

# Seeing Beyond Sound: Visualization and Abstraction in Audio Data Representation

Final Draft 6

Ashlae Blum'e

PhD Candidate, Math Department Victoria University of Wellington

[a.h.blum@gmail.com](mailto:a.h.blum@gmail.com)

Keywords: Audio information, data visualization, audio signal processing, design philosophy, human-computer interaction, bioacoustics, interface design, software design.

L'étude profonde de la nature est la source la plus féconde de découvertes mathématiques.

Jean-Baptiste Joseph Fourier

« Et voici que, à la fin du vingtième siècle, sous la pression de l'information moderne et sous l'effet de l'évolution des techniques de traitement, la graphique franchit une étape nouvelle et décisive. La grande différence entre la représentation graphique d'hier, mal dégagée de l'image figurative, et la graphique de demain, c'est la disparition de la fixité congénitale de l'image.

Lorsque l'on peut superposer, juxtaposer, transposer, permuer des images graphiques de façon à obtenir des groupements et des classements, l'image graphique quitte l'image morte, "l'illustration" pour devenir l'image vivante, l'instrument de recherche à la portée de tous qu'elle est en train de devenir. La graphique n'est plus seulement la "représentation" d'une simplification finale, elle est le point de départ de la découverte de ces simplifications et le moyen de leur justification. »

Jacques Bertin, *Semiologie Graphique*, 1967

"And now, at the end of the twentieth century, under the pressure of modern information and the advances of data processing, graphics is taking a new and decisive step. The great difference between the graphic representation of yesterday, which was poorly separated from the figurative image, and the graphics of tomorrow, is the disappearance of the congenital fixity of the image.

When one can superimpose, juxtapose, transpose, and permute graphic images in such a way as to obtain groupings and classifications, the graphic image leaves the dead image, the 'illustration,' to become the living image, the research tool within everyone's reach that it is becoming. Graphics is no longer only the 'representation' of a final simplification, it is the point of departure for the discovery of these simplifications and the mode of their justification."

0. Abstract: See – Smell – Taste – Touch – Hear. A picture is worth a thousand words. Seeing is believing.

The interpretation of complex data is epistemically linked to human perceptual frameworks. In audio information research, sound is represented and transformed using visual elements that highlight abstract patterns detached from the physical experience of sonic perception. Fields such as bioacoustics, music information retrieval, and auditory science employ a wide array of tools to transform theoretical knowledge into applied science. However, these tools carry with them hidden assumptions that have become masked through their adoption to new scientific contexts. Many such tools have inherited domain-specific technical conventions that nod to their historical origins in experimental media, entertainment, communications, and defense research. Cognitive psychology suggests that the way information is presented strongly influences an individual's ability to form complex associations. We argue that re/designing tools to align with emergent needs of modern users will improve both analytical as well as creative outputs due to an increased affinity for using them. This paper

explores the potentials associated with adding dimensionality back into visualizations to facilitate positive social effects in the use of audio information visualization tools.

## I. Introduction

Advanced data visualization techniques enable scientists to interpret complex datasets by transforming high-dimensional data and metadata into abstract visual elements. This serves not only to reveal patterns in information, but also to build narratives that enhance our collective understanding of the world around us. Such representations are often mediated by software and tools designed for specific domains, embedding assumptions that, while optimized in one context, may inhibit another. For audio data, waveforms and spectrograms form the basis of our visual knowledge. These rely on two-dimensional visualizations of the time-frequency domain that are mathematically well-defined, but often lack intuitive correspondence with the multisensory nature of auditory perception. The advent of the digital audio workstation (DAW) provided users a familiar template for audio interaction. With origins in the software revolution of the 1970s, its design elements persist in today's interfaces that span from the film industry to scientific research. More recently, the rise of programming literacy and the expansion of audio research have evolved alongside the need and interest in low-level control. Libraries such as Librosa (Python), Web Audio API (JavaScript), and tuneR (R) have arrived on the scene, with enthusiastic online userbases that connect communities across the internet, and the world. The broadening scope of creative coding has bridged science and art to expand the worlds of the technical and the expressive into expansive layers of abstraction; apps and games built to facilitate music-making and sound exploration proliferate; sound art and sound design are now well-established as legitimate commercial fields. In short, the spectrum of use cases in which audio is being transformed from numbers into something else is ever-expanding, and so, too, must the ways in which we interact with it.

## II. A Brief History of Audio Visualization

Modern audio analysis software is an amalgamation of design principles, applied scientific theory, and physical constraints that has been continuously refined over the last century or so. Early hardware inventions that modeled sound signals were built using analog electronics to implement theoretical concepts from harmonic and spectral analysis. Ranging from exploratory to practical, these devices were physical embodiments of the understanding of sound as a medium of the times. Due to their inherent physicality, they also carried with them necessary limitations and operational conventions that have persisted in the shift from analog to digital audio analysis. In today's software, such assumptions are now often overlooked, as analog origins have largely been superseded by their digital descendants. DAW-like analysis software, such as Audacity, Raven, and Sonic Visualiser, are but some examples at the heart of audio workflows that propel scientific inquiry. However, the presets embedded in such tools often assume specific use cases. Without knowledge of their existence, it can be easy to generate results using numerical parameters intended for another domain. To better assess the contemporary landscape, we first review the historical origins of modern audio visualization tools.

### 1. The Steampunk Origins of Sound Science

The steampunk origins of sound science grew out of the electromechanical age (1830-1940), in which the use of electricity to process and transform information revolutionized all aspects of society. Fourier's seminal works on harmonic analysis (1807, 1822) [1,2] had laid the mathematical foundations for audio signal processing, yet practical applications of these theories took time to crystallize. The earliest mechanical devices to record and play sound were the phonautograph in 1857 [3] and the phonograph (1877) [4], both of which were powered by hand. The phonautograph recorded sound waves by etching them on glass or paper [3]. The phonograph etched its sounds on tin foil, and could additionally play back audio from the etching [4]. The invention of the telegraph (1837) [5,6] marked a transition as the first electric device to transmit sound, followed by the telephone (1876) [7], gramophone (1887) [8], loudspeaker (1925) [9], and sound spectrograph (1946) [10,11]. These devices were built from mechanical and analog electric components, their design and use constrained by both material and human limitations. Friction and inertia of mechanical parts, short-circuits, and overheating are but some of the factors that impacted their smooth operation. These limitations were far from hidden; they were explicit, tactile, and fundamentally affected the user's interaction with and interpretation of sound.

## 2. Theoretical Foundations: Let's Get Digital

The development of the Fast-Fourier Transform (FFT) in 1965 [12] formed the backbone of signal processing algorithms as digital computing became ubiquitous through the rest of the century, and beyond. FFT-based methods impacted a wide variety of industries, for example, telecommunications (DSL modem [13,14], cell phones [15,16]), medicine (MRI [17,18], EEG [19,20]), and music (reverb [21,22], phase vocoder [23]), their implementations often remaining domain-specific. Now essential, Digital Signal Processing (DSP) algorithms form the building blocks of audio analysis software, and are intricately linked to our fundamental understanding of sound. Yet they, too, are built upon necessary parametric limits and assumptions inherited from their origins. For example, there is always a tradeoff of knowable information about a signal, as described by the Gabor uncertainty principle. This places a lower limit on the amount time-frequency uncertainty, and affects nfft and hop length parameter choices. Music production, speech analysis, and sonar engineering serve as specific examples where innovative uses of DSP algorithms left a lasting impact. Many of these pivotal technologies were gradually incorporated into the greater lexicon of digital audio analysis software, where they now live side-by-side as part of an unassuming digital toolkit.

## 3. How We Interact With Sound: Interface Design, Use Cases, and The Rise of the DAW

Like many of its digital audio counterparts, the modern DAW was also originally a piece of hardware. Arguably, the first DAW was the Soundstream Digital Editing System (1977), which operated on a minicomputer that ran custom software called the Digital Audio Processor (DAP). It was used to edit master tapes in the commercial audio industry, and featured hard disk recording, an interactive screen for waveform editing, and both analog and digital interfaces. The Fairlight CMI (1979) was another groundbreaking technology that became famous for its “Page R” sequencing environment, displaying rows of blocks that represented notes and audio – a precursor to today’s MIDI sequencing capabilities. Text-based DAWs of the 1980s, such as the Commodore 64 and Keyboard Computer System (KCS), supported multiple MIDI tracks using lists and drop-down menus. The visual layout of the Steinberg Pro-16 (1987), developed for the Atari, was the predecessor to today’s DAW interface. It was a MIDI sequencer that visually looked like physical hardware mixing consoles, complete with playback and routing controls, and horizontal arrangement views to allow for multiple perspectives within the audio environment. This took the abstract concept of sequencing, previously done using manual list entry, and made it look and feel like working with physical audio hardware. Since computer processors at the time could not yet support multi-track recording or playback, these early workstations were MIDI-only. Computers became more powerful throughout the 1990s as the semiconductor industry enabled processor technology to become cheaper, faster, and smaller. Consequently, computationally-expensive audio functions such hard-disk audio processing could live side-by-side with sequencing. Prominent examples include Sound Tools (1992), with its limited audio recording; Cubase (1992), with its MIDI and audio visible in the same interface; and the invention of the Virtual Studio Technology (VST, 1996) plugin, which allowed digital effects to be applied to individual channels.

## 4. How We Perceive Sound: Sensory, Perceptual, and Cognitive Considerations

How we view sound is profoundly affected by not only the physical experience of perceiving an image on a screen, but by a broader sense of cognition and perception about its fundamental nature. For most humans, sound is one of five core senses we experience throughout our lives. Our relationship with it changes as we age, and as we add information to our sensory network through lived experiences. A number of tools are used to visualize sound, some of which strive to depict spatialized relationships between its components, and others which employ layers of abstraction to expand its sphere of perceptible information. Oscilloscopes plot time-amplitude waveforms by reading the voltage from a transducer (microphone) to display pressure oscillations. A spectrogram uses the Short-Time Fourier Transform (STFT) to sum windowed segments of a signal, trading temporal precision for frequency resolution: lower time-resolution allows the calculation of finely-grained frequency evolution, and vice-versa. Mel-Frequency Cepstral Coefficients (MFCCs) represent spectral energy as a series of coefficients scaled exponentially to align with the human auditory system. These types of audio tools

are optimized for quantitative feature extraction, however, they can obscure higher-order perceptual structures such as timbre, microtonality, or the fullness of polyphonic sound.

5. One major challenge in data visualization is mapping high-dimensional features to visual variables in a way that intuitively makes sense when you look at it. Many tools from statistics, such as scatter plots and time-series graphs, are precise and well-established, yet they require an input of low-dimensional data. Audio features, which are highly multidimensional (e.g. dozens of MFCCs, spectral and temporal centroids, entropy scores), require correspondingly advanced encodings. There are innovative efforts across many domains that strive to expand and explore the nature of data visualization, and to unify multidimensional and interactive visualizations with cognition. The adoption of topological data analysis can reveal the underlying shape of a dataset of whale songs. [1] Activation maps in convolutional neural networks (CNNs) trained on audio show the features a model learns to detect. [2] Through exposure, use, and familiarization, these visual innovations have become part of standard audio data visualization workflows.

6. As experimental graphics research continues to push the boundaries of technology, media domains such as virtual reality (VR), augmented reality (AR), and 360 video offer expanded formats for multisensory immersion. These technologies, often referred to as experiences, prioritize full sensory immersion and interactivity. While VR gaming centers and AR media experiences have become increasingly common, their uses outside of entertainment have been slower to catch on. Examples include the use of haptics for blind or deaf people, VR for therapy and training, and 3D sculpture as a tool for design. For audio, 3D time-frequency embeddings attempt to visualize timbral similarity by projecting features into a spatial manifold. This embedding allows researchers to see clusters of similar bird calls or phonetic units [3]. Similarly, sonic labyrinths use interactive 3D structures to represent sound, where navigation corresponds to spectral exploration [4]. From science to media, innovations in audio data visualization proliferate as technology facilitates the accessible transformation of multisensory information.

### III. Addressing Specific Knowledge Gaps

Using 1) the setup from the historical origins, 2) the transition from physical and analog to perceptual and digital, and 3) an overview of physical limitations and inherited conventions, we now illustrate some very specific examples found in software that illustrate these limitations.

#### a. Hidden assumptions: software as a black-box

The metaphor of the black-box comes from the aviation industry, where flight data recorders in airplanes were housed in a literal metal box that had been painted black. These were comprised of analog and electromechanical components that engraved flight metrics onto metal foil [5]. The black-box metaphor has since become an analogy for the study of a closed system without prior knowledge of its inner workings, relying solely on knowledge of input, and observation of output, to evaluate its structure and evolution [6].

Comprising anywhere from hundreds to ten-thousands of lines of code and more, it becomes necessary to treat software as a black-box, or we would never get anything done. Since code is more often read than it is written, especially for free, libre, and open-source software (FLOSS), it is seen as a best practice to leave a clear, well-documented paper trail in the form of in-line notes, for posterity. Along with a (hopefully) clear set of instructions on how to use the software, these notes, known colloquially as documentation, are essential so that others who use it thereafter can follow the design and flow of logic, and possibly to understand features that may be only partially implemented, or future scaling intentions. This facilitates not only a deeper understanding of such tools, but also the ability to change, edit, or repurpose the software for either similar, or far-flung and imaginative notions (use cases?). Also, in an area of intensive development where people are often working independently, documentation serves as a form of communication and connectedness between developers who may never meet each other in real life, adding an additional layer of cognition aside from just a functional or utilitarian need.

b) Parameters, presets, and preconceived notions.

Design transparency openly acknowledges such choices, providing access to customization that may liberate the user from the constraints of domain-specific applications. Knowledge of equations from signal processing, population dynamics, or neuroscience can allow for backtracking through lines of dense programming languages. Portability and translateability are also facilitated by transparency, since at times one can simply replace one equation with another or add it to a centralized dictionary of options. The forms that such equations often take (in the code) are direct, if dense, translations into formal logic through layers of abstraction known as standard software libraries (e.g. numpy, librosa, fftw). As with all equations that govern the empirical sciences, numerical parameters must be chosen to allow mathematical computation to occur. This is the starting point, from which it is assumed that values will be changed to suit the particular needs of a specific application at-hand. However, as meta-uses compound, this implied reliance on presets or parameters can become buried, obscured, or forgotten. Therein runs a risk of making assumptions that may not be appropriate for a specific domain's application. In the following section, we focus primarily on a comparison of FLOSS tools and their hardcoded assumptions that have been noticed firsthand while reading through source code. See Appendix for a more complete list of audio-specific software and libraries that incorporate presets.

- Praat was developed specifically to study the human voice, and has pre-emphasis filtering that boosts frequencies above 50 Hz. This alters the relationship between frequency content in the signal, and can be problematic for many animals that communicate using low-frequency information. Specific examples include elephants, rhinos, whales, large ungulates, big cats, bears, wolves, seals, sea lions, manatees, and some fish. [appendix?]

- Praat's preset limits the visual display of audio clips greater than a certain specific duration of time.

- More fully-featured software, such as Audacity, Sonic Visualiser, Avisoft (proprietary), and Raven (proprietary), represent a spectrum of graphical DAW-like tools that have developed specialized use cases in audio information domains. Their workflows are rooted in temporal manipulation, which is often (but not always) a stepping-stone in audio information science. For example, the purpose of cutting audio at annotation points is to then perform other calculations on that audio slice, i.e. feature extraction.

- Horizontal vs. vertical layouts are tied to workflows from the audio recording industry. For scientific use cases, comparing many small files along horizontal timelines feels clunky when looking to broadly assess their similarities and differences. This is different from when we want to view the audio as a time sequence, where (horizontal) temporal continuity may be useful.

- Interacting with all files (or annotated slices) at once can be labor-intensive, often requiring manual interaction with each one. There is not always a way to batch import many files vertically along independent channels. Files may be required to be loaded individually, or the batching of such files might be for a calculation or analysis that is hidden in the software's algorithms.

- If batch loading and viewing is indeed possible, interacting with all files simultaneously can require the manual labor of clicking each single track to turn such a feature on. Repetitive clicking with a mouse or trackpad is not physically ergonomic and can cause physical harm over time if done too frequently.

- A need for effects batching further exemplifies the manual-selection issue. If, for example, a bandpass filter is required to eliminate some machine noise or a natural event such as an earthquake, it is far more efficient to apply this same effect to all files at the same time. Instead, one might have to select a checkbox or button, or add a VST device onto every single audio channel – a task that, when required for thousands of files, quickly becomes tiresome.

- In scikit-maad, a 4th-order Butterworth (infinite-impulse response) filter is the preset for automated feature and region of interest (roi) selection. It preferences frequency precision with a flat passband and -24dB/octave

rolloff, but limits temporal precision due to its phase-nonlinearity. Since different frequency components of a signal travel at different rates, this type of filtering shifts the timing of low- and high-frequency information differently within the same acoustic event. The infinite filter response can also create acausal pre-event artifacts that interfere with the detection of onset transients. To mitigate this, `maad` defaults to the zero-phase `filtfilt`, but this choice is inappropriate when high temporal precision is needed. Examples include measuring intervals between syllables (such as echolocation clicks), sample-accurate onset detection, or fine-scale waveform comparison. Using `scipy.signal` can allow for more finely-grained control.

- Librosa's native sample rate is set to 22.05 kHz, and its STFT parameter defaults are set to a `nfft` value of 2048 and hop length of 512. Unless you know about this, you may be performing calculations with incorrect assumptions.

- Audacity's power spectrum calculation limits `nfft` value choices based on signal length; as such, the same `nfft` value can't be chosen for all files in a batch if they are of non-uniform lengths. Also, spectral analysis can only be performed by clicking through a series of sub-menus, and can only be done on one sound clip at a time. The low-level libraries that supposedly allow for batch processing of files to do this task don't actually work as described in the online documentation.

- Audacity's Fourier transform (`pffft`) relies on a translation of Fortran 77 code from FFTPACK that was written in 1985 [1]. These algorithms are very powerful, but may be difficult to integrate with modern software, and may not behave as expected, since they were designed to operate on hardware that had different limitations.

- The number of different FFT algorithms that have been written and re-written for specific uses is at this point an unofficial meme in signal processing. This is evident across many different packages with amusing names such as "Pretty Fast Fast Fourier Transform" (`pffft`), "Keep It Simple Stupid Fast Fourier Transform" (`kissfft`), "Fastest Fourier Transform in the West" (`fftw`), and others. This can be overwhelming to keep up with when choosing algorithms.

In short, when it comes numerical computation, there will always be hidden assumptions that form a collection of presets, whether for parameter values, expected modes of user interaction, or conceptual approaches to sound. Tool choice is often made based on the baked-in assumptions that align most closely with a task at hand. This is neither inherently good nor bad, but a phenomenon of engaging in real-world problem-solving.

#### IV. Proposed Technical Solutions – Conceptual Reimaginings

In the previous section, we outlined a technical wish-list based upon issues we have encountered in our use of audio analysis software. Informed in tandem with historical perspectives and conceptual extensions, we present a variety of solutions to the problem of Schrodinger's Audio Data Visualization Conundrum that go beyond the aforementioned specific issues into an evaluation of the landscape of contemporary cognition. <We propose that giving users access to independence and agency facilitates an increased ability to form complex cognitive associations.> (In a sense, this concept goes slightly beyond software into the domain of pedagogy, however, we strive to refine our focus toward the field of audio information visualization.) In the argument for this shift, we identify three fundamental principles/elements that are necessary to enable this:

**Transparency** – a clear-box approach, rather than a black-box approach, can empower the user to make their own appropriate choices for their intended use. This can involve presenting available options as visual cues at the point of interaction, rather than making decisions for the user or simply leaving all instructions in the documentation. It could also involve informing the user as to why certain design choices were made, and provide options for real-time reconfiguration.

**Flexibility** – the ability to configure an environment that best aligns with an individual's task requirements or work style can give a sense of agency over workflows. Sometimes, it is especially useful to have multiple perspectives when trying to understand a complex situation. The difficulty of working with time-series data is no

exception; the ability to switch seamlessly between analogous options, and even to compare them side-by-side, can be very informative. Adaptable design principles make tools easier to use across a wide variety of scenarios, and may encourage users to stick with one familiar tool, rather than switching frequently between diverse workflows.

Robustness – tool-based environments should be able to handle a wide variety of contexts, and should be as agnostic as possible to the types of data that are input. Like a hammer, which has a variety of uses, or a multi-tool, which has even more, these persistent objects are two divergent examples design principles. One is strikingly simple, and as such, is almost neutral to its application. The other is quite complex – a fusion of many different small devices that are convenient to have in the same place, ingeniously attached through a shared ‘handle’, and made compact and portable, ready for action. While clearly intended for specific tasks, each of these tools enjoy a wide variety of uses and applications in everyday human life.

## 1. Cognition and Visual Design:

The theoretical benefits of incorporating an updated set of modern design principles into audio visualization workflows have far-reaching implications outside of simply being less annoyed while performing daily tasks. Studies across cognitive psychology and design theory show that increased perceptual connections can enhance pattern recognition []. The following examples demonstrate how spatial and temporal representations of information impact mental processes such as comprehension, memory, and learning.

### Cognitive Load Theory:

Split-attention effects show that having to combine information from multiple, individual, spatially-separated sources inhibits learning [sweller]. These effects are also found in scenarios where information is presented simultaneously, but in different formats [sw]. This implies, conversely, that if information is visibly close together, and/or presented simultaneously but in the same format, learning will be easier. In audio software, we can draw an analogy to split-screen views that show waveforms, spectrograms, and power spectral density on separate screens. Users are required to constantly switch back and forth between views, trying to remember what they previously saw on the last screen as they translate information from one format to another. (This is an actual, real problem in Audacity; see section IV-b.) Such display issues limit a user’s mental availability to make intuitive inferences, since one must search for and map visual elements back to each other while holding prior information in working memory. The demands on cognitive load also increase when information is presented sequentially [], rather than in staggered or simultaneous formats. Furthermore, information complexity is modulated not just by the total number of elements, but also by their interactions []. Simultaneous information streams require greater load on working memory []; therefore, the more interconnected a group of elements is, the more complex the information they represent. From this, we can conclude that sequential formats are not ideal for processing complex interconnected information. Some studies have even shown that the maximum amount of new or interacting information that can be processed at the same time ranges from two to four chunks []. Outside of cognitive psychology, inefficiency in linear and sequential information processing is also well-known in the communications, network theory, energy, and computing industries [qed?]. Since humans are the architects of these systems, the phenomenon that preferences a non-simultaneity of information processing could even be a function of human cognition, but that is outside of the scope of this paper to explore.

2. The effects of visual elements on perception have been explored systematically through a variety of principles that govern design theory. The visual variables framework describe position and size as the principal factors that express quantitative differences []. Color, as a variable, is broken into the values of hue, which describes the qualitative difference of category, and value, which describes the quantitative difference of order []. Together with shape, orientation, and texture, these visual variables describe a hierarchy of information with levels that are either associative or dissociative []. This means that visual characteristics can be used to deconstruct the emergent patterns that inform meaningful group characteristics. That is to say, when objects are perceived as being part of a group, visual variables provide a basis for distinction. To extend these thoughts to audio software and visualization, we can thereby conclude that the ability to identify patterns in abstract representations, such as

those used for audio visualization, can be facilitated by making visual design choices that correctly map visual elements to meaningful features. This is consistent with the approaches for dimensionality reduction often seen in modern data visualization strategies.

### 3. Practical implementation:

Through the lenses of cognitive and visual design theory, we show that associations between visual elements and the human psyche are intrinsically linked through the perceptual continuum that is bodied sensory experience. The inner workings of human cognition and psychology fundamentally demand an interactive format to give context to complex information. We can therefore project that for audio information visualization design, users may benefit from access to tools and workflows that allow for a perceptually diverse engagement with sound. This could include nonlinear workflows, reorienting information along different axes, using new metrics to scale information, or interchanging relationships between variables. The incorporation of diverse design principles into audio analysis tools and workflows can expand the boundaries of both technical analysis and creative sound exploration. Practically, it takes time to implement new tools. To be adopted, novel visualizations may require a shift in representational paradigms. When presented with something new, not everyone is ready to accept the change. Users must first overcome cognitive dissonance and resistance to change, followed by the learning curve associated with performing any new task []. As familiarity and then mastery is attained, these tools can become completely streamlined into existing workflows. We may even struggle to remember what life was like before we had access to them; such is the curse of convenience. However, increased technical literacy begets the benefits of speed, efficiency, and creative flexibility.

### V. Impact, intended audience, who this benefits

There are endless ways to explore the theoretical effects of applied design philosophy, but what about their impact? When a new tool or technique is deployed, who will actually use it? Who will it benefit? Where and how will it be used? Especially now, in the age of Big Data, there is an accelerated need to include non-domain experts and citizen science participants in the validation and annotation of data. Tools designed specifically with interaction and visualization in mind can make it more accessible for people to interact with data in ways that are relatable, intuitive, and familiar. The tactile experiences of everyday digital tools, such as apps and games, can be modeled and expanded upon to create user experiences that feel familiar while not being too distracting. Such tools can also give people a sense of agency over what they're doing – they may reveal the 'secret elements' that are often reserved for specialists, increasing transparency, building institutional trust, and generating a sense of community investment. Furthermore, tools that are fun and interesting to use generate conversations outside of their initial use/community. When everyday people get excited enough about wild bird audio annotation apps to discuss them at coffee shops or networking events, for example, this can be viewed as a sign of success that such a tool is connected with social values. Thus, there are diverse practical reasons in favor of increasing the accessibility of audio analysis and exploration to both technical and non-technical audiences. The following are some examples of benefits to specific groups:

- People who already use data viz tools regularly for their jobs, such as scientists, data scientists and analysts will certainly benefit from increased efficiency and intuition, allowing them to see audio information in new ways.
- Citizen scientists who participate in valuable tasks such as data annotation and validation, species identification, symptom reporting, noise pollution assessment, can have a way to easily annotate in real-time that may allow them to feel included as an essential part of a team, gives them more knowledge about the science and behind the scenes, which could encourage them to become more excited and involved from a scientific standpoint. This is triply beneficial because science education is essential as people need to work together to address many urgent problems in fields such as conservation, medicine, and society.
- Accessibility by including things that are interesting or fun to look at, listen to, and interact with, especially for non-experts, can provide entertainment as well as social values. The possibility of gamification can also increase



audience reach, and can be used to collect feedback about what does and doesn't work, as well as who tends to use the tools and how, which are valuable insights for any tool designer.

- We can imagine a use where, for a large dataset that needs annotation, the dataset can be broken up into smaller pieces and distributed among a group of people to lessen the workload. Then, it is essential for all users to be sure they are referring to the same phenomena, and the same features, across the same interface.

- AI users in particular, who may not be used to working with real data, or who may work with many different types of data, need assistance in understanding the nuances of datasets when they are not familiar with the subject matter. In the rising proliferation of AI outside of experimental and research domains, the number of people working with audio data will increase dramatically, as will the use of AI as an everyday tool in its own right. Such human individuals (and, more dangerously, their AI counterparts) can make incorrect assumptions about properties or characteristics of sound if they are not informed in a way that is fast, efficient, and intuitive. This also factors into the field of ethics, since the dangers of making assumptions can proliferate quickly in cases where a small effect may spiral out of control over a massive dataset like those seen in Big Data, or may propagate into models through training, or affect other datasets through extracted metadata.

- Audio visualization tools can act as intermediary steps between the many people involved along the way in the process of scientific and artistic inquiry. It places control in the hands of the user, and reconfigures the hierarchy that limits niche knowledge to be held solely by domain experts. Increased agency can build a sense of community, and strengthens the ties that people feel to their work or special interest.

Many people will continue to be affected by today's rapid advancements in audio data visualization as the Age of Information spirals outwards. We hope that with this expanded consideration of the implications and impacts of new tools on their audiences, the case for incorporating a broader set of user-centric design principles may be compelling.

## VI. Conclusions.

**Problem:** Audio information visualization software carries inherited assumptions that interfere with users' ability to engage in efficient and creative workflows.

**Solution:** Update software using design principles that are aligned with modern needs and cognition.

### Perception, Information, and Visualization

Sound perception is multidimensional and its study requires abstraction. In the transformation of audio information from perceived experiences to visual media, care must be taken to ensure that features are mapped meaningfully to corresponding visual elements. From cognitive psychology, we know that information accessibility affects learning and understanding. Separated and sequential information is more difficult to process than combined and multithreaded information, and adjusting their presentation can increase cognitive efficiency. Visual variables further describe patterns and groups that seem to form in abstract visual representations, associations which are an intrinsic part of human cognition. Aligning design principles with human experience can make it easier to see patterns in abstract information.

### The History and Impact of Software Design

Software inherits design choices from the past. DAWs and other software tools were originally physical hardware that transitioned with the rise of digital technology. The influence of past physical constraints persists in modern digital software as hidden assumptions. This black-box model of software design can inhibit efficiency and creativity in modern contexts due to its use transference. Various specific hidden assumptions made in audio info viz (AIV?) software have perks and drawbacks. In modern software design, transparency facilitates comprehension, communication, and connectedness. Design transparency allows for technical flexibility. Preset selection is inevitable and useful. By re-evaluating the historical limitations of software design, we can create tools better suited for modern users.

### Theoretical Outcomes

Transparency, Flexibility, and robustness are three core design principles that should be followed to maximize agency and independence, and thus efficiency and creativity. Giving users control of their tools facilitates agency and cognitive capacity, and promotes positive social values. Practical implementations take time, and cognitive dissonance may slow the adoption of new tools. Successful implementation of such tools can increase speed, efficiency, and creative flexibility. Updated designs will make it easier to meet the practical demands of real-world applied data visualization because tools will be easier and more enjoyable for people to use. Freedom begets cognitive benefits. Cross-domain extensions such as experimental media expand creative and technical limits.