It's about Time: Adopting Theoretical Constructs from Visualization for Sonification

Kajetan Enge
St. Pölten Univ. of Applied Sciences
Univ. of Music & Performing Arts Graz
Austria
kajetan.enge@fhstp.ac.at

Alexander Rind
St. Pölten Univ. of Applied Sciences
Inst. of Creative\Media/Technologies
Austria
alexander.rind@fhstp.ac.at

Michael Iber
St. Pölten Univ. of Applied Sciences
Inst. of Creative\Media/Technologies
Austria
michael.iber@fhstp.ac.at

Robert Höldrich
Univ. of Music & Performing Arts Graz
Inst. of Electronic Music & Acoustics
Austria
robert.hoeldrich@kug.ac.at

Wolfgang Aigner
St. Pölten Univ. of Applied Sciences
Inst. of Creative\Media/Technologies
Austria
wolfgang.aigner@fhstp.ac.at

ABSTRACT

Both sonification and visualization convey information about data by effectively using our human perceptual system, but their ways to transform the data could not be more different. The sonification community has demanded a holistic perspective on data representation, including audio-visual analysis, several times during the past 30 years. A design theory of audio-visual analysis could be a first step in this direction. An indispensable foundation for this undertaking is a terminology that describes the combined design space. To build a bridge between the domains, we adopt two of the established theoretical constructs from visualization theory for the field of sonification. The two constructs are the *spatial substrate* and the *visual mark*. 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. 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.

CCS CONCEPTS

• **Human-centered computing** → **Auditory feedback**; **Visualization theory, concepts and paradigms**; *Sound-based input / output*; • **Applied computing** → **Sound and music computing**.

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. Thus, a theory of sonification has already an articulated set of design constructs at its disposal [31]. However, we argue that constructs at a more basic level are missing from the current stage of scientific dialog.

This paper proposes *marks* in a *substrate* as basic constructs for designing sonifications. The theoretical model is adopted from visualization literature [3, 7, 30], where visual marks in a spatial substrate are widely used. 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., [30]), software tools (e.g., [45]) and toolkits (e.g., [41, 51]), as well as automatic recommendation of visualizations (e.g., [24, 29]).

Theoretical cross-pollination between visualization and sonification is most reasonable because both fields share very similar goals. While sonification is "the use of nonspeech audio to convey information" [21], visualization is defined as "the use of computer-supported, interactive, visual representations of abstract data to amplify cognition" [7] . Unsurprisingly, sonifications are often employed together with visualizations in real-world scenarios, such as by diagnostic ultrasonic devices. However, far too little attention has been paid to the theoretical underpinnings of audio-visual data analysis approaches [48]. Such approaches essentially use both our vision and our auditory sense in combination to convey information about data sets. A combined design theory with compatible basic constructs is even more reasonable as a step towards bridging terminological barriers between the research communities and making progress in both fields.

There are, however, fundamental differences between our visual and auditory perception [48]. For example, auditory perception is

less precise in space than visual perception is [4]. Sound is an inherently temporal phenomenon [10, 20, 42]. Therefore, adaptations of the model of marks and substrate are needed.

This paper starts with related work (Section 2) and an overview of marks in the visualization literature (Section 3). Section 4 investigates, how an equivalent mark and its substrate can be modeled in the sonification domain.

With this paper, we propose a new way to think about sonification design and, in the future, audio-visual representation of data.

2 RELATED WORK

There are numerous examples of designs that combine sonification and visualization. Hildebrandt et al. [18] combined visualization and sonification to analyze business process execution data. Rabenhorst et al. [38] augmented a vector field visualization with sonification. Chang et al. used an audio-visual approach to explore the activity of neurons in the brain [2]. In 2003, Hermann et al. presented 'AVDisplay' [17], a system for monitoring processes in complex computer network systems including both sonifications and visualizations. In 2007, MacVeigh and Jacobson described "a way to incorporate sound into a raster-based classified image." They augmented a classical map with further dimensions through sonification [26].

Taken together, the abovementioned works support the notion that visualization and sonification can be combined for effective data analysis. They, however, remain rather small steps on the road toward a combined design theory for audio-visual analysis. In the early 2000s, Nesbitt introduced a taxonomy for the multi-modal design space, which apparently did not have a lasting impact on the community [32–36]. Nesbitt proposed essentially two ways to describe the multi-modal design space, including haptic displays. The first one is an extension of the reference model for visualization by Card and Mackinlay [7], which we choose also as our reference in this paper. In his extended Card-Mackinlay 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 [34]. These categories take into account the inherent temporal structure of sound, which is not the case with the extended Card-Mackinlay design space. While Nesbitt introduced a new description of the multi-modal design space, in this paper, we suggest using time instead of space as the substrate of sonification and adopting the vocabulary from visualization theory.

Compared to visualization, sonification is a considerably younger discipline [11]. This might be one of the causes why its theoretical foundation is not as developed even though both disciplines pursue very similar goals. 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 [19], Barrass' dissertation in 1997 [1], the sonification report in 1999 [21], Walker's work on magnitude estimation of conceptual data dimensions in 2002 [47], Herman's dissertation on 'Sonification for Exploratory Data Analysis' [14], the book 'Ecological Psychoacoustics', edited by Neuhoff in 2004 [37], de Campo's design space map in 2007 [9], Hermann's taxonomy in 2008 [15], the 'Sonification Handbook' in 2011 [16], and
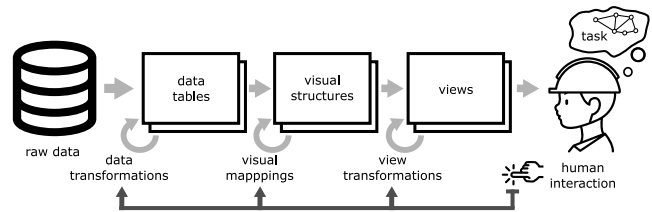

Figure 1: The reference model for visualization [7] introduces visual structures as an intermediate state in mapping data to visual representations. Reusing the icon "engineer" by Pawinee E. from Noun Project, CC BY 3.0.

Worrall's 'Sonification Design' in 2019 [52]. Nevertheless, in 2019 Nees [31, 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 the work by Gregor and Jones [12] as a possibility for the development of a sonification design theory. Gregor and Jones describe eight components that any design theory should include. "Constructs" are one of these eight components. The authors describe the constructs in [12, p.33]: "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 [31].

In this paper, we intend to contribute to the development of a design theory for sonification 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.

3 BASIC THEORETICAL CONSTRUCTS IN VISUALIZATION THEORY

Since the design space of possible visualization techniques is extensive, the visualization community has worked on theoretical models to formalize design knowledge [30]. Based on Bertin's seminal 'Semiology of Graphics' [3], many visualization models (e.g., [6, 7, 24, 30, 51]) are centered around marks as the basic building blocks of visualization techniques. In general terms, a mark is a geometric object that represents a data object's attributes by position, color, or other visual features.

The widely adopted reference model for visualization [7] 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 are *visual structures* that consist of marks 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 (Figure 1).

3.1 Defining visual structures

Next, we introduce the three components of a visual structure: a spatial substrate, marks, and channels.
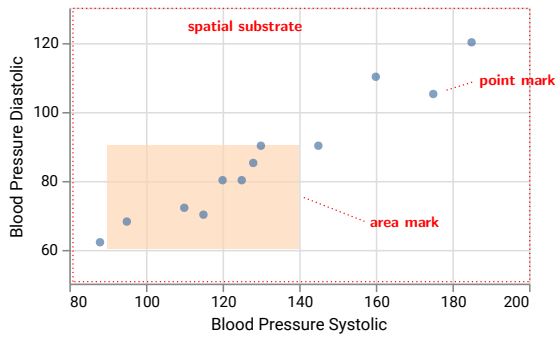
Figure 2: Example scatter plot with blood pressure measurements as *points* and a rectangle representing the *area* of normal systolic and diastolic blood pressure.

Figure 3: Example scatter plot using the channels size and shape. Note that rectangles and circles represent *point* marks.

Channels such as position and color encode the information of the data table's attributes into the visual features of the marks. 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 [30, 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" [7, p. 26]. Besides spatial position, Bertin [3] 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 [30]).

The *spatial substrate* is the container where marks are positioned in a conceptual space. 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 (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 Ⴑ) 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 (Figure 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.

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 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 perceptible.
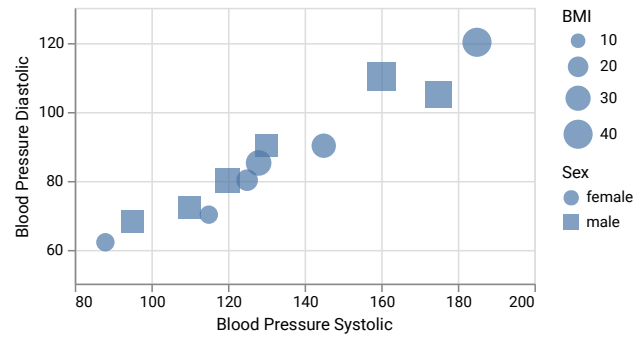
Since the spatial extent of a point mark does not convey information, the mark is not constrained and can use the channel size to encode another data attribute. Yet another data attribute can be mapped to the channel shape, so that one category is shown as square and another as circle (Figure 3). Neither the size nor the shape channel can be mapped to an area mark (cp. Figure 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 (Figure 3) or a data object spanning an area in the spatial substrate (Figure 2). To correctly interpret such graphics, contextual information is necessary that visualization designers need to provide via legends, annotations, or other onboarding approaches [44].

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 (Figure 2) and to invent completely new techniques. Several visualization software frameworks apply these constructs to specify the visual encoding: e.g., Tableau [45], ggplot2 [50], or Vega-Lite [41].

The usage 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 [3]. For example, proximity on the spatial substrate and similarity of the color channel can be perceived as a Gestalt. However, not every combination of marks, substrates and channels results in an effective representation of its underlying data. Yet, this conceptual model helps to systematically investigate the effectiveness of its components. For example, the experiments by Cleveland and McGill [8] 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 (e.g., [24, 30]) or integrated into tools (e.g., [24, 25, 29]).

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 for formalizing sonification techniques as well and, thus, pave 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. These constructs define the terminology that is 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", and describe their meta-meaning: 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 the conceptual expansion in their substrate. Channels are the parameters of a data representation encoded in a mark, carrying the information.

Next, this section investigates what the possible analogies for these constructs in sonification are. On the one hand, in sonification, the construct of channels is relatively familiar with parameters such as loudness, pitch, or timbre [16, 47]. On the other hand, the two constructs substrate and marks are not commonly used to describe a sonification. Since marks can expand conceptually within their substrate, these two constructs are closely intertwined. Visualization uses space as a substrate, so we first explore the potential of space as the substrate of sonification. However, the potential of time as the substrate of sonification has shown to be more promising, see subsection 4.2.

4.1 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 strongly characterized by their dimensionality. Generally, they can comprise 1-, 2-, and 3-dimensional representations defining the environment in which items can be displayed. In the field of audio reproduction, we commonly speak of mono-, stereo-, surround- and 3D-reproduction of signals which provide the adjustable dimensionality that is required as a pre-condition to qualify as a spatial substrate in visualization. Following this rationale, 0-dimensional mono sound sources correspond to point marks, 1-dimensional stereo sources to line marks, 2-dimensional surround sources to area marks, and 3D audio sources to volumetric marks. All of these sources could be embedded into spatial auditory substrates with equivalent or higher dimensionality.

What at first view seems to be a perfectly matching analogy reveals major drawbacks at second sight. 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 [4], Fastl and Zwicker [53], and Bregman [5]. 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 also are 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, 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 his 'Theory of Indispensable Attributes,' he states that it is not the direction that helps us identify an auditory object, but its temporal and spectral properties [22, 23].

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

4.2 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 in support of time as the substrate of sonification. Time is a dimension inherently necessary to perceive sound. When we think of headphones that project a sound wave directly into our ear canal, space is not conceptually necessary to perceive sound. Time, on the other hand, is a dimension that we cannot even conceptually switch off while listening. Just as space is not necessary to convey information via sonification, most visualization designs do not use time as a dimension [30]. Thus, in a static visualization, we can think of time as conceptually "switched off". Albeit visual perception is a construction process over a certain amount of time the visualization itself is not changing over time. Within sonification, one can think of sounds being "positioned in time", just like visual marks are positioned in space. In visualization, we can localize multiple spatially distributed visual marks, even when they look identical, hence they use the same non-positional channel values. In sonification, on the other hand, we cannot necessarily localize several identical sound sources when they are presented simultaneously. Also, 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 [4].

For these reasons, we consider time to be a suitable substrate for sonification and refer to it as the "temporal substrate." We should clarify that the temporal substrate is only a subset of time itself. The temporal substrate refers to the period of time that passes during a sonification, just as the spatial substrate in visualization refers only to the subset of physical space available to a visualization. For the temporal substrate, it is not relevant whether the sonification is just listened to or whether somebody interacts with it. 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.
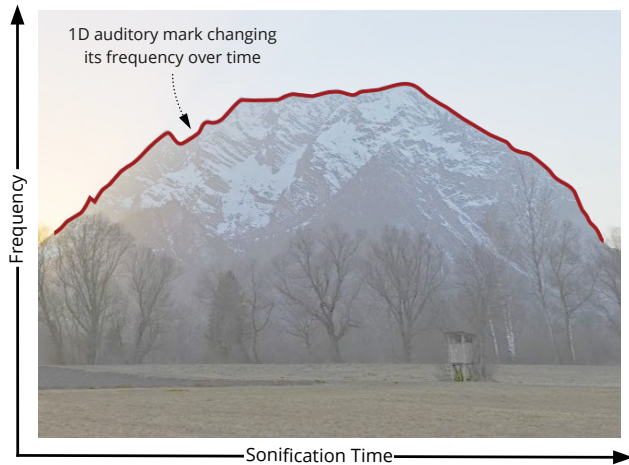
Figure 4: The silhouette of the mountain Grimming in Austria. A 1D auditory mark maps the horizontal positions of the silhouette to time, and the height of the silhouette to the frequency of a sine. The horizontal positions correspond to the sortable attributes k and the height values to the attributes x from Figure 5 and Equation 3. Function $g(x_i)$ from Equation 3 maps the height values x_i to the time dependent channel $\mathring{c}(\mathring{t}_i)$, the frequency of the sine.

Figure 5: An unsorted data set is sorted and sonified to an 1D auditory mark, evolving over sonification time.

We know that visualization theory distinguishes its visual marks by their conceptual dimensionality, i.e. their conceptual expansion within the spatial substrate. As has been shown, conceptual expansion is not equal to physical expansion. Visual marks need to occupy space to become visible, even if conceptually they do not expand [3]. Correspondingly, we want to be able to distinguish auditory marks by their conceptual expansion within 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" [7]. They represent all the possibilities for conceptual spatial expansion from 0D (no conceptual expansion) up to 3D (maximum conceptual expansion). While space is three-dimensional, time is one-dimensional. Thus, we define auditory marks that are 0D (no conceptual expansion) or 1D (maximum conceptual expansion). There are no 2D or 3D auditory marks, since time does not provide second and third dimensions. We consider an auditory mark as 0D if it does not conceptually expand in time just like 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. 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.

4.2.1 1D auditory mark: A 1D auditory mark represents the data via a development over time. More precisely: The *temporal evolution*
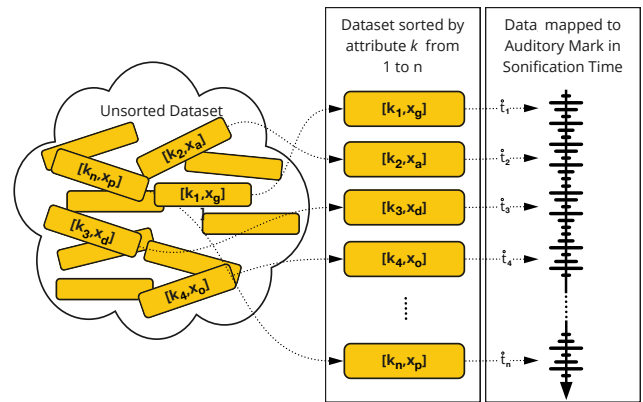
of a 1D auditory mark represents a dataset along one of the set's sorted attributes. It does so by evolving its channel(s) over time according to the sort, thus representing the evolution of attributes in the dataset. We regard the 1D auditory mark as "conceptually expanded in time" as it conveys information over time. The sorted attribute has to be a key attribute. A key attribute is a unique identifier for all of the items in a dataset. In a table, it could be, for example, the row number. This ensures that every item in the dataset gets mapped to time bijectively.

An example of such a 1D auditory mark is shown in Figure 4 via the silhouette of a mountain as a red line. Imagine a parameter mapping sonification [13], conveying information about the shape of the silhouette. The sonification maps the horizontal and the vertical positions of the silhouette to the time and the frequency 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 this case, we usually speak of an auditory graph as a special version of a parameter mapping sonification [27, 43]. In this example, the sonification uses a one-dimensional auditory mark as its channel (frequency) evolves over time according to the development of the vertical position sorted along the horizontal position in the dataset.

We can describe the one-dimensional auditory mark in a more general mathematical way: When we discuss the mathematical description, we can still use our silhouette example as a reference. Think of a dataset that holds many items with at least two attributes each. Figure 5 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 it is a unique identifier that can be used to look up all the items in a dataset [30].

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

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 Equation 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 the $\mathring{t}$ 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 [40], and then developed further by Vogt and Höldrich [46]. 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 sonification variable frequency (compare [13, p. 368]). To be called a sonification, this transformation needs to be systematic, objective, and reproducible [15].

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

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 sonification time $\mathring{t}$ and the time-dependent channel $\mathring{c}(\mathring{t})$:

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

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 for the 0D auditory mark. Every sonification has to expand in time, but not all of them convey information over time. Mathematically speaking, $\mathring{y}$ always depends on $\mathring{t}$, but $\mathring{c}$ does not have to depend on $\mathring{t}$. Auditory icons and Earcons, for example, are sonification techniques that convey information without an inherent dependency on developments in the data [16]. They usually inform their users about states.

4.2.2 0D auditory mark: 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. This can be the case if, for example (1), there is no sortable attribute in the data, or if (2) the sorted data set is not mapped to sonification time. For further explanation, we construct two examples.

Table 1: Substrates and Mark Types

Domain	Substrate	Mark Types
Visualization	Space	0D: Point 1D: Line 2D: Area 3D: Volume
Sonification	Time	0D: State in time 1D: Development over time

A so-called "Earcon" [28] can typically be described as a 0D auditory mark. The sound your computer makes when an error occurs 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 moment 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 sorted data items to frequency instead of time would also result in a 0D auditory mark. To explain this, we can re-use the silhouette example from before. The abscissa in Figure 4 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.

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 Equation 5 and Equation 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.

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

5 PARALLELS BETWEEN VISUALIZATION AND SONIFICATION

Using time as the substrate of sonification and defining marks to conceptually expand in time reveals several parallels 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.

A parallel 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 channels, but they do not have to be. In sonification, the moment and duration of an auditory mark can be channels, but they also do not have to be. In both domains, these parameters do not define the type of the mark. The type depends on the conceptual expansion in their substrate.

It is another parallel between visualization and sonification that information can be encoded both in the marks, and in Gestalten [49] that reveal themselves through a group of marks with related channels. The correlation of two data sets 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, in both domains, 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 Gestalt of the dots suggests a line phenomenon. The same applies to sonification. In granular synthesis, [39] the close positioning of many grains (0D) can trigger the perception of one continuously developing sound, hence of a 1D auditory mark.

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. If we saw a green dog, we would not separately perceive the dog and the attribute "greenness". The attribute belongs to the object [5]. Bregman [5, p. 11] states that "the stream plays the same role in auditory mental experience as the object does in visual." 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.

Last but not least, just like visualization needs to deal with spatial clutter, sonification needs to deal with temporal masking.

6 CONCLUSION

This paper provided an overview of fundamental theoretical constructs from visualization theory and adopted two of 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, and they are the second theoretical construct that has been adopted for the field of sonification. The construct of channels has not been adopted in this paper. Our work shows that time qualifies as the substrate of sonification, we therefore call it temporal substrate. Just like visual marks have positions in space, auditory marks have positions in time. We also investigated the possibility to use space as a substrate for sonification but rejected the model due to several drawbacks regarding spatial auditory perception. With time as the substrate of sonification, many parallels to visualization theory reveal themselves. One parallel is the possibility to think of marks as conceptually expanded in their substrate.

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. Furthermore, our work introduces new terminology to describe sonifications in general. It can also feed back into visualization theory with regard to the temporal description of data visualizations.

Our next step will be to closely investigate the possible channels in a combined audio-visual design space.

ACKNOWLEDGMENTS

REFERENCES

[1] Stephen Barrass. 1997. *Auditory Information Design*. Ph.D. Dissertation. Australian National University, Canberra. https://openresearch-repository.anu.edu.au/bitstream/1885/46072/16/02whole.pdf
[2] Jonathan Berger, Ge Wang, and Mindy Chang. 2010. Sonification and Visualization of Neural Data. In *Proceedings of the 16th International Conference on Auditory Display (ICAD-2010)*. Georgia Institute of Technology, 201–205.
[3] Jacques Bertin. 1983. *Semiology of Graphics: Diagrams Networks Maps*. University of Wisconsin, Madison. Originally published in 1967 in French.
[4] Jens Blauert. 1996. *Spatial Hearing: The Psychophysics of Human Sound Localization*. MIT press.
[5] Albert S. Bregman. 1990. *Auditory Scene Analysis: The Perceptual Organization of Sound*. MIT press.
[6] Stuart K. Card and Jock Mackinlay. 1997. The Structure of the Information Visualization Design Space. In *Proc. IEEE Symp. Information Visualization, InfoVis*. 92–99. https://doi.org/10.1109/INFVIS.1997.636792
[7] Stuart K. Card, Jock Mackinlay, and Ben Shneiderman (Eds.). 1999. *Readings in Information Visualization: Using Vision to Think*. Morgan Kaufmann, San Francisco.
[8] William S. Cleveland and Robert McGill. 1984. Graphical Perception: Theory, Experimentation, and Application to the Development of Graphical Methods. *J. American Statistical Association* 79, 387 (1984), 531–554.
[9] Alberto de Campo. 2007. Toward a Data Sonification Design Space Map. In *Proceedings of the 13th International Conference on Auditory Display*. Georgia Institute of Technology.
[10] David Freides. 1974. Human information processing and sensory modality: Cross-modal functions, information complexity, memory, and deficit. *Psychological Bulletin* 81, 5 (1974), 284–310.
[11] Steven P. Frysinger. 2005. A Brief History of Auditory Data Representation to the 1980s. In *Proceedings of ICAD 05-Eleventh Meeting of the International Conference on Auditory Display*. Georgia Institute of Technology.
[12] Shirley Gregor and David Jones. 2007. The Anatomy of a Design Theory. *Journal of the Association for Information Systems* 8, 5, Article 1 (2007).
[13] Florian Grond and Jonathan Berger. 2011. Parameter Mapping Sonification. In *The Sonification Handbook*, Thomas Hermann, Andy Hunt, and John G. Neuhoff (Eds.). 363–397.
[14] Thomas Hermann. 2002. *Sonification for Exploratory Data Analysis*. Ph.D. Dissertation. Bielefeld, Germany.
[15] Thomas Hermann. 2008. Taxonomy and Definitions for Sonification and Auditory Display. In *Proceedings of the 14th International Conference on Auditory Display*.
[16] Thomas Hermann, Andy Hunt, and John G. Neuhoff (Eds.). 2011. *The Sonification Handbook*. Logos, Bielefeld.
[17] Thomas Hermann, Christian Niehus, and Helge Ritter. 2003. Interactive Visualization and Sonification for Monitoring Complex Processes. In *Proceedings of the 2003 International Conference on Auditory Display*.
[18] Tobias Hildebrandt, Felix Amerbauer, and Stefanie Rinderle-Ma. 2016. Combining Sonification and Visualization for the Analysis of Process Execution Data. In *2016 IEEE 18th Conference on Business Informatics (CBI)*, Vol. 2. 32–37.
[19] Gregory Kramer (Ed.). 1994. *Auditory Display: Sonification, Audification and Auditory Interfaces*. Addison-Wesley, Reading, Mass.
[20] Gregory Kramer. 1994. Some Organizing Principles for Representing Data with Sound. In *Auditory Display: Sonification, Audification and Auditory Interfaces*, Gregory Kramer (Ed.). Addison-Wesley, Reading, Mass, 185–221.
[21] Gregory Kramer, Bruce Walker, Terri Bonebright, Perry Cook, John H Flowers, Nadine Miner, John Neuhoff, et al. 1999. Sonification Report: Status of the Field and Research Agenda. (1999).
[22] Michael Kubovy. 1981. Concurrent-Pitch Segregation and the Theory of Indispensable Attributes. In *Perceptual Organization*. Routledge, 55–98.

[23] Michael Kubovy and David Van Valkenburg. 2001. Auditory and visual objects. *Cognition* 80, 1-2 (2001), 97–126. https://doi.org/10.1016/S0010-0277(00)00155-4

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

[25] Jock D. Mackinlay, Pat Hanrahan, and Chris Stolte. 2007. Show Me: Automatic Presentation for Visual Analysis. *IEEE Trans. Visualization and Computer Graphics* 13, 6 (2007), 1137–1144. https://doi.org/10.1109/TVCG.2007.70594

[26] Ryan MacVeigh and R. Daniel Jacobson. 2007. Increasing the dimensionality of a geographic information system (GIS) using auditory display. In *Proceedings of the 13th International Conference on Auditory Display*.

[27] Douglass L Mansur, Merra M Blattner, and Kenneth I Joy. 1985. Sound graphs: A numerical data analysis method for the blind. *Journal of medical systems* 9, 3 (1985), 163–174.

[28] David McGookin and Stephen Brewster. 2011. Earcons. In *The Sonification Handbook*, Thomas Hermann, Andy Hunt, and John G. Neuhoff (Eds.). 339–361.

[29] Dominik Moritz, Chenglong Wang, Gregory Nelson, Halden Lin, Adam M. Smith, Bill Howe, and Jeffrey Heer. 2018. Formalizing Visualization Design Knowledge as Constraints: Actionable and Extensible Models in Draco. *IEEE Trans. Visualization and Computer Graphics* 25, 1 (2018), 438–448. https://doi.org/10.1109/TVCG.2018.2865240

[30] Tamara Munzner. 2015. *Visualization Analysis and Design*. CRC Press.

[31] Michael A. Nees. 2019. Eight Components of a Design Theory of Sonification. In *Proceedings of the 25th International Conference on Auditory Display (ICAD 2019)*. 176–183. https://doi.org/10.21785/icad2019.048

[32] Keith V. Nesbitt. 2000. A Classification of Multi-Sensory Metaphors for Understanding Abstract Data in a Virtual Environment. In *Proc. IEEE Conf. Information Visualization (IV)*. 493–498. https://doi.org/10.1109/IV.2000.859802

[33] Keith V. Nesbitt. 2004. MS-Taxonomy: a conceptual framework for designing multi-sensory displays. In *Proc. Eighth International Conference on Information Visualisation, 2004. IV 2004*. 665–670. https://doi.org/10.1109/IV.2004.1320213

[34] Keith V. Nesbitt. 2006. Modelling Human Perception to Leverage the Reuse of Concepts across the Multi-Sensory Design Space *(APCCM '06)*. Australian Computer Society, Inc., Australia, 65–74.

[35] Keith V. Nesbitt and Stephen Barrass. 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.

[36] Keith V Nesbitt and Stephen Barrass. 2004. Finding Trading Patterns in Stock Market Data. *IEEE Computer Graphics and Applications* 24, 5 (2004), 45–55. https://doi.org/10.1109/MCG.2004.28

[37] John G. Neuhoff (Ed.). 2004. *Ecological Psychoacoustics*. Elsevier Academic Press.

[38] David A Rabenhorst, Edward J Farrell, David H Jameson, Thomas D Linton Jr, and Jack A Mandelman. 1990. Complementary visualization and sonification of multidimensional data. In *Extracting Meaning from Complex Data: Processing, Display, Interaction*, Vol. 1259. International Society for Optics and Photonics, 147–153.

[39] Curtis Roads. 2001. *Microsound*. MIT Press, Cambridge, Mass.

[40] Julian Rohrhuber. 2010. S–Introducing sonification variables. In *Proceedings of the Supercollider Symposium*.

[41] Arvind Satyanarayan, Dominik Moritz, Kanit Wongsuphasawat, and Jeffrey Heer. 2017. Vega-Lite: A Grammar of Interactive Graphics. *IEEE Trans. Visualization and Computer Graphics* 23, 1 (2017), 341–350. https://doi.org/10.1109/TVCG.2016.2599030

[42] Norman Sieroka. 2018. *Philosophie der Zeit: Grundlagen und Perspektiven*. Vol. 2886. CH Beck.

[43] Tony Stockman, Louise Valgerour Nickerson, and Greg Hind. 2005. Auditory graphs: A summary of current experience and towards a research agenda. In *Proceedings of the 11th International Conference on Auditory Display*, Eoin Brazil (Ed.). Georgia Institute of Technology, 420–422. https://smartech.gatech.edu/handle/1853/50097

[44] Christina Stoiber, Florian Grassinger, Margit Pohl, Holger Stitz, Marc Streit, and Wolfgang Aigner. 2019. Visualization Onboarding: Learning How to Read and Use Visualizations. In *IEEE Workshop on Visualization for Communication*. OSF Preprints. https://doi.org/10/gh38zd

[45] Chris Stolte, Diane Tang, and Pat Hanrahan. 2002. Polaris: A System for Query, Analysis, and Visualization of Multidimensional Relational Databases. *IEEE Trans. Visualization and Computer Graphics* 8, 1 (2002), 52–65. https://doi.org/10.1109/2945.981851

[46] Katharina Vogt and Robert Höldrich. 2012. Translating Sonifications. *Journal of the Audio Engineering Society* 60, 11 (2012), 926–935.

[47] Bruce N Walker. 2002. Magnitude estimation of conceptual data dimensions for use in sonification. *Journal of Experimental Psychology. Applied* 8, 4 (2002), 211.

[48] Bruce N. Walker and Gregory Kramer. 2004. Ecological Psychoacoustics and Auditory Displays: Hearing, Grouping, and Meaning Making. In *Ecological Psychoacoustics*, John G. Neuhoff (Ed.). Elsevier Academic Press, 150–175.

[49] Max Wertheimer. 1923. Untersuchungen zur Lehre von der Gestalt. II. *Psychologische Forschung* 4, 1 (1923), 301–350.

[50] Hadley Wickham. 2010. A Layered Grammar of Graphics. *Journal of Computational and Graphical Statistics* 19, 1 (2010), 3–28. https://doi.org/10.1198/jcgs.2009.07098

[51] Leland Wilkinson. 2005. *The Grammar of Graphics* (second ed.). Springer.

[52] David Worrall. 2019. *Sonification Design: From Data to Intelligible Soundfields*. Springer, Cham. https://doi.org/10.1007/978-3-030-01497-1

[53] Eberhard Zwicker and Hugo Fastl. 1999. *Psychoacoustics: Facts and Models*. Springer Series in Information Sciences, Vol. 22. Springer.