Bioacoustic and Ecoacoustic Data in Audiovisual Core

Ed Baker ‡

‡ Natural History Museum, London, United Kingdom

Corresponding author: Ed Baker (edwbaker@gmail.com)

Abstract

Audiovisual Core (Audiovisual Core Maintenance Group 2023), is the TDWG standard for metadata related to biodiversity multimedia. The Audiovisual Core Maintenance Group has been working to expand the standard to provide the terms necessary for handling sound recordings. Audiovisual Core can now handle acoustic metadata related to biodiversity from single species (bioacoustics) to the ecosystem scale (ecoacoustics).

Bioacoustics

The Natural History Museum, London has a significant collection of recorded insect sounds (Ragge and Reynolds 1998) that are often directly linked to museum specimens (Fig. 1). The sound collection has previously been digitised and made available electronically through the BioAcoustica project (Baker et al. 2015). The BioAcoustica platform allows for annotation of audio files with tags including "Call" for deliberate sound made by an organism, "Voice Introduction" for metadata, and "Extraneous Noise." These boundaries are defined by their start and end times (in seconds) relative to the start of the file (Fig. 2).

Ecoacoustics

Ecoacoustics deals with the sounds present within an entire soundscape or ecosystem. The calls of individual species form the biological part of the soundscape (biophony) alongside sounds produced by non-living natural sources (geophony) and humans (anthropophony). Individual components are often defined by date and time boundaries, and sometimes by upper and lower frequency limits (Fig. 3).

Regions of Interest

The recently added concept of a "Region of Interest" (ROI) allows for the annotation of sound files, identifying multiple regions within a single recording with time and/or frequency bounds. However, the vocabulary of ROI*[1] is not just intended for sounds. Equivalent terms also allow for regions to be specified with images and videos.

The use of well-defined annotations has the potential to generate large amounts of training data for machine learning models and provide a standard for generating observation records from these models (e.g., BirdNet, see Kahl et al. 2021), which can be verified by linking them to audio segments within a much larger recording.

The development of a metadata standard for regions of interest has several interesting possibilities, including linking multiple observation records to a single soundscape recording (the recording acts similarly to a voucher specimen) and aggregating regions across multiple datasets to create larger corpora for training machine learning models.

Keywords

biodiversity information standards, biodiversity media, soundscapes

Presenting author

Ed Baker

Presented at

SPNHC-TDWG 2024

Acknowledgements

Conflicts of interest

The authors have declared that no competing interests exist.

References

- Audiovisual Core Maintenance Group (2023) Audiovisual Core Term List. https://ac.tdwg.org/termlist/2023-02-24. Accessed on: 2024-9-09.
- Baker E, Price B, Rycroft SD, Hill J, Smith V (2015) BioAcoustica: a free and open repository and analysis platform for bioacoustics. Database 2015 https://doi.org/10.1093/database/bav054
- Kahl S, Wood C, Eibl M, Klinck H (2021) BirdNET: A deep learning solution for avian diversity monitoring. Ecological Informatics 61 https://doi.org/10.1016/j.ecoinf.2021.101236

- Ragge D, Reynolds WJ (1998) The Songs of the Grasshoppers and Crickets of Western Europe. Harley Books, Colchester. https://doi.org/10.1163/9789004632189

Endnotes

*1   https://ac.tdwg.org/termlist/#711-region-of-interest-vocabulary

Figure 1.

Left: label data from the Natural History Museum, London collection, including an orange label referencing the recorded sound collection. Right: Specimens in the collection showing one with an orange label indicating a recording of its song is in the recorded sound collection. Images by author, © Trustees of the Natural History Museum, London.

Figure 2.

Annotated regions of an audio file are colour-coded by the type of annotation (blue for voice introductions, red for extraneous noise and green for calls). Regions may overlap. From: http://bio.acousti.ca/node/11778, Baker et al. (2015).
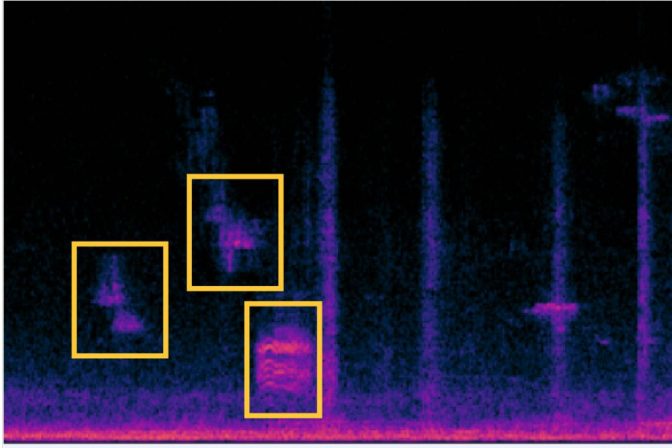
Figure 3.

Spectrogram (frequency vs. time) plot, with three regions of interest highlighted. Recording by author, visualised in Audacity, © Trustees of the Natural History Museum, London.