Newcastle University

Open Research Case Study

An Open Approach to Engaging with Data through Sound

Chris Harrison

Reader in Astrophysics, School of Mathematics, Statistics and Physics.

Introduction and Research context

Sonification is the process of turning data into sound, as an alternative, or complement, to visualisation. We are a group of astrophysicists who launched a new sonification research project in December 2021. Our goals are to use sound for: enhancing researchers' exploration of complex datasets; creating immersive and accessible educational resources; and helping the public be more engaged with data. Our ethos since launching Audio Universe (www.audiouniverse.org), is to be open and inclusive in everything we do.

Open practices used

We employ multiple open practises:

We use **open access publications and a pre-print repository** (arXiv.org) so all research results are accessible, cost free. All papers are locatable through our website.

Our sonification code, STRAUSS, is **open source, through github,** and is extensively documented. This Python-based code underpins all our research and engagement. We encourage participation in development, primarily through github issues, email suggestions and in-person discussions.

We have **a dedicated public data repository** for data associated with publications, **released with CC-BY licences**. We also deposit engagement resources (video, audio etc.) under CC-BY, except where external artists require CC-BY-ND.

We practise **open and participatory methods** through tutorial notebooks and example codes, available via github and/or our website. Many can be accessed via Google Colab, meaning users can run them directly in the browser. We've also used notebooks during in-person training, focussed on: inter-disciplinary sonification design (Audible Universe,2022), solar physics (WISA, 2023) and communication of big data (dotAstronomy, 2024).

For these, we created bespoke notebooks using datasets most relevant to the participants.

We repurpose public datasets by presenting them through sound. This includes several examples using public astrophysics data. For example, taking the publicly available data on gravitational wave events and turning it into an immersive, sonification-based virtual reality experience. In other domains, we have taken public temperature anomaly data from the UK Met Office to make an emotive sonification addition to the famous "warming stripes" visualisation. A showreel of examples of some public data sonifications is available on our YouTube channel.

Benefits

Our sonification approaches provide a novel method to enable people feel more engaged with public datasets, often being less intimidating than traditional visual methods, such as charts, graphs or tables. With our open access code we've had unexpected applications. For example, STRAUSS was used by Dr Perez Montero for sonifications to represent themes around inclusion and accessibility.

Our code was used by a group to create a Citizen Science project to search for exoplanets. We've also been invited to explore sonification for research in other domains with complex data, including ecological networks (Prof. D. Evans, Newcastle University) and Earth observation data (Mike Smith, ASPIA space).

The code has seen improvements due to participatory approaches. Examples include more integrated flexibility to different data types as new users explore their data. Another example comes from a blind American student who wished to use the code for their lab work; their queries/requests led to new code features.

Based on the transparency and reproducibility of our publications, they are being referenced by other groups as good practice for demonstrating sonification efficacy (e.g., Casado et al. 2024).

Challenges

One challenge has been ensuring our code is accessible for a wide user-base with different platforms, levels of coding experience, etc. A number of technical hurdles have been overcome. One solution we have found has been using browser-based Google Colab notebooks. This has enabled us to reach users who otherwise would not have been confident to use the code. We need to continue to develop these example tutorials to demonstrate the full potential of the code; we will need to devote considerable time to this. This includes supporting initiatives to make more widely accessible notebooks in general terms ("Notebooks for All"). With our participatory research we've had to understand GDPR and ethics processes, which were new to us. We previously only worked on astrophysics research, where these issues were not relevant. Thanks to excellent support from university colleagues, we've learnt about these processes and broadened our own skills/ expertise.

Lessons learned

We have seen substantial benefits from embedding open practices into a project right from the outset. Not only has this made data more accessible, engaging, and understandable for everyone, it has clear benefits for our research, resulting in a better code, exposed us to new applications and broader impact. We advocate for anyone embarking on a new project to integrate open research practises from its inception!