MONYC: Music of New York City Dataset

Magdalena Fuentes[1], Danielle Zhao[1], Vincent Lostanlen[2], Mark Cartwright[3], Charlie Mydlarz[1], Juan Pablo Bello[1]

[1]New York University, USA, [2]CNRS, Laboratoire des Sciences du Numérique de Nantes (LS2N), France, [3]Department of Informatics, New Jersey Institute of Technology

🏠 https://magdalenafuentes.github.io/
✉ mf3734@nyu.edu

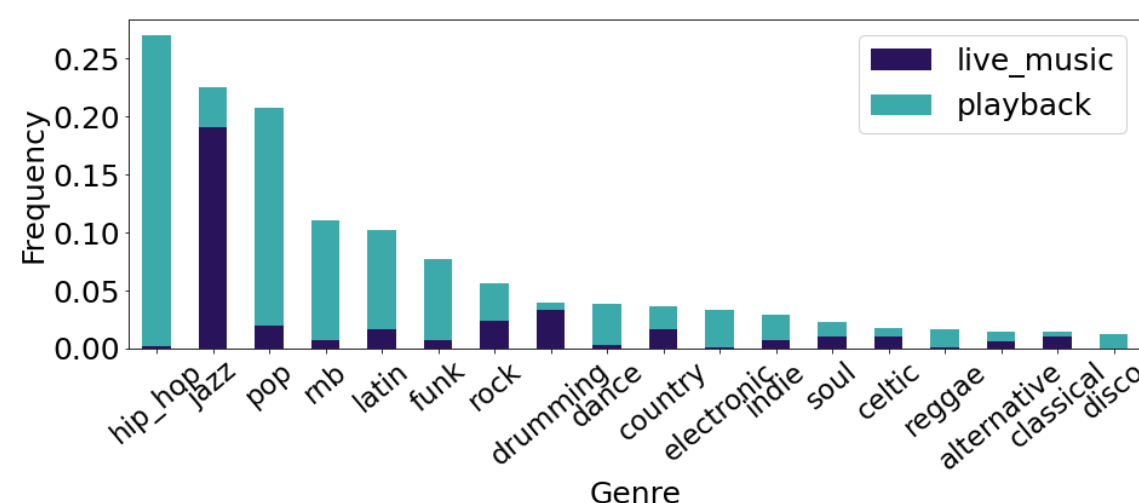DCASE 2021 WORKSHOP

1. Motivation

- Music is important in human cultures and is an integral part of urban soundscapes

- To make sense of these soundscapes, machine listening models should be able to detect and classify street music

- The **lack of well-curated resources** for training and evaluating models currently hinders their development. To mitigate this, we introduce MONYC, the first-of-its-kind open dataset of music in urban settings

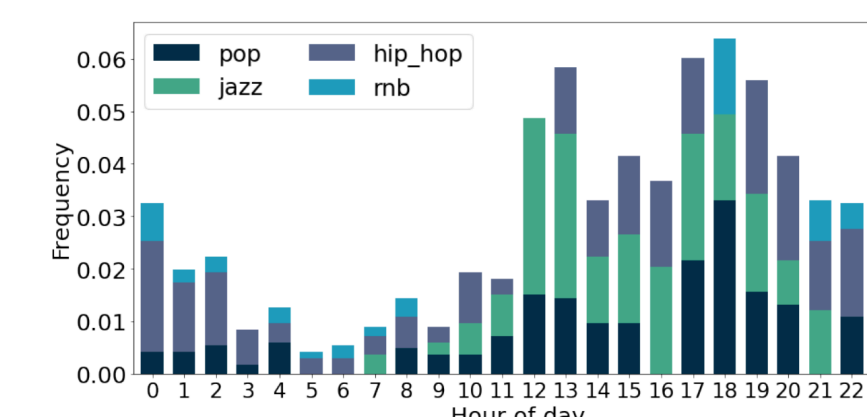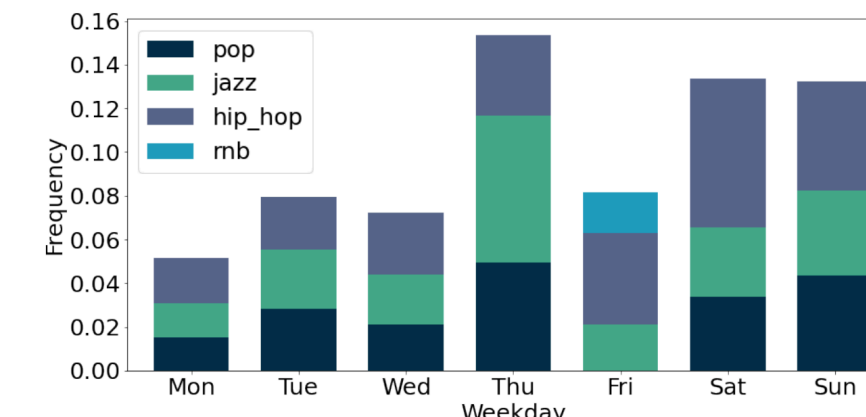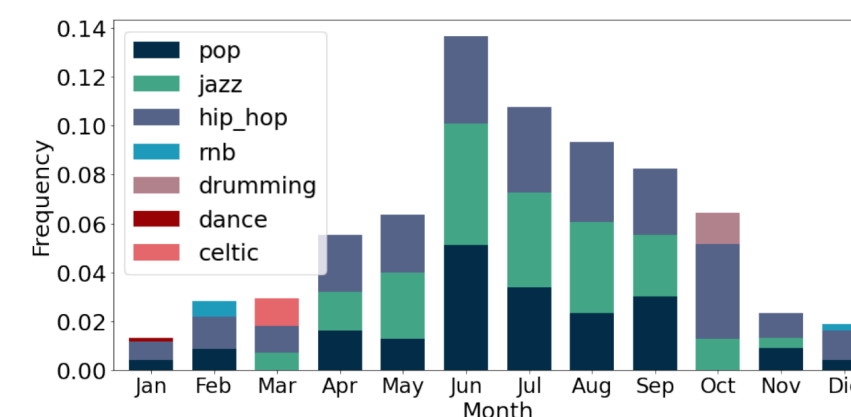## 2. Data curation and annotations

stage	SONYC	2017	subsampled	15 sensors	music	uniform	DPP	annotation	agreement
# clips	250M	30M	10M	5.8M	94k	30k	3k	1.7k	1.5k

- We use data-driven and self-supervised methods for sampling and curating a diverse set of music clips in different stages, using recordings from the SONYC sensor network [1], from 250M clips to the final 1.5k

- Annotations include **genre tags**, **live vs. playback music**, **multi- vs. single-instrument**, **loud vs. quiet** indicators
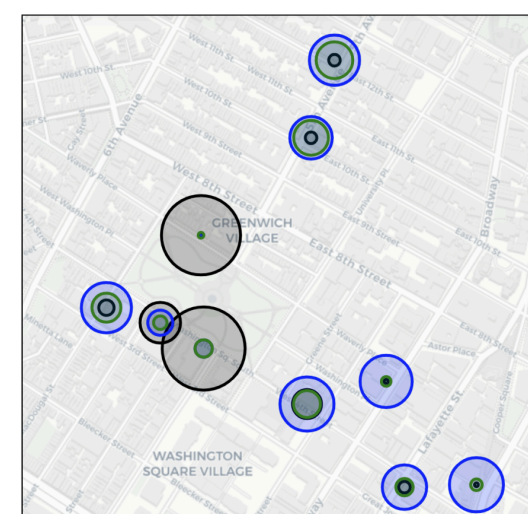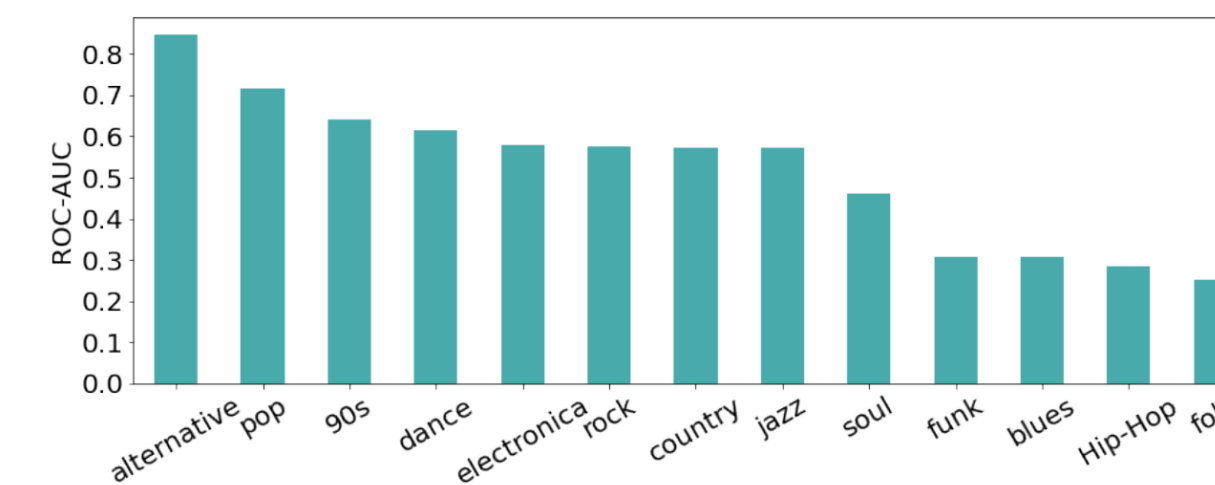
3. Dataset overview

- The top genre in MONYC is hip hop

- Most genres are play-backed, except for jazz and drumming

- We identified more music clips towards the Summer months (June, July), and considerably less in Winter (November, December and January)

- Genre presence usually correlates with events, e.g. Celtic music in March due to St. Patrick's Day

- Less street music at the beginning of the week, more towards the weekend

- With MONYC we can look at the spatial distribution of genres such as hip hop (blue), jazz (black) and pop (green)

- Jazz is mostly live, so is concentrated around Washington Square Park more than the other genres

- We did an experiment with the off-the-shelf music tagger musicnn [2], and observed poor performance compared to standard MIR genre datasets (where mean performance is on the order of 50%). This is to expect since the model is not trained in music with low Signal-to-Noise Ratio and external source interference as MONYC clips are.

- The model performed 8-12% worse in average in those recordings labeled with high interference of sources

4. Conclusions and Future work

- We present MONYC, the first-of-its-kind open dataset of music in urban settings

- MONYC opens the possibility to develop and evaluate machine listening models for the classification of street music

- Such models for the classification of street music offer the opportunity to dig into behavioral patterns related to human activities in urban, such as nightlife, festivals, street celebrations, among others.

- https://magdalenafuentes.github.io/monyc/
- https://github.com/soundata/soundata

[1] https://wp.nyu.edu/sonyc/

[2] J. Pons and X. Serra, "musicnn: Pre-trained Convolutional Neural Networks for Music Audio Tagging," arXiv preprint arXiv:1909.06654, 2019.