foresee whether there will be something interesting to say about guidelines at the beginning of a design study project. [...] In the process of reflecting on design guidelines, confirming and refining well-known guidelines such as “Boo 3D for abstract data”, “Boo rainbow colormaps”, or “Boo piecharts” seems to be the easier case. [Sed16, p. 4]

While Sedlmair focuses on the potential contributions of a design study, Meyer and Dykes [MD20] spent four years of research to be able to formulate criteria for rigor. When discussing the "informed"-criterion, they state that:

> It is important to approach design study with a *prepared mind* [40], that is, with a broad awareness of visualization idioms, design guidelines and methods, and assessment techniques [96]; the disciplinary underpinnings of visualization – core topics and area boundaries, ontological and epistemological positions, socio-cultural views [1, 91]; and relevant design materials like datasets, code, software, hardware, and physically manipulable materials. [MD20, p. 91]

In the selected articles, both Sedlmair [Sed16] as well as Meyer and Dykes [MD20] consider it to be relevant that guidelines and a broad understanding of the field do exist before conducting a design study. The objective to "refine visualization design guidelines" is even baked into the very definition of design studies [SMM12]. Regarding audio-visual analytics, we cannot consider the existing literature to provide the community with a "broad awareness of audio-visual analytics idioms, design guidelines and methods, and assessment techniques." On the contrary, only recently did the community have little structured knowledge about existing audio-visual analytics idioms, design guidelines, methods, or assessment techniques. Such a knowledge base existed in visualization and sonification research on an individual but not on an integrated level.

While the community was (and still is) poorly equipped with guidelines and assessment techniques, the situation towards theory building resembles the famous "chicken and egg" problem. In design science research, do we need a design theory before we can do design studies, or do we need design studies to build a design theory? Intuitively, both approaches seem plausible. The studies composing this thesis prioritized the development of fundamental theoretical knowledge over specific knowledge on solutions for particular domain problems, which can be considered to be in line with Meyer and Dykes' call for a "prepared mind" when conducting design studies [MD20]. With the theoretical constructs [ERI*23], we have provided a systematic approach to classify and discuss audio-visual analytics idioms. With the STAR [EEC*24], we have, for the first time, gained structured knowledge about the existing design practice in the field. With the lab study on Parallel Chords, we have suggested a method for evaluating audio-visual analytics designs under lab conditions [EER*24]. All of these contributions can be considered steps toward a more mature field of audio-visual analytics, one that will eventually be ready to conduct rigorous design studies. Whether or not that is the case by now is a discussion we should have in our community. An extensive corpus of design studies will, eventually, help find robust answers to RQ 3, discussed in the following section.

5.3 Discussion of the Research Questions

This section discusses how the presented publications contribute to answering the three research questions posed in Section 1.3.

Research Question 1

Can we identify basic theoretical constructs that build a bridge between sonification and visualization? How do we define such constructs to offer terminology for the formal analysis and description of AVA idiom design?

With our theoretical contributions in [ERI*23], we identify and discuss three theoretical constructs that enable the formal description of audio-visual analytics idioms on a fundamental and structural level. The three constructs are the marks, channels, and substrates. While these are not the only possible constructs that could potentially bridge the fields, they are widely established in the visualization literature and lend themselves to the description of sonification idioms just as well. With the definition of auditory and visual marks, channels, and substrates, we provide a low-level design space description of AVA idioms that proved useful to our research group, as well as to other researchers who adopted the constructs into their work.

Research Question 2

What is the current state of the art of audio-visual analytics, both with respect to its idiom design space and its research community? Can we identify apparent research gaps from the current state of the art?

Our state-of-the-art report [EEC*24] provides, for the first time in the field, a structured overview of the current design space in audio-visual analytics research. A team of international and interdisciplinary researchers studied AVA idioms, introducing and employing a classification system for such idioms. The classification justifies the conclusion that, in general, the design space of AVA idioms is diverse with respect to both sonification and visualization parts. The vast majority of sonification techniques are parameter mapping sonifications. Nevertheless, within those, the authors employ different channels and approaches to convey their data. A research gap that became apparent was the integration of sonification into XR visualization settings. The branch of research that studies the potential of XR for visual analytics should lend itself perfectly to the integration with sonification. A common challenge here is the spatially restricted view of users who could be guided using spatial audio technologies [ZF19]. With the

STAR, we also shed light on the state of collaboration in the community. We identified the community as composed mainly of disjoint and small research groups rarely working together on multiple publications. Nevertheless, the development of publication numbers in recent years and the fact that domain experts seem to be involved in many of the studied project teams indicate a promising future in the field. The report also identifies and discusses the adjacent topics of monitoring, accessibility, and data art as inspirational fields, even if not covered systematically in the article.

Research Question 3

In the context of exploratory data analysis, can sonification support visualization challenges and offer solutions? Can we describe examples and what advantages emerge from combined designs?

To satisfy an extensive answer to this question, the design and evaluation of various expository instantiations of AVA idioms is required. With a considerable amount of examples, we will likely be able to identify classes of visualization challenges and solutions that are offered by integrating sonification into the idiom design. Also, a classification of advantages that emerge from such combinations will likely be derived from a corpus of expository instances. While it goes beyond the scope of the present thesis to design and evaluate a large corpus of examples, our article on "Parallel Chords" [EER*24] describes such an expository instantiation of an audio-visual analytics idiom. The article introduces an idiom integrating parallel coordinate visualization with interactive parameter mapping sonification. The article discusses the design process of the idiom, describes a usage scenario with prototypical data patterns, and offers results from a user study. The challenge we tackled in this specific case was that a user could only see the relationship between adjacent axes in the single view of a parallel coordinates plot. The additional sonification enables the exploration of relationships between non-adjacent axes without updating the visual view. Speaking of advantages of AVA idioms, we should distinguish between those that can be captured with user performance metrics, such as task completion times of just noticeable differences, and those that should be studied on a qualitative level, such as user engagement or affective response. On a user performance level, we did not identify an advantage of the sonification when compared individually to the visualization. On the user experience level, we see an indication of increased enjoyment and confidence when using sonification and visualization together. Studying such "soft" idiom qualities will be a relevant topic when thinking of a fruitful future in the field. This and other pressing topics are described in the following section.

6. Imagining a Future in Audio-Visual Data Analytics

Drawing from the articles presented earlier and in relation to the eight components of a design theory, this section discusses a selected set of topics that are relevant to be studied by the community in their collective future work. Some of the topics are described in our STAR as well. Nevertheless, they are relevant enough to be mentioned here again.

Towards a Culture of Evaluation and Documentation

It will be essential for the thriving future of audio-visual analytics research to establish a culture of evaluation. Out of 57 papers covered in our STAR [EEC*24], 20 papers do not offer any evaluation, only eleven offer a user experience evaluation, and only twelve studied user performance. This lack of evaluations is critiqued regularly within the sonification community, and Nees [Nee19] has related evaluation to the testable propositions component by Gregor and Jones [GJ07]. If the sonification community wants to establish such a culture, then one option is to implement explicit rules for publishing full papers at ICAD and ACM Audio Mostly. Such a rule could be that a paper can only be accepted as a full paper if an evaluation has been conducted. Otherwise, it must be presented as a poster contribution. With ICAD and ACM Audio Mostly happening together for the first time in 2025, this might be an excellent moment to instantiate new rules for the future of both conferences.

In our STAR, we found only 29 out of 57 projects that provided demos of their designs, which are still available online today. Such demos can be, for example, videos, audio recordings, or interactive websites. They are especially valuable whenever researchers wish to understand the design decisions that were made in someone else's work. Documentation is closely related to the topic of replicability. In the sonification community, the topic of replicability has been discussed previously by Degara et al. [DQNH13], Nees [Nee19], and recently in our replication study [EFH24] of Flower's et al. study from 1997 [FBT97]. In a project led by Katharina Groß-Vogt [GEm23], we developed criteria for "effective sonification" and, to do so, studied a large amount of sonification literature. Working through the literature made us realize most papers could not be replicated from the available information alone.

Affective Engagement: An Alternative Objective

Currently, the communities' research efforts seem to focus on "hard metrics," such as task completion times or just noticeable differences. In contrast, some studies have found only small or no perceptual advantages when combining sonification and visualization. On the other hand, they show differences in enjoyment and other "soft metrics" [EER*24, RJ16, Rön21]. While the community should continue to study potential perceptual advantages of combining sonification and visualization, we should also investigate the potential of other objectives, such as increasing a user's affective engagement with the display and the data.

Definitions of engagement are vast: In a review of 351 articles on engagement in HCI, Dohrethy and Dohrethy [DD19] found 102 different definitions of engagement. We generally distinguish three kinds of engagement [KH18]: cognitive, behavioral, and emotional or affective engagement. A low-level definition of affective engagement with visualization is provided by Hung and Pearsons [HP18], describing it as "the user's emotional involvement or investment while interacting with a visualization." In data analysis, and related to the definition above, we often wish for users to spend more time with their data instead of less. This can be true in an industrial context where unexpected phenomena might be hidden in the data, as well as for readers of online data journalism seeking to understand specific phenomena. Research shows that emotionally captured users will likely spend more time with the display as well [KH18]. With sound being an inherently affective medium, it is worth studying if AVA idioms have the potential to capture a user emotionally, potentially increasing the time spent with exploration.

In the core visualization community, systematically studying user engagement and affective visualization is fairly new [MKK15, LWC24]. Nevertheless, in recent years, research on affective design has gained visibility and interest within the community. With their recently published state-of-the-art report, Lan et al. show that the number of publications on the topic has become a multitude of what was the case a decade ago [LWC24].

Two of the research opportunities that Lan et al. [LWC24] discuss are especially relevant also to the future work of the audio-visual analytics community: (O1) How to evoke emotion responsibly and ethically, and (O2) how techniques such as multimodal interaction, including sound, influence emotion [LWC24].

Intentionally emotionalizing viewers comes with the necessity to discuss ethical implications, which is especially relevant considering modern issues around misinformation [LPLK23]. In the context of data analysis, we seem to consider

emotion as opposed to rationality. While this perspective is plausible at first glance, research shows that "emotion is not the enemy of rationality [LWC24]." Even if emotion and rationality were considered integrated concepts, ethical implications and how to responsibly engage one's audience are topics of utmost importance.
Regarding the second opportunity (O2) mentioned by Lan et al. [LWC24] and with respect to sonification, Ballora [Bal14] discussed the 'wow'-effect of sonification. Ballora states that "sonification's potential value, like much of the scientific visualisation content, probably lies less in hard facts and more in how it may serve as a stimulant for curiosity" [Bal14, p. 30]. Ballora's statement relates to the term "exploranation" coined by Ynnerman et al. [YLT18]. In their publication, Ynnerman et al. discuss the fusion of exploration and explanation to a "new science communication paradigm" they call exploranation. Using interactive scientific visualization on touch displays, the authors increased the behavioral engagement of museum audiences with digital and physical artifacts. They could show that audiences spent more time studying the actual exhibit if they had previously explored the digital version of the same artifact. Again, with sound having inherently affective qualities, a visualization accompanied by sonification could potentially increase a viewer's engagement with the display and indirectly with the data. Studying the phenomena around (affective) user engagement when combining the two forms of display will be a timely endeavor for our community.

Effective Task Distributions

The data from our STAR revealed that most existing AVA idioms use visualization to provide an overview of the data, while sonification conveys details about the data. It is worth studying the potential of breaking this pattern to provide a user with an auditory overview and visual details. Beyond that, the design space of AVA idioms is vast, and the options to distribute tasks to the senses are manifold. When we have a corpus of evaluated design studies available in the future, we will (hopefully) be able to better understand how to distribute data analysis tasks to the senses most effectively.

AVA Idioms and Extended Reality Environments

Again, the data from our STAR shows that only four out of 57 papers combined visualization and sonification in an extended reality environment. The possibilities we have to provide a user with three-dimensional sound via headphones [ZF19] suggest the combination of sonification and visualization in XR. In XR, a user often cannot keep an overview of all their spatially distributed data, which is a problem that could be met using sonification and auditory guidance towards interesting data points.

Integrating Sonification and Scientific Visualization

While this thesis focuses on the subfield of information visualization, the integration of sonification and scientific visualization is especially promising. Scientific visualization often deals with the display of inherently spatial and spatiotemporal data. When it comes to exploring our daily surroundings, our ears and eyes are phenomenal instruments. This circumstance and selected examples from the literature [TMN*21, MNW*18] are inspirational to study the potential of this combination in our collective future work.

A Final Thought

While the sonification field does not lack psychoacoustical evidence for the potential of sound to be an information carrier, we still do not see sonification evolve into a mature field with numerous users outside of the academic context. Visualization is the predominant and widely established technique to represent abstract data, and it is only plausible that integration with visualization might be fruitful soil for sonification. This also becomes clear through the fact that many sonifications implicitly come with a visual representation of data or a visual interface. Hence, visual displays are so dominant that even a community explicitly studying another form of display inherently falls back into using the visual sense. At the same time, few sonification designs have explicitly considered such visualizations as an integral element of their designs.

Subjectively speaking, the sonification community sometimes seems to fall into a "smaller sibling" mode when discussing visualization. While a "we can do this too" mentality can be justified occasionally, embracing visualization as a useful type of display does not disqualify sonification. Explicitly combining sonification and visualization could elevate both kinds of display into a form that can be "more than the sum of the two" and could even help identify the core value of sonification.

Bibliography

[AB04] Alais D., Burr D.: The Ventriloquist Effect Results from Near-Optimal Bimodal Integration. *Current Biology 14*, 3 (Feb. 2004), 257–262. doi:10.1016/j.cub.2004.01.029.

[ACJ24] Armitage J., Crozzoli M., Jones D.: Artificial Life in Integrated Interactive Sonification and Visualisation: Initial Experiments with a Python-Based Workflow. In *The 29th International Conference on Auditory Display (ICAD 2024)* (Troy, NY, USA, 2024). doi:10.21785/icad2024.003.

[AEI*22] Aigner W., Enge K., Iber M., Rind A., Elmqvist N., Höldrich R., Rönnberg N., Walker B. N.: Workshop on Audio-Visual Analytics. In *Proceedings of the 2022 International Conference on Advanced Visual Interfaces* (Frascati, Rome, Italy, 2022), ACM. doi:10.1145/3531073.3535252.

[ALC*23] Akbaba D., Lange D., Correll M., Lex A., Meyer M.: Troubling Collaboration: Matters of Care for Visualization Design Study. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (Hamburg, Germany, Apr. 2023), ACM, pp. 1–15. doi:10.1145/3544548.3581168.

[BAA*15] Brand A., Allen L., Altman M., Hlava M., Scott J.: Beyond authorship: Attribution, contribution, collaboration, and credit. *Learned Publishing 28*, 2 (Apr. 2015), 151–155. doi:10.1087/20150211.

[Bal14] Ballora M.: Sonification, Science and Popular Music: In search of the 'wow'. *Organised Sound 19*, 1 (Apr. 2014), 30–40. doi:10.1017/S1355771813000381.

[Ber83] Bertin J.: *Semiology of Graphics: Diagrams Networks Maps.* University of Wisconsin Press, Madison, Wisconsin, USA, 1983.

[Bla96] Blauert J.: *Spatial Hearing: The Psychophysics of Human Sound Localization.* The MIT Press, Oct. 1996. doi:10.7551/mitpress/6391.001.0001.

[Bre90] Bregman A. S.: *Auditory Scene Analysis: The Perceptual Organization of Sound.* MIT press, Cambridge, MA, USA, 1990.

[BSK*24] Boucher M., Stoiber C., Kejstova M., Kandlhofer M., Ertl A., Kriglstein S., Aigner W.: Mapping the Landscape of Data Visualizations in Schools and Educational Resources. In *EuroVis 2024* (2024), The Eurographics Association. doi:10.2312/eved.20241052.

[CLR22] Caiola V., Lenzi S., Riccò D.: Audiovisual sonifications: A design map for multisensory integration in data representation. In *DRS2022: Bilbao* (2022), Lockton D., Sara L., Hekkert P., Oak A., Sádaba J., Lloyd P., (Eds.). doi:10.21606/drs.2022.380.

[CM97] Card S. K., Mackinlay J.: The structure of the information visualization design space. In *Proc. IEEE Symp. Information Visualization, InfoVis* (1997), pp. 92–99. doi:10.1109/INFVIS.1997.636792.

[CMS99] Card S. K., Mackinlay J., Shneiderman B. (Eds.): *Readings in Information Visualization: Using Vision to Think.* Morgan Kaufmann, San Francisco, USA, 1999.

[Cro01] Cross N.: Design/Science/Research: Developing a Discipline. *Proceedings of the Korea Society of Design Studies Conference* (2001), 16–24.

[CWB10] Chang M. H., Wang G., Berger J.: Sonification and visualization of neural data. In *Proceedings of the 16th International Conference on Auditory Display (ICAD-2010)* (2010), pp. 201–205.

[DD19] Doherty K., Doherty G.: Engagement in HCI: Conception, Theory and Measurement. *ACM Computing Surveys 51*, 5 (Sept. 2019), 1–39. doi:10.1145/3234149.

[DQNH13] Degara Quintela N., Nagel F., Hermann T.: Sonex: An Evaluation Exchange Framework for Reproducible Sonification. In *Proceedings of the 19th International Conference on Auditory Display* (Łódź, Poland, 2013). doi:10.13140/2.1.4998.3680.

[EE22] Elmquist E., Enge K.: Towards the Combination of Visualization and Sonification for Cylindrical Displays. In *AVI 2022 Workshop on Audio-Visual Analytics* (Frascati, Rome, Italy, 2022), p. 5. doi:10.5281/zenodo.6553824.

[EEB*24] Elmquist E., Ejdbo M., Bock A., Thaler D. S., Ynnerman A., Rönnberg N.: Birdsongification: Contextual and Complementary Sonification for Biology Visualization. In *The 29th International Conference on Auditory Display (ICAD 2024)* (Troy, NY, USA, 2024). doi:10.21785/icad2024.006.

[EEC*24] Enge K., Elmquist E., Caiola V., Rönnberg N., Rind A., Iber M., Lenzi S., Lan F., Höldrich R., Aigner W.: Open Your Ears and Take a Look: A State-of-the-Art Report on the Integration of Sonification and Visualization. *Computer Graphics Forum (EuroVis '24) 43*, 3 (2024), e15114. doi:10.1111/cgf.15114.

[EER*24] Elmquist E., Enge K., Rind A., Navarra C., Höldrich R., Iber M., Bock A., Ynnerman A., Aigner W., Rönnberg N.: Parallel Chords: An Audio-Visual Analytics Design for Parallel Coordinates. *Personal and Ubiquitous Computing* (May 2024). doi:10.1007/s00779-024-01795-8.

[EF10] Elmqvist N., Fekete J.-D.: Hierarchical Aggregation for Information Visualization: Overview, Techniques, and Design Guidelines. *IEEE Transactions on Visualization and Computer Graphics 16*, 3 (May 2010), 439–454. doi:10.1109/TVCG.2009.84.

[EFH24] Enge K., Fabry L., Höldrich R.: Flowers Revisited: A preliminary replication of Flowers et al. 1997. In *The 29th International Conference on Auditory Display (ICAD 2024)* (Troy, NY, USA, 2024). doi:10.48550/arXiv.2407.11992.

[EK22] Engeln L., Keck M.: Exploring Sketch-based Sound Associations for Sonification. In *AVI 2022 Workshop on Audio-Visual Analytics (WAVA22)* (Frascati, Rome, Italy, June 2022), Zenodo. doi:10.5281/zenodo.6576697.

[ERI*21] Enge K., Rind A., Iber M., Höldrich R., Aigner W.: It's about Time: Adopting Theoretical Constructs from Visualization for Sonification. In *Proceedings of the 16th International Audio Mostly Conference (AM'21)* (2021), ACM, pp. 64–71. doi:10.1145/3478384.3478415.

[ERI*22] Enge K., Rind A., Iber M., Höldrich R., Aigner W.: Towards Multimodal Exploratory Data Analysis: SoniScope as a Prototypical Implementation. In *Proc. 24th Eurographics Conference on Visualization (EuroVis 2022) - Short Papers* (Rome, 2022), The Eurographics Association, pp. 67–71. doi:10.2312/evs.20221095.

[ERI*23] Enge K., Rind A., Iber M., Höldrich R., Aigner W.: Towards a unified terminology for sonification and visualization. *Personal and Ubiquitous Computing 27*, 5 (Oct. 2023), 1949–1963. doi:10.1007/s00779-023-01720-5.

[FBT97] Flowers J. H., Buhman D. C., Turnage K. D.: Cross-Modal Equivalence of Visual and Auditory Scatterplots for Exploring Bivariate Data Samples. *Human Factors: The Journal of the Human Factors and Ergonomics Society 39*, 3 (Sept. 1997), 341–351. doi:10.1518/001872097778827151.

[FdLM09] Franchin W. J., de Souza Watanabe L., Levkowitz H., Minghim R.: Development, Implementation, and Evaluation of Sonification Tools for Point-and-Surface-Based Data Exploration. In *Proc. 2nd Int. Conf. Visualisation* (Barcelona, 2009), IEEE, pp. 3–9. doi:10.1109/VIZ.2009.36.

[FG03] Findlay J. M., Gilchrist I. D.: *Active Vision: The Psychology of Looking and Seeing.* Oxford University Press, Aug. 2003.

[Flo05] Flowers J. H.: Thirteen years of reflection on auditory graphing: Promises, pitfalls, and potential new directions. In *Proceedings of ICAD 05-Eleventh Meeting of the International Conference on Auditory Display,* (Limerick, Ireland, July 2005), Georgia Institute of Technology.

[GEm23] Gross-Vogt K., Enge K., m. Zmölnig Io.: Reflecting on qualitative and quantitative data to frame criteria for effective sonification design. In *Proceedings of the 18th International Audio Mostly Conference* (New York, NY, USA, Oct. 2023), AM '23, ACM, pp. 93–100. doi:10.1145/3616195.3616233.

[GJ07] Gregor S., Jones D.: The anatomy of a design theory. *Journal of the Association for Information Systems 8*, 5 (2007), 312–335. doi:10.17705/1jais.00129.

[Gre15] Gregory R. L.: *Eye and Brain : The Psychology of Seeing - Fifth Edition.* Princeton University Press, 2015.

[HAR16] Hildebrandt T., Amerbauer F., Rinderle-Ma S.: Combining Sonification and Visualization for the Analysis of Process Execution Data. In *2016 IEEE 18th Conference on Business Informatics (CBI)* (Aug. 2016), vol. 02, pp. 32–37. doi:10.1109/CBI.2016.47.

[HHN11] Hermann T., Hunt A., Neuhoff J. G. (Eds.): *The Sonification Handbook.* Logos, Bielefeld, 2011.

[HP18] Hung Y.-H., Parsons P.: *Affective Engagement for Communicative Visualization: Quick and Easy Evaluation Using Survey Instruments.* Oct. 2018.

[HR21] Hermann T., Reinsch D.: Sc3nb: A Python-SuperCollider Interface for Auditory Data Science. In *Proceedings of the 16th International Audio Mostly Conference* (New York, NY, USA, Oct. 2021), AM '21, ACM, pp. 208–215. doi:10.1145/3478384.3478401.

[Hun07] Hunter J. D.: Matplotlib: A 2D Graphics Environment. *Computing in Science & Engineering 9*, 3 (May 2007), 90–95. doi:10.1109/MCSE.2007.55.

[ID90] Inselberg A., Dimsdale B.: Parallel coordinates: A tool for visualizing multi-dimensional geometry. In *Proc. 1st Conference on Visualization '90* (1990), pp. 361–378. doi:10.1109/VISUAL.1990.146402.

[IIJ*13] Isenberg T., Isenberg P., Jian Chen, Sedlmair M., Moller T.: A Systematic Review on the Practice of Evaluating Visualization. *IEEE Transactions on Visualization and Computer Graphics 19*, 12 (Dec. 2013), 2818–2827. doi:10.1109/TVCG.2013.126.

[Ins85] Inselberg A.: The plane with parallel coordinates. *The Visual Computer 1*, 2 (Aug. 1985), 69–91. doi:10.1007/BF01898350.

[JP14] Johannesson P., Perjons E.: *An Introduction to Design Science.* Springer, 2014.

[KH18] Kennedy H., Hill R. L.: The Feeling of Numbers: Emotions in Everyday Engagements with Data and Their Visualisation. *Sociology 52*, 4 (Aug. 2018), 830–848. doi:10.1177/0038038516674675.

[Koh95] Kohonen T.: *Self-Organizing Maps*, vol. 30 of *Springer Series in Information Sciences.* Springer, Berlin, Heidelberg, 1995. doi:10.1007/978-3-642-97610-0.

[Kra94] Kramer G. (Ed.): *Auditory Display: Sonification, Audification and Auditory Interfaces.* Addison-Wesley, Reading, Mass, 1994.

[Kub81] Kubovy M.: Concurrent-pitch segregation and the theory of indispensable attributes. In *Perceptual Organization.* Routledge, London, 1981, pp. 55–98.

[KvKM17] Kondo H. M., van Loon A. M., Kawahara J.-I., Moore B. C. J.: Auditory and visual scene analysis: An overview. *Philosophical Transactions of the Royal Society B: Biological Sciences 372*, 1714 (Feb. 2017), 20160099. doi:10.1098/rstb.2016.0099.

[KVV01] Kubovy M., Van Valkenburg D.: Auditory and visual objects. *Cognition 80*, 1 (June 2001), 97–126. doi:10.1016/S0010-0277(00)00155-4.

[KWB*99] Kramer G., Walker B., Bonebright T., Cook P., Flowers J. H., Miner N., Neuhoff J., et al.: *Sonification Report: Status of the Field and Research Agenda*. Report for the NSF, International Community for Auditory Display, 1999.

[LC24] Lenzi S., Ciuccarelli P.: Designing tools for designers: The Data Sonification Canvas. In *DRS2024: Boston* (June 2024). doi:10.21606/drs.2024.730.

[Lev71] Levitt H.: Transformed Up-Down Methods in Psychoacoustics. *The Journal of the Acoustical Society of America 49*, 2B (Feb. 1971), 467–477. doi:10.1121/1.1912375.

[LPLK23] Lisnic M., Polychronis C., Lex A., Kogan M.: Misleading Beyond Visual Tricks: How People Actually Lie with Charts. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (Hamburg Germany, Apr. 2023), ACM, pp. 1–21. doi:10.1145/3544548.3580910.

[LWC24] Lan X., Wu Y., Cao N.: Affective Visualization Design: Leveraging the Emotional Impact of Data. *IEEE Transactions on Visualization and Computer Graphics 30*, 1 (Jan. 2024), 1–11. doi:10.1109/TVCG.2023.3327385.

[LZ24] Linke S., Ziemer T.: Somson — Sonification of Multidimensional Data in Kohonen Maps. In *The 29th International Conference on Auditory Display (ICAD 2024)* (2024). doi:10.21785/icad2024.008.

[Mac86] Mackinlay J.: Automating the design of graphical presentations of relational information. *ACM Trans. Graphics 5*, 2 (1986), 110–141. doi:10.1145/22949.22950.

[MD18] Meyer M., Dykes J.: Reflection on Reflection in Applied Visualization Research. *IEEE Computer Graphics and Applications 38*, 6 (Nov. 2018), 9–16. doi:10.1109/MCG.2018.2874523.

[MD20] Meyer M., Dykes J.: Criteria for Rigor in Visualization Design Study. *IEEE Transactions on Visualization and Computer Graphics 26*, 1 (Jan. 2020), 87–97. doi:10.1109/TVCG.2019.2934539.

[MF95] Minghim R., Forrest A.: An illustrated analysis of sonification for scientific visualisation. In *Proceedings Visualization '95* (Oct. 1995), pp. 110–117. doi:10.1109/VISUAL.1995.480802.

[MKK15] Mahyar N., Kim S.-H., Kwon B. C.: Towards a taxonomy for evaluating user engagement in information visualization. In *Workshop on Personal Visualization: Exploring Everyday Life* (2015), vol. 3, p. 4.

[MM76] Mcgurk H., Macdonald J.: Hearing lips and seeing voices. *Nature 264*, 5588 (Dec. 1976), 746–748. doi:10.1038/264746a0.

[MNW*18] MacDonald D. E., Natarajan T., Windeyer R. C., Coppin P., Steinman D. A.: Data-Driven Sonification of CFD Aneurysm Models.

In *Proc. 24th International Conference on Auditory Display (ICAD 2018)* (Houghton, Michigan, USA, June 2018), The International Community for Auditory Display, pp. 28–33. `doi:10.21785/icad2018.010`.

[Mun15] Munzner T.: *Visualization Analysis and Design.* CRC Press, Boca Raton, 2015.

[NB02] Nesbitt K. V., Barrass S.: Evaluation of a multimodal sonification and visualisation of depth of market stock data. In *Proceedings of the 8th International Conference on Auditory Display* (Kyoto, Japan, 2002).

[Nee18] Nees M. A.: Auditory Graphs Are Not the "Killer App" of Sonification, But They Work. *Ergonomics in Design: The Quarterly of Human Factors Applications 26*, 4 (Oct. 2018), 25–28. `doi:10.1177/1064804618773563`.

[Nee19] Nees M. A.: Eight Components of a Design Theory of Sonification. In *Proceedings of the 25th International Conference on Auditory Display (ICAD 2019)* (Newcastle upon Tyne, Great Britain, June 2019), Northumbria University, pp. 176–183. `doi:10.21785/icad2019.048`.

[Nes01] Nesbitt K. V.: Modeling the multi-sensory design space. In *Proceedings of the 2001 Asia-Pacific Symposium on Information Visualisation - Volume 9* (AUS, Dec. 2001), APVis '01, Australian Computer Society, Inc., pp. 27–36.

[Nes04] Nesbitt K.: MS-taxonomy: A conceptual framework for designing multi-sensory displays. In *Proceedings. Eighth International Conference on Information Visualisation, 2004. IV 2004.* (London, England, 2004), IEEE, pp. 665–670. `doi:10.1109/IV.2004.1320213`.

[Nes06] Nesbitt K. V.: Modelling human perception to leverage the reuse of concepts across the multi-sensory design space. In *Proceedings of the 3rd Asia-Pacific Conference on Conceptual Modelling – Volume 53* (Darlinghurst, Australia, 2006), Apccm '06, Australian Computer Society, pp. 65–74.

[Neu04] Neuhoff J. G. (Ed.): *Ecological Psychoacoustics.* Elsevier Academic Press, San Diego, USA, 2004.

[Owe98] Owen C. L.: Design research: Building the knowledge base. *Design Studies 19*, 1 (Jan. 1998), 9–20. `doi:10.1016/S0142-694X(97)00030-6`.

[PLP*24] Pile S., Lesota O., Peter S. D., Humer C., Gasser M.: Spin-Wave Voices: Sonification of Nanoscale Spin Waves as an Engagement and Research Tool. In *The 29th International Conference on Auditory Display (ICAD 2024)* (Troy, NY, USA, June 2024).

[RC15] Rosli M. H. W., Cabrera A.: Gestalt Principles in Multimodal Data Representation. *IEEE Computer Graphics and Applications 35*, 2 (Mar. 2015), 80–87. `doi:10.1109/MCG.2015.29`.

[REI*24] Rind A., Enge K., Iber M., Rönnberg N., Lenzi S., Elmquist E., Caiola V., Lan F., Höldrich R., Aigner W.: Integrating Sonification and Visualization - But Why? In *EuroVis 2024 - Panel* (2024), Schulz H.-J., Isenberg T., (Eds.), The Eurographics Association. `doi:10.2312/evt.20241100`.

[RF22] Rönnberg N., Forsell C.: Questionnaires assessing usability of audio-visual representations. In *AVI 2022 Workshop on Audio-Visual Analytics (WAVA22)* (Frascati, Rome, Italy, June 2022), Zenodo. `doi:10.5281/zenodo.6555676`.

[RFJ*90] Rabenhorst D. A., Farrell E. J., Jameson D. H., Linton Jr T. D., Mandelman J. A.: Complementary visualization and sonification of multidimensional data. In *Extracting Meaning from Complex Data: Processing, Display, Interaction* (1990), vol. 1259, International Society for Optics and Photonics, pp. 147–153. `doi:10.1117/12.19981`.

[RH22] Reinsch D., Hermann T.: Interacting with sonifications: The mesonic framework for interactive auditory data science. In *Proceedings of the 7th Interactive Sonification Workshop (ISon)* (2022), CITEC, Bielefeld University, p. 65.

[RH23] Reinsch D., Hermann T.: Sonecules: A Python Sonification Architecture. In *Proceedings of the 28th International Conference on Auditory Display (ICAD2023)* (Linköping, Sweden, June 2023), International Community for Auditory Display, pp. 62–69. `doi:10.21785/icad2023.5580`.

[RJ16] Rönnberg N., Johansson J.: Interactive Sonification for Visual Dense Data Displays. In *Proc. 5th Interactive Sonification Workshop, ISon* (2016), CITEC, Bielefeld University, pp. 63–67.

[Rön21] Rönnberg N.: Sonification for conveying data and emotion. In *Proceedings of the 16th International Audio Mostly Conference* (New York, NY, USA, 2021), AM '21, ACM, pp. 56–63. `doi:10.1145/3478384.3478387`.

[RW10] Roberts J. C., Walker R.: Using All Our Senses: The need for a Unified Theoretical Approach to Multi-sensory Information Visualization. In *IEEE VisWeek 2010 Workshop: The Role of Theory in Information Visualization, Salt Lake City, Utah* (Oct. 2010).

[SAR22] Svoronos-Kanavas I., Agiomyrgianakis V., Rönnberg N.: An exploratory use of audiovisual displays on oceanographic data. In *AVI 2022 Workshop on Audio-Visual Analytics (WAVA22)* (Frascati, Rome, Italy, June 2022), Zenodo. `doi:10.5281/zenodo.6555839`.

[Sed16] Sedlmair M.: Design Study Contributions Come in Different Guises: Seven Guiding Scenarios. In *Proceedings of the Sixth Workshop on Beyond Time and Errors on Novel Evaluation Methods for Visualization* (New York, NY, USA, Oct. 2016), BELIV '16, ACM, pp. 152–161. `doi:10.1145/2993901.2993913`.

[Shn03] Shneiderman B.: The Eyes Have It: A Task by Data Type Taxonomy for Information Visualizations. In *The Craft of Information Visualization*, Bederson B. B., Shneiderman BEN., (Eds.), Interactive Technologies. Morgan Kaufmann, San Francisco, USA, Jan. 2003, pp. 364–371. `doi:10.1016/B978-155860915-0/50046-9`.

[Sim08] Simon H. A.: *The Sciences of the Artificial*, 3 ed. MIT Press, Cambridge, Mass., USA, 2008.

[SKS02] Shams L., Kamitani Y., Shimojo S.: Visual illusion induced by sound. *Cognitive Brain Research 14*, 1 (June 2002), 147–152. doi:10.1016/S0926-6410(02)00069-1.

[SMM12] Sedlmair M., Meyer M., Munzner T.: Design Study Methodology: Reflections from the Trenches and the Stacks. *IEEE Transactions on Visualization and Computer Graphics 18*, 12 (Dec. 2012), 2431–2440. doi:10.1109/TVCG.2012.213.

[SMS*24] Schütz L., Matinfar S., Schafroth G., Navab N., Fairhurst M., Wagner A., Wiestler B., Eck U., Navab N.: A Framework for Multimodal Medical Image Interaction. *IEEE Transactions on Visualization and Computer Graphics* (2024), 1–11. doi:10.1109/TVCG.2024.3456163.

[SMWH17] Satyanarayan A., Moritz D., Wongsuphasawat K., Heer J.: Vega-Lite: A grammar of interactive graphics. *IEEE Trans. Visualization and Computer Graphics 23*, 1 (2017), 341–350. doi:10.1109/TVCG.2016.2599030.

[Spe07] Spence R.: *Information Visualization: Design for Interaction.* Pearson/Prentice Hall, 2007.

[TC05] Thomas J. J., Cook K. A.: *Illuminating the Path: The Research and Development Agenda for Visual Analytics.* National Visualization and Analytics Ctr, 2005.

[TMN*21] Temor L., MacDonald D. E., Natarajan T., Coppin P. W., Steinman D. A.: Perceptually-motivated sonification of spatiotemporally-dynamic CFD data. In *Proc. 26th International Conference on Auditory Display (ICAD 2021)* (Virtual Conference, June 2021), International Community for Auditory Display, pp. 202–209.

[Tuk77] Tukey J. W.: *Exploratory Data Analysis.* Addison-Wesley, 1977.

[TVY90] Takeda H., Veerkamp P., Yoshikawa H.: Modeling Design Process. *AI Magazine 11*, 4 (Dec. 1990), 37–37. doi:10.1609/aimag.v11i4.855.

[UN 24] UN DESA: *The Sustainable Development Goals Report 2024.* UN DESA, New York, NY, USA, June 2024. URL: https://unstats.un.org/sdgs/report/2024/.

[vdH03] van der Heijden A. H. C.: *Attention in Vision: Perception, Communication and Action.* Psychology Press, London, Dec. 2003. doi:10.4324/9780203491522.

[VK15] Vaishnavi V., Kuechler W.: *Design Science Research Methods and Patterns: Innovating Information and Communication Technology, 2nd Edition.* CRC Press, 2015.

[Weg90] Wegman E. J.: Hyperdimensional Data Analysis Using Parallel Coordinates. *Journal of the American Statistical Association 85*, 411 (Sept. 1990), 664–675. doi:10.1080/01621459.1990.10474926.

[Wer23] Wertheimer M.: Untersuchungen zur Lehre von der Gestalt. II. *Psychologische Forschung 4*, 1 (1923), 301–350. doi:10.1007/BF00410640.

[Wil05] Wilkinson L.: *The Grammar of Graphics*, 2 ed. Springer, New York, NY, USA, 2005.

[WWES92] Walls J. G., Widmeyer G. R., El Sawy O. A.: Building an Information System Design Theory for Vigilant EIS. *Information Systems Research 3*, 1 (Mar. 1992), 36–59. doi:10.1287/isre.3.1.36.

[YLT18] Ynnerman A., Löwgren J., Tibell L.: Exploranation: A New Science Communication Paradigm. *IEEE Computer Graphics and Applications 38*, 3 (May 2018), 13–20. doi:10.1109/MCG.2018.032421649.

[ZF19] Zotter F., Frank M.: *Ambisonics: A Practical 3D Audio Theory for Recording, Studio Production, Sound Reinforcement, and Virtual Reality.* Springer Nature, Graz, Austria, 2019.

[ZLN24] Zhao Y., Lu J., Nacenta M. A.: Speech-based Mark for Data Sonification. In *Proceedings of the 26th International ACM SIGACCESS Conference on Computers and Accessibility* (New York, NY, USA, Oct. 2024), ASSETS '24, ACM, pp. 1–5. doi:10.1145/3663548.3688514.

A. List of Publications

The following papers were published during my PhD research:

- K. Enge, A. Rind, M. Iber, R. Höldrich, and W. Aigner, "It's About Time: Adopting Theoretical Constructs from Visualization for Sonification," in Proceedings of the 16th International Audio Mostly Conference (AM'21), ACM, 2021, pp. 64–71. doi: 10.1145/3478384.3478415.
- K. Enge, A. Rind, M. Iber, R. Höldrich, and W. Aigner, "Towards Multimodal Exploratory Data Analysis: SoniScope as a Prototypical Implementation," in Proc. 24th Eurographics Conference on Visualization (EuroVis 2022) - Short Papers, Rome: The Eurographics Association, 2022, pp. 67–71. doi: 10.2312/evs.20221095.
- E. Elmquist and K. Enge, "Towards the Combination of Visualization and Sonification for Cylindrical Displays," in AVI 2022 Workshop on Audio-Visual Analytics, Frascati, Rome, Italy, 2022, doi: 10.5281/zenodo.6553824.
- K. Enge, A. Rind, M. Iber, R. Höldrich, and W. Aigner, "Towards a Unified Terminology for Sonification and Visualization," Pers. Ubiquit. Comput., 2023, doi: 10.1007/s00779-023-01720-5.
- K. Groß-Vogt, K. Enge, and IOhannes. m. Zmölnig, "Reflecting on qualitative and quantitative data to frame criteria for effective sonification design," in Proceedings of the 18th International Audio Mostly Conference, in AM '23. New York, NY, USA: Association for Computing Machinery, Oct. 2023, pp. 93–100. doi: 10.1145/3616195.3616233.
- E. Elmquist, K. Enge, et al., "Parallel Chords: an audio-visual analytics design for parallel coordinates," Pers. Ubiquit. Comput., 2024, doi: 110.1007/s00779-024-01795-8.
- K. Enge et al., "Open Your Ears and Take a Look: A State-of-the-Art Report on the Integration of Sonification and Visualization," Computer Graphics Forum (EuroVis '24), vol. 43, no. 3, p. e15114, 2024, doi: 10.1111/cgf.15114.
- K. Enge, L. Fabry, and R. Höldrich, "Flowers Revisited: A preliminary replication of Flowers et al. 1997," in The 29th International Conference on Auditory Display (ICAD 2024), Troy, New York, 2024. doi: 10.48550/arXiv.2407.11992

B. *CRediT* author statement

The Contributor Roles Taxonomy (*CRediT*) enables teams of co-authors to make individual contributions transparent. The original publication [BAA*15], as well as the *CRediT* wesbite[1] provide explanations of the individual types of contributions. Table 1 displays my (not necessarily exclusive) contributions to the individual articles composing this thesis.

Publications	Conceptualization	Data Curation	Formal Analysis	Funding Acquisition	Investigation	Methodology	Project Administration	Resources	Software	Supervision	Validation	Visualization	Writing – Original Draft	Writing – Review & Editing
Enge et al. (2023) [ERI*23]	✓	✓	✓	✓	✓	✓	✓	✓			✓	✓	✓	✓
Enge et al. (2024) [EEC*24]	✓	✓	✓	✓	✓	✓	✓	✓	✓		✓	✓	✓	✓
Elmquist, Enge et al. (2024) [EER*24] *	✓	✓	✓	✓	✓	✓	✓	✓	✓		✓	✓	✓	✓
Enge et al. (2022) [ERI*22]	✓	✓	✓	✓	✓	✓	✓	✓	✓		✓	✓	✓	✓
Elmquist, Enge (2022) [EE22] *	✓			✓	✓	✓	✓				✓	✓	✓	✓
Groß-Vogt et al. (2023) [GEm23]	✓		✓	✓	✓	✓						✓		✓
Enge et al. (2024) [EFH24]	✓		✓	✓		✓				✓	✓	✓	✓	✓

Table 1: Publications and individual (not necessarily exclusive) contributions. The top three journal articles formally compose the present thesis. The Contributor Role Taxonomy is described in detail in [BAA*15]. Articles marked with an asterisk (*) are published with shared first authorship.

In more detail: In [EER*24], Elias Elmquist contributed most to the software, while I contributed most of the formal statistical analysis. [EFH24] originates from the bachelor's thesis of Liam Fabry, who designed the software and collected and curated all of the data. In [EEC*24], the original drafts of Sections 5.1 Accessibility, 5.2 Monitoring, and 5.3 Arts were written by Niklas Rönnberg, Michael Iber, and Sara Lenzi, respectively.

[1] credit.niso.org (accessed Nov. 20th, 2024)

C. List of AVAC Events

Between 2021 and 2024, I co-organized and co-hosted several events at international conferences to bridge the gap between the sonification and the visualization research communities. For this purpose, we started the Audio-Visual Analytics Community (AVAC). Furthermore, in 2022, I hosted three online meet-ups that allowed selected community members to present their current work, followed by an extended discussion.

- **September 3, 2021:** W. Aigner, K. Enge et al., "Workshop on Audio-Visual Analytics – Identifying Research Gaps for Integrating Sonification and Visualization," in Audio Mostly Workshops, 2021.
- **October 25, 2021:** W. Aigner, K. Enge et al., "Workshop on Audio-Visual Analytics - Towards a Research Agenda for Integrating Sonification and Visualization," in IEEE VIS Workshops, 2021.
- **June 7, 2022:** W. Aigner, K. Enge et al., "Workshop on Audio-Visual Analytics," in Proceedings of the 2022 International Conference on Advanced Visual Interfaces, in AVI 2022. New York, NY, USA: ACM, Jun. 2022, p. 92:1-92:4. doi: 10.1145/3531073.3535252
- **October 20, 2022:** W. Aigner, K. Enge et al., Audio-Visual Analytics: Potential Applications of Combined Sonifications and Visualizations [Application Spotlight]. IEEE VIS 2022, Oklahoma City, US.
- **June 29, 2023:** M. Iber, K. Enge et al., "Audio-Visual Analytics Community: Bringing Senses to Mind | Panel Discussion at ICAD 2023," presented at the 28th International Conference on Auditory Display, 2023, Norrköping, Sweden
- **May 30, 2024:** A. Rind, K. Enge et al., "Integrating Sonification and Visualization - But Why?," EG EuroVis 2024 Panel Discussion, The Eurographics Association, 2024. doi: 10.2312/evt.20241100
- **January 25, 2022 | Meet-Up:** Victor Schetinger & Elias Elmquist
- **April 6, 2022 | Meet-Up:** Rafael Bresciani & Jordan Wirfs-Brock and Maxene Graze
- **May 25, 2022 | Meet-Up:** Iason Svoronos-Kanavas & Sara Lenzi and Paolo Ciuccarelli

Original Article 1

Towards a unified terminology for sonification and visualization

The following article [ERI*23] was published in the Journal on Personal and Ubiquitous Computing[1] and is an extended version of our contribution to the proceedings of the 2021 Audio Mostly Conference "It's about Time: Adopting Theoretical Constructs from Visualization for Sonification" [ERI*21].

[1]link.springer.com/journal/779

Personal and Ubiquitous Computing
https://doi.org/10.1007/s00779-023-01720-5

ORIGINAL ARTICLE

Towards a unified terminology for sonification and visualization

Kajetan Enge[1,2] · Alexander Rind[1] · Michael Iber[1] · Robert Höldrich[2] · Wolfgang Aigner[1]

Received: 31 March 2022 / Accepted: 19 March 2023

Abstract
Both sonification and visualization convey information about data by effectively using our human perceptual system, but their ways to transform the data differ. Over the past 30 years, the sonification community has demanded a holistic perspective on data representation, including audio-visual analysis, several times. A design theory of audio-visual analysis would be a relevant step in this direction. An indispensable foundation for this endeavor is a terminology describing the combined design space. To build a bridge between the domains, we adopt three of the established theoretical constructs from visualization theory for the field of sonification. The three constructs are the *spatial substrate*, the *visual mark*, and the *visual channel*. In our model, we choose time to be the *temporal substrate* of sonification. *Auditory marks* are then positioned in time, such as visual marks are positioned in space. *Auditory channels* are encoded into auditory marks to convey information. The proposed definitions allow discussing visualization and sonification designs as well as multi-modal designs based on a common terminology. While the identified terminology can support audio-visual analytics research, it also provides a new perspective on sonification theory itself.

Keywords Sonification theory · Visualization theory · Audio-visual data analysis

1 Introduction

Designers of sonification systems can nowadays base their work on a solid foundation of research on auditory perception and several sonification techniques such as auditory icons, parameter mapping, and model-based sonification [2, 3]. Thus, a theory of sonification already has an articulated set of design constructs at its disposal [4]. However, we argue that constructs at a more basic level are missing from the current stage of scientific dialogue. This seems to be especially relevant for the design, description, and evaluation of combinations of sonification and visualization.

This article[1] proposes *channels* encoded into *marks* that are positioned in a *substrate* as basic constructs for designing sonifications. The theoretical model is adopted from the visualization literature [5–7], where channels, marks, and spatial substrate are widely used constructs. They allow the description of the extensive design space of visualization approaches using only a small set of atomic building blocks, and have thus been successfully used as framework for guidelines (e.g., [7]), software tools (e.g., [8]), and toolkits (e.g., [9, 10]), as well as automatic recommendation of visualizations (e.g., [11–13]).

Theoretical cross-pollination between visualization and sonification is most reasonable because both fields share similar goals. While sonification is "the use of nonspeech audio to convey information" [14], visualization is defined as "the use of computer-supported, interactive, visual representations of abstract data to amplify cognition" [6].

Kajetan Enge
kajetan.enge@fhstp.ac.at

1 Institute of Creative Media Technologies, FH St. Pölten, Campusplatz 1, St. Pölten 3100, Austria

2 Institute of Electronic Music and Acoustics, University of Music and Performing Arts Graz, Leonhardstraße 15, Graz 8010, Austria

1 This article is an extended version of our contribution to the proceedings of the 2021 Audio Mostly Conference *It's about Time: Adopting Theoretical Constructs from Visualization for Sonification* (https://doi.org/10.1145/3478384.3478415) [1]

Unsurprisingly, sonifications are often employed together with visualizations in real-world scenarios, for instance, by diagnostic ultrasonic devices. However, too little attention has been paid to the theoretical underpinnings of audio-visual data analysis approaches [15]. Such approaches essentially use both our vision and our auditory sense in combination to convey information about data. Bridging terminological barriers between the research communities is a reasonable step towards a combined design theory with compatible basic constructs and making progress in both fields.

There are, however, fundamental differences between our visual and auditory perception [15]. For example, with regard to spatial resolution, auditory perception is less accurate than visual perception [16]. Sound is an inherently temporal phenomenon [17–20] unlike vision. Therefore, adaptations of the model of channels, marks, and the substrate are needed.

This article starts with related work (Section 2) and an introduction to the constructs of the substrate, marks, and channels from visualization literature (Section 3). Section 4 investigates how equivalent constructs can be defined for the sonification domain and provides a mathematical description of auditory marks. In Section 5, we discuss analogies between sonification and visualization practice emerging from our model and analyze existing designs from sonification and visualization literature with our model. Before we conclude in Section 7, we argue for the rejection of space and frequency as substrates for sonification in Section 6.

With this article, we propose a new way to describe combinations of visualization and sonification. A terminology that uses the same basic constructs will help members of both communities with discussing their work and with combining their knowledge.

Our original paper [1] has been extended by

- A discussion of the construct of auditory channels,
- A discussion of frequency as a potential substrate for sonification, and
- A demonstration of the unified terminology by describing existing work using the adopted constructs.

2 Related work

There are numerous examples of designs that combine sonification and visualization and many of them can be found via the "Data Sonification Archive" via https://sonification.design. Recently, Caiola et al. [21] analyzed 80 examples of audio-visual designs leading towards their definition of an "audiovisual design map," meant to support the integration of sonification and visualization. Hildebrandt et al. [22] combined visualization and sonification to analyze business process execution data. Rabenhorst et al. [23] augmented a vector field visualization with sonification. Chang et al. used an audio-visual approach to explore the activity of neurons in the brain [24]. In 2003, Hermann et al. presented "AVDisplay" [25], a system for monitoring processes in complex computer network systems including both sonifications and visualizations. MacVeigh and Jacobson [26] described "a way to incorporate sound into a raster-based classified image." They augmented a map with further dimensions through sonification.

Taken together, the abovementioned works support the notion that visualization and sonification can be combined for effective data analysis. Nesbitt introduced a taxonomy for the multi-modal design space [27–31]. He proposed essentially two ways to describe the multimodal design space, including haptic displays. The first is an extension of the reference model for visualization by Card, Mackinlay, and Shneiderman [6], which we also choose as our reference in this article. In his extended design space, Nesbitt uses space as the substrate for visual, auditory, and haptic displays. His second description of the multi-modal design space is based on three types of metaphors: spatial metaphors, temporal metaphors, and direct metaphors [31]. These categories take into account the inherent temporal structure of sound. While Nesbitt introduced a new description of the multimodal design space, in this article, we suggest using time instead of space as the substrate of sonification and adopting the vocabulary from visualization theory, as will be argued in the following.

Compared to visualization, sonification is a considerably younger discipline [32]. This might be one of the causes why its theoretical foundation is not as developed even though both disciplines pursue very similar goals [4]. In sonification, some of the milestones in theory development have been the "Proceedings of the 1st Conference on Auditory Display" in 1992, which were edited in the book *Auditory Display* in 1994 [33], marking the beginning of systematic research on sonification by the international community for auditory display. Barrass' dissertation in 1997 [34] introduced task analysis, data characterization, and a case-based design method to the community. The sonification report in 1999 [14] provided an overview of the field at the time and a definition of sonification that is still widely used. Walker [35] worked on magnitude estimation and mapping-polarity of conceptual data dimensions in 2002 and Hermann [36] studied sonification in the context of exploratory data analysis. The book *Ecological Psychoacoustics*, edited by Neuhoff in 2004 [37], provides a more holistic perspective on psychoacoustics than conventional laboratory studies could offer. The design space map introduced by de Campo in 2007 [38] helps a designer decide on an appropriate sonification technique with respect to the number of data items and attributes to be sonified. Hermann's taxonomy from 2008 [3] provides a detailed definition of sonification and auditory display in a scientific context. The *Sonification Handbook* gave another overview of the field in 2011 [2], and Worrall's *Sonification Design* [39] put another focus on both theory and design of sonifications in 2019.

However, in 2019, Nees [4, p. 176] stated that "[...] sonification theory remains so underdeveloped that even the path to advance theory-building for sonification remains unclear." He then refers to an article by Gregor and Jones [40] as inspiration for the development of a sonification design theory. Gregor and Jones describe eight components that any design theory should include, specifically, (1) purpose and scope, (2) constructs, (3) principle of form and function, (4) artifact mutability, (5) testable propositions, (6) justificatory knowledge, (7) principles of implementation, and (8) expository instantiation.

In this sense, our article focuses on the *constructs* of a design theory, as they are especially relevant for a combined terminology of sonification and visualization. Gregor and Jones [40, p.33] describe the constructs: "The representations of the entities of interest in the theory [...] are at the most basic level in any theory. These entities could be physical phenomena or abstract theoretical terms." The state of the art of the eight components for a design theory of sonification is well described in the 2019 paper by Nees [4].

In our work, we intend to contribute to the development of a design theory for the combination of sonification and visualization by offering low-level constructs for the description of sonification designs. We do so by adopting some of the elaborated theoretical constructs from visualization theory for the domain of sonification. In the following section, we introduce these constructs: the spatial substrate, the mark, and the channel.

3 Basic theoretical constructs in visualization theory

Since the design space of possible visualization solutions is extensive, the visualization community has worked on theoretical models to formalize design knowledge [7]. Based on Bertin's seminal book *Semiology of Graphics* [5], many visualization models (e.g., [6, 7, 9, 11, 41]) are centered around marks as the basic building blocks of visualization techniques. In general terms, a mark is a geometric object that represents the attributes of a data object by position, color, or other visual features.

The widely adopted reference model for visualization by Card, Mackinlay, and Shneiderman [6] provides the more specific formalism needed for a transfer to the field of sonification. It dissects visualization as a pipeline of data transformations from raw data to a visual form perceived by humans. In the center of this pipeline, there are *visual structures* that consist of marks positioned in a spatial substrate and channels that encode information to the marks' features. These visual structures are created from data tables and subsequently projected onto a view for display (Fig. 1).

3.1 Defining visual structures

The three components of a visual structure are the spatial substrate, marks, and channels.

Channels such as position and color encode the information of the data table's attributes into the visual features of the marks. Besides spatial position, Bertin [5] enumerates six non-positional channels: size, color hue, color gray scale value, shape, orientation/angle, and texture; yet further channels are possible (e.g., color saturation, curvature, motion [7]). The reference model originally refers to channels as "graphical properties" and the visualization literature contains a number of further synonyms such as "perceptual attributes" or "visual variables," yet "channel" seems to be most widely used [7, p. 96]. Since spatial position allows very effective encoding for visual perception, the reference model conceptualizes it as a substrate "into which other parts of a Visual Structure are poured" [6, p. 26].

The *spatial substrate* is the conceptual space where marks are positioned. While it is most often a two-dimensional (2D) space, a conceptual three-dimensional (3D) spatial substrate can also be projected on a 2D view for display on a computer screen or viewed on a virtual reality device. Different types of axes and nesting mechanisms subdivide the spatial substrate.

The reference model distinguishes four elementary types of *marks*: points (zero-dimensional, 0D), lines

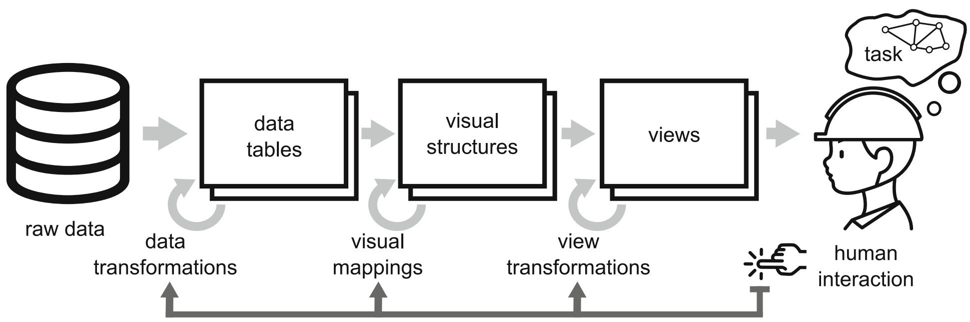


Fig. 1 The reference model for visualization [6] introduces visual structures as an intermediate state in mapping data to visual representations (figure from [1], CC BY). Reusing the icon "engineer" by Pawinee E. from Noun Project, CC BY 3.0

(one-dimensional, 1D), areas and surfaces (2D), and volumes (3D). Marks can have as many dimensions as their containing substrate; therefore, surfaces and volumes occur only in 3D substrates. Furthermore, the visualization reference model introduces special mark types to encode connection (e.g., in a node-link diagram) and containment (e.g., in a Venn diagram). For example, the dots in a 2D scatter plot are point marks (0D) positioned along two orthogonal quantitative axes, and in the same plot, an area mark (2D) can represent a range of values along both axes (Fig. 2). The countries in a choropleth map are also area marks positioned in a geographical spatial substrate. An example of 1D marks is the line in a line plot or isolines on a geographic map.

The distinction between mark types depends not only on their visual form but also on the data object represented by the mark—whether the data object encodes information for a point in the spatial substrate, or it encodes information about some extent of the spatial substrate. In fact, the rendered marks need to have some extent in all dimensions of the spatial substrate (e.g., 2D) because an infinitely small point or an infinitely thin line would not be visible.

Since the spatial extent of a point mark does not convey information per se, the mark is not constrained and can use the channel *size* to encode a data attribute. Yet another data attribute can be mapped to the channel *shape*, so that one category is shown as square and another as circle (Fig. 3). Neither the size nor the shape channel can be mapped to an area mark (cp. Fig. 2) because its spatial extent is constrained by the represented information.

Finally, these examples illustrate how the same visual form, in this case a rectangle, can represent either a data object positioned at a point with size and shape (Fig. 3) or a data object spanning an area in the spatial substrate (Fig. 2). To correctly interpret such graphics, contextual information is necessary that visualization designers need to provide via legends, annotations, or other onboarding approaches [42].

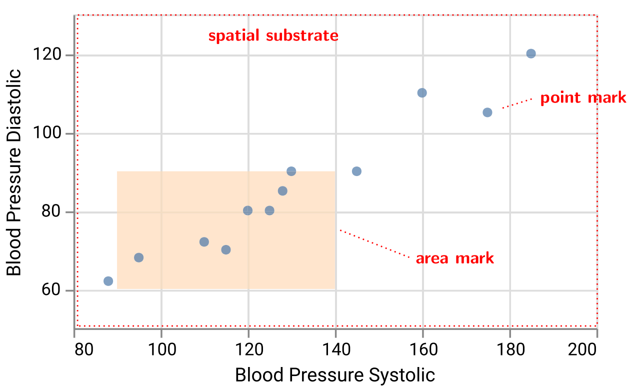


Fig. 2 Example scatter plot with blood pressure measurements (artificial data) as *points* (0D) and a rectangle representing the *area* (2D) of normal systolic and diastolic blood pressure (figure from [1], CC BY)

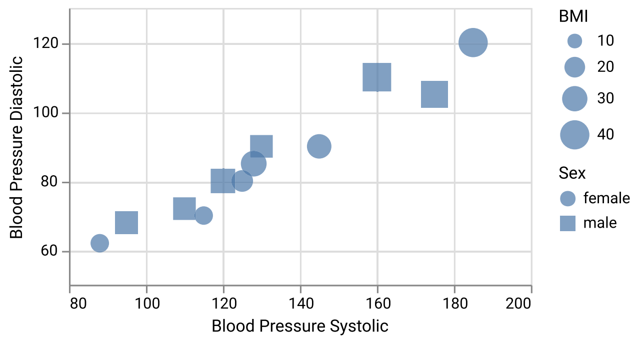


Fig. 3 Example scatter plot (artificial data) using size and shape as two channels. Note that rectangles and circles represent *point* marks (0D) (figure from [1], CC BY)

3.2 Applying visual structures

Within this conceptual model, the design space of visualization techniques stretches over all possible combinations of marks, spatial substrates, and channels. It provides a terminology to characterize existing techniques such as the scatter plot (Fig. 2) and to invent completely new techniques. Several visualization software frameworks apply these constructs to specify the visual encoding: e.g., Tableau [8], ggplot2 [43], RAWGraphs [44], or Vega-Lite [10].

The use of spatial substrates, marks, and channels ensures a consistent mapping from data to visual form, and thus promotes visual pattern recognition. The resulting graphic can be read as a whole, as individual marks, and at multiple intermediate levels [5]. For example, proximity on the spatial substrate and similarity of the color channel can be perceived as Gestalt. However, not all combinations of marks, substrates, and channels result in an effective representation of its underlying data. Yet, this conceptual model helps to systematically investigate the effectiveness of the visualization's components. For example, the experiments by Cleveland and McGill [45] found that the position channel was superior to length or angle in terms of accuracy.

Such results from empirical work can be distilled to design knowledge that is published as guidelines. For example, Mackinlay [11] ranks channels by their accuracy on perceptual tasks with quantitative, ordinal, and nominal data. Thus, he compares channels not only by their effectiveness, but also by their expressiveness. In another design guideline, Munzner [7] distinguishes *magnitude channels*, expressing quantitative or ordinal data, from *identity channels*, expressing categorical data, and ranks both by their relative effectiveness (Fig. 4). The position, size, and tilt of visual marks are conventional magnitude channels that inform about "how much of something there is"[7, p.99]. Color hue and shape are often used as identity channels, informing users about "what something is" [7, p.99]. Likewise, design knowledge is integrated into tools such as APT

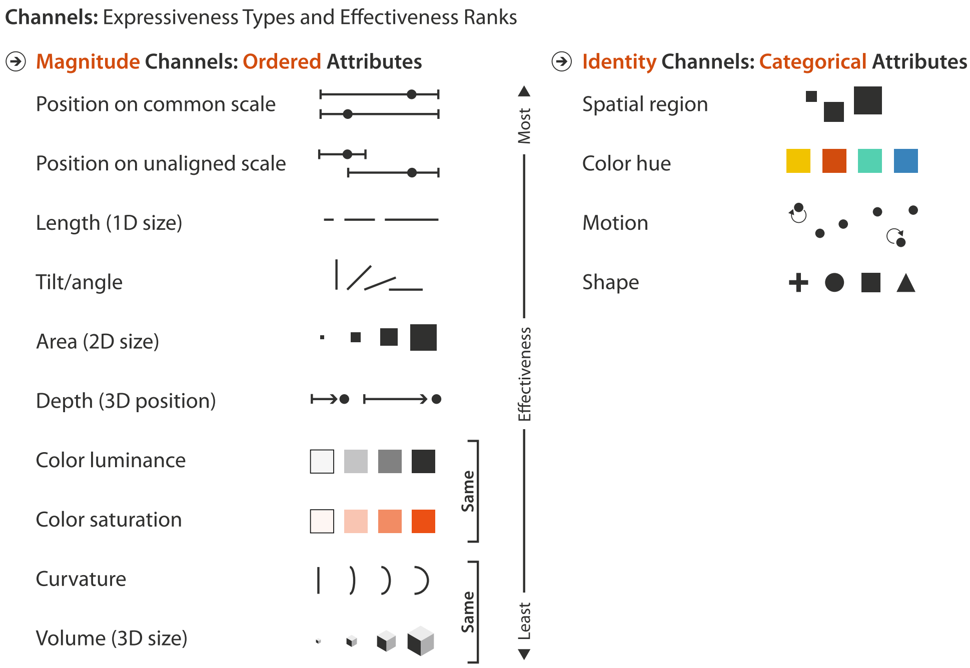


Fig. 4 Munzner's guideline to group visual channels into magnitude and identity channels and rank them by effectiveness [7, p. 102] (figure from "Visualization Analysis and Design" [7] by Tamara Munzner, with illustrations by Eamonn Maguire, AK Peters Visualization Series, CRC Press, 2014, CC BY 4.0.)

[11], Tableau [12], and Vega-Lite [13] for automated visualization recommendations.

Overall, marks, spatial substrates, and channels have shown to work well as a formal model for visualization techniques. We assume that these constructs lend themselves to formalizing sonification techniques as well, thus paving the way for creating audio-visual techniques for data analysis.

4 Adopting the constructs for sonification

To develop a combined design theory for audio-visual analytics, it is important to use common theoretical constructs. Such constructs define the terminology necessary to discuss audio-visual techniques at a conceptual level. In this section, we adopt the theoretical constructs that have been established in the visualization community for the field of sonification. First, we generalize the three constructs "substrate," "mark," and "channel": The substrate is the conceptual space on which a data representation is instantiated; it "holds" the marks. Marks are the perceptual entities of a data representation that can be distinguished by their conceptual expansion within their substrate. Channels are the parameters of a data representation encoded in a mark, carrying the information.

Next, this section investigates possible analogies for these constructs in sonification. On the one hand, in sonification, the construct of channels is relatively familiar with parameters such as loudness, pitch, or timbre [2, 35]. However, the two constructs of substrate and marks are not commonly used to describe a sonification. Since marks expand conceptually within their substrate, these two constructs are closely intertwined. As visualization uses space as a substrate, we will discuss the potentials and limitations of space and frequency as possible substrates for sonification in Section 6. However, the potential of time as the substrate for sonification has shown to be more promising.

4.1 Time as the substrate of sonification

Next to space, we have another fundamental dimension at our disposal: time. If we compare the dimensions space and time against each other, we find several arguments and analogies in support of time as the substrate for sonification.

Both time and space are physical dimensions inherently bound to our visual and auditory perception. However, with respect to sonification, spatial locatability is not necessary for the perception of a sound. When we hear a mono sound originating in front of us, we will hear it from the position of the loudspeaker. When we hear the same sound over headphones we will perceive it within our head (internalization). Our perception of the sound itself will not be altered; hence, the sonified information we perceive is consistent. Therefore, we argue that the perceived acoustic space is not inherently necessary for sonification. Time, on the other hand, is a dimension that we cannot even conceptually "switch off" while listening: A sonification that does not expand over time is not imaginable.

The opposite holds true for visualizations and space as their substrate: A visualization without spatial extent is not imaginable, while time is a dimension that can be conceptually "switched off" as long as the visualization is static (i.e., not using informative animation). Even though scanning a visualization involves eye movements at a rate between two to five saccades per second [46, p. 144–145] and analyzing a dataset is an iterative visual search process, the static visualization itself does not change over time.

Using this analogy, one can think of sounds being "positioned in time" in a sonification, just as visual marks are positioned in space. This is also supported by the fact that, with our eyes, we have a precise resolution for the relative spatial position of two visual objects, while with our ears, we have a far better temporal resolution for the relative position of two sounds. Furthermore, the temporal structure of sound is perceivable with only one ear, while generally we have to use both of our ears to detect spatial cues [16].

For these reasons, we consider time to be a suitable substrate for sonification and refer to it as the "temporal substrate." For the temporal substrate, it is not relevant whether the sonification is passively listened to or whether somebody interacts with it. In our model, time as a dimension is always considered to be linear. The follow-up question must be how to define types of auditory marks in a temporal domain.

4.2 Auditory marks

We know that visualization theory distinguishes its visual marks by their conceptual dimensionality, i.e., their conceptual extent within the spatial substrate. As has been shown, conceptual expansion does not have to be equal to physical expansion. Visual marks need to occupy space to become visible, even if conceptually they do not expand [5]. Correspondingly, it should be possible to distinguish auditory marks by their conceptual expansion within their substrate, time. Two more questions arise: How do we define conceptual expansion in time, and how many different types of auditory marks exist?

In visualization theory, the four mark types are "points," "lines," "areas," and "volumes" [6]. They represent all the possibilities for conceptual spatial expansion from 0D (no conceptual expansion) up to 3D (maximal possible conceptual expansion). While space is three-dimensional, time is one-dimensional. Thus, we define auditory marks that are 0D (no conceptual expansion) or 1D (maximal possible conceptual expansion). We cannot define 2D or 3D auditory marks, since time does not provide a second or third dimension for the marks to unfold in. We consider an auditory mark as 0D if it *does not conceptually expand* in time, just as a visual mark that does not expand in space is 0D. If an auditory mark *conceptually expands in time*, it is considered as 1D, equivalent to the definition of a visual mark.

For better readability, whenever we speak of an auditory mark, we automatically mean a temporal auditory mark, and whenever we speak of a visual mark, we mean a spatial mark. Following this logic, audio-visual data representations can use both visual marks, positioned on the spatial substrate, and auditory marks, positioned in the temporal substrate. Next, we will formally define 1D and 0D auditory marks and provide mathematical descriptions of both types.

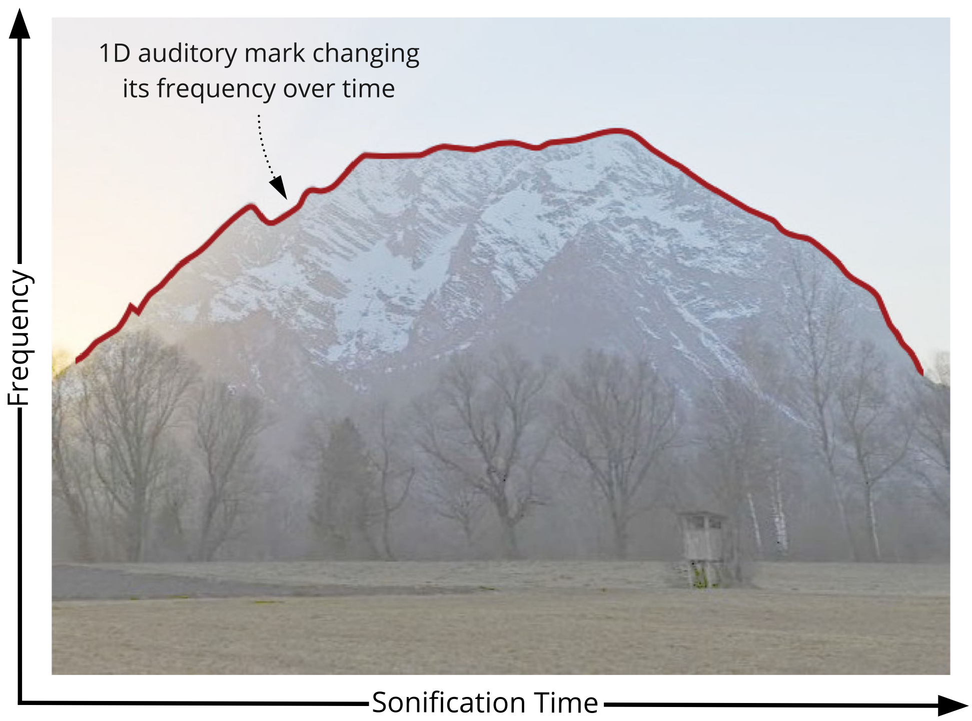


Fig. 5 The silhouette of the mountain "Grimming" in Austria. A 1D auditory mark maps the horizontal positions of the silhouette to time, and the height of the silhouette to the frequency of a sine wave. The horizontal positions correspond to the sortable attributes *k* and the height values to the attributes *x* from Fig. 6 and Eqn. 3 (figure from [1], CC BY)

4.2.1 1D auditory marks

A 1D auditory mark represents the data via its development over time. More precisely, the *temporal evolution* of a 1D auditory mark represents a dataset along one of the set's sorted attributes. It does so by evolving its channel(s) over time according to the sort, thus representing the evolution of attributes in the dataset. We regard the 1D auditory mark as "conceptually expanded in time" as it conveys information over time. The sorted attribute has to be a key attribute. A key attribute is a unique identifier for all items in a dataset. In a table, it could be, for example, the row number. This ensures that every item in the dataset is mapped to time bijectively.

An illustrative example of such a 1D auditory mark is shown in Fig. 5 via the silhouette of a mountain as a red line. Imagine a *parameter mapping sonification* [47], conveying information about the shape of the silhouette. The sonification maps the horizontal and vertical positions of the silhouette to the temporal and spectral evolution of a sine wave: Moving along the silhouette from west to east results in rising frequency whenever the mountain has an uphill slope, and falling frequency whenever it has a downhill slope. In such a case, we speak of an auditory graph as a special version of a parameter mapping sonification [48, 49]. In this example, the sonification uses a one-dimensional auditory mark, since its channel (frequency) evolves over time according to the development of the vertical position sorted along the horizontal position in the dataset.

We now have defined the theoretical construct of a 1D auditory mark that conceptually expands in its substrate, in time. We still have to provide a definition of the 0D auditory mark. Every sonification has to expand in time, but not all of them convey information over time. Auditory icons and earcons, for example, are sonification techniques that convey information without an inherent dependency on developments in the data [2]. They usually inform their users about states and will be further discussed in Section 4.3

4.2.2 0D auditory marks

A 0D auditory mark represents the data as a state in time, not as a development over time. More precisely, the *temporal evolution* of a 0D auditory mark does not represent a dataset along one of the set's sorted key-attributes. The 0D auditory mark still needs to physically expand in time to become audible, but its temporal evolution is not bijectively representing the data over time. This can be the case if, for example, (1) there is no sortable attribute in the data, or if (2) the sorted dataset is not mapped to sonification time. For further explanation, we construct two examples.

A so-called earcon [50] can typically be described as a 0D auditory mark. The sound of a computer after an error is such an earcon and its precise temporal evolution is not informative. Instead, the meaning of such a sound has to be learned as a whole. The earcon conveys information about a state in time, not a development over time. The instant in time that the sound occurs is a channel, just like the position of a visual mark in space is a channel. The auditory mark itself conceptually does not expand in time; therefore, we identify it as zero-dimensional.

Mapping the sorted data items to frequency instead of time would also result in a 0D auditory mark. To explain this, we can reuse the silhouette example from before. The abscissa in Fig. 5 would not be the sonification time but a frequency axis, and the ordinate would not be a frequency axis but the power spectral density. In this case, the silhouette bijectively maps to the shape of a sound's power spectral density, and the information is not encoded over time but into the spectral envelope of a static sound. This static sound is the 0D auditory mark, not evolving over time and therefore conceptually not expanded.

4.2.3 Mathematical description of auditory marks

We first want to describe the one-dimensional auditory mark in a more general mathematical way. Figure 6 shows an unsorted dataset that is first sorted and then transformed to become a 1D auditory mark. We refer to one of the attributes as k and to the other one as x. The attribute k is a key attribute, which means that it is a unique identifier that can be used to look up all items in a dataset [7].

$$k_i \neq k_j, \qquad \forall i \neq j. \tag{1}$$

To produce a one-dimensional auditory mark, k has to be sorted and mapped to sonification time via a strictly monotonically increasing function f (compare (2)). Sonification time is understood as the physical time which evolves during a sonification and is denoted as $\mathring{t}$. The ring symbol on top of $\mathring{t}$ helps to distinguish between sonification variables and domain variables. In our example, the domain variables are the horizontal and vertical positions k_i and x_i, while $\mathring{t}$ denotes the physical time that passes while listening to the auditory mark. This convention was first introduced by Rohrhuber [51], and then developed further by Vogt and Höldrich [52]. In the silhouette example, we used the horizontal positions k_i to sort the vertical positions x_i from west to east.

$$\mathring{t}_i = f(k_i), \tag{2}$$

We have now defined which position is mapped to which point in time. In the next step, we need to define the channel through which the mapping is realized. In our example, the channel $\mathring{c}(\mathring{t}_i)$ is the time-dependent frequency of a sine wave. Function $g(x_i)$ transforms the domain variable x, the vertical position, to the auditory channel *frequency* (compare [47,

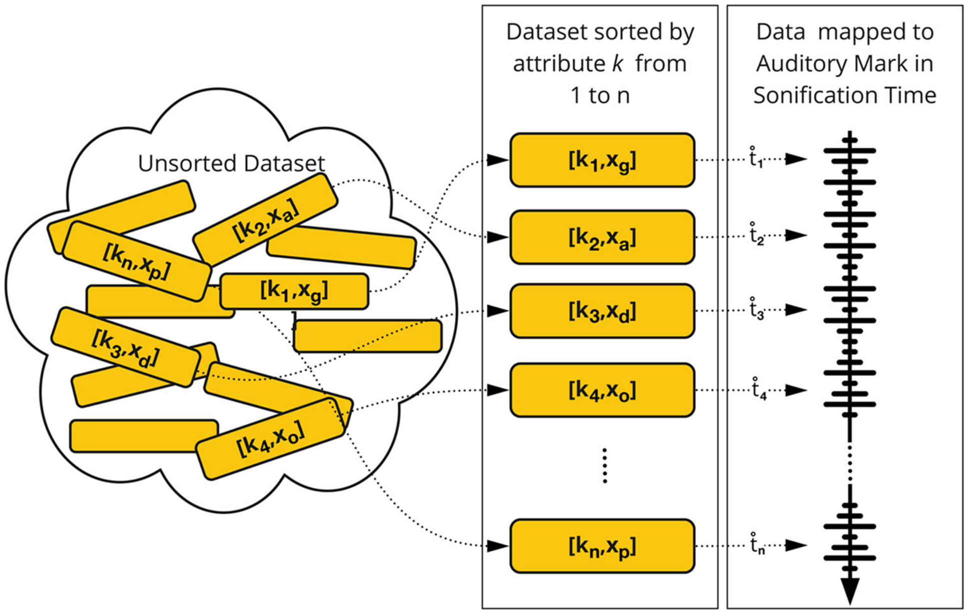


Fig. 6 An unsorted dataset is sorted and sonified to a 1D auditory mark, evolving over sonification time (figure from [1], CC BY)

p. 368]). To be called sonification, this transformation must be systematic, objective, and reproducible [3].

$$\mathring{c}(\mathring{t}_i) = \mathring{c}(f(k_i)) = g(x_i) \tag{3}$$

We usually deal with discrete data; therefore, some kind of interpolation between $\mathring{t}_i$ and $\mathring{t}_{i+1}$ will often be necessary. It is not necessary for $\mathring{t}_i$ to be equidistant, neither is it necessary for the interpolation to be linear. However, the mapping from the sorted attribute to sonification time has to be bijective; hence, every position on the silhouette must map to exactly one point in sonification time. Equation 4 formalizes the interpolation process with

$$\mathring{c}(\mathring{t}) = interp\left(\mathring{t}; \{\mathring{c}(\mathring{t}_i)\}\right), \quad \forall\, \mathring{t}_i < \mathring{t} < \mathring{t}_{i+1}. \tag{4}$$

Finally, the physical realization of a 1D auditory mark $\mathring{y}$ depends on the sonification time $\mathring{t}$ and the time-dependent channel $\mathring{c}(\mathring{t})$:

$$\text{1D auditory mark} = \mathring{y}\left(\mathring{t}; \mathring{c}(\mathring{t})\right) \tag{5}$$

A mathematical description is also possible for the 0D auditory mark. Function g is not mapping the attributes x_i to sonification time $\mathring{t}$, which leads to time-independent channels $\mathring{c}$.

$$\mathring{c} = g(x_i) \tag{6}$$

The comparison between (5) and (7) shows that 1D and 0D auditory marks differ in the time-dependency of their channels. The channels of 1D auditory marks are time-dependent; the channels of 0D auditory marks are not. Mathematically speaking, $\mathring{y}$ always depends on $\mathring{t}$, but $\mathring{c}$ does not have to depend on $\mathring{t}$.

$$\text{0D auditory mark} = \mathring{y}(\mathring{t}; \mathring{c}) \tag{7}$$

4.3 Auditory channels

The third construct we intend to adopt from visualization theory is the *channel*. Munzner [7, p. 96] describes a visual channel as “a way to control the appearance of marks, independent of the dimensionality of the geometric primitive.” Sonification designers also control the appearance of sounds (auditory marks) using parameters such as pitch, loudness, panning/spatial position, duration, or timbre [2, 53]. The sonification community has used several terms for these parameters, such as auditory or acoustic dimensions, auditory or acoustic parameters, sound dimensions, sound parameters, sonification parameters, display parameters, or perceptual parameters [2, 14, 52–56].

With the objective of a unified design theory of combined sonification and visualization, we argue for the usage of the same terminology in both fields: visual and auditory channels. Using this terminology, it is essential to distinguish between the auditory channel in the current context and the auditory channel as a synonym for auditory perception or even the ear canal. We use the term channel with inspiration from information theory, mapping information from a source (the data) to a receiver (the human) [57].

Following the description of visual channels, we describe auditory channels as “a way to control the appearance of auditory marks, independent of their dimensionality.” We argue that also in sonification we can distinguish between magnitude channels and identity channels. Pitch and loudness are often used magnitude channels, conveying information about “how much” of something there is. Timbre (e.g., instrumentation) is a conventional identity channel, informing the listener about “what” something is [53].

While it seems reasonable to describe a parameter mapping sonification with the construct of auditory channels, it is less intuitive to use them for the description of auditory icons [58, 59] or earcons [50]. According to Gaver [59], an auditory icon uses everyday sounds to represent information that is inherently connected to that everyday sound. Deleting a document on a Mac computer, for example, triggers the sound of paper being crumpled. Earcons, on the other hand, are [60, p. 7] “abstract, synthetic tones that can be used in structured combinations to create sound messages [...].” Examples are the tone sequences played back by PCs when connecting or removing a USB drive. These series of tones do not use a single distinct acoustic parameter but still convey (categorical) information.

The essence of auditory icons and earcons is their iconic and symbolic qualities [50, 58]. Auditory icons, as icons in general, resemble their referent by an ecological connection. Symbols and earcons, on the other hand, represent their referent by a connection that has to be learned first. Everyday sounds (auditory icons) as well as tone sequences and instrumentations (earcons) could be connotated and perceived in a biased way depending on sociocultural contexts. Nevertheless, independently of such potential biases, we argue that both techniques generally use identity channels like they are described by Munzner [7].

The recording of an everyday sound such as a bird chirp can be used as an auditory icon, being interpretable due to its ecological connection to our memory of bird sounds. We identify an audio recording of a bird sound as resembling a bird, just as we identify the visual icon of a bird because it resembles the shape of a bird. Following this logic, auditory icons use an identity channel constructed from the timbre of the sound. Based on the definition given by the Acoustic Society of America, Pratt and Doak [61] refine the term timbre as "that attribute of acoustic sensation whereby a listener can judge that two sounds are dissimilar using any criteria other than pitch, loudness or duration." To think of timbre as an identity channel is also supported by a connection between “color” and “timbre” in the German language. The German word for timbre is “Klangfarbe,” which can be literally translated to “sound color.” Hence, the German language enables us to differentiate between “colors of sounds” by using a term that typically describes the acoustic qualities of instruments. It is a common practice in sonification to use different timbres (e.g., different instruments) to differentiate between items or attributes of data.

Both in visualization and in sonification, marks can combine identity channels and magnitude channels to encode more attributes. A visual point mark can use color hue as identity channel and size as magnitude channel, and in sonification an auditory mark can combine the timbre of an oboe (identity) with variable pitch (magnitude). While an auditory icon is inherently using an identity channel, it can still be parameterized with a magnitude channel, as shown, for example, by the sonification of planetary data of Elmquist et al. [62]. In such a case, an auditory icon would use for example the loudness as an additional magnitude channel to convey continuous data.

Now that we have discussed the three constructs of substrates, marks, and channels, we will explore analogies between visualization and sonification and describe examples from the literature using the terminology we have found.

5 Analogies and examples

Using time as the substrate of sonification and defining marks to conceptually expand in time reveals several analogies between visualization theory and sonification. First of all, the two domains use the two most fundamental dimensions in physics, space and time, as their substrates. Table 1 shows substrates and mark types for both domains in a compact form. An analogy shows itself regarding the restrictions for a mark’s expansion. The size of a point mark does not have to be informative, so it could expand freely in size, without changing its meaning. A line mark, on the other hand, cannot change its length without changing its meaning. In our temporal definition of 0D and 1D auditory marks, we see a similar situation: A 0D auditory mark is free to expand in time, without changing its meaning, but a 1D auditory mark is not. Its duration is tied to the amount of data to be sonified. The position and size of a visual mark can be used as channels. In sonification, the instant in time and duration of an auditory mark can be channels. However, both in visualization and sonification, these parameters do not define the type of a mark. The type of mark depends on the conceptual expansion in their substrate. It is another analogy between visualization and sonification that information can be perceived on two levels: on the one hand from the appearance of individual marks, and on the other hand from Gestalts [63] that form perceptual artifacts through a group of marks with related channels. The correlation of two datasets resulting in a diagonal scatter plot is a typical example for a Gestalt in a visualization. A rhythmical pattern or a harmonic structure can be perceived as an auditory Gestalt in a sonification. Furthermore, both in visualization and in sonification, a gradual transition takes place from the sum of many 0D marks to a single 1D mark. In visualization, the best example is a dotted line: Even if every dot could have individual meaning, the

Table 1 Substrates, mark types, and channels

Domain	Substrate	Mark types	Possible channels
Visualization	Space	0D: Point	position, size, color hue,...
		1D: Line	
		2D: Area	
		3D: Volume	
Sonification	Time	0D: State in time	pitch, loudness, timbre,...
		1D: Development over time	

Gestalt of the dots suggests a line phenomenon. The same applies to sonification and auditory perception. A violinist, to give an example from the field of music, can play a melody with the note transitions tightly tied together ("legatissimo"), or play each of them short and strictly separated ("staccato"). In both cases, a listener will recognize the tone sequence as one unit, as one Gestalt. In visualization, the different marks are perceived as individual entities, as objects with visual features. This is also reflected by the way we generally perceive our visual surroundings as humans. Bregman used the example of a green dog: We would not separately perceive a dog and the attribute “greenness”, i.e., the attribute belongs to the object [64]. He also states that “the stream plays the same role in auditory mental experience as the object does in visual” [64, p. 11]. Basically, an auditory stream is perceived to be originating from one sound source. To design effective sonifications, it is therefore necessary to be well informed about the effects that influence our perception of auditory streams.

Both in visualization and in sonification, we can define channels that encode information into the marks and can distinguish between identity channels and magnitude channels. Last but not least, just as visualization needs to deal with spatial clutter, sonification needs to deal with temporal masking.

We now want to discuss existing visualizations, sonifications, and combinations using the model of substrates, marks, and channels. These specific cases have been chosen because they give an overview of the design space that can be described and analyzed with our unified terminology.

Examples from the visualization domain

1. Example 1: **Node-link network diagrams** with force-directed placement [65] combine 0D point marks for network nodes with 1D line marks for their connections. An algorithm places the point marks by simulating physical forces that move connected nodes towards and unconnected nodes away from each other. In contrast to a scatter plot (Fig. 2), the position of point marks in the spatial substrate does not directly encode data attributes. Yet, the resulting placement is often effective in indicating network clusters by their proximity of marks in the spatial substrate, although cluttered areas can also be due to artifacts [7, p. 204]. Additional data attributes can be encoded with the color, size, and shape channel of point marks, as well as the color, width, and dashing of line marks.
2. Example 2: **Parallel coordinates** [66, 72] represent multivariate data as 1D line marks. On the spatial substrate, one vertical axis for each attribute is placed in parallel across the available horizontal space. The line marks, actually polygonal paths, connect the positions encoded by attribute values between adjacent axes. In addition, color hue can be used as an identity channel. The resulting plot can provide overview of multiple attributes and indicate correlation between adjacently placed attributes.
3. Example 3: The **treemap** [68] represents hierarchical data using nested rectangular area marks (2D). An algorithm iteratively divides the available spatial substrate into rectangles while mapping the size of each rectangle to an attribute summed up from the contained items. Treemaps can be applied for stock market data with stocks hierarchically grouped by sector. The marks use the size channel for market capitalization and the color channel for the relative change in stock price [69, 70].

Examples from the sonification domain

1. Example 1: A conventional **auditory graph** [48, 71] translates the visual representation of a linechart to an auditory representation by using a *one-dimensional auditory* mark in the temporal substrate. The auditory channel *pitch* conveys information about the data while the auditory mark evolves over time. This example shows a direct translation of a one-dimensional visual mark into a one-dimensional auditory mark by translating horizontal and vertical spatial position into temporal position and pitch.
2. Example 2: Baier et al. [72] used 0D auditory marks on the temporal substrate to encode **information about EEG signals**. To do so, they used several different auditory channels such as timbre, pitch, and duration, mapping signal parameters such as the duration between peaks in the EEG signals to auditory channels. The sonifi-

cation can be listened to via their supplementary material [73].

3. Example 3: Bywater and Middelton **sonified amino acid sequences** "as a string of musical notes with sound qualities that reflect the properties of these residues" [74, p. 18]. They used 0D auditory marks ("musical notes") in the form of marimba sounds and placed them equally distributed on the temporal substrate. Pitch was used as an auditory channel ("sound qualities") to convey information about amino acid values in the studied sequences. The authors state that they would use other channels like timbre, dynamics, and articulation in future investigations.

Examples from combined designs

1. Example 1: Enge et al. [75] presented **SoniScope**, a tool that combines a visual scatterplot with interactive parameter mapping sonification. The visualization uses 0D point marks in the spatial substrate, using the channel *position* to communicate two of the data attributes. The sonification displays a third and non-visible attribute with 0D auditory marks (short marimba sounds) positioned in the temporal substrate, using the auditory channel of pitch.
2. Example 2: **Listen To Wikipedia** [76] is a website built by Stephen LaPorte and Mahmoud Hashemi enabling users to monitor changes to Wikipedia in real-time through both visualization and sonification. Whenever someone edits Wikipedia, the tool displays a 0D visual mark somewhere on the spatial substrate using the visual channels of size and color. The size encodes the size of the edit, and the color encodes whether the edit was done by an automated bot (purple), an unregistered (green), or a registered user (white). The channel *timbre* of the sounds (identifying either a bell or a string instrument) is used to communicate added (bell sounds) or removed (plucked string sounds) content on Wikipedia. The channel *pitch* again encodes the size of the edit, representing larger edits with lower pitch.
3. Example 3: Rönnberg and Johansson [77] combined a **parallel coordinates visualization with a parameter mapping sonification** to investigate the potential of sonification for the exploration of dense and visually cluttered areas. The visualization used one-dimensional line marks on the spatial substrate, encoding information via the visual channels color and position. The sonification used one-dimensional auditory marks in their temporal substrate, representing the densities of two data clusters via the auditory channel of volume of two synth sounds. The two synths represented two data clusters via the identity channel of pitch.

6 Reflections on space and frequency as potential substrates for sonification

While our model uses time as the substrate for sonification, we want to discuss two other parameters especially relevant to sonification: space and frequency. Both of them come to mind when we search for a concept that can be described as "the container" of sonification. We now want to reflect on our decision to not model space and/or frequency as substrates for sonification.

6.1 Why space is not the substrate of sonification

The ability to spread over both time and space is an essential attribute of sound. In regard to the concept of spatial substrates in visualization it may seem self-evident to assign space equally as a substrate in the sonification domain. Spatial substrates in visualization are characterized by their dimensionality. In most cases, the spatial substrate is two-dimensional, like a piece of paper or a computer screen. Three-dimensional substrates can be used in virtual reality applications or conceptually via a projection to a conventional screen. Such two- or three-dimensional spatial substrates can contain zero- to three-dimensional visual marks. In the field of audio reproduction, we commonly speak of mono, stereo, surround, and 3D reproduction of signals, thus providing the dimensionality that is required as a precondition to qualify as an equivalent to the concept of a spatial substrate in visualization.

Following this rationale, a spatially 0D auditory mark corresponds to a point mark in visualization and could be rendered using a single loudspeaker at a specific location. A spatially one-dimensional auditory mark would correspond to a line mark in visualization. Such a mark would convey different auditory information from the different spatial positions on the stereo panorama. Technically, this could be displayed with a stereo speaker setup or with a line of speakers positioned next to each other. 2D and 3D auditory marks would then be defined accordingly and could be rendered with respective surround or 3D audio systems (such as Ambisonics [78]).

What at first sight seems to be a perfectly matching analogy reveals major drawbacks at closer analysis. Spatial substrates in visualization provide clearly determined and delimited environments. Marks can be uniquely perceived and identified within these substrates. The perception of sound, however, relies heavily on psychoacoustic phenomena as they have been described by Blauert [16], Fastl and Zwicker [79], and Bregman [64]. For instance, for the stereo projection of a sound source, we utilize so-called phantom sources composite of sonic contributions of a left-hand ($-30°$) and a right-hand ($30°$) loudspeaker in relation to a listener in order for them to be perceived at specific positions between the two speakers. Even a slight turn of the listener's head could alter the localization of the sound and change its perceived timbre. Besides the impact the coherence of sonic signals has on their localizability, overlaying sounds are also often indistinguishable for listeners, perceptually amalgamating to one compound sound. Psychoacoustic effects such as the precedence effect also contribute to the unreliability of auditory spatial perception.

Furthermore, according to Kubovy and Van Valkenburg, space is not central for the formation of auditory objects as it is not relevant from *where* a sound approaches us, but *what* sounds. In their 'Theory of Indispensable Attributes,' they state that it is not the direction that helps us identify an auditory object, but its temporal and spectral properties [20, 80].

Considering these ambiguities, we argue that auditory space does not qualify as a spatial substrate in analogy to its visual counterpart.

6.2 Why frequency is not the substrate for sonification

Kubovy's and Van Valkenburg's work on indispensable attributes [20, 80] inspires one to think about pitch or frequency as potential substrates for sonification. Kubovy et al. plausibly argue for time and frequency as two indispensable attributes of auditory objects [20, 80]. In their original paper [20], the authors mistakenly talk about "pitch" but corrected the wording later to "frequency" [81]. They essentially state that "a *perceptual object* is that which is susceptible to figure-ground segregation" [20, p. 102] and that "an attribute (or dimension) is defined as indispensable if and only if it is a prerequisite of perceptual numerosity" [20, p. 108]. In a much earlier publication [82], Kubovy argued for pitch as a medium and a potential equivalent of space in audition. He refers to Attneave and Olson [83] with the example of a pitch-shifted melody keeping its perceptual identity.

We argue, on the other hand, that to be considered as a substrate of visualization or sonification it is relevant that a dimension enables translation-invariant placing of marks. Hence, a mark that is placed at different positions of its substrate should appear identical. It is not enough for an auditory mark to "keep its perceptual identity" like a pitch-shifted melody would, it should appear identical.

It is a quality of space that a visual mark does not change its individual appearance if it has another position on the spatial substrate. A red point in a scatter plot looks the same whether it is in the lower left corner or the upper right corner of the substrate. It conveys different information but its individual appearance is not altered by a shift in position. In search of an analog concept in sonification, we are looking for a substrate that offers the same quality to auditory marks. While time offers this quality (a sound that is only played back later will have the same individual appearance), frequency or pitch do not. A change in frequency or pitch changes the individual appearance of any sound. We want to discuss this phenomenon with two brief examples: a musical melody and everyday sounds. There are two possibilities for shifting a sound in the spectral dimension: pitch shifting or frequency shifting. A melody indeed can be transposed and still be "the same" melody, but only if the transposition happens with respect to the pitch of the individual notes. If one would change the frequencies of all the notes in a melody by a constant value, the melody would change and could not be recognized.

We humans have learned to recognize environmental sounds by listening to them over and over. That is essentially what the sonification technique of auditory icons uses to convey information to us. If one of those auditory icons would be shifted to a totally different frequency range, we would lose our environmental connection to that sound and most probably would not recognize it anymore. In such cases, even the perceptual identity of a sound would be lost.

Space and time are two dimensions that have no physical borders to our perception, while frequencies below 20 Hz and above 20kHz cannot be perceived by humans. It should be able to place an auditory mark anywhere in its substrate without losing the ability to perceive it as humans.

Due to these arguments, in our model, we do not think of frequency or pitch as adequate pendants for the spatial substrate.

7 Conclusion and future work

This paper provided an overview of three fundamental theoretical constructs from visualization theory and adopted them for the field of sonification. One is the spatial *substrate*; hence, the space a visualization uses to place visual entities on. These visual entities are called *marks*; they are positioned in the spatial substrate and have visual *channels* such as size or color encoded into them. Our work shows that time qualifies as the substrate of sonification; we, therefore, call it temporal substrate. Just as visual marks have positions in space, auditory marks have positions in time. Auditory marks use auditory channels to encode information about their identity or their magnitude. We also investigated the possibility to use space or frequency as potential substrates for sonification but rejected the models due to several drawbacks. With time as the substrate of sonification, we discussed emerging analogies between sonification and visualization theory and showed how our model can be used to describe existing designs.

The possibility to use consistent theoretical constructs for the description of audio-visual data analysis techniques fosters mutual understanding and can help the visualization and sonification communities with the further development of a combined design theory. The identified constructs proved to be useful for the authors of this article in the development of two audio-visual analytics approaches: one for scatterplots [75] and one for parallel coordinates [84]. We found the common language helpful to efficiently discuss ideas while minimizing misunderstandings between the visualization and sonification experts in our team. Furthermore, our work introduces new terminology to systematically describe sonification designs and could also feed back into visualization theory concerning the temporal description of data visualizations. One strategy to evaluate the practical usability of the identified theoretical constructs would be to conduct a systematic review of cases from the literature, similar to the recent work by Caiola et al. [21].

In our future research, we will continue with the design, implementation, and evaluation of combined designs of sonification and visualization, using the theoretical underpinnings of the presented unified terminology. We will investigate how different visual and auditory channels can be combined in corresponding or complementary ways to help users explore their data. One specific next step is to tackle the known challenges of parallel coordinates, i.e., visual clutter, outlier detection, and comparability of non-adjacent axis [85, 86] with sonification. Furthermore, we will use our concept and framework of SoniScope [75] to test different combinations of visualization and sonification. Thus, we will proceed to testable propositions as another component of a design theory according to Gregor and Jones [40].

While a fundamental discussion of the possibilities for combined audio-visual designs and suggestions for novel mappings is out of the scope of this article, we want to emphasize the need for future research regarding these questions. To design expressive audio-visual displays, it will be necessary for our community to study and consider cross-modal effects on the human perception of data representations as well as Gestalt- and auditory streaming phenomena. We expect our unified terminology to support the description and communication of future guidelines in such a way that both communities can contribute to the development of an audio-visual design theory.

Acknowledgements This research was funded in whole, or in part, by the Austrian Science Fund (FWF): P33531-N and the Gesellschaft für Forschungsförderung Niederösterreich (GFF NÖ) SC20-006. For the purpose of open access, the author has applied a CC BY public copyright license to any Author Accepted Manuscript version arising from this submission.

Funding Open access funding provided by FH St. Pölten - University of Applied Sciences.

Declarations

Conflict of interest The authors declare no competing interests.

References

1. Enge K, Rind A, Iber M, Höldrich R, Aigner W (2021) It's about time: adopting theoretical constructs from visualization for sonification. In: Audio Mostly 2021, pp 64–71. ACM New York. https://doi.org/10.1145/3478384.3478415
2. Hermann T, Hunt A, Neuhoff JG (2011) (ed) The Sonification Handbook. Logos, Bielefeld
3. Hermann T (2008) Taxonomy and definitions for sonification and auditory display. In: Proceedings of the 14th international conference on auditory display. http://hdl.handle.net/1853/49960
4. Nees MA (2019) Eight components of a design theory of sonification. In: Proceedings of the 25th international conference on auditory display (ICAD 2019), pp 176–183. https://doi.org/10.21785/icad2019.048
5. Bertin J (1983) Semiology of Graphics Diagrams Networks Maps. University of Wisconsin, Madison. Originally published in 1967 in French
6. Card SK, Mackinlay J, Shneiderman B (1999) Readings in Information Visualization: using Vision to Think. Morgan Kaufmann, San Francisco
7. Munzner T (2014) Visualization Analysis and Design. CRC Press, Boca Raton
8. Stolte C, Tang D, Hanrahan P (2002) Polaris A system for query, analysis, and visualization of multidimensional relational databases. IEEE Trans. visualization and computer graphics 8(1):52–65. https://doi.org/10.1109/2945.981851
9. Wilkinson L (2005) The Grammar of Graphics, 2nd edn. Springer, New York
10. Satyanarayan A, Moritz D, Wongsuphasawat K, Heer J (2017) Vega-Lite: a grammar of interactive graphics. IEEE Trans. visualization and computer graphics 23(1):341–350. https://doi.org/10.1109/TVCG.2016.2599030
11. Mackinlay J (1986) Automating the design of graphical presentations of relational information. ACM Trans. Graphics 5(2):110–141. https://doi.org/10.1145/22949.22950
12. Mackinlay JD, Hanrahan P, Stolte C (2007) Show Me: automatic presentation for visual analysis. IEEE Trans. visualization and computer graphics 13(6):1137–1144. https://doi.org/10.1109/TVCG.2007.70594
13. Moritz D, Wang C, Nelson G, Lin H, Smith AM, Howe B, Heer J (2018) Formalizing visualization design knowledge as constraints Actionable and extensible models in Draco. IEEE Trans. visualization and computer graphics 25(1):438–448. https://doi.org/10.1109/TVCG.2018.2865240

14. Kramer G, Walker B, Bonebright T, Cook P, Flowers JH, Miner N, Neuhoff J et al (1999) Sonification report Status of the field and research agenda. Report for the NSF, international community for auditory display. http://www.icad.org/websiteV2.0/References/nsf.html Accessed 31 March 2022
15. Walker BN, Kramer G (2004) Ecological psychoacoustics and auditory displays Hearing, grouping, and meaning making. In: Psychoacoustics Ecological (ed) Neuhoff JG. Elsevier Academic Press, San Diego, pp 150–175
16. Blauert J (1996) Spatial Hearing The Psychophysics of Human Sound Localization, Revised. MIT Press, Cambridge, MA, USA
17. Kramer G (1994) Some Organizing Principles for Representing Data with Sound. In: Kramer G (ed) Auditory Display Sonification, Audification and Auditory Interfaces. Addison-Wesley, Reading, Mass, pp 185–221
18. Freides D (1974) Human information processing and sensory modality: Cross-modal functions, information complexity, memory, and deficit. Psychological bulletin 81(5):284–310
19. Sieroka N (2018) Philosophie der Zeit Grundlagen und Perspektiven, vol 2886. CH Beck, München
20. Kubovy M, Van Valkenburg D (2001) Auditory and visual objects. Cognition 80(1–2):97–126. https://doi.org/10.1016/S0010-0277(00)00155-4
21. Caiola V, Lenzi S, Riccò, D (2022) Audiovisual sonifications: a design map for multisensory integration in data representation. In: Lockton D, Sara L, Hekkert P, Oak A, Sádaba J, Lloyd P. (eds.) DRS2022 Bilbao. https://doi.org/10.21606/drs.2022.380
22. Hildebrandt T, Amerbauer F, Rinderle-Ma S (2016) Combining sonification and visualization for the analysis of process execution data. In: 2016 IEEE 18th Conference on business informatics (CBI), vol. 2:pp 32–37
23. Rabenhorst DA, Farrell EJ, Jameson DH, Linton TD, Mandelman JA (1990) Complementary visualization and sonification of multidimensional data. In: Extracting meaning from complex data processing, display, interaction, vol. 1259:pp 147–153. International Society for Optics and Photonics
24. Berger J, Wang G, Chang M (2010) Sonification and visualization of neural data. In: Proceedings of the 16th international conference on auditory display (ICAD-2010), pp 201–205. Georgia Institute of Technology
25. Hermann T, Niehus C, Ritter H (2003) Interactive visualization and sonification for monitoring complex processes. In: Proceedings of the 2003 international conference on auditory display, p 247. http://hdl.handle.net/1853/50467
26. MacVeigh R, Jacobson RD (2007) Increasing the dimensionality of a geographic information system (gis) using auditory display. In: Proceedings of the 13th international conference on auditory display, pp 530–535. http://hdl.handle.net/1853/50005
27. Nesbitt KV (2000) A classification of multi-sensory metaphors for understanding abstract data in a virtual environment. In: Proc. IEEE Conf. information visualization (IV), pp 493–498. https://doi.org/10.1109/IV.2000.859802
28. Nesbitt KV, Barrass S (2002) Evaluation of a multimodal sonification and visualisation of depth of market stock data. In: Proceedings of the 8th international conference on auditory display, Kyoto. http://hdl.handle.net/1853/51355
29. Nesbitt KV, Barrass S (2004) Finding trading patterns in stock market data. IEEE Computer graphics and applications 24(5):45–55. https://doi.org/10.1109/MCG.2004.28
30. Nesbitt KV (2004) MS-Taxonomy a conceptual framework for designing multi-sensory displays. In: Proc. eighth international conference on information visualisation. IV 2004, pp 665–670. https://doi.org/10.1109/IV.2004.1320213
31. Nesbitt KV (2006) Modelling human perception to leverage the reuse of concepts across the multi-sensory design space. In: Proceedings of the 3rd asia-pacific conference on conceptual modelling Vol. 53 APCCM ’06, pp 65–74. Australian Computer Society, Darlinghurst, Australia
32. Frysinger SP, (2005) A brief history of auditory data representation to the 1980s. In: Proceedings of the international conference on auditory display. http://hdl.handle.net/1853/50089
33. Kramer G (ed) (1994) Auditory Display Sonification. Audification and Auditory Interfaces, Addison-Wesley, Reading, Mass
34. Barrass S (1997) Auditory Information Design. PhD thesis, Australian National University, Canberra. https://doi.org/10.25911/5d7a2acd2b71d
35. Walker BN (2002) Magnitude estimation of conceptual data dimensions fruse insonification. Journal of experimental psychology applied 8(4):211–221. https://doi.org/10.1037/1076-898X.8.4.211
36. Hermann T (2002) Sonification for exploratory data analysis. PhD thesis, Bielefeld University, Bielefeld, Germany
37. Neuhoff JG (ed) (2004) Ecological Psychoacoustics. Elsevier Academic Press, San Diego
38. de Campo A (2007) Toward a data sonification design space map. In: Proceedings of the 13th international conference on auditory display, pp. 342–347. http://hdl.handle.net/1853/50042
39. Worrall D (2019) Sonification Design From Data to Intelligible Soundfields. Springer, Cham, Human-computer interaction series. https://doi.org/10.1007/978-3-030-01497-1
40. Gregor S, Jones D (2007) The anatomy of a design theory. Journal of the association for information systems 8(5):312–335. https://doi.org/10.17705/1jais.00129
41. Card SK, Mackinlay J (1997) The structure of the information visualization design space. In: Proc. IEEE Symp. information visualization, infovis, pp. 92–99. https://doi.org/10.1109/INFVIS.1997.636792
42. Stoiber C, Grassinger F, Pohl M, Stitz H, Streit M, Aigner W (2019) Visualization onboarding: Learning how to read and use visualizations. In: IEEE Workshop on visualization for communication. OSF Preprints, Vancouver, BC, Canada. https://doi.org/10/gh38zd
43. Wickham H (2010) A layered grammar of graphics. Journal of computational and graphical statistics 19(1):3–28. https://doi.org/10.1198/jcgs.2009.07098
44. Mauri M, Elli T, Caviglia G, Uboldi G, Azzi M (2017) RAWGraphs: A visualisation platform to create open outputs. In: Proceedings of the 12th biannual conference on italian sigchi chapter. CHItaly ’17. ACM, New York, NY, USA. https://doi.org/10.1145/3125571.3125585
45. Cleveland WS, McGill R (1984) Graphical perception: Theory, experimentation, and application to the development of graphical methods. J American statistical association 79(387):531–554
46. Ware C (2021) Information Visualization Perception for Design, 4th edn. Morgan Kaufmann, Cambridge
47. Grond F, Berger J (2011) Parameter mapping sonification. In: Hermann T, Hunt A, Neuhoff JG, (eds.) The Sonification Handbook, pp 363–397
48. Mansur DL, Blattner MM, Joy KI (1985) Sound graphs: a numerical data analysis method for the blind. Journal of medical systems 9(3):163–174. https://doi.org/10.1007/BF00996201
49. Stockman T, Nickerson LV, Hind G (2005) Auditory graphs: a summary of current experience and towards a research agenda. In: Brazil E, (ed.) Proceedings of the 11th international conference on auditory display, pp 420–422. http://hdl.handle.net/1853/50097
50. McGookin D, Brewster S (2011) Earcons. In: Hermann T, Hunt A, Neuhoff JG, (eds.) The Sonification Handbook, pp 339–361
51. Rohrhuber J (2010) $\mathring{S}$ - introducing sonification variables. In: Proceedings of the supercollider symposium. http://wertlos.org/~rohrhuber/articles/Rohrhuber_Sonvar.pdf Accessed 31 March 2022

52. Vogt K, Höldrich R (2012) Translating sonifications. Journal of the audio engineering society 60(11):926–935
53. Dubus G, Bresin R (2013) A systematic review of mapping strategies for the sonification of physical quantities. PLoS ONE 8(12):82491
54. Ferguson J, Brewster SA (2017) Evaluation of psychoacoustic sound parameters for sonification. In: Proceedings of the 19th ACM international conference on multimodal interaction. ICMI 2017, pp 120–127. ACM, Glasgow UK. https://doi.org/10.1145/3136755.3136783
55. Neuhoff JG, Knight R, Wayand J (2002) Pitch change, sonification, and musical expertise: Which way is up? In: Proceedings of the 2002 international conference on auditory display. http://hdl.handle.net/1853/51370
56. Walker BN, Lane DM (2001) Psychophysical scaling of sonification mappings A comparison of visually impaired and sighted listeners. In: Proceedings of the 2001 international conference on auditory display, pp 90–94. http://hdl.handle.net/1853/50630
57. Shannon CE (1948) A Mathematical Theory of Communication. The bell system technical journal 27(3):379–423. https://doi.org/10.1002/j.1538-7305.1948.tb01338.x
58. Brazil E, Fernström M (2011) Auditory Icons. In: Hermann T, Hunt A, Neuhoff J (eds) The Sonification Handbook. Logos, Berlin, pp 325–338
59. Gaver WW (1989) The SonicFinder: an interface that uses auditory icons. Human-computer interaction 4(1):67–94. https://doi.org/10.1207/s15327051hci0401_3
60. Brewster SA (1994) Providing a structured method for integrating non-speech audio into human-computer interfaces. PhD thesis, University of York
61. Pratt RL, Doak PE (1976) A subjective rating scale for timbre. Journal of sound and vibration 45(3):317–328. https://doi.org/10.1016/0022-460X(76)90391-6
62. Elmquist E, Ejdbo M, Bock A, Rönnberg N (2021) Openspace sonification: Complementing visualization of the solar system with sound. In: Proceedings of the 26th international conference on auditory display (ICAD 2021), pp 135–142. https://doi.org/10.21785/icad2021.018
63. Wertheimer M (1923) Untersuchungen zur Lehre von der Gestalt. II. Psychologische forschung 4(1):301–350
64. Bregman AS (1990) Auditory Scene Analysis: The Perceptual Organization of Sound. MIT press, Cambridge, MA, USA
65. Fruchterman TMJ, Reingold EM (1991) Graph drawing by force-directed placement. Software practice and experience 21(11):1129–1164. https://doi.org/10.1002/spe.4380211102
66. Inselberg A, Dimsdale B (1990) Parallel coordinates: a tool for visualizing multi-dimensional geometry. In: Proc. 1st conference on visualization '90, pp 361–378. https://doi.org/10.1109/VISUAL.1990.146402
67. Wegman EJ (1990) Hyperdimensional data analysis using parallel coordinates. Journal of the american statistical association 85(411):664–675. https://doi.org/10.1080/01621459.1990.10474926
68. Shneiderman B (1992) Tree visualization with Tree-Maps: 2-D space-filling approach. ACM Trans. graph. 11(1):92–99. https://doi.org/10.1145/102377.115768
69. Wattenberg M (1999) Visualizing the stock market. In: CHI '99 Extended abstracts on human factors in computing systems, pp 188–189. https://doi.org/10.1145/632716.632834
70. FinViz S&P 500 Map. https://finviz.com/map.ashx Accessed 31 March 2022
71. Grond F, Hermann T (2012) Singing function: exploring auditory graphs with a vowel based sonification. Journal on multimodal user interfaces 5(3–4):87–95. https://doi.org/10.1007/s12193-011-0068-2
72. Baier G, Hermann T, Stephani U (2007) Multi-channel sonification of human EEG. In: Proceedings of the 13th international conference on auditory display, Montreal, Canada, pp 491–496. http://hdl.handle.net/1853/50013
73. Baier G, Hermann T, Stephani U (2007) Supplementary Material for "Multi-Channel Sonification of Human EEG". https://pub.uni-bielefeld.de/record/2699375 Accessed 31 March 2022
74. Bywater RP, Middleton JN (2016) Melody discrimination and protein fold classification. Heliyon 2(10):00175. https://doi.org/10.1016/j.heliyon.2016.e00175
75. Enge K, Rind A, Iber M, Höldrich R, Aigner W (2022) Towards multimodal exploratory data analysis: SoniScope as a prototypical implementation. In: Agus M, Aigner W, Hoellt T, (eds.) Proceedings of Eurovis 2022 pp 67–71. Eurographics, Goslar, Germany. https://doi.org/10.2312/evs.20221095
76. LaPorte S, Hashemi M (2022) Listen to Wikipedia. http://listen.hatnote.com/#en Accessed 31 March 2022
77. Rönnberg N, Jimmy J (2016) Interactive sonification for visual dense data displays. In: ISon 2016, 5th Interactive sonification workshop, CITEC, Bielefeld University, Germany, December 16, 2016, pp 63–67
78. Zotter F, Frank M (2019) Ambisonics: a Practical 3D Audio Theory for Recording, Studio Production, Sound Reinforcement, and Virtual Reality. Springer Nature, Graz. https://doi.org/10.1007/978-3-030-17207-7
79. Zwicker E, Fastl H (1999) Psychoacoustics: facts and Models. Springer series in information sciences, vol. 22. Springer, Berlin, Heidelberg
80. Kubovy M (1981) Concurrent-pitch segregation and the theory of indispensable attributes. Perceptual organization. Routledge, London, pp 55–98
81. Van Valkenburg D, Kubovy M (2003) In defense of the theory of indispensable attributes. Cognition 87(3):225–233. https://doi.org/10.1016/S0010-0277(03)00005-2
82. Kubovy M (1988) Should we resist the seductiveness of the space time vision audition analogy? Journal of experimental psychology human perception and performance 14(2):318–320. https://doi.org/10.1037/0096-1523.14.2.318
83. Attneave F, Olson RK (1971) Pitch as a Medium: a New Approach to Psychophysical Scaling. The american journal of psychology 84(2):147–166. https://doi.org/10.2307/1421351
84. Elmquist E, Enge K (2022) Towards the combination of visualization and sonification for cylindrical displays. In: AVI 2022 workshop on audio-visual analytics (WAVA22), Frascati, Rome, Italy. https://doi.org/10.5281/zenodo.6553825
85. Heinrich J, Weiskopf D (2013) State of the art of parallel coordinates. Eurographics (State of the Art Reports), 95–116
86. Johansson J, Forsell C (2015) Evaluation of parallel coordinates: overview, categorization and guidelines for future research. IEEE Transactions on visualization and computer graphics 22(1):579–588

Original Article 2

Open Your Ears and Take a Look: A State-of-the-Art Report on the Integration of Sonification and Visualization

The following article [EEC*24] was published in the Computer Graphics Forum[2] after being presented at the *EuroVis 2024* conference in Odense, Denmark.

[2]onlinelibrary.wiley.com/journal/14678659

EUROVIS 2024
C. Garth, A. Kerren, and R. Raidou
(Guest Editors)

COMPUTER GRAPHICS *forum*
Volume 43 (2024), Number 3
STAR – State of The Art Report

Open Your Ears and Take a Look: A State-of-the-Art Report on the Integration of Sonification and Visualization

K. Enge[1,2], E. Elmquist[3], V. Caiola[4], N. Rönnberg[3], A. Rind[1], M. Iber[1], S. Lenzi[5,6], F. Lan[7], R. Höldrich[2], and W. Aigner[1]

[1]Institute of Creative\Media/Technologies (IC\M/T), St. Pölten University of Applied Sciences, St. Pölten, Austria
[2]Institute of Electronic Music and Acoustics (IEM), University of Music and Performing Arts Graz, Graz, Austria
[3]Division for Media and Information Technology (MIT), Linköping University, Linköping, Sweden
[4]School of Creative Media, City University of Hong Kong, Kowloon, Hong Kong
[5]Ikerbasque Basque Foundation for Science, Bilbao, Spain
[6]Faculty of Engineering, Universidad de Deusto, Bilbao, Spain
[7]Scientific Computing and Imaging Institute (SCI), University of Utah, Salt Lake City, USA

Abstract

The research communities studying visualization and sonification for data display and analysis share exceptionally similar goals, essentially making data of any kind interpretable to humans. One community does so by using visual representations of data, and the other community employs auditory (non-speech) representations of data. While the two communities have a lot in common, they developed mostly in parallel over the course of the last few decades. With this STAR, we discuss a collection of work that bridges the borders of the two communities, hence a collection of work that aims to integrate the two techniques into one form of audiovisual display, which we argue to be "more than the sum of the two." We introduce and motivate a classification system applicable to such audiovisual displays and categorize a corpus of 57 academic publications that appeared between 2011 and 2023 in categories such as reading level, dataset type, or evaluation system, to mention a few. The corpus also enables a meta-analysis of the field, including regularly occurring design patterns such as type of visualization and sonification techniques, or the use of visual and auditory channels, showing an overall diverse field with different designs. An analysis of a co-author network of the field shows individual teams without many interconnections. The body of work covered in this STAR also relates to three adjacent topics: audiovisual monitoring, accessibility, and audiovisual data art. These three topics are discussed individually in addition to the systematically conducted part of this research. The findings of this report may be used by researchers from both fields to understand the potentials and challenges of such integrated designs while hopefully inspiring them to collaborate with experts from the respective other field.

1. Introduction

Over the course of the last few decades, two research communities have developed largely in parallel: one studying data visualization and one studying sonification. While the visualization community is primarily interested in "the use of computer-supported, interactive, visual representations of abstract data to amplify cognition" [CMS99], the sonification community studies "the use of non-speech audio to convey information" [KWB*99]. To this day, the communities seem to be largely disjoint, despite their shared goals promising fruitful collaboration. Both the theoretical cross-pollination [ERI*23, CLR22] and the practical integration and combination of sonification and visualization offer potential for interesting research outcomes. Therefore, with this state-of-the-art report (STAR), we want to shed light on audiovisual display idioms that systematically integrate data visualization and sonification. Informed by the definition of the visual idiom by Munzner [Mun15], we think of an *audiovisual display idiom* as "a distinct approach to creating and manipulating audiovisual representations of data."

In our daily lives, we perceive our surroundings in inherently multimodal and transmodal ways. We see, and we hear, we read books, and we listen to music. We use our senses to understand and explore the world around us. Although we are multisensorial beings, the predominant data analysis idioms are unimodal, often using visualization only. Sight and hearing are inherently different, with different strengths and challenges, and are most probably suitable for different approaches in data representation. Human visual perception can look upon a visual representation in a non-linear fashion, and the use of different types of graphs, charts, and other visual formats can reveal patterns, correlations, and trends in data that are often not as noticeable in numerical form. Visual representations of data can also be experienced as more engaging and memorable to a user compared to tables with numbers [FPS*21]. Sonification, on the other hand, exploits the excellent ability of the human auditory system to recognize temporal changes and patterns. It is, therefore, useful when displaying complex patterns, such as changes in time and warnings for immediate action. In real-time environments, sonification allows a controller to perceive information without the constant monitoring of visual displays. Sonification also enables the communication of data and information for visually impaired individuals [WM10].

The human visual perception has some challenges that can be supported by sonification, and similarly, the auditory system has other challenges that, in turn, can be supported by visualization. Therefore, we believe that a well-designed audiovisual representation can be more than the mere sum of a visual and an auditory representation. To put this into practice, we refer to a typical real-world situation where the combination of visual and auditory inputs aids us in reaching a more informed conclusion: imagine rain falling outside of a window. It is often difficult to correctly estimate the density of rain by just looking outside a closed window. It is also not easy to estimate the amount of rain when only listening to it with your eyes closed. It is the holistic audiovisual perspective, perceptually integrating both of our senses, that allows us to best determine whether we should use an umbrella or even stay indoors.

Inspired by the capabilities of the human visual and auditory systems and the possibility of integrating visualization and sonification, this STAR covers academic contributions from both the visualization and the sonification communities that blend sonification and visualization within the context of data exploration and data presentation. We hope this STAR will help both visualization and sonification researchers realize the potential of such combinations and foster future collaborations between two often disjoint communities.

1.1. Sonification Background

This STAR being published at a visualization venue calls for a brief introduction to the sonification techniques that are part of our data (for a more comprehensive description of the different techniques, see the Sonification Handbook [HHN11]). Overall, we can distinguish the five main techniques of *audification*, *parameter mapping sonification*, *model-based sonification*, *auditory icons*, and *earcons*:

Audification is a technique to represent typically long sequences of data values (often time-series) by interpreting them as digital audio waveforms and directly playing them back over a loudspeaker. The resulting sound is a translation of the data values into the audible domain in terms of frequency and loudness. An example of audification is the playback of seismographic data with increased playback speed, such that the original low-frequency signal gets pitch-shifted into a range that is audible to humans. A design challenge for audification is the influence of the chosen playback speed on the salience of emerging auditory patterns that should inform a listener about their data.

Parameter mapping sonification (PMS) is a technique that involves the association of data values with auditory parameters such as pitch, timbre, and loudness. The technique is conceptually closest to many standard visualization techniques as it employs the direct mapping of data values to auditory channels of a carrying sound. If a visualization utilizes a visual mark to represent information by the mark's position or its color, then a parameter mapping sonification utilizes a sound to carry information via its pitch, loudness, or other auditory channels. Similar to a visualization, such a sound is dependent on the mapping function between the data and the auditory channel as well as the nature of the mapping, for example, being linear to linear or linear to exponential. An open challenge in sonification research is the task-dependent and appropriate selection of auditory channels or the perceptual influence of one channel on another. In terms of visualization, this is comparable to the influence of the spatial size of a mark on our perception of its color.

The technique of *model-based sonification* (MBS) is inspired by the fact that most of the interactions with our physical environment result in an acoustic response. These acoustic responses, such as the sound of a drum being hit, inform us about the state of the object that we interact with (in the case of a drum, its tuning). Model-based sonification is a general term for sonification techniques that make use of dynamic models describing changes in a system over time. These dynamic models are "tuned" by the data that an analyst wants to explore. To listen to their data, a user is required to excite the model with an interaction, such as when a drummer needs to hit their instrument to hear its tuning. Another example is the excitation of a mass-spring model, where the mass and spring parameters are determined by high-dimensional data, defining the sound of the model when excited by a user. The technique is meant to foster exploratory data analysis, as the type and place of interaction strongly influence the acoustic response of the model. In this context, an ongoing research challenge is how to establish an intuitive relation between excitation modes (where and how hard to hit the drum) and typical interactions during exploratory data analysis, such as zooming or filtering of the data.

Also inspired by real-world sounds, *auditory icons* are short, distinctive sounds present in everyday life that can be compared to visual icons. This means that there is an inherent association between the auditory icon and the event they represent. A classic example of an auditory icon is the sound of a piece of paper being crumpled and thrown into a bin, representing deleting a file on a computer. A research challenge is the possibility of cultural differences between listeners, as they can lead to confusion or misinterpretation of auditory icons [JLSP15], similar to the way visual icons can be context-dependent.

Earcons, comparable to visual symbols, are short, distinctive sounds or melodies that are often used to represent specific events. These sounds are usually synthesized tones or sound patterns and can be described as designed or composed sound symbols. Since there is no inherent association with the real world, the meaning of the earcon needs to be learned before it can be beneficially used. Examples of such earcons are the sounds provided to pilots in the cockpit of an airplane, alerting them about events that require attention.

While other techniques for the sonification of data exist, these five are the most prominent ones. Within the scope of our STAR, we identified all five of these techniques, even though the vast majority of papers utilize parameter mapping sonification. Some of the papers also combine techniques (for example, parameter mapping sonification and auditory icons, see Table 1).

A brief history of sonification: While the discipline of visualization has a relatively long history [Fri08], the research field of sonification is younger [Fry05]. In 1982, Bly submitted her PhD thesis Sound and computer information presentation [Bly82], where she suggested methods of encoding information into sound. Ten years later, the first International Conference on Auditory Display convened, which is regarded as the birth of the International Community For Auditory Display (ICAD). When its proceedings were published in 1994, the book Auditory Display [Kra94] was a reflection on the potential of the newborn field of research that encompasses sonification. Early on, Barrass [Bar97] presented a task taxonomy for auditory displays, called *TaDa!*, which stands for Tasks and Data. The *TaDa!* taxonomy is especially relevant in the context of this STAR, as it is well-aligned with taxonomies from the visualization literature [Ber83, BM13, SNHS13, Shn96, YKSJ07], and it also functions as inspiration for the classification applied later in this STAR.

The book Ecological Psychoacoustics, edited by Neuhoff in 2004 [Neu04], challenged many psychoacoustical studies (which is the part of psychophysics that involves the scientific study of sound perception, traditionally conducted in controlled laboratory environments). Neuhoff promoted an ecological sound approach to sonification from a holistic perspective, which echoes the aims of the BELIV workshop series established in 2006 at the Advanced Visual Interfaces conference [BPS06]. Neuhoff's intervention underscores the need to consider real-world contexts for transforming design principles and methodologies for auditory displays from theory to practice. This perspective emphasized integrating ecological factors in sonification to increase effectiveness and deepen the connection between auditory stimuli and real-world experiences.

The *sonification design space map* was introduced in 2007 by deCampo [dC07], guiding a designer's decision-making process of selecting an appropriate sonification technique for their task. The map creates a two-dimensional space between the number of data properties a designer intends to sonify and the number of data points that are necessary for the different sonification techniques to be employed adequately. Retrospectively, another milestone within the sonification community was the introduction of a now widely accepted definition of sonification as a scientific technique for representing data, presented by Hermann [Her08] in 2008. Before the introduction of this definition, it was less clear where to draw the border between artistic and scientific mappings from data to sound (which brings to mind the discussion that data visualization is more than just pretty pictures). As it reflects our understanding of the term sonification, we want to refer to the full definition below:

"A technique that uses data as input, and generates sound signals (eventually in response to optional additional excitation or triggering) may be called sonification, if and only if

- *The sound reflects objective properties or relations in the input data.*
- *The transformation is systematic. This means that there is a precise definition provided of how the data (and optional interactions) cause the sound to change.*
- *The sonification is reproducible: given the same data and identical interactions (or triggers) the resulting sound has to be structurally identical.*
- *The system can intentionally be used with different data, and also be used in repetition with the same data."*

The Sonification Handbook [HHN11], published in 2011, provided the first general and overarching perspective on the field of sonification, discussing both the theory and practice of sonification. The Handbook is still the most comprehensive collection of sonification work, and therefore, its publication year in 2011 also marks the beginning of the time period considered in our STAR.

Over the years, various design frameworks for sonification have been proposed. The design framework proposed by Barrass [Bar12] emphasizes the fusion of aesthetics and functionality to improve the accessibility and meaningfulness of sonifications for a broader audience. The work of Worrall in 2019 [Wor19b] formalized sonification techniques into a framework that also highlighted the challenges and advantages of these sonification techniques, as well as the importance of understanding processes and choices that influence sound representation. The *sonification design canvas*, introduced by Lenzi in 2021 [Len21], is a contribution to the construction of a more comprehensive design framework, with the aim of integrating all aspects into a cohesive design tool. Despite these efforts, developing a comprehensive protocol that systematically considers end-users at each stage of the design process has yet to be achieved.

More than 30 years after the beginning of systematic sonification research, we saw a considerable number of theoretical contributions to the field [Kra94, KWB*99, VH06, dC07, NW08, HHN11, GH12, Sup12, Nee19, Neu19, Wor19b, Len21]. Explicit work integrating sonification and visualization theory is rare, but has been called for [RW10]. In an attempt to find a common language and, consequently, build a theoretical bridge between the visualization and the sonification communities, Enge et al. [ERI*23] introduced three theoretical constructs to formally describe audiovisual display idioms. They defined the "auditory mark" inspired by the visual mark, the "auditory channel" inspired by the visual channel, and the "substrate of sonification" inspired by the spatial substrate that is used in visualization theory [Ber83, CMS99]. The definition of *time* as the substrate of sonification allows a description of sonification designs with auditory marks placed in time, with data encoded into their auditory appearance using auditory channels such as pitch or loudness. These definitions allow a high-level discussion and categorization of both the visual and the auditory part of an audiovisual display idiom. The theoretical constructs proved useful, so we adopted the term "auditory channels" for classification in this STAR as well.

1.2. Motivation and related work

Seminal visualization texts such as Wilkinson's Grammar of Graphics [Wil05] and Spence's textbook [Spe07] made clear statements that their understanding of visualization respectively graphics is not limited to vision but that data can be encoded for other sensory modalities such as sound. Already at the 6th IEEE Visualization Conference, Minghim and Forrest [MF95] postulated areas where sonification can help tackle visualization challenges such as adding complementary or redundant dimensions, natural mapping for time-oriented data, or improved memory of data. Some research agendas

[TC05] suggest multimodal approaches so that one sensory modality can overcome problems that others may have. In 1990, Grinstein and colleagues [GS90, SBG90] presented an audiovisual interface for the exploration of multivariate data using icons and parameter mapping sonification. Also, several works presented at early visualization conferences have integrated sound into visualizations for surfaces, volumes, and fluid dynamics [MF95, LWS96, RN96, VG97, FBZ*99]. The *Data Mountain* interface [RCL*98] augmented its spatial document management space with auditory cues that indicated how many pages were moving. Published as early as 1989, Gaver presented the *Sonic Finder,* which was the auditory user interface used in Apple computers and coined the term "auditory icon."

The auditory perception has an exceptional ability to detect temporal changes and patterns [GGB05, RG71]. Also, human hearing is capable of perceiving and distinguishing between several sounds simultaneously, at least between three auditory streams at the same time [SW13]. Another capability of our auditory perception is the possibility to detect and focus on events that spatially occur all around the listener, which can enable an information display to convey peripheral information to a user. Furthermore, the auditory system enables quick reactions when performing certain types of tasks due to the different processing times of the senses [JBKS15].

The combination of audio and visual data representations has shown to be advantageous, particularly for a number of applications. First, situations where the visual modality is busy with another task, such as monitoring [Ibe20, NB02], lend themselves beneficial for being complemented with sonification. Second, the combination of visual and auditory techniques has been shown to better facilitate learning [May14, SKS06], since it has the potential to increase working memory capacity and retention of information while also reducing cognitive load. Third, using sonification also shows benefits for data exploration. Flowers et al. [FBT97] published a seminal article in the sonification community that demonstrated that auditory scatter plots, where data is mapped to onset time and pitch of sounds instead of horizontal and vertical position of visual marks, can offer similar performance as visual representations.

Research has shown that visual and auditory perception is naturally integrated with each other, which can be observed, for example, with so-called crossmodal correspondences [Spe11, SDS23]. In our context, crossmodal correspondences describe a phenomenon where we perceive different visual and auditory stimuli as inherently related to each other. Such correspondences could provide an opportunity to use the strengths of the auditory modality to support and enhance visualizations, creating more effective and compelling representations of data [KBBG07]. Rosli and Cabrera [RC15], for example, have identified the potential of integrated designs to form a more concise, general representation of the data set compared to individual stimuli alone. Rubab et al. recently explored relationships between auditory channels and visual channels and suggested factors that impact their effectiveness [RYTW23].

Overall, the research described above suggests that the integration of the auditory modality has the potential to remedy challenges that exist for visual perception. Challenges, such as simultaneous brightness contrast [War19] or the Mach band phenomenon [LWP99], impact perception of visual representations [SGS*18, ZTSS23]. It has been demonstrated that various auditory channels can be successfully linked and related to visual channels [FB18, WHT06, CH04], which could offer a way of substituting visual channels with auditory channels to address these challenges.

1.3. Related Surveys

In general, systematic state-of-the-art reports are less established in the sonification community. A rare exception is the "*systematic review of mapping strategies for the sonification of physical quantities*" by Dubus and Bresin [DB13], which, however, just covers sonification-only contributions. Much earlier, in 2001, Walker and Lane [WL01] provided a website enabling researchers to search for sonification mappings that have been used in scientifically evaluated designs. Unfortunately, the website is no longer available. Another more timely and exhaustive exploration of sonification literature emerges in Andreopoulou and Goudarzi's 2021 publication [AG21]. The authors reviewed 456 papers from the International Conference on Auditory Display proceedings. This incisive analysis exposes compelling trends, ranging from the sonification domains to the diverse publication venues. The report reveals linguistic trends and explores the balance between research and artistic contributions. In addition, it illuminates the landscape of tools, methodologies, and evaluation practices that have led to sonification's multifaceted evolution. Marking the beginning of an important sociocultural reflection within the sonification community, in 2017, Andreopoulou and Goudarzi [AG17] also studied the "*representation of female researchers and artists in the conferences of the International Community for Auditory Display (ICAD)*". Their findings showed that only about 18% of ICAD papers were co-authored by women, with stagnant numbers between the years 1994 and 2016.

With respect to combinations of visualization and sonification, Caiola et al. [CLR22] recently presented an analysis of visual and auditory channels commonly used in audiovisual display idioms. Their survey includes combined idioms that map data attributes redundantly to both a visual channel (such as position) and an auditory channel (such as pitch). Analyzed works stem from the Sonification Archive (described below) and a Google keyword search using sonification-related terms exclusively. The *Sonification Archive*, widely known in the sonification community, is a curated collection of sonification designs, often related to other modes of representation, such as visualization. The Sonification Archive holds both artistic and academic contributions, as well as designs from data journalism.

Searching through the visualization literature, we were not able to find any STAR or survey focused on the integration of sonification and visualization. We explored the survey of surveys [ML17] but could not identify any related contributions. Therefore, to the best of our knowledge, this is the first systematic STAR dedicated to academic contributions in the intersection of sonification and visualization for data exploration and presentation.

1.4. How to Use This Survey

With this STAR, we intend to provide an overview of an emerging research field, as well as connect two mostly disjoint research communities. We hope to reach researchers from both communities, inspiring them to intertwine sonification and visualization in their future research. We see several ways of using this STAR:

- using it as an overview, intended for researchers who seek a summary of the field.
- using it to find research opportunities and existing gaps in the field.
- using our supplemental material to study the existing meta-data in more detail, such as identifying authors from the respective other field for potential collaboration. Furthermore, we provide a public *Zotero* library, holding all relevant publication metadata, our tags, and all open access PDFs.

This STAR will be structured as follows: section 2 describes the methodology used to search and filter the literature identified as potentially relevant. In section 3, we describe our classification system and use it to discuss the survey literature. In section 4, we apply a meta-perspective on the survey data, describing correlations between individual tags, as well as the co-author network of the field. In section 5 we introduce the three adjacent topics of accessibility, monitoring, and arts, which are related to our STAR, but were not studied systematically. Finally, in section 6, we offer a concluding discussion focusing on future work.

2. Method

In this section, we discuss our inclusion and exclusion criteria and the methods we used to search for the relevant literature. We used a five-stage pipeline to construct a corpus of research that is at the intersection of visualization and sonification for data exploration and presentation (see Figure 1).

2.1. Scope of the Surveyed Literature

Sonification and visualization share the aim of making data interpretable to their users and observers. With this shared goal, combinations of the two can be designed for numerous possible applications and contexts. Our research interest in this STAR is the combination of sonification and visualization in the context of data analysis, covering both data exploration and presentation. A work that is relevant to our STAR must include both visualization and sonification of data. Therefore, a sonification with a visual interface that does not represent data is not enough to be considered relevant, and neither is a visualization with sounds that do not represent data. The work must be an academic paper published between the years 2011 and 2023 and must be peer-reviewed to be considered in our STAR.

Thinking more broadly about the combination of sonification and visualization, three additional areas of application come to mind: (1) accessibility, (2) real-time monitoring, and (3) arts. All three areas are vast, and a detailed classification of them is beyond the scope of our STAR. However, we find them relevant and inspirational for our field. Thus, we provide a brief introduction to the fields of accessibility, monitoring, and artistic contributions in section 5. In the context of accessibility, sonification can be used to support the collaboration between blind and sighted users by mapping data to both an auditory and a visual display. In the same manner, such a design could support the collaboration between deaf individuals and individuals without hearing impairment. Nevertheless, our research interest is the combination of sonification and visualization for the integrated analysis of data. Therefore, in our STAR, we consider only designs intended to be used with both the visual and the auditory senses fully available to a user. The application of real-time monitoring, such as medical monitoring, critical infrastructure monitoring, alarms, or real-time feedback on body movement, is vast and distinct from the purpose of data exploration. Especially with respect to sonification and auditory display, the field is well-researched (e.g., [KIK19, SJMT19, VRGM20, WMY*17, HRM15]), and we will not cover such designs in this STAR. Artistic contributions have the potential to be highly inspirational for our field but require a different search method and, most likely, a different system of classification. Again, we decided not to systematically cover artistic contributions in this STAR but to provide a subsection discussing a list of representative works that may serve as a starting point for future research interests.

2.2. Search Strategy and Filtering

We base our corpus of literature on (1) publications that the authors have already been aware of from their previous work in this field and (2) an extensive online literature search. The online search was a keyword-based search in the digital libraries of IEEE Xplore, ACM Digital Library, and Springer Link, which include work published at IEEE VIS, CHI, and other VGTC- and SIGCHI-sponsored venues. Furthermore, we searched the digital libraries of Eurographics, ICAD, ISon, Organised Sound, and the Sound and Music Computing Community. Figure 1 provides an overview of the different stages we used to systematically filter our database for relevant publications. To keep track of the progress, we used a Google sheet document. The final database, including all papers and all tags by the authors, is provided in the supplemental material as a CSV file.

Stage 1 – The search query we used in stage 1 was the following: *("Visualization" OR "Visualisation" OR "Visual Analytics") AND ("Sonification" OR "Auditory Display") AND NOT (centrifug* OR lys* OR homogeniz*).* It consists of an AND combination of a visualization term and a sonification term combined with the exclusion of specific word beginnings. While the terms “visualization” and “visual analytics” are well known, our audio-related search terms are likely to be less well-known overall. Nevertheless, we decided not to include search terms such as music, sound, or tone in our online search. Searching for publications with such broad terminology would have resulted in too many papers to be scanned. Also, we argue that academic publications interested in the integration of visualization and sonification are likely using the appropriate terminology. The reason to exclude papers that hold words starting with *centrifug**, *lys**, and *homogeniz** is the fact that sonification, also called sonication, is a term also used in biology to describe a process where sound is used to agitate particles in a sample. As papers using sonification in this context are not relevant to this STAR, we identified terms, including the three mentioned above, that are often associated with this meaning of sonification. The search in the Springer Link database was especially sensitive to the exclusion of these terms. As there could be false negatives using the exclusion of the three terms, we manually reviewed the paper titles that were excluded due to this strategy and restored five potentially relevant papers. This online search, combined with papers we were already aware of from our previous work in the field, resulted in a database holding 1498 papers. We used the literature management software Zotero to download the respective papers and to make them available to all co-authors.

Stage 2 – In the second phase of our literature search, each paper title was read by two of the authors and classified into *potentially relevant* or *irrelevant*. We agreed to use an inclusive mindset for this stage, i.e., we tagged vague titles mostly as potentially relevant so as not to overlook many papers at this early stage. For papers that were tagged differently by two people, the two people came to an agreement, or the paper was taken to the next stage. Stage 2 resulted in a database holding 500 papers.

Stage 3 – In the third phase of our literature search, each abstract was read by two of the authors and classified into *potentially relevant* or *irrelevant*. Again, we used an inclusive mindset and tagged vague abstracts as potentially relevant. For papers that were tagged differently by two people, the two people came to an agreement, or the paper was taken to the next stage. Stage 3 resulted in a database holding 163 papers.

Stage 4 – In the fourth phase of our literature search, each paper was read by one of the authors and classified as *relevant* or *irrelevant*. As the authors had a solid common understanding of the inclusion and exclusion criteria by this stage, and as a person had the full paper information available to make a decision, one person decided on the relevance at this stage. The papers identified as relevant in stage 4 were classified using the tags explained in detail in Table 2. Each paper was first classified by one of the co-authors, and their tags were later verified by a second co-author. Whenever two co-authors initially disagreed on a specific tag, they came to an agreement, or the first author made a decision. Stage 4 resulted in a database holding 47 papers.

Stage 5 – Furthermore, we extended the corpus of relevant papers by snowballing [Woh14], checking all incoming and outgoing references of the articles matching our inclusion criteria. Snowballing was done by one of the co-authors with an exclusive mindset towards the paper titles, meaning that a vague title was not considered relevant. Stage 5 resulted in our database holding 57 papers overall, adding ten papers to the prior stage. During the final two stages, whenever we identified an audiovisual idiom that was published in more than one paper, we retained the most extensive version in our STAR. We identified two such cases where a design was previously published in a short paper but later expanded [YH18, MAFP19b]. It is notable that a considerable number of papers that are part of our final scope were published at the International Conference on Auditory Display (ICAD). This is not surprising as this venue is the most recognized venue to publish sonification work holding the largest single corpus of work in the field [BB12, AG21, GVEZ23]. It is also not surprising that it is a sonification venue that is prominent, as there is a structural imbalance between the domains. Many sonification designs, in general, also include some sort of visualization in their design, but that is not true the other way around.

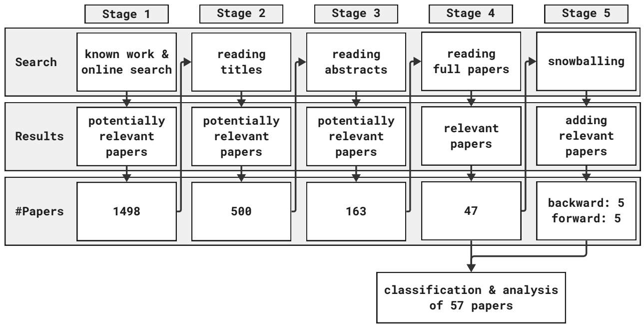


Figure 1: *The literature search was conducted in five stages, ranging from an online keyword search to using the snowballing technique on the papers identified as relevant. Figure inspired by [YM22].*

3. Categorization and Results

In this section, we will discuss the relevant literature in detail from the perspectives of our classification. The systematic integration of sonification and visualization is a wide and diverse research field that is difficult to classify using only a handful of categories. Therefore, we decided to apply an extensive list of tags to the literature to be able to present diverse perspectives on the field, mostly concerning basic research, i.e., the basic principles that distinguish the designs/idioms from each other.

To give readers an initial thematic overview of the field, we will start by briefly describing each of the 57 selected papers in subsection 3.1. We will do so by clustering the literature to the following topics: astronomy, medicine and health, molecular science, earth science, domain agnostic data displays, and other topics. We will then continue with a discussion on the purpose that an idiom can be designed for (subsection 3.2), followed by an analysis of idiom design possibilities (subsection 3.3). We will study several technical perspectives that were tagged individually for the sonification and the visualization aspects of each paper: the reading levels (subsection 3.4) suggested by Bertin [Ber83], the search levels (subsection 3.5) suggested by Munzner [Mun15], as well as the dataset types [Mun15] including the levels of measurement of the displayed data (subsection 3.6). We will then review different levels of mapping redundancy (subsection 3.7), different evaluation methods (subsection 3.8), different target platforms, and various possibilities of interacting with audiovisual display idioms (subsection 3.9). Finally, we study the diverse user groups and the possible goals of designers (subsection 3.10).

All the above categories and subcategories are concisely presented in two tables. Table 1 shows an overview of all papers and a subset of their most relevant tags. Table 2 provides a detailed description for each category and their respective subcategories. We also report on the total number of papers within each subcategory. The four technical categories mentioned above have distinct visualization and sonification tags, which are represented in the left and right boxes in the "Num." column. The luminance of the boxes encodes the total number of papers in each subcategory.

3.1. Thematic Corpus Overview

Before we employ the classification system described in Table 2, we want to provide a thematic corpus overview. This overview is intended for readers who are interested in a special field of application, such as astronomy or earth sciences. While we do not intend to

Paper	Theme	Demo Link	Purpose			Redundancy			Visualization Idiom	Sonification Technique					Goal			
			Exploration	Presentation	Both	Redundant	Mixed	Complementary		PMS	MBS	Audification	Auditory Icons	Earcons	Education	Data Analysis	Research	Public Engagement
[EEBR21]	Astronomy	[<—>]			•		•		volume rendering	•								•
[Rib19]	Astronomy	[<—>]		•		•			line chart, scatter plot	•					•			
[HTHB22]	Astronomy	[<—>]		•		•			point cloud	•								•
[HPDW23]	Astronomy	[<—>]			•			•	scatter plot	•		•			•			
[Rib18]	Astronomy	[<—>]			•	•			volume rendering	•						•		
[RS22]	Astronomy	[<—>]		•			•		3D map	•								•
[TB23]	Astronomy			•				•	3D scatter plot	•								•
[BTB23]	Domain Agnostic Display		•			•			line chart, parallel coordinates	•						•		
[CWM21]	Domain Agnostic Display	[<—>]			•	•			line chart	•					•	•		•
[DLVDCG22]	Domain Agnostic Display	[<—>]			•	•			line chart	•						•		
[DCM*18]	Domain Agnostic Display				•	•			convex hull, other, bar chart	•						•		•
[ERI*22]	Domain Agnostic Display	[<—>]	•			•	•	•	scatter plot	•						•		
[FBC12]	Domain Agnostic Display				•	•			line chart	•						•		
[GKW21]	Domain Agnostic Display	[<—>]	•					•	network	•					•			•
[KLTW17]	Domain Agnostic Display	[<—>]		•			•		line chart	•								•
[LF21]	Domain Agnostic Display	[<—>]			•	•			line chart	•					•	•		•
[MAFP19b]	Domain Agnostic Display	[<—>]*	•			•			heatmap	•						•		
[PCB23]	Domain Agnostic Display	[<—>]		•		•			line chart	•					•			
[PC19]	Domain Agnostic Display			•		•			line chart	•						•		
[RJ16]	Domain Agnostic Display	[<—>]	•			•			scatter plot, parallel coordinates	•						•		
[YH18]	Domain Agnostic Display	[<—>]	•					•	scatter plot		•					•		
[FN18]	Domain Agnostic Display		•				•		line chart	•						•		
[Bal15]	Earth Science			•			•		heatmap	•								•
[Bea11]	Earth Science	[<—>]	•			•			map	•						•		
[BF12]	Earth Science	[<—>]	•			•			map	•						•		
[GDAS*18]	Earth Science				•	•			map	•							•	
[HK22]	Earth Science			•		•			geographic scatter plot	•								•
[HCTP14]	Earth Science			•		•			heatmap			•			•			
[MMU16]	Earth Science		•					•	map	•		•				•		
[NRL*12]	Earth Science	[<—>]	•					•	map	•					•			
[PFH*22]	Earth Science	[<—>]			•		•		dot map			•					•	
[SAR22]	Earth Science	[<—>]		•		•			fluid-like simulation	•								•
[WW15]	Earth Science	[<—>]		•		•			line chart			•						•
[GRK*16]	Medicine and Health		•					•	slicing, volume rendering	•						•		
[GR11]	Medicine and Health		•			•			slicing	•						•		
[LSB*23]	Medicine and Health	[<—>]	•				•		map	•						•		
[MNW*18]	Medicine and Health	[<—>]	•					•	volume rendering	•						•		
[RFM13]	Medicine and Health		•			•			slicing	•						•		
[TMN*21]	Medicine and Health	[<—>]	•				•		volume rendering	•						•		
[AJB*18]	Molecular Science	[<—>]	•					•	3D molecule rendering	•						•		
[BBV16]	Molecular Science		•					•	3D molecule rendering	•			•			•		
[BM20]	Molecular Science		•			•			volume rendering	•						•		
[RFK*15]	Molecular Science		•					•	volume rendering	•			•			•		
[LLW21]	Others	[<—>]	•					•	3D network	•								•
[NSC16]	Others	[<—>]	•				•		gantt chart	•				•		•		
[CB17]	Others			•			•		bar chart	•						•		
[MMM18]	Others	[<—>]		•			•		individual circular design	•								•
[ASH*12]	Others	[<—>]	•			•			volume rendering	•						•		
[Her20]	Others	[<—>]	•			•			network	•							•	
[PBM15]	Others		•					•	3D network	•			•			•		
[PBV14]	Others		•					•	network	•						•		
[HAR16]	Others		•				•		dotted chart visualization	•				•		•		
[JMP13]	Others				•			•	3D scatter plot	•						•		
[KAV21]	Others	[<—>]		•		•			3D point cloud	•								•
[Rön21]	Others	[<—>]		•				•	bar chart	•								•
[AK11]	Others			•			•		map	•						•		
[BB19]	Others		•			•			point grid	•						•		

Table 1: *A table showing all 57 entries in our database, sorted by their thematic cluster. In the columns for the used sonification technique, "PMS" stands for parameter mapping sonification, and "MBS" stands for model-based sonification. Where available, the demo links point to the supplemental material of the papers (last accessed on* 22^{nd} *of December 2023). *There are two additional demo videos associated with this paper: [<—>] and [<—>]; all three demos require the password: icad2019.*

provide detailed descriptions of the individual articles at this point, we will present more detailed descriptions in the later subsections 3.2 to 3.10.

Scanning our database, we identified seven **astronomy** related papers. Riber [Rib18] presented a prototypical virtual and interactive audio synthesizer called *Planethesizer* that enables its users to design sonifications, especially focused on planetary data. *Sonifigrapher* [Rib19] is a virtual synthesizer that lets users sonify the light curves data from NASA's exoplanet archive. Also, the recently presented *Sonified Hertzsprung-Russel Diagram* [HPDW23] sonifies the light curves, with the diagram acting as both the visualization and the interface to choose a star to be sonified. With this design, hearing a constant pitch will inform a user about the rotation of a star. The rotation, temperature, and other parameters of planets in our solar system were also sonified by Elmquist et al. [EEBR21] in *OpenSpace Sonification*. Their design can be used both with conventional computer desktop environments and in planetarium settings, and they are tailored towards public outreach and science communication. Public outreach is also the core of the publication *Audio Universe* by Harrison et al. [HTHB22]. The publication describes the design of a 35-minute audiovisual show about the solar system integrating visualization and sonification, as well as an audiovisual animation displaying the stars in the same order they appear to our eyes during dusk. Similarly, Russo and Santaguida [RS22] collaborated with NASA, celebrating the discovery of the 5000th exoplanet. Their design displays the exoplanets as they were discovered over the years. Recently, Traver presented another audiovisual installation where users can control the auditory representation of the planets using a Midi controller [TB23].

We identified six **medicine and health** related topics in our database, out of which three are related to brain scans [GR11, RFM13, GRK*16], two are audiovisually displaying blow flow and aneurysm models [MNW*18, TMN*21], and one is concerned with Covid-19 data [LSB*23].

Our database holds four idioms that we relate to **molecular science**. In an idiom presented by Rau et al. [RFK*15], scientific visualization of a molecular simulation is enhanced using parameter mapping sonification and auditory icons. Among other things, their design guides the attention of a user towards visually occluded phenomena using sonification. Ballweg et al. [BBV16] use sonification with the intention of supporting chemists and structural biologists with drug design. For their sonification plug-in, they focused on tasks that were not well supported visually in a software for the interactive visualization of molecular structures called "UCSF Chimera." In the context of biomolecules simulation, Arbon et al. [AJB*18] developed a sonification displaying characteristics of the "free energy landscape," a map used to study the properties of biomolecular systems. Their technique allows a user to visually inspect the physical configuration of a biomolecule while listening to their corresponding free energy landscape. Exploring the possibilities of 3D sound, Bouchara and Mones [BM20] suggested a work-in-progress immersive sonification model to study protein surfaces.

The **earth science** cluster in our survey data holds eleven publications covering topics that range from oceanographic data [SAR22], wildfires [HK22], hurricanes [Bal15], and climate change [Bea11], to sonophenology [NRL*12], seismology [HCTP14, WW15, MMU16, PFH*22], and geospatial data displays [BF12, GDAS*18].

The category of **domain agnostic data display** idioms in our survey data holds 15 papers. These idioms are not designed to support users from a specific domain but are implementations tackling problems across multiple domains. Six of the papers describe software frameworks that are intended to help people design sonifications along with visual representations of their data [PC19, PCB23, LF21, DLVDCG22, CWM21, KLTW17]. Their unifying core goal is the democratization of sonification as a technique to represent data, hence making it accessible to more people, both professionals as well as domain experts. Other publications focus on basic research combining different sonification techniques such as parameter mapping sonification or model-based sonification with basic information visualizations such as scatterplots [ERI*22, RJ16, YH17], parallel coordinates plots [BTB23, RJ16], or line charts [FBC12, FN18] to study the potentials of audiovisual display idioms. Groppe et al. [GKW21] studied network visualization and sonification through their design, while Malikova et al. [MAFP19b] show the potential of sonification to help users identify smallest symmetry differences in scalar fields visualizations. While most studies focus on metrics such as precision, error rates, or task completion times, the study by Du et al. [DCM*18] explicitly investigates the sonification's influence on user engagement.

Finally, we want to provide a brief overview of the remaining **other** 14 publications not part of the above thematic clusters. The topics related to the publications tagged as "other" are diverse, ranging from a multimodal system for analyzing business process execution data [HAR16] or Git version control data [NSC16], to an audiovisual representation of the Portuguese consumption patterns [MMM18], to a multimodal implementation of "Game of Life" [KAV21], and a musical sonification aimed at conveying information about running data and emotion [Rön21]. Several publications focus on audiovisual data representation in virtual or extended realities [CB17, PBV14, PBM15, BB19, JMP13] and two publications focus on explainable AI [LLW21, Her20]. Alonso et al. [ASH*12] presented an interface for product design that communicates a virtual object's geometrical shape using visual, haptic, and auditory stimuli, and Adhitya and Kuuskankare [AK11] proposed a prototype that offers a sonification-based approach to urban design planning.

As we have now presented a thematic overview of all 57 papers that are part of our database, we can focus on the meta-level classification in the following subsections. With these more high-level descriptors, we intend to provide a number of versatile perspectives on the literature. They will help us identify research gaps and opportunities for the systematic integration of sonification and visualization for the future work of both research communities in section 6. Along the discussion of these perspectives, we will present selected papers that are representative examples for the respective category. By doing so, we intend to provide the reader with (1) a broad exploration of the field overall and (2) insights into the content of the actual papers themselves.

Category	Num.	Subcategory	Explanation
Purpose	29	Exploration	designs used for data exploration
	17	Presentation	designs used for data presentation
	11	Both	designs used for both purposes above
Visualization Design		Idiom	the visualization idiom, such as scatter plot or line chart
		Identity channels	the name of the visual channel, such as color hue, or shape, that is used to communicate the identity of an item (i.e., "What" something is)
		Magnitude channels	the name of the visual channel such as position or length that is used to communicate the magnitude of an attribute (i.e., "How much" something is)
Sonification Design		Technique	the sonification technique such as parameter mapping sonification or earcons
		Identity channels	the auditory channel such as timbre that is used to communicate the identity of an item (i.e., "What" something is)
		Magnitude channel	the auditory channel such as pitch or duration that is used to communicate the magnitude of an attribute (i.e., "How much" something is)
Reading Level	50 29	Whole	designs that display *all* datapoints
	13 29	Group	designs that display a *group* of datapoints
	4 17	Single	designs that display *single* datapoints
Search Level	2 2	Lookup	the user knows the location and the target of the search (see Figure 4)
	9 6	Browse	the user knows the location but not the target of the search
	7 7	Locate	the user doesn't know the location but the target of the search
	40 42	Explore	the user doesn't know the location or the target of the search
	11 8	None	none of the above
Dataset Type	29 40	Table	data constructed from items and attributes (spreadsheets; see Figure 5)
	5 3	Network	data constructed from items (nodes), links and potentially attributes
	9 7	Field	data constructed from grids (positions) and attributes
	20 9	Geometry	data constructed from items and positions
Level of Measurement	22 13	Nominal	data that builds categories (such as different fruits)
	8 8	Ordinal	data that builds ordered categories (such as t-shirt sizes)
	27 31	Interval	data that has equal intervals, such as the time on the clock
	27 34	Ratio	data that has equal intervals and a meaningful zero point, such as length or weight
Level of Redundancy	17	Redundant	a design mapping all displayed information to both senses redundantly (see Figure 6)
	14	Complementary	a design mapping part of the information exclusively to the visualization and another part of the information exclusively to the sonification
	28	Mixed	a design mapping some information redundantly, some information complementary
Evaluation System	12	User Performance	evaluations collecting metrics such as error rates or task completion times
	11	User Experience	evaluations collecting user feedback, typically done in a usability test
	2	Algorithmic Performance	evaluations doing measurements without users such as rendering speed
	17	Qualitative Result Inspection	evaluations providing subjective arguments on the quality of the result
	20	None	no evaluation
Target Platform	43	Desktop Computer Display	conventional screen on desktop computer
	4	XR	extended reality settings such as virtual or augmented reality glasses
	10	Physical Environments	environments that foster collaboration such as a CAVE
	5	Touch Display	interactive screen that users can interact with via touch
User Interaction	37	Yes	designs that require user interaction other than pressing "play"
	20	No	designs that require no user interaction
Users	25	Domain Experts	domain experts that are not visualization or sonification researchers
	7	Researchers	visualization or sonification researchers
	31	General Public	the interested general public
Goals	8	Education	idioms designed for education
	35	Data Analysis	idioms designed for data analysis (incl. data exploration)
	3	Research	idioms designed for visualization and sonification research
	17	Public Engagement	idioms designed for public engagement
Demo	29	Yes	the paper links to a demo such as an interactive website, a video, or an audio recording
	23	No	the paper doesn't link to a demo
	5	Yes, but not online	the paper provides a link that is not online anymore

Table 2: *Descriptions and distributions of all used classification tags. For the categories with two separate tags under the "Num." column, the left and right boxes represent the visualization and sonification distributions, respectively.*

3.2. Purpose

In this section, we will discuss two different purposes an audiovisual idiom can be designed for, inspired by the taxonomy described in [Mun15]: exploration and presentation. The *purpose* should be read as the general design goal of a tool with a broad perspective. Therefore, the term exploration also covers what is widely referred to as data analysis.

3.2.1. Exploration

A wide range of audiovisual idioms in our database support users with the exploration of data (see Table 2). We refer to the purpose of exploration in cases where a user intends to acquire new knowledge from their data by using an audiovisual display idiom.

Exploratory data analysis is the core aim of model-based sonification [Her02]. The *Mode Explorer* is an interactive audiovisual display idiom, combining a scatterplot visualization and model-based sonification to explore high-dimensional data [YH18]. To explore the data, a user can "scratch" a scatter plot of dimensionality-reduced data with an interactive pen. The scratching introduces a virtual particle to the high-dimensional space, which will follow a gradient descent to the nearest mode in the data. A user will be able to hear the "kinetic energy" of this virtual particle. While the particle travels closer and closer to its final mode, the sound gradually turns more and more harmonic, finally resulting in a clear pitch when a mode is reached.

AIive combines visualization and sonification in a virtual reality environment [LLW21]. The idiom helps users understand the basic concepts of neural networks. Users can manipulate the weights of the nodes of a virtually displayed 3D neural network by dragging the nodes in 3D space. In real-time, the sonification displays the loss and accuracy of the neural network, therefore enabling a user to explore different constellations of node weights. Users are also able to add or delete nodes, hence experimenting with the complexity of the neural network. While the authors do not provide a user study in their paper, it is most plausible the exploratory character of the design can support people in understanding the basics of neural networks.

In a recent publication, Lemmon et al. presented an audiovisual map idiom that seeks to tackle some of the sociotechnical challenges associated with epidemiological mapping [LSB*23]. Using their idiom, users can interactively explore Suffolk County's experience with Covid-19. The black population and associated case numbers are displayed on the left audio channel, while the white population and associated case numbers are displayed on the right channel via the pitch of sine and triangle waves. A correlation between black and white population numbers and COVID-19 case numbers and their dependency on different regions becomes clearly apparent when listening to the sonification while brushing the map using the computer mouse.

3.2.2. Presentation

We refer to the purpose of presentation when a user intends to present prior knowledge to others or in cases where a user is presented with information that is new to them but not to the designer.

One potential application of a presentation-only approach is in citizen science, exemplified by the communication of significant discoveries, such as NASA's announcement of the 5000th exoplanet [RS22]. This representation is explicitly crafted for communication with the general public, primarily through social media channels. Translating data into various sensory modalities is straightforward, while significant attention is devoted to achieving a pleasing and harmonious aesthetic. The sonification, complemented by two animated videos, enhances visual comprehension, vividly portraying celestial spheres with dynamic elements representing exoplanet characteristics. NASA's strategic dissemination on social media platforms garnered substantial engagement, with positive feedback indicating the presentation's broad appeal.

The presentation objective appears linked to eliciting emotional engagement from end users, as evident in Rönnberg's research [Rön21]. This representation involves assessing running statistics, weather data, and associated emotions. Presented as an animated visualization synchronized with sonification, it depicts weekly runs emphasizing evoking emotional responses rather than precise data interpretation—a practice denoted by the author as "musification" (see further discussion about sonification and music in Vickers, 2016 [Vic16]), leveraging sound's dual capacity to illustrate data and engender a musical experience. The study assesses the representation's efficacy through a user study, wherein participants watched a video. Results indicate that the sonification effectively conveys intended emotions but at the expense of a less accurate data representation.

Recent works delve into environmental concerns and the promotion of awareness. One such noteworthy contribution is presented in [HK22]. This publication centers around a multimodal museum installation designed to foster public engagement with wildfire forecasts in specifically chosen California and South Korea regions. The project incorporates interactive data visualization, sonification, and 3D-printed sculptures. Combining these elements conveys wildfire data, creating a comprehensive and immersive experience. The installation allows audiences to explore the representation freely through contactless interaction. Without user engagement, the wildfire data representation seamlessly loops across the screen. Showcased at the ARKO Art Center in Seoul, this project successfully captivated the attention of 20 users. The outcomes revealed a heightened level of interest and comprehension regarding the impact of wildfires through the effective utilization of multimodal interaction. The project emphasizes observations and prompts audience reflection, leveraging new media to enhance public climate change comprehension.

3.2.3. Exploration and Presentation

Some of the designs we identified were designed both to explore and to present data. Inclusion in this category requires the designer to put similar weight on the exploration and presentation of the data. To present a few examples, Paté et al. [PFH*22] demonstrate their audiovisual display idiom with three seismic data sets, where the sonification methods are adapted to the specific properties of each data set. A multi-scale audification method is presented where different speed factors are used depending on the size of each dataset, and different sound designs are used to highlight specific properties of the data. Huppenkothen et al. [HPDW23] created an audiovisual version of the Hertzsprung-Russell Diagram where the user can listen and compare auditory representations of each type of star that is included in the diagram (see Figure 7(c)). It is also possible to filter the diagram according to certain criteria, which, in turn, filters which sonifications can be listened to. Elmquist et al. [EEBR21] created complementing sonifications of the planets in the solar system, which were integrated with a visualization, where a user can explore the properties of the planets and make comparisons between them. The user can listen to all of the sonifications for each planet at the same time or enable specific sonifications for each planet to compare specific properties. Du et al. [DCM*18] conducted a study investigating the enhanced visualization of basketball player movement data during a game. The visualization is designed to convey the offensive and defensive dynamics of a team. The interface primarily enables users to specify a particular time range of interest, providing a more comprehensive view of information within that timeframe. Notably, sonification is integrated exclusively during the exploration phase.

3.3. Audiovisual Idiom Design

Designers of audiovisual idioms have countless possibilities to combine their visualizations and sonifications. They can choose from a vast number of established visualization idioms and sonification techniques. Having chosen two designs to integrate with each other, they are free to choose the visual and auditory channels they want to map their data to. During the classification of the surveyed literature, we used six different categories to capture the state-of-the-art of audiovisual idiom design. The categories are the visualization idiom, the sonification technique, the visual identity- and magnitude channels, as well as the auditory identity- and magnitude channels. The constructs of identity- and magnitude channels are used to distinguish between encodings that communicate "what" something is and "how much of" something there is [Mun15, ERI*23]. Typical visual identity channels are the shape of a visual mark or its color hue. Typical visual magnitude channels are the position or the length of a visual mark. A typical auditory identity channel is the timbre of a sound, often generated using different musical instruments. Typical auditory magnitude channels are pitch and loudness [DB13, CLR22]. In the following, we will make use of these descriptions to shed light on some of the designs in our database.

SoniScope [ERI*22] is an interactive audiovisual display idiom that combines a scatterplot visualization with an interactive parameter mapping sonification (compare Figure 2). Inspired by the technique of auscultation, the SoniScope acts as a stethoscope for data, providing an auditory lens to "listen into" one's data. The exploratory character of the idiom is most apparent when the sonification displays non-visual data dimensions. To interact with the idiom, a user can brush the data using a visual lens. Clicking into the scatterplot will then trigger the sonification of a non-visual attribute of the selected data items. Regarding the used magnitude channels, we see a common combination of visual position and the pitch of the sounds, also employed in publications such as [BF12, DLVDCG22, FBC12, FN18, KLTW17].

Bearman [Bea11] studied the possibilities of displaying the uncertainty in future climate projections for the UK. The proposed audiovisual idiom shows a map of the UK with a heat map overlay

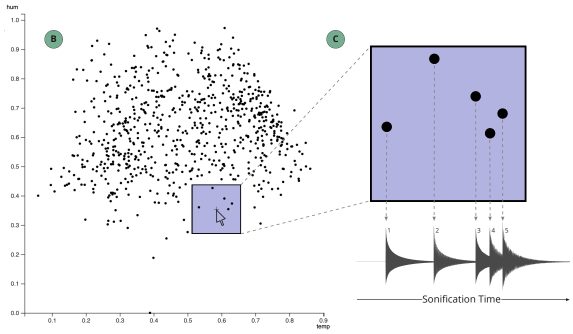
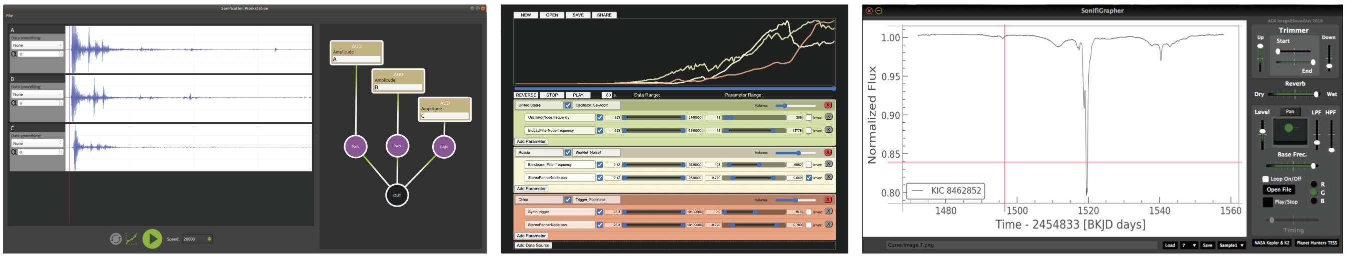


Figure 2: *The SoniScope audiovisual display idiom [ERI*22]. Users can interactively select a region in a scatterplot to sonify an additional data attribute for the respective items.*

displaying the projected temperature values. This visualization is combined with the sonification technique of interactive parameter mapping. People in a user study were asked to hover over the heat map with a mouse, triggering the sonification of the respective region. A higher pitch of a trumpet sound communicated a higher uncertainty for that region. Hence, for their design, Bearman used the spatial position as a visual identity channel and color as a visual magnitude channel. This sonification did not use an identity channel (only one sound could be heard) but used pitch as a magnitude channel.

Several papers in our database describe frameworks that are explicitly developed for the design of sonifications (see examples in Figure 3), always in combination with a visualization [PC19, PCB23, LF21, DLVDCG22, Rib19, CWM21]. As a representative and flexible example we want to discuss the *Highcharts Sonification Studio* [CWM21], a collaboration between the company Highcharts and the Georgia Tech Sonification Lab. Regarding the design of the visualization, the environment offers line- and area charts, as well as scatter plots, bar charts, and pie charts, all with their respective standard visual channels such as position, length, angle, or color hue. The sonification is done using parameter mapping with several options for auditory magnitude channels such as pitch, spatial position in the stereo field, loudness, or harmonic range. Different attributes in the data can be distinguished using different musical instruments, hence the employed auditory identity channel is the timbre of the different sounds. The default option of the Highcharts sonification studio is set to combine a line chart with an auditory graph, which traditionally plays back the line from left to right using pitch over time.

In their paper, Winters et al. [WW15] describe the design of the visualization and the sonification of the 2011 Tohoku earthquake in Japan. The design combines four line charts of four seismographic recording stations in Japan with an audification of the same data streams. The essence of the audification is that the low-frequency recordings of the earthquake are played back at a faster tempo, making them audible to the human ear. This very direct mapping between the physical (non-audible) phenomenon and the audification to the audible range results in a rich auditory impression that becomes informative in an ecologically valid manner. The paper



(a) (b) (c)

Figure 3: *Examples of frameworks for sonification in combination with visualization. (a) The Sonification Workstation by Phillips and Cabrera [PC19] is designed for general sonification tasks. (b) The WebAudioXML Sonification Toolkit (WAST) user interface by Lindetorp and Falkenberg [LF21] also aimed for general sonification tasks. (c) The Sonifigrapher virtual synthesizer interface by Riber [Rib19] for sonifying light curves. All screenshots CC BY 4.0.*

mostly discusses the popularity of this sonification on YouTube, currently with around 90k views, explicitly mentioning the relevance of combining the sonification with visualization for its success in public outreach. The visualization uses position as both its identity and its magnitude channel. The sonification employs the spatial stereo position of the four channels to distinguish between them, hence as their identity channel. The magnitude channels are a mix of pitch and timbre, resulting from the direct mapping between the physical phenomenon and the sound.

Discussion: From the data, we see that parameter mapping is by far the most used sonification technique for the design of audiovisual idioms – 53 out of 57 papers in the database employ some sort of parameter mapping sonification. Only seven of them combine parameter mapping sonification with other techniques such as audification [MMU16, HPDW23], earcons [HAR16, NSC16], or auditory icons [BBV16, RFK*15, PBM15]. The technique of audification is part of the database three times on its own [PFH*22, HCTP14, WW15], and once in combination with earcons in a second design of a paper [PFH*22]. Yang and Hermann are the only authors employing model-based sonification with their *Mode Explorer* design [YH17]. The most prominent auditory identity channel in our literature corpus is timbre (used 21 times), and the most used auditory magnitude channel is pitch (47 times), also used in combination with other channels. These findings are just in line with other meta-analyses of the field of sonification [CLR22, DB13]. The visualizations in our corpus employ idioms such as line charts (10 times), scatter plots (8 times), and maps (7 times), as well as networks, bar charts, heatmaps, and several other idioms. These publications use the identity channels color hue (27 times), position (11 times), as well as shape (11 times). The employed visual magnitude channels are position (33 times), color hue (15 times), size (9 times), and other channels such as opacity, tilt, or animation.

Tagging the surveyed literature made us realize that using the concept of the "channel" for techniques such as audification or model-based sonification is not trivial. With these techniques, the designer of a sonification, to some extent, loses control over its sonic outcome. This contradicts the definition of a visual channel as "a way to control the appearance of marks" [Mun15]. Instead, thinking of the "channel" as "the quality of a mark that transports information," also allows for the description of audifications and

Sonification	Visualization		
	whole	group	single
whole	27	3	1
group	24	12	1
single	14	4	3

Table 3: *The number of entries for Bertin's different reading levels [Ber83] and their distribution to the two display types.*

model-based sonifications. While these techniques typically result in highly complex sound sequences, it will often be qualities such as pitch or timbre that are informative to the listener. In general, we see a quite diverse field of different audiovisual idiom designs in our corpus.

3.4. Reading Levels

The classification of designs with respect to reading levels is inspired by the taxonomy by Bertin in his seminal book Semiology of Graphics [Ber83]. The three reading levels describe the amount of data a user studies using a specific tool. The "Whole" level describes tools that enable the user to ask questions about "all" of the data under consideration. The "Group" level describes tools that enable the user to ask questions about a "subgroup" of the data under consideration, and the "Single" level describes tools that enable the user to study "single" items. For this category, we decided to assign two tags per entry in our database: One for the visualization part and one for the sonification part of the idiom. The classification using reading levels aims towards potential differences in the distribution of tasks between sonification and visualization, such as overview and detail phenomena.

Rau et al. [RFK*15] presented an audiovisual idiom that lets users interactively explore molecular structures by using a "virtual microphone" that can be placed inside a 3D molecular visualization. The visual design provides an overview using the "whole" level while displaying details at the "group" level via the sonification. At the same time, the design provides a user with information about potentially visually occluded data. Hence, it makes the user aware of the existence of a phenomenon that they could, if relevant, study

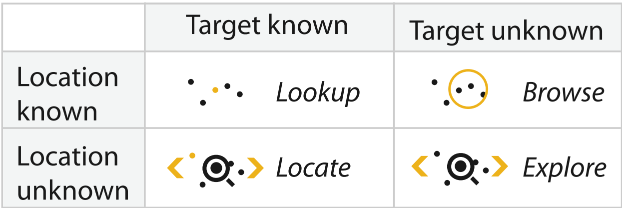


Figure 4: *The four search levels suggested by Munzner: lookup, browse, locate, and explore. (Figure from "Visualization Analysis and Design" [Mun15] by Tamara Munzner, with illustrations by Eamonn Maguire, CC BY 4.0.)*

Sonification	Visualization explore	browse	locate	lookup
explore	37	4	5	5
browse	3	2	2	2
locate	4	3	5	5
lookup	1	0	1	2

Table 4: *The number of entries for Munzner's different search levels [Mun15] and their distribution to the two display types.*

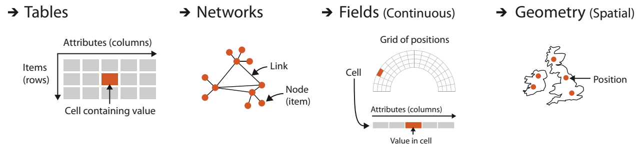


Figure 5: *Four dataset types suggested by Munzner: tables, networks, fields, and geometries. (Figure from "Visualization Analysis and Design" [Mun15] by Tamara Munzner, with illustrations by Eamonn Maguire, CC BY 4.0.)*

Sonification	Visualization table	network	field	geometry
table	23	1	4	12
network	0	2	0	0
field	0	0	7	1
geometry	1	0	0	9

Table 5: *The number of entries for the different dataset types [Mun15] and their combinations between the two display techniques. Notably, the visualization of geometries is often combined with the sonification of table data.*

in detail at a later point. The metaphor of the microphone to be used to "listen into the data" is similar to the one of the stethoscope in [ERI*22], where the visualization provides an overview while users can interactively choose a subset of data to display acoustically. An example of an idiom working at the sonification "whole" level and the visualization "group" level is the *Mode Explorer* [YH17]. The design sonifies a high-dimensional data space while visualizing a two-dimensional projection of that space with a scatterplot.

Discussion: Out of all possible combinations of reading levels and display type, the majority of cases (27) use the "Whole" level for both the sonification and the visualization (some of them are [EEBR21, MMM18, RS22, SAR22, WW15]). The next biggest group in our database holds 24 papers that can display data at the "Whole" level using their visualization and at the "group" level using their sonification, such as [BF12, BTB23, FBC12, GDAS*18, RFK*15]. Generally speaking, we can observe that the visualization parts are mostly covering the same or a higher level of reading than the sonification parts of a design (compare Table 3). This phenomenon can be directly related to the distribution of tasks on the two display types, with visualization more often providing an overview and the sonification rather displaying details. In section 6, we will argue for breaking such patterns in the future work of the community.

3.5. Search Level

Inspired by the taxonomy suggested by Munzner [Mun15], we tagged four different search levels, again assigned individually for the sonification and the visualization parts of the idiom. The type of search a user applies depends on their prior knowledge (see Figure 4). Users who know what they are looking for and where they can find it will do a *lookup*. Searching for an unknown pattern at a known location is called *browsing* while *locating* is a search without knowing the location but the pattern one is looking for. A search uninformed both with regards to the location and the type of pattern one is looking for is called *exploration*. Designs that combine sonification and visualization could serve each of the described search levels as well as a combination of them.

Malikova et al. [MAFP19b], for example, presented an idiom that helps users explore scalar fields. The regional magnitude of small areas of a scalar field is sonified using pitch. From the search level perspective, a user of this idiom does not generally know what to look or listen for at the beginning of the analysis, and therefore, both modalities are used to explore the presented data. The system could, for example, reveal the existence of small symmetry differences in scalar fields. What visually seems to be a perfectly symmetrical field could become apparent as not quite symmetric when two similar pitches result in an acoustic phenomenon called "frequency beating." The clearly audible phenomenon makes a listener aware of the non-symmetry of the field, potentially resulting in them taking a closer look at their data.

Discussion: Generally speaking, many idioms can be used for more than one search level, and which one they are used for seems to be dependent on the user's prior knowledge (both about the data and about the idiom). Nevertheless, the majority of papers in the surveyed literature offer the search level of "explore" and, in general, most papers use the same search level for both their visualization and their sonification parts (see Table 2 and Table 4). Some examples of designs employing the same search level are [CB17, DCM*18, ERI*22, FBC12, Rib19, GR11, HK22], while fewer combine different search levels for the two modalities [HAR16, RFK*15, RJ16, AK11, Bal15, CWM21].

Sonification	**Visualization**			
	ratio	interval	ordinal	nominal
ratio	21	11	0	10
interval	7	22	2	12
ordinal	0	2	8	7
nominal	4	2	5	12

Table 6: *The number of entries for the different levels of measurement and their combinations between the two display techniques. Most idioms use identical levels of measurement. Notably, 42 cases use a higher level of measurement with their sonification than with their visualization (numbers above the main diagonal). Twenty cases use the levels of measurement the other way around (numbers under the main diagonal).*

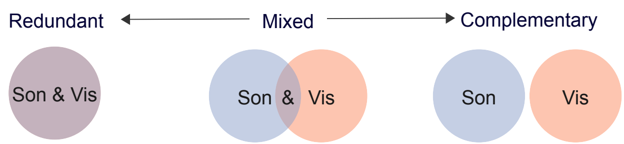


Figure 6: *The three levels of mapping-redundancy: Redundant mappings display the same information via visualization and sonification. Complementary mappings display different information to the two senses, and mixed mappings map some information redundantly and others in a complementary way.*

3.6. Dataset Type and Level of Measurement

Two more categories that we tagged individually for the sonification parts and the visualization parts are the dataset type (see Figure 5 and Table 2) and the level of measurement (see Table 2). We distinguish between the four dataset types of tables, networks, fields, and geometries [Mun15] and between the four levels of measurement of nominal, ordinal, interval, and ratio scale.

It is reasonable that most idioms display data from identical dataset types with both their visualization and their sonification. The most prominent dataset type is the table, used in designs such as [BTB23, DLVDCG22, DCM*18, ERI*22, FBC12, HAR16, MMM18, RJ16]. Table 5 reveals a notable exception from the dominance of identical dataset types: the combination of visualized geometry in combination with a sonified tables, often related to maps and supplemented regional information [BF12, LSB*23, MMU16, NRL*12, RS22].

Concerning the levels of measurement of the displayed data, we observe, again, that most idioms display data from the same level (see Table 5). We list a selection of cases for ratio data [ASH*12, DCM*18, GRK*16, PFH*22, Rön21], for interval data [BB19, DLVDCG22, Rib19, Her20, MMU16], for ordinal data [AK11, LF21, PCB23], and for nominal data [GKW21, HAR16, NSC16], and now will continue with a discussion of three methods to map the data to the senses, the levels of redundancy.

3.7. Level of Redundancy

When combining sonification and visualization, designers have three options – displayed in Figure 6 – to distribute the information they want to transport to the senses. They can have the same information represented in both the sonification and the visualization (redundant mappings), they can display one part of the data exclusively visually and another exclusively auditorily (complementary mappings), or they can map some parts of the data redundantly and some complementary (mixed mappings). While this continuum seems to be an intuitive description of audiovisual display idioms, we should distinguish between “technical redundancy” and “communicative redundancy.”

Technical redundancy describes the technical mapping from the data to a visual and auditory representation. If all the displayed information is mapped to channels of both modalities, then the audiovisual display idiom employs a technically redundant mapping. An example is the auditory line graph: We see and hear identical information via spatial position (visualization) and pitch (sonification). In some cases, it is not enough to use a purely technical description of redundancy without incorporating our way of perceiving the different displays as humans. Therefore, we introduce the term "communicative redundancy."

Communicative redundancy describes the fact that technically redundant designs might encode the same information with different perceptual qualities. Hence, we could identify different patterns in data by listening to them than by looking at them. Communicative redundancy will usually be strongly related to technical redundancy, but there are exceptions, such as the combination of a WAV form visualization (line chart) and an audification, such as in the Sonification of the Tohoku Earthquake in Japan in 2011, also described in [WW15]. Technically, those two are fully redundant (the same data attributes are displayed both visually and auditorily), but communicatively, they complement each other.

In the following, when we speak of redundancy, we use the definition of technical redundancy. Communicative redundancy is most likely also dependent on the individual receiver, which is why we would not be able to consistently assign a tag to each case. Nevertheless, when implementing an audiovisual display idiom, a designer should consider its communicative redundancy just as well.

Rönnberg and Johannson [RJ16] present a technically redundant mapping of the density of a scatterplot or a parallel coordinates plot to auditory channels. Users would be able to hover over the plot and listen to an auditory representation of the density in the data. It is technically redundant as we are generally able to see the density of the plot by looking at it. In a communicative sense, the design is not redundant, which is why their evaluation shows that the additional sonification helps users identify especially dense areas. The eyes' ability to assess the visual density is supported by the sound of their design.

Dedicated environments that offer the design of audiovisual display idioms frequently employ redundant mappings [PC19, Rib18, CWM21, DLVDCG22, KLTW17, Rib19]. One of them is the *Sonifigrapher* (see Figure 3(c)), presented by García [Rib19], that combines a line chart and a parameter mapping sonification of the light curves from NASA's exoplanet archive. The vertical position of a line is mapped to the pitch of a musical sound, essentially "playing back" the line from left to right (compare Figure 7).

The work presented by Rau et al. [RFK*15] complementarily enhances the visualization of molecular simulations by using auditory icons and parameter mapping sonification. A user can position a "virtual microphone" inside of a 3D-rendered visual representation of a molecule and listen to visually occluded processes.

Temor et al. [TMN*21] presented a mixed mapping approach for the audiovisual analysis of computational fluid dynamics tailored towards the prediction of cerebral aneurysm ruptures, hence in the medical context. Their sonification design is psychoacoustically motivated in a way that amplifies the differences between different simulations of their spatiotemporally dynamic data. The authors explicitly mention their choice to apply a mixed mapping in their design. They motivate their decision by the observation that the "*spatiotemporal fluctuations are highlighted in a way that seems to be superior to the presentation of different information to different sensory modalities, which is in line with how we make sense of spatiotemporally-dynamic stimuli in everyday scenarios.*" We consider this observation highly plausible and inspirational for potential future research and will reflect upon it in the concluding discussion of this STAR.

Maças & Martins, and Machado present an audiovisual display idiom that displays consumption patterns collected from Portuguese supermarkets over the course of two years. The authors, just in line with Temor et al. [TMN*21], also argue for the employment of a mixed design, their teams having made prior experience with complementary designs that seem to have been less effective.

Discussion: Regarding the level of mapping redundancy, a high-level observation is the following: When designing an audiovisual display idiom, mixed mapping seems to show a special potential. The mapping overlap between the sonification and the visualization seems to help the user perceive an idiom as integrated rather than as two displays existing next to each other. There are different options to design a mixed mapping, out of which a promising one seems to be to synchronize a visual animation with a sonification [TMN*21, MNW*18]. In a similar manner, synchronizing the spatial position of the visual display with the direction that a user perceives the sound from can be helpful to perceptually integrate the two stimuli [CB17]. An audiovisual analysis idiom that utilizes complementary or mixed mappings enables a designer to choose the employment of sound instead of a second view. Data that would conventionally be made visible via a second view can be represented using sound instead, which can modify the design of the visual view itself. Such situations can become especially apparent whenever the visualization needs to fit on small screens such as on smartwatches.

3.8. Evaluation Approaches

The evaluation of designs is a pressing issue in both the sonification and visualization fields. Audiovisual idioms likewise need to be evaluated. Rönnberg and Forsell even argue for the standardization of questionnaires that assess the usability of audiovisual representations that could be used in combination with other measurements [RF22]. To study the current practice of evaluation of audiovisual designs, we applied four of the classes of evaluation techniques suggested by Isenberg et al. [IIC*13]. The four classes are user performance (UP), user experience (UE), algorithmic performance (AP), and qualitative result inspection (QRI).

When evaluations measure how specific features in a visualization or a sonification affect user performance with a system, these approaches belong to the evaluation class UP [IIC*13]. Controlled experiments using various time measurements and accuracy are typical example methods in this class. Rönnberg and Johansson [RJ16] explored the combination of visualization and sonification, where the users had to identify the visual area with the highest density in visual representations with and without the support of sonification. The authors measured accuracy and response time, and the results showed that the combination of visualization and sonification increased accuracy in comparison with visualization only but also that response times were longer when sonification was used.

The class of UE includes evaluations that focus on subjective feedback on and experiences with a visualization or a sonification [IIC*13]. Interviews and/or various questionnaires are common evaluation methods. In our corpus of papers, this was a common approach for evaluating the audiovisual representations and was often used in combination with other evaluation approaches. The paper by Ballweg et al., [BBV16] presents a study where users answered a questionnaire about their experience of an audiovisual system for drug design on a 5-point Likert scale. By comparing responses given before and after the study, it could be determined how the system could be integrated into the users' workflow, if sonification could have a positive effect on solving the task, and in what way the system could be improved.

Few papers in the systematic literature review used AP as an evaluation approach. Evaluation approaches in this class should contain a quantitative study of the performance or quality of visualization and/or sonification algorithms [IIC*13]. However, in our corpus of papers, this approach was only found in a few cases and was employed to determine that an audiovisual algorithm could produce sufficient quality rather than determining a certain level of quality or comparing different algorithms. As an example, in the paper by Kariyado et al. [KAV21], it is stated that the evaluation showed that the system presented in the paper allowed for a default amount of 255 audio sources with any audio drop-outs.

Evaluations in the class QRI aim to draw conclusions based on qualitative discussions and assessments of audiovisual representations [IIC*13]. In contrast to UE, these types of evaluations do not involve end users but instead, ask the user to assess the representation for themselves. In the work of Bru et al. [BTB23], an approach to a combined audiovisual representation based on parallel coordinates or dense line charts is presented. Several different attributes and characteristics in the representations are discussed based on the researchers' own reflections, but no formal user evaluation has been performed.

Discussion: We have now demonstrated there are various approaches for evaluating audiovisual representations. The aim of the work and the research questions that are asked determine the class of evaluation chosen. In some cases, in the corpus of papers, two of these evaluation approaches have been combined to provide a better and more detailed analysis of the outcome of the study findings [BBV16, GRK*16, PFH*22]. Yet about a third of the studies included in the systematic review did not include an evaluation at all. The absence of evaluation might lead to a situation where promising audiovisual design ideas might be rejected because clear and convincing evidence of their usability, benefits, and function is not presented in a paper, or less useful ideas are overrated and promoted.

3.9. Target Platforms and Interactivity

Audiovisual display idioms can be displayed on different target platforms and in different environments (see examples in Figure 7). A typical display for the combination of sonification and visualization is the computer desktop environment, as this is also the most commonly used environment for visualization-only designs. About two-thirds of the designs included in this survey are developed for a desktop environment.

The second largest category of target platforms was physical and/or multi-user environments. This includes dome theaters and planetariums, which are commonly used for audiovisual display idioms toward topics of astronomy (as previously mentioned in subsection 3.1) [HTHB22, RS22, EEBR21]. Another type of physical environment is a dedicated immersive environment for collaborative data analysis, where several users can take part in the data exploration. One example of such an environment is the *CRAIVE-Lab* (Collaborative-Research Augmented Immersive Virtual Environment Laboratory), where a design visualized and sonified market data of 128 corporations by using the large panoramic display and the 128-channel loudspeaker array of the environment [CB17] (see Figure 7(b)).

A low-cost alternative for physical environments is a virtual one, which is most commonly enabled through head-mounted displays. An approach for an audiovisual display idiom in this type of environment is to let the user navigate the dataset from a first-person perspective and use spatialized sonification, which dynamically changes depending on where the user is positioned in the dataset. An example is the design of Berger & Bill (see Figure 7(a)), which facilitates an immersive exploration of urban noise standards [BB19] by creating a virtual environment of a city where the sonification allows the user to listen to the collected noise levels by navigating the environment.

The use of touch displays and other tangible interfaces for an audiovisual display idiom can allow for a unified integration of sonification and visualization. A recurring approach is to use the visualization as a graphical user interface that enables the sonification upon interaction. Through a touch display, the design by Ferguson et al. [FBC12] enables the user to filter a line graph with multi-touch gestures to select what data should be sonified. As a more analog approach, the system *Sonophenology* by Ness et al. [NRL*12] lets the user select areas on a printed color-coded paper map through fiducial markers, which in turn selects what information should be conveyed by the sonification.

About two-thirds of the papers in the database include some form of interaction. One of the most recurring objectives of the interaction was to make a selection of the dataset that would be sonified [FBC12]. This can include selecting a geographic region to which the dataset is attributed to [NRL*12]. For example, the audiovisual display idiom of Matsubara et al. [MMU16] creates an interactive sonification system for exploration of seismic data through a horizontal and vertical range slider to specify a geographic region. Other forms of interaction include simpler tasks such as browsing sonifications of individual or groups of data objects [Rib19, LSB*23]. Huppenkothen et al. [HPDW23] created an interactive multimedia version of the Hertzsprung-Russell Diagram, where a user can select individual data entries of the diagram to listen to their sonification. On the other hand, there also exist more complex forms of interactions such as navigating a 3D environment [BB19, JMP13, MNW*18].

Papachristodoulou et al. [PBM15] created a design for navigating complex datasets of brain networks, where the sonification conveyed information about different brain regions of the dataset. Within the sonification community, it was the introduction of model-based sonification [Her02], that made user interaction an integral part of the data analysis process. An example of a model-based sonification is the *Mode Explorer* by Yang and Hermann [YH18], where scratching-interactions with a pencil on a scatter plot enable the user to investigate different modes and their properties through the sonification.

Discussion: When displaying an audiovisual display idiom on a target platform other than a desktop environment, it often creates opportunities for interacting with the system in a novel manner [EEBR21, HK22, NRL*12, PBM15]. In the context of data analysis, interaction is also a particularly relevant part of the user experience [BBV16, Bea11, EEBR21, HK22].

3.10. Users and Goals

Just as with visualization-only designs, audiovisual display idioms are often developed toward specific users and end goals. The most commonly occurring goal for the papers in the database was “data analysis.” This could either be to present a dataset to the user or letting the user explore one or several datasets with the idiom, as described in subsection 3.2. The benefit of using visualization and sonification in this regard is that the user can gain different perspectives of the data through the two sensory modalities. For example, Alonso-Arevalo et al. [ASH*12] created a multimodal interface for curve shapes and curvature, where the user can evaluate the quality of a three-dimensional shape by using both the visual and auditory, and in this case even their haptic perception. The curve shape was mapped to the fundamental frequency of different carrier sounds, and the idiom offered different sound designs to convey the information.

The most recurring pair of users and goals in the database was data analysis for domain experts, which most commonly would involve an audiovisual display idiom to convey data in a specific application domain in the natural sciences. Temor et al. [TMN*21] and MacDonald et al. [MNW*18] created an auditory complement to a scientific visualization of computational fluid dynamics, where the sonification aided in understanding the temporal changes in the visual animation. Gionfrida et al. [GRK*16] suggest combining the visualization of brain scans (in the context of Alzheimer’s dementia research) with a parameter mapping sonification that they call *Triple-Tone Sonification*. The design makes use of the fact that the human

(a) (b) (c)

Figure 7: *Examples of different target platforms and environments. (a) The combination of VR visualization and sonification for an immersive exploration of noise in urban environments by Berger and Bill [BB19]. (b) An immersive virtual environment for audiovisual spatialized data sonification presented by Chabot and Braasch [CB17]. (c) A screenshot showing the interactive interface for the sonified Hertzsprung-Russel diagram by Huppenkothen et al. [HPDW23]. All screenshots CC BY 4.0.*

ear will perceive the sound of two or more very similar frequencies as "frequency beating," which is also employed in [MAFP19a]. For the purpose of process execution data analysis, Hildebrandt et al. [HAR16] demonstrated how incorporating sound could enhance the process of identifying anomalies or conducting root cause analysis for irregularities and errors. A design for a more general user case was created by North et al. [NSC16], which created a an idiom to convey Git version control data. The temporal nature of the data lent itself well to be sonified by sequentially going through the data, where an earcon is played when a commit occurs, and drum sounds are played whenever a conflict occurs in the data.

The second most commonly occurring pair of users and goals was public engagement for the general public, where the use of an audiovisual display idiom has the potential to enable higher engagement by conveying information through both the visual and auditory modality. This would commonly be targeted towards popular science topics such as climate change [HK22, Bal15] and astronomy [HTHB22, TB23]. As previously mentioned in subsection 3.2, the work by Russo et al. [RS22] is an example of a sonification targeted towards the general public for public engagement, where the response of the resulting videos on social media was in part used as a metric for the public engagement of the design.

Discussion: Regarding the successful design of audiovisual idioms that support their users in analyzing data, it is most relevant to include them in the design process. Both the visualization community and the sonification community have individually studied the relevance of including domain experts in their design process [SMM12, Gou17], and the same will be necessary for the integration of sonification and visualization. The co-author network shown in Figure 9 displays quite many domain experts being collaborators in the surveyed literature, which can be regarded as a promising sign in general.

4. Survey Data Analyses

So far, we have focused on the discussion of the audiovisual idiom designs themselves. In this section, we want to take an even broader perspective on the collected data. We will study existing relationships between different tags along a correlation matrix in subsection 4.1, as well as the network of co-authors and its implication for the field in subsection 4.2.

4.1. Correlation Matrix Analysis

To understand potential relationships between different tags within our classification, we computed a correlation matrix, shown in Figure 8. We want to highlight some of the found correlations in the data out of which most are not surprising, nevertheless, they are not obvious and our data enables us to take such a meta-perspective.

Close to the main diagonal of the correlation matrix, we highlighted four smaller matrices **(1)**, showing an interesting phenomenon: Whenever the tagging of the papers was done individually for the visualization and the sonification parts, the two were mostly tagged with the same label. Hence, the visualizations and the sonification, in many cases, use the same *reading levels* [Ber83], the same *search levels* [Mun15], the same *dataset types*, and the same *levels of measurement*. This finding can motivate future research regarding the reading levels and the search levels. Studies could investigate the potential of other distributions of the reading level onto the senses.

Marked with **(2)**, we highlighted four phenomena concerning the purpose of an audiovisual idiom. Idioms designed to present rather than explore data are likely to also use the reading level “whole” for their sonification ($r = 0.41$). On the other hand, idioms that are designed for exploration likely use the group reading level ($r = 0.3$). We see a correlation between the purpose of an idiom and its interactivity, with idioms for exploration more likely to be interactive ($r = 0.23$) and idioms for presentation more likely not to be interactive ($r = -0.48$). Also, domain experts use more idioms for exploration ($r = 0.44$), while the general public is more likely to be ‘just’ presented with data ($r = 0.50$), going hand in hand with their general goals of data analysis ($r = 0.52$) or public engagement ($r = 0.50$).

Two more relationships with the interaction tag emerge from the correlation matrix **(3)**. A sonification that uses the reading levels of “group” or “single” is likely to offer user interaction ($r_{group} = 0.45$ and $r_{single} = 0.24$), while a sonification with the reading level of “whole” does not usually offer user interaction ($r_{whole} = -0.28$). Furthermore, designs that map data to the senses in a complementary way are more likely to offer interaction to their users ($r = 0.32$).

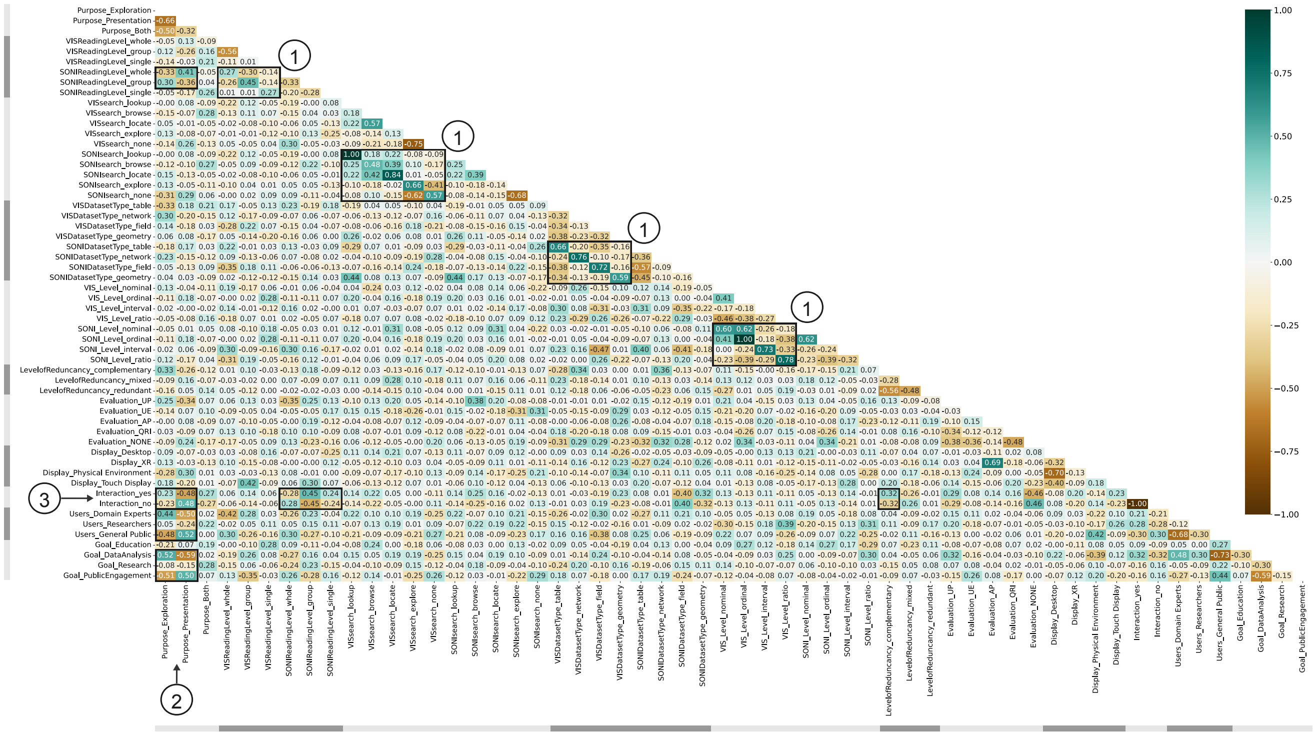


Figure 8: *The correlation matrix of a selection of classification tags. A diverging color palette is used with white in the center representing no correlation, teal (up) increasingly positive correlations, and brown (down) increasingly negative correlations. (1) reveals similar tags for visualization and sonification parts; (2) shows correlations between purpose, reading level, interactivity, and intended audience; (3) uncovers relationships with the interaction tag. A high-resolution version of this figure is part of the supplemental material.*

Overall, most of the computed correlation values in the correlation matrix are close to zero. Also, some of the existing higher values are likely not significant due to the low amount of data for the respective tags (such as with the AP evaluation, which has been used in only two papers). Therefore, we only discussed three perspectives that have considerable correlation values, are plausible, and that we consider most relevant to the bigger picture.

4.2. Co-Author Network and Development Over Time

As part of our meta-analyses, we studied the co-author network within our database, providing a good picture of the state of the research community interested in audiovisual analytics. Each node in the co-author network represents an author with a total of 165 nodes (in 48 teams), and a link connects two nodes if the corresponding authors have collaboratively contributed to a publication. We assigned a color to each node using the primary discipline of the author and classified them, to the best of our abilities, into one of four disciplines: sonification, visualization, domain, and unclassified. For each author, we considered their main publication focus, particularly around the time of their contribution to the works in our database, along with their background, education, their role and input to the relevant papers. Authors classified under “domain” are recognized as domain experts, and the category “unclassified” encompasses scenarios where the author’s primary discipline does not fall into

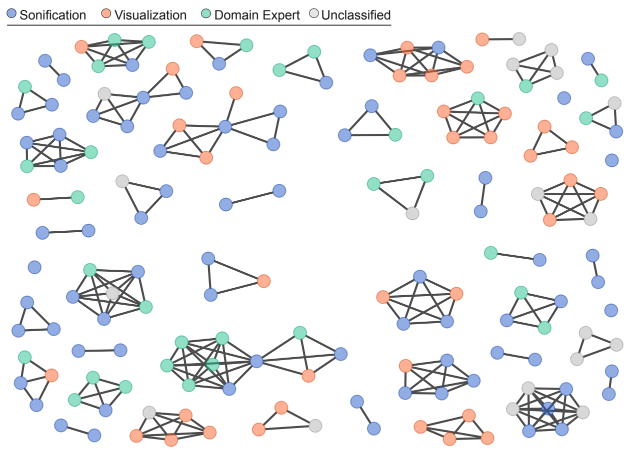


Figure 9: *The co-author network, displaying 165 authors of the 57 papers in the database. The links represent papers written by a team of co-authors, and the colors display their primary disciplines. An interactive version of the network can be explored using the Orange workflow, which is part of the supplemental material.*

any of the other three categories, the author is in the early stages of their career without a defined research focus, or that the information about the author is not available.

The network offers several insights into the structure of the field. The majority of co-author teams are individual teams without interconnections. Only in four cases have authors collaborated with different (groups of) co-authors, usually because one person worked with different collaborators. This phenomenon differs from networks in other fields, such as the one described in a recent STAR on the use of embeddings in visual analytics [HWKK23], showing a network with clusters up to the size of 75 authors (note that the STAR on embeddings has 122 entries). We interpret the dominance of many disjoint groups as a sign for the rather young field and hope, with this STAR, to contribute to the future development of the field growing closer together. We identified 14 teams or individual authors contributing to the corpus that consist of sonification experts exclusively. Two of the teams consist only of visualization researchers, and 21 of the teams are collaborations between the communities, including domain experts in many cases. The inclusion of domain experts and the quite large number of diverse groups also including domain experts can be considered a promising sign for the future development of the field.

Our data also enables us to take a basic temporal perspective on the development of the field. During the years 2011 and 2023, overall, 57 papers were published. The majority of papers were published in the second half of this time span, pointing towards a generally increasing trend. A drop in publications in the year 2020 is likely a phenomenon related to the COVID-19 pandemic, being dominant worldwide, especially during the year 2020.

5. Adjacent Topics

As mentioned earlier, three distinct topics are adjacent to the scope of our STAR: audiovisual idioms in the context of accessibility, real-time monitoring, and arts. Although the three areas are related, they might be better served by a classification system different from the one we used. Still, they are relevant and can be inspirational for our field. The following three sections present work from those adjacent fields that have been curated by our team of co-authors. The sections do not claim completeness or the usage of a systematic approach but are intended as a general introduction for the interested reader.

5.1. Accessibility

Even though visualization is one of the most common ways of communicating data, many visualizations are inaccessible to readers with visual impairments. Some studies have suggested the use of sonification to support readers with visual impairments and a recent STAR on accessibility research in visualization identified 16 out of 56 papers that utilize sonification [KJRK21]. During the filtering and classification of our assembled literature, a category emerged exploring sonification for accessibility of visualization. The literature that was sorted into this category focused on the design and evaluation of sonification to improve the accessibility of visualizations for visually impaired people. As the literature in this category did not explore integrated audiovisual representations for data analysis in general but rather used sonification as a support technique, these are not included in the systematic part of our STAR.

Sonification for accessibility in this way often suggests sonification designs leaning towards auditory graphs. Auditory graphs are the auditory equivalent to visual representations, such as plots, graphs, and charts [HHN11], but instead of mapping data properties to line positions or the size of an area, they are mapped to various auditory parameters such as pitch or loudness. Auditory graphs are suggested to be useful as sensory substitution and as a means of presenting information when line of sight is not possible [SNH05], and using sound to represent data makes data analysis more accessible [WM10]. To give the reader a brief introduction to this category, here are a few recent studies that have explored sonification for accessibility in visualization.

One study explored the accessibility of data visualizations presenting data to blind and visually impaired users [FFSR*23]. It was found that sonification might make discerning trends in the data more accessible, but also that the lack of experience in using sonification could lead to misinterpretations of the presented data. The study was conducted using screen readers, and the users often used alt-text information to support and validate what was heard through the sonification. Another study explored the use of natural sounds mapped to data visualizations, bar charts, and line charts [HEUHEB23]. The reason for using natural sounds was to support users without musical training in understanding the sonification and auditory graph, and it was found that these natural sounds could support the understanding of categorical data and were most useful for users without musical training. Infographics, the combination of visualization and text information, has been explored with an auditory-only approach [HGI*22]. An interactive approach, infosonic, was explored to facilitate accessibility of data to blind and low-vision users, using spoken introduction and annotation, and non-speech sonification. The study shows that the sonification approach supported understanding and forming a mental image of the data.

Often, accessibility studies tend to blend sonification, i.e., non-speech sounds, with generated speech or screen readers. This makes the interpretation of the results from these studies somewhat challenging. The findings, i.e., the experiences and the understanding of the sonification, might be due to the sonification itself or depending on other auditory stimuli used. Also, the visual representation, the user interaction, and the user interface are of importance to the study outcome. In visual data analysis, interaction is essential for exploring the data [CGM19], and similarly, for sonification to be a useful tool, dynamic human interaction is necessary [HH04]. Sonification for accessibility of visualization for visually impaired users is an exciting approach and some of these findings could suggest interesting approaches also for sighted users in supporting visual perception with sonification [Rön19a, Rön19b], or reducing cognitive load [ZPR16] on the visual modality by sonification (see for example [MLS95]).

5.2. Monitoring

A comprehensive overview or a systematic literature review of audiovisual process monitoring is far beyond the scope of this report.

This section is intended to provide some insights into this research field and to distinguish the various scientific communities involved.

The combination of visual and auditory displays for process monitoring purposes has been well-established in real-world applications and interfaces. The audiovisual representation of Morse code for an exchange of information serves as an early example of a multimodal monitoring display and interface. In supervising and controlling the states of various devices and processes, audiovisual interfaces combine the advantages of both modes: volatile alerting sound signals enhance situation awareness, while visual cues provide additional information about the situation, enabling users to take action. Implementations of audiovisual interfaces are ubiquitous in both professional and day-life environments. For instance, consider the warning beep, which can be heard when the temperature outside your car drops below 4°C, alerting you to the possibility of slippery road conditions. A quick glance at the visual display will provide you with the exact temperature and the opportunity to decide whether to take action or not.

Besides the approaches to audiovisual data analysis described in the previous sections of this report, audiovisual monitoring interfaces play an integral role in various domains, including air traffic control [EBL*23], control rooms [SFLD22, HHG*12], anesthesia [AQH*21, RMS*22], neurology [LCS*28], dermoscopy [WRK*19], surgery [Zie23], network monitoring [Wor19a, AHvR*21], automotive [JDS15, XWX*22], to name a few.

During the literature search for this STAR, several publications were labeled with the tag "monitoring." After closer investigation, which involved excluding multiple publications written by identical authors on very similar topics, as well as publications addressing topics of accessibility and artistic media installations, only a relatively small representative subset of 14 findings qualifies for audiovisual process monitoring within the search criteria of this STAR.

The reasons for this rather limited number of retrieved articles in this category are manifold. Firstly, most research on monitoring applications used for auditory feedback of human behavior in the fields of sports, therapy, or rehabilitation focuses on the design, impact, and evaluation of interfaces that utilize the non-intrusive attributes of auditory displays. Hence, they enable users to perceive information without any interfering actions, such as adjusting their head or body position to view a visual display [SJMT19, VRGM20, MNS16]. While some of these interfaces include visualizations at a low scale level, they hardly qualify as audiovisual monitoring devices in terms of providing a balanced contribution from both modalities.

In industrial and surveillance contexts, factors such as cognitive workload, perception organization, situational awareness, alarm fatigue, and deafness play crucial roles in the successful implementation of infrastructures. However, these attributes are discussed in other scientific communities, including Human Factors and Applied Ergonomics, as well as several areas of medicine and health, particularly in anesthesia, rehabilitation, sports, and cognitive psychology. In addition to performance comparisons between auditory, visual, and audiovisual modalities [AG19], particular interest is given to the dual-task paradigm, which evaluates, for instance, the ability to identify an event requiring action in a secondary domain while simultaneously performing a primary task [LWP*13, HHRM16, TML*15, NGJW14].

(Non-)systematic literature reviews have been published in several of the mentioned areas [WMY*17, KIK19, SMS21, HRM15]. To the best knowledge of the authors, no comprehensive STAR covering the entire spectrum of audiovisual or related multimodal monitoring applications and approaches has been conducted thus far. However, there are some fundamental publications treating the specific attributes and criteria for research and development in the field, such as [Ibe20, Joh04, WS07].

5.3. Arts

During the classification of the corpus, an "art" category emerged, defining projects combining visualization and sonification for artistic purposes. We understand the emergence of art practices that intertwine sonification and visualization not only as inspiring for designers and audiences alike but also relevant to the broader discourse around data representation towards a better, more efficient, and engaging human-data relationship.

Working with the literature, we understood artistic endeavors call for a different kind of classification than idioms intended for audiovisual data analysis. We decided not to include such contributions in the systematic part of this STAR and excluded papers where authors made their intentions of an artistic contribution explicit in the paper.

Nevertheless, to give the reader a brief introduction to the field, we present representative cases of an emerging art category that could form the basis of future research on the integration of sonification and visualization. In addition to the artistic cases that emerged from our literature search, we use four more sources to curate a collection of representative cases for artistic endeavors: The Data Sonification Archive (DSA), the Computer Music Journal, the Leonardo Music Journal, and the Ars Electronica Festival. The Data Sonification Archive is an online crowd-sourced collection of data sonification projects. Out of 455 cases currently hosted in the DSA, two of the co-authors of this report recently categorized 139 cases as "art", i.e., as having their primary purpose in creating and delivering an artistic experience to an audience [LLC23]. A preliminary search on some of the major venues for academic artistic publications – Computer Music Journal and Leonardo Music Journal – accounts for 182 and 118 cases, respectively, in line with the volume of art-related cases in the Data Sonification Archive. A search of the archive of the renowned Ars Electronica Festival on projects that translate data into an audiovisual idiom returns 32 cases, again, in line with the 23 projects currently hosted on the DSA.

Typically, data-driven artworks that use an audiovisual design address a non-expert audience and can take different forms, such as live performances, multimedia installations, video productions, or web experiences where the sonification and visualization of the same dataset are combined. The DSA classification highlights recurring topics such as climate change and/or the representation of environmental data, as well as individual and collective interaction in the urban space, internet, and social media. The predominance of socially relevant topics shows that the emerging art category presents borderline characteristics with the adjacent category (in the DSA classification system, see [LLC23]) of "public engagement".

This latter category identifies projects that, although being often presented in the form of an artistic experience, have explicitly and primarily the goal of increasing awareness – even fostering activism – among the general public on a specific phenomenon.

This is the case for projects that present data related to climate change, such as *Klima|Anlage—Performing climate data* [GHJ*19], one of the cases identified as artistic in our literature search. In this work, climate data and model predictions between 1950 and 2100 can be chosen by the listener for twelve selected regions of the world and interactively sonified and visualized in an immersive installation environment with the goal of contributing to "the urgent need to inform the general public about climate change".

In *Too Blue* [Foo15], hosted on the DSA, the author created an audiovisual experience (in the form of a video) where music generated by tracking the land loss in coastal Louisiana over 78 years due to man-made levees, drilling and dredging for oil and gas, and climate change is combined with an aerial color-coded visual map that shows the increase in the "blues" colored areas, corresponding to the sea, in the coastal land of Louisiana. In *Heat and the Heartbeat of the City* [Pol06], the urgency to communicate climate change to the broader audience combines with a reflection on the urban space through the simultaneous sonification and visualization of actual and projected data on increased temperature in New York City. In *aqua_forensic* [SŠ18] data on invisible pollutants in the Danube river's water, and the effect on its ecosystem, are collected, sonified, and visualized 'to present the results of this scientific research to the wide audience in a poetic and artistic way.' Other prominent social issues such as personal data collection and privacy in the age of social media are at the center of multi-awarded art projects such as *Digital Violence* [FE21] where sonification is explicitly used to increase the audience's engagement with the visualized dataset on the issue of illegal government surveillance practices. From the collective to the individual experience, projects such as *Orbuculum* [Ylm23], and *Deep Sync* [KOR23] invite the public to interactively engage with personal psychological and physiological dimensions to create an immersive auditory and visual experience. In a more-than-human perspective, works such as *Spider Web Sonification* [SQS*20] and *Biota Beats* [KGS*20] shift the focus from humans to other species in an effort to support the public engagement with science.

Data art can stimulate an "artistic affectivization" that can contribute to building shared perspectives within a community [BMLA22]. The engagement of the audience with the phenomenon behind the data, however, also implies an information and knowledge transfer [MVC*10], i.e., the public not only has to connect emotionally with the data but also has to understand their meaning. The emergence of a category of artistic projects that combine sonification and visualization with the explicit intent to activate the audience on a socially relevant issue shows that multimodal approaches have the potential to generate impact, not only for education and awareness but also for stimulating action.

6. Concluding Discussion and Future Work

In this STAR, we have provided an overview of the field of audiovisual idiom design tailored toward data exploration and presentation. We have used a variety of perspectives to classify the existing literature, and, in section 5, we offered insights into three topics that are adjacent to our report and can be highly inspirational for future developments. For our categorizations, we also provided brief discussions or reflections as part of the individual subsections 3.2 to 3.10.

Before we adopt a global viewpoint and discuss broader insights, challenges, and exciting future research opportunities, we now want to mention some of the key insights and results from previous sections. Overall, our corpus holds a quite diverse field of combined designs, employing visual idioms such as scatter plots, maps, and networks, as well as sonification techniques such as parameter mapping sonification, audification, and earcons.

The analysis of employed reading levels [Ber83] of the different designs revealed an imbalance between the two modalities. The visualization part of the identified audiovisual idioms mostly works at higher reading levels than their sonification counterparts. This phenomenon suggests that many designs provide an overview of the data using their visualization, while the sonification supports the analysis of details.

Such design decisions are also related to the level of redundancy an audiovisual idiom employs, and mixed approaches, which represent parts of the data redundantly and other parts in a complementary manner, seem to show special potential. Such a selective alignment of mapping between the sonification and visualization might assist us in perceiving an idiom as integrated rather than as two separate displays coexisting next to each other.

Furthermore, the option to use sonification as part of an idiom's design can directly influence the visual view as well. Thinking of idioms with "multiple views," one of the views could, for example, be covered by an "auditory view." For visualization designers, such considerations are especially relevant when space is limited, as is the case, for example, on smartwatches [LDIC21].

The culture of providing supplemental material: A thriving future for the field of audiovisual idiom design also depends on the communities' culture when providing demos for their designs as supplemental material. Out of 57 articles in our surveyed literature, only 29 provided a demo that is still available today. Five provided a demo that is no longer available, and 23 provided no demo at all. Online repositories such as Zenodo, hosted by CERN, the platform OSF, maintained by the Center for Open Science, or long-term archives such as Phaidra, hosted by the University of Vienna, allow the upload of videos and even the assignment of DOIs. While making sound available in a paper used to be an explicit challenge to the sonification field, it is not the case anymore. An additional visual representation of the sound can also be included in the paper, using a spectrogram of the sound itself or an iconic representation of it.

Evaluating audiovisual idioms: It is often challenging to conduct systematic, well-informed, and, above all, properly performed and reported evaluations of audiovisual idioms, especially when the goal is to compare different studies. As mentioned earlier, Rönnberg and Forsell suggested the standardization of questionnaires tailored to the field in order to increase the comparability between different studies [RF22]. Establishing this comparability is particularly difficult because the two senses are used simultaneously in an audiovisual display, and the evaluation methods must adapt to this characteristic of the design. Specifically, the existing methods focus on several aspects. One is user performance, which studies quantitative metrics such as task completion times and precision of responses. Another is user experience, which focuses more on qualitative metrics such as user engagement and memorability. From the data we collected in our STAR, we can draw high-level conclusions regarding the type of evaluations that have been used within the corpus of literature.

Among the papers that included a user evaluation (21 of the 57 papers), most reported a subjectively stated benefit of sonification in perceiving and interpreting the visualization. Some subjective ratings and feedback also suggested that the use of sonification was engaging (six papers considered this [HK22, RS22, EEBR21, KAV21, Rön21, DCM*18]), could provide emotional content (mainly two papers [RS22, Rön21]), was experienced as interesting (five papers [HK22, RS22, EEBR21, Rön21, DCM*18]), as well as pleasant (four papers [RS22, EEBR21, Rön21, DCM*18]), and in perceived improved quality of visual information (six papers [PFH*22, EEBR21, KAV21, FN18, RJ16, BF12]). Other papers also reported subjective and objective measurements of improved interpretation and higher understanding of data when sonification and visualization were used in combination (16 papers [PFH*22, CWM21, EEBR21, DCM*18, FN18, GDAS*18, YH18, BBV16, GRK*16, NSC16, RJ16, PBM15, PBV14, ASH*12, BF12, Bea11]). This was shown in terms of higher scores or increased accuracy when sonification was used. However, this increase in accuracy was reported to be related to increased task completion times (see, for example, in [RJ16]). Similar effects have also been shown in other studies not included in this systematic review [MBKSM16, Rön19a, Rön19b]. The case might be that the use of the additional information provided by sonification that makes the increased accuracy possible takes a longer time to process and assess. While traditionally increased task completion times are considered a disadvantage, in some contexts, this can be seen as positive. Slowing down the user, resulting in them spending more time with their data, studying it from more and from different perspectives (including the auditory perspective), might reveal structures or patterns in a dataset that would have stayed silent and unseen if analyzed with a conventional unimodal display [HAS11].

These results indicate that there is, in general, a benefit of combining visualization and sonification for data exploration and analysis, regardless of the specific visualization and/or sonification design. Nevertheless, we consider it a major task for the future research of our community to systematically study what kinds of benefits the integration of sonification and visualization can provide.

The potential of unconventional task distribution: The data from our STAR shows that reading levels [Ber83] are not equally distributed to the two senses: The vast majority – 50 out of 57 papers used the reading level "whole" for their visualization (see Table 2), while only about half of the papers use the same reading level for their sonification parts. This observation might be directly related to the way audiovisual idioms distribute the tasks of overview and detail to the two senses. The currently more popular distribution seems to be to use visualization to display an overview of the data, while sonification is employed to provide details. While this distribution seems like an intuitive choice, future research should study the potential of switching the roles between sonification and visualization. It is our daily lives and the way we experience the world around us that suggests sonifying an overview and visualizing details. Our auditory sense constantly screens our 360° environment while our visual sense actively focuses on details in our environment. The field of auditory process monitoring makes use of exactly this situation when alerting us of a status, regardless of our current visual focus. Nevertheless, we identified only three cases where the reading level for the sonification was "whole" and the one for the visualization was "group" [DLVDCG22, FBC12, YH18].

A second phenomenon we observe in the data is that many audiovisual display idioms employ the same search level [Mun15] (see Figure 8), and most of them employ the search level of "explore" (see Table 2). Again, it seems like a fruitful future endeavor to study the opportunities of designs that explicitly break such a pattern.

The potential of integrating sonification and scientific visualization: Scientific visualization typically displays inherently spatial phenomena and is often used to represent data that varies over time. The inherently spatio-temporal quality of such displays constitutes their potential to be integrated with sonification, as sonification is a display technique tailored toward temporal data structures [GGB05, ERI*23]. We relate this argument back to the real-world example from our introduction, where the combination of our visual and auditory senses offers a better understanding of a phenomenon than each sense alone: Rain falling outside of a closed window. Opening the window to not only see but also hear the sound of the rain gives us a better estimation of the amount and intensity of the rainfall. Abstractly, in such a situation, we perceive a spatio-temporal phenomenon. It is most plausible that adding sound to a time series of scientific visualization can increase its holistic interpretability in an ecologically valid manner [Neu04]. This phenomenon is shown in two publications in the database, in the work by MacDonald et al. [MNW*18], and the work by Temor et al. [TMN*21]. Similarly, these ideas also apply to animations in information visualization, where temporal aspects of the data might be visualized and sonified.

Design frameworks: When it comes to the possibilities of designing audiovisual display idioms, most researchers, as well as most domain experts, will not be able to develop their own designs. Interdisciplinary knowledge bridging visualization, sonification, interactive design, and human perception is necessary to design effective, engaging, and re-usable audiovisual display idioms. Therefore, we cannot expect a domain expert to be able to quickly draft a prototype the same way they may be able to do using established visualization-only frameworks such as *matplotlib* [Hun07].

In our database, we identified several design frameworks tackling this challenge [PC19, PCB23, LF21, DLVDCG22, KLTW17, CWM21]. Their primary focus is the design of the sonification part of the idiom, which is why we cannot speak of truly balanced contributions to both senses. A promising endeavor to enable truly balanced designs, with both the visualization and the sonification being equally well-informed, might be to not design them using one but two separate frameworks, each specialized for its purpose. In this regard, the recent work by Reinsch and Hermann [HR21, RH22, RH23] is promising, as it provides a sonification design framework embedded in the *Python* environment and is conceptually inspired by visualization design tools such as *matplotlib*. With both the sonification design and the visualization design happening in the same environment, in this case, for example, a *Jupyter Notebook*, both can be developed with the required quality to meet their individual standards. Regarding the introduction of users to a new design environment, be it for visualization, sonification, or their combination, effective "onboarding" is a pressing issue. In the visualization community, the topic has been studied in recent years [SPA23, SWG*23], and the same will be necessary for sonification and audiovisual display idioms. Again, when it comes to sonification as a technique to display data, the prior experience of domain experts or the general public is low. This lack of experience can, in the short term, only be met with carefully designed onboarding processes.

Adjacent topics as inspiration to the field: Our brief look into the adjacent fields of monitoring, accessibility, and arts shows a special potential for inspiration for our communities' future work. The field of monitoring offers established evaluation methods that could potentially be adopted to support the evaluation of audiovisual idioms tailored toward exploratory data analysis. During our unstructured investigation into the field, we also identified the potential of a systematic STAR as a future contribution to the field of monitoring.

Also, the adjacent field of accessibility can inspire future research, where understanding the analysis patterns of visually impaired individuals can inform design decisions for idioms tailored towards sighted people [WM10]. While the field of accessibility can be informative to the field of audiovisual idioms, the same holds true the other way around. Especially the designs from our STAR that use redundant mapping to both the visual and the auditory senses can inspire future designs for visually impaired users. Furthermore, such designs have the potential to foster successful collaborative data analysis involving both sighted and blind users.

The identified artistic contributions show great potential to be inspirational to the field of audiovisual idiom design as well. In the future, effort should be made by the research community to develop a specific framework for the evaluation and analysis of audiovisual idioms at the crossing of art and public engagement. Such efforts need to explore how specific design strategies (e.g., interactivity and embodiment, as used in many of the cases presented in subsection 5.3) can be combined to support both affect and sense-making. Again, a systematic STAR dedicated to the artistic perspective would be a highly timely contribution to the community.

An audiovisual analytics community: With this state-of-the-art report, we hope to contribute to the future establishment of systematic research on audiovisual display idiom design. The current development concerning the studied co-author network and the temporal development of publication numbers in the field is promising. We hope to reach researchers from both the visualization and sonification communities, inspiring them for future interdisciplinary collaborations, and to *open their ears and take a look*.

Supplemental Material

As supplemental material to this STAR, we provide at https://phaidra.fhstp.ac.at/o:5541:

- corpus.bib and corpus.rdf: publication metadata of the surveyed literature in BibTeX and Zotero RDF format.
- corpus-tagging.csv: a table holding the surveyed literature and all of the used tags. Interested readers may use the table to identify articles that fall into specific combinations of classes.
- authors.csv: a table holding the names of the authors in our STAR corpus and their primary discipline. Interested readers may use it to identify potential future collaborators.
- corrmat.pdf: a high-resolution version of Figure 8.
- authornet.ows: the Orange Data Mining workflow file, used to generate the co-author network in Figure 9 (incl. the used color scheme file authornet_colorscheme.colors).

In addition, the corpus of relevant papers with tags, publication metadata, and full text of open access work is available as a *Zotero* library at https://www.zotero.org/groups/integrationsonificationvisualization/items.

Acknowledgments

We would like to thank PerMagnus Lindborg, School of Creative Media, City University of Hong Kong, for valuable feedback. This research was funded in part by the Austrian Science Fund (FWF): 10.55776/P33531 and the Gesellschaft für Forschungsförderung Niederösterreich (GFF) SC20-006, as well as by the Knut and Alice Wallenberg Foundation (grant KAW 2019.0024).

References

[AG17] ANDREOPOULOU A., GOUDARZI V.: Reflections on the representation of women in the International Conferences on Auditory Displays (ICAD). In *Proceedings of the 23rd International Conference on Auditory Display (ICAD 2017)* (Pennsylvania State University, June 2017), Georgia Institute of Technology, pp. 43–48. doi:10.21785/icad2017.031. 4

[AG19] AUDRY E., GARCIA J.: Congruent audio-visual alarms for supervision tasks. In *Proceedings of the 25th International Conference on Auditory Display (ICAD 2019)* (Newcastle upon Tyne, June 2019), Department of Computer and Information Sciences, Northumbria University, pp. 7–11. doi:10.21785/icad2019.022. 20

[AG21] ANDREOPOULOU A., GOUDARZI V.: Sonification first: The role of icad in the advancement of sonification-related research. In *Proceedings of the 26th International Conference on Auditory Display* (June 2021), Georgia Institute of Technology, pp. 65–73. URL: http://hdl.handle.net/1853/66335. 4, 6

[AHvR*21] AXON L., HAPPA J., VAN RENSBURG A. J., GOLDSMITH M., CREESE S.: Sonification to support the monitoring tasks of security operations centres. *IEEE Transactions on Dependable and Secure Computing 18*, 3 (2021), 1227–1244. doi:10.1109/TDSC.2019.2931557. 20

[AJB*18] ARBON R. E., JONES A. J., BRATHOLM L. A., MITCHELL T., GLOWACKI D. R.: Sonifying stochastic walks on biomolecular energy landscapes. In *Proc. 24th International Conference on Auditory Display (ICAD 2018)* (Michigan Technological University, June 2018), Georgia Institute of Technology, pp. 232–239. doi:10.21785/icad2018.032. 7, 8

[AK11] ADHITYA S., KUUSKANKARE M.: The sonified urban masterplan (sum) tool: Sonification for urban planning and design. In *Proc. 17th International Conference on Auditory Display (ICAD-2011)* (Budapest, Hungary, June 2011). http://hdl.handle.net/1853/51918. 7, 8, 13, 14

[AQH*21] ANDRADE E., QUINLAN L., HARTE R., BYRNE D., FALLON E., KELLY M., CASEY S., KIRRANE F., O'CONNOR P., O'HORA D., SCULLY M., LAFFEY J., PLADYS P., BEUCHÉE A., ÓLAIGHIN G.: Augmenting critical care patient monitoring using wearable technology: Review of usability and human factors. *JMIR Human Factors 8*, 2 (2021), e16491. doi:10.2196/16491. 20

[ASH*12] Alonso-Arevalo M. A., Shelley S., Hermes D., Hollowood J., Pettitt M., Sharples S., Kohlrausch A.: Curve shape and curvature perception through interactive sonification. *ACM Trans. Appl. Percept. 9*, 4 (Oct. 2012). doi:10.1145/2355598.2355600. 7, 8, 14, 16, 22

[Bal15] Ballora M.: Two examples of sonification for viewer engagement: Hurricanes and squirrel hibernation cycles. In *Proc. 21st International Conference on Auditory Display (ICAD-2015)* (Graz, Austria, July 2015), Georgia Institute of Technology, pp. 300–301. http://hdl.handle.net/1853/54172. 7, 8, 13, 17

[Bar97] Barrass S.: *Auditory Information Design*. PhD thesis, Australian National University, Canberra, 1997. 3

[Bar12] Barrass S.: The aesthetic turn in sonification towards a social and cultural medium. *AI & Society 27*, 2 (May 2012), 177–181. doi:10.1007/s00146-011-0335-5. 3

[BB12] Bearman N., Brown E.: Who's sonifying data and how are they doing it? a comparison of ICAD and other venues since 2009. In *Proceedings of the 18th International Conference on Auditory Display* (2012), pp. 231–232. URL: http://hdl.handle.net/1853/44402. 6

[BB19] Berger M., Bill R.: Combining vr visualization and sonification for immersive exploration of urban noise standards. *Multimodal Technologies and Interaction 3*, 2 (May 2019), 34:1–34:15. doi:10.3390/MTI3020034. 7, 8, 14, 16, 17

[BBV16] Ballweg H., Bronowska A. K., Vickers P.: Interactive sonification for structural biology and structure-based drug design. In *Proceedings of ISon 2016, 5th Interactive Sonification Workshop, CITEC, Bielefeld University* (Bielefeld, Germany, 2016). 7, 8, 12, 15, 16, 22

[Bea11] Bearman N.: Using sound to represent uncertainty in future climate projections for the united kingdom. In *Proc. 17th International Conference on Auditory Display (ICAD-2011)* (Budapest, Hungary, June 2011). http://hdl.handle.net/1853/51922. 7, 8, 11, 16, 22

[Ber83] Bertin J.: *Semiology of Graphics: Diagrams Networks Maps*. University of Wisconsin, Madison, 1983. Originally published in 1967 in French. 3, 6, 12, 17, 21, 22

[BF12] Bearman N., Fisher P. F.: Using sound to represent spatial data in arcgis. *Computers & Geosciences 46* (Sept. 2012), 157–163. doi:10.1016/j.cageo.2011.12.001. 7, 8, 11, 13, 14, 22

[Bly82] Bly S. A.: *Sound and computer information presentation*. PhD thesis, University of California, Davis, 1982. 3

[BM13] Brehmer M., Munzner T.: A multi-level typology of abstract visualization tasks. *IEEE Trans. Visualization and Computer Graphics 19*, 12 (2013), 2376–2385. doi:10.1109/TVCG.2013.124. 3

[BM20] Bouchara T., Montès M.: Immersive sonification of protein surface. In *2020 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)* (Mar. 2020), pp. 380–383. doi:10.1109/VRW50115.2020.00082. 7, 8

[BMLA22] Buening R., Maeda T., Liew K., Aramaki E.: Between fact and fabrication: How visual art might nurture environmental consciousness. *Frontiers in Psychology 13* (2022). doi:10.3389/fpsyg.2022.925843. 21

[BPS06] Bertini E., Plaisant C., Santucci G. (Eds.): *Proceedings of the 2006 AVI workshop on BEyond time and errors: novel evaluation methods for information visualization*. ACM, 2006. doi:10.1145/1168149. 3

[BTB23] Bru E., Trautner T., Bruckner S.: Line Harp: Importance-driven sonification for dense line charts. In *Proceedings of IEEE Visualization and Visual Analytics (VIS) 2023 – Short Papers* (2023), pp. 186–190. arXiv:2307.16589, doi:10.1109/VIS54172.2023.00046. 7, 8, 13, 14, 15

[CB17] Chabot S., Braasch J.: An immersive virtual environment for congruent audio-visual spatialized data sonifications. In *Proc. 23rd International Conference on Auditory Display (ICAD–2017)* (Pennsylvania State University, June 2017), Georgia Institute of Technology, pp. 203–210. doi:10.21785/icad2017.072. 7, 8, 13, 15, 16, 17

[CGM19] Ceneda D., Gschwandtner T., Miksch S.: A review of guidance approaches in visual data analysis: A multifocal perspective. *Computer Graphics Forum 38*, 3 (2019), 861–879. doi:10.1111/cgf.13730. 19

[CH04] Collier W. G., Hubbard T. L.: Musical scales and brightness evaluations: Effects of pitch, direction, and scale mode. *Musicae Scientiae 8* (2004), 151–173. doi:10.1177/102986490400800203. 4

[CLR22] Caiola V., Lenzi S., Riccò D.: Audiovisual sonifications: A design map for multisensory integration in data representation. In *DRS2022: Bilbao* (2022). doi:10.21606/drs.2022.380. 1, 4, 11, 12

[CMS99] Card S. K., Mackinlay J. D., Shneiderman B. (Eds.): *Readings in Information Visualization: Using Vision to Think*. Morgan Kaufmann, San Francisco, 1999. 1, 3

[CWM21] Cantrell S. J., Walker B. N., Moseng Ø.: Highcharts sonification studio: An online, open-source, extensible, and accessible data sonification tool. In *Proc. 26th International Conference on Auditory Display (ICAD 2021)* (Virtual Conference, June 2021), Georgia Institute of Technology, pp. 210–216. doi:10.21785/icad2021.005. 7, 8, 11, 13, 14, 22

[DB13] Dubus G., Bresin R.: A Systematic Review of Mapping Strategies for the Sonification of Physical Quantities. *PLoS ONE 8*, 12 (2013), e82491. doi:10.1371/journal.pone.0082491. 4, 11, 12

[dC07] de Campo A.: Toward a data sonification design space map. In *Proceedings of the 13th International Conference on Auditory Display* (2007), pp. 342–347. URL: http://hdl.handle.net/1853/50042. 3

[DCM*18] Du M., Chou J.-K., Ma C., Chandrasegaran S., Ma K.-L.: Exploring the role of sound in augmenting visualization to enhance user engagement. In *2018 IEEE Pacific Visualization Symposium (PacificVis)* (Kobe, Apr. 2018), IEEE, pp. 225–229. doi:10.1109/PacificVis.2018.00036. 7, 8, 11, 13, 14, 22

[DLVDCG22] De La Vega G., Dominguez L. M. E., Casado J., García B.: Sonouno web: An innovative user centred web interface. In *HCI International 2022 – Late Breaking Posters* (Cham, 2022), Stephanidis C., Antona M., Ntoa S., Salvendy G., (Eds.), Communications in Computer and Information Science, Springer Nature Switzerland, pp. 628–633. doi:10.1007/978-3-031-19679-9_79. 7, 8, 11, 14, 22

[EBL*23] Elmquist E., Bock A., Lundberg J., Ynnerman A., Rönnberg N.: SonAir: the design of a sonification of radar data for air traffic control. *Journal on Multimodal User Interfaces 17*, 3 (2023), 137–149. doi:10.1007/s12193-023-00404-x. 20

[EEBR21] Elmquist E., Ejdbo M., Bock A., Rönnberg N.: Openspace sonification: Complementing visualization of the solar system with sound. In *Proc. 26th International Conference on Auditory Display (ICAD 2021)* (Virtual Conference, June 2021), International Community for Auditory Display, pp. 135–142. doi:10.21785/icad2021.018. 7, 8, 11, 13, 16, 22

[ERI*22] Enge K., Rind A., Iber M., Höldrich R., Aigner W.: Towards multimodal exploratory data analysis: Soniscope as a prototypical implementation. In *Proc. 24th Eurographics Conference on Visualization (EuroVis 2022) - Short Papers* (Rome, 2022), The Eurographics Association, pp. 67–71. doi:10.2312/evs.20221095. 7, 8, 11, 13, 14

[ERI*23] Enge K., Rind A., Iber M., Höldrich R., Aigner W.: Towards a unified terminology for sonification and visualization. *Journal on Personal and Ubiquitous Computing 27* (2023), 1949–1963. doi:10.1007/s00779-023-01720-5. 1, 3, 11, 22

[FB18] Ferguson J., Brewster S. A.: Investigating perceptual congruence between data and display dimensions in sonification. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (2018), ACM, p. 611:1–611:9. doi:10.1145/3173574.3174185. 4

[FBC12] Ferguson S., Beilharz K., Calò C. A.: Navigation of interactive sonifications and visualisations of time-series data using multi-touch computing. *Journal on Multimodal User Interfaces 5*, 3 (May

2012), 97–109. doi:10.1007/s12193-011-0075-3. 7, 8, 11, 13, 14, 16, 22

[FBT97] FLOWERS J. H., BUHMAN D. C., TURNAGE K. D.: Cross-modal equivalence of visual and auditory scatterplots for exploring bivariate data samples. *Human Factors 39* (1997), 341–351. doi:10.1518/001872097778827151. 4

[FBZ*99] FROHLICH B., BARRASS S., ZEHNER B., PLATE J., GOBEL M.: Exploring geo-scientific data in virtual environments. In *Proceedings Visualization '99 (Cat. No.99CB37067)* (1999), pp. 169–173. doi:10.1109/VISUAL.1999.809884. 4

[FE21] FORENSIC ARCHITECTURE, ENO B.: Digital violence, 2021. Webpage: https://sonification.design/#DigitalViolence. Last accessed: Dec 21st 2023. 21

[FFSR*23] FAN D., FAY SIU A., RAO H., KIM G. S.-H., VAZQUEZ X., GRECO L., O'MODHRAIN S., FOLLMER S.: The accessibility of data visualizations on the web for screen reader users: Practices and experiences during covid-19. *ACM Transactions on Accessible Computing 16*, 1 (2023), 1–29. 19

[FN18] FITZPATRICK J., NEFF F.: Stream segregation: Utilizing harmonic variance in auditory graphs. In *Proc. 15th Sound and Music Computing Conference (SMC2018)* (Limassol, Cyprus, 2018), Zenodo, pp. 52–59. doi:10.5281/zenodo.1422501. 7, 8, 11, 22

[Foo15] FOO B.: Too Blue, 2015. Webpage: https://sonification.design/#TooBlue. Last accessed: Dec 21st 2023. 21

[FPS*21] FRANCONERI S. L., PADILLA L. M., SHAH P., ZACKS J. M., HULLMAN J.: The science of visual data communication: What works. *Psychological Science in the public interest 22*, 3 (2021), 110–161. 2

[Fri08] FRIENDLY M.: A Brief History of Data Visualization. In *Handbook of Data Visualization*, Chen C.-h., Härdle W., Unwin A., (Eds.), Springer Handbooks Comp.Statistics. Springer, Berlin, Heidelberg, 2008, pp. 15–56. doi:10.1007/978-3-540-33037-0_2. 3

[Fry05] FRYSINGER S. P.: A brief history of auditory data representation to the 1980s. In *Proceedings of the International Conference on Auditory Display* (2005). URL: http://hdl.handle.net/1853/50089. 3

[GDAS*18] GUNE A., DE AMICIS R., SIMÕES B., SANCHEZ C. A., DEMIREL H. O.: Graphically hearing: Enhancing understanding of geospatial data through an integrated auditory and visual experience. *IEEE Computer Graphics and Applications 38*, 4 (July 2018), 18–26. doi:10.1109/MCG.2018.042731655. 7, 8, 13, 22

[GGB05] GUTTMAN S. E., GILROY L. A., BLAKE R.: Hearing what the eyes see: auditory encoding of visual temporal sequences. *Psychological science 16*, 3 (2005), 228–235. 4, 22

[GH12] GROND F., HERMANN T.: Singing function: Exploring auditory graphs with a vowel based sonification. *Journal on Multimodal User Interfaces 5*, 3-4 (May 2012), 87–95. doi:10.1007/s12193-011-0068-2. 3

[GHJ*19] GROSS-VOGT K., HERMANN T., JURY M. W., STEINER A. K., KARTADINATA S.: KlimaIAnlage—performing climate data. In *Addressing the Challenges in Communicating Climate Change Across Various Audiences*, Leal Filho W., Lackner B., McGhie H., (Eds.). Springer, Cham, 2019, pp. 339–355. doi:10.1007/978-3-319-98294-6_21. 21

[GKW21] GROPPE S., KLINCKENBERG R., WARNKE B.: Sound of databases: Sonification of a semantic web database engine. *Proc. VLDB Endow. 14*, 12 (July 2021), 2695–2698. doi:10.14778/3476311.3476322. 7, 8, 14

[Gou17] GOUDARZI V.: *Systematic Procedure to Develop Sonifications*. PhD thesis, University of Music and Performing Arts Graz, 2017. 17

[GR11] GOMEZ I., RAMIREZ R.: A data sonification approach to cognitive state identification. In *Proc. 17th International Conference on Auditory Display (ICAD-2011)* (Budapest, Hungary, June 2011). http://hdl.handle.net/1853/51569. 7, 8, 13

[GRK*16] GIONFRIDA L., ROGINSKA A., KEARY J., MOHANRAJ H., FRIEDMAN K. P.: The triple tone sonification method to enhance the diagnosis of alzheimer's dementia. In *Proc. 22nd International Conference on Auditory Display (ICAD–2016)* (Canberra, Australia, July 2016). doi:10.21785/icad2016.023. 7, 8, 14, 16, 22

[GS90] GRINSTEIN G. G., SMITH S.: The perceptualization of scientific data. In *Extracting Meaning from Complex Data: Processing, Display, Interaction* (1990), vol. 1259, SPIE, pp. 190–199. doi:10.1117/12.19986. 4

[GVEZ23] GROSS-VOGT K., ENGE K., ZMÖLNIG I.: Reflecting on qualitative and quantitative data to frame criteria for effective sonification design. In *Proceedings of the 18th International Audio Mostly Conference* (New York, NY, USA, 2023), AM '23, Association for Computing Machinery, p. 93–100. doi:10.1145/3616195.3616233. 6

[HAR16] HILDEBRANDT T., AMERBAUER F., RINDERLE-MA S.: Combining sonification and visualization for the analysis of process execution data. In *2016 IEEE 18th Conference on Business Informatics (CBI)* (Aug. 2016), vol. 02, pp. 32–37. doi:10.1109/CBI.2016.47. 7, 8, 12, 13, 14, 17

[HAS11] HULLMAN J., ADAR E., SHAH P.: Benefitting InfoVis with visual difficulties. *IEEE Transactions on Visualization and Computer Graphics 17*, 12 (Dec. 2011), 2213–2222. doi:10.1109/TVCG.2011.175. 22

[HCTP14] HOLTZMAN B., CANDLER J., TURK M., PETER D.: Seismic sound lab: Sights, sounds and perception of the earth as an acoustic space. In *Sound, Music, and Motion* (Cham, 2014), Aramaki M., Derrien O., Kronland-Martinet R., Ystad S., (Eds.), Lecture Notes in Computer Science, Springer International Publishing, pp. 161–174. doi:10.1007/978-3-319-12976-1_10. 7, 8, 12

[Her02] HERMANN T.: *Sonification for Exploratory Data Analysis*. PhD thesis, Bielefeld University, Bielefeld, Germany, 02 2002. 10, 16

[Her08] HERMANN T.: Taxonomy and definitions for sonification and auditory display. In *Proceedings of the 14th International Conference on Auditory Display* (2008). URL: http://hdl.handle.net/1853/49960. 3

[Her20] HERRMANN V.: Visualizing and sonifying how an artificial ear hears music. In *Proceedings of the NeurIPS 2019 Competition and Demonstration Track* (Dec. 2020), Escalante H. J., Hadsell R., (Eds.), vol. 123 of *Proceedings of Machine Learning Research*, PMLR, pp. 192–202. https://proceedings.mlr.press/v123/herrmann20a.html. 7, 8, 14

[HEUHEB23] HOQUE M. N., EHTESHAM-UL-HAQUE M., ELMQVIST N., BILLAH S. M.: Accessible data representation with natural sound. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (2023), pp. 826:1–826:19. doi:10.1145/3544548.3581087. 19

[HGI*22] HOLLOWAY L. M., GONCU C., ILSAR A., BUTLER M., MARRIOTT K.: Infosonics: Accessible infographics for people who are blind using sonification and voice. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems* (2022), pp. 480:1–480:13. doi:10.1145/3491102.3517465. 19

[HH04] HUNT A., HERMANN T.: The importance of interaction in sonification. In *Proc. of the 10th Meeting of the International Conference on Auditory Display (ICAD 2004)* (Sydney, Australia, 2004), pp. 1–8. 19

[HHG*12] HÖFERLIN B., HÖFERLIN M., GOLOUBETS B., HEIDEMANN G., WEISKOPF D.: Auditory support for situation awareness in video surveillance. In *Proceedings of the 18th International Conference on Auditory Display,* (Atlanta, GA, USA,, June 2012), Georgia Institute of Technology, pp. 156–163. URL: http://hdl.handle.net/1853/44426. 20

[HHN11] HERMANN T., HUNT A., NEUHOFF J. G. (Eds.): *The Sonification Handbook*. Logos, Bielefeld, 2011. 2, 3, 19

[HHRM16] HILDEBRANDT T., HERMANN T., RINDERLE-MA S.: Continuous sonification enhances adequacy of interactions in peripheral process monitoring. *International Journal of Human-Computer Studies 95* (2016), 54–65. doi:10.1016/j.ijhcs.2016.06.002. 20

[HK22] HAN Y. C., KHANDUJA A.: The future is red: Visualizing wildfire predictions using contactless interaction. In *Extended Abstracts of the 2022 CHI Conference on Human Factors in Computing Systems* (New York, NY, USA, 2022), CHI EA '22, Association for Computing Machinery. doi:10.1145/3491101.3519903. 7, 8, 10, 13, 16, 17, 22

[HPDW23] HUPPENKOTHEN D., PAMPIN J., DAVENPORT J. R. A., WENLOCK J.: The sonified hertzsprung-russell diagram. In *Proc. 28th International Conference on Auditory Display (ICAD 2023)* (Norrköping, Sweden, June 2023), Georgia Institute of Technology, pp. 272–279. doi:10.21785/icad2023.6263. 7, 8, 10, 12, 16, 17

[HR21] HERMANN T., REINSCH D.: Sc3nb: A Python-SuperCollider Interface for Auditory Data Science. In *Proceedings of the 16th International Audio Mostly Conference* (New York, NY, USA, Oct. 2021), AM '21, Association for Computing Machinery, pp. 208–215. doi:10.1145/3478384.3478401. 22

[HRM15] HILDEBRANDT T., RINDERLE-MA S.: Server sounds and network noises. In *Cognitive Infocommunications (CogInfoCom), 2015 6th IEEE International Conference on* (2015), IEEE, pp. 45–50. 5, 20

[HTHB22] HARRISON C., TRAYFORD J., HARRISON L., BONNE N.: Audio universe: Tour of the solar system. *Astronomy & Geophysics 63*, 2 (Apr. 2022), 2.38–2.40. doi:10.1093/astrogeo/atac027. 7, 8, 16, 17

[Hun07] HUNTER J. D.: Matplotlib: A 2d graphics environment. *Computing in Science & Engineering 9*, 3 (2007), 90–95. doi:10.1109/MCSE.2007.55. 22

[HWKK23] HUANG Z., WITSCHARD D., KUCHER K., KERREN A.: VA + Embeddings STAR: A State-of-the-Art Report on the Use of Embeddings in Visual Analytics. *Computer Graphics Forum 42*, 3 (2023), 539–571. doi:10.1111/cgf.14859. 19

[Ibe20] IBER M.: Auditory display in workplace environments. In *Foundations in Sound Design for Embedded Media: A Multidisciplinary Approach*, Filimowicz M., (Ed.). Routledge, New York, 2020, pp. 131–154. 4, 20

[IIC*13] ISENBERG T., ISENBERG P., CHEN J., SEDLMAIR M., MÖLLER T.: A systematic review on the practice of evaluating visualization. *IEEE Trans. Visualization and Computer Graphics 19*, 12 (2013), 2818–2827. doi:10.1109/TVCG.2013.126. 15

[JBKS15] JAIN A., BANSAL R., KUMAR A., SINGH K.: A comparative study of visual and auditory reaction times on the basis of gender and physical activity levels of medical first year students. *International Journal of Applied and Basic Medical Research 5* (2015), 124. doi:10.4103/2229-516X.157168. 4

[JDS15] JAKUS G., DICKE C., SODNIK J.: A user study of auditory, head-up and multi-modal displays in vehicles. *Applied Ergonomics 46* (2015), 184–192. doi:10.1016/j.apergo.2014.08.008. 20

[JLSP15] JEON M., LEE J.-H., STERKENBURG J., PLUMMER C.: *Cultural Differences in Preference of Auditory Emoticons: USA and South Korea*. Georgia Institute of Technology, July 2015. 2

[JMP13] JOLIAT N., MAYTON B., PARADISO J. A.: Spatialized anonymous audio for browsing sensor networks via virtual worlds. In *Proc. 19th International Conference on Auditory Display (ICAD2013)* (Lodz, Poland, July 2013), Georgia Institute of Technology, pp. 67–75. http://hdl.handle.net/1853/51643. 7, 8, 16

[Joh04] JOHANNSEN G.: Auditory displays in human-machine interfaces. *Proceedings of the IEEE 92*, 4 (2004), 742–758. doi:10.1109/JPROC.2004.825905. 20

[KAV21] KARIYADO Y., AREVALO C., VILLEGAS J.: Auralization of three-dimensional cellular automata. In *Artificial Intelligence in Music, Sound, Art and Design* (Cham, 2021), Romero J., Martins T., Rodríguez-Fernández N., (Eds.), Lecture Notes in Computer Science, Springer International Publishing, pp. 161–170. doi:10.1007/978-3-030-72914-1_11. 7, 8, 15, 22

[KBBG07] KASAKEVICH M., BOULANGER P., BISCHOF W. F., GARCIA M.: Augmentation of visualisation using sonification: A case study in computational fluid dynamics. In *Proc. IPT-EGVE Symposium* (Germany, Europe, 2007), The Eurographics Association, pp. 89–94. doi:10.2312/PE/VE2007Short/089-094. 4

[KGS*20] KIM C., GUO A., SALHOTRA G., SPRINKHUIZEN S., SHETTY K., KONG D. S.: Sonifying data from the human microbiota: Biota beats. *Computer Music Journal 44*, 1 (Mar. 2020), 51–70. doi:10.1162/comj_a_00552. 21

[KIK19] KANEV I. K., ILIEV I. T., KRASTEVA V. T.: Sonification – an alternative presentation of the electrocardiogram: A systematic literature review. In *2019 IEEE XXVIII International Scientific Conference Electronics (ET)* (2019), IEEE, pp. 1–4. doi:10.1109/ET.2019.8878650. 5, 20

[KJRK21] KIM N. W., JOYNER S. C., RIEGELHUTH A., KIM Y.-S.: Accessible visualization: Design space, opportunities, and challenges. *Computer Graphics Forum 40*, 3 (2021), 173–188. doi:10.1111/cgf.14298. 19

[KLTW17] KONDAK Z., LIANG T. A., TOMLINSON B., WALKER B. N.: Web sonification sandbox - an easy-to-use web application for sonifying data and equations. In *Proc. Web Audio Conference WAC-2017* (London, UK, Aug. 2017). https://qmro.qmul.ac.uk/xmlui/handle/123456789/26083. 7, 8, 11, 14, 22

[KOR23] KIESENHOFER S., OELSCH A., RAMMER D.: Deep sync, 2023. Webpage: https://ars.electronica.art/futurelab/en/projects-deep-sync/. Last accessed: Dec 28st 2023. 21

[Kra94] KRAMER G. (Ed.): *Auditory Display: Sonification, Audification and Auditory Interfaces. Proceedings of the First International Conference on Auditory Display (ICAD) 1992*. Addison-Wesley, Reading, Mass, 1994. 3

[KWB*99] KRAMER G., WALKER B., BONEBRIGHT T., COOK P., FLOWERS J. H., MINER N., NEUHOFF J., ET AL.: *Sonification Report: Status of the Field and Research Agenda*. Report for the NSF, International Community for Auditory Display, 1999. URL: http://www.icad.org/websiteV2.0/References/nsf.html. 1, 3

[LCS*28] LIN J.-W., CHEN W., SHEN C.-P., CHIU M.-J., KAO Y.-H., LAI F., ZHAO Q., CICHOCKI A.: Visualization and sonification of long-term epilepsy electroencephalogram monitoring. *Journal of Medical and Biological Engineering 38*, 6 (2028), 943–952. doi:10.1007/s40846-017-0358-6. 20

[LDIC21] LEE B., DACHSELT R., ISENBERG P., CHOE E. K. (Eds.): *Mobile Data Visualization*. Chapman and Hall/CRC, New York, Dec. 2021. doi:10.1201/9781003090823. 21

[Len21] LENZI S.: *The Design of Data Sonification. Design Processes, Protocols and Tools Grounded in Anomaly Detection*. PhD thesis, Politecnico di Milano, 2021. https://www.politesi.polimi.it/handle/10589/177079. 3

[LF21] LINDETORP H., FALKENBERG K.: Sonification for everyone everywhere: Evaluating the webaudioxml sonification toolkit for browsers. In *Proc. 26th International Conference on Auditory Display (ICAD 2021)* (Virtual Conference, June 2021), Georgia Institute of Technology, pp. 15–21. http://hdl.handle.net/1853/66351. 7, 8, 11, 12, 14, 22

[LLC23] LINDBORG P., LENZI S., CHEN M.: Climate data sonification and visualization: An analysis of topics, aesthetics, and characteristics in 32 recent projects. *Frontiers in Psychology 13* (2023). doi:10.3389/fpsyg.2022.1020102. 20

[LLW21] LYU Z., LI J., WANG B.: Aiive: Interactive visualization and sonification of neural networks in virtual reality. In *2021 IEEE International Conference on Artificial Intelligence and Virtual Reality (AIVR)* (Nov. 2021), pp. 251–255. doi:10.1109/AIVR52153.2021.00057. 7, 8, 10

[LSB*23] LEMMON E., SCHEDEL M., BILKHU I., ZHU H., ESCOBAR L., AUMOITHE G.: Mapping in the emergency: Designing a hyperlocal and socially conscious sonified map of Covid-19 in Suffolk County, New York. In *Proceedings of ISon 2022, 7th Interactive Sonification Workshop,* *BSCC, University of Bremen* (Jan. 2023), Zenodo. doi:10.5281/ZENODO.7552257. 7, 8, 10, 14, 16

[LWP99] LOTTO R. B., WILLIAMS S. M., PURVES D.: Mach bands as empirically derived associations. *Proceedings of the National Academy of Sciences 96*, 9 (1999), 5245–5250. 4

[LWP*13] LU S. A., WICKENS C. D., PRINET J. C., HUTCHINS S. D., SARTER N., SEBOK A.: Supporting interruption management and multimodal interface design: Three meta-analyses of task performance as a function of interrupting task modality. *Human Factors 55*, 4 (2013), 697–724. doi:10.1177/0018720813476298. 20

[LWS96] LODHA S. K., WILSON C. M., SHEEHAN R. E.: Listen: Sounding uncertainty visualization. In *Visualization '96. Proceedings.* (San Francisco, CA, USA, 1996), ACM, pp. 189–195. doi:10.1109/VISUAL.1996.568105. 4

[MAFP19a] MALIKOVA E., ADZHIEV V., FRYAZINOV O., PASKO A.: Visual-auditory representation and analysis of molecular scalar fields. In *Proc. EUROGRAPHICS 2019* (2019), pp. 25–26. doi:10.2312/egp.20191051. 17

[MAFP19b] MALIKOVA E., ADZHIEV V., FRYAZINOV O., PASKO A.: Visual-auditory volume rendering of scalar fields. In *Proc. 25th International Conference on Auditory Display (ICAD 2019)* (Northumbria University, June 2019), Georgia Institute of Technology, pp. 147–154. doi:10.21785/icad2019.004. 6, 7, 8, 13

[May14] MAYER R. E.: The cambridge handbook of multimedia learning. *The Cambridge Handbook of Multimedia Learning, Second Edition* (1 2014), 1–930. doi:10.1017/CBO9781139547369. 4

[MBKSM16] METATLA O., BRYAN-KINNS N., STOCKMAN T., MARTIN F.: Sonification of reference markers for auditory graphs: Effects on non-visual point estimation tasks. *PeerJ Computer Science 2* (2016), e51. 22

[MF95] MINGHIM R., FORREST A.: An illustrated analysis of sonification for scientific visualisation. In *Proceedings Visualization '95* (1995), pp. 110–117. doi:10.1109/VISUAL.1995.480802. 3, 4

[ML17] MCNABB L., LARAMEE R. S.: Survey of Surveys (SoS) - Mapping The Landscape of Survey Papers in Information Visualization. *Computer Graphics Forum 36*, 3 (2017), 589–617. doi:10.1111/cgf.13212. 4

[MLS95] MOUSAVI S. Y., LOW R., SWELLER J.: Reducing cognitive load by mixing auditory and visual presentation modes. *Journal of educational psychology 87*, 2 (1995), 319. 19

[MMM18] MAÇÃS C., MARTINS P., MACHADO P.: Consumption as a rhythm: A multimodal experiment on the representation of time-series. In *2018 22nd International Conference Information Visualisation (IV)* (July 2018), pp. 504–509. doi:10.1109/iV.2018.00093. 7, 8, 13, 14

[MMU16] MATSUBARA M., MORIMOTO Y., UCHIDE T.: Collaborative study of interactive seismic array sonification for data exploration and public outreach activities. In *Proceedings of ISon 2016, 5th Interactive Sonification Workshop, CITEC, Bielefeld University* (Bielefeld, Germany, 2016). 7, 8, 12, 14, 16

[MNS16] MACULEWICZ J., NILSSON N. C., SERAFIN S.: An investigation of the effect of immersive visual and auditory feedback on rhythmic walking interaction. In *Proceedings of the Audio Mostly 2016* (2016), AM '16, Association for Computing Machinery, pp. 194–201. doi:10.1145/2986416.2986429. 20

[MNW*18] MACDONALD D. E., NATARAJAN T., WINDEYER R. C., COPPIN P., STEINMAN D. A.: Data-driven sonification of cfd aneurysm models. In *Proc. 24th International Conference on Auditory Display (ICAD 2018)* (Houghton, Michigan, June 2018), The International Community for Auditory Display, pp. 28–33. doi:10.21785/icad2018.010. 7, 8, 15, 16, 22

[Mun15] MUNZNER T.: *Visualization Analysis and Design*. CRC Press, Boca Raton, Jan. 2015. 1, 6, 10, 11, 12, 13, 14, 17, 22

[MVC*10] MASUD L., VALSECCHI F., CIUCCARELLI P., RICCI D., CAVIGLIA G.: From data to knowledge – visualizations as transformation processes within the data-information-knowledge continuum. In *2010 14th International Conference Information Visualisation* (July 2010), pp. 445–449. doi:10.1109/IV.2010.68. 21

[NB02] NESBITT K. V., BARRASS S.: Evaluation of a multimodal sonification and visualisation of depth of market stock data. In *Proceedings of the 8th International Conference on Auditory Display* (Kyoto, 2002). URL: http://hdl.handle.net/1853/51355. 4

[Nee19] NEES M. A.: Eight components of a design theory of sonification. In *Proceedings of the 25th International Conference on Auditory Display (ICAD 2019)* (2019), pp. 176–183. doi:10.21785/icad2019.048. 3

[Neu04] NEUHOFF J. G. (Ed.): *Ecological Psychoacoustics*. Elsevier Academic Press, San Diego, 2004. 3, 22

[Neu19] NEUHOFF J. G.: Is sonification doomed to fail. In *Proceedings of the 25th International Conference on Auditory Display (ICAD 2019)* (2019), pp. 327–330. 3

[NGJW14] NEES M. A., GABLE T. M., JEON M., WALKER B. N.: Prototype auditory displays for a fuel efficiency driver interface. In *Proceedings of the 20th International Conference on Auditory Display* (New York, USA, June 2014), Georgia Institute of Technology. Publisher: Georgia Institute of Technology. URL: http://hdl.handle.net/1853/52089. 20

[NRL*12] NESS S., REIMER P., LOVE J., SCHLOSS W. A., TZANETAKIS G.: Sonophenology. *Journal on Multimodal User Interfaces 5*, 3 (May 2012), 123–129. doi:10.1007/s12193-011-0066-4. 7, 8, 14, 16

[NSC16] NORTH K. J., SARMA A., COHEN M. B.: Understanding git history: A multi-sense view. In *Proceedings of the 8th International Workshop on Social Software Engineering* (New York, NY, USA, 2016), SSE 2016, Association for Computing Machinery, pp. 1–7. doi:10.1145/2993283.2993285. 7, 8, 12, 14, 17, 22

[NW08] NEES M. A., WALKER B. N.: Encoding and representation of information in auditory graphs: Descriptive reports of listener strategies for understanding data. In *Proceedings of the International Conference on Auditory Display* (Paris, FR, June 2008). URL: http://hdl.handle.net/1853/49907. 3

[PBM15] PAPACHRISTODOULOU P., BETELLA A., MANZOLLI J.: Augmenting the navigation of complex data sets using sonification: A case study with brainx3. In *2015 IEEE 2nd VR Workshop on Sonic Interactions for Virtual Environments (SIVE)* (Mar. 2015), pp. 1–6. doi:10.1109/SIVE.2015.7361284. 7, 8, 12, 16, 22

[PBV14] PAPACHRISTODOULOU P., BETELLA A., VERSCHURE P.: Sonification of large datasets in a 3d immersive environment: A neuroscience case study. In *Proc. ACHI 2014 : The Seventh International Conference on Advances in Computer-Human Interactions* (Jan. 2014), IARIA, pp. 35–40. 7, 8, 22

[PC19] PHILLIPS S., CABRERA A.: Sonification workstation. In *25th International Conference on Auditory Display (ICAD 2019)* (Northumbria University, June 2019), Georgia Institute of Technology, pp. 184–190. http://hdl.handle.net/1853/61529. 7, 8, 11, 12, 14, 22

[PCB23] PENG T., CHOI H., BERGER J.: Siren: Creative and extensible sonification on the web. In *Proc. 28th International Conference on Auditory Display (ICAD 2023)* (Norrköping, Sweden, June 2023), Georgia Institute of Technology, pp. 78–84. https://hdl.handle.net/1853/72879. 7, 8, 11, 14, 22

[PFH*22] PATÉ A., FARGE G., HOLTZMAN B. K., BARTH A. C., POLI P., BOSCHI L., KARLSTROM L.: Combining audio and visual displays to highlight temporal and spatial seismic patterns. *Journal on Multimodal User Interfaces 16*, 1 (Mar. 2022), 125–142. doi:10.1007/s12193-021-00378-8. 7, 8, 10, 12, 14, 16, 22

[Pol06] POLLI A.: Heat and the Heartbeat of the City: Sonifying Data Describing Climate Change. *Leonardo Music Journal 16* (Dec. 2006), 44–45. doi:10.1162/lmj.2006.16.44. 21

[RC15] ROSLI M. H. W., CABRERA A.: Gestalt principles in multimodal data representation. *IEEE Computer Graphics & Applications 32* (2015), 80–87. doi:10.1109/MCG.2015.29. 4

[RCL*98] ROBERTSON G., CZERWINSKI M., LARSON K., ROBBINS D. C., THIEL D., VAN DANTZICH M.: Data mountain: Using spatial memory for document management. In *Proceedings of the 11th Annual ACM Symposium on User Interface Software and Technology* (New York, NY, USA, 1998), UIST '98, ACM, pp. 153–162. doi:10.1145/288392.288596. 4

[RF22] RÖNNBERG N., FORSELL C.: Questionnaires assessing usability of audio-visual representations. In *Proc. AVI 2022 Workshop on Audio-Visual Analytics* (Rome, 2022). doi:10.5281/zenodo.6555676. 15, 21

[RFK*15] RAU B., FRIESS F., KRONE M., MULLER C., ERTL T.: Enhancing visualization of molecular simulations using sonification. In *2015 IEEE 1st International Workshop on Virtual and Augmented Reality for Molecular Science (VARMS@IEEEVR)* (Mar. 2015), pp. 25–30. doi:10.1109/VARMS.2015.7151725. 7, 8, 12, 13, 15

[RFM13] ROGIŃSKA A., FRIEDMAN K., MOHANRAJ H.: Exploring sonification for augmenting brain scan data. In *Proc. 19th International Conference on Auditory Display (ICAD2013)* (Lodz, Poland, July 2013), Georgia Institute of Technology, pp. 95–101. http://hdl.handle.net/1853/51653. 7, 8

[RG71] RUBINSTEIN L., GRUENBERG E. M.: Intramodal and crossmodal sensory transfer of visual and auditory temporal patterns. *Perception & Psychophysics 9* (9 1971), 385–390. doi:10.3758/BF03210235. 4

[RH22] REINSCH D., HERMANN T.: Interacting with sonifications: The mesonic framework for interactive auditory data science. In *Proceedings of the 7th Interactive Sonification Workshop (ISon)* (2022), CITEC, Bielefeld University, p. 65. 22

[RH23] REINSCH D., HERMANN T.: Sonecules: A Python sonification architecture. In *Proceedings of the 28th International Conference on Auditory Display* (Norrköping, Sweden, June 2023), Georgia Institute of Technology, pp. 62–69. doi:10.21785/icad2023.5580. 22

[Rib18] RIBER A. G.: Planethesizer: Approaching exoplanet sonification. In *Proc. 24th International Conference on AuditoryDisplay (ICAD 2018)* (Michigan Technological University, June 2018), Georgia Institute of Technology, pp. 219–226. http://hdl.handle.net/1853/60073. 7, 8, 14

[Rib19] RIBER A. G.: Sonifigrapher: Sonified light curve synthesizer. In *Proc. 25th International Conference on Auditory Display (ICAD 2019)* (Northumbria University, June 2019), Georgia Institute of Technology, pp. 62–66. http://hdl.handle.net/1853/61497. 7, 8, 11, 12, 13, 14, 16

[RJ16] RÖNNBERG N., JOHANSSON J.: Interactive sonification for visual dense data displays. In *Proc. 5th Interactive Sonification Workshop, ISon* (2016), CITEC, Bielefeld University, pp. 63–67. 7, 8, 13, 14, 15, 22

[RMS*22] ROCHE T. R., MAAS E. J. C., SAID S., BRAUN J., MACHADO C., SPAHN D. R., NOETHIGER C. B., TSCHOLL D. W.: Anesthesia personnel's visual attention regarding patient monitoring in simulated non-critical and critical situations, an eye-tracking study. *BMC Anesthesiology 22*, 1 (2022), 167:1–167:10. doi:10.1186/s12871-022-01705-6. 20

[RN96] ROSSITER D., NG W.-Y.: A system for the complementary visualization of 3d volume images using 2d and 3d binaurally processed sonification representations. In *Proceedings of Seventh Annual IEEE Visualization '96* (1996), pp. 351–354. doi:10.1109/VISUAL.1996.568129. 4

[Rön19a] RÖNNBERG N.: Musical sonification supports visual discrimination of color intensity. *Behaviour & Information Technology 38*, 10 (2019), 1028–1037. doi:10.1080/0144929X.2019.1657952. 19, 22

[Rön19b] RÖNNBERG N.: Sonification supports perception of brightness contrast. *Journal on Multimodal User Interfaces 13*, 4 (2019), 373–381. doi:10.1007/s12193-019-00311-0. 19, 22

[Rön21] RÖNNBERG N.: Sonification for conveying data and emotion. In *Proceedings of the 16th International Audio Mostly Conference* (New York, NY, USA, 2021), AM '21, Association for Computing Machinery, pp. 56–63. doi:10.1145/3478384.3478387. 7, 8, 10, 14, 22

[RS22] RUSSO M., SANTAGUIDA A.: 5000 exoplanets: Listen to the sounds of discovery. In *Proc. 27th International Conference on Auditory Display (ICAD 2022)* (Virtual Conference, June 2022), Georgia Institute of Technology, pp. 64–68. http://hdl.handle.net/1853/67384. 7, 8, 10, 13, 14, 16, 17, 22

[RW10] ROBERTS J. C., WALKER R.: Using all our senses: The need for a unified theoretical approach to multi-sensory information visualization. In *Workshop on the Role of Theory in Information Visualization* (2010). 3

[RYTW23] RUBAB S., YU L., TANG J., WU Y.: Exploring Effective Relationships Between Visual-Audio Channels in Data Visualization. *Journal of Visualization 26*, 4 (Aug. 2023), 937–956. doi:10.1007/s12650-023-00909-3. 4

[SAR22] SVORONOS-KANAVAS I., AGIOMYRGIANAKIS V., RÖNNBERG N.: An exploratory use of audiovisual displays on oceanographic data. In *Proc. AVI 2022 Workshop on Audio-Visual Analytics (WAVA22)* (Frascati, Italy, 2022), Zenodo. doi:10.5281/zenodo.6555839. 7, 8, 13

[SBG90] SMITH S., BERGERON R. D., GRINSTEIN G. G.: Stereophonic and surface sound generation for exploratory data analysis. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (New York, NY, USA, 1990), CHI '90, Association for Computing Machinery, pp. 125–132. doi:10.1145/97243.97264. 4

[SDS23] SPENCE C., DI STEFANO N.: Sensory translation between audition and vision. *Psychonomic Bulletin & Review* (Oct. 2023). doi:10.3758/s13423-023-02343-w. 4

[SFLD22] SIRKKA A., FAGERLÖNN J., LINDBERG S., DELSING K.: The design of an auditory alarm concept for a paper mill control room. In *Advances in Ergonomics In Design, Usability & Special Populations: Part III* (2022), vol. 20, AHFE Open Acces. ISSN: 27710718 Issue: 20. doi:10.54941/ahfe1001301. 20

[SGS*18] SCHLOSS K. B., GRAMAZIO C. C., SILVERMAN A. T., PARKER M. L., WANG A. S.: Mapping color to meaning in colormap data visualizations. *IEEE transactions on visualization and computer graphics 25*, 1 (2018), 810–819. 4

[Shn96] SHNEIDERMAN B.: The eyes have it: A task by data type taxonomy for information visualizations. In *Proceedings of the IEEE Symposium on Visual Languages, VL* (1996), pp. 336–343. doi:10.1109/VL.1996.545307. 3

[SJMT19] SCHAFFERT N., JANZEN T. B., MATTES K., THAUT M. H.: A review on the relationship between sound and movement in sports and rehabilitation. *Frontiers in Psychology 10* (2019). doi:10.3389/fpsyg.2019.00244. 5, 20

[SKS06] SEITZ A. R., KIM R., SHAMS L.: Sound facilitates visual learning. *Current biology : CB 16* (7 2006), 1422–1427. doi:10.1016/J.CUB.2006.05.048. 4

[SMM12] SEDLMAIR M., MEYER M., MUNZNER T.: Design study methodology: Reflections from the trenches and the stacks. *IEEE Trans. Visualization and Computer Graphics 18*, 12 (2012), 2431–2440. doi:10.1109/TVCG.2012.213. 17

[SMS21] SCHLOSSER P. D., MATTHEWS B., SANDERSON P. M.: Head-worn displays for healthcare and industry workers: A review of applications and design. *International Journal of Human-Computer Studies 154* (2021), 102628. doi:10.1016/j.ijhcs.2021.102628. 20

[SNH05] STOCKMAN T., NICKERSON L. V., HIND G.: Auditory graphs: A summary of current experience and towards a research agenda. In *International Conference on Auditory Display (ICAD2005), Limerick, Ireland* (2005). 19

[SNHS13] SCHULZ H.-J., NOCKE T., HEITZLER M., SCHUMANN H.: A design space of visualization tasks. *IEEE Trans. Visualization and Computer Graphics 19*, 12 (Dec. 2013), 2366–2375. doi:10.1109/TVCG.2013.120. 3

[SPA23] STOIBER C., POHL M., AIGNER W.: Design actions for the design of visualization onboarding methods. In *EduVis Workshop, IEEE VIS* (Melbourne, Australia, 2023). doi:10.31219/osf.io/wjp5x. 23

[Spe07] SPENCE R.: *Information Visualization: Design for Interaction*, 2nd ed. Prentice Hall, Upper Saddle River, NJ, 2007. 3

[Spe11] SPENCE C.: Crossmodal correspondences: A tutorial review. *Attention, Perception, & Psychophysics 2011 73:4 73* (1 2011), 971–995. doi:10.3758/S13414-010-0073-7. 4

[SQS*20] SU I., QIN Z., SARACENO T., BISSHOP A., MÜHLETHALER R., ZIPORYN E., BUEHLER M. J.: Sonification of a 3-D spider web and reconstitution for musical composition using granular synthesis. *Computer Music Journal 44*, 4 (Dec. 2020), 43–59. doi:10.1162/comj_a_00580. 21

[SŠ18] SEBJANIČ R., ŠUTIĆ G.: Aqua_forensic, 2018. Webpage: https://ars.electronica.art/aeblog/en/2018/11/07/aquaforensic/. Last accessed: Dec 21st 2023. 21

[Sup12] SUPPER A.: *Lobbying for the Ear : The Public Fascination with and Academic Legitimacy of the Sonification of Scientific Data*. Doctoral thesis, Maastricht University, 2012. doi:10.26481/dis.20120606as. 3

[SW13] SCHUETT J. H., WALKER B. N.: Measuring comprehension in sonification tasks that have multiple data streams. In *Proc. of the Audio Mostly Conference* (2013), pp. 11:1–11:6. doi:10.1145/2544114.2544121. 4

[SWG*23] STOIBER C., WAGNER M., GRASSINGER F., POHL M., STITZ H., STREIT M., POTZMANN B., AIGNER W.: Visualization onboarding grounded in educational theories. In *Visualization Psychology*. Springer Nature, 2023. doi:10.1007/978-3-031-34738-2. 23

[TB23] TRAVER P., BERGH E.: Harmonices solaris - sonification of the planets. In *Proc. 28th International Conference on Auditory Display (ICAD 2023)* (Norrköping, Sweden, June 2023), Georgia Institute of Technology, pp. 242–248. doi:10.21785/icad2023.204. 7, 8, 17

[TC05] THOMAS J. J., COOK K. A. (Eds.): *Illuminating the Path: The Research and Development Agenda for Visual Analytics*. IEEE, Los Alamitos, CA, USA, 2005. 4

[TML*15] TARDIEU J., MISDARIIS N., LANGLOIS S., GAILLARD P., LEMERCIER C.: Sonification of in-vehicle interface reduces gaze movements under dual-task condition. *Applied Ergonomics 50* (2015), 41–49. doi:10.1016/j.apergo.2015.02.004. 20

[TMN*21] TEMOR L., MACDONALD D. E., NATARAJAN T., COPPIN P. W., STEINMAN D. A.: Perceptually-motivated sonification of spatiotemporally-dynamic CFD data. In *Proc. 26th International Conference on Auditory Display (ICAD 2021)* (Virtual Conference, June 2021), Georgia Institute of Technology, pp. 202–209. http://hdl.handle.net/1853/66343. 7, 8, 15, 16, 22

[VG97] VOLPE C. R., GLINERT E. P.: Auralization of streamline vorticity in computational fluid dynamics data. In *Visualization '97., Proceedings* (1997), pp. 51–57. doi:10.1109/VISUAL.1997.663856. 4

[VH06] VICKERS P., HOGG B.: Sonification abstraite/sonification concrète: An 'æsthetic persepctive space' for classifying auditory displays in the ars musica domain. In *Proceedings of the 12th International Conference on Auditory Display* (London, UK, June 2006), Georgia Institute of Technology, pp. 210–216. URL: http://hdl.handle.net/1853/50641. 3

[Vic16] VICKERS P.: Sonification and music, music and sonification. In *The Routledge Companion to Sounding Art*, Cobussen M., Meelberg V., Truax B., (Eds.). Routledge, 2016, pp. 135–144. 10

[VRGM20] VAN RHEDEN V., GRAH T., MESCHTSCHERJAKOV A.: Sonification approaches in sports in the past decade: a literature review. In *Proceedings of the 15th International Audio Mostly Conference* (2020), ACM, pp. 199–205. doi:10.1145/3411109.3411126. 5, 20

[War19] WARE C.: *Information visualization: perception for design*. Morgan Kaufmann, 2019. 4

[WHT06] WARD J., HUCKSTEP B., TSAKANIKOS E.: Sound-colour synaesthesia: To what extent does it use cross-modal mechanisms common to us all? *Cortex 42* (2006), 264–280. doi:10.1016/S0010-9452(08)70352-6. 4

[Wil05] WILKINSON L.: *The Grammar of Graphics*, second ed. Springer, New York, 2005. 3

[WL01] WALKER B. N., LANE D. M.: Sonification mappings database on the web. In *Proceedings of the 2001 International Conference on Auditory Display* (Espoo, Finland, 2001), Georgia Institute of Technology, p. 281. 4

[WM10] WALKER B. N., MAUNEY L. M.: Universal design of auditory graphs: A comparison of sonification mappings for visually impaired and sighted listeners. *ACM Transactions on Accessible Computing (TACCESS) 2*, 3 (2010), 1–16. 2, 19, 23

[WMY*17] WANG Q., MARKOPOULOS P., YU B., CHEN W., TIMMERMANS A.: Interactive wearable systems for upper body rehabilitation: a systematic review. *Journal of NeuroEngineering and Rehabilitation 14*, 1 (2017), 20. doi:10.1186/s12984-017-0229-y. 5, 20

[Woh14] WOHLIN C.: Guidelines for snowballing in systematic literature studies and a replication in software engineering. In *Proceedings of the 18th International Conference on Evaluation and Assessment in Software Engineering* (London England United Kingdom, May 2014), ACM, pp. 1–10. doi:10.1145/2601248.2601268. 6

[Wor19a] WORRALL D.: Polymedia design for network metadata monitoring. In *Sonification Design: From Data to Intelligible Soundfields*, Worrall D., (Ed.). Springer, Cham, 2019, pp. 253–273. doi:10.1007/978-3-030-01497-1_9. 20

[Wor19b] WORRALL D.: *Sonification Design: From Data to Intelligible Soundfields*. Springer, Cham, 2019. doi:10.1007/978-3-030-01497-1. 3

[WRK*19] WALKER B. N., REHG J. M., KALRA A., WINTERS R. M., DREWS P., DASCALU J., DAVID E. O., DASCALU A.: Dermoscopy diagnosis of cancerous lesions utilizing dual deep learning algorithms via visual and audio (sonification) outputs: Laboratory and prospective observational studies. *EBioMedicine 40* (2019), 176–183. doi:10.1016/j.ebiom.2019.01.028. 20

[WS07] WATSON M., SANDERSON P.: Designing for attention with sound: Challenges and extensions to ecological interface design. *Human Factors 49*, 2 (2007), 331–346. doi:10.1518/001872007X312531. 20

[WW15] WINTERS R. M., WEINBERG G.: Sonification of the tohoku earthquake: Music, popularization & the auditory sublime. In *Proc. The 21th International Conference on Auditory Display (ICAD–2015* (July 2015), Georgia Institute of Technology, pp. 273–280. http://hdl.handle.net/1853/54149. 7, 8, 11, 12, 13, 14

[XWX*22] XIN X., WANG Y., XIANG G., YANG W., LIU W.: Effectiveness of multimodal display in navigation situation. In *The Ninth International Symposium of Chinese CHI* (2022), Chinese CHI 2021, Association for Computing Machinery, pp. 50–62. doi:10.1145/3490355.3490361. 20

[YH17] YANG J., HERMANN T.: Mode explorer: Using model-based sonification to investigate basins of attraction. In *Proceedings of the 12th International Audio Mostly Conference on Augmented and Participatory Sound and Music Experiences* (New York, NY, USA, 2017), AM '17, Association for Computing Machinery. doi:10.1145/3123514.3123525. 8, 12, 13

[YH18] YANG J., HERMANN T.: Interactive mode explorer sonification enhances exploratory cluster analysis. *Journal of the Audio Engineering Society 66*, 9 (Sept. 2018), 703–711. doi:10.17743/jaes.2018.0042. 6, 7, 10, 16, 22

[YKSJ07] YI J. S., KANG Y. A., STASKO J. T., JACKO J. A.: Toward a deeper understanding of the role of interaction in information visualization. *IEEE Trans. Visualization and Computer Graphics 13*, 6 (2007), 1224–1231. doi:10.1109/TVCG.2007.70515. 3

[Ylm23] YLMAZ S.: Orbuculum, 2023. Ars Electronica Festival 2023. Webpage: https://ars.electronica.art/who-owns-the-truth/de/orbuculum/. Last accessed: Dec 28th 2023. 21

[YM22] YESHCHENKO A., MENDLING J.: A survey of approaches for event sequence analysis and visualization using the ESeVis framework, 2022. arXiv:2202.07941. 6

[Zie23] ZIEMER T.: Three-dimensional sonification for image-guided surgery. In *Proceedings of the 28th International Conference on Auditory Display* (Norrköping, Sweden, June 2023), Georgia Institute of Technology, pp. 8–14. doi:10.21785/icad2023.2324. 20

[ZPR16] ZAGERMANN J., PFEIL U., REITERER H.: Measuring cognitive load using eye tracking technology in visual computing. In *Proceedings of the sixth workshop on beyond time and errors on novel evaluation methods for visualization* (2016), ACM, pp. 78–85. doi:10.1145/2993901.2993908. 19

[ZTSS23] ZIMNICKI C., TSENG C., SZAFIR D. A., SCHLOSS K. B.: Effects of data distribution and granularity on color semantics for colormap data visualizations. In *2023 IEEE Visualization and Visual Analytics (VIS)* (2023), pp. 11–15. doi:10.1109/VIS54172.2023.00011. 4

Original Article 3

Parallel Chords: an audio-visual analytics design for parallel coordinates

The following article [EER*24] was published in the Journal on Personal and Ubiquitous Computing[3] and is the result of a collaboration between the St. Pölten University of Applied Sciences, Austria, and Linköping University, Sweden.

[3]link.springer.com/journal/779

Personal and Ubiquitous Computing
https://doi.org/10.1007/s00779-024-01795-8

ORIGINAL PAPER

Parallel Chords: an audio-visual analytics design for parallel coordinates

Elias Elmquist[1] · Kajetan Enge[2,3] · Alexander Rind[2] · Carlo Navarra[1] · Robert Höldrich[3] · Michael Iber[2] · Alexander Bock[1] · Anders Ynnerman[1] · Wolfgang Aigner[2] · Niklas Rönnberg[1]

Received: 6 July 2023 / Accepted: 27 February 2024

Abstract

One of the commonly used visualization techniques for multivariate data is the parallel coordinates plot. It provides users with a visual overview of multivariate data and the possibility to interactively explore it. While pattern recognition is a strength of the human visual system, it is also a strength of the auditory system. Inspired by the integration of the visual and auditory perception in everyday life, we introduce an audio-visual analytics design named *Parallel Chords* combining both visual and auditory displays. *Parallel Chords* lets users explore multivariate data using both visualization and sonification through the interaction with the axes of a parallel coordinates plot. To illustrate the potential of the design, we present (1) prototypical data patterns where the sonification helps with the identification of correlations, clusters, and outliers, (2) a usage scenario showing the sonification of data from non-adjacent axes, and (3) a controlled experiment on the sensitivity thresholds of participants when distinguishing the strength of correlations. During this controlled experiment, 35 participants used three different display types, the visualization, the sonification, and the combination of these, to identify the strongest out of three correlations. The results show that all three display types enabled the participants to identify the strongest correlation — with visualization resulting in the best sensitivity. The sonification resulted in sensitivities that were independent from the type of displayed correlation, and the combination resulted in increased enjoyability during usage.

Keywords Audio-visual analytics · Sonification · Parallel coordinates · User evaluation

1 Introduction

In the context of analyzing multivariate data, visual analytics has proven to be effective for the detection of patterns and trends between variables. Parallel coordinates [44, 88] is a widely used visualization technique for analyzing multivariate data, where individual variables are presented as vertical axes that are evenly spaced parallel to each other. Data items are represented by polylines that intersect the axes at the value of the respective data item.

Even though the parallel coordinates technique is well-suited to visualize multivariate data, a number of challenges exist [37]. The individual axes of parallel coordinates being positioned next to each other allow one variable to be compared to at most two direct neighbors at a time. Axis order algorithms can be applied to set the optimal axis order to reveal a certain pattern, e.g., clusters [7], but the problem still persists when making more bivariate comparisons than the parallel coordinates plot allows. To overcome this limitation, parallel coordinates and other multivariate visualizations are often part of a multiple-view system to provide additional perspectives of the data. One approach is to display several parallel coordinates plots simultaneously as a matrix, showing all pairwise relations for the variables of the dataset [38]. Another approach is to accompany the parallel coordinates plot with bivariate visualizations. This allows displaying more variable comparisons within the same context, and can also make use of the strengths of different types of visualizations. For example, Li et al. [54] conducted a comparative user experiment for scatter plots and parallel coordinates and found that users can distinguish twice as many correlation levels in a scatter plot than in a parallel coordinates plot, and that users overestimate negative correlations with paral-

Elias Elmquist and Kajetan Enge both contributed equally to this work.

Elias Elmquist
elias.elmquist@liu.se

Kajetan Enge
kajetan.enge@fhstp.ac.at

Extended author information available on the last page of the article

Published online: 03 May 2024

lel coordinates. On the other hand, Kanjanabose et al. [47] showed that participants performed better with clustering and outlier detection tasks when using a parallel coordinates plot compared to using a scatter plot. This shows that different types of visualizations can complement each other when used in conjunction. However, presenting several types of visualizations simultaneously in a multiple-view system can lead to a loss of context and focus when switching between the views [59]. It also leaves less space for each visualization on the display, which especially impacts displays with limited screen size. Therefore, other ways of conveying additional bivariate information for multivariate visualizations should be explored. The added information and perspectives could instead be conveyed by another modality, such as the auditory, which has specific differences from the visual modality. Numerous studies on sonification as a technique for auditory representation of data have demonstrated its potential for the recognition of coherent patterns [21, 31, 60].

In this article, we investigate how the addition of the auditory modality through sonification can be beneficially used in conjunction with a parallel coordinates visualization by creating an audio-visual analytics design, named *Parallel Chords*. This is demonstrated by presenting how the sonification would convey different prototypical patterns, while also demonstrating it in a usage scenario. As a first step to validate the approach, a controlled experiment was conducted towards the most foundational type of patterns, i.e., correlations (positive and negative), to evaluate the sensitivity thresholds and experience of participants when distinguishing similar correlations through visualization, sonification, and by using both in combination. Through the demonstration of an audio-visual analytics design and analysis of experiment results, the following contributions are presented:

- An audio-visual analytics design for parallel coordinates, which is demonstrated with the use of prototypical data patterns and a usage scenario.
- Quantitative and qualitative evaluation results of participants distinguishing similar correlations using parallel coordinates, an auditory scatter plot, and the combined usage of both.

The results of this study will support both the visualization and the sonification communities to better understand the implications of audio-visual designs, what to consider during sound design, and in what context to combine sonification and visualization.

2 Related work

Previous research from both the visualization and the sonification areas is relevant to this work. We present an overview of visualization idioms for multivariate data, and specifically what solutions exist to mitigate the challenges of the parallel coordinates technique. We present how sonification can complement visualization, and what audio-visual analytics designs currently exist.

2.1 Visualization of multivariate data

Visualization idioms for multivariate data focus on datasets that encompass a large number of items, three or more quantitative variables, and no further relational information between the items. Numerous visualization idioms have been designed for such data [80, 87] and we will compare parallel coordinates plots to six idiom families. *Axis-based* idioms represent data items in a layout that maps variables to axes. With points, the scatter plots for each pair of variables are shown together in a grid, the scatter plot matrix, or sequentially one after another, the grand tour [4]. Both arrangements of the scatter plots depict each data item as multiple point marks either spatially or temporally and require the user to mentally connect the marks for multivariate patterns. In the parallel coordinates plot [44, 88], each data item is represented as one polyline, which, at least in principle, visually connects all the variable values. A radar plot is comparable to the parallel coordinates plot but places the axes in a radial arrangement. Polyline-based idioms allow the analysis of bivariate patterns best between neighboring axes. All axis-based idioms preserve the quantitative distribution of the variable values but they can suffer from overplotting. *Table-based* idioms such as Table Lens [66] simply follow the visual layout of a spreadsheet's rows and columns while densely encoding data values, e.g., as a bar chart. With interactive sorting, the user can identify patterns between two or more variables. They avoid overplotting but the y-position does not directly express a variable value. A third family of idioms algorithmically transforms multivariate data to simple visual marks. Dimension reduction algorithms like multi-dimensional scaling, primary component analysis, or stochastic neighbor embedding can reduce the dataset to the two dimensions of a scatter plot [76]. The RadViz [42] or Dust & Magnet [93] idiom applies a force-based metaphor so that point marks for items are positioned closer to the variable marks, for which they have a large value. The Andrews' curves [2] idiom displays a multivariate item as the plot of a finite Fourier series. *Glyph-based* idioms such as profile glyphs, stick figures, or Chernoff faces represent each data item with a visual entity that encodes multiple variables to the visual channels of one or more visual marks [8, 34]. Transformation-based and glyph-based idioms focus on the holistic representation of the data items whereas parallel coordinates also display patterns across one or more variables. *Pixel-based* idioms such as VisDB [49] and *nested* visualizations such as mosaic plots and dimensional stacking

[52] rely on some relational information such as a temporal dimension or categories to arrange the multivariate data and are thus not directly comparable to parallel coordinates.

2.2 Enhancements of parallel coordinates

While parallel coordinates plots provide an overview of multivariate data expressing all variable values along scaled axes, they have limitations from overplotting and can play their strengths best for bivariate patterns between two variables that are displayed on neighboring axes. These limitations can be addressed, to some extent, by interactively filtering items and changing the order and direction of axes. The effects of overplotting can be mitigated by histograms on top of axes, semi-transparent polylines, rasterization that preserves line orientations [58], or showing clusters of items as bands [33, 61]. Various algorithmic approaches, e.g., [3, 20, 64, 92], propose an axes order that reduces overplotting or better reveals patterns such as clusters. Blumenschein et al. [7] identified 32 reordering approaches and found out that there is a trade-off between these goals. Furthermore, the users' background knowledge about the data and the task may still call for a different combination of variables to be analyzed together.

Other visualization idioms derived from parallel coordinates abandon the sequential arrangement of the axes in order to better show bivariate patterns. TimeWheel [81], 3D TimeWheel [82], and CMRPC [46] place one axis in the center and connect it with all other axes that are positioned around it. In the 2D version, the axes cannot be placed parallel. Lind et al. [55] replicated these axes and arranged them as polygons around a central variable. These show all bivariate combinations of variables with a pair of parallel axis in 2D. The Parallel Scatterplot Matrix [85] can be interactively rotated from a scatter plot matrix into a multi-row parallel coordinates plot. The Parallel Coordinates Matrix [38] combines as many rows of parallel coordinates plots as needed to display all pairs of variables as neighbors. Claessen and van Wijk [15] give users the flexibility to create custom layouts by drawing and linking axes on a 2D canvas. All these derived approaches have drawbacks similar to the scatter plot matrix: the data item is no longer represented by a single visual mark and the subdivision of display space decreases the visual resolution. Some of the approaches are further restricted by rotated axes or a 3D projection.

2.3 Sonification and audio-visual analytics

Sonification is the auditory equivalent to visualization, i.e., the transformation of data into sound or the mapping of data characteristics to auditory channels [39, 65]. Sonification can be used for data exploration and there are a number of studies that evaluate auditory graphs [30, 57, 60, 79]. It has also been demonstrated that sonification can support visual perception [1, 70, 71], and various auditory channels can be successfully linked and related to visual channels [17, 21, 27, 57, 86]. Sounds, in sonification, can convey a multitude of information to listeners quickly [83], without adding visual clutter [13]. This suggests that sonification and visualization can fruitfully be combined, and previous studies have explored this combination [11, 21, 23, 25, 30, 41, 48, 57, 60, 67, 69]. While these studies indicate that using sonification together with visualization supports a user in various data analysis tasks, also theoretical bridges between the visualization and the sonification communities exist. Theoretical constructs from visualization research — the "spatial substrate," the "mark," and the "channels" [59] — have been adopted to sonification. The theoretical framework characterizes sonification using time as its "substrate" and zero-dimensional and one-dimensional "auditory marks" that use "auditory channels" to control their auditory appearance [24, 26].

Auditory scatter plots, where data attributes are mapped to note onset, duration, and pitch, can provide information resulting in almost identical estimates of correlation magnitude and correlation coefficients as for visual representations [31]. Judgments of correlation magnitude are also similarly affected by single outliers in visual and auditory scatter plots [31]. Auditory scatter plots where density levels have been mapped to auditory channels have also been proven to successfully provide information about density levels as well as data distribution and clustering [68, 91]. Sonification can also be used to support the perception of visually dense areas in cluttered visual displays, both scatter plots and parallel coordinates representations, by sonifying density levels as well as different datasets in the visual representation [72]. Furthermore, sonification can be used together with parallel coordinates plots for perception-based classification of individual data records in a relational dataset [62, 63]. Finally, sonification has been proposed to be able to complement parallel coordinates in terms of visual cluster overlapping, visual representations in general for high dimensionality data, challenges for the visual perception in color distinction, and limitations by screen resolution [36]. Based on these findings we argue that sonification can be successfully used as a supplement to visual representations, and can be used to discover classes of data and data features (see further discussions in [30]).

3 Parallel chords

To explore the combination of sonification and visualization in the context of parallel coordinates, workshops were conducted to decide on the final design, which are described in Section 3.1. Based on these discussions, we created an audio-visual analytics design named *Parallel Chords*. It enables users to explore multivariate data through both visualization and sonification by interacting with the axes of a parallel coordinates plot. Through the *Parallel Chords* interface, the user can select two axes by sequentially clicking on two axis labels (see Fig. 1). The variable selected first is mapped to the temporal onset of auditory marks in the sonification, and the variable of the second selected axis is mapped to their pitch. As soon as the user has selected the second axis, the interrelationship between the two axes is sonified. After this, two new axes can be selected for comparison.

This section first summarizes our design process and then presents the visualization and sonification components of *Parallel Chords* in detail using the constructs of the unified theoretical framework for audio-visual analytics designs by Enge et al. [24, 26]: *Visual marks* are placed in space (the *substrate* of visualization) using *channels* to encode information. *Auditory marks* (i.e., the individual sounds) are placed in time (the *substrate* of sonification) using *auditory channels* to encode information. In *Parallel Chords* each data item is represented both as a visual polyline and as an auditory mark that uses two auditory channels to encode information: the auditory mark's onset time and its pitch. In Section 4, we demonstrate the design by applying it to commonly occurring patterns in visual analytics, and in Section 4.1, we present a usage scenario demonstrating how the design can be used to convey information about non-adjacent axes in a parallel coordinates plot.

3.1 Design process

To investigate how sonification could support a parallel coordinates visualization, workshops were conducted where two visualization researchers and three sonification researchers, which are co-authors of this article, were interviewed as a group around how sonification could be used to benefit the visualization. The first workshop was an ideation workshop to identify challenges and generate ideas related to the integration of sonification with parallel coordinates. The second workshop, again within our group, was a concept workshop discussing different prototypical implementations that tackle the challenges of parallel coordinates identified in the ideation workshop. Both these workshops were led by the two first authors of this article and were conducted as semi-structured group interviews. Additionally, we conducted individual workshops with one interaction design researcher and one visualization researcher with a specific interest in parallel coordinates. These meetings were conducted as semi-structured interviews as well and provided an external and well-informed perspective onto our design ideas.

The notes that were taken during all workshops were categorized by the two first authors to summarize the outcome. The main take-aways from the ideation workshop was that sonification had the possibility to prevent specific shortcomings of parallel coordinates. One of the identified shortcomings concerned visual clutter due to the overplotting of an abundance of data items. Another challenge was the axis ordering problem and the limitation of pair-wise comparison of axes, which led to non-adjacent axes not being comparable. A number of sonification mapping strategies were created to concretize the ideas from the ideation workshop, which would aid in the shortcomings of the visualization. These mapping strategies were realized as sonification concepts and were demonstrated at the concept workshop.

The mapping strategies focused on the two main challenges of parallel coordinates that were highlighted in the

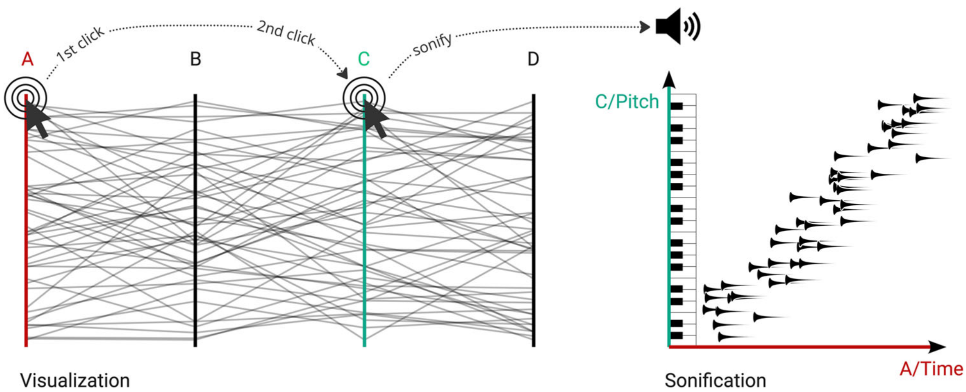


Fig. 1 Illustration of the audio-visual analytics design *Parallel Chords* when exploring the relationship of non-adjacent axes in a parallel coordinates plot. A user selects one axis (1) which is mapped to the temporal onsets of individual sounds in the sonification, and a second axis (2) which is mapped to the pitch of the individual sounds. In this case, the sonification conveys a positive correlation between axes *A* and *C* to the user represented via increasing pitches as time progresses

ideation workshop, namely overplotting and axis ordering. The mapping strategies for mitigating axis ordering focused on representing several data dimensions at the same time or in quick succession to get an overview of the dataset through the sonification. By mapping each attribute in a dataset to an individual auditory channel, it would enable a user to become aware of patterns in the dataset by listening to changes in the sonification. Another mapping approach was to use the same sonification mapping for each attribute but spatialize them to the corresponding axis in the parallel coordinates plot to be able to distinguish between them.

The mapping strategies for overplotting focused on representing the lines in the parallel coordinates plot in an auditory manner to aid in detecting patterns in a large dataset. One suggested mapping strategy was to convey the angle of the lines in a parallel coordinates plot through the sonification to perceive correlations in the dataset. Another mapping strategy involved a dynamic line selection, where lines in the parallel coordinates plot would only be sonified if they diverged from the general pattern between two axes.

While many of the mapping strategies had the potential of complementing a parallel coordinates visualization, we decided to use an established sonification mapping approach to act as an initial exploration of the combined design. Based on the discussions during the workshops, we decided for combining and studying mappings that are well researched individually, also due to the fact that we are investigating a relatively new field with this study. Therefore, the sonification was created to work as an auditory equivalent of a scatter plot, which is an established and well researched technique in the sonification community [30, 31, 57, 60, 79].

3.2 Visualization design

The visualization design consists of a standard parallel coordinates plot which was created by using the Data-Driven Documents (D3.js) information visualization library [9]. The parallel coordinates plot contains generic axis labels without tick marks to put focus on the data patterns themselves rather than specific data values. The lines are colored in black and were drawn with 30% opacity to allow blending of the lines. Users explore different variable relations by dragging the axes of the parallel coordinates plot. As this design will be used in a controlled experiment that investigates the identification of global data patterns such as correlations or outliers, it does not provide interaction for filtering or details on demand. This includes brushing and zooming of axes that filter the data range of an axis and highlighting data lines that give additional details of each data item. However, these common types of interactions should be provided when applying the design in a real-world use case and evaluating it in future application-oriented studies.

3.3 Sonification design

While the parallel coordinates plot displays all attributes at once in a certain axis order, the sonification works as an auditory scatter plot of two user-selected attributes. Onset time is used to convey the values of the first selected axis, where lower values lead to earlier onset times for the respective auditory mark. This is scaled in relation to the value range of the variable, such that value gaps in the data are noticeable as pauses. The overall playback duration of the sonification is set to range between one and two seconds. The auditory channel pitch is used to represent the values of the second selected axis, where a higher data value results in a higher pitch for the auditory mark, creating a positive polarity mapping. The pitch ranges from MIDI note number 55 (G3, fundamental frequency 196 Hz) to 91 (G6, fundamental frequency 1567.98 Hz). Together with the onset mapping, the change of pitch over time enables the user to detect auditory patterns in the data. A positive correlation between two variables, for example, would result in a sequence of auditory marks with later marks having higher pitch than earlier ones. In the course of this article, the axis that is used to sort the onset times will be called the *time axis*, and the axis that is used to sort the pitches of the individual auditory marks will be called the *pitch axis*.

The synthetic model of a mallet instrument comparable to a marimbaphone is used for the sonic representation of the auditory marks. The sound design was chosen for its clear distinction of temporal onsets, short decay time, and aesthetic qualities. The clear temporal onset of a mallet instrument aids in discerning individual data points when played in rapid succession. The short decay times of the individual sounds help to avoid temporal clutter during the sonification. If the chosen instrument needed a longer decay time to sound plausible, the individual sounds would soon mask each other. Using a sound design that resembles a real instrument was a dedicated decision. The timbre of a real instrument could be perceived as more aesthetic and therefore could increase its acceptability in comparison to the characteristics of a pure sine wave. Choosing such a sound design is also in line with all four design criteria for effective sonification design recently presented by Groß-Vogt et al. [35]. The four criteria are (1) to use easily perceptible sounds, (2) not to contradict data metaphors, (3) to follow a natural mapping, and (4) to use sounds appropriate to the task. The sonification model was implemented using SuperCollider [56], a real-time sound synthesis software environment that is commonly used for sonification purposes [12].

4 Prototypical patterns

To demonstrate *Parallel Chords*, we present commonly occurring types of data patterns [37] to serve as examples for the design. These patterns are positive and negative correlations, clusters, outliers, and sine. A compilation of the prototypical patterns in a parallel coordinates plot and in an illustrated auditory scatter plot can be seen in Fig. 2. The parallel coordinates plot and scatter plot create distinctly different outputs for each type of pattern. Since the sonification adopts an auditory scatter plot approach, different perspectives of the data can be gained by using both the visualization and the sonification simultaneously. An audiovisual demonstration of the patterns shown in Fig. 2 and additional examples can be viewed in *Video 1*.[1] The video includes more variations for each type of pattern compared to what is shown in Fig. 2, and also includes versions of the patterns where noise has been added to convey how the sonification behaves for less clear patterns. The video displays all of the patterns sequentially and uses an arrow to guide the viewer during the playback of the sonification. The rest of this section describes each type of pattern with respect to its statistical meaning, what variations of the pattern exist, how it could be identified with the audio-visual design, and how a user could benefit from using *Parallel Chords* to detect the patterns.

Correlation Correlation conveys to which degree two variables are linearly related to each other, ranging from perfect negative to perfect positive correlations. Positive and negative correlations are polar opposites from a mathematical perspective and create distinctly different types of patterns in a parallel coordinates plot. A perfect positive correlation is visually displayed as parallel lines between two axes, while a negative correlation is displayed in the shape of an "X" or diabolo [54]. Since the sonification has an auditory scatter plot approach, the two types of correlations are more alike, where a positive correlation is identified as a sequential increase of pitch of the auditory marks, while a negative correlation is identified as a sequential decrease in pitch. To determine the strength of a correlation, positive or negative, a user would analyze how sorted the auditory sequence is regarding pitch. The sonification aids in estimating correlations by offering a second perspective. The exploration of positive and negative correlations is what was studied in the controlled experiment of this study, presented in Section 5.

Clusters Clusters are groups of data items that share similarities with regard to at least one of their variables. In a parallel coordinates plot, a cluster is visually represented as a structure of lines that appear spatially grouped together, and usually creates several distinct visual patterns when many clusters are present. Auditory clusters can either be identified through the temporal or pitch grouping in the sonification, depending on the characteristics of the cluster. If clusters are present on the *pitch axis*, the distinction of the clusters is identified by a sudden and bigger difference in pitch. If clusters are present on the *time axis*, they are identified by a pause that separates groups of auditory marks. If clusters are occurring on both axes, a combination of these two effects is perceived. Since the two mappings of the sonification could be easier to distinguish between, compared to the visualization, it could support the user in detecting and identifying clusters in the dataset.

Outliers An outlier is characterized as a data item that differs substantially from the rest of the dataset. This can be due to it being outside of the general range of values in the dataset, or that it differs from the pattern that is displayed when comparing two variables with each other. An outlier is identified with the sonification by listening for a deviation in the overarching pattern of the dataset. Similarly to clusters, if an outlier is present on the *pitch axis*, it is identified by a sudden and big difference in pitch. If an outlier is present on the *time axis*, there will be a temporal pause separating the outlier from the rest of the auditory marks. Since the auditory modality has a high sensitivity to both temporal changes and changes in pitch [94], the sonification supports the user in detecting outliers in the dataset.

Sine Some types of patterns could be easier to identify audibly due to their inherent development over time, such as sine functions. Although these types of patterns are not as commonly occurring in real-life multivariate data, it gives an example of how the temporal perspective of the sonification can contribute to the analysis of the data. When visually inspecting the pattern of a sine function through a parallel coordinates plot, one might not associate the image with a sine function. Through the sonification, however, there is a signature sound of a periodical increase and decrease of pitch over time. Furthermore, the number of periods can be identified by counting how many times the auditory pattern is repeated.

4.1 Usage scenario

The presented prototypical patterns have demonstrated the use of *Parallel Chords* for commonly occurring patterns in a parallel coordinates plot with two axes. The following usage scenario illustrates how *Parallel Chords* could be used in practice by a fictional analyst, Dr. B., exploring botanical data. While this fictional user story can not validate our design, it serves to clarify our vision for such a tool. Dr. B. has

[1] Video of prototypical patterns: https://www.youtube.com/watch?v=T2n7JV9Qvog

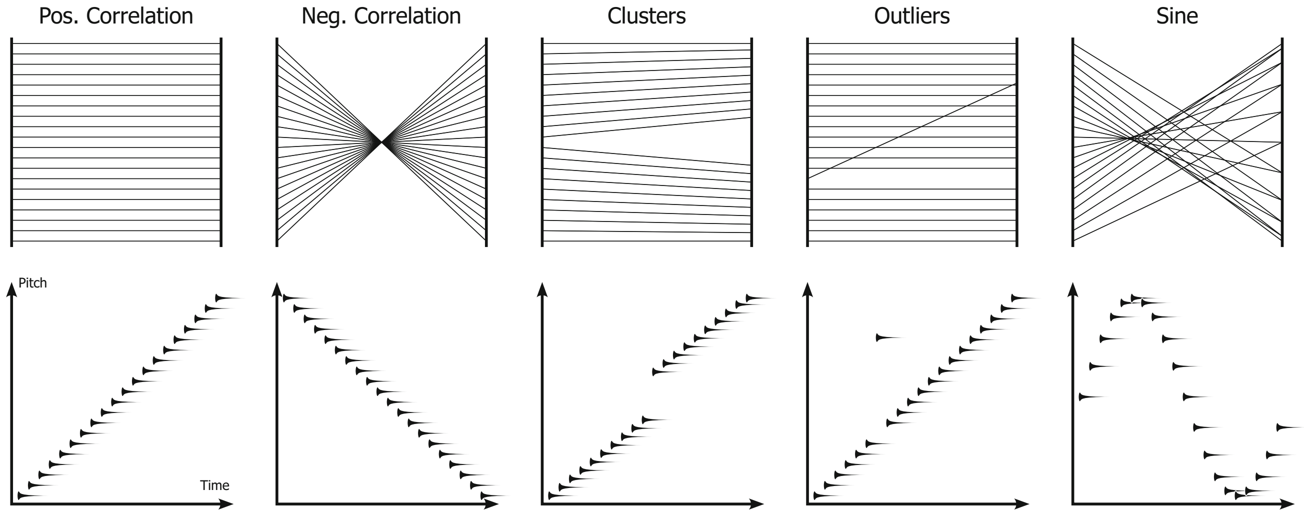


Fig. 2 Prototypical patterns: positive and negative correlations, clusters, outliers, and, additionally, a sine function. The patterns are displayed in a parallel coordinates plot and in an illustrated auditory scatter plot to demonstrate the audio-visual mapping of *Parallel Chords*. While the left axes of the individual parallel coordinates plots are used as the *time axes* of the auditory scatter plots, the right axes of the parallel coordinates plots are used as the *pitch axis*. A positive correlation, therefore, results in a sound sequence of individual tones rising in pitch over time. See *Video 1* to listen to the sonification of the patterns, including more variations and noisy versions

musical experience and is a data expert that has prior experience with using sonification and visualization to analyze and explore her data. She wears headphones and uses a standard screen in her office to interact with *Parallel Chords*. The supplemental material provides a video of the interactions that are relevant in the usage scenario of *Parallel Chords*.[2]

At the beginning of her analysis, Dr. B checks if the system is up and running by clicking axis D and E. She chose those two axes as this is were she sees two distinct clusters that are not overlapping. Therefore, she expects to hear a cluster of lower tones in the beginning, then a pause and a cluster of higher notes afterward. The system seems to work and Dr. B continues her investigation. From her previous experience using *Parallel Chords*, she knows that sometimes patterns can be more audible than visible. She is interested in the relationship between axis A and B and listens to them. The sonification reveals the presence of three clusters. After Dr. B. has heard the three clusters, she takes a closer look and now also sees them. For more detailed analysis, she decides to use a scatter plot at a later point and continues with her analysis using *Parallel Chords.* Dr. B is expecting axis A and D to hold clusters but does not know how many exist. From her experience, she knows that the relationship between A and D is only relevant if the two axes hold three or more clusters with each other. To quickly check the number of clusters between the two non-adjacent axes, without changing the visual view onto her data, Dr. B. sonifies their relationship and hears three distinct clusters. She now wants to understand the relationship better and drags axis D next to axis A to visualize their connection. The visualization helps her to understand that the lowest cluster is strictly separated from the other two, while the two remaining clusters slightly overlap each other.

In our scenario, Dr. B. listened to two visually adjacent axes (A and B) because she suspected a pattern that was not clearly visible to her. The sonification led her to take another look directed toward clusters which made her see them. We expect such an interaction between visualization and sonification to be dependent on prior experience and gained intuition using a tool. To better understand the occurrence of such phenomena, follow-up studies will be necessary as they are out of the scope of this article. Dr. B. also used the sonification in a complementary manner with the visualization by sonifying non-adjacent axes (A and D). By using only the sonification, non-adjacent axes can be explored separately from the visualization to aid the user in making multiple bivariate comparisons for the same axis, something that is otherwise a shortcoming with parallel coordinates. This can also be useful if an axes order algorithm has been applied to the dataset since axes comparisons can be made by the user while keeping the axis ordering of the visualization untouched. The user can explore each combination of axes with the sonification, and choose to confirm or sharpen their impression using the visualization by dragging the axes together. The user can also alternate between assigning the two axes of a pair to become either the *time axis* or the *pitch axis* to get an additional auditory perspective.

[2] https://www.youtube.com/watch?v=BYZVYHx56P8 - The displayed data is the iris dataset [29]

5 User evaluation

As a first step to validate *Parallel Chords*, we performed a controlled experiment for the most foundational type of pattern, namely correlations (positive and negative). The participants were asked to identify the strongest out of three correlations. This experiment task was chosen to understand the participant's ability to correctly interpret the mapping of the sonification and to study the participants' sensitivity in distinguishing small differences in the presented data.

Positive and negative correlations were tested separately to find potential differences between distinguishing the strength of the two types of correlations. Participants were tasked with identifying the strongest correlation using three display types: *Visualization*, *Sonification*, and by using both in *Combination*. *Visualization* would be used as a benchmark for regular use with a parallel coordinates plot. *Sonification* reflects how a user would use *Parallel Chords* for non-adjacent axes since there would not be any visual representation of the correlations. *Combination* would reflect how *Parallel Chords* is used for adjacent axes with combined visual and auditory representations of the correlations and would capture the sensitivity threshold of the participants when using both modalities at the same time.

5.1 Method

The experiment tasks were structured as "three alternative forced choice tests." A parallel coordinates plot with four axes was presented, where the participant would select which out of the three axes *B*, *C*, or *D* had the strongest correlation with axis *A*. For *Visualization*, the participant would interact by dragging the axes, as is commonly done with parallel coordinates, to compare and select the strongest correlation. For *Sonification*, the polylines of the parallel coordinates plot were not visible and the participant clicked on either axis *B*, *C*, or *D* to listen to their correlations with axis *A*. The *time axis* for the sonification was always assigned to axis *A*, which meant that the participant only needed to assign the *pitch axis* to any of the other axes. For *Combination*, the participant used the dragging interaction which would trigger the sonification when an axis was released next to axis A. It was also possible to click on the currently compared axis to listen to the sonification again. A screenshot of the experiment interface can be seen in Fig. 3.

The sensitivity threshold of the participant was measured using a staircase test design [22, 53], where the difficulty of the tasks changed depending on the prior responses of the participant. The difficulty, in this context, relates to the difference in the correlation coefficient between the correct axis, which had the strongest correlation with axis *A*, and the two incorrect axes, which had a weaker correlation with axis *A*. Smaller differences in the strength of correlation result in a more difficult detection task. The staircase procedure followed a one-up two-down procedure [53], meaning that the difficulty would increase after responding correctly two times in a row, and decrease after responding incorrectly once. Each increase in difficulty would span two levels of difficulty, while each decrease would span only one level of difficulty. Furthermore, the difficulty would increase by three levels after responding correctly four times in a row to enable the participant to converge to their individual sensitivity threshold with fewer number of tasks. The staircase started with a low level of difficulty, i.e., with large differences between the correct and wrong answers, so that it can be assumed most participants would be able to correctly respond to the first few examples before reaching their individual thresholds. One staircase procedure was used for each of the negative and positive correlation tasks to be able to analyze the results separately. The positive and negative staircases were interleaved such that every other task displayed negative correlations and the tasks in between displayed positive correlations. Additionally, the task completion times and the number of interactions needed to submit an answer were recorded. Qualitative aspects were obtained from a questionnaire that contained questions to be answered before and after the experiment, capturing the participant's subjective experience when performing the experiment. Pilot tests were conducted in preparation for the experiment to determine suitable correlation strengths and the number of tasks for each participant.

5.2 Datasets

Synthetic data was created to be used in the controlled experiment. Out of the three selectable axes (*B*, *C*, *D*), one axis would hold a stronger correlation with axis *A*. Selecting that axis was the correct answer, and selecting one of the other two axes, holding a weaker correlation with axis *A*, was an incorrect answer. Equations 1, 2, 3, and 4 describe the generation of the vectors $v1 - v4$ that were displayed on axis *A*-*D*. Vector $v1$ was always displayed on axis *A* and vectors $v2 - v4$ were randomly assigned to axis *B, C*, and *D*.

Axis *A*, displayed as the left-most axis during the experiment, held the vector $v1$, a linearly spaced vector holding 300 entries with values between 0 and 1, that had the reference noise of 7% (defined here as Gaussian noise with $\sigma = 0.07$) added to it. The axis that held the strongest correlation with axis *A* (with a Pearson correlation of about 0.95) had the same properties as axis *A*, but since it was individually generated it would not hold identical values. The two other axes were both individually generated with additional added noise ($\Delta\sigma$), so that they had the same properties but with a weaker correlation with axis *A*.

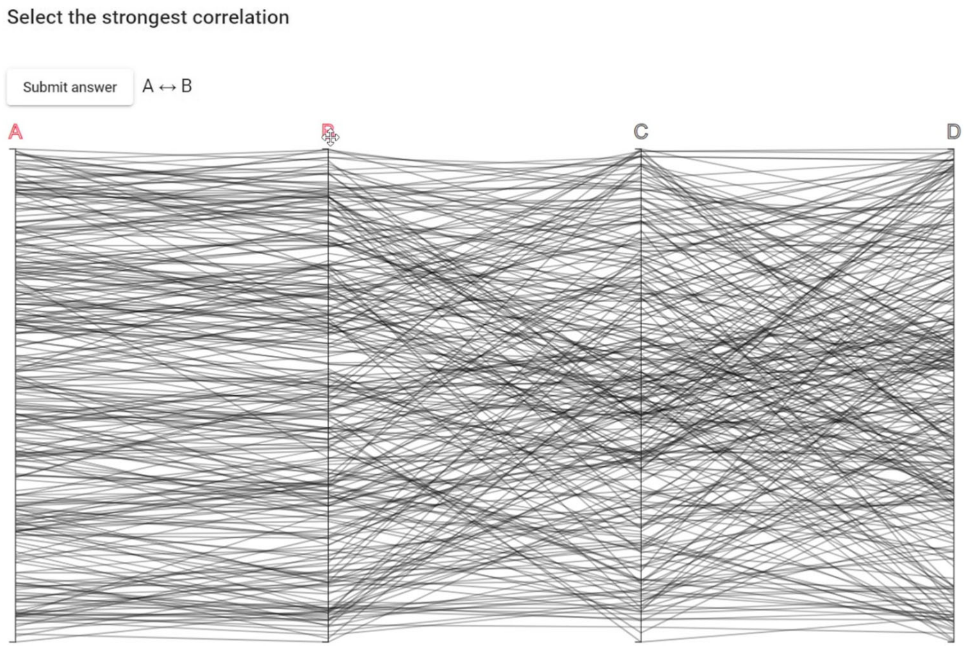


Fig. 3 The experiment interface, where the participant has currently selected the correlation between axis *A* and *B* as their answer. The task is to select which out of the three axes *B*, *C*, or *D* holds the strongest correlation with axis *A*. Using *Sonification* the 300 polylines were not visible and the interface only consisted of the axes and their labels. See the tutorial video in the supplemental material for a video of the example

$$v_1(i) = \mathcal{N}(\frac{i}{300}, \sigma^2) \quad (1)$$

$$v_2(i) = \mathcal{N}(\frac{i}{300}, \sigma^2) \quad (2)$$

$$v_3(i) = \mathcal{N}(\frac{i}{300}, (\sigma + \Delta\sigma)^2) \quad (3)$$

$$v_4(i) = \mathcal{N}(\frac{i}{300}, (\sigma + \Delta\sigma)^2) \quad (4)$$

where

$$\begin{aligned} i &= 0, 1, 2, ..., 299 \\ \mathcal{N}(\mu, \sigma^2) &= \text{normal distribution} \\ \sigma &= \text{reference noise } (0.07) \\ \Delta\sigma &= \text{additionally added noise } (0.02 - 20). \end{aligned}$$

The level of $\Delta\sigma$ determined the difficulty of the tasks. A lower $\Delta\sigma$ would result in more similar stimuli, which increased the difficulty of the task. 20 different levels of $\Delta\sigma$ were used to serve as the difficulty levels of the staircase test, ranging from $\Delta\sigma = 20\%$ (least difficult) down to $\Delta\sigma = 0.2\%$ (most difficult). Negatively correlated datasets were generated in the same manner, only that the selectable axes had their data entries inverted. Whenever a generated value happened to be out of range [0–1] due to added noise, a new random value with $\mu = \frac{i}{300}$ was generated and replaced with the original one. This was necessary for the datasets not to become sparsely populated around the edges of the axes after normalizing them to the same value range in the final step. The Jupyter Notebook we used to generate the datasets is available in the supplemental materials.

5.3 Procedure

The experiment procedure took participants around one hour to complete and was divided into five parts: pre-test questionnaire, tutorial, training, test, and post-test questionnaire (see Fig. 4). The experiment started with the participant filling in pre-test questions in a paper-based questionnaire concerning their age, gender, and possible perceptual impairments. Self-rated knowledge of the concept of correlation, their familiarity with parallel coordinates, their musical experience, and their familiarity with sonification were collected through a 5-point Likert scale. The experiment questionnaire can be viewed in the supplemental materials. The experiment proceeded on the computer with a tutorial on how to analyze correlations with visualization and sonification, respectively, and on how to interact with the interface to compare the different axes. Advice for distinguishing the strongest correlation was given for both the visualization and the sonification. The advice for visual analysis was to look for the "uniformity of

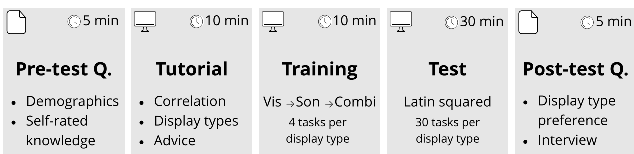


Fig. 4 The five parts that were included in the experiment procedure, including how long each part took to complete. The questionnaires were done on paper. The tutorial, training, and test were done on a computer

the patterns." The advice for auditory analysis was to listen to "how sorted the sounds are regarding their pitch." At this stage, the participant was also able to confirm that the sound volume was set at an appropriate level based on the tutorial video, which had the same sound volume as in the training part and the test. The tutorial video can be viewed in *Video 3*[3] which shows the experiment interface for *Sonification* and *Combination*. At the end of the tutorial, the participant was informed that all user inputs would be recorded for the training and test session and that this data would be stored, analyzed, and reported anonymously. No audio or video recordings of the participants were done, and the participant was able to leave the evaluation session at any point.

A training session followed, familiarizing the participant with the experiment interface by providing the same tasks as in the test, while also giving feedback if an answer was correct or incorrect. If the answer was incorrect, the participant would get further attempts at the task until they were correct. 12 training tasks were presented, with four training tasks for each display type, starting with *Visualization*, followed by *Sonification*, and finally using both in *Combination*. The test session started after the training, containing 30 tasks for each display type, with a total of 90 tasks for the three display types. The presentation order of the display types was structured with a Latin square design, such that the participants would be presented with three different orders. The three possible orders were used equally with the participants to even out any learning or order effects. The tasks alternated between negative and positive correlation for every task, with 15 tasks for each type of correlation. When the test was completed, a post-test questionnaire was filled in by the participant which included questions about the experience of using the different display types. An open-ended interview concluded the evaluation to complement the ratings made in the questionnaire, where the participant was asked to share details of their experience after performing the test. We report on the most frequent comments in Section 6.2.

5.4 Apparatus and architecture

The experiment was performed in a closed-off room on a standard desktop computer setup. The setup consisted of a 25″ computer screen with a resolution of 2560 × 1440 pixels, placed at arm's length from the participant's seating position. A pair of Beyerdynamic DT 770 PRO headphones was used for sound playback. A mouse was used to navigate the website displaying the experiment interface. The website was presented in full-screen mode to avoid any visual distractions. The experiment was performed in three separate locations but with the same setup.

[3] Tutorial video: https://www.youtube.com/watch?v=fp6dOphJ5FI

To connect the sonification with the visualization we used web technologies with a client-server architecture. The server, implemented with Node.js, performs the client communication by exchanging JSON messages via the WebSocket protocol and communicates with SuperCollider by sending messages using the Open Sound Control [90] protocol. On the client side, the experiment platform was built in TypeScript using the Angular framework.

5.5 Participants

35 participants took part in the controlled experiments (13 female, 19 male, 2 non-binary, 1 gender apathetic) which ranged from 20 to 60 years old (average age of 32.26, *SD* = 9.71). The self-rated knowledge of the participants regarding the concept of correlation, their familiarity with parallel coordinates, their musical experience, and their familiarity with sonification can be seen in Fig. 5. Overall, the participants had more knowledge in correlation and parallel coordinates compared to their knowledge in sonification and musical experience.

6 Results from the user evaluation

Statistical analysis was performed on the observed sensitivity thresholds and on the participants' task completion times. According to two different normality-tests [18, 19, 77], neither sensitivity threshold nor task completion time consistently follow normal distributions. For further analysis, we used a Friedman test [32] and Wilcoxon signed rank tests [89]. Holm-Bonferroni correction [43] for multiple comparisons was used where necessary. In addition to the

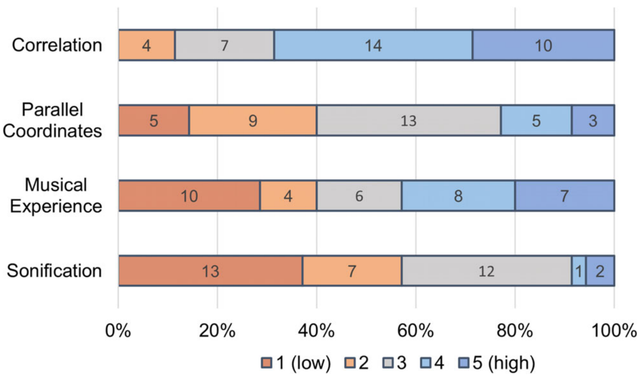


Fig. 5 Self-rated knowledge of the experiment participants through a 5-point Likert scale, where 1 corresponded to a low level of knowledge in the subject, and 5 corresponded to a high level of knowledge. See the specific description of the scale items in the evaluation questionnaire as part of the supplemental material

significance analysis, the effect size measure "Cliff's Delta" (Δ) [16] is reported.[4]

The p-values and Δ-values of the analysis are displayed in Table 1. They provide the results for both main effects (the type of correlation and the type of display) and all potentially interesting pairwise comparisons of conditions. For each participant, both the sensitivity threshold and the task completion time metrics were calculated by averaging the last eight of fifteen responses, i.e., the second half of each staircase. Even though some participants might not have converged towards their individual sensitivity threshold within the first half of each staircase, we confirmed that all results for the sensitivity thresholds reported in Table 1 are robust. The p-values and Δ-values would not change in a relevant way if one of the three different metrics was applied: (1) only the last value of each staircase, (2) the average over the last 4 responses, and (3) the average over all 15 responses.

Thematic analysis [10] was performed on the notes from the interviews with the participants by coding the notes based on a number of topics. These topics included comments on their approach for making a decision during the test, how they perceived the sonification, and for which condition they were most or least confident. The comments in each category were then counted by how many times a specific opinion was mentioned, where the most frequent comments are presented as results in Section 6.2.

6.1 Quantitative results

This section presents the quantitative results for the participants' sensitivity thresholds when they distinguished different strengths of correlations, results with respect to task completion times, and reflections on the influence of prior experience on the results. Kruskal-Wallis tests [50] did not reveal any significant differences regarding the order of presentation of display types (sensitivity: $p > 0.16$; task completion time: $p > 0.08$).

6.1.1 Sensitivity threshold

We found significant main effects of the correlation type ($p \ll 0.01$) and the display type ($p \ll 0.01$) on the participants' sensitivity thresholds and no significant interaction between the two factors ($p = 0.59$). Pairwise comparisons in Table 1 show differences between correlation types within each display type, and between display types within each correlation type.

Main effect 1: correlation type To study the main effect of the correlation type, we accumulated the data for the display type. The correlation type influenced the participants' sensitivity threshold significantly (Main Effect 1: $p \ll 0.01$, $\Delta = -0.3$). Being presented with negative correlations, the participants reached a lower threshold (i.e., they had better sensitivity) than with positive correlations displayed to them.

[4] Δ-values around 0.11 can be considered as "small effects," Δ-values around 0.28 as "medium effects," and Δ-values around and above 0.43 as "large effects" [84].

Table 1 p-values and Cliff's Delta values for all comparisons regarding the sensitivity thresholds of participants

	Comparisons		p	Δ
Main effect 1	Neg	Pos	$\ll$0.01	−0.3
Main effect 2	Vis	Son	$\ll$0.01	−0.69
	Son	Combi	$\ll$0.01	0.63
	Vis	Combi	0.07	−0.11
Pairwise	Vis Neg	Vis Pos	$<$ 0.01	−0.45
	Son Neg	Son Pos	0.28	−0.09
	Combi Neg	Combi Pos	$<$ 0.01	−0.40
	Vis Neg	Son Neg	$\ll$0.01	−0.83
	Vis Neg	Combi Neg	0.15	−0.19
	Son Neg	Combi Neg	$\ll$0.01	0.78
	Vis Pos	Son Pos	$<$ 0.01	−0.57
	Vis Pos	Combi Pos	0.24	−0.10
	Son Pos	Combi Pos	$<$ 0.01	0.48

The metric used for calculation is the average of the last 8 out of 15 responses in each staircase. "$\ll$" means smaller than 10^{-5} and "$<$" is smaller than 10^{-3}

Main effect 2: display type To study the main effect of the display type, we accumulated the data for the correlation type and ran a Friedman test ($p \ll 0.01$), followed by Wilcoxon signed rank tests for the three pairwise comparisons. While there is no significant difference between *Visualization* and *Combination* ($p = 0.07$, $\Delta = -0.11$), both of them significantly differ from *Sonification* (Vis: $p \ll 0.01$, $\Delta = -0.69$; Son: $p \ll 0.01$, $\Delta = 0.63$).

Interaction Using an aligned rank transform [40] enabled the application of a repeated measures two-way ANOVA, which did not reveal any interaction between the independent variables "Correlation type" and the "Display type" ($p = 0.59$) [5, 28]. In line with this result, the distributions show parallel trends for positive and negative correlations, i.e., the display type affects both correlation types similarly (see Fig. 6).

Pairwise comparisons Within the two display types *Visualization* and *Combination* a Wilcoxon signed rank pairwise comparison revealed a significant difference between negative and positive correlations (Vis: $p < 0.01$, $\Delta = -0.45$; Combi: $p < 0.01$, $\Delta = -0.40$). Using *Visualization* and *Combination* participants were able to distinguish significantly smaller differences whenever they were presented with negative correlations. For *Sonification*, a pairwise comparison revealed no significant difference between negative and positive correlations ($p = 0.28$, $\Delta = -0.09$).

Within the group of negative correlations, pairwise comparisons revealed significant differences between *Visualization* and *Sonification* ($p \ll 0.01$, $\Delta = -0.83$), between *Sonification* and *Combination* ($p \ll 0.01$, $\Delta = 0.78$), but not between *Visualization* and *Combination* ($p = 0.15$, $\Delta = -0.19$).

Within the group of positive correlations, pairwise comparisons revealed significant differences between *Visualization* and *Sonification* ($p < 0.01$, $\Delta = -0.57$), between *Sonification* and *Combination* ($p < 0.01$, $\Delta = -0.48$), but not between *Visualization* and *Combination* ($p = 0.24$, $\Delta = -0.10$).

6.1.2 Task completion time

Regarding the participants' task completion times, we found a significant main effect only of the correlation type ($p \ll 0.01$, $\Delta = -0.18$). The difference is small and is dominated by the results of *Visualization* and the *Combination* conditions. None of the other comparisons are robust against changing the metric from the average of the last 8 values. In most cases other metrics would lead to not significant differences and small effect sizes, hence, we do not consider them reportable. Table 2 shows the medians and standard deviations of the participants' task completion times, all being in similar ranges.

6.1.3 Self-rated knowledge analysis

In the questionnaire accompanying the experiment, participants rated their prior experience regarding four topics: the concept of correlation, parallel coordinates plots, their musicality, and the method of sonification (see Fig. 5).

Whether prior knowledge was affecting the sensitivity threshold of the participants was studied by comparing groups of participants with low experience to groups with high experience. The grouping was done such that the data would be distributed as balanced as possible between the two groups.[5] A visual comparison of box plots and the analysis of Cliff's Delta values revealed that the only condition potentially affected was *Sonification* for negative correlations ($\Delta = -0.35$ for sonification experience, -0.3 for musical experience, -0.14 for experience with parallel coordinates, and -0.36 for correlation experience). All other effect sizes were too small to be considered relevant. As the sample size of the two compared distributions is small the reliability of these results is vague. Therefore, the phenomenon we see in the data can only be considered an indication of a possible effect.

In addition to the grouping of sensitivities, a global cluster analysis of all participants revealed that four participants showed especially different behavior. While these four participants answered very differently than the more homogeneous rest of the participants, they also rated their sonification experience and musical experience as low. For the sake of transparency and a diverse sample of participants, we decided to not exclude those four participants from any statistical analysis presented in this study. Nevertheless, we did run all the analyses also with a sample of only 31 participants, in general revealing the same phenomena as the sample with 35 participants: The outliers for the *Sonification* conditions in Fig. 6 would disappear, i.e., it is the same distinct group of people causing the especially low sensitivity threshold observations when it comes to sonification. With the four participants excluded, the comparison of the display types *Sonification* and *Combination* would be significantly different ($p = 0.04$) but still show a small effect ($\Delta = -0.14$).

Table 2 Median values for sensitivity thresholds and task completion times ± their standard deviations for the six different conditions

Condition		Sensitivity thresholds [$\Delta\sigma$]	TC-times [s]
Visualization	Neg	0.8 ± 0.4	14.6 ± 5.9
	Pos	1.5 ± 2	18.5 ± 10.3
Sonification	Neg	3.4 ± 3.4	14.8 ± 4.9
	Pos	3.9 ± 3.8	17 ± 5
Combination	Neg	1.1 ± 0.5	16.3 ± 4.8
	Pos	2 ± 1.8	18.7 ± 8.9

The metric used for calculation is the average of the last 8 out of 15 responses in each staircase

6.2 Subjective ratings and experiences

The results from the post-test questionnaire can be seen in Fig. 7, where the participants answered which modality was preferred for making a decision when using *Combination*, which display type the participant was most confident using, which display type was easier to understand, and which display type was the most enjoyable to use. Overall, the participants felt most confident when using *Combination* (22 of 35), but reported that *Visualization* was easier to understand (19 of 35), and that *Combination* was the most enjoyable display type to use (24 of 35). The participants rated that they either used just the visualization or a combination of the visualization and sonification to reach a decision for their answer (both 14 of 35).

[5] Grouping for correlation ratings 2|3 vs. 4|5; parallel coordinates 1|2 vs. 3|4|5; musical experience 1|2 vs. 3|4|5; sonification 1|2 vs. 3|4|5.

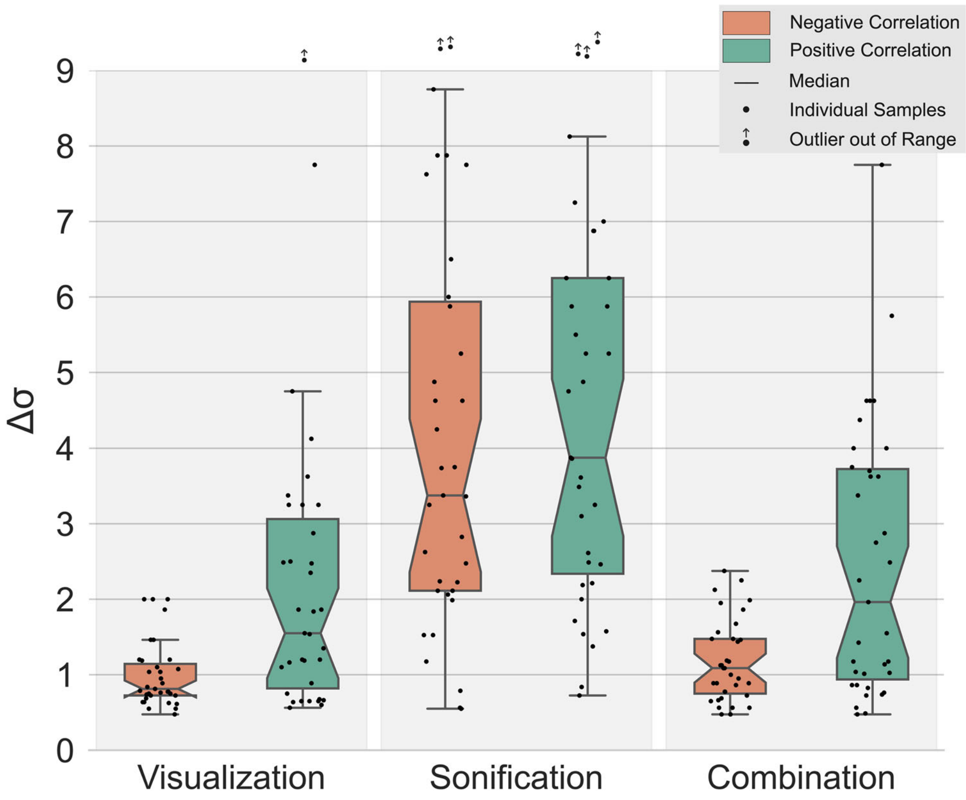


Fig. 6 The sensitivity thresholds for 35 participants and six different conditions are displayed as box plots. Participants are generally more sensitive using *Visualization* and *Combination*. Only for *Sonification*, the sensitivity threshold is not affected by the type of correlation

The open-ended interviews provided additional insights into the participants' experience. When comparing the difficulty of analyzing the two types of correlations across the display types, eleven participants explicitly stated that using *Visualization* for negative correlations was the easiest to interpret. For *Sonification*, four participants stated that the positive correlation and negative correlation were equally difficult to interpret, while three other participants stated that the positive correlation was easier to interpret compared to the negative correlation for *Sonification*. Three participants stated that the negative correlations were easier to interpret across all of the three display types.

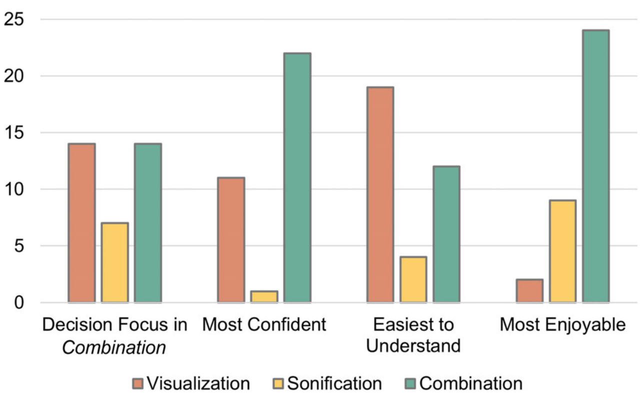


Fig. 7 Number of responses to the post-test questionnaire regarding the preference of display type. When using *Combination*, 14 participants decided to focus more on the visualization, 6 focused more on the sonification, and 14 used both for their decision

Regarding the decision strategy when using *Combination*, eight participants stated that they used the visualization for the first and easier tasks, and when tasks became more difficult they started to also use the sonification in their decision-making. Seven participants stated that they used the visualization to make an initial decision, and then used the sonification do double-check their decision. One participant mentioned that *"the audio helped with intuitive decisions and the visuals helped with analytic decisions."*

Regarding the sonification, five participants expressed that they were not confident with how they were supposed to interpret the sonification. A participant with high self-rated musical experience mentioned that they thought more about the harmony of the sound while analyzing with the sonification, referring to the tonal relationship between the first and last auditory marks. Some participants mentioned how the sonification affected them emotionally, where one participant stated that *"it felt satisfying to hear the strongest correlation,"* and *"I found Visualization way more tedious/ boring – with the sound the time of the task felt shorter."*

6.3 Summary of the results

The analysis presented in Section 6.1 shows that participants were able to identify the strongest correlation with all three types of display but with different sensitivity thresholds. For *Visualization* and *Combination*, the participants reached a lower sensitivity threshold, i.e., they had better sensitivity (see Fig. 6 and Table 2). With negative correlations, participants were even able to distinguish datasets that only differed by an amount of $0.8 \pm 0.4\%$ added noise and below. Using *Sonification* enabled participants to reach a threshold of 3–4% of added noise (with standard deviations of also 3–4%). This is in line with the subjective feedback from several of the participants stating that using *Visualization* for negative correlations was the easiest to interpret, mostly because they were able to compare the density of lines in the central area between the two axes. As the analysis did not show a statistically significant difference between the two display types *Visualization* and *Combination*, it can be assumed that, when having both modalities available, the visualization was the dominant representation to take a decision. While the thresholds for *Visualization* and *Combination* were dependent on the type of correlation, the thresholds for *Sonification* did not show a dependency on the type of correlation. A general user, comparing non-adjacent axes using sonification, would be similarly sensitive to changes of both positive and negative correlation coefficients.

The use of the three display types did not influence the task completion time significantly. While previous research showed higher task completion times with sonification [71], such a phenomenon did not appear in this study. The type of correlation, on the other hand, was influential on the participants' task completion times. Negative correlations did lead to slightly but significantly faster responses than positive correlation examples, which is a plausible outcome with negative correlations resulting also in better sensitivity thresholds.

In the beginning of the user evaluation, the participants were asked to report on their prior experience using sonification and on their musicality. Our results indicate that the participants' prior knowledge of these two was influential on only one specific condition: using *Sonification* to distinguish negative correlations. Participants with more sound experience were more sensitive than the ones without prior knowledge ($\Delta = -0.35$ for sonification experience and -0.3 for musical experience). It is plausible that prior experience and familiarity with sound affect the sensitivity threshold of participants with the sonification conditions, but while the phenomenon seems to exist for negative correlations, it does not for positive correlations. Masking effects as they are described by Schnupp et al. [75] might be relevant for the explanation of such a phenomenon. As the sound design, however, is based on semitones, it is not reasonable that masking effects are solely responsible for the phenomenon that we see in the data, and future research is needed to explain this observation.

The subjective ratings from Section 6.2 show that using *Combination* made participants more confident in their answers, and it was also the most enjoyable display type to use. The answers from the interviews indicate that participants used the sonification either to form an intuitive decision or to double-check their decision based on the visualization. Furthermore, the interviews indicated that sonification has the potential not only to serve as an analysis method but also to increase the engagement and emotional involvement.

7 Discussion

The evaluation of the present design was conducted by combining three approaches described by Isenberg et al. [45]. First, we studied the user performance (UP) by collecting data on the sensitivity, task completion times, as well as number of interactions done by subjects to find an answers. Second, we studied the user experience (UE) by collecting data on the subjects' confidence and enjoyment while using Parallel Chords. Third, we provide a qualitative result inspection (QRI) by describing the prototypical patterns and presenting a usage scenario.

To the best of our knowledge, we are not aware of a study that showed the dependency of users' sensitivity for differences in correlations on the direction of the correlations using parallel coordinates plots. The perceptive advantage of negative correlation patterns cannot be considered a surprise to the visualization community. Nevertheless, our data confirms that phenomenon. While we see a dependency on the type of correlation in the visualization condition, we could not observe such a dependency in the sonification condition. Again, we are not aware of a study in the sonification literature that tested differences in sensitivity for very similar correlation strengths. It is necessary to put this result into context: For the chosen sound design (which is in general a widely used one) and for the chosen base level of correlation around $r = 0.95$ ($\sigma = 0.07$), we are able to report a sensitivity threshold between three and four percent of added Gaussian noise. Furthermore, we were able to show that the type of correlation (positive or negative) does not affect the sensitivity when participants use the sonification. To understand these phenomena in more detail, a follow-up study would be necessary, testing sensitivities at different base correlation levels.

When relating our evaluation results to existing literature in a more general sense, we can see that some align with our results, while others deviate. Similar to our results, Stahl and Vogt [78] showed, in contrary to their initial hypothesis, that the audiovisual condition of an experiment with serial spatial stimuli did not show improved learnability of spatial positions compared to a visual-only condition. Similarly, when augmenting a visualization for a guidance task with sonification, Roodaki et al. [73] conducted evaluations which showed that participants performed better when using the visual-only technique in comparison with the audiovisual technique. On the other hand, a study of Rönnberg and Johansson [72] showed that participants gave more precise answers but that the response time was longer when using an audio-visual display, compared to a visual-only display. In our study, we observed lower sensitivity and no reportable difference in response time. The results of the experiment by Flowers et al. [31] also suggest that there is a cross-modal equivalence of visual and auditory scatterplots for exploring bivariate data samples, which strengthens the motivation of presenting a scatter plot in the auditory domain.

Regarding the subjective aspect of the evaluation results, most participants rated that the visualization was the easiest to understand out of the three display types. This can be associated with the challenge for sonification as a data representation technique being less known and used in everyday life [6]. The lower sonification literacy can affect the amount of training needed for participants to get accustomed to the mappings [74]. This might also have influenced the participants' strategy when taking a decision in the experiment, as several of the participants stated that they focused on the visualization as their primary method for reaching a decision, while the sonification was used to double-check or only used for the more difficult tasks. Still, the sonification influenced the overall experience of performing the experiment tasks, considering that most participants felt more confident and found the tasks more enjoyable when using the visualization together with the sonification compared to using the visualization by itself. These findings are emphasized by some of the participants' feedback of the sonification making them more involved and satisfied when hearing the strongest correlation, which, in turn, suggests that the sonification positively contributes to the user experience. Assessing enjoyability acts as an initial exploration of whether the use of *Parallel Chords* can promote more engagement of the user, which could extend the use of the tool beyond analytics and towards public engagement as well for future applications.

7.1 Applicability

The results from the present study show that it is possible to distinguish between different strengths of correlation using all three display types but with different sensitivity thresholds. Whenever the participants were able to use the visualization they reached lower sensitivity thresholds, i.e., they performed better. The participants' ability to correctly interpret the presented datasets with sonification shows the potential for applicability for other purposes since distinguishing between the strength of correlations is a more complex task than just being aware of the existence or the quality of a pattern. Therefore, the study results suggest *Parallel Chords* will be suitable at conveying complementary and more high-level information about a dataset. Rather than distinguishing between the strength of correlations, *Parallel Chords* can be used to convey an overview of a dataset, to make the user aware of the existence of patterns in the dataset. As demonstrated with prototypical patterns in Section 4, it is possible to distinguish between several types of patterns while using the same sonification mappings. This can be beneficially used when searching for patterns occurring between non-adjacent axes in a parallel coordinates plots, as demonstrated in the usage scenario in Section 4.1, which can alleviate the challenge of axis ordering with parallel coordinates.

As an audio-analytics design, *Parallel Chords* can draw from the advantages of both sonification and visualization. The sonification can extend a traditional parallel coordinates plot by giving non-adjacent axes information while keeping the same view in the visualization, which could be beneficial when considering that interactive reordering costs time and cognitive resources [51]. On the other end, other multi-dimensional visualization conveys more dimensions by displaying more plots of the data, such as the scatter plot matrix. With our design, it is possible to display additional dimensions while still keeping the same screen estate, which can be more suitable for smaller visual interfaces. By conveying the data through sound, it offers a temporal aspect which can lead to new insights of the data. In the context of our design, the temporal perspective of sonification could therefore aid domain experts in the analysis of multi-dimensional time-series data. Moreover, the fact that sound can be perceived all around the user leads to that it can be utilized in settings where the visual modality is limited. Monitoring tasks are one such situation, where multi-dimensional temporal data can be conveyed to the user through the focused visual display as well as in the periphery through the sonification. This is also the case in virtual reality and other immersive environments, where the sonification component of our design could guide the user to interesting attributes in the dataset, which would otherwise be occluded or be out of sight in a 3D environment. The *Parallel Chords* design could also be of use in conventional data analysis and decision-taking processes, as they are common in industrial design contexts. An example for the use of parallel coordinates supporting decision-taking between multi-criteria alternatives is described by Cibulski et al. [14].

When relating our design to other similar designs in the literature, we find examples of how our design provides a different approach. The sonification design of Rönnberg and

Johansson [72] extends a parallel coordinates display with sonification to support a user in identifying areas of different densities. In comparison to their design, ours sonifies the data items individually. While the density display provides information about one specific phenomenon, *Parallel Chords* is designed to help users identify several different types of patterns in their data. Parson et al. [62] used a sonification approach that treated a parallel coordinates plot as a waveform, and changed the timbre of the sonification based on the average value of the attributes of the dataset. While this facilities an overview of the dataset, it does not allow the user to focus on specific attributes with the sonification. Through the *Parallel Chords* design it is possible to get specific attribute information by selecting different axes in the parallel coordinates plot.

7.2 Limitations

Parallel Chords is not intended to increase the amount of data that can be displayed using parallel coordinates, rather it aims to support the use of parallel coordinates for data exploration. Generally, *Parallel Chords* representing data both spatially and temporally affects the scalability of the design in two ways. While the conventional limitations regarding the visual overplotting need to be considered, *Parallel Chords* also is limited by temporal constraints. We evaluated the design by sonifying 300 items with their onsets happening within one second. Considering three to four seconds as the maximum feasible time for effective sonification, the design scales up to about 1000 data items. Considering different sound designs with, e.g., shorter sounds or additional spatial positioning of auditory marks could increase the scalability of the design.

While one central application of sonification concerns the accessibility of visual display, this was not the focus of this study. Nevertheless, the results regarding sensitivity threshold and the different auditory patterns can inspire the design of accessible visualizations. The design implications for the sonification are not limited to parallel coordinates plots in that regard.

The method to generate the data for the controlled experiment of this study also implies one of its limitations. Some participants were able to distinguish datasets that only differed by a very small amount of added Gaussian noise, i.e., by a standard deviation of only 1% and below. Such small differences can not only be too small to be perceived as different, but also to be considered as statistically different at all. To make sure to not ask participants to detect differences between two statistically equal datasets, we re-generated the staircase datasets whenever their Pearson r values would not be monotonically decreasing over the course of the staircase. It is unlikely that users would ever want to distinguish such small differences between correlations ($\Delta\sigma < 1\%$) using a parallel coordinates plot. Therefore, when participants were able to distinguish such small differences, they can be considered as perfectly sensitive.

To rule out the possibility of systematic higher sensitivity thresholds due to our sonification design, we also studied the potential influence of the chosen MIDI quantization of the *pitch axis* on the evaluation results. A comparison between the Pearson r correlation coefficients with and without the 36-step quantization of one of the axes showed a neglectable difference. We can conclude that the observed sensitivity thresholds with sonification, and therefore also the evaluation results, were not significantly impacted by the MIDI quantization.

8 Conclusion

We presented *Parallel Chords*, an audio-visual analytics design for parallel coordinates that combines visual and auditory displays to aid the user in finding and determining patterns in multivariate data. Through a set of prototypical patterns, we demonstrated how the sonification of *Parallel Chords* can be interpreted to identify patterns together with a parallel coordinates plot. With a usage scenario of a real dataset, we showed how *Parallel Chords* can be used to convey patterns between non-adjacent axes. The results of a controlled user evaluation showed that participants were able to distinguish differences of correlations, but with different sensitivities when only using visualization or sonification, and when using a combination of both.

While in this article we focused on only one of many possible designs to combine parallel coordinates plots and sonification, future work will cover several other aspects. In this study, synthetic data was used as stimuli for participants in the user evaluation to act as a first step to validate *Parallel Chords*. Future experiments will need to use datasets and the expertise from data analysts of different domains to evaluate the real-world applicability of the current design. To allow for a more extensive exploratory data analysis approach, *Parallel Chords* can be extended to more efficiently compare several non-adjacent axes. This could be done by sequentially sonifying every pairwise relation of one axis through one interaction of the user. Alternatively, the variables of a dataset could be mapped to individual auditory channels (like pitch, spatial position, timbre, duration, and loudness) to enable a user to become aware of the existence of patterns in their data by only listening to the sequence of polylines once. A controlled experiment would then reveal if a user would also identify a correlation by hearing a complex sound moving from one speaker to the other, or by the sounds getting louder over time.

The results suggest that *Parallel Chords* can be a useful audio-visual analytics tool, even if more research is needed to fully explore and evaluate it. The work has not only led to novel knowledge about audio-visual analytics but also, to some extent, bridged the distance between the visualization and the sonification research communities.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s00779-024-01795-8.

Funding Open access funding provided by Linköping University. This research was funded in part by the Austrian Science Fund (FWF): 10.55776/P33531 and the Gesellschaft für Forschungsförderung Niederösterreich (GFF) SC20-006, as well as by the Knut and Alice Wallenberg Foundation (grant KAW 2019.0024).

Data availability The datasets generated during and/or analyzed during the current study are available in the osf repository, https://osf.io/z9vnw/?view_only=bac1e61f5b1e4b3fb7c08720ef5d7355

Declarations

Conflict of interest The authors declare no competing interests.

References

1. Ahmetovic D, Avanzini F, Baratè A et al (2019) Sonification of rotation instructions to support navigation of people with visual impairment. In: 2019 IEEE International conference on Pervasive Computing and Communications (PerCom), IEEE, pp 1–10, https://doi.org/10.1109/PERCOM.2019.8767407
2. Andrews DF (1972) Plots of high-dimensional data. Biometrics 28(1):125–136. https://doi.org/10.2307/2528964
3. Ankerst M, Berchtold S, Keim D (1998) Similarity clustering of dimensions for an enhanced visualization of multidimensional data. In: Proc. IEEE Symp. on Information Visualization. IEEE, pp 52–60, https://doi.org/10.1109/INFVIS.1998.729559
4. Asimov D (1985) The grand tour: a tool for viewing multidimensional data. SIAM J Sci Stat Comput 6(1):128–143. https://doi.org/10.1137/0906011
5. Beasley TM, Zumbo BD (2009) Aligned rank tests for interactions in split-plot designs: distributional assumptions and stochastic heterogeneity. J Mod Appl Stat Methods 8(1):16–50. https://doi.org/10.22237/jmasm/1241136180
6. Bijsterveld K (2019) Sonic skills: listening for knowledge in science, medicine and engineering (1920s-present). Palgrave Macmillan London. https://doi.org/10.1057/978-1-137-59829-5
7. Blumenschein M, Zhang X, Pomerenke D et al (2020) Evaluating reordering strategies for cluster identification in parallel coordinates. Comput Graph Forum 39(3):537–549. https://doi.org/10.1111/CGF.14000
8. Borgo R, Kehrer J, Chung DHS et al (2013) Glyph-based visualization: foundations, design guidelines, techniques and applications. In: Sbert M, Szirmay-Kalos L (eds) Eurographics 2013 - State of the Art Reports. The Eurographics Association, pp 39–63, https://doi.org/10.2312/conf/EG2013/stars/039-063
9. Bostock M, Ogievetsky V, Heer J (2011) D^3 data-driven documents. IEEE Trans Vis Comput 17(12):2301–2309. https://doi.org/10.1109/TVCG.2011.185
10. Braun V, Clarke V (2006) Using thematic analysis in psychology. Qual Res Psychol 3:77–101. https://doi.org/10.1191/1478088706QP063OA
11. Caiola V, Lenzi S, Riccò D (2022) Audiovisual sonifications: a design map for multisensory integration in data representation. In: Lockton D, Sara L, Hekkert P et al (eds) DRS2022: Bilbao, https://doi.org/10.21606/drs.2022.380
12. de Campo A, Rohrhuber J, Bovermann T et al (2011) Sonification and auditory display in supercollider. In: Wilson S, Cottle D, Collins N (eds) The SuperCollider Book. MIT Press, Cambridge, Massachusetts, pp 381–408
13. Case A, Day A (2018) Designing with sound: fundamentals for products and services. O'Reilly Media
14. Cibulski L, May T, Schmidt J et al (2014) COMPO*SED: composite parallel coordinates for co-dependent multi-attribute choices. IEEE Trans Vis Comput Graph 29(10):4047–4061. https://doi.org/10.1109/TVCG.2022.3180899
15. Claessen JHT, van Wijk JJ (2011) Flexible linked axes for multivariate data visualization. IEEE Trans Vis Comput Graph 17(12):2310–2316. https://doi.org/10.1109/TVCG.2011.201
16. Cliff N (1993) Dominance statistics: ordinal analyses to answer ordinal questions. Psychol Bull 114:494–509. https://doi.org/10.1037/0033-2909.114.3.494
17. Collier WG, Hubbard TL (2004) Musical scales and brightness evaluations: effects of pitch, direction, and scale mode. Musicae Scientiae 8:151–173. https://doi.org/10.1177/102986490400800203
18. D'Agostino R, Pearson ES (1973) Tests for departure from normality. Empirical results for the distributions of b2 and $\sqrt{}$b1. Biometrika 60(3):613–622. https://doi.org/10.1093/biomet/60.3.613
19. D'Agostino RB (1971) An omnibus test of normality for moderate and large size samples. Biometrika 58(2):341–348. https://doi.org/10.1093/biomet/58.2.341
20. Dasgupta A, Kosara R (2010) Pargnostics: screen-space metrics for parallel coordinates. IEEE Trans Vis Comput Graph 16(6):1017–1026. https://doi.org/10.1109/TVCG.2010.184
21. Dubus G, Bresin R (2013) A systematic review of mapping strategies for the sonification of physical quantities. PloS one 8(12):e82491. https://doi.org/10.1371/journal.pone.0082491
22. Elliott MA, Nothelfer C, Xiong C et al (2020) A design space of vision science methods for visualization research. IEEE Trans Vis Comput Graph 27(2):1117–1127. https://doi.org/10.1109/TVCG.2020.3029413
23. Elmquist E, Ejdbo M, Bock A et al (2022) Openspace sonification: complementing visualization of the solar system with sound. In: Proceedings of the 26th International Conference on Auditory Display (ICAD 2021). International Community for Auditory Display, pp 135–142, https://doi.org/10.21785/icad2021.018
24. Enge K, Rind A, Iber M et al (2021) It's about time: adopting theoretical constructs from visualization for sonification. In: Proceedings of the international audio mostly conference. ACM, pp 64–71, https://doi.org/10.1145/3478384.3478415
25. Enge K, Rind A, Iber M et al (2022) Towards multimodal exploratory data analysis: soniscope as a prototypical implementation. In: Agus M, Aigner W, Hoellt T (eds) EuroVis 2022 - Short

Papers. The Eurographics Association, pp 67–71, https://doi.org/10.2312/evs.20221095
26. Enge K, Rind A, Iber M et al (2023) Towards a unified terminology for sonification and visualization. J Pers Ubiquitous Comput 27(5):1949–1963. https://doi.org/10.1007/s00779-023-01720-5
27. Ferguson J, Brewster SA (2018) Investigating perceptual congruence between data and display dimensions in sonification. In: Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems. ACM, pp 611:1–611:9, https://doi.org/10.1145/3173574.3174185
28. Feys J (2016) New nonparametric rank tests for interactions in factorial designs with repeated measures. J Mod Appl Stat Methods 15(1):78–99. https://doi.org/10.22237/jmasm/1462075500
29. Fisher RA (1988) Iris. UCI machine learning repository, https://doi.org/10.24432/C56C76
30. Flowers JH (2005) Thirteen years of reflection on auditory graphing: promises, pitfalls, and potential new directions. In: Proc. International Conference on Auditory Display, (ICAD 2005), Georgia Institute of Technology, pp 406–409
31. Flowers JH, Buhman DC, Turnage KD (1997) Cross-modal equivalence of visual and auditory scatterplots for exploring bivariate data samples. Hum Factors 39:341–351. https://doi.org/10.1518/001872097778827151
32. Friedman M (1937) The use of ranks to avoid the assumption of normality implicit in the analysis of variance. J Am Stat Assoc 32(200):675–701. https://doi.org/10.2307/2279372
33. Fua YH, Ward MO, Rundensteiner EA (1999) Hierarchical parallel coordinates for exploration of large datasets. In: Proc. Conf. Visualization. IEEE Computer Society, Washington, DC, USA, pp 43–50, https://doi.org/10.1109/VISUAL.1999.809866
34. Fuchs J, Isenberg P, Bezerianos A et al (2017) A systematic review of experimental studies on data glyphs. IEEE Trans Vis Comput Graph 23(7):1863–1879. https://doi.org/10.1109/TVCG.2016.2549018
35. Groß-Vogt K, Enge K, m. zmölnig I (2023) Reflecting on qualitative and quantitative data to frame criteria for effective sonification design. In: Proceedings of the 18th International Audio Mostly Conference. Association for Computing Machinery, New York, NY, USA, AM '23, pp 93–100, https://doi.org/10.1145/3616195.3616233
36. Halim Z, Baig R, Bashir S (2006) Sonification: a novel approach towards data mining. In: 2006 International Conference on Emerging Technologies, IEEE, pp 548–553, https://doi.org/10.1109/ICET.2006.336029
37. Heinrich J, Weiskopf D (2013) State of the art of parallel coordinates. In: Sbert M, Szirmay-Kalos L (eds) Eurographics 2013 – State of the Art Reports. The Eurographics Association, https://doi.org/10.2312/conf/EG2013/stars/095-116
38. Heinrich J, Stasko J, Weiskopf D (2012) The parallel coordinates matrix. In: Eurographics Conference on Visualization – Short Papers, pp 37–41, https://doi.org/10.2312/PE/EuroVisShort/EuroVisShort2012/037-041
39. Hermann T, Hunt A, Neuhoff JG (2011) The sonification handbook. Logos Publishing House, Berlin, Germany
40. Higgins JJ, Blair RC, Tashtoush S (1990) The aligned rank transform procedure. Conf Appl Stat Agriculture. https://doi.org/10.4148/2475-7772.1443
41. Hildebrandt T, Amerbauer F, Rinderle-Ma S (2016) Combining sonification and visualization for the analysis of process execution data. In: 2016 IEEE 18th Conference on Business Informatics (CBI), IEEE, pp 32–37, https://doi.org/10.1109/CBI.2016.47
42. Hoffman P, Grinstein G, Marx K et al (1997) DNA visual and analytic data mining. In: Proc. Visualization. IEEE, pp 437–441, https://doi.org/10.1109/VISUAL.1997.663916
43. Holm S (1979) A simple sequentially rejective multiple test procedure. Scandinavian Journal of Statistics 6(2):65–70. https://www.jstor.org/stable/4615733
44. Inselberg A (1985) The plane with parallel coordinates. Vis Comput 1(2):69–91. https://doi.org/10.1007/BF01898350
45. Isenberg T, Isenberg P, Chen J et al (2013) A systematic review on the practice of evaluating visualization. IEEE Trans Vis Comput Graph 19:2818–2827. https://doi.org/10.1109/TVCG.2013.126
46. Johansson J, Cooper M, Jern M (2005) 3-dimensional display for clustered multi-relational parallel coordinates. In: Proc. 9th Int. Conf. on Information Visualisation (IV). IEEE, pp 188–193, https://doi.org/10.1109/IV.2005.1
47. Kanjanabose R, Abdul-Rahman A, Chen M (2015) A multi-task comparative study on scatter plots and parallel coordinates plots. Comput Graph Forum 34(3):261–270. https://doi.org/10.1111/cgf.12638
48. Kasakevich M, Boulanger P, Bischof WF et al (2007) Augmentation of visualisation using sonification: a case study in computational fluid dynamics. In: Proc. IPT-EGVE Symposium. The Eurographics Association, Germany, Europe, pp 89–94, https://doi.org/10.2312/PE/VE2007Short/089-094
49. Keim D, Kriegel HP (1994) VisDB: database exploration using multidimensional visualization. IEEE Comput Graph Appl 14(5):40–49. https://doi.org/10.1109/38.310723
50. Kruskal WH, Wallis WA (1952) Use of ranks in one-criterion variance analysis. J Am Stat Assoc 47(260):583–621. https://doi.org/10.1080/01621459.1952.10483441
51. Lam H (2008) A framework of interaction costs in information visualization. IEEE Trans Vis Comput Graph 14(6):1149–1156. https://doi.org/10.1109/TVCG.2008.109
52. LeBlanc J, Ward M, Wittels N (1990) Exploring N-dimensional databases. In: Proc. First IEEE Conference on Visualization. IEEE, pp 230–237, https://doi.org/10.1109/VISUAL.1990.146386
53. Levitt H (1971) Transformed up-down methods in psychoacoustics. J Acoust Soc Am 49(2B):467–477. https://doi.org/10.1121/1.1912375
54. Li J, Martens JB, van Wijk JJ (2010) Judging correlation from scatterplots and parallel coordinate plots. Inf Vis 9(1):13–30. https://doi.org/10.1057/ivs.2008.13
55. Lind M, Johansson J, Cooper M (2009) Many-to-many relational parallel coordinates displays. In: Proc. 13th Int. Conf. Information Visualisation. IEEE, pp 25–31, https://doi.org/10.1109/IV.2009.43
56. McCartney J (2002) Rethinking the computer music language: supercollider. IEEE Comput Graph Appl 26:61–68. https://doi.org/10.1162/014892602320991383
57. Metatla O, Bryan-Kinns N, Stockman T et al (2016) Sonification of reference markers for auditory graphs: effects on non-visual point estimation tasks. PeerJ Comput Sci 2:e51. https://doi.org/10.7717/peerj-cs.51
58. Muigg P, Hadwiger M, Doleisch H et al (2011) Visual coherence for large-scale line-plot visualizations. Comput Graph Forum 30(3):643–652. https://doi.org/10.1111/j.1467-8659.2011.01913.x
59. Munzner T (2015) Visualization analysis and design. CRC Press, Boca Raton
60. Nees MA (2018) Auditory graphs are not the "killer app" of sonification, but they work. Ergonomics in Design 26(4):25–28. https://doi.org/10.1177/1064804618773563
61. Novotny M, Hauser H (2006) Outlier-preserving focus+context visualization in parallel coordinates. IEEE Trans Vis Comput Graph 12(5):893–900. https://doi.org/10.1109/TVCG.2006.170
62. Parson DE, Hoch DE, Langley H (2016) Timbral data sonification from parallel attribute graphs. The 31st Annual Spring Conference of the Pennsylvania Computer and Information Science Educators (PACISE). Kutztown University of PA, Kutztown, PA, pp 22–31

63. Parson DE, Malke W, Langley H et al (2017) Mapping data visualization to timbral sonification and machine listening. Computer Science and Information Technology Faculty at Kutztown University 2017(4). https://research.library.kutztown.edu/cisfaculty/4
64. Peng W, Ward M, Rundensteiner E (2004) Clutter reduction in multi-dimensional data visualization using dimension reordering. In: Proc. IEEE Symp. Information Visualization. IEEE, pp 89–96, https://doi.org/10.1109/INFVIS.2004.15
65. Philipsen L, Kjærgaard RS (2018) The aesthetics of scientific data representation: more than pretty pictures. Routledge Advances in Art and Visual Studies, Denmark, Europe
66. Rao R, Card SK (1994) The table lens: merging graphical and symbolic representations in an interactive focus+context visualization for tabular information. In: Proc. SIGCHI Conference on Human Factors in Computing Systems. ACM, New York, USA, pp 318–322, https://doi.org/10.1145/191666.191776
67. Rau B, Frieß F, Krone M, et al (2015) Enhancing visualization of molecular simulations using sonification. In: Proc. IEEE 1st International Workshop on Virtual and Augmented Reality for Molecular Science (VARMS@IEEEVR 2015). IEEE, pp 25–30, https://doi.org/10.1109/VARMS.2015.7151725
68. Riedenklau E, Hermann T, Ritter H (2010) Tangible active objects and interactive sonification as a scatter plot alternative for the visually impaired. In: Proc. 16th International Conference on Auditory Display (ICAD 2010). International Community for Auditory Display, Germany, Europe, pp 1–7
69. Rind A, Iber M, Aigner W (2018) Bridging the gap between sonification and visualization. In: Proc. MultimodalVis'18 Workshop at AVI 2018, pp 1–4, https://doi.org/10.5281/zenodo.6510342
70. Rönnberg N (2019) Musical sonification supports visual discrimination of color intensity. Behav Inf Technol 38(10):1028–1037. https://doi.org/10.1080/0144929X.2019.1657952
71. Rönnberg N (2019b) Sonification supports perception of brightness contrast. Journal on Multimodal User Interfaces pp 1–9. https://doi.org/10.1007/s12193-019-00311-0
72. Rönnberg N, Johansson J (2016) Interactive sonification for visual dense data displays. In: Proceedings of ISon 2016, 5th Interactive Sonification Workshop. CITEC, Bielefeld University, pp 63–67
73. Roodaki H, Navab N, Eslami A et al (2017) Sonifeye: sonification of visual information using physical modeling sound synthesis. IEEE Trans Vis Comput Graph 23:2366–2371. https://doi.org/10.1109/TVCG.2017.2734327
74. Sawe N, Chafe C, Treviño J (2020) Using data sonification to overcome science literacy, numeracy, and visualization barriers in science communication. Front Commun 5:46. https://doi.org/10.3389/fcomm.2020.00046
75. Schnupp J, Nelken I, King A (2011) Auditory neuroscience: making sense of sound. MIT press
76. Sedlmair M, Munzner T, Tory M (2013) Empirical guidance on scatterplot and dimension reduction technique choices. IEEE Trans Vis Comput Graph 19(12):2634–2643. https://doi.org/10.1109/TVCG.2013.153
77. Shapiro SS, Wilk MB (1965) An analysis of variance test for normality (complete samples). Biometrika 52(3/4):591–611. https://doi.org/10.2307/2333709
78. Stahl B, Vogt K (2015) The effect of audiovisual congruency on short-term memory of serial spatial stimuli: a pilot test. In: Proceedings of the 21st International Conference on Auditory Display, Extended Abstracts. Georgia Institute of Technology, http://hdl.handle.net/1853/54200
79. Stockman T, Nickerson LV, Hind G (2005) Auditory graphs: a summary of current experience and towards a research agenda. In: Proc. International Conference on Auditory Display, (ICAD 2005), Georgia Institute of Technology, pp 420–422
80. Tominski C, Schumann H (2020) Interactive visual data analysis. AK Peters Visualization Series, CRC,. https://doi.org/10.1201/9781315152707
81. Tominski C, Abello J, Schumann H (2004) Axes-based visualizations with radial layouts. In: Proc. ACM Symp. Applied Computing (SAC). ACM, Nicosia, Cyprus, pp 1242–1247, https://doi.org/10.1145/967900.968153
82. Tominski C, Abello J, Schumann H (2005) Interactive poster: 3D axes-based visualizations for time series data. In: Poster at IEEE Symp. Information Visualization (InfoVis)
83. Tsuchiya T, Freeman J, Lerner LW (2006) Data-to-music API: real-time data-agnostic sonification with musical structure models. In: Proc. 21st International Conference on Auditory Display (ICAD 2015). Georgia Institute of Technology, pp 244–251
84. Vargha A, Delaney HD (2000) A critique and improvement of the CL Common Language effect size statistics of McGraw and Wong. J Educ Behav Stat 25(2):101–132. https://doi.org/10.3102/10769986025002101
85. Viau C, McGuffin MJ, Chiricota Y et al (2010) The FlowVizMenu and parallel scatterplot matrix: hybrid multidimensional visualizations for network exploration. IEEE Trans Vis Comput Graph 16(6):1100–1108. https://doi.org/10.1109/TVCG.2010.205
86. Ward J, Huckstep B, Tsakanikos E (2006) Sound-colour synaesthesia: to what extent does it use cross-modal mechanisms common to us all? Cortex 42:264–280. https://doi.org/10.1016/S0010-9452(08)70352-6
87. Ward MO, Grinstein G, Keim D (2015) Interactive data visualization: foundations, techniques, and applications, 2nd edn. AK Peters/CRC Press, Boca Raton
88. Wegman EJ (1990) Hyperdimensional data analysis using parallel coordinates. J Am Stat Assoc 85(411):664–675. https://doi.org/10.1080/01621459.1990.10474926
89. Wilcoxon F (1945) Individual comparisons by ranking methods. Biometrics Bull 1(6):80–83. https://doi.org/10.2307/3001968, https://www.jstor.org/stable/3001968
90. Wright M (2005) Open sound control: an enabling technology for musical networking. Organised Sound 10(3):193. https://doi.org/10.1017/S1355771805000932
91. Yang J, Hermann T (2018) Interactive mode explorer sonification enhances exploratory cluster analysis. J Audio Eng Soc 66(9):703–711. https://doi.org/10.17743/jaes.2018.0042
92. Yang J, Peng W, Ward M, et al (2003) Interactive hierarchical dimension ordering, spacing and filtering for exploration of high dimensional datasets. In: Proc. IEEE symp. information visualization. IEEE, pp 105–112, https://doi.org/10.1109/INFVIS.2003.1249015
93. Yi JS, Melton R, Stasko J et al (2005) Dust & magnet: multivariate information visualization using a magnet metaphor. Inf Vis 4(4):239–256. https://doi.org/10.1057/palgrave.ivs.9500099
94. Zwicker E, Fastl H (1999) Psychoacoustics: facts and models, springer series in information sciences, vol 22. Springer, Berlin

Authors and Affiliations

Elias Elmquist[1] · Kajetan Enge[2,3] · Alexander Rind[2] · Carlo Navarra[1] · Robert Höldrich[3] · Michael Iber[2] · Alexander Bock[1] · Anders Ynnerman[1] · Wolfgang Aigner[2] · Niklas Rönnberg[1]

Alexander Rind
alexander.rind@fhstp.ac.at

Carlo Navarra
carlo.navarra@liu.se

Robert Höldrich
robert.hoeldrich@kug.ac.at

Michael Iber
michael.iber@fhstp.ac.at

Alexander Bock
alexander.bock@liu.se

Anders Ynnerman
anders.ynnerman@liu.se

Wolfgang Aigner
wolfgang.aigner@fhstp.ac.at

Niklas Rönnberg
niklas.ronnberg@liu.se

1 Linköping University, Linköping, Sweden

2 St. Pölten University of Applied Sciences, St. Pölten, Austria

3 University of Music and Performing Arts Graz, Graz, Austria