

Matching human vocal imitations to birdsong: An exploratory analysis

Kendra Oudyk, Yun-Han Wu, Vincent Lostanlen, Justin Salamon, Andrew Farnsworth and Juan Bello

2019-08-29

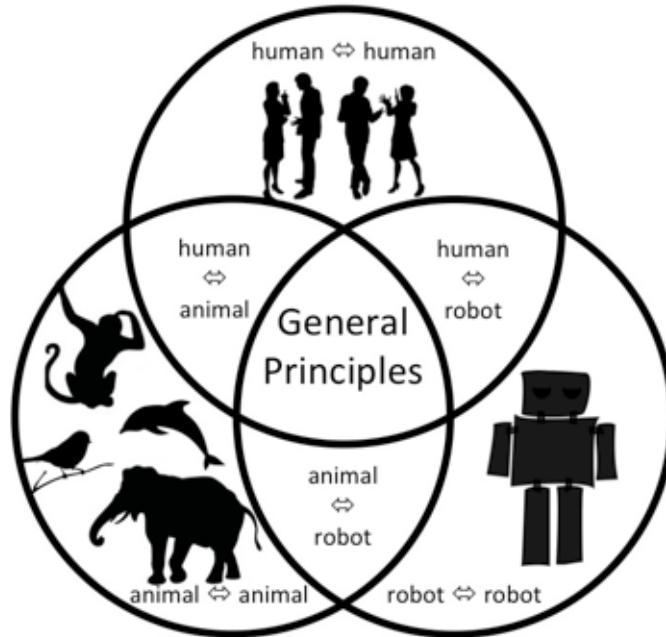
VIHAR 2019, London, UK

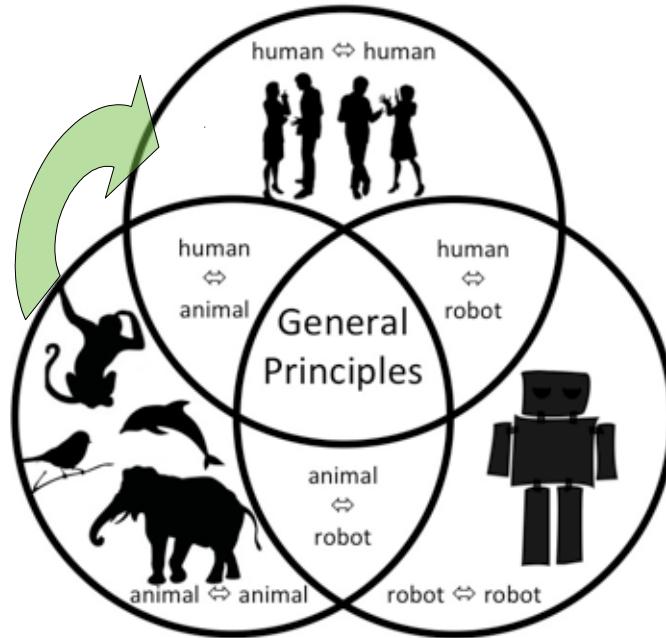
kendra.oudyk@mail.mcgill.ca

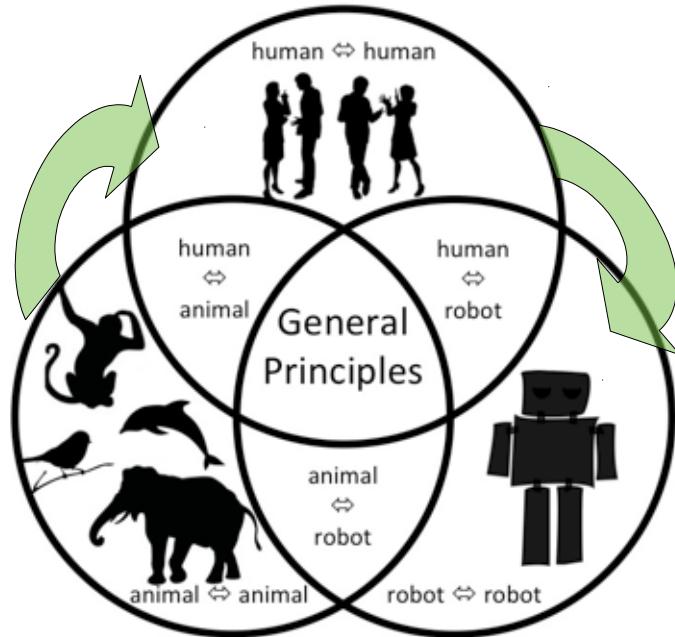


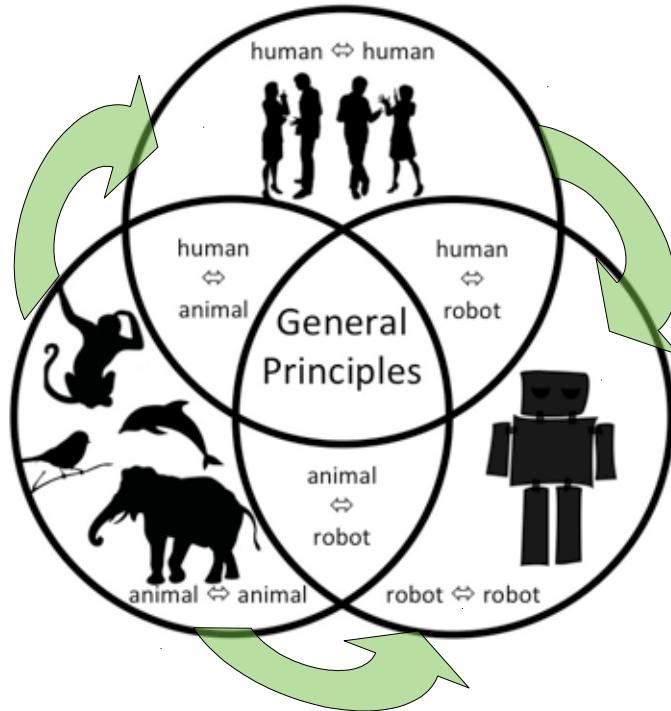
JYVÄSKYLÄN YLIOPISTO
UNIVERSITY OF JYVÄSKYLÄ

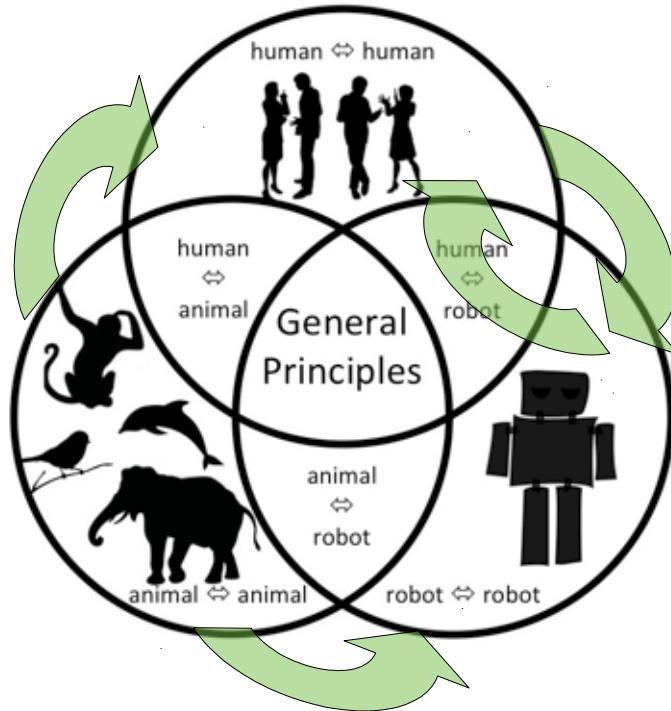




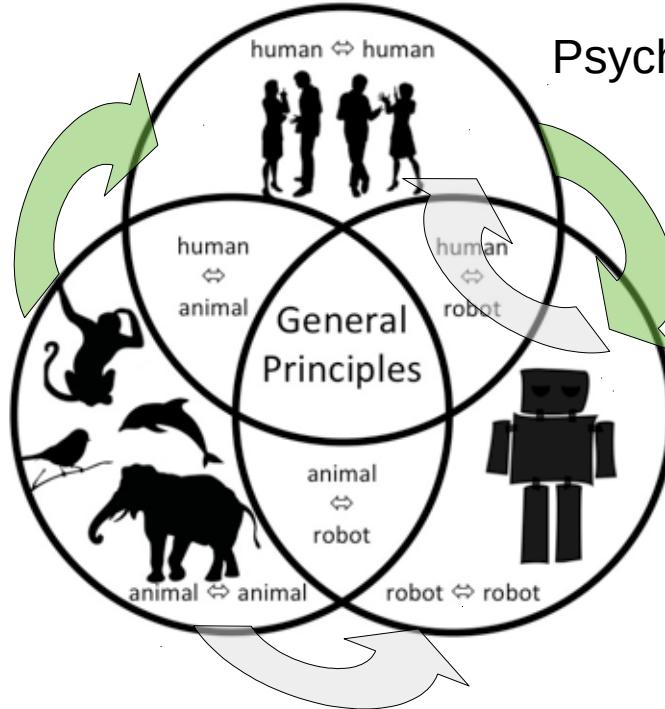


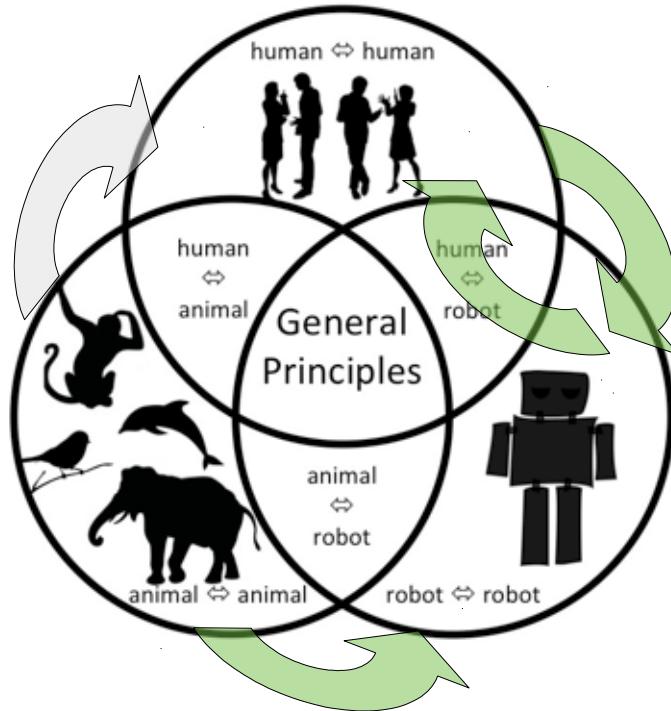




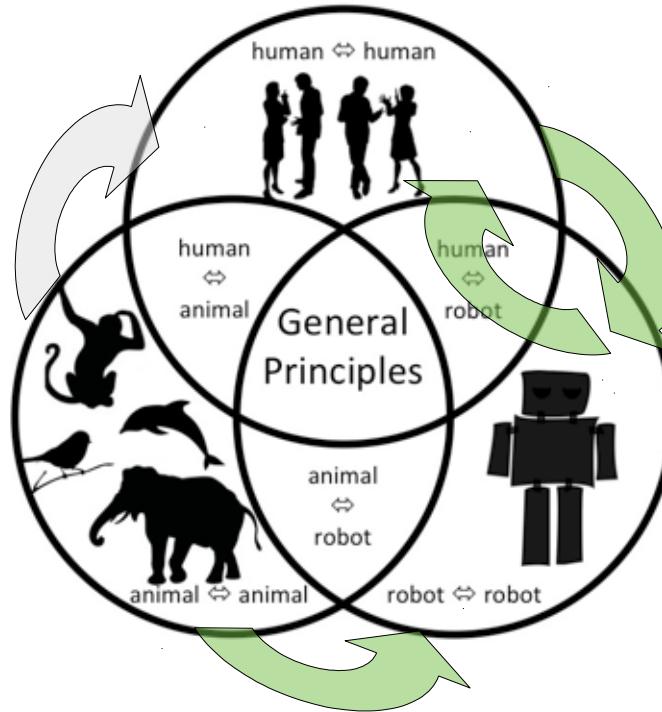


Psychology





Machine listening



Efficient

Methods Summary

- Data collection
 - Birdsong
 - Imitations
- Exploratory analyses
 - Spectral
 - Temporal
 - Spectrotemporal

Data Collection

xeno-canto



Search or browse

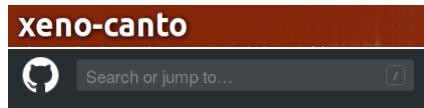
koudyk / birdsong-downloadXC

SONIC VISUALISER



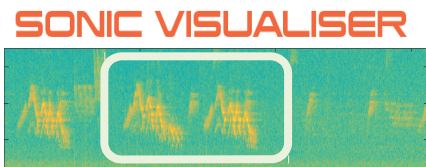
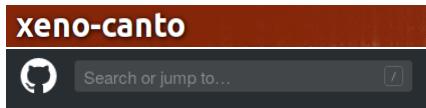
Data Category	N	Mean (StDev) length (sec)	Total (HHMMSS)
Original Recordings	1216	76.2 (97.0)	25:45:14
Excerpts	6659	3.2 (1.3)	05:55:48
Stimuli	100	3.9 (1.4)	00:06:32
Imitations	1700	5.9 (1.4)	00:09:52

Data Collection



Data Category	N	Mean (StDev) length (sec)	Total (HHMMSS)
Original Recordings	1216	76.2 (97.0)	25:45:14
Excerpts	6659	3.2 (1.3)	05:55:48
Stimuli	100	3.9 (1.4)	00:06:32
Imitations	1700	5.9 (1.4)	00:09:52

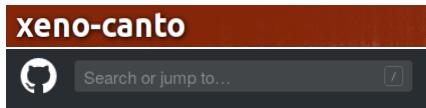
Data Collection



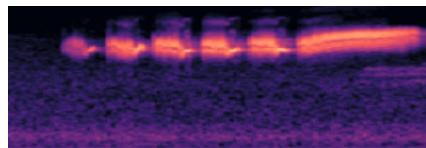
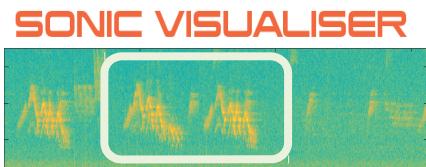
Data Category	N	Mean (StDev) length (sec)	Total (HHMMSS)
Original Recordings	1216	76.2 (97.0)	25:45:14
Excerpts	6659	3.2 (1.3)	05:55:48
Stimuli	100	3.9 (1.4)	00:06:32
Imitations	1700	5.9 (1.4)	00:09:52

Cannam, C., Landone, C., and Sandler, M. (2010). Sonic visualiser: An open source application for viewing, analysing, and annotating music audio files. In Proceedings of the ACM Multimedia 2010 International Conference, pages 1467–1468, Firenze, Italy.

Data Collection

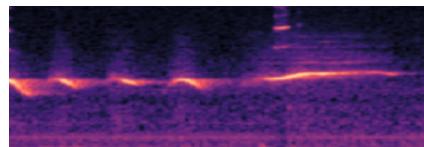
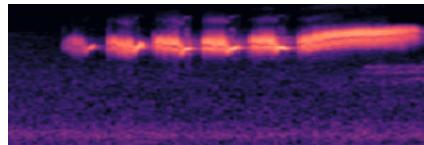
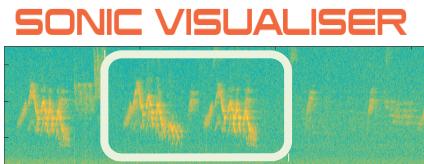
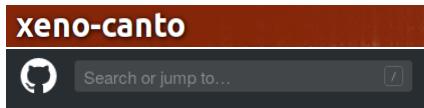


koudyk / birdsong-downloadXC

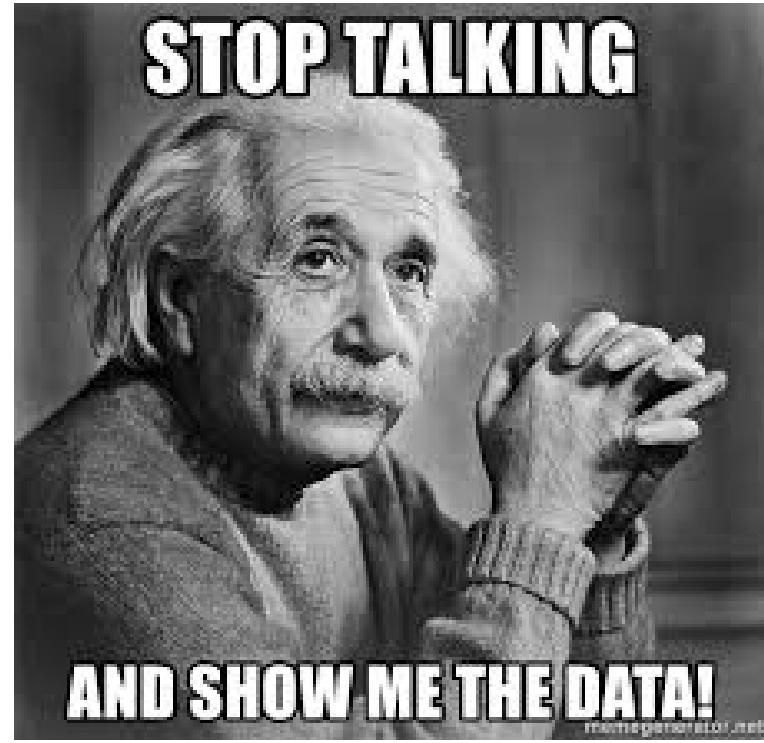


Data Category	N	Mean (StDev) length (sec)	Total (HHMMSS)
Original Recordings	1216	76.2 (97.0)	25:45:14
Excerpts	6659	3.2 (1.3)	05:55:48
Stimuli	100	3.9 (1.4)	00:06:32
Imitations	1700	5.9 (1.4)	00:09:52

Data Collection



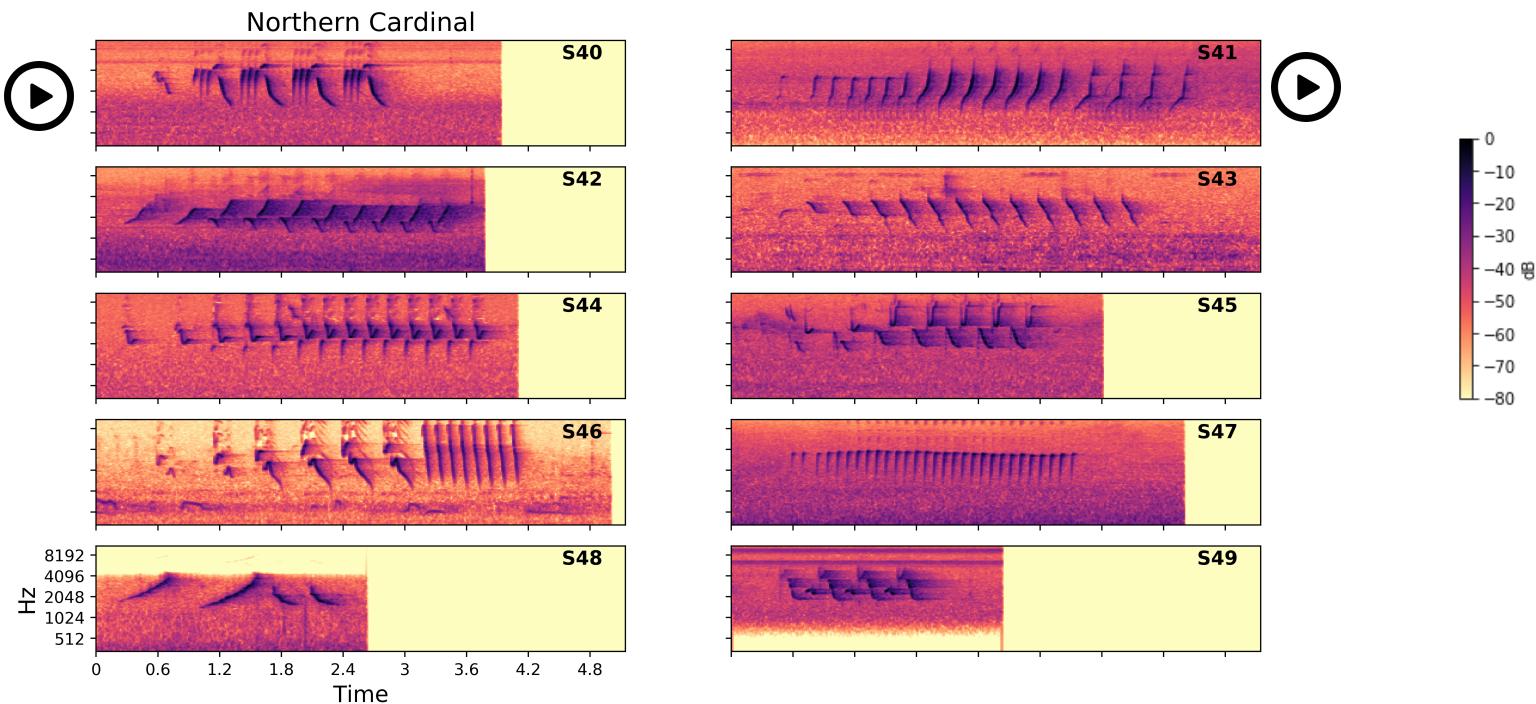
Data Category	N	Mean (StDev) length (sec)	Total (HHMMSS)
Original Recordings	1216	76.2 (97.0)	25:45:14
Excerpts	6659	3.2 (1.3)	05:55:48
Stimuli	100	3.9 (1.4)	00:06:32
Imitations	1700	5.9 (1.4)	00:09:52



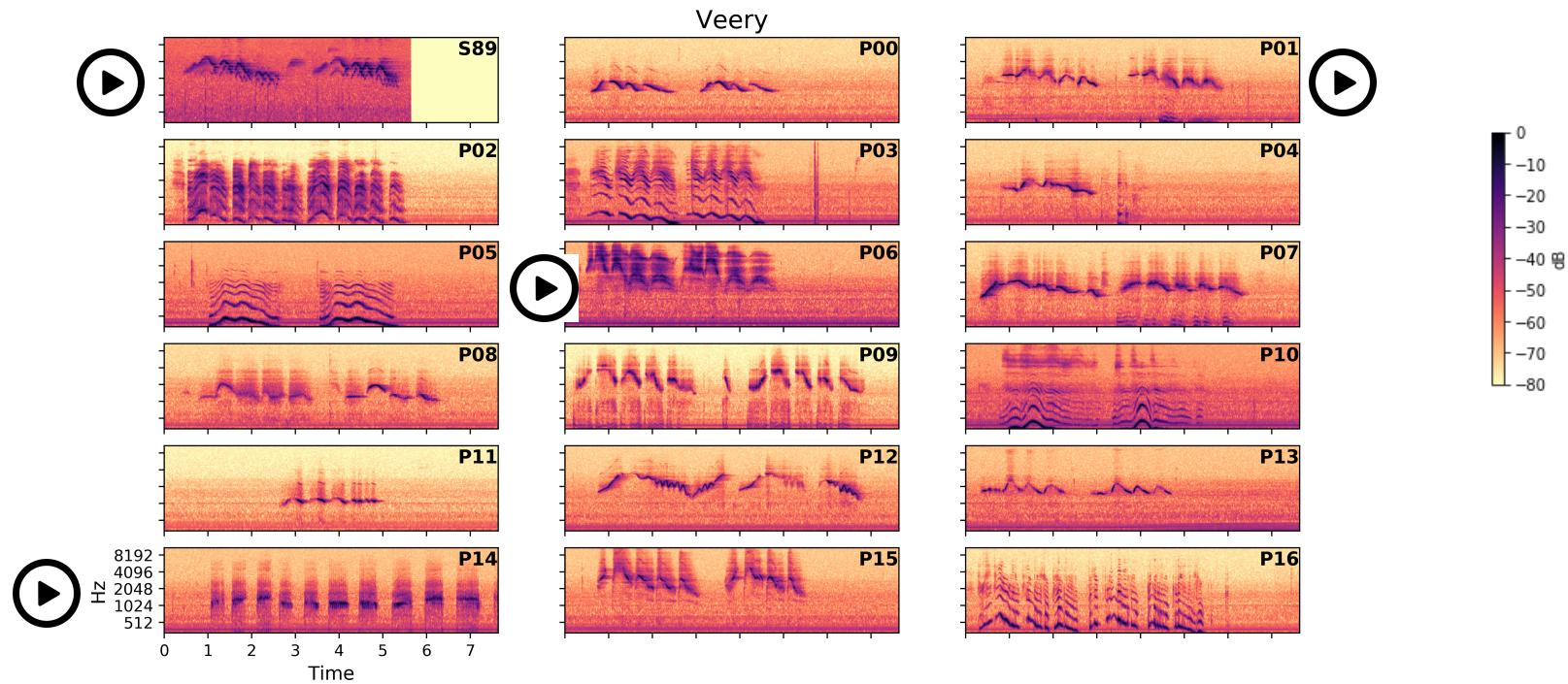
memegenerator.net

<https://www.google.com/url?sa=i&rct=j&q=&esrc=s&source=images&cd=&cad=rja&uact=8&ved=2ahUKEwil4eewtZTkAhVsU98KHUQZAB8Qjhx6BAgBEAI&url=https%3A%2F%2Fmemegenerator.net%2Finstance%2F65236303%2Falbert-einstein-stop-talking-and-show-me-the-data&psig=AOvVaw2EuolF8SApE3FOiXXIXY1N&ust=1566492454437579>

Variation within species



Variation between participants



Data

- Nested variation

Data

- Nested variation
- Can a human solve this problem?

Data

- Nested variation
- Can a human solve this problem?
 - Which stimulus is being imitated?

Stimulus 1 Stimulus 2 Stimulus 3



Imitation



Data

- Nested variation
- Can a human solve this problem?
 - Which stimulus is being imitated?

Stimulus 1 Stimulus 2 Stimulus 3



Imitation



Data

- Nested variation
- Can a human solve this problem?
 - Which stimulus is being imitated?



Imitation



Exploratory analyses

Exploratory analyses

- Spectral
- Temporal
- Spectrotemporal

Spectral analysis

- Goal: Separate species in the spectral feature space of imitations

Spectral analysis

- Goal: Separate species in the spectral feature space of imitations
- Pipeline:
 - (based on Kapur, A., Benning, M., & Tzanetakis, G., 2004)

Spectral analysis

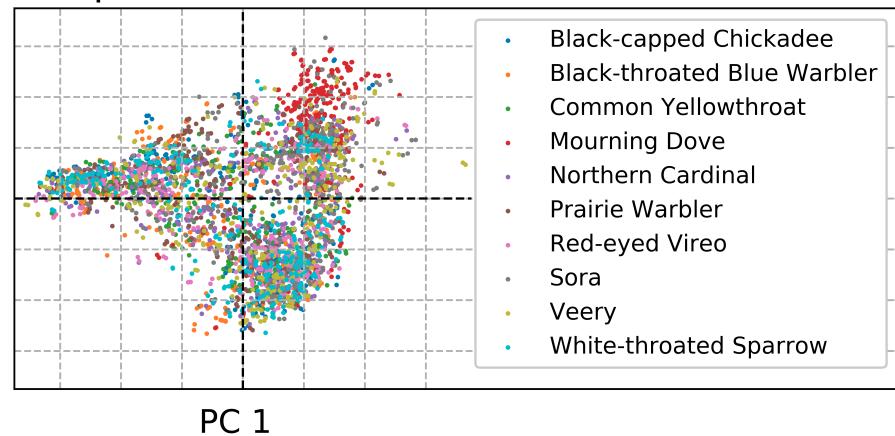
- Goal: Separate species in the spectral feature space of imitations
- Pipeline:
 - (based on Kapur, A., Benning, M., & Tzanetakis, G., 2004)
 - 2 highest-energy spectrogram frames
 - 12 MFCCs

Spectral analysis

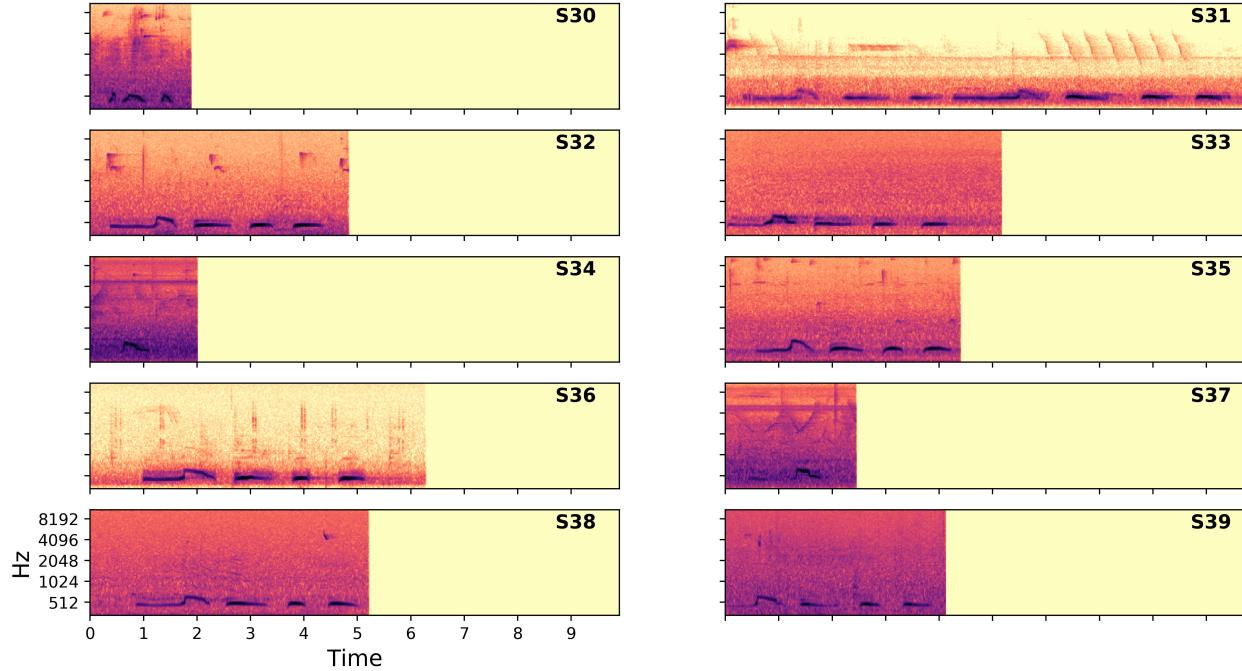
- Goal: Separate species in the spectral feature space of imitations
- Pipeline:
 - (based on Kapur, A., Benning, M., & Tzanetakis, G., 2004)
 - 2 highest-energy spectrogram frames
 - 12 MFCCs
 - PCA → 2 PCs
 - Visualize species separability

Spectral analysis - Results

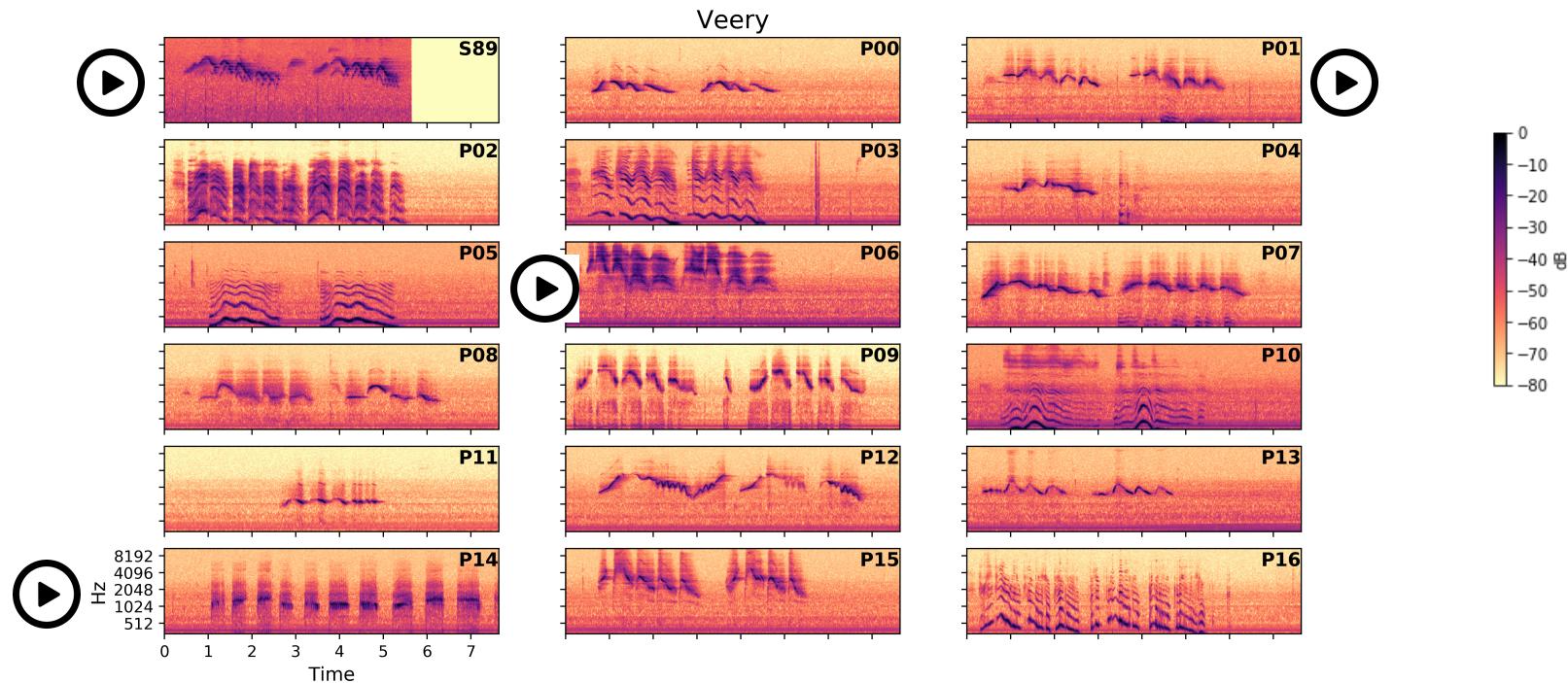
A) Species imitated



Mourning Dove

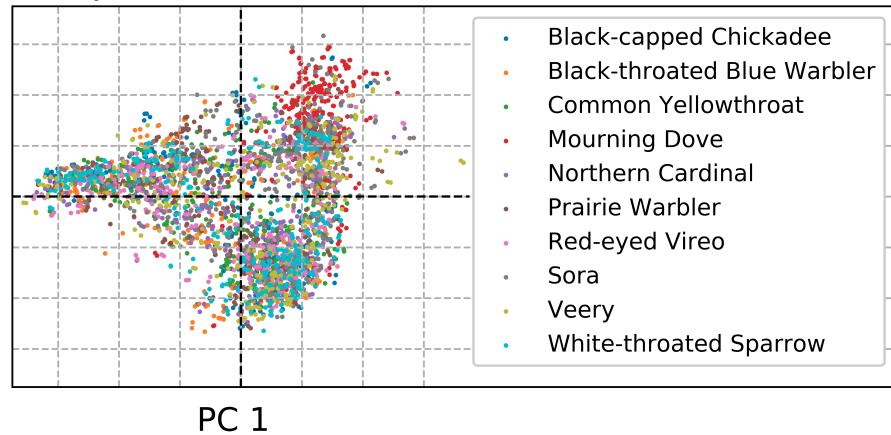


Variation between participants

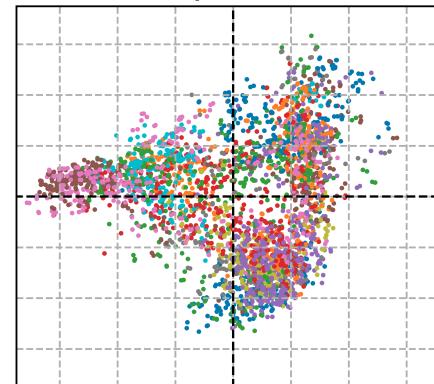


Spectral analysis - Results

A) Species imitated

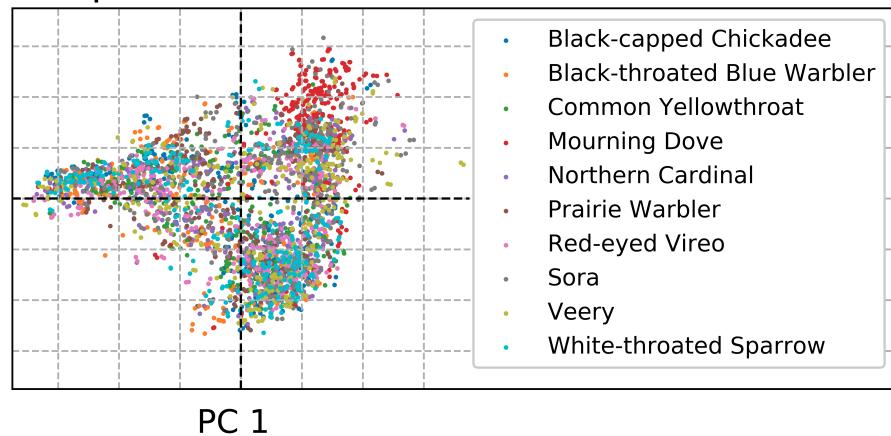


B) Participants

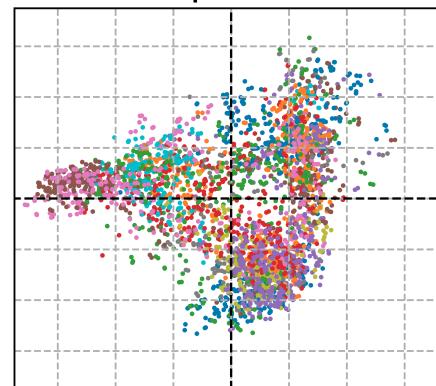


Spectral analysis - Results

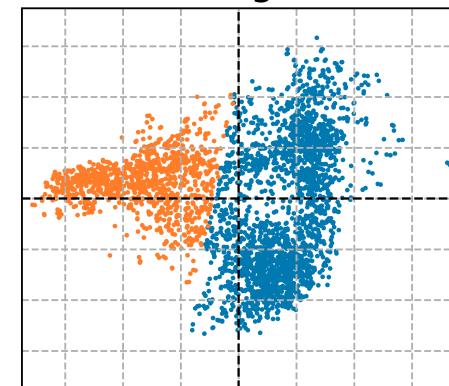
A) Species imitated



B) Participants



C) Clustering result

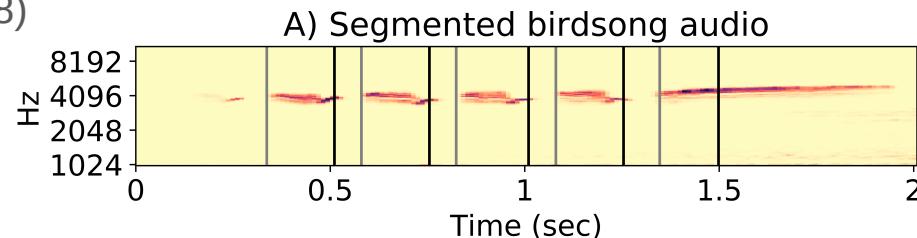


Temporal analysis

- Goal: compare **number of events** in stimuli vs imitations

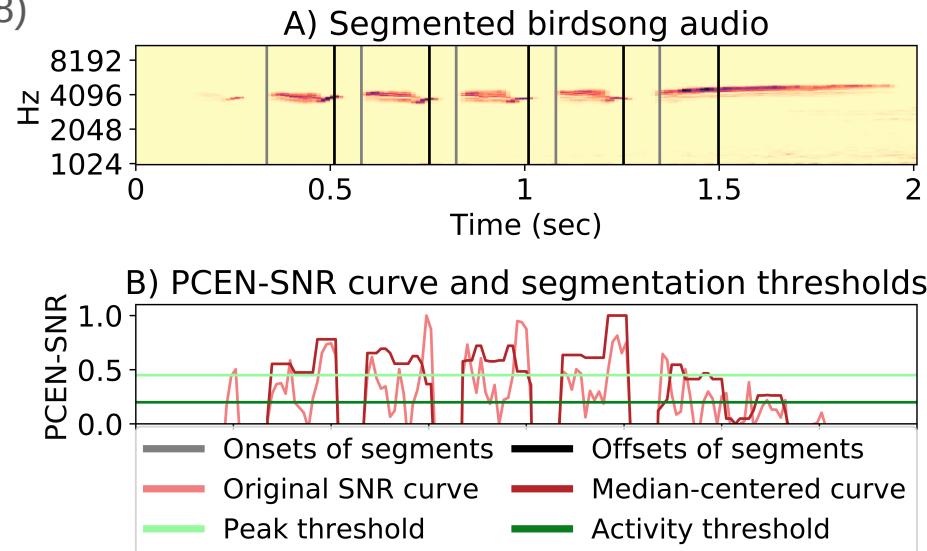
Temporal analysis

- Goal: compare **number of events** in stimuli vs imitations
- Pipeline:
 - Per-channel energy normalized mel spectrogram
(Wang et al., 2017; Lostanlen et al., 2018)



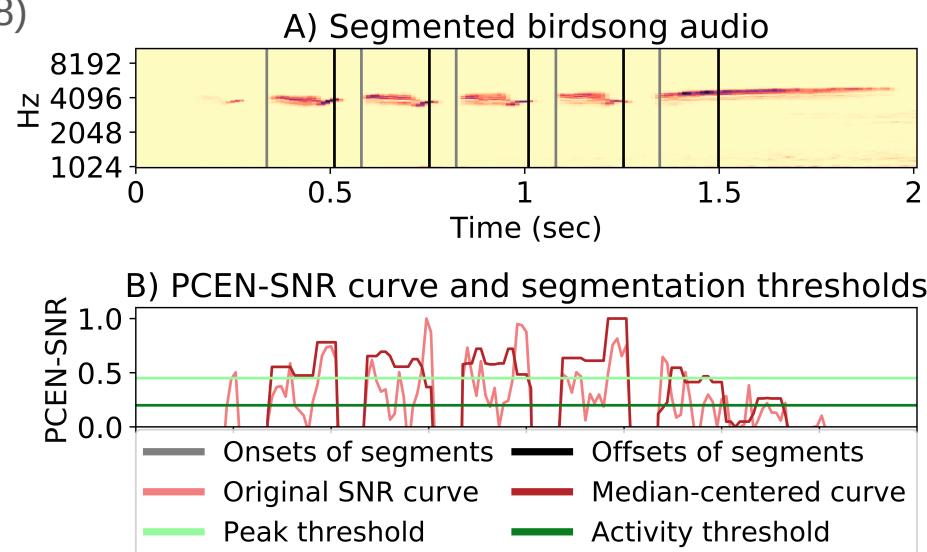
Temporal analysis

- Goal: compare **number of events** in stimuli vs imitations
- Pipeline:
 - Per-channel energy normalized mel spectrogram (Wang et al., 2017; Lostanlen et al., 2018)
 - Vocal activity detection

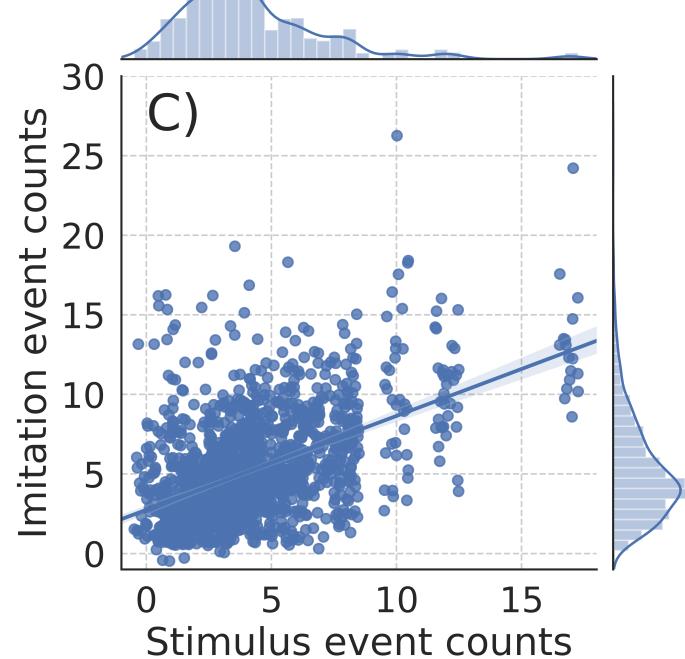


Temporal analysis

- Goal: compare **number of events** in stimuli vs imitations
- Pipeline:
 - Per-channel energy normalized mel spectrogram (Wang et al., 2017; Lostanlen et al., 2018)
 - Vocal activity detection
 - Segmentation
 - Compare stimulus & imitation counts



Temporal analysis - Results

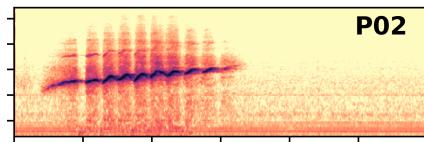
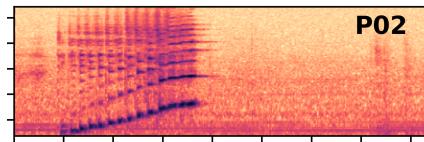
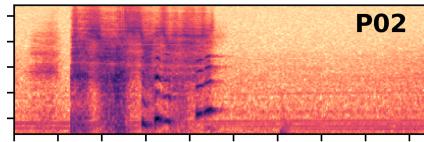


Spectrotemporal Analysis

- Goal: Compare **pitch contours** between stimuli and imitations

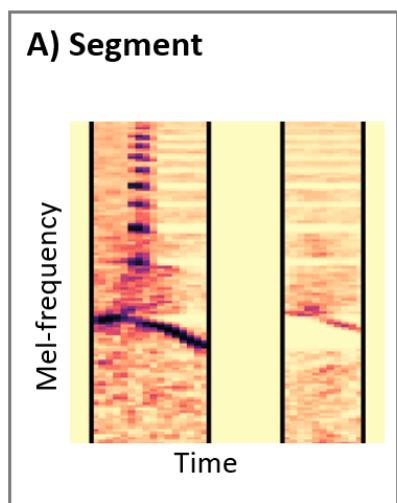
Spectrotemporal Analysis

- Goal: Compare **pitch contours** between stimuli and imitations
- Focus on 6 participants and 4 species with clearer pitch



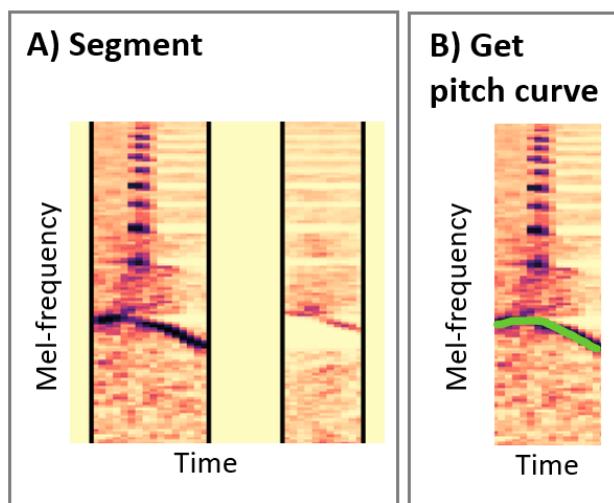
Spectrotemporal Analysis

- Goal: Compare **pitch contours** between stimuli and imitations
- Focus on 6 participants and 4 species with clearer pitch
- Pipeline:



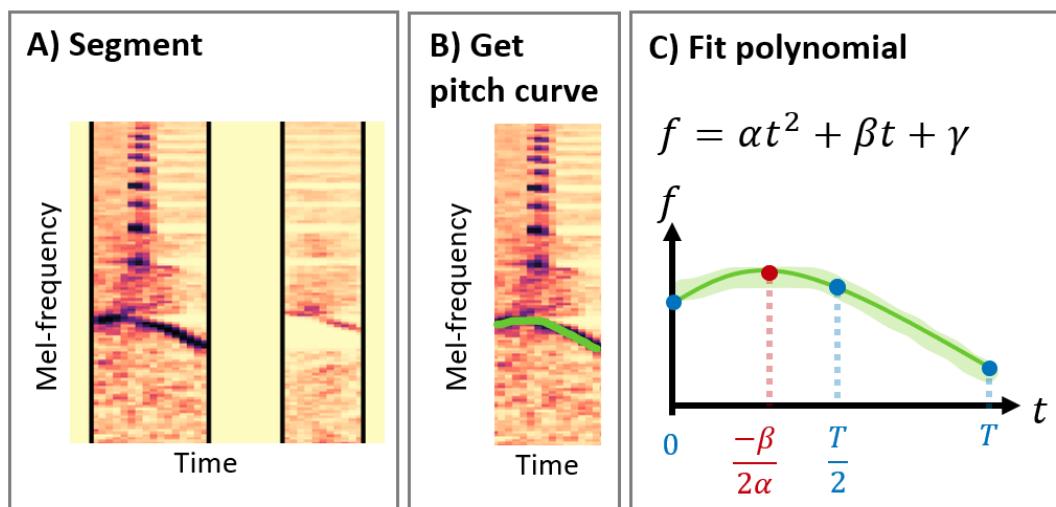
Spectrotemporal Analysis

- Goal: Compare **pitch contours** between stimuli and imitations
- Focus on 6 participants and 4 species with clearer pitch
- Pipeline:



Spectrotemporal Analysis

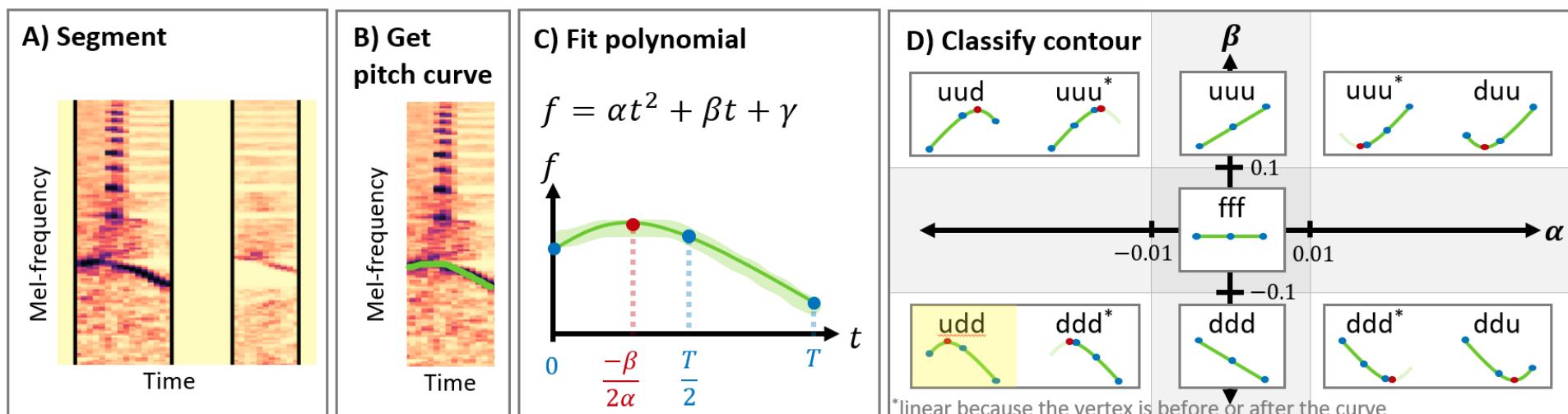
- Goal: Compare **pitch contours** between stimuli and imitations
- Focus on 6 participants and 4 species with clearer pitch
- Pipeline:



(Bittner et al., 2017)

Spectrotemporal Analysis

- Goal: Compare **pitch contours** between stimuli and imitations
- Focus on 6 participants and 4 species with clearer pitch
- Pipeline:



(Bittner et al., 2017)

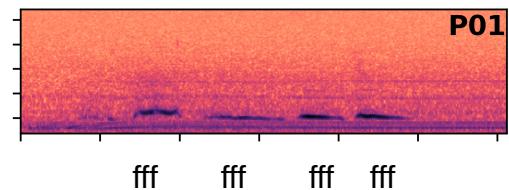
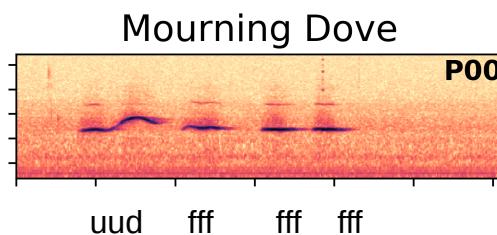
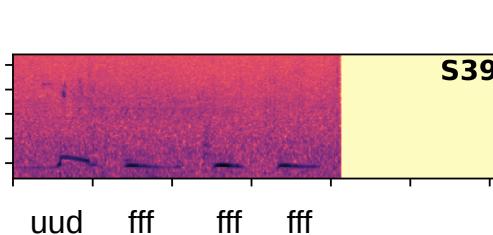
(Adams, 1976; Salamon et al., 2012)

Spectrotemporal Analysis

- Goal: Compare **pitch contours** between stimuli and imitations
- Focus on 6 participants and 4 species with clearer pitch
- Pipeline:
 - Segment
 - Get pitch curve
 - Fit polynomial (Bittner et al., 2017)
 - Classify contour (Adams, 1976; Salamon et al., 2012)

Spectrotemporal Analysis

- Goal: Compare **pitch contours** between stimuli and imitations
- Focus on 6 participants and 4 species with clearer pitch
- Pipeline:
 - Segment
 - Get pitch curve
 - Fit polynomial (Bittner et al., 2017)
 - Classify contour (Adams, 1976; Salamon et al., 2012)
 - Levenschtein distance (e.g., Lemström and Ukkonen, 2000)



Spectrotemporal analysis

- Consider 3 distances:
 - a) stimulus \leftrightarrow its 6 imitations
 - b) stimulus \leftrightarrow 6 imitations of a similar song from the same species
 - c) stimulus \leftrightarrow 6 random imitations from other species

Spectrotemporal analysis

- Consider 3 distances:
 - a) stimulus \leftrightarrow its 6 imitations
 - b) stimulus \leftrightarrow 6 imitations of a similar song from the same species
 - c) stimulus \leftrightarrow 6 random imitations from other species
- If the distance between pitch contours is useful, then $a \leq b < c$

Spectrotemporal analysis - Results

- Consider 3 distances:
 - a) stimulus ↔ its 6 imitations
 - b) stimulus ↔ 6 imitations of a similar song from the same species
 - c) stimulus ↔ 6 random imitations from other species
- If the distance between pitch contours is useful, then $a \leq b < c$
 - True for 71% of selected data

Discussion

- Results summary
 - Spectral analysis – identifies strategy
 - Temporal analysis – some correspondence
 - Spectrotemporal analysis – most promising, at least for pitched imitations

Discussion

- Limitations
 - Exploratory
 - Convenience sample
 - Small sample

Discussion

- Future directions
 - More data
 - Online data collection?



<https://imgflip.com/i/iqu9>

Discussion

- Future directions
 - More data
 - Online data collection?
 - Different analyses for different imitation strategies?

Discussion

- Future directions
 - More data
 - Online data collection?
 - Different analyses for different imitation strategies?
 - Deep adversarial learning?

Imperfect open science

- Data collection, analysis notebooks, paper, slides:
https://github.com/BirdVox/oudyk_vihar2019
- Scraping bird sounds from Xeno-Canto.org:
<https://github.com/koudyk/birdsong-downloadXC>
- Data:
Coming on <https://birdvoximitation.weebly.com>
Example on https://github.com/BirdVox/oudyk_vihar2019/tree/master/data

Acknowledgments

- Participants
- BirdVox team at NYU and CLoO
- Course instructors at JYU
- Supervisor at McGill, J.B. Poline
- Funding
 - Leon Levy Foundation
 - National Science Foundation's Big Data grant 1633206
 - Travel grant from the University of Jyväskylä (KO)



JYVÄSKYLÄN YLIOPISTO
UNIVERSITY OF JYVÄSKYLÄ



Mixed Linear Model Regression Results

```
=====
```

Model: MixedLM Dependent Variable: imitation_counts
No. Observations: 1700 Method: REML
No. Groups: 10 Scale: 7.0076
Min. group size: 170 Likelihood: -4084.1087
Max. group size: 170 Converged: Yes
Mean group size: 170.0

```
-----
```

	Coef.	Std.Err.	z	P> z	[0.025	0.975]
Intercept	3.243	0.308	10.545	0.000	2.641	3.846
stimulus_counts	0.472	0.027	17.396	0.000	0.419	0.525
Group Var	0.776	0.148				

```
=====
```