

# 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



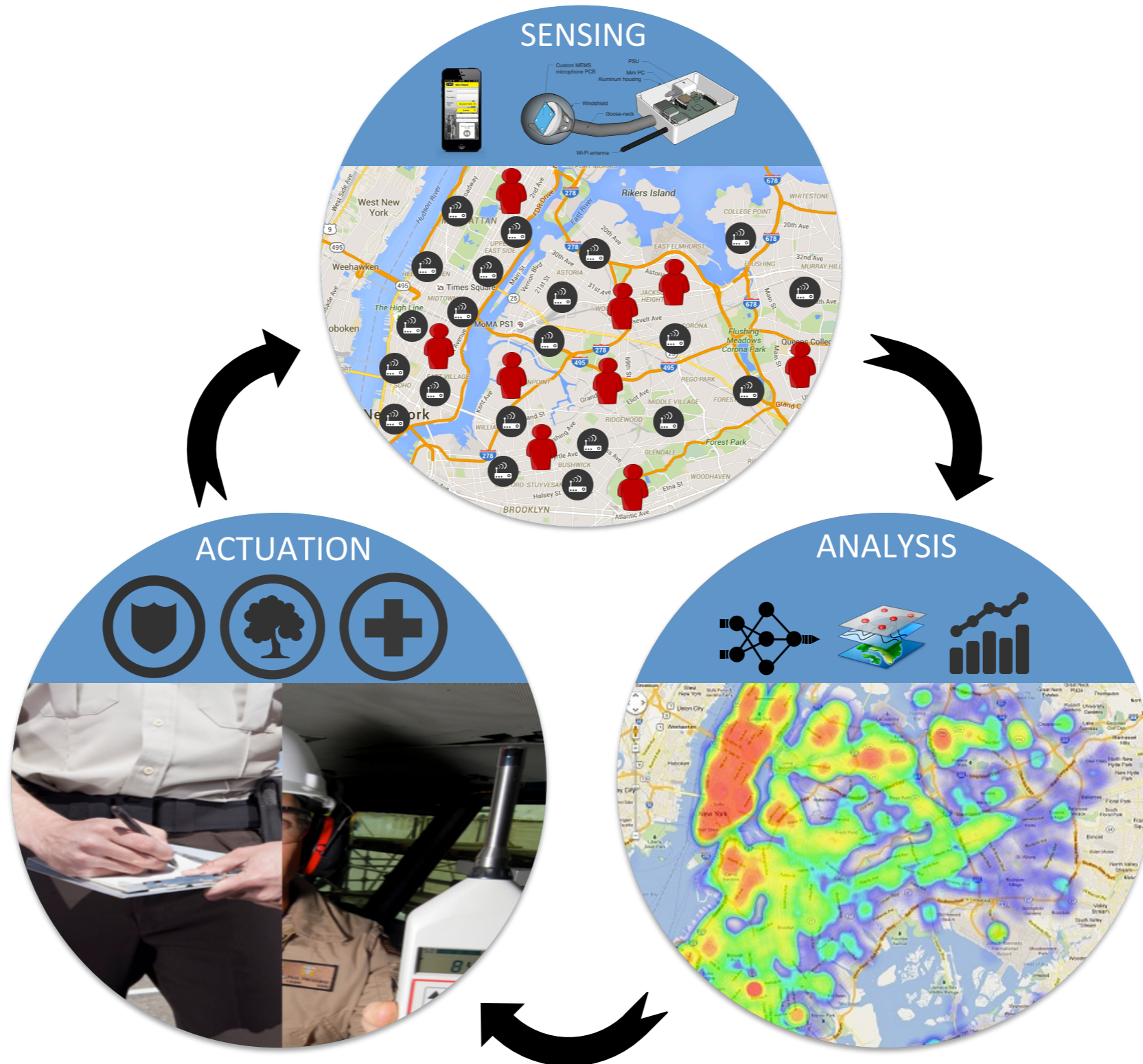
# Crowdsourcing Multi-label Audio Annotation Tasks with Citizen Scientists

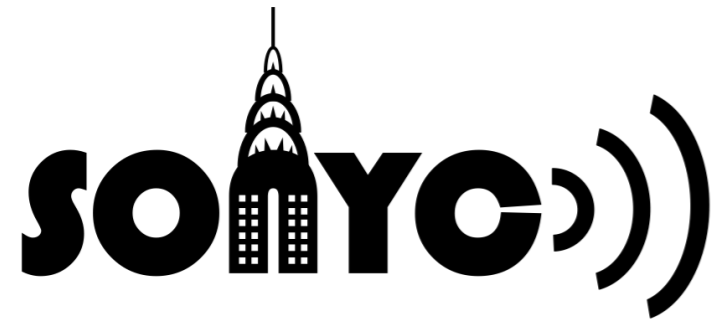
---

**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



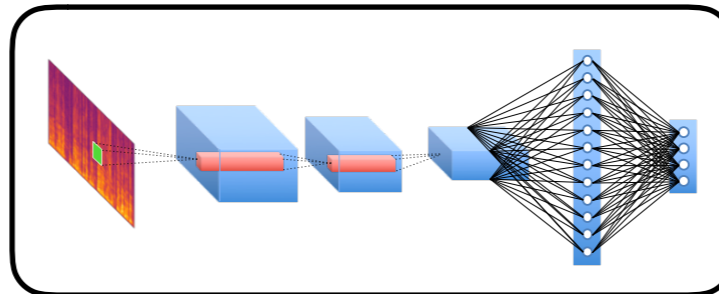




## Acoustic Sensor Network

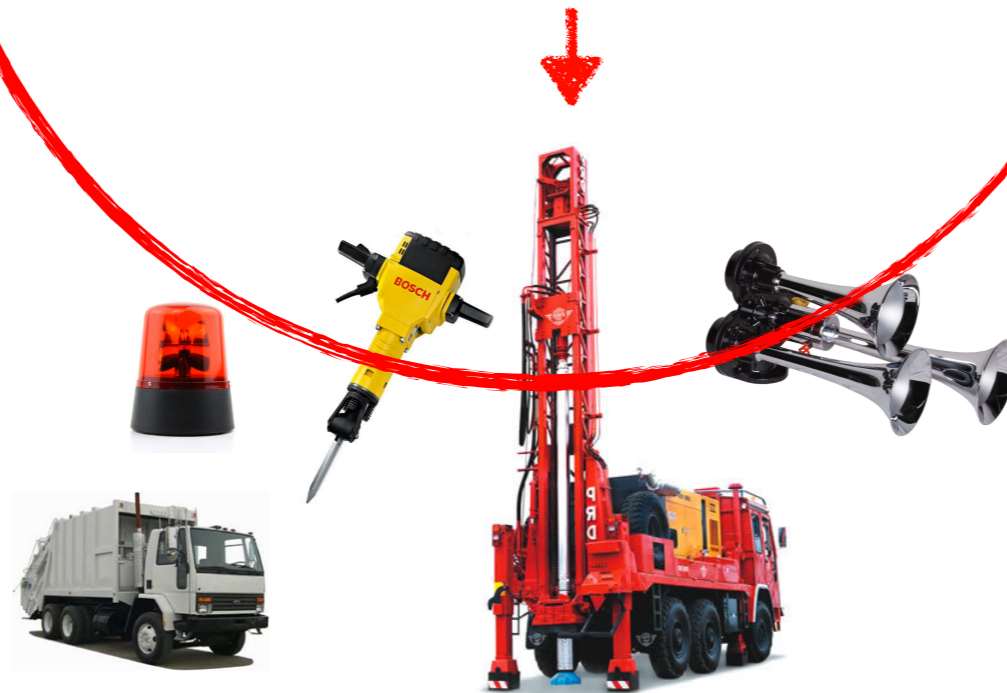
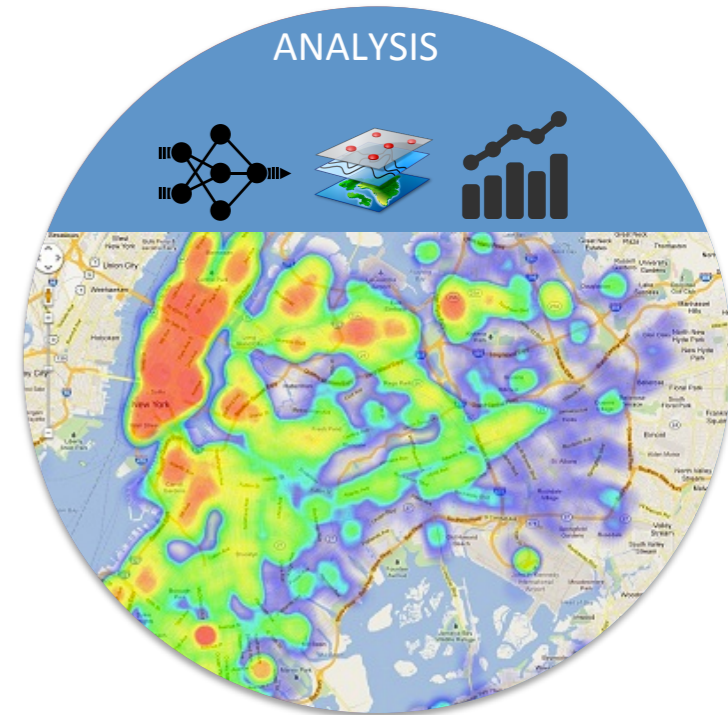


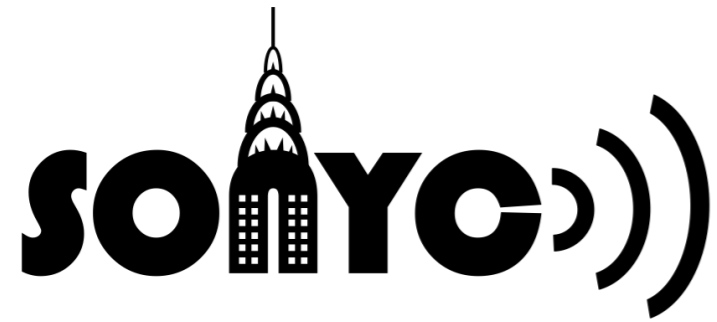
## Machine Listening



## Data Science

ANALYSIS





**60**

Sensors

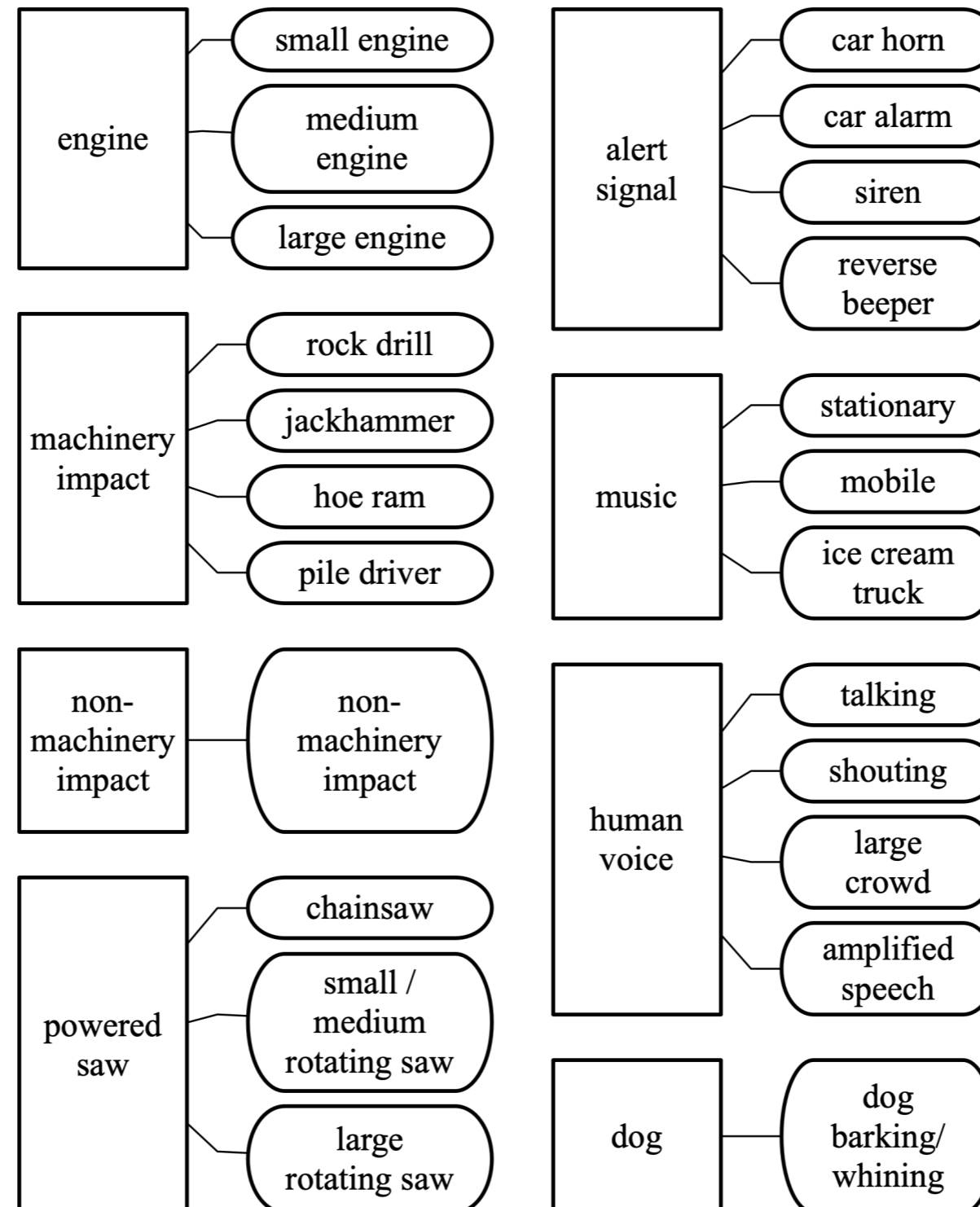
**130,000,000**

Recordings

**41**

Years of Audio

# SONYC Urban Sound Tagging Classes



# Citizen Science Audio Annotation Campaign

---

**SONYC** SOUNDS OF NEW YORK CITY (SONYC) ABOUT CLASSIFY TALK COLLECT RECENTS LAB

**Noise pollution has become one of the greatest problems in New York City. Please help us identify the sounds of the city so we can fight against noise pollution!**

[Learn more](#) [Get started](#)

# 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 Proceedings 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 Society for Music Information Retrieval Conference.




# Binary-labeling Annotation Task

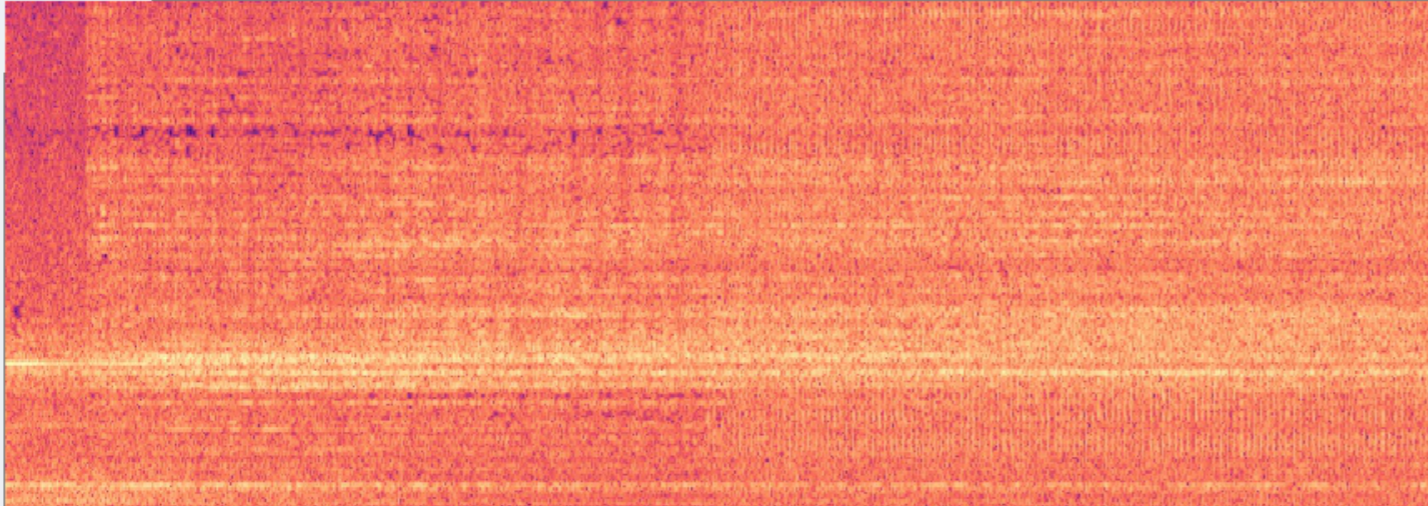
The screenshot displays the Sounds of New York City (SONYC) website interface. At the top left is the SONYC logo and the text "Sounds of New York City (SONYC)". The top navigation bar includes links for "ABOUT", "CLASSIFY", "TALK", "COLLECT", "RECENTS", and "LAB". The main content area features a large spectrogram of an audio recording on the left, with a video player interface below it showing a progress bar at 0:00 / 0:10. On the right, there is a task panel with two tabs: "TASK" and "TUTORIAL". The "TASK" tab is active, displaying the question "Is there a jackhammer present in the recording?". Below the question are two radio button options: "Yes" and "No". The "Yes" option is highlighted with a red rectangular border. Below the options is a section titled "NEED SOME HELP WITH THIS TASK?" with a green "Done" button and a settings gear icon. A vertical "FIELD GUIDE" label is positioned on the far right edge of the interface.



# Multi-label Annotation Task


Sounds of New York City (SONYC) ✔

ABOUT
CLASSIFY
TALK
COLLECT
RECENTS
LAB



▶ 0:00 / 0:10🔊

ⓘ
♥
☰

TASK
TUTORIAL

Category		
Small-sounding engine	Large rotating saw	Other/unknown music
Medium-sounding engine	Other/unknown saw	Person or small group talking
Large-sounding engine	Car horn	Person shouting
Other/unknown engine	Car alarm	Crowd
Rock drill	Siren	Amplified speech
Jackhammer	Reverse beeper	Dog barking/whining
Hoe ram	Other/unknown alert signal	Other/unknown human or animal vocalization sound
Pile driver	Stationary music	Artificial/Interference Noise
Other/unknown impact sound	Mobile music	Other/unknown construction sound
Chainsaw	Ice cream truck	Other/unknown sound
Small/medium rotating saw		

Showing 31 of 31 🗑️ Clear filters

Done & Talk
Done
⚙️

FIELD GUIDE

# Hierarchical Multi-label Annotation Task

The screenshot displays the Sounds of New York City (SONYC) web application. At the top left is the SONYC logo and the text "Sounds of New York City (SONYC)". The top navigation bar includes links for "ABOUT", "CLASSIFY", "TALK", "COLLECT", "RECENTS", and "LAB". The main content area features a large spectrogram on the left and a classification task panel on the right. The spectrogram shows a complex sound waveform with a time axis at the bottom ranging from 0:00 to 0:10. The classification panel is divided into "TASK" and "TUTORIAL" sections. The "TASK" section lists categories: Engines, Impact sounds, Powered sawing tools, Alert signals, and Music. The "TUTORIAL" section lists: Humans and animal vocalization sounds, Artificial/Interference Noise, Other/unknown construction sound, and Other/unknown sound. A "Done" button is visible at the bottom of the task panel. A vertical "FIELD GUIDE" label is on the right edge. At the bottom center, there is a link to "SWITCH TO DARK THEME".

# Hierarchical Multi-label Annotation Task

The screenshot displays the Sounds of New York City (SONYC) website interface. At the top left is the SONYC logo and the text "Sounds of New York City (SONYC)". To the right are navigation links: ABOUT, CLASSIFY, TALK, COLLECT, RECENTS, and LAB. The main content area is split into two sections. On the left is a spectrogram of an audio recording, with a playback control bar below it showing "0:00 / 0:10". On the right is a classification task panel with two tabs: "TASK" and "TUTORIAL". The "TASK" tab is active and shows a list of sound categories with small icons: "Rock drill", "Jackhammer", "Hoe ram", and "Pile driver". The "Jackhammer" entry is highlighted with a red border. To the right of the "TASK" list, the "TUTORIAL" tab shows a list of hierarchical labels: "Artificial/Interference Noise", "Other/unknown impact sound", and "Other/unknown sound". Below the list is a "Showing 7 of 7" indicator and a "Clear filters" button. At the bottom of the panel is a green "Done" button and a settings gear icon. On the far right edge of the interface, there is a vertical "FIELD GUIDE" label.

# Hierarchical Multi-label Annotation Task

The screenshot displays the Sounds of New York City (SONYC) website interface. At the top left is the SONYC logo and the text "Sounds of New York City (SONYC)". The navigation menu includes "ABOUT", "CLASSIFY", "TALK", "COLLECT", "RECENTS", and "LAB". The main content area features a spectrogram on the left and a classification task panel on the right. The spectrogram shows a frequency spectrum with a prominent horizontal band of energy. The classification panel has two tabs: "TASK" and "TUTORIAL". Under the "TASK" tab, there is a list of sound categories with corresponding icons: "Chainsaw", "Small/medium rotating saw", and "Large rotating saw". The "Large rotating saw" category is highlighted with a red box. Below the list, it says "Showing 6 of 6" and "Clear filters". A green "Done" button is at the bottom of the panel. At the bottom of the page, there is a "SWITCH TO DARK THEME" link.

SONYC Sounds of New York City (SONYC)

ABOUT CLASSIFY TALK COLLECT RECENTS LAB

TASK	TUTORIAL
Chainsaw	Artificial/Interference Noise
Small/medium rotating saw	Other/unknown saw
Large rotating saw	Other/unknown sound

Showing 6 of 6 Clear filters

Done

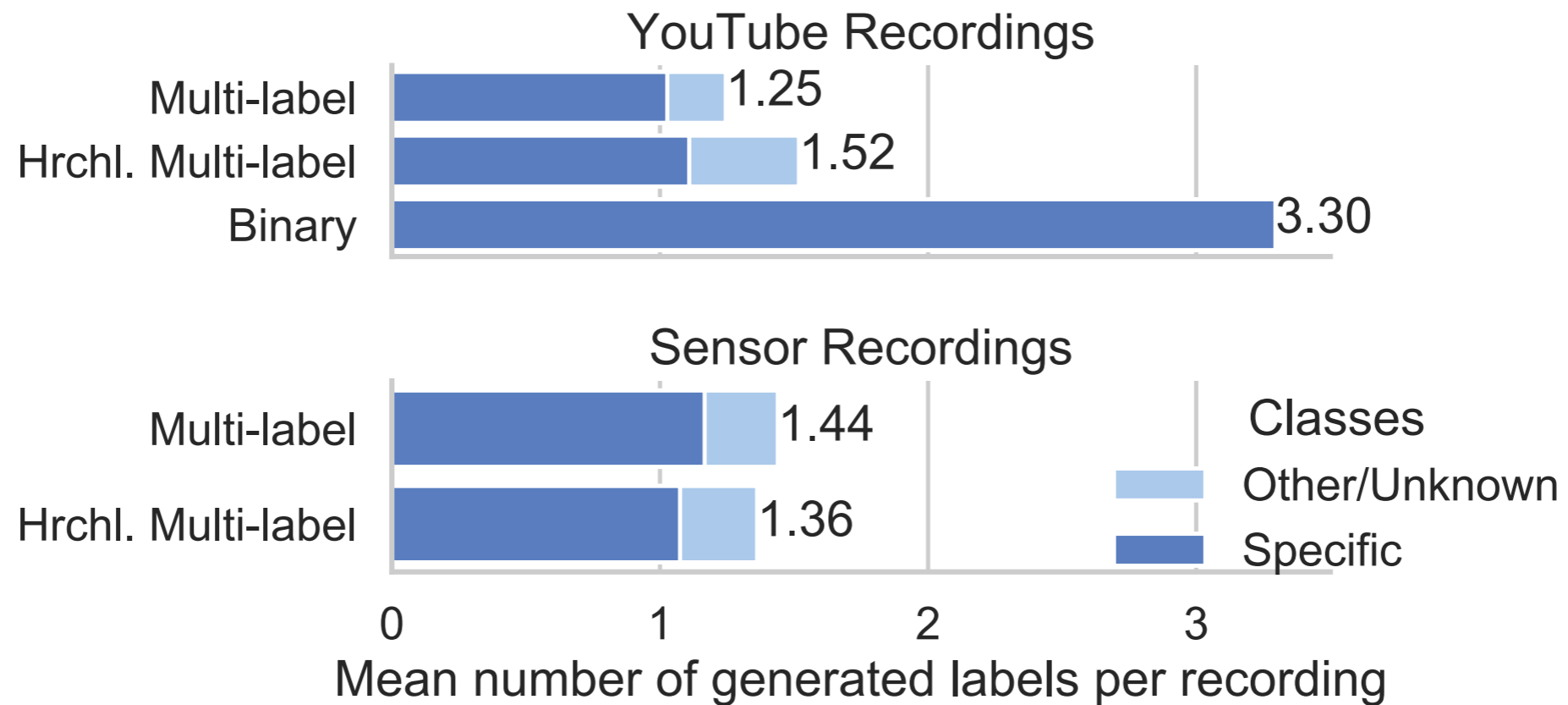
SWITCH TO DARK THEME

FIELD GUIDE

# Annotation Throughput

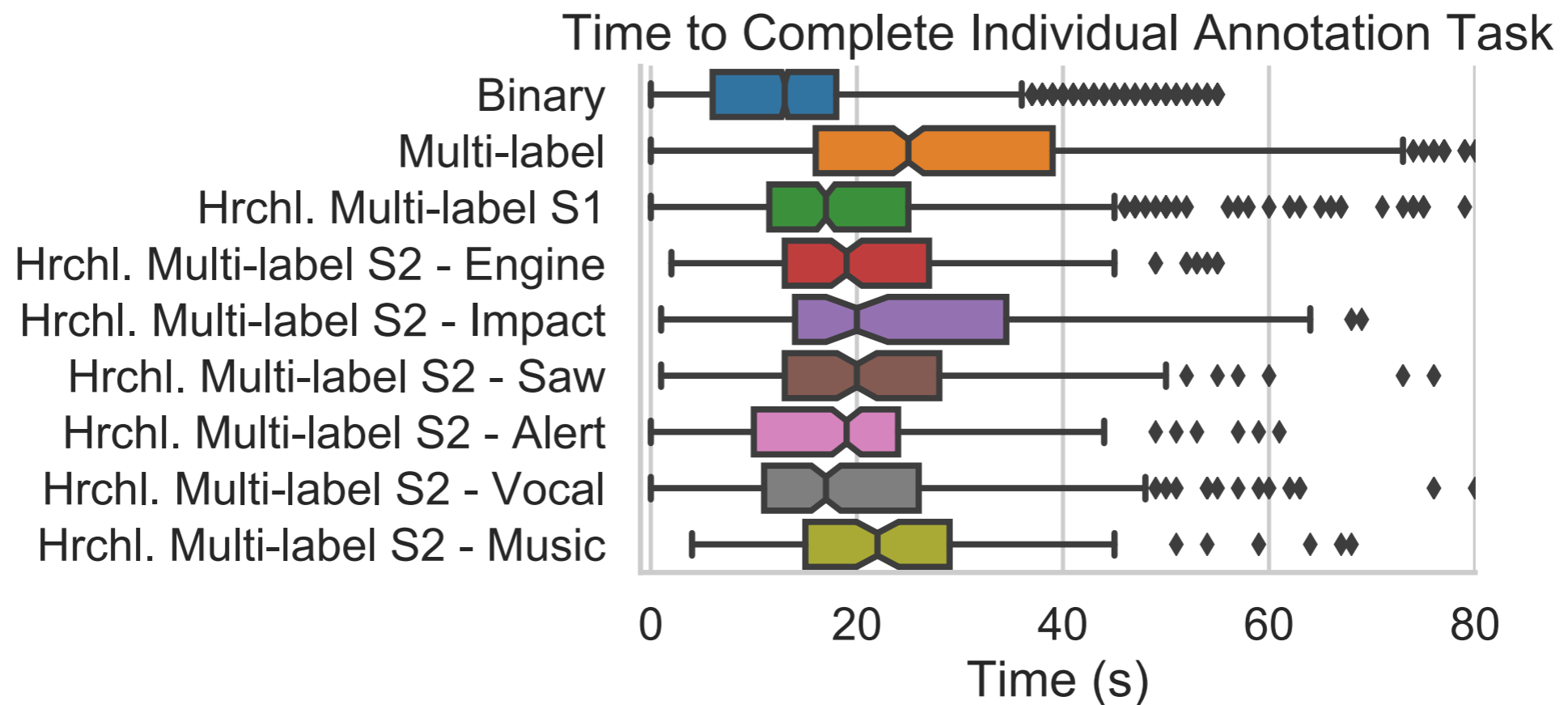
---

- Binary labeling task generated more overall positive labels per recording



# Annotation Throughput

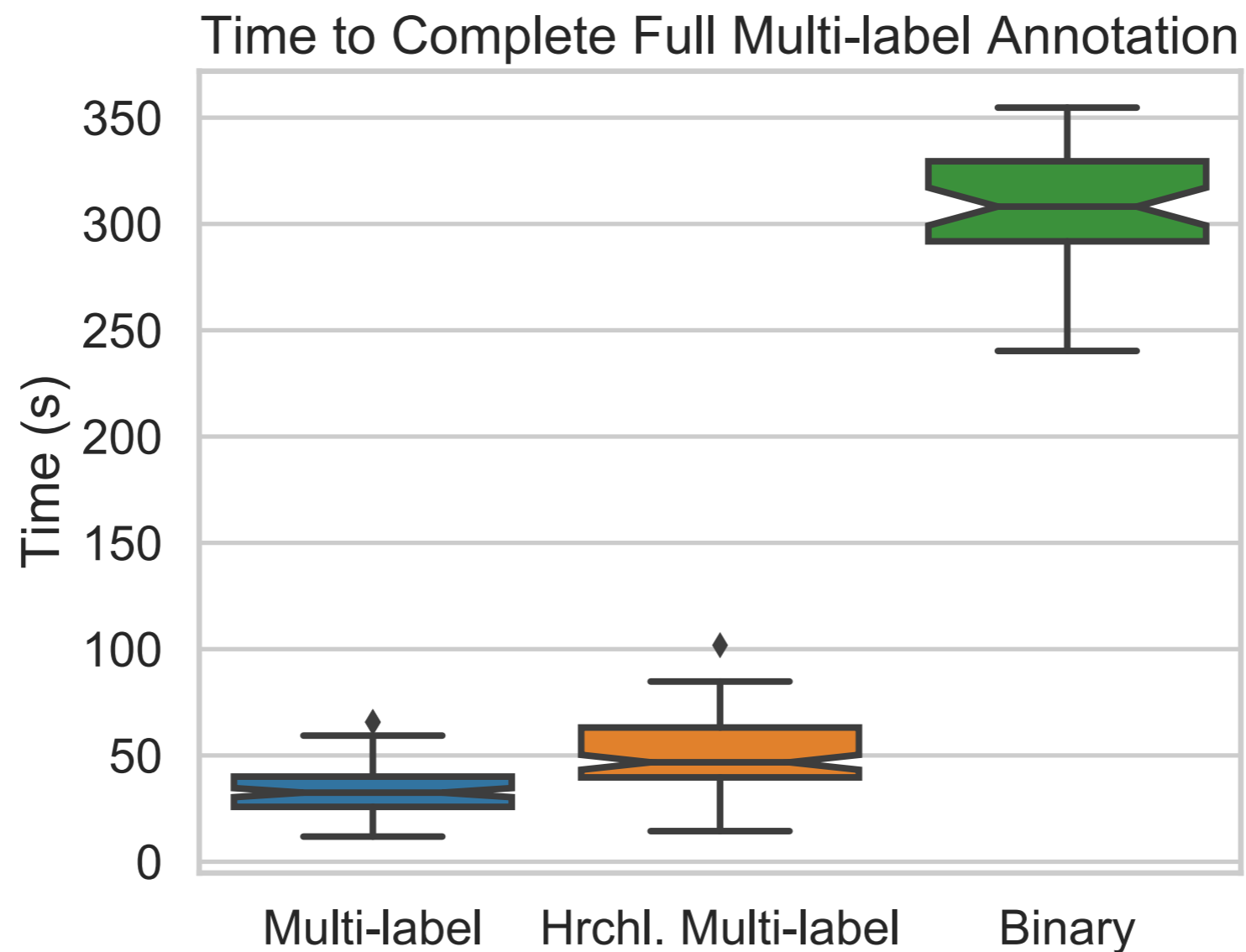
- Binary labeling task took half as long as multi-label for an individual annotation



# Annotation Throughput

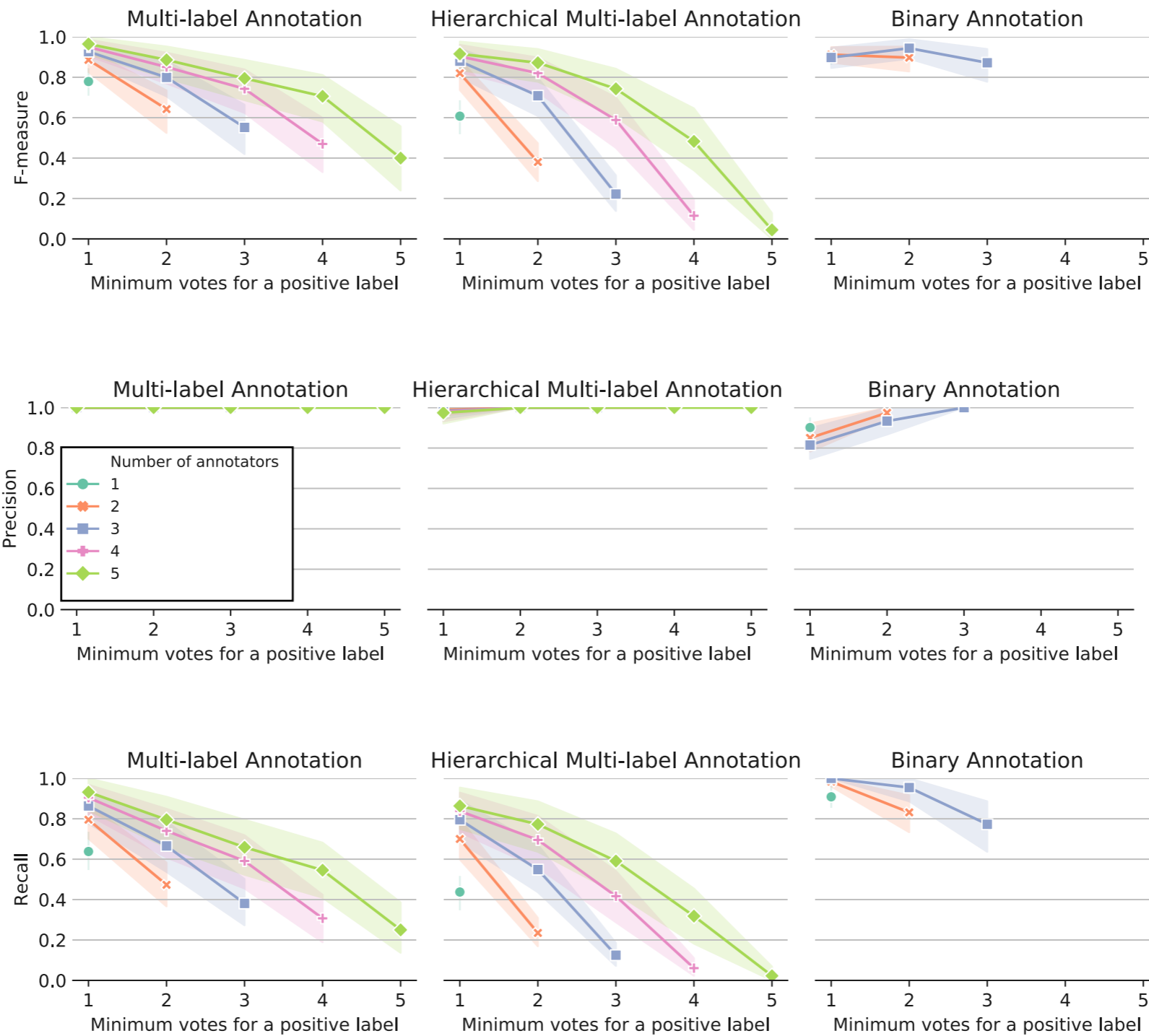
---

- 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 propagate 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

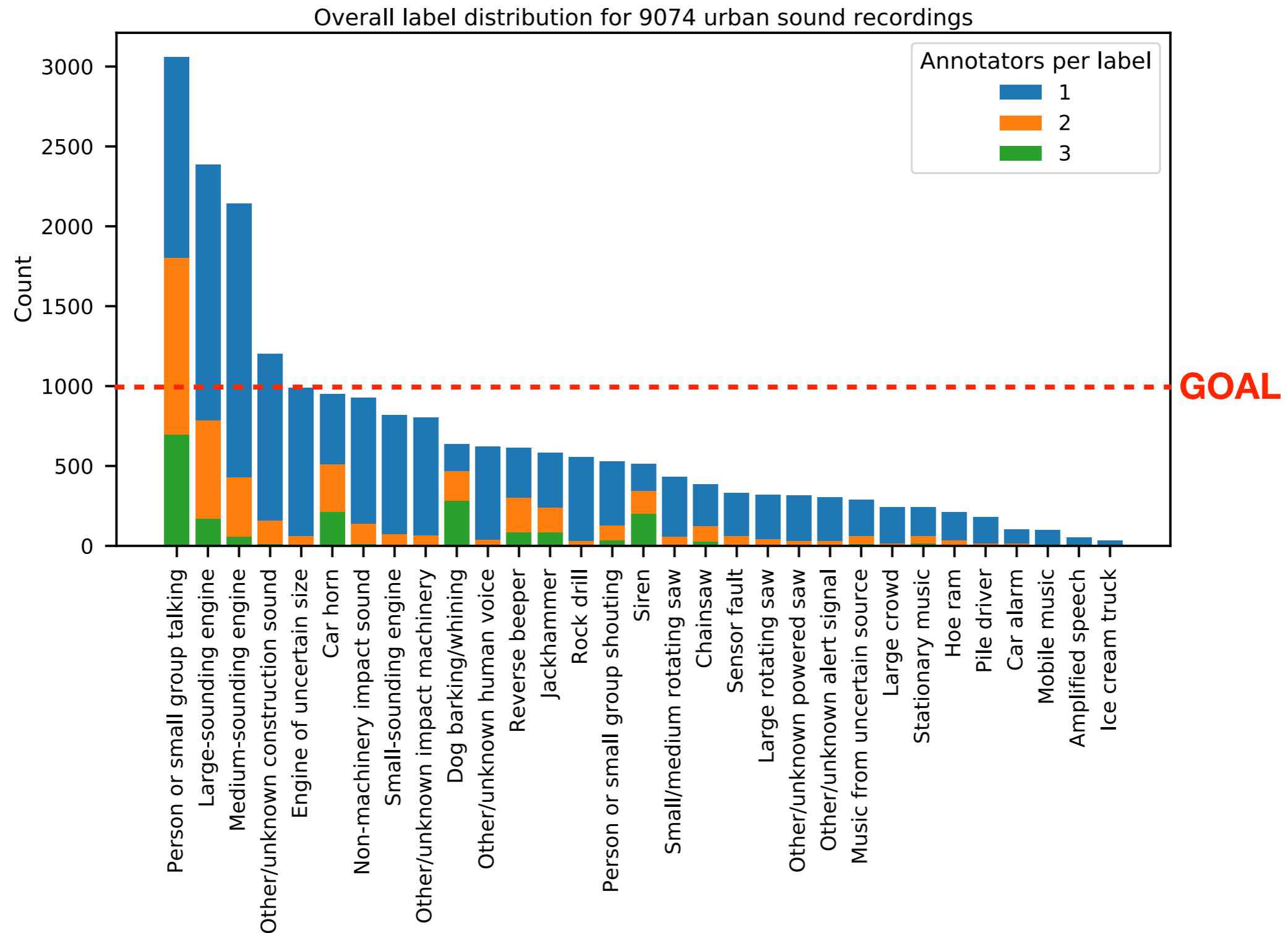
**30,376**

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>

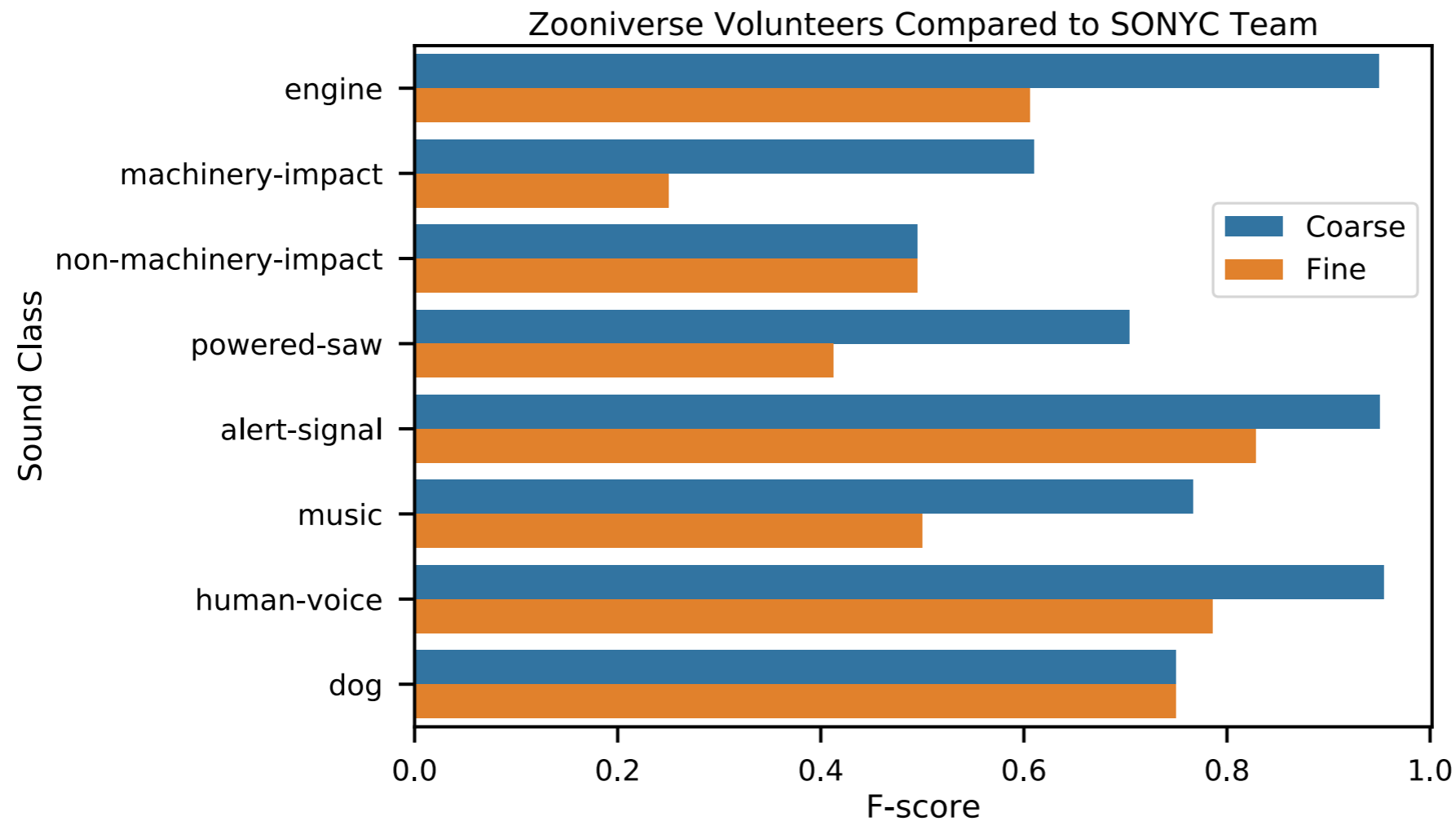


## DCASE 2019 Challenges Tasks:

-  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 propagate 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.