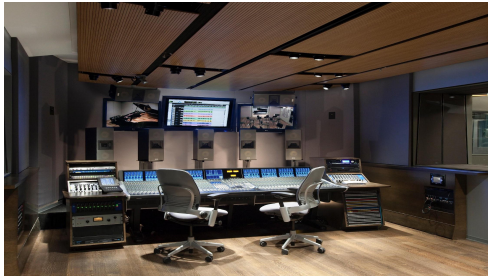
Brian McFee
New York University

NEMISIG-19, Brooklyn College, Brooklyn, NY, USA

2018 at NYU's Music and Audio Research Lab



MARL: founded in late 2008, moved to new facilities in 2009, **moving again in 2019!**
14+ researchers, Funded by NSF, IMLS, NSERC, Fulbright, Industry, NYU

MARL: Areas of Interest

(Anna Xambó)

(Music Cognition, Music Psychology)

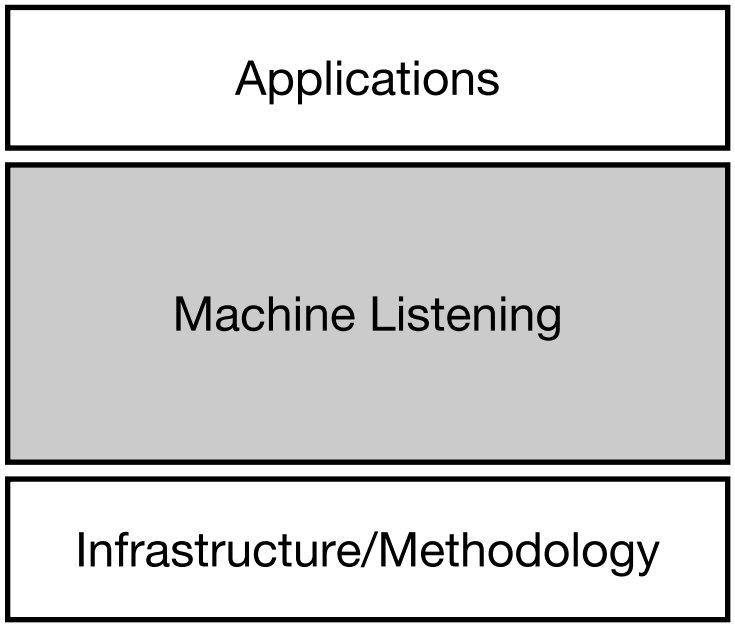
(Computer Music, Audio Engineering)

(User Experience/Education)

(Music Information Retrieval, Music Informatics)

MARL students, post-docs, faculty, and visitors in 2018*
*photo not actually from 2018

Music/Audio Informatics: What we do?



Acoustic scene analysis, computational musicology, interactive music systems

Chord recognition, music structure analysis, source/instrument identification, melody extraction, downbeat extraction, rhythm similarity

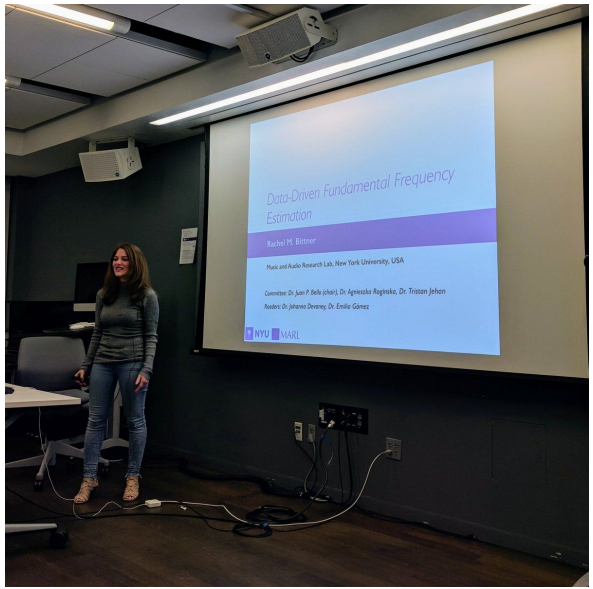
Datasets, evaluation frameworks, data formats, methodological studies

2018 at MARL

#PhDone

- Dr. Rachel Bittner

  Data-Driven Fundamental Frequency Estimation



- Dr. Finn Upham

  Detecting the Adaptation of Listener's Respiration to Heard Music

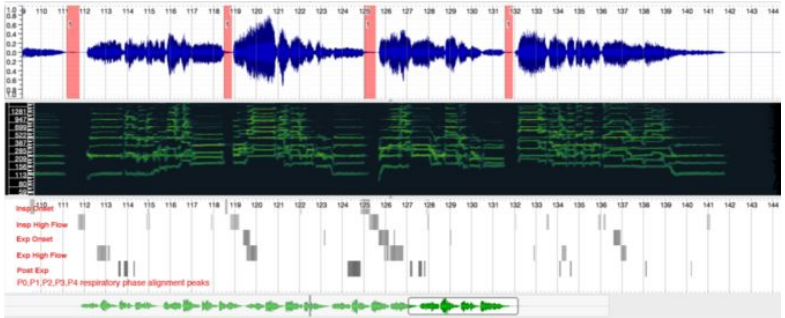


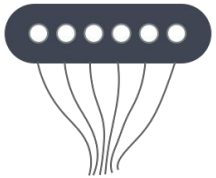
Figure VII.2: Visualizations of the last two verses of Stimulus 106, *Le rosier de trois couleurs de roses* by Strada (Traditional, 2002), along with concurrent alignment peaks in respiratory phase components across all 60 repeated response listenings. On the stereo butterfly sound wave amplitude, red highlighted regions mark audible inspirations in the recording. The spectrogram presents frequencies on a log scale. On the full sound file amplitude outline below each set of panes, the enlarged excerpt is framed.

GuitarSet: A dataset for guitar transcription

Qingyang Xi, Rachel Bittner, Johan Pauwels, Xuzhou Ye, Juan Bello (ISMIR 2018)

https://guitarset.weebly.com/

ACOUSTIC GUITAR RECORDINGS



HEXAPHONIC PICKUP

3 Progressions via Lead Sheets
12 Bar Blues
Autumn Leaves
Pachelbel Canon

⊗

5 Styles
Rock
Singer-Songwriter
Bossa Nova
Jazz
Funk

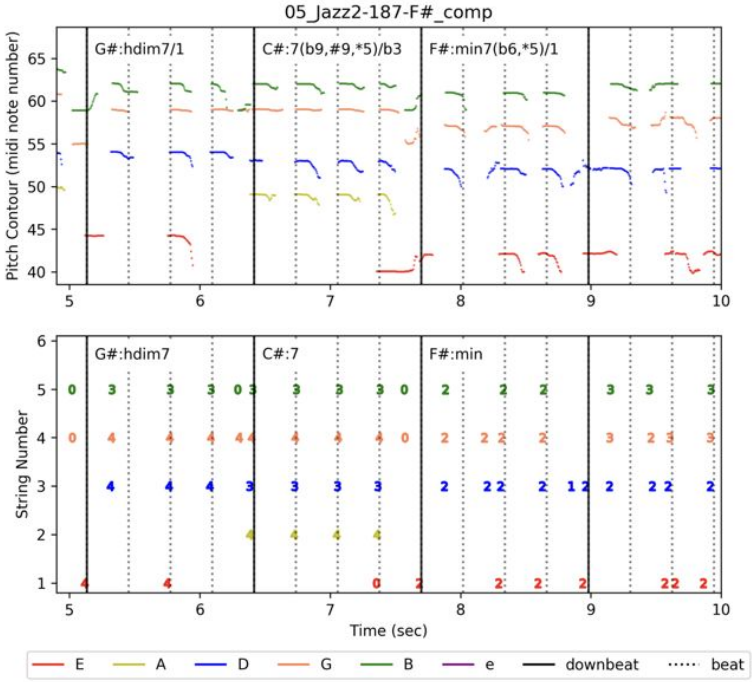
⊗

2 Tempi
Fast
Slow

⊗

2 Version
Comp
Solo

⊗ From Each of the **6 Players** =
360 clips of ~30sec.
Total **~3 Hr. of Audio**

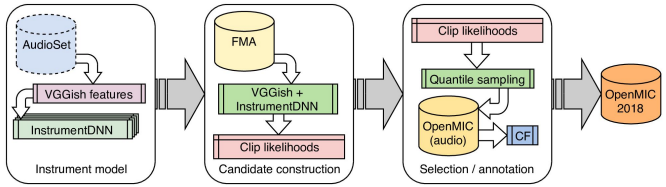


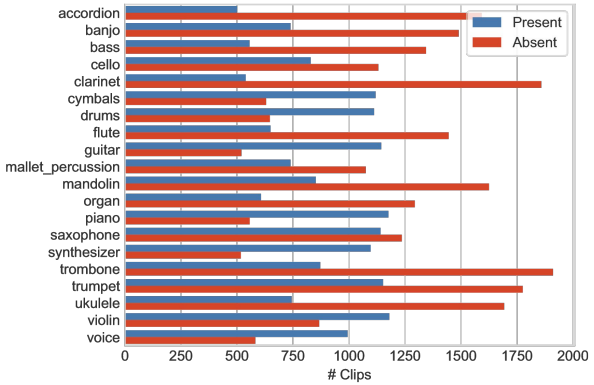
OpenMIC 2018

http://bit.ly/openmic-2018

E. Humphrey, S. Durand, B. McFee (ISMIR 2018)

- **20,000 polyphonic audio clips**
  - CC-licensed (Free Music Archive)
- **20 instrument classes**
- Crowd-sourced annotations
  - **≥ 40,000** consensus labels
- Crowd-source for strong **positives** and **negatives**
  Per class, **≥ 500** positive examples, **≥ 1500** total

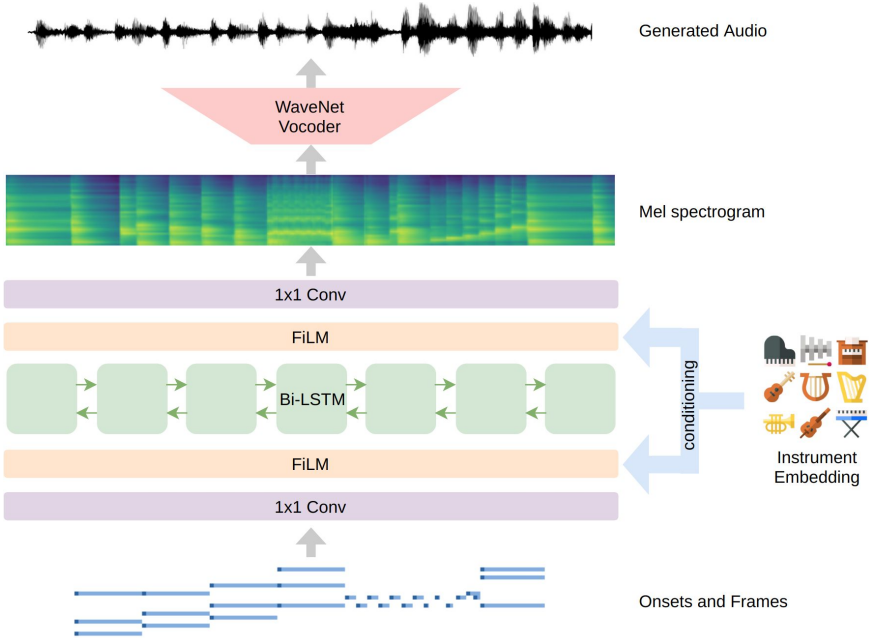


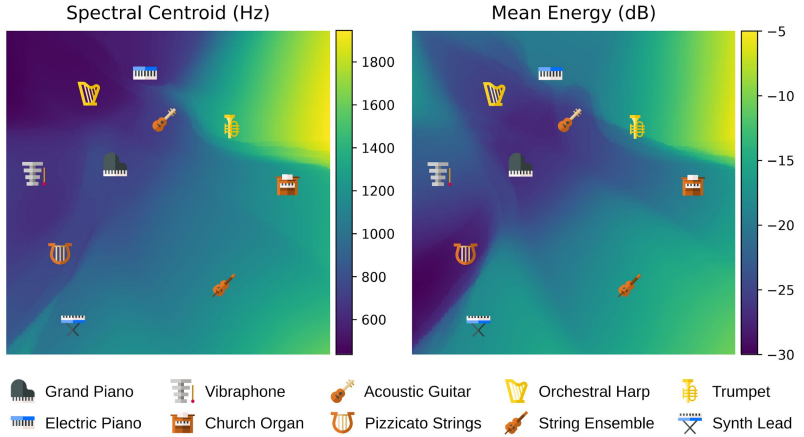


Neural Music synthesis for Flexible Timbre Control

Jong Wook Kim, Rachel Bittner, Aparna Kumar, Juan Pablo Bello (ICASSP 2019)

Combines piano rolls and timbre embedding to synthesize music



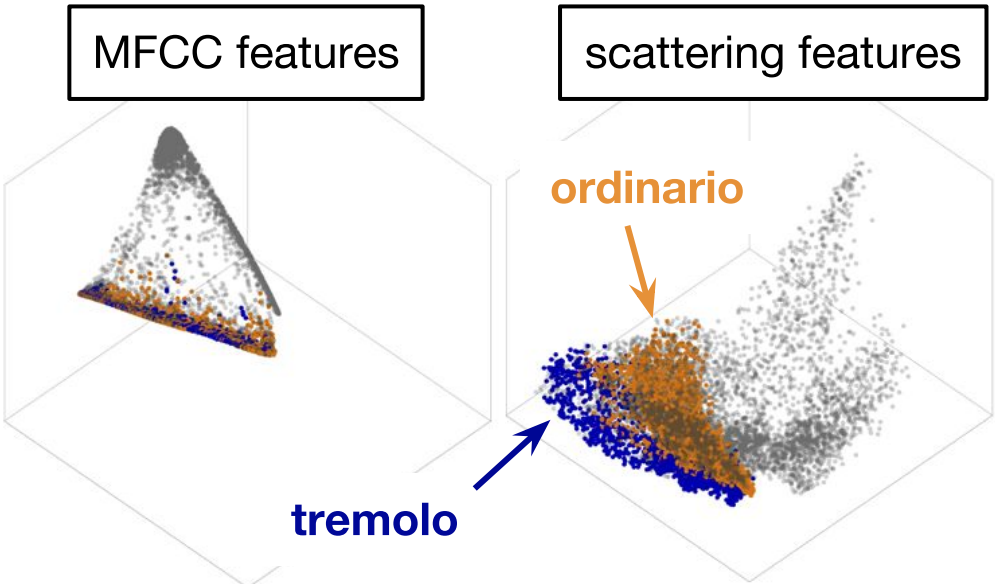


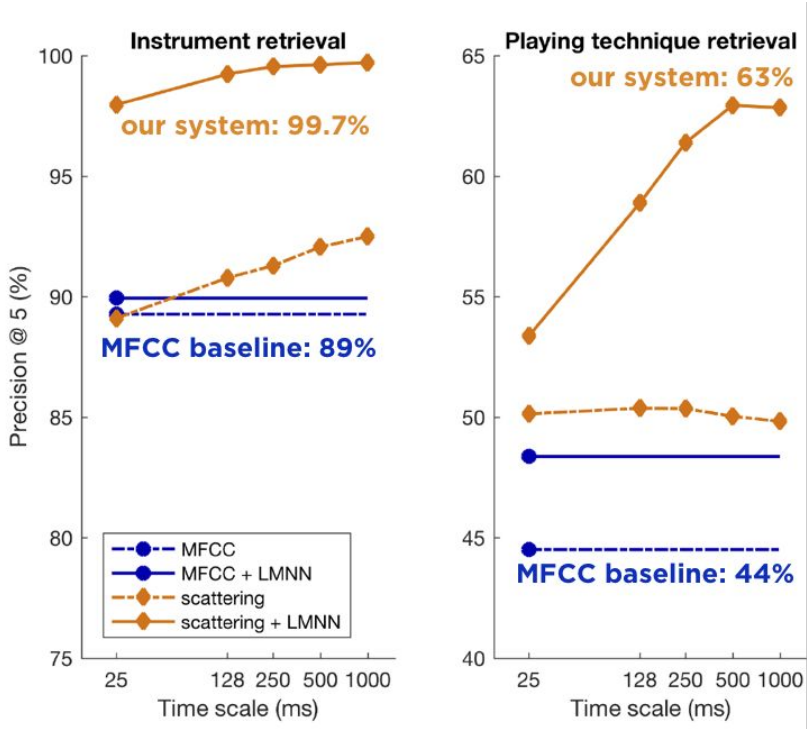
Interactive Demo at:
https://neural-music-synthesis.github.io

Scattering transform as a representation of timbre

MARL researcher: Vincent Lostanlen.
Collaborators: Joakim Andén (Flatiron Institute), Mathieu Lagrange (CNRS).

EURASIP JASMP 2018. DLfM 2018
Isomap embedding of playing techniques



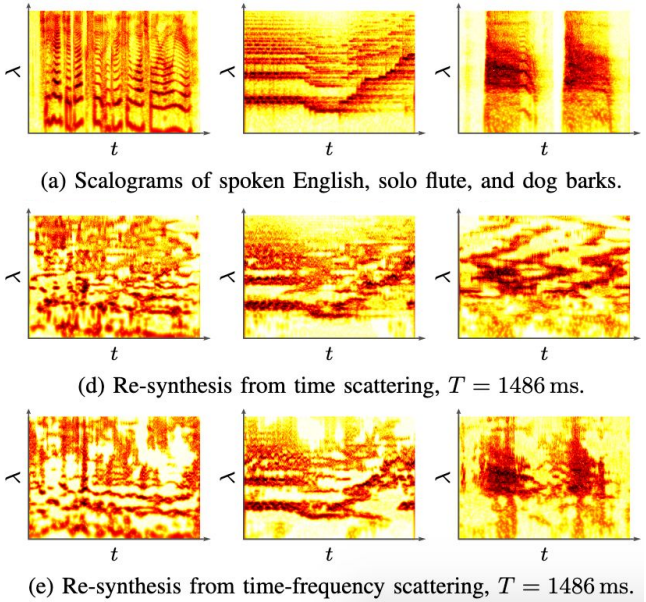


Scattering transform as a representation of texture

MARL researcher: Vincent Lostanlen.

Collaborators: Joakim Andén (Flatiron Institute), Florian Hecker (U. Edinburgh).

IEEE TSP, to appear.
Gradient backpropagation
of time-frequency scattering

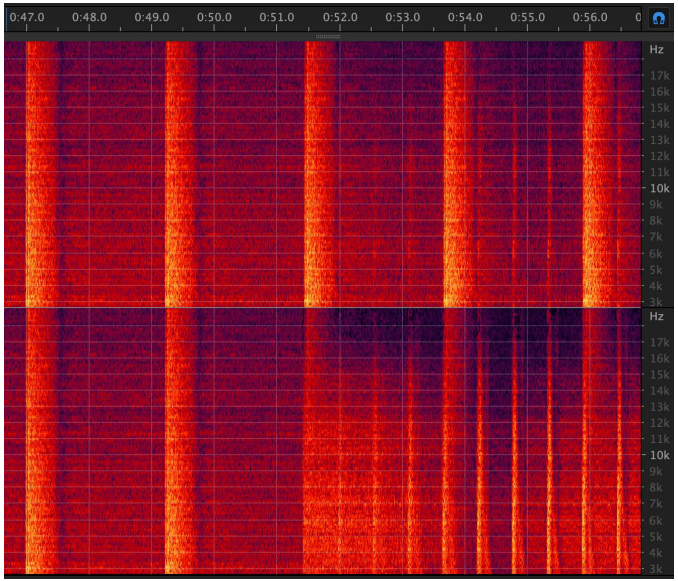


(a) Scalograms of spoken English, solo flute, and dog barks.

(d) Re-synthesis from time scattering, $T = 1486$ ms.

(e) Re-synthesis from time-frequency scattering, $T = 1486$ ms.

WARP 425, July 2018.
"Scattering remix" of Lorenzo Senni's
The Shape of Trance to Come

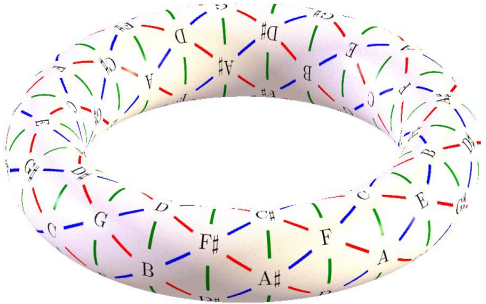


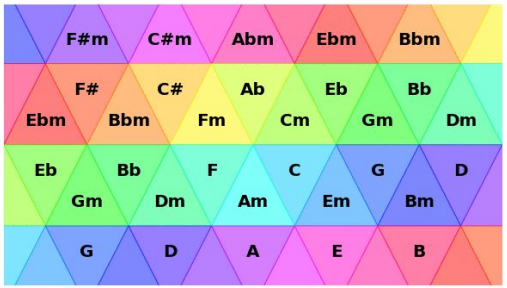
Scattering transform as a representation of harmony

MARL researcher: Vincent Lostanlen, Johanna Devaney.

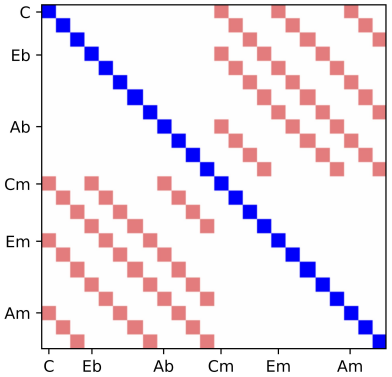
"Eigenprogressions" on the Tonnetz by Graph Laplacian diagonalization. ISMIR-LBD

Torus structure of 24 triads

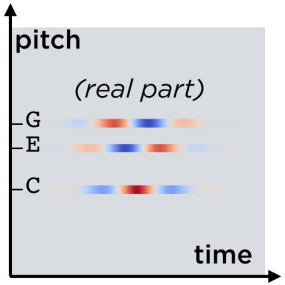


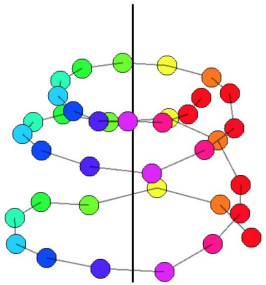


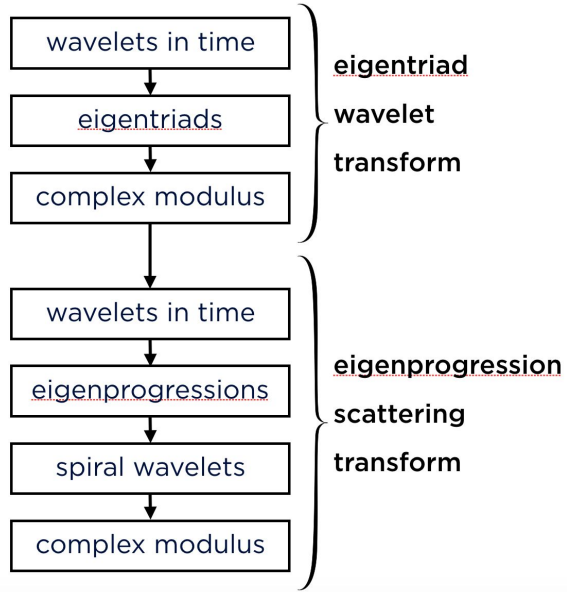
Laplacian matrix



Application: composer recognition



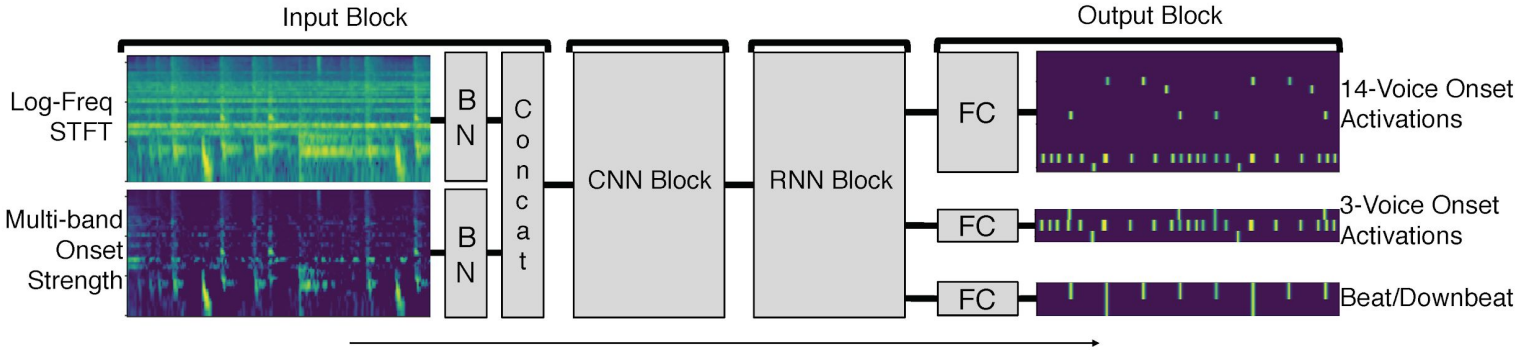


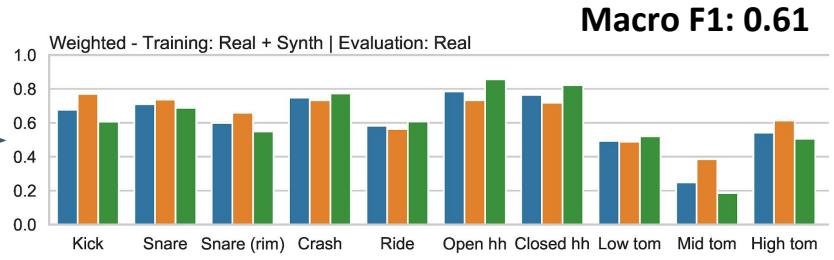


Drum Transcription

Mark Cartwright, Juan P. Bello (DAFx 2018)

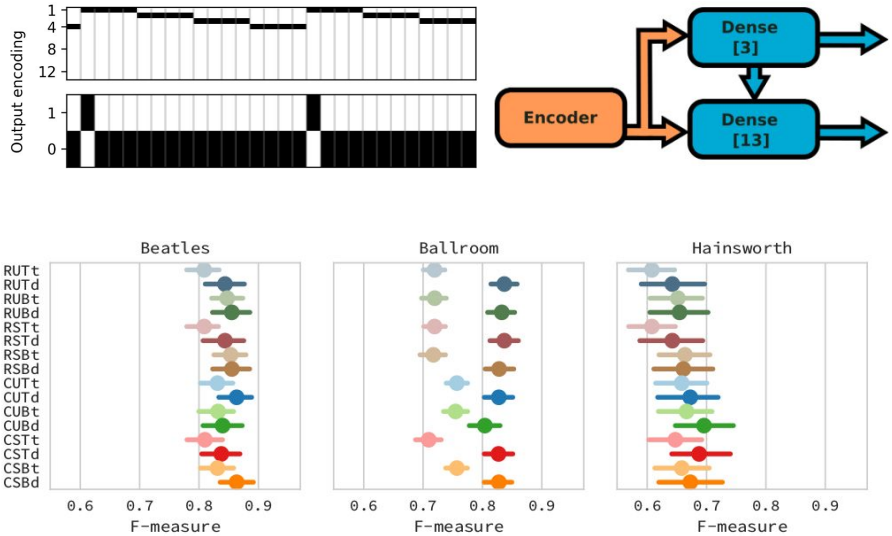
Synthetic + real data = transcription of more drum voices!



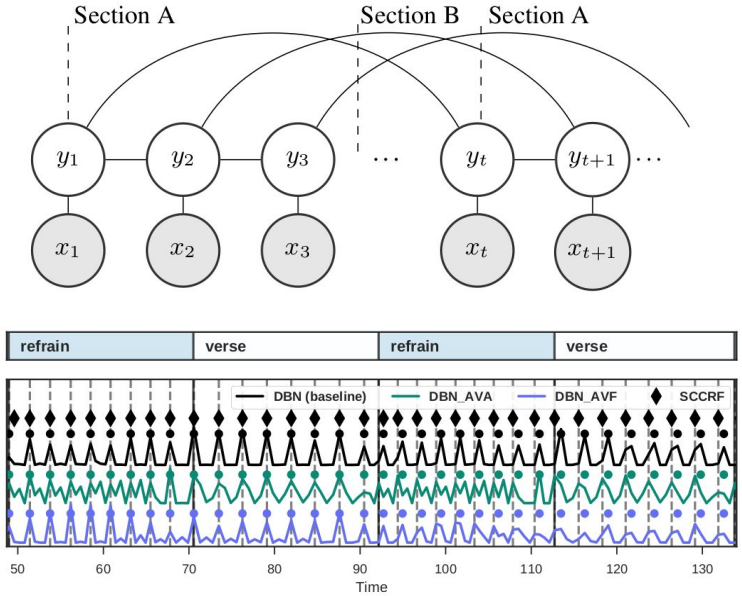


Structured Downbeat Prediction

MARL researchers: Brian McFee, Juan Bello
Collaborators: Magdalena Fuentes, Slim Essid, Hélène C. Crayencour

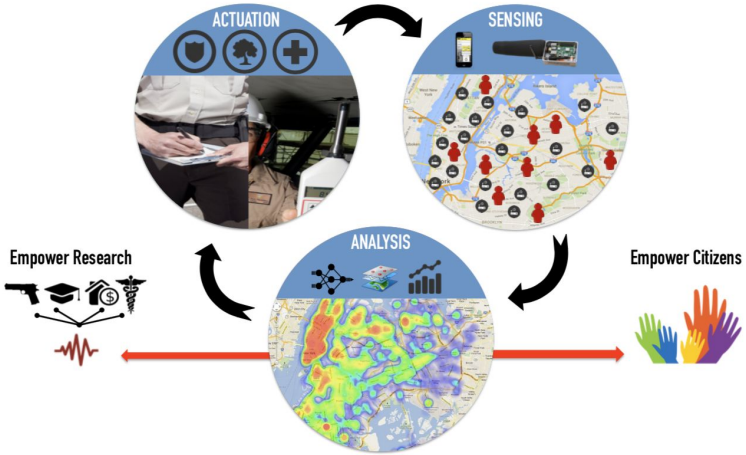


Analysis of Common Design Choices in Deep Learning Systems for Downbeat Tracking (ISMIR 2018)



A Music Structure Informed Downbeat Tracking System Using Skip-Chain Conditional Random Fields and Deep Learning (ICASSP 2019)

Beyond Music: Sounds of New York City (SONYC)



SONYC: A System for Monitoring, Analyzing, and Mitigating Urban Noise Pollution
J. P. Bello, C. Silva, O. Nov, R. L. Dubois, A. Arora, J. Salamon, C. Mydlarz, H. Doraiswamy
Communications of the ACM, February 2019, Vol. 62 No. 2, Pages 68-77

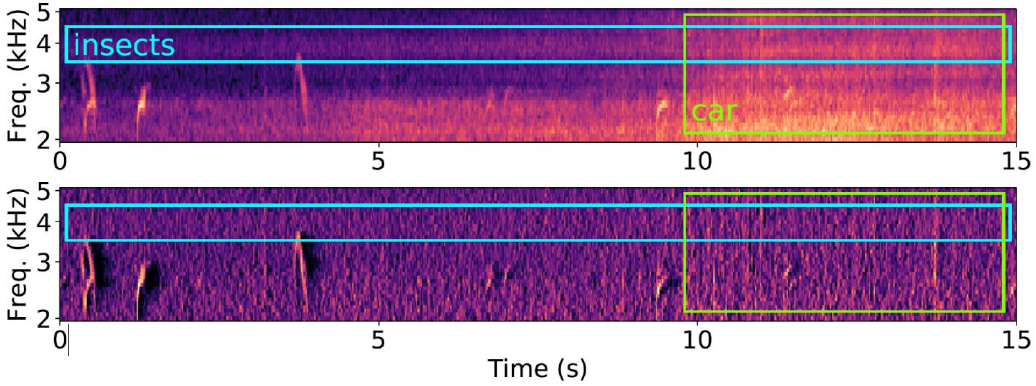
Sound analysis in smart cities
J. P. Bello, C. Mydlarz, and J. Salamon.
In T. Virtanen, M. D. Plumbley, and D. P. W. Ellis, editors, Computational Analysis of Sound Scenes and Events, pages 373–397. Springer International Publishing, 2018.

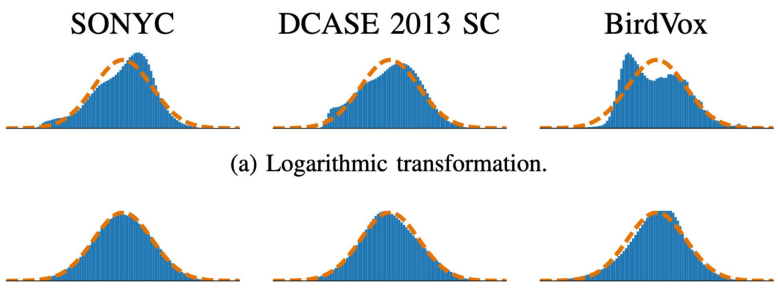
Acoustic sensor networks

MARL researchers: Vincent Lostanlen, Justin Salamon, Mark Cartwright, Brian McFee.

Per-Channel Energy Normalization: Why and How. IEEE SPL, January 2019.

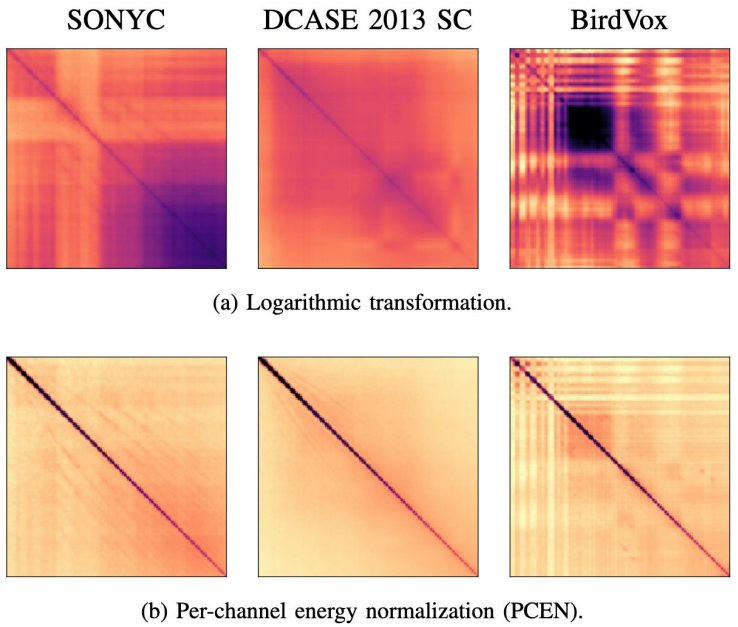
PCEN Gaussianizes magnitudes, decorrelates subbands, and enhances foreground contrast.





(a) Logarithmic transformation.

(b) Per-channel energy normalization (PCEN).



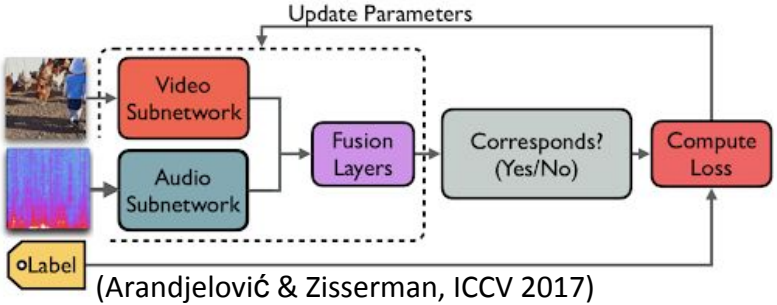
(a) Logarithmic transformation.

(b) Per-channel energy normalization (PCEN).

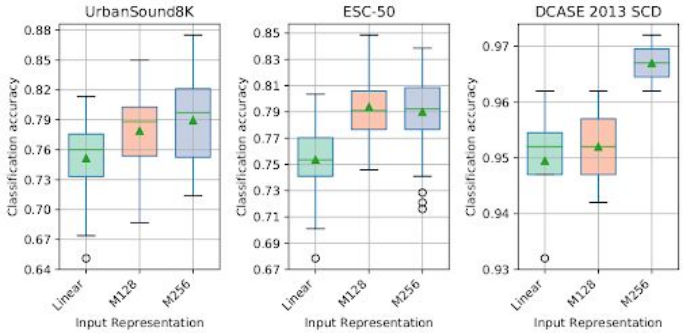
Deep Audio Embeddings

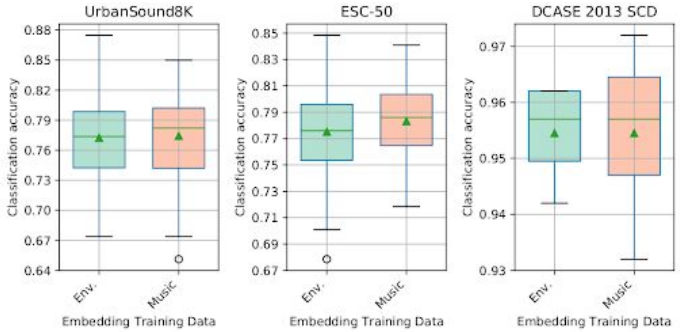
$ pip install openl3

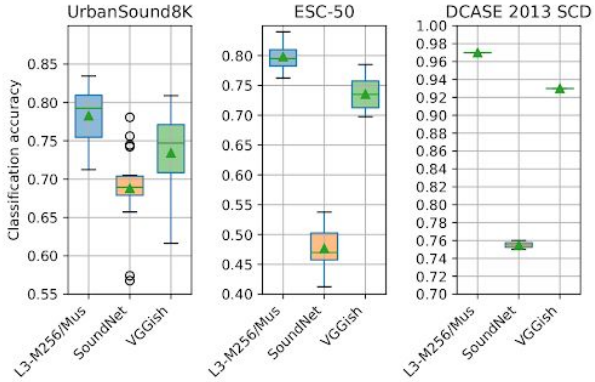
MARL researchers: Jason Cramer, Ho-Hsiang Wu, Justin Salamon, Juan Pablo Bello (ICASSP 2019)



(Arandjelović & Zisserman, ICCV 2017)

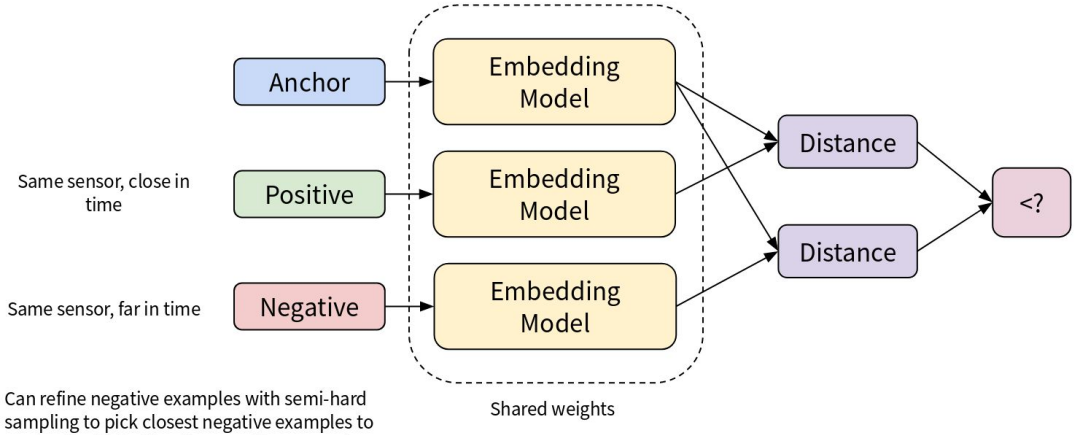




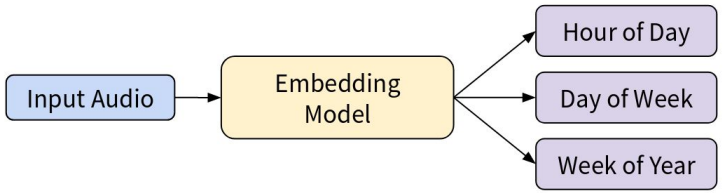


Deep Audio Embeddings

MARL researchers: Mark Cartwright, Justin Salamon, Jason Cramer



(Jansen, Plakal, et al., ICASSP 2018)



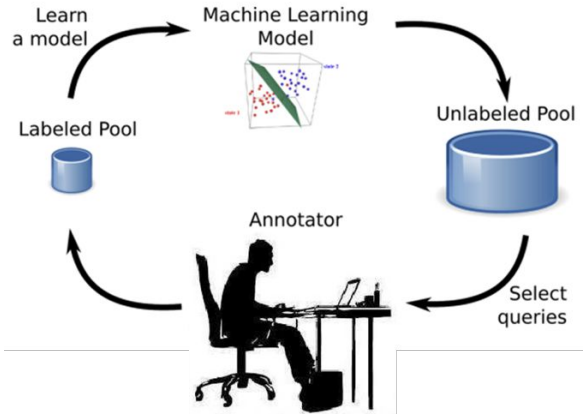
Active Learning: efficient annotation and classification with unlabeled data

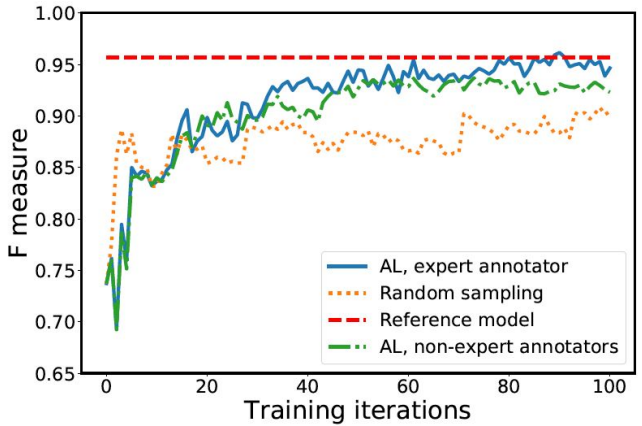
Yu Wang, Ana Elisa Mendez Mendez, Mark Cartwright, Juan Pablo Bello (ICASSP 2019)

Goal: build a classifier to detect distorted artefact in SONYC data

Large amount of unlabeled data + Rare class =

Active learning with desired sampling strategy:

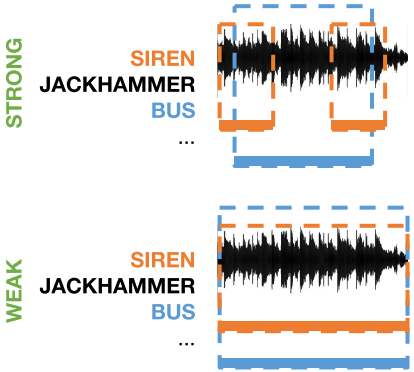


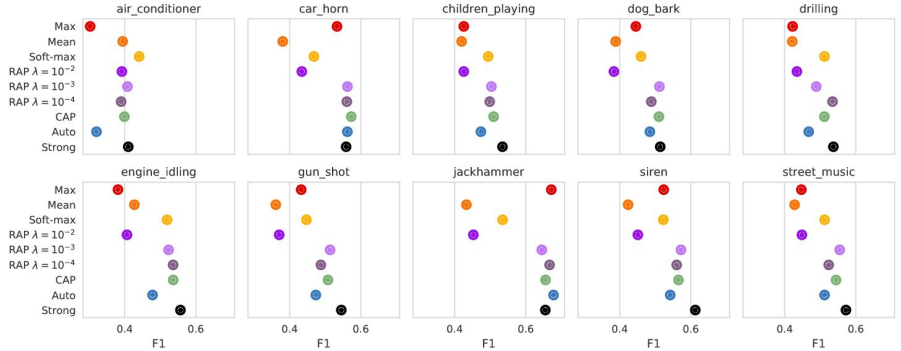


MIL and Learning from weak labels

$ pip install autopool

B. McFee, J. Salamon, J.P. Bello (IEEE-TASLP, 2018)





Can we learn a strong predictor from weak labels?

New prediction **pooling operator** that automatically adapts to label characteristics

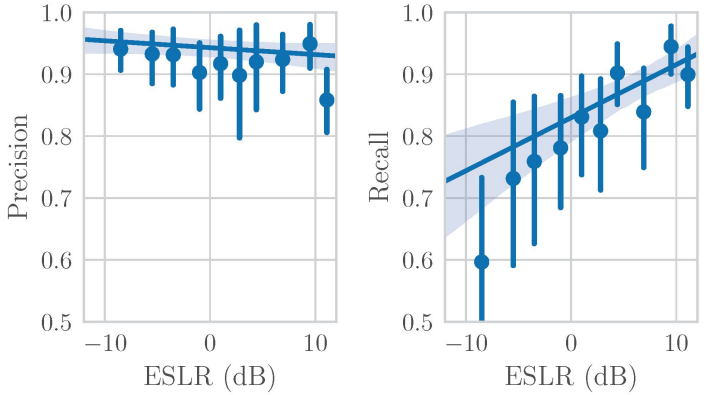
Autopool + **weak training** nearly matches strong training on both weak and strong evaluation

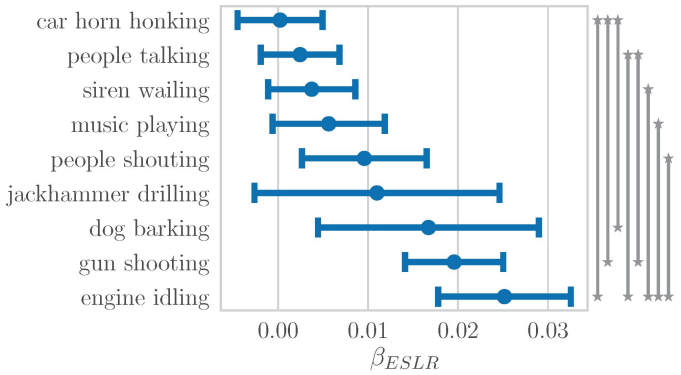
		WEAK TESTING			STRONG TESTING		
	Model	F_1	P	R	F_1	P	R
WEAK TRAINING	Max	0.742	0.774	0.717	0.463	0.774	0.330
	Mean	0.543	0.726	0.436	0.408	0.280	0.751
	Soft-max	0.630	0.772	0.537	0.492	0.397	0.646
	RAP 10^{-2}	0.544	0.719	0.449	0.419	0.296	0.717
	RAP 10^{-3}	0.746	0.790	0.711	0.529	0.584	0.484
	RAP 10^{-4}	0.754	0.754	0.756	0.526	0.650	0.442
	CAP	0.754	0.781	0.732	0.533	0.622	0.466
	Auto	0.757	0.784	0.739	0.504	0.738	0.382
	Strong	0.762	0.708	0.822	0.551	0.693	0.458

Crowdsourcing Audio Annotations

M. Cartwright, J. Salamon, A. Seals, O. Nov, J.P. Bello (ICASSP 2018)

Investigating the Effect of Sound-Event Loudness on Annotation Quality



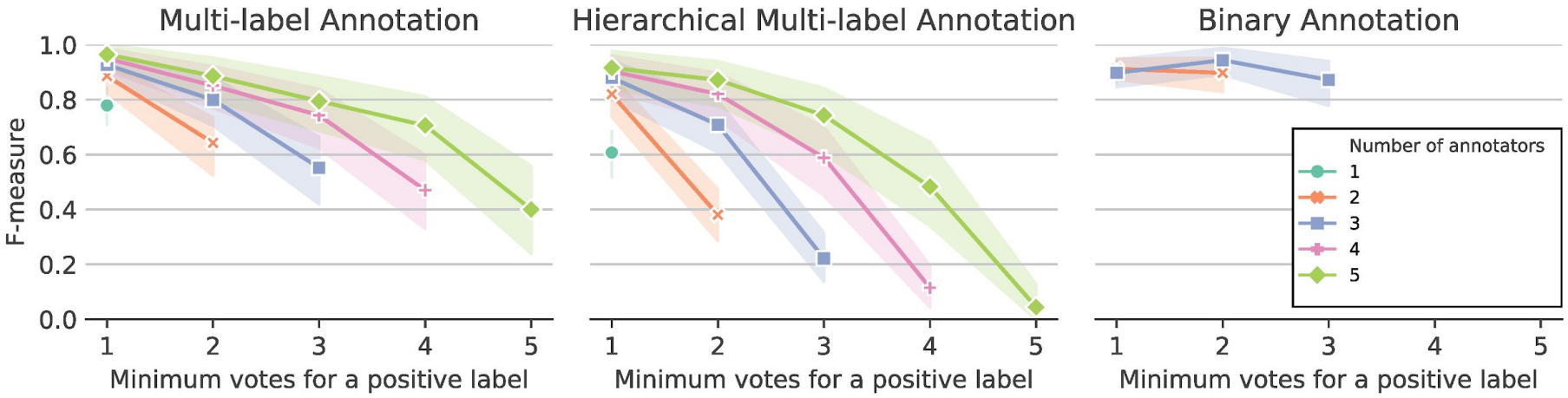


- Relative sound-event loudness affects recall but has only minor effect on precision
- The effect of relative loudness to annotation performance is class-dependent

Citizen Science Audio Annotations

M. Cartwright, G. Dove, A.E.M. Mendez, J.P. Bello, O. Nov (CHI 2019)

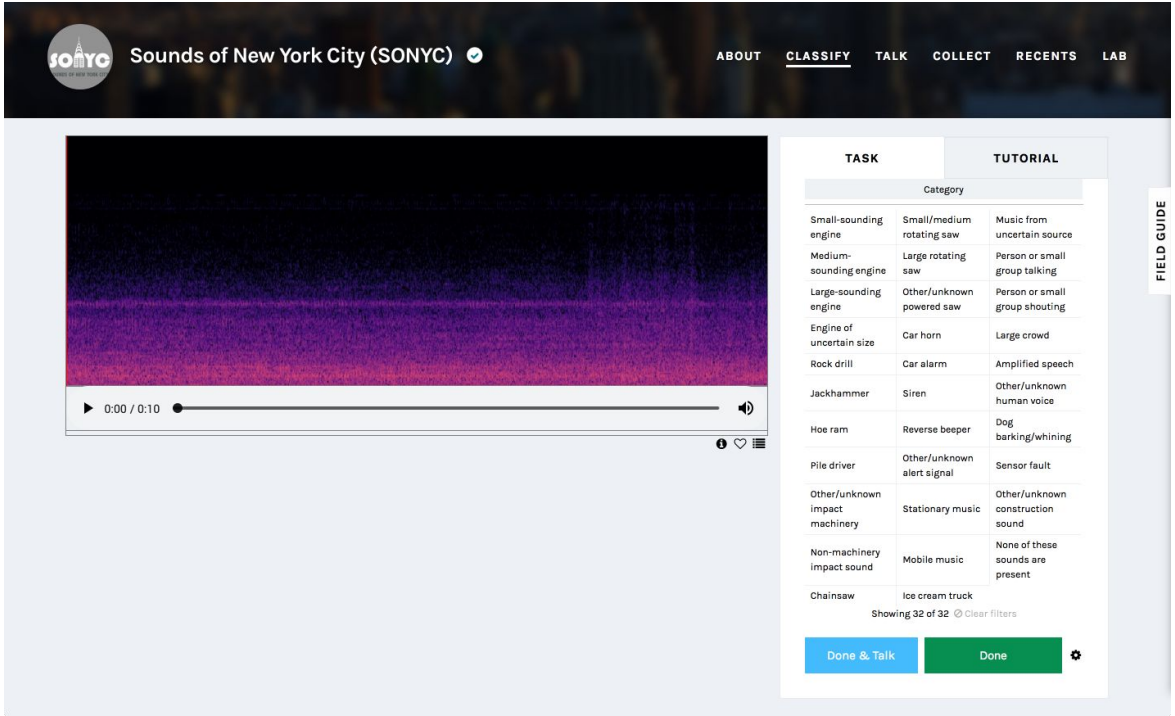
- Volunteers “over annotate” (low precision) in binary labeling and “under annotate” (low recall) in multi-labeling
- Can balance during aggregation
- Throughput higher with multi-labeling



SONYC Annotation Campaign on Zooniverse

MARL researchers: Mark Cartwright, Ana Elisa Mendez Mendez, Juan Bello
Collaborators: Graham Dove, Oded Nov

Tag SONYC recordings on Zooniverse.org to help fight urban noise pollution!



Open source

https://github.com/marl

- **openl3** - deep audio embeddings
  https://openl3.readthedocs.io
  $ pip install openl3
- **crepe** - convolutional representation for pitch estimation
  https://marl.github.io/crepe/
  $ pip install crepe
- **Kymatio** - scattering transforms in pytorch
  https://kymatio.readthedocs.io
  $ pip install kymatio
- **AutoPool** - adaptive pooling operators in Keras
  https://autopool.readthedocs.io
  $ pip install autopool
- **librosa** 0.6 series released, 0.7 in 2019
  https://librosa.github.io
  $ pip install librosa

Open data

- **GuitarSet** 3 hours of string separated acoustic guitar recordings with detailed note-level annotations.
  https://guitarset.weebly.com/

- **OpenMIC-2018** - 20K clips
  http://bit.ly/openmic-2018

- **Medley-solos-DB** cross-collection dataset for musical instrument recognition
  http://zenodo.org/record/1344103

- **BirdVox-scaper-1M** synthetic audio for terabyte-scale avian bioacoustics
  http://wp.nyu.edu/birdvox + see poster by Elizabeth Mendoza today

DCASE 2019 Workshop

General chairs: Juan Pablo Bello, Mark Cartwright
Technical chairs: Dan P. W. Ellis, Justin Salamon, Michael Mandel
Local chair: Vincent Lostanlen

October 25-26, 2019

NYU Tandon School of Engineering (Brooklyn, NY)

Topics

- Tasks in computational environmental audio analysis
- Methods for computational environmental audio analysis
- Resources, applications, and evaluation of computational environmental audio analysis

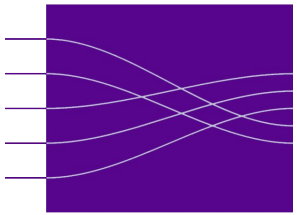
Challenge

DCASE 2019 Workshop

General chairs: Juan Pablo Bello, Mark Cartwright
Technical chairs: Dan P. W. Ellis, Justin Salamon, Michael Mandel
Local chair: Vincent Lostanlen

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**



MARL

Many thanks, and apply to join us!

For more info visit: http://steinhardt.nyu.edu/marl/