

Using Sound to Enhance Data Interpretation



Dan Damelin
(ddamelin@concord.org)
is a senior scientist.

By Dan Damelin

The use of sound to communicate nonverbally predates written records and is deeply embedded in human history. The simple rhythmic sounds from early drums, bells, whistles, and horns evolved to become complex communication data. Today, graphs are the most common way information from data is conveyed. The COVID-Inspired Data Science Education through Epidemiology (CIDSEE) project, a collaboration between Tumblehome, Inc. and the Concord Consortium, is developing and researching new tools that add sound to graphs to help students explore and make sense of data.

Sonification is the process of turning information into sound, for example, to mark time with bells or relay battlefield information with drums or trumpets. It also enhances scientific discovery by providing a multimodal approach to data analysis, communicates meaning through artistic interpretations of data, and increases the accessibility of data visualizations.

A short history of sonification

One of the first modern examples of sound being used to represent numerical data occurred in the 1920s when geoscientists turned seismic data into sound, allowing researchers to “hear” earthquake patterns. (Seismographs themselves resemble sound waves [Figure 1].) In the 1930s and 1940s, auditory signals provided situational awareness for pilots and navigators, including information about the condition of their planes and ships, as well as proximity warnings.

From the 1950s to the 1970s, even as computers were just being developed, sound synthesis and a rudimentary exploration of representing data using sound laid the foundation of modern sonification. Bell Labs developed one of the first programs for digital sound synthesis. While primarily focused on creating music, this work demonstrated how numerical data could be translated into sound, inspiring collaborations between artists and scientists who explored data-driven composition and influencing later scientific sonification. About the same time, NASA researchers also began experimenting with mapping data to sound to interpret large datasets collected during space missions. Such sonification allowed scientists to explore patterns in data that were difficult to discern visually.

In 1992, Gregory Kramer organized the first International Conference on Auditory Display, marking a major milestone in the history of sonification. The event brought together researchers from diverse disciplines to discuss sonification and formalized sonification as an academic field. Kramer and others developed frameworks for how sonification could be systematically applied to scientific and practical contexts. The International Community for Auditory Display, which began as a result of that inaugural conference, is still active today.

In recent years, as computers have become more powerful, sonification has become more ubiquitous. It has been used across multiple disciplines and applications from EEG monitoring to sifting through the enormous datasets generated by the CERN particle collider to hearing trends in climate data.

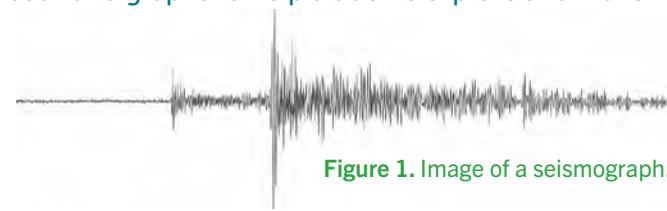


Figure 1. Image of a seismograph.

CODAP and the Sonify plugin

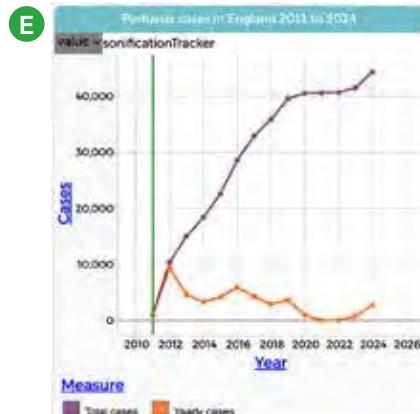
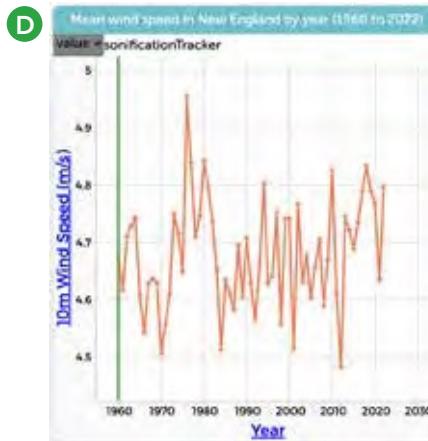
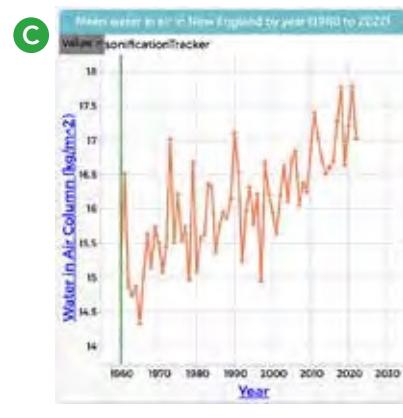
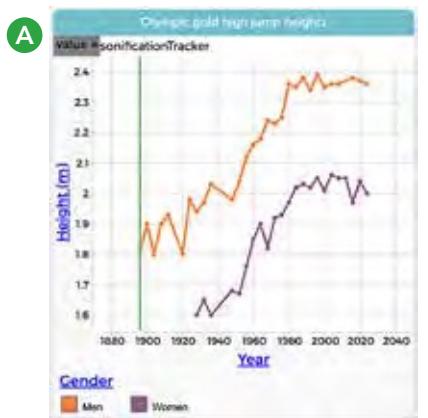
The Common Online Data Analysis Platform (CODAP) is a free, web-based tool for supporting students in sensemaking with data and learning data science practices. CODAP’s functionality can be extended through plugins. In 2018, Takahiko Tsuchiya, then a graduate student at the Georgia Institute of Technology, began building a plugin with the help of one of the CODAP developers, Jonathan Sandoe, to facilitate sonification of data in CODAP. The National Science Foundation-funded CIDSEE project has extended and improved the Sonify plugin, which is available from the plugin menu on the CODAP toolbar.

Figure 2 shows an example of the Sonify plugin, configured to work with a dataset of CO₂ measurements over time from NOAA’s Mauna Loa observatory. The graph of that data shows a classic upward trend (Figure 3). To sonify this data, the user selects what attribute to map to pitch and what attribute to map to date/time. After the user clicks “create graph,” the plugin instructs CODAP to generate the graph of CO₂ vs. date. When the user clicks “Play,” the green line on the graph sweeps from left to right and the plugin produces a sonic representation of the data with higher values of CO₂ generating higher-pitched tones.



Figure 2. The Sonify plugin in CODAP.

Figure 3. Graph of CO₂ vs. date at the Mauna Loa observatory in Hawaii.



CIDSEE and sound graphs

Since 2021, the CIDSEE project has empowered over 1,000 underserved middle school students in afterschool and summer camp “Data Detectives Clubs” to explore the field of epidemiology through the use of scientific data and innovative data tools. By engaging with real-world epidemiological challenges, students cultivate critical skills and an appreciation for the importance of data in understanding public health issues. The program is designed to enhance students’ confidence, interest, knowledge, and abilities in working with datasets and data tools, including CODAP and its Sonify plugin, to investigate pandemics. The goal is to inspire student interest in epidemiology and related STEM fields, including research, modeling, data analysis, and science communication, by providing hands-on experiences with the data-driven work of epidemiologists.

Through a supplemental award from NSF, the CIDSEE project is extending its work to study how and when sonification might aid student understanding of time-series graphs, such as the spread of COVID over time, and improve data interpretation with other time-series phenomena. To explore the value of sonification, we asked students to interpret a variety of graphs with and without sound (Figure 4, A-E):

- A** a comparison between two groups (e.g., men’s vs. women’s winning Olympic high jump heights over the years)
- B** a simple phenomenon with an easy description (e.g., it goes up then levels off)
- C** a clear trend with some significant variability (e.g., water content in the air over time)
- D** no clear trend (e.g., change in wind speed over time)
- E** a cumulative vs. marginal comparison (e.g., yearly vs. cumulative pertussis cases)

Figure 4. Graphs were presented with and without sonification.

The CIDSEE work described in this article is based on the combined efforts of many individuals, including Jan Mokros, Penny Noyce, Jacob Sagrans, Barnas Monteith, and Bill Finzer.

While the analysis of the study is ongoing, some preliminary observations have emerged:

- Most students easily grasp the basic concept of sonification, understanding that higher pitches correspond to higher values and lower pitches to lower values.
- Students appreciate having the ability to adjust the playback speed, allowing them to listen to the graph at a faster or slower pace.
- Overall, students find sonification appealing—it’s a novel and engaging way to interact with data.
- Students find it useful to have the visual line scan across the graph during sonification, helping them track their position on the graph while listening.
- Most students find it challenging to interpret the sonification through sound alone. They need to see the graph alongside the audio to make meaningful sense of the data.

Accessibility and sonification

A new NSF-funded project called Data By Voice is building upon lessons learned by the CIDSEE project to explore ways to make data sonification more understandable, especially for blind and low-vision (BLV) users. The project is using artificial intelligence to facilitate sensemaking with data in CODAP by creating a voice interface for interacting with CODAP and a large language model to describe data representations created by BLV students. In addition to verbal descriptions generated by AI, students can request that a graph they made be sonified. With AI, the future of sonification is just beginning.

LINKS

CODAP—codap.concord.org

CIDSEE—tumblehome.org/research/

Data By Voice—concord.org/data-by-voice