The Journal of Open Source Software

Crowsetta: A Python tool to work with any format for annotating animal vocalizations and bioacoustics data.

David Nicholson [1¶]

1 Independent Research, USA ¶ Corresponding author

DOI: 10.21105/joss.05338

Software

- Review
- Repository
- Archive

Editor: Olivia Guest

Reviewers:

- @oliviaguest

Submitted: 29 March 2023
Published: 12 April 2023

Summary

Studying how animals communicate with sound allows researchers to answer a wide range of questions, from "What species of birds live in this area?", to "How do mice mothers protect their young?". Some animals learn their vocalizations, and by studying the vocal behavior of these animals we can investigate questions like "How did speech evolve?". Answering these questions require computational methods and big team science across many disciplines, such as ecology, ethology, bioacoustics, psychology, neuroscience, linguistics, and genomics.

Analyses of animal acoustic communication often require annotating the sounds animals make. To annotate animal sounds, researchers typically use graphical user interfaces (GUIs), that enable them to annotate audio and/or spectrograms. Such annotations usually include the times when sound events start and stop, and labels that assign each sound to some set of classes chosen by the annotator. GUI applications save the annotations in many different file formats. This Python package, crowsetta, can parse the most widely used formats, and it provides software abstractions that make it easy to extend the library to parse new formats. In this way, crowsetta allows researchers to work with data annotated in a wider variety of formats. Additionally, crowsetta helps users convert annotations to simple file formats, such as csv files, that do not require detailed knowledge of the annotation format itself. This facilitates loading the annotations with widely used libraries for data analysis (e.g., Pandas in Python), and also promotes sharing data. Overall, crowsetta supports the interdisciplinary collaboration required for the study of animal acoustic communication.

Statement of need

Studying how animals communicate with sound allows researchers to answer a wide range of questions. For example: "What is the language faculty, and how did it evolve?" (Hauser et al., 2002), and "What is the basis of learned vocalizations in animals?" (Wirthlin et al., 2019). Answering these questions will require big team science and true interdisciplinary collaborations (Hauser et al., 2002; Wirthlin et al., 2019). The methods used to answer these questions are becoming ever more computational and data driven. As one example of such methods, deep learning models have recently proliferated in the field of bioacoustics (Stowell, 2022) and adjacent fields like animal behavior (Berman, 2018; Pereira et al., 2020), specifically studies of vocal behavior (Sainburg & Gentner, 2021). These models have also become common in the field of neuroscience (Cohen, Engel, et al., 2022), and more specifically in studies of the neural bases of vocal communication (Coffey et al., 2019; Cohen, Nicholson, et al., 2022; Goffinet et al., 2021; Sainburg et al., 2020; Steinfath et al., 2021). The performance of these data-driven deep learning models depends crucially on annotated data. High-quality annotations from humans are needed to train supervised learning models to predict annotations (Coffey et al., 2019; Cohen, Nicholson, et al., 2022; Steinfath et al., 2021), or to provide inputs to unsupervised models that perform dimensionality reduction and compute measures

of similarity (Goffinet et al., 2021; Sainburg et al., 2020). In spite of this clear-cut need for high quality, easily accessible human annotations of animal acoustic communication data, it is surprisingly difficult to work with these annotations in code. Ironically, researchers studying animal acoustic communication face challenges when communicating with each other.

This barrier arises in part because of the lack of a generalized schema for annotating audio datasets, which in turn results in a proliferation of annotation formats. A standardized format for annotations would greatly facilitate interoperability, one of the guiding FAIR principles (Wilkinson et al., 2016) There are some existing schema for bioacoustics datasets that include annotations (Baskauf et al., 2022; Fukuzawa, 2022; Recalde, n.d., 2023; Roch et al., n.d.) but there has been no broad effort at standardization. Likewise, in animal behavior and neuroscience, specifically in areas that study acoustic communication, there have been proposals for datasets structures, file formats, and databases schema, e.g. Dragly et al. (2018), that would by necessity include annotations. But again there has been little formalization across research groups and again, no wider effort to make these interoperable with audio annotations more generally. The most direct work on standardization that I am aware of is the Json Annotated Music Specification (JAMS) (Humphrey et al., 2014; McFee et al., n.d.), a standard for annotation in music information retrieval research. One implementation of this standard has been provided in a Python library of the same name (Humphrey et al., 2014). These issues were discussed at the first Audio Across Domains (AudioXd) conference (https://kitzeslab.github.io/audioxd/), and a generalizable schema for audio annotation is being proposed (in preparation). In the meantime, the lack of standardization creates a clear need for software tools that provide a sort of interoperability layer.

Currently, there is little tooling available that makes it easy to convert between annotation formats, and to share annotations in widely used simple formats such as a csv file. Many tools have been developed to work with specific formats, e.g., Praat Textgrid (Buschmeier & Włodarczak, n.d.; Jadoul et al., 2018) or Raven selection tables saved as text files (in Python (HAUPERT et al., 2022), and in R (Araya-Salas, 2020)), but to the best of my knowledge there are no tools that focus specifically on interoperability of formats. This lack of tooling stands in contrast to the clear-cut need for researchers to be able to collaborate across disciplines when working with annotated audio datasets.

Crowsetta addresses the clear need for a tool that allows for interoperability between the many existing annotation formats, and that makes it possible to flexibly access annotations within Python for development of downstream applications. Crowsetta also meets the needs of researchers to easily share annotations and to use them within Python for imperative code, e.g., scientist-coder scripts used to fit statistical models or analyze behavior. To address these needs all these needs, the package has built-in support for many widely used formats such as Audacity (Audacity Team, 2019) label tracks, Praat (Paul Boersma & David Weenink, 2021) TextGrid files, and Raven (Charif et al., 2006; Program, 2016) selection tables exported to text files. The design of crowsetta also focuses on interoperability. It allows researchers to convert annotations loaded into built-in formats to more generic formats: for example, a generic "sequence-like" format that can represent annotated sequences of speech and animal vocalizations. This generic format can be saved as a csv file, making data easier to share and easier to work with through widely-used data analysis libraries like Pandas (McKinney, 2010; team, 2020), In this way crowsetta minimizes the need for specialized knowledge of tool-specific formats. In sum, the package provides a Pythonic way to work with annotation formats for animal vocalizations and bioacoustics data.

Originally, crowsetta was developed for use with vak (Nicholson & Cohen, 2022), a neural network framework for researchers studying animal acoustic communication. Crowsetta made it possible to work with several annotation formats when using vak to benchmark a neural network architecture that automates annotation of birdsong, TweetyNet (Cohen, Nicholson, et al., 2022; Cohen & Nicholson, 2023). Since then, crowsetta has been used in tandem with vak by several research groups in neuroscience (Goffinet et al., 2021; McGregor et al., 2022) and bioacoustics (Provost et al., 2022).

Acknowledgements

I would like to acknowledge support from Yarden Cohen for development of crowsetta as part of the VocalPy ecosystem. I would also like to acknowledge contributions to crowsetta during the pyOpenSci review, made by the two reviewers, Tessa Rhinehart and Sylain Haupert, as well as expert advice from Yannick Jadoul on support for TextGrid, with guidance of Chia Marmo as the editor.

References

Araya-Salas, M. (2020). *Rraven: Connecting R and Raven bioacoustic software. R package version 1.0.9.*

Audacity Team. (2019). *Audacity*. https://www.audacityteam.org/

Baskauf, S., Desmet, P., Klazenga, N., Blum, S., Baker, E., Morris, B., Webbink, K., danstowell, Döring, M., & Junior, M. (2022). *Tdwg/ac: Audubon Core standard 2022-02-23 version*. Zenodo. https://doi.org/10.5281/zenodo.6590205

Berman, G. J. (2018). Measuring behavior across scales. *BMC Biology*, *16*(1), 23. https://doi.org/10.1186/s12915-018-0494-7

Buschmeier, H., & Włodarczak, M. (n.d.). *TEXTGRIDTOOLS: A TEXTGRID PROCESSING AND ANALYSIS TOOLKIT FOR PYTHON*.

Charif, R., Ponirakis, D., & Krein, T. (2006). Raven Lite 1.0 user's guide. *Cornell Laboratory of Ornithology, Ithaca, NY*.

Coffey, K. R., Marx, R. E., & Neumaier, J. F. (2019). DeepSqueak: A deep learning-based system for detection and analysis of ultrasonic vocalizations. *Neuropsychopharmacology*, *44*(5), 859–868. https://doi.org/10.1038/s41386-018-0303-6

Cohen, Y., Engel, T. A., Langdon, C., Lindsay, G. W., Ott, T., Peters, M. A. K., Shine, J. M., Breton-Provencher, V., & Ramaswamy, S. (2022). Recent Advances at the Interface of Neuroscience and Artificial Neural Networks. *Journal of Neuroscience*, *42*(45), 8514–8523. https://doi.org/10.1523/JNEUROSCI.1503-22.2022

Cohen, Y., & Nicholson, D. (2023). *Tweetynet*. Zenodo. https://doi.org/10.5281/zenodo.7627197

Cohen, Y., Nicholson, D. A., Sanchioni, A., Mallaber, E. K., Skidanova, V., & Gardner, T. J. (2022). Automated annotation of birdsong with a neural network that segments spectrograms. *Elife*, *11*, e63853.

Dragly, S.-A., Hobbi Mobarhan, M., Lepperød, M. E., Tennøe, S., Fyhn, M., Hafting, T., & Malthe-Sørenssen, A. (2018). Experimental Directory Structure (Exdir): An Alternative to HDF5 Without Introducing a New File Format. *Frontiers in Neuroinformatics*, *12*. https://doi.org/10.3389/fninf.2018.00016

Fukuzawa, Y. (2022). *Computational methods for a generalised acoustics analysis workflow: A thesis presented in partial fulfilment of the requirements for the degree of Master of Science in Computer Science at Massey University, Auckland, New Zealand* [{PhD} {Thesis}]. Massey University.

Goffinet, J., Brudner, S., Mooney, R., & Pearson, J. (2021). Low-dimensional learned feature spaces quantify individual and group differences in vocal repertoires. *eLife*, *10*, e67855. https://doi.org/10.7554/eLife.67855

HAUPERT, S., Ulloa, J. S., Gil, J. F. L., scikit-maad, & Suarez, G. A. P. (2022). *Scikit-maad/scikit-maad: Stable Release : v1.3.12*. Zenodo. https://doi.org/10.5281/zenodo.

7324324

Hauser, M. D., Chomsky, N., & Fitch, W. T. (2002). The Faculty of Language: What Is It, Who Has It, and How Did It Evolve? *Science*, *298*(5598), 1569–1579. https://doi.org/10.1126/science.298.5598.1569

Humphrey, E. J., Salamon, J., Nieto, O., Forsyth, J., Bittner, R. M., & Bello, J. P. (2014). *JAMS: A JSON ANNOTATED MUSIC SPECIFICATION FOR REPRODUCIBLE MIR RESEARCH*. 6.

Jadoul, Y., Thompson, B., & Boer, B. de. (2018). Introducing Parselmouth: A Python interface to Praat. *Journal of Phonetics*, *71*, 1–15. https://doi.org/10.1016/j.wocn.2018.07.001

McFee, B., Humphrey, E. J., Nieto, O., Salamon, J., Bittner, R., Forsyth, J., & Bello, J. P. (n.d.). *PUMP UP THE JAMS: V0.2 AND BEYOND*. 8.

McGregor, J. N., Grassler, A. L., Jaffe, P. I., Jacob, A. L., Brainard, M. S., & Sober, S. J. (2022). Shared mechanisms of auditory and non-auditory vocal learning in the songbird brain. *eLife*, *11*, e75691. https://doi.org/10.7554/eLife.75691

McKinney, Wes. (2010). Data Structures for Statistical Computing in Python. In Stéfan van der Walt & Jarrod Millman (Eds.), *Proceedings of the 9th Python in Science Conference* (pp. 56–61). https://doi.org/10.25080/Majora-92bf1922-00a

Nicholson, D., & Cohen, Y. (2022). *Vak*. Zenodo. https://doi.org/10.5281/zenodo.6808839

Paul Boersma, & David Weenink. (2021). *Praat: Doing phonetics by computer*. https://doi.org/10.1097/aud.0b013e31821473f7

Pereira, T. D., Shaevitz, J. W., & Murthy, M. (2020). Quantifying behavior to understand the brain. *Nature Neuroscience*, *23*(12), 1537–1549. https://doi.org/10.1038/s41593-020-00734-z

Program, B. R. (2016). *Raven Lite: Interactive Sound Analysis Software (Version 2.0)*. The Cornell Lab of Ornithology Ithaca, NY.

Provost, K. L., Yang, J., & Carstens, B. C. (2022). The impacts of fine-tuning, phylogenetic distance, and sample size on big-data bioacoustics. *PLOS ONE*, *17*(12), e0278522. https://doi.org/10.1371/journal.pone.0278522

Recalde, N. M. (n.d.). *Pykanto: A python library to accelerate research on wild bird song*.

Recalde, N. M. (2023). *Pykanto: A python library to accelerate research on wild bird song*. arXiv. https://doi.org/10.48550/arXiv.2302.10340

Roch, M. A., Baumann-Pickering, S., Batchelor, H., Širovi, A., Berchok, C. L., Cholewiak, D., Oleson, E. M., & Soldevilla, M. S. (n.d.). *Tethys: A workbench and database for passive acoustic metadata*. 5.

Sainburg, T., & Gentner, T. Q. (2021). Toward a Computational Neuroethology of Vocal Communication: From Bioacoustics to Neurophysiology, Emerging Tools and Future Directions. *Frontiers in Behavioral Neuroscience*, *15*, 811737. https://doi.org/10.3389/fnbeh.2021.811737

Sainburg, T., Thielk, M., & Gentner, T. Q. (2020). Finding, visualizing, and quantifying latent structure across diverse animal vocal repertoires. *PLOS Computational Biology*, *16*(10), e1008228. https://doi.org/10.1371/journal.pcbi.1008228

Steinfath, E., Palacios-Muñoz, A., Rottschäfer, J. R., Yuezak, D., & Clemens, J. (2021). Fast and accurate annotation of acoustic signals with deep neural networks. *eLife*, *10*, e68837. https://doi.org/10.7554/eLife.68837

Stowell, D. (2022). *Computational bioacoustics with deep learning: A review and roadmap*. 46.

team, T. pandas development. (2020). *Pandas-dev/pandas: pandas* (latest). Zenodo. https://doi.org/10.5281/zenodo.3509134

Wilkinson, M. D., Dumontier, M., Aalbersberg, Ij. J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., Silva Santos, L. B. da, Bourne, P. E., Bouwman, J., Brookes, A. J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C. T., Finkers, R., … Mons, B. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data*, *3*(1), 160018. https://doi.org/10.1038/sdata.2016.18

Wirthlin, M., Chang, E. F., Knörnschild, M., Krubitzer, L. A., Mello, C. V., Miller, C. T., Pfenning, A. R., Vernes, S. C., Tchernichovski, O., & Yartsev, M. M. (2019). A Modular Approach to Vocal Learning: Disentangling the Diversity of a Complex Behavioral Trait. *Neuron*, *104*(1), 87–99. https://doi.org/10.1016/j.neuron.2019.09.036