Crowdsourcing Multi-label Audio Annotation Tasks with Citizen Scientists

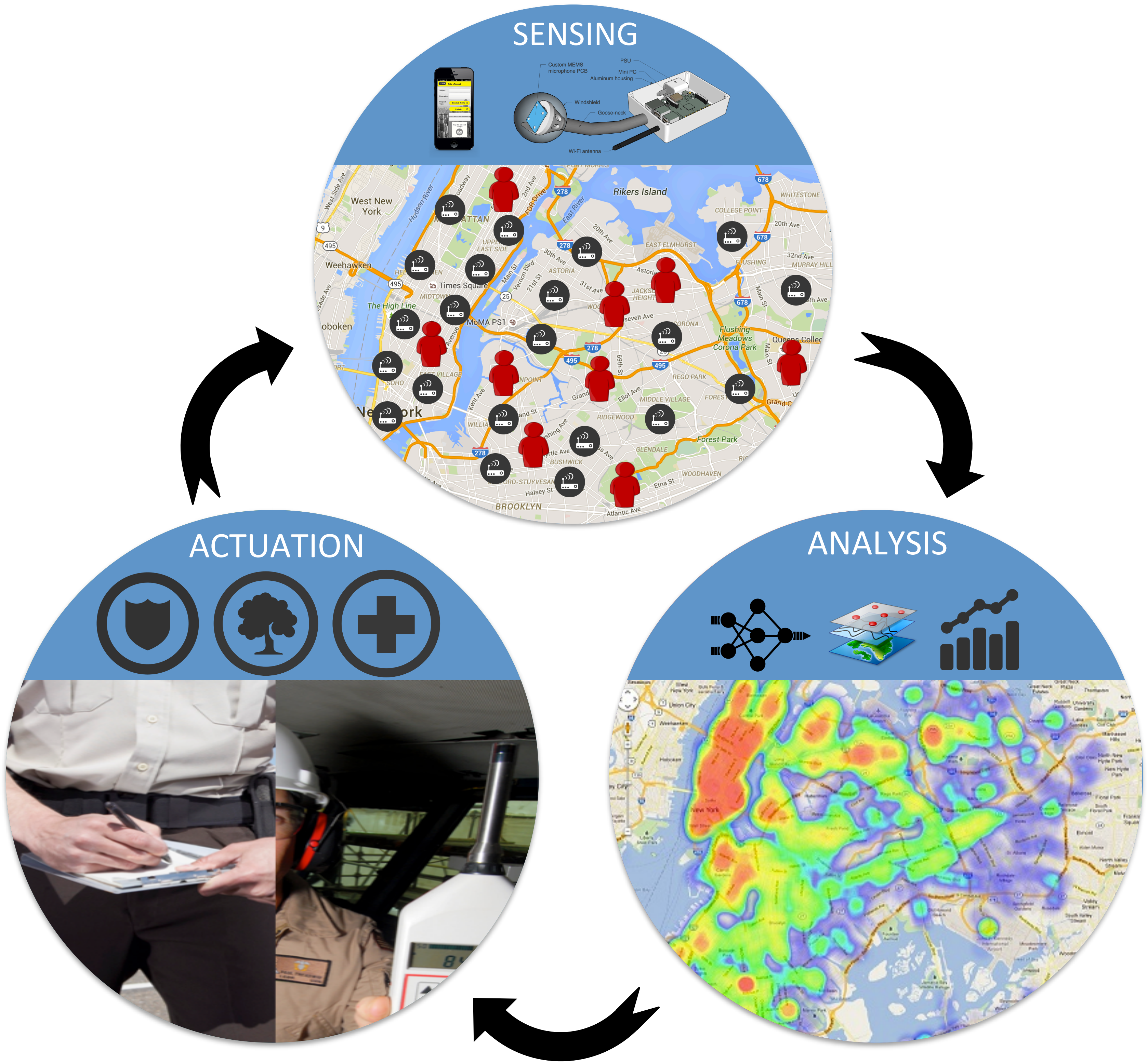
Mark Cartwright, Graham Dove, Ana Elisa Méndez Méndez,
Juan P. Bello, Oded Nov

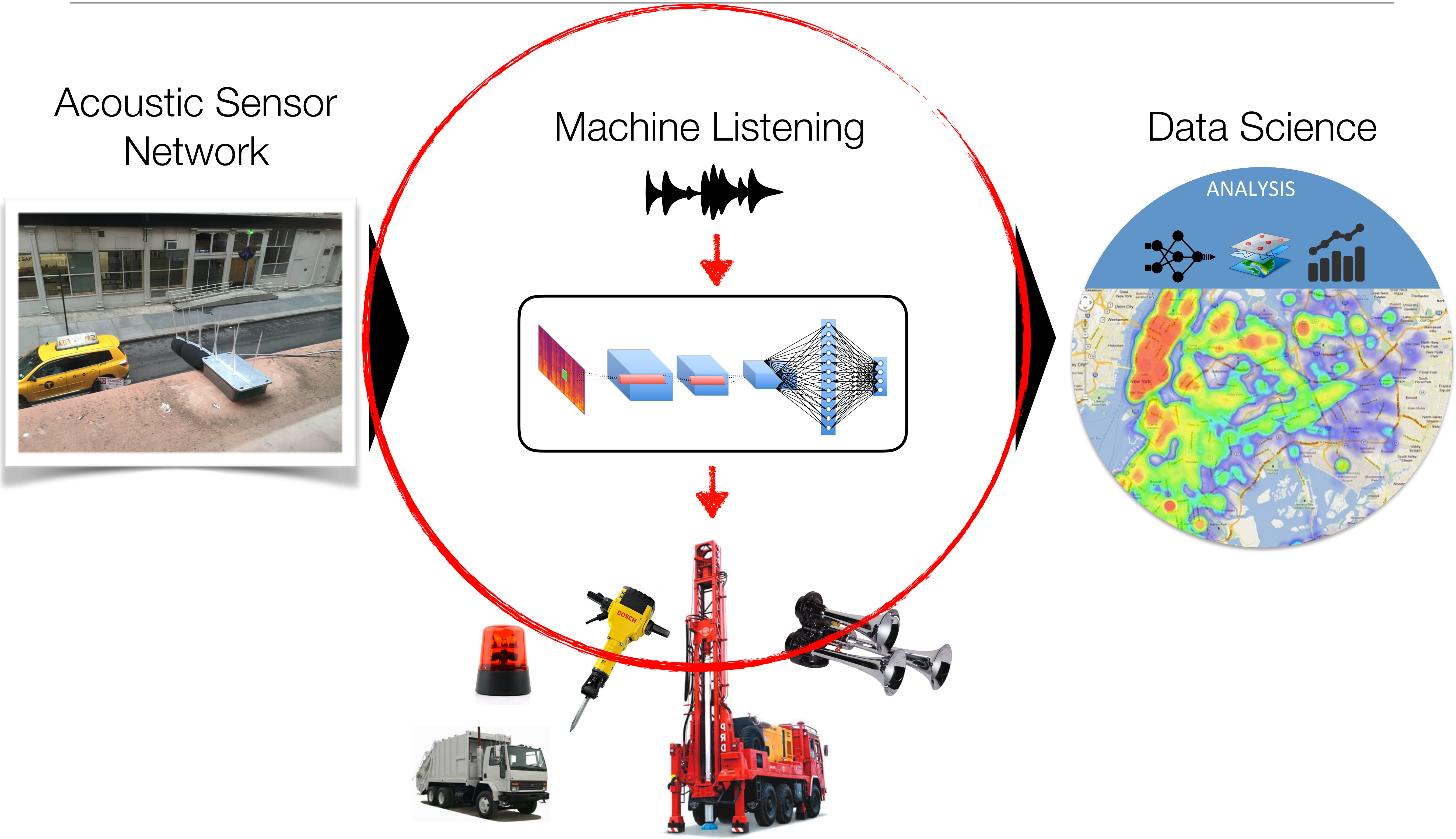
New York University
Music and Audio Research Lab
Department of Computer Science and Engineering

Crowdsourcing Multi-label Audio Annotation Tasks with Citizen Scientists

Mark Cartwright, Graham Dove, Ana Elisa Méndez Méndez,
Juan P. Bello, Oded Nov

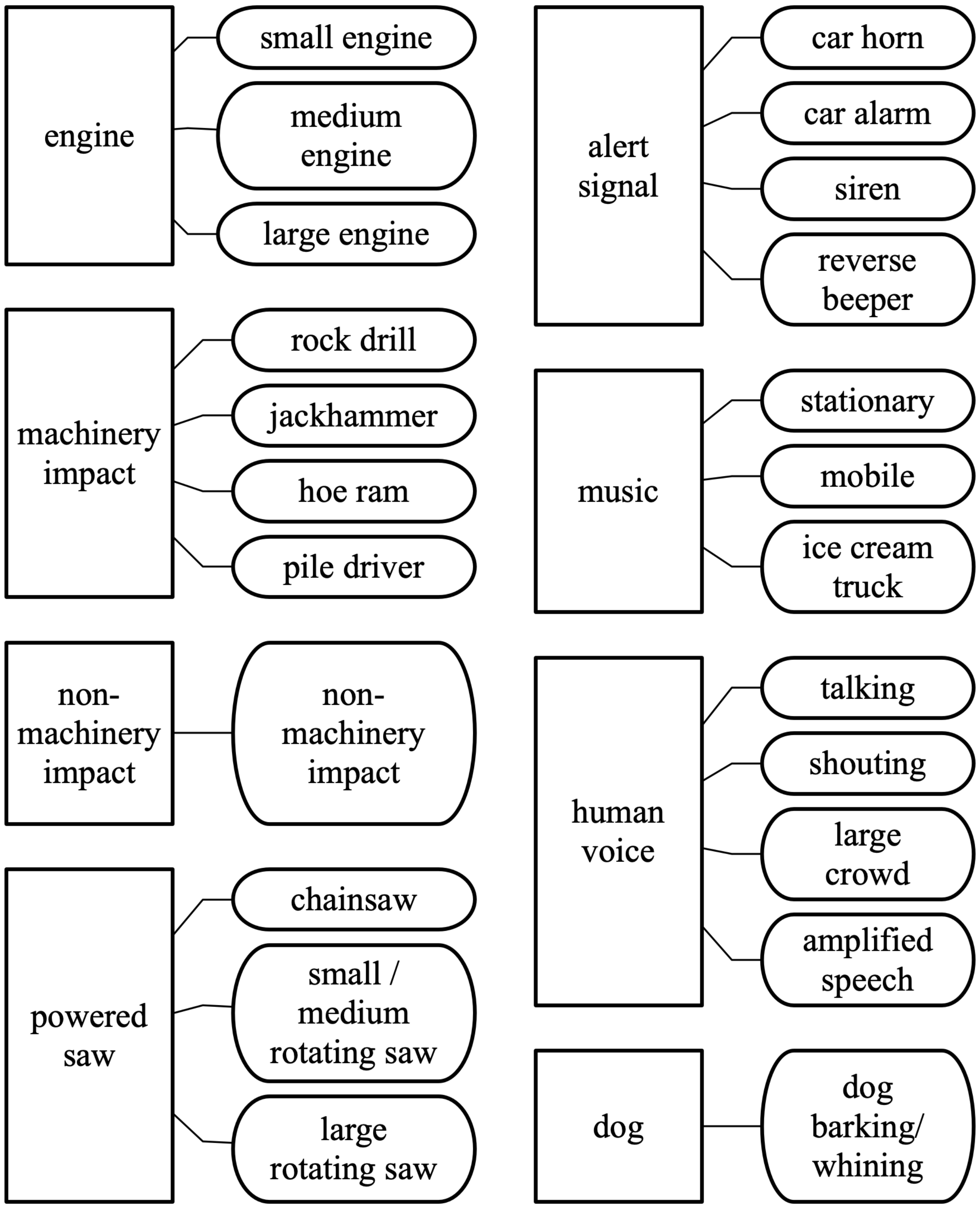
New York University
Music and Audio Research Lab
Department of Computer Science and Engineering


SENSING
ACTUATION
ANALYSIS

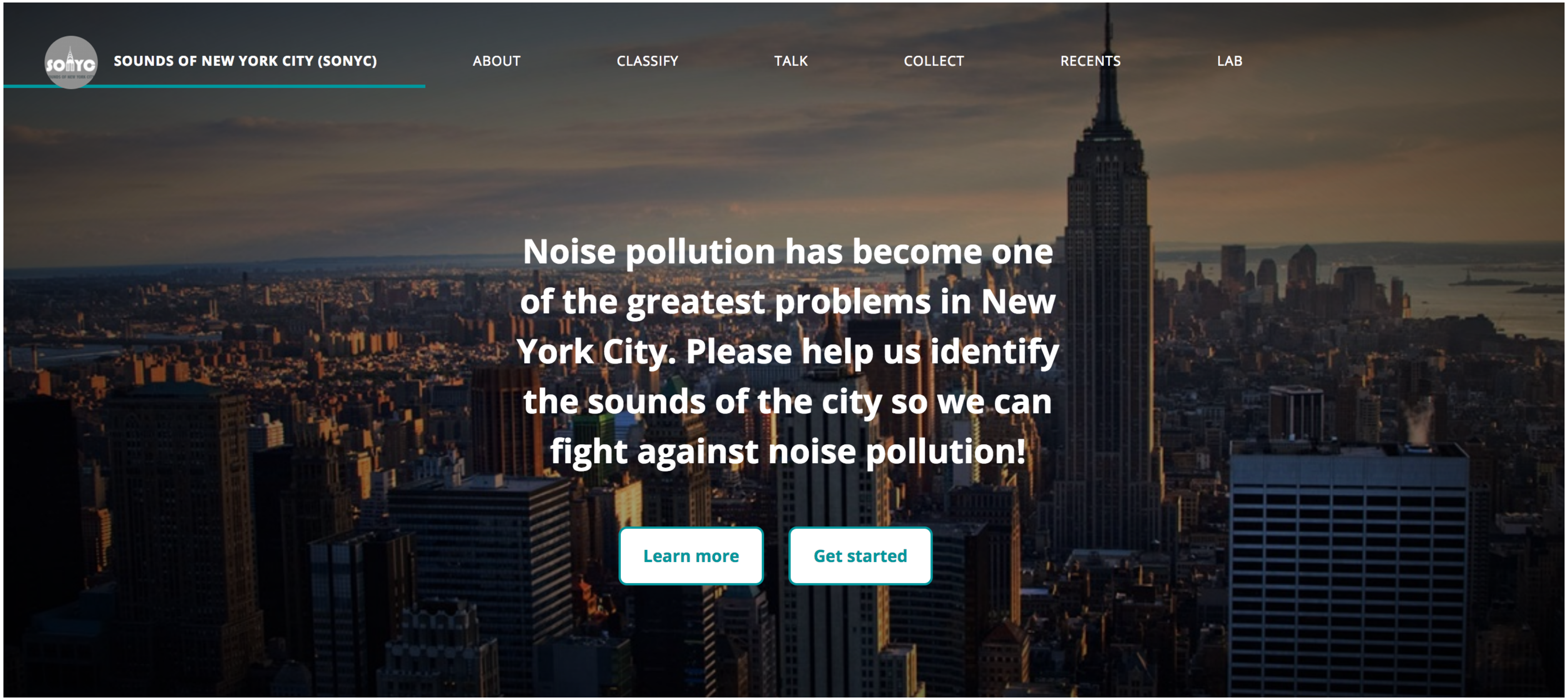

Acoustic Sensor Network
Machine Listening
Data Science
ANALYSIS
BOSCH

60	130,000,000	41
Sensors	Recordings	Years of Audio

SONYC Urban Sound Tagging Classes



Citizen Science Audio Annotation Campaign



How does the type of multi-label annotation task affect throughput and quality?

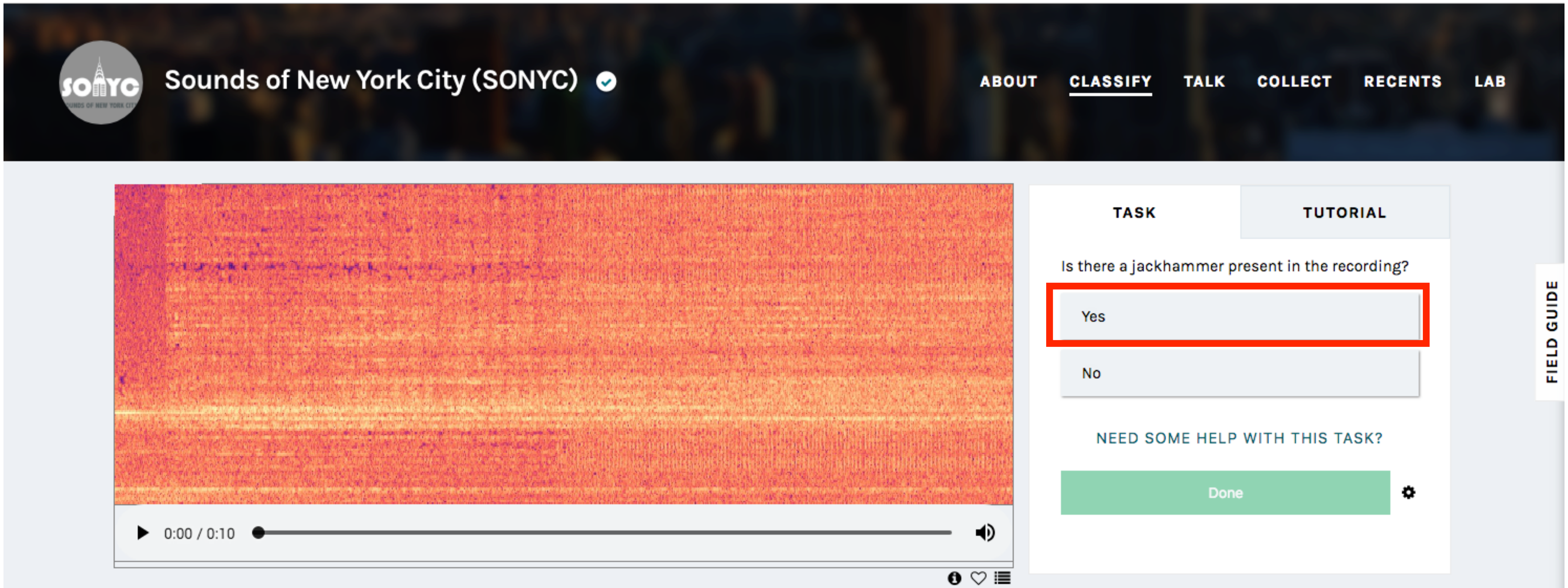
- Do we adopt norms of paid crowdsourcing audio tasks* and break annotation into ***multiple binary annotation*** tasks?

- Or do we adopt norms of image annotation with citizen scientists and use ***multi-label annotation*** tasks?

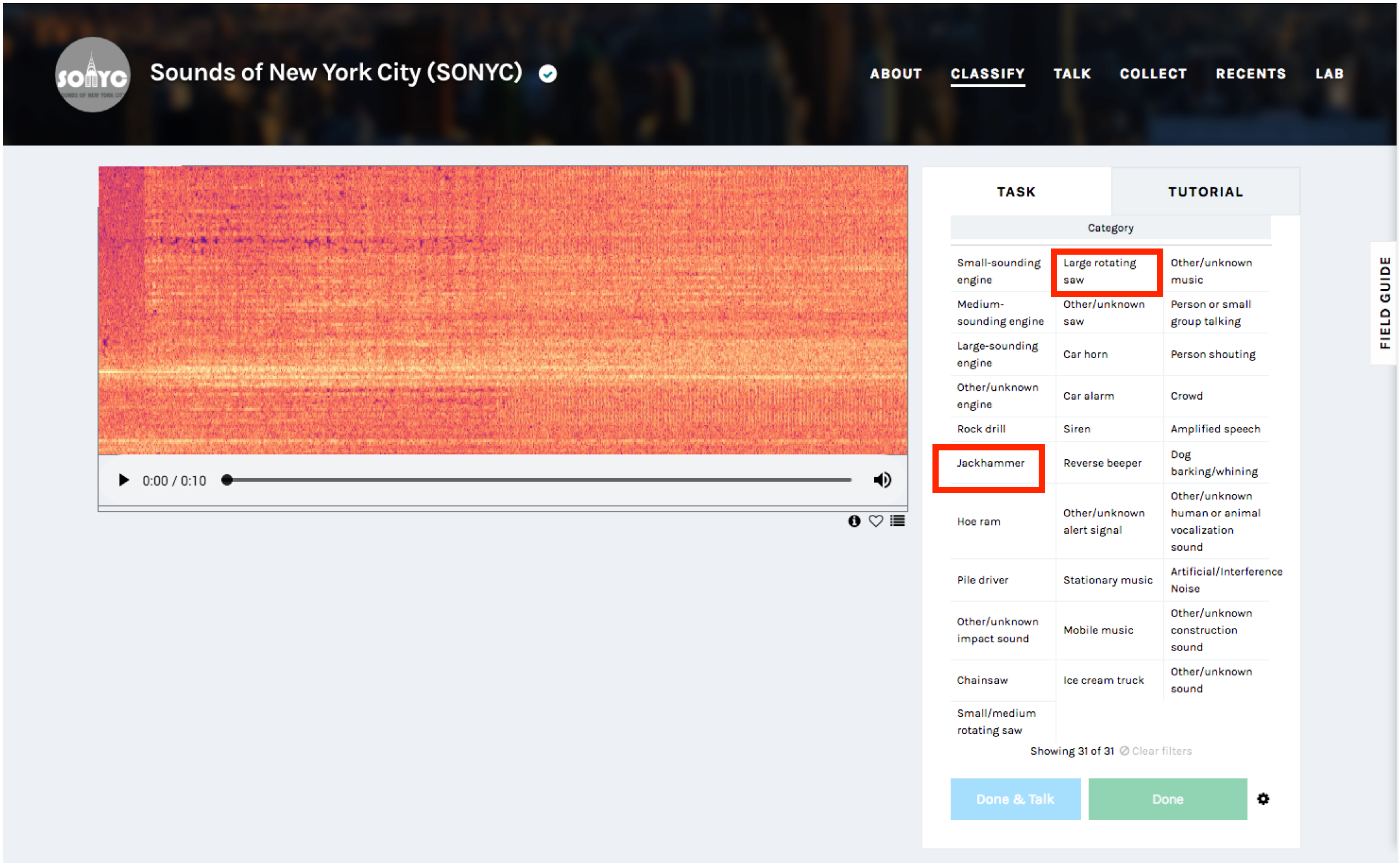
*Lawrence, R Channing Moore, Manoj Plakal, and Marvin Ritter. 2017. *Audio Set: An ontology and human-labeled dataset for audio events*. In Proceedigns of the IEEE International Conference on Acoustics, Speech, and Signal Processing

*Eric Humphrey, Simon Durand, and Brian McFee. 2018. *OpenMIC-2018: an open dataset for multiple instrument recognition*. In Proceedings of the International Societv for Music Information Retrieval Conference.

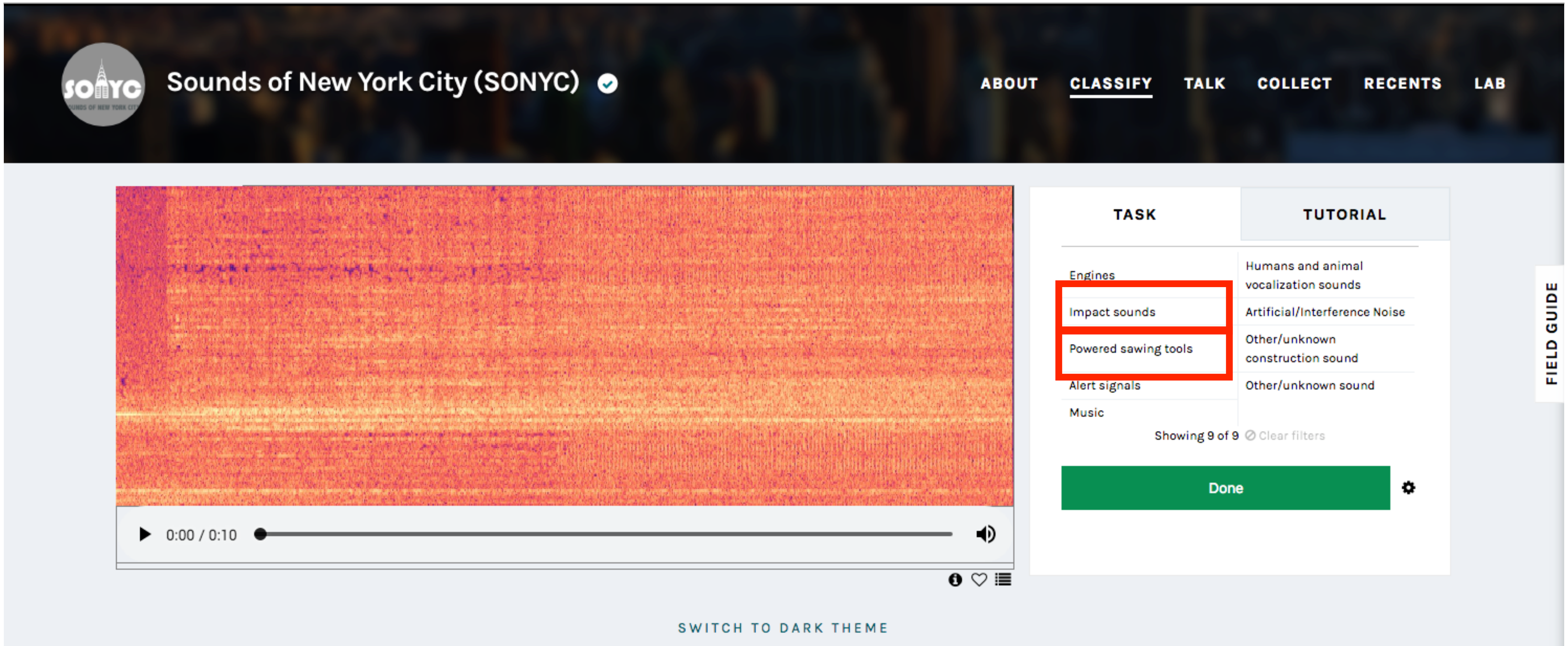
Binary-labeling Annotation Task



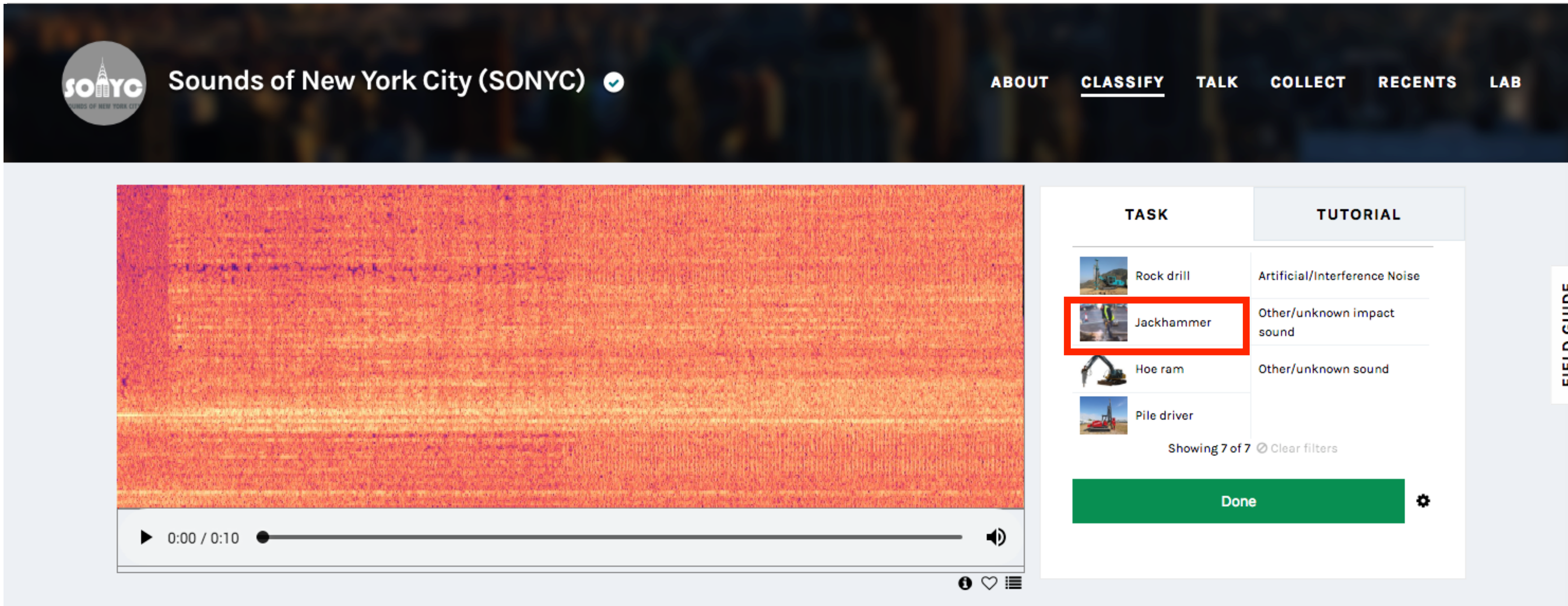
Multi-label Annotation Task



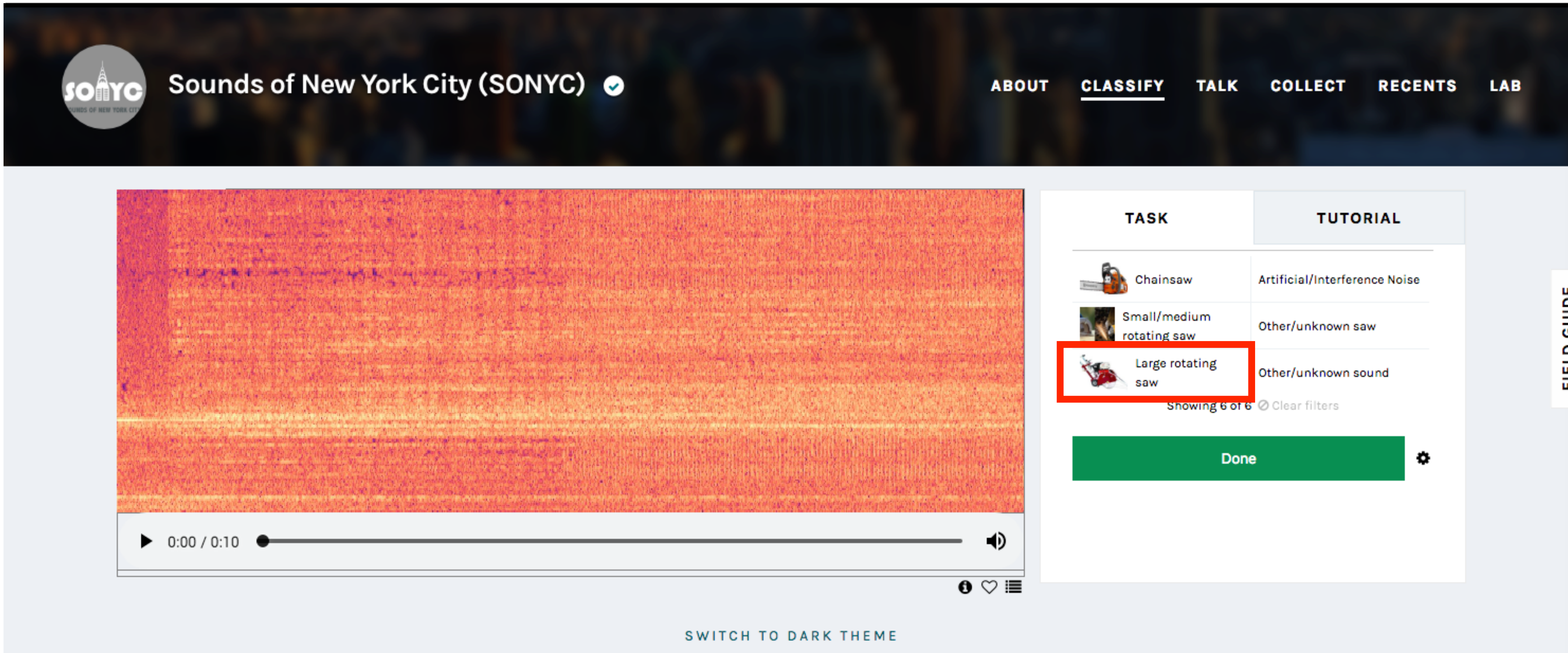
Hierarchical Multi-label Annotation Task



Hierarchical Multi-label Annotation Task

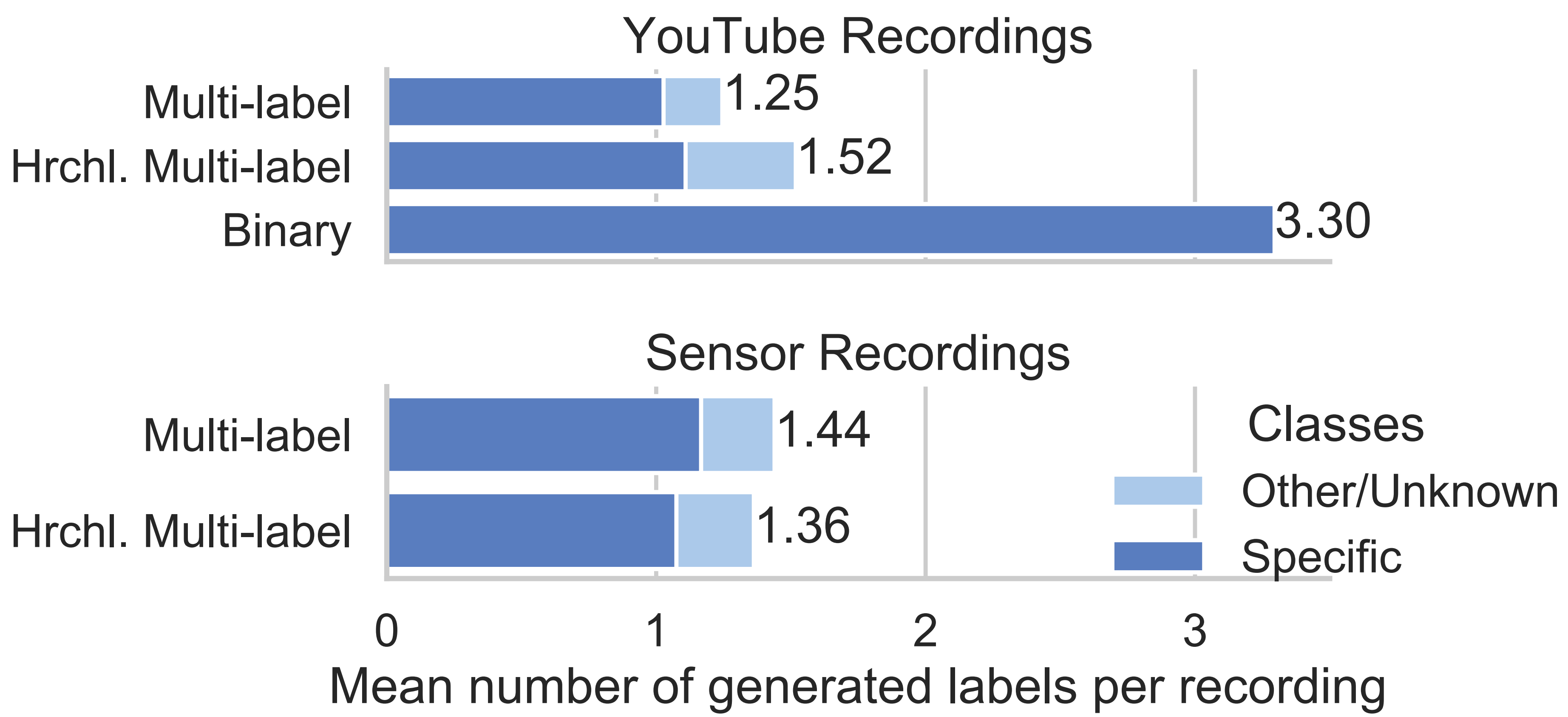


Hierarchical Multi-label Annotation Task



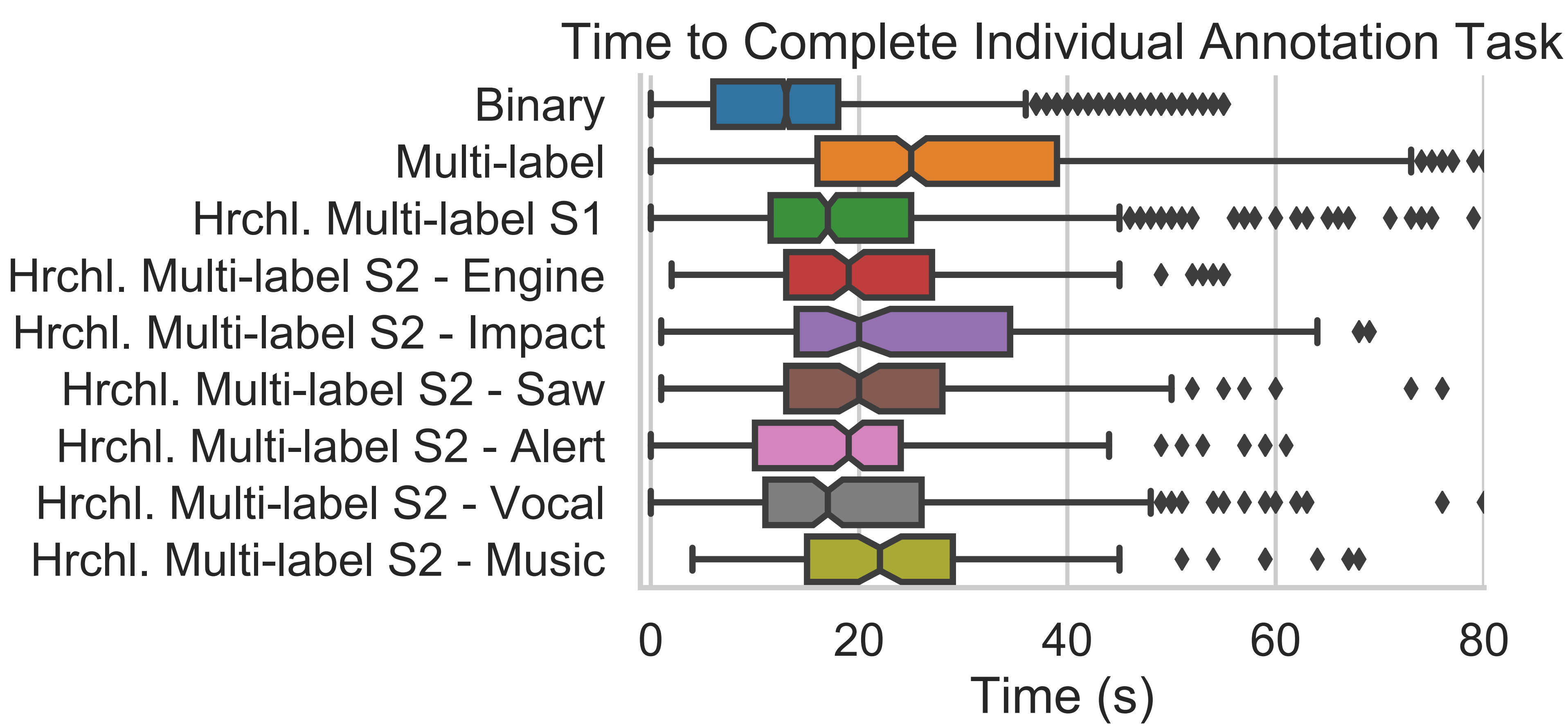
Annotation Throughput

- Binary labeling task generated more overall positive labels per recording



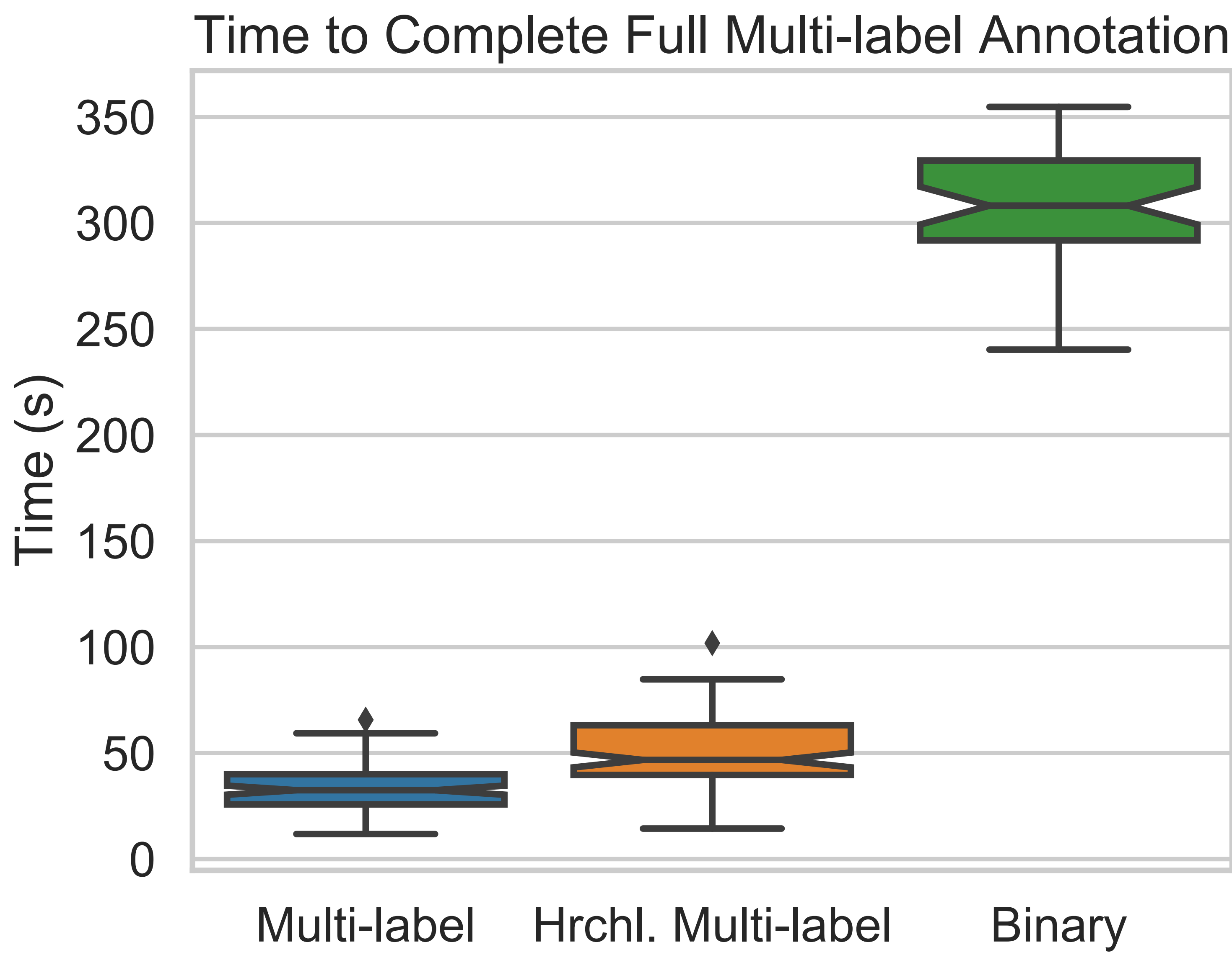
Annotation Throughput

- Binary labeling task took half as long as multi-label for an individual annotation

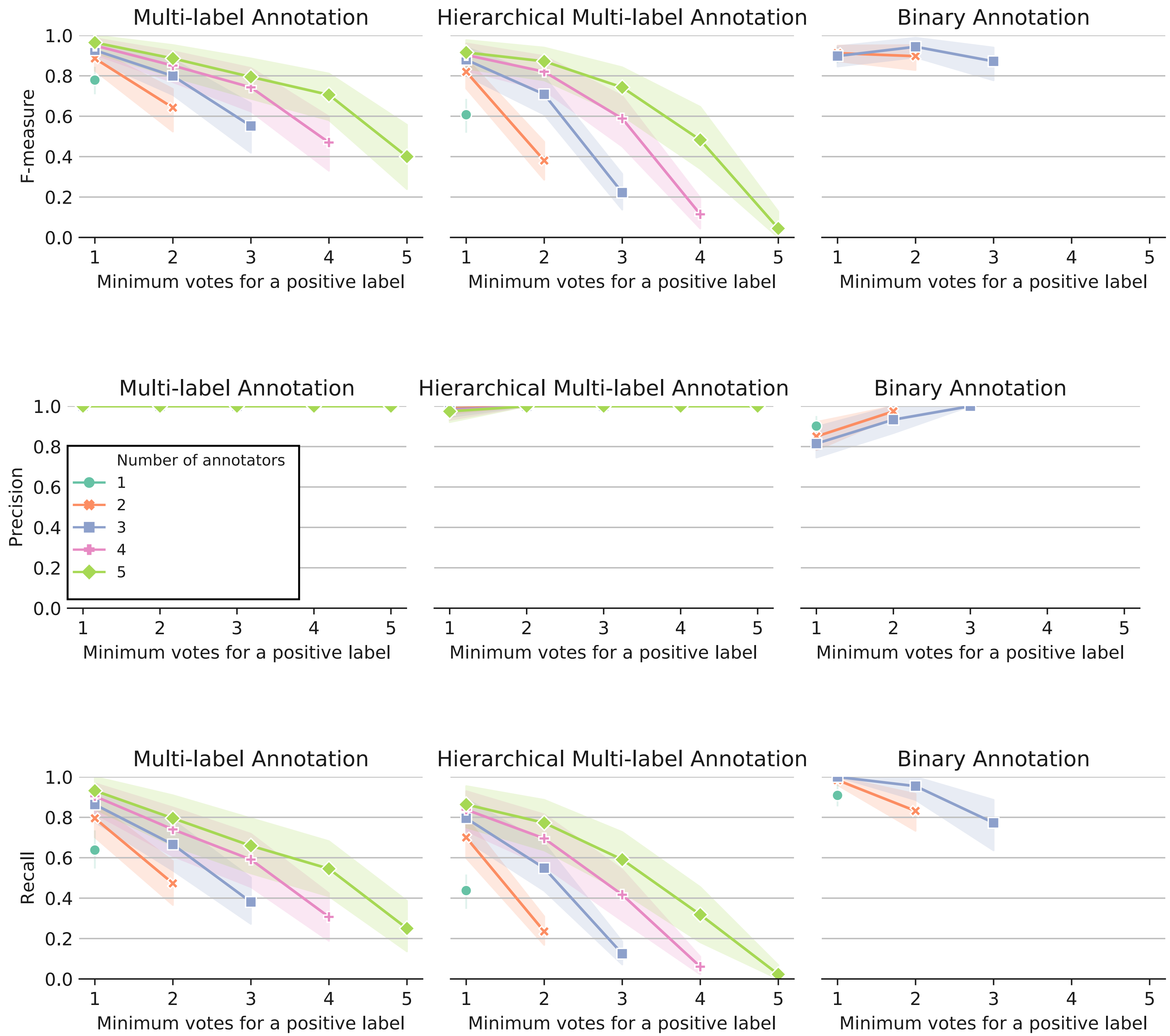


Annotation Throughput

- However, for a full 23 class multi-label annotation binary labeling took 9x as long as multi-labeling



Annotation Quality



Feedback from Participants (Binary Labeling)

- "*There might be a better way than is that X sound yes or no to classify quicker. People will get tired of listening to sound clips faster than other quick options, like the animal diaries.* ***You want to squeeze as much data out of each audio clip***."

- *"I hear drums, observer/audience yelling applause, at least one large size dog that is very unhappy about the noise. This takes place outside.* ***I have no way to label more than two features, so it will probably be more frustrating than I can deal with to participate****."*

- *"In my opinion,* ***this project should*** *use the same model as the animal camera trap projects, that is,* ***have a list of sound categories that one can click on for each clip****, and give the opinion to choose more than one category."*

Conclusions of Study

- Overall quality of multi-label annotations from binary and multi-label tasks are comparable. They have differences but they can be balanced.

- Multi-label is much more efficient, but only if you need full multi-label annotation

- Hierarchical multi-label tends to propagates error, leading to lower recall

- Informal feedback indicates that volunteers much preferred multi-label, opposite of paid crowdworkers

- Results side with the common practice of citizen science image annotation rather than that of paid audio crowdsourcing.

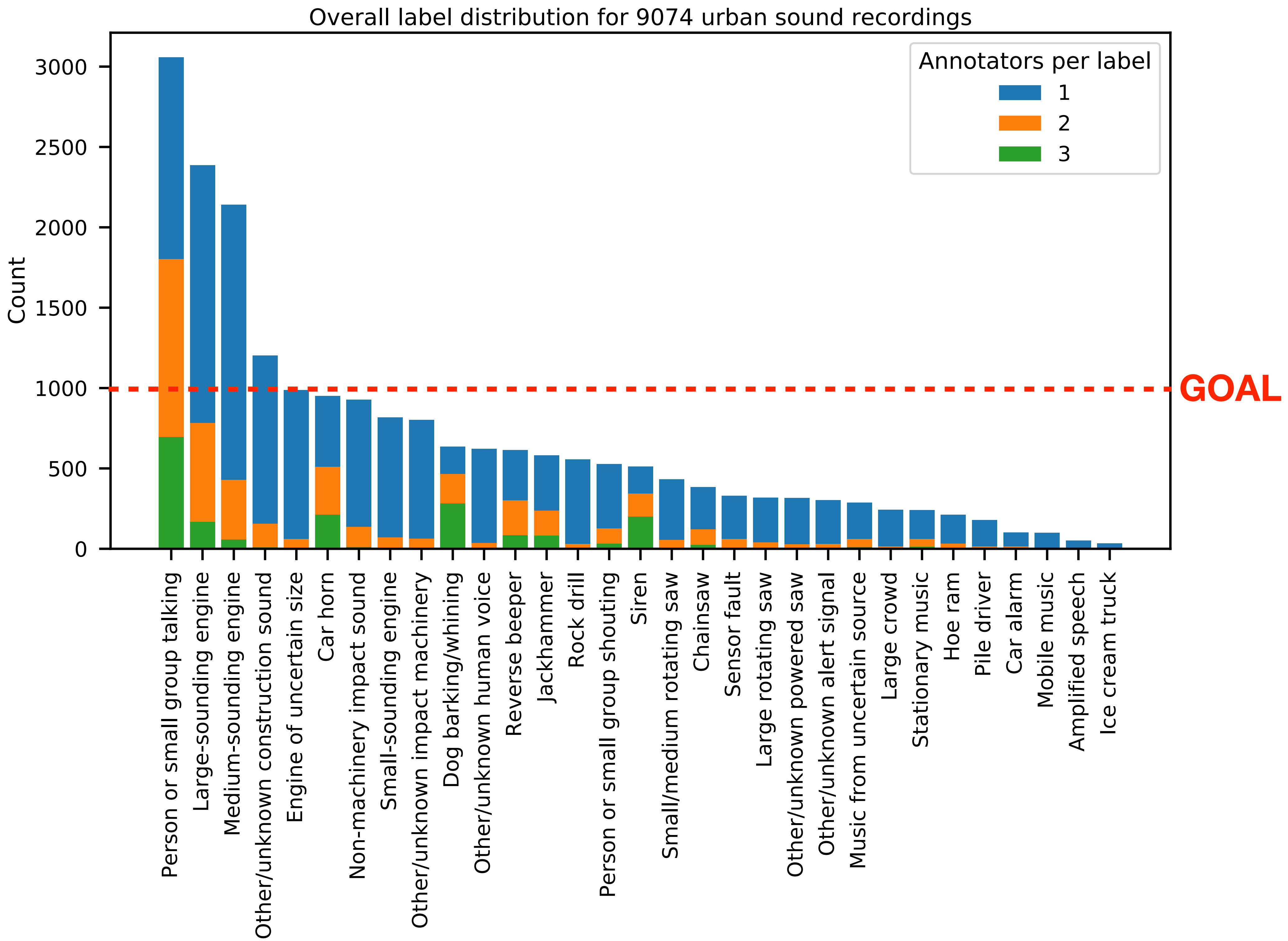
Ongoing Citizen Science Annotation Campaign

1,051
Registered Annotators

30,376
Full Multi-label Annotations

9,765
Completed Recordings

Ongoing Citizen Science Annotation Campaign



SONYC Urban Sound Tagging Dataset

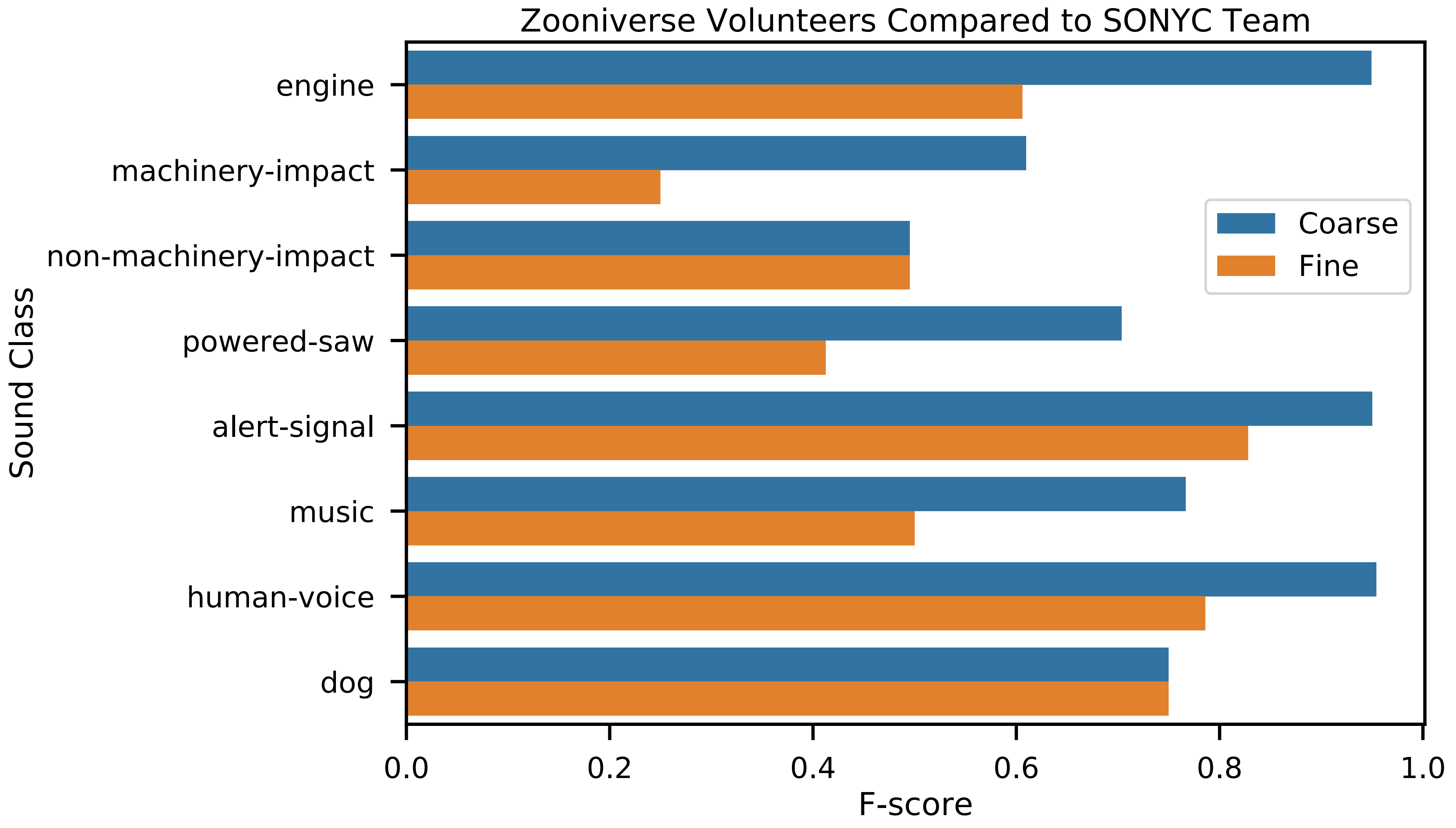
- Released in March
- 2351 training recordings and 443 validation
- Multi-label annotation on 23 classes
- 3 Zooniverse annotators per recording
- Validation set annotated by SONYC team
- https://doi.org/10.5281/zenodo.2590742

DCASE 2019 Challenges Tasks:

- Acoustic scene classification
- Audio tagging with noisy labels and minimal supervision
- Sound event localization and detection
- Sound event detection in domestic environments
- **Urban Sound Tagging**

SONYC Urban Sound Tagging Dataset

- How do annotations from Zooniverse volunteers compare to those of the SONYC team?



Conclusions of Study

- Overall quality of multi-label annotations from binary and multi-label tasks are comparable. They have differences but they can be balanced.

- Multi-label is much more efficient, but only if you need full multi-label annotation

- Hierarchical multi-label tends to propagates error, leading to lower recall

- Informal feedback indicates that volunteers much preferred multi-label, opposite of paid crowdworkers

- Results side with the common practice of citizen science image annotation rather than that of paid audio crowdsourcing.