BirdVox-Imitation: A dataset of birdsong imitations with potential as a testbed for machine listening

Kendra Oudyk*[1,2,3], Vincent Lostanlen[3], Justin Salamon[4,5], Andrew Farnsworth[3], and Juan Pablo Bello[4,5]

*[1]Integrated Program in Neuroscience, McGill University, Canada; [2]Department of Music, Art, and Culture Studies, University of Jyväskylä, Finland; [3]Cornell Lab of Ornithology, Cornell University, USA; [4]Centre for Urban Science and Progress, New York University, USA; [5]Music and Audio Research Lab, New York University, USA; *kendra.oudyk@gmail.com*

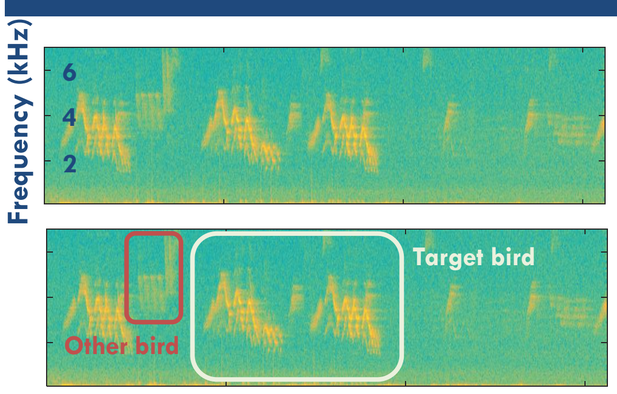
INTRO

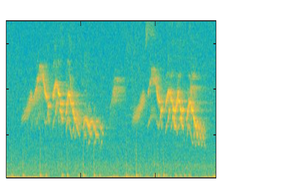
- Bird watchers imitate bird sounds to find birds
- **Bird imitation is challenging for humans;** bird vocalizations can have
  - higher/wider pitch **range**
  - faster **timing**
  - complex, noisy, multi-toned **timbres**
- There are different strategies for learning/imitating bird sounds, e.g.:
  - mnemonic phrases – e.g., "who cooks for you" (owl)
  - visualizing spectrograms
  - vocalizations: whistling, humming, singing
  - use of hands as a resonating chamber

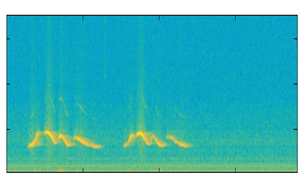
PURPOSE

- To build a dataset of human imitations of birdsong
- This is part of a larger project to build a retrieval-by-imitation system for birdsong
- Open science efforts:
  - dataset will be released on Zenodo
  - implementation of YIN-bird[1] on GitHub
  - birdsong was scraped from Xeno-Canto.org, a citizen science platform for sharing bird sounds

DATA COLLECTION







	Data Category	Data collection procedure	N	Mean (SD) length (sec)	Total (HHMMSS)
	Original Recordings	Field recordings were scraped using an API from Xeno-Canto - 10 species - Quality: (highest) A B C D E (lowest) - Type of vocalization: Song	1216	76.2 (97.0)	25:45:14
	Excerpts	Excerpts were manually annotated from the original recordings - 'Clean' - Appropriate length for imitation (~2-10 sec)	6659	3.2 (1.3)	05:55:48
	Stimuli	Quasi-randomly selected from excerpts - 10 per species, some repeats - Same stimuli for all participants	100	3.9 (1.4)	00:06:32
	Imitations	17 participants, incl. - 11 with birdwatching experience - 11 musicians	1700	5.9 (1.4)	00:09:52

Procedure: 1) Hear birdsong 2) Hear clap 3) Perform imitation

SAMPLE DATA

Birdsong excerpts:

Human imitations:

POTENTIAL APPLICATIONS:

Testbed for machine listening

- This is possibly the first audio dataset that is both
  - **Multimodal** (bird vs. human)
  - **Domain-adversarial** (imitation strategy, e.g., humming, whistling, singing)
- This gives plenty of room for inventing new methods in machine listening

e.g., Birdsong retrieval-by-imitation

- Based on pitch-curve features[3,4]
- Pitch-curve estimation with YIN-bird[1]
- Project in progress

YIN-bird[1] algorithm (figure adapted from [1])

FUNDING

- University of Jyväskylä Travel Grant
- National Science Foundation: Big Data program grant

REFERENCES

- [1] O'Reilly, C., & Harte, N. (2017). Pitch tracking of bird vocalizations and an automated process using YIN-bird. *Cogent Biology*, 3(1), 1322025
- [2] Bird photos from https://www.allaboutbirds.org/guide/
- [3] Salamon, J., Rocha, B. M. M., & Gómez, E. (2012, March). Musical genre classification using melody features extracted from polyphonic music signals. In *ICASSP* (pp. 81-84).
- [4] Panteli, M., Bittner, R., Bello, J. P., & Dixon, S. (2017, March). Towards the characterization of singing styles in world music. In *ICASSP* (pp. 636-640).