Learning auditory similarities between instrumental techniques

Vincent Lostanlen[1], Joakim Andén[2], Grégoire Lafay[3,4], Mathieu Lagrange[3,5]

[1]Music and Audio Research Lab, New York University
[2]Flatiron Institute, Simons Foundation
[3]École Centrale de Nantes [4]Lonofi [5]CNRS

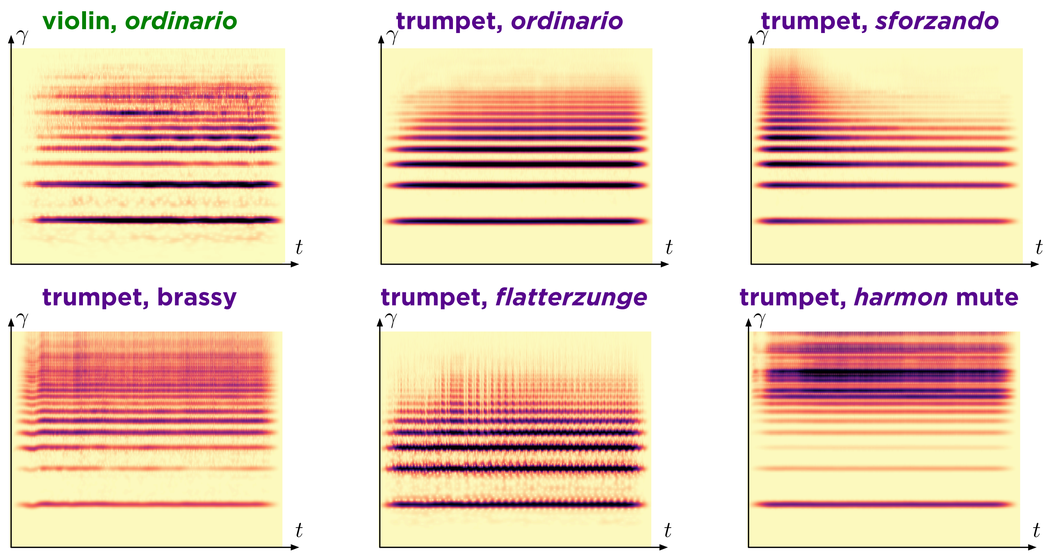
www.lostanlen.com
vincent.lostanlen@nyu.edu

- Motivation: modeling **timbre** in the European instrumentarium.
- Context: computer-assisted **spectralist orchestration** [Maresz].
- Prior work with multidimensional scaling from judgments…
  - revealed the **limitations of short-term** features [McAdams].
- Prior work with spectrotemporal receptive fields (**STRF**) …
  - but with a single ***ordinario*** sample per instrument [Patil].
- How to **generalize** to a broader timbral palette?

Use a corpus of **extended playing techniques** as stimuli.

Why playing techniques (PT) are interesting

- In contemporary music, *ordinario* is only one of many options.
- Beyond erudite music, PT are often a part of **musical folklore**:
  - Klezmer **slide**, Irish **fiddle**, jazz **growl**, rockabilly **slap**, etc.
- If instruments are the "*what?*" of timbre, PT are the ***"how?"***.
- PT induce a **sensation of motion** without sight nor touch.
- Instrument recognition on *ordinario* samples may be solved…
  - but **playing technique retrieval is ripe for more MIR research**.



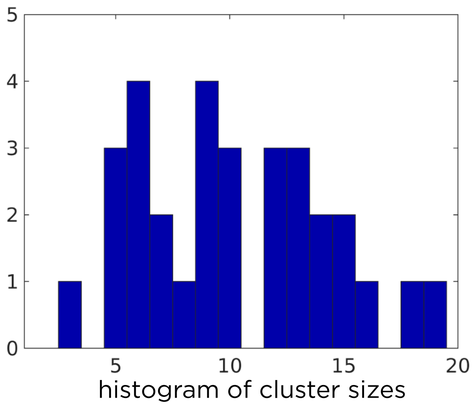
- On a spectrogram, there can be more **shape similarity** between two *ordinari* than between two PT from the same instrument.

Expressing perceived similarity by clustering

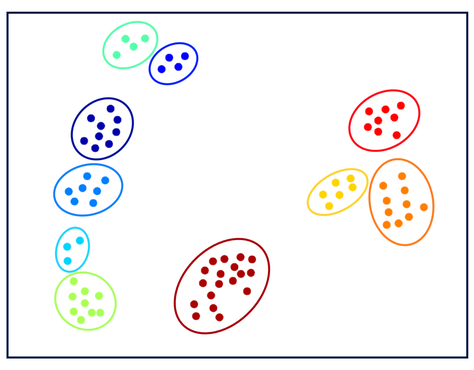
- How to collect similarity judgments between *n* stimuli?
- Rating scale: quantitatively ill-defined, requires n^2 ratings.
- "Odd-one-out": more intuitive but requires n^3 ratings.

Use **hard cluster assignments** as a proxy for similarity.

- Audio: 78 PT from 16 instruments in SOL dataset (Ircam).
- Subjects: 31 adult students from the Paris Conservatory.
- Number of clusters: minimum = 3, maximum = 19, median = 10.

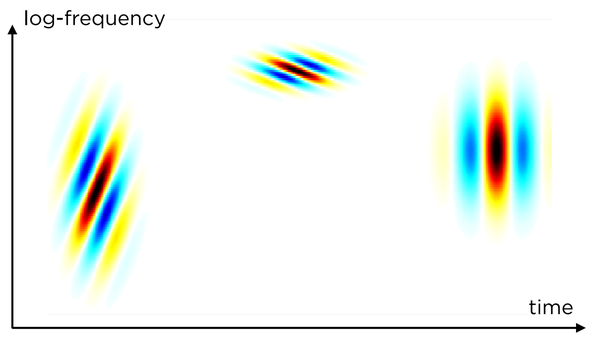


histogram of cluster sizes


annotation interface with 10 clusters

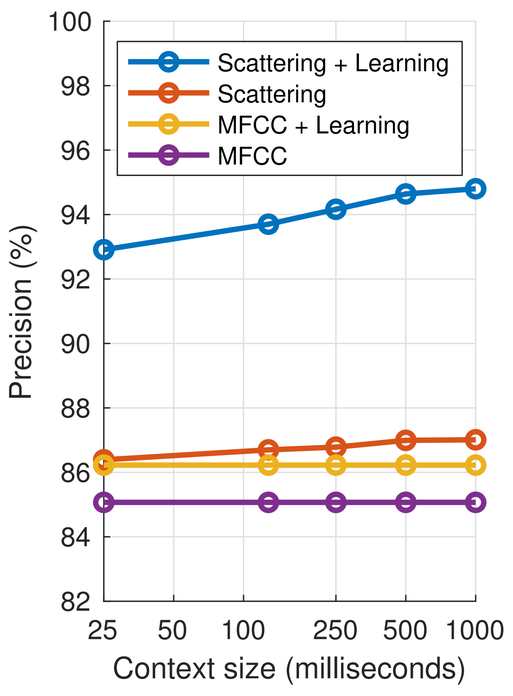
Time–frequency scattering with wavelets

- **Neuro-inspired**: accelerated STRF
- **convnet** architecture yet without training.
- Euclidean distances encode **deformations** in the time-frequency domain [Mallat].



- but do they reflect **perception**?

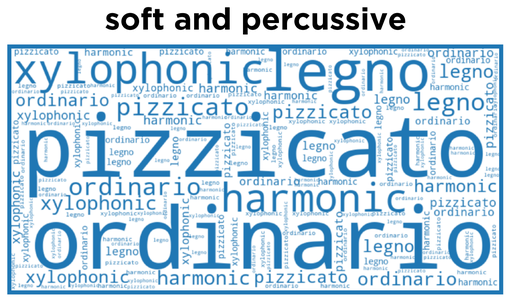
Incorporating subjective judgments



- **Metric learning** algorithm: Large-Margin Nearest Neighbors. (LMNN)
- We find a consensus between subjects with **hypergraph partitioning**.
- Euclidean scattering outperforms adapted MFCC.
- Best fit with scattering and LMNN combined.
- **Long context** (~500 ms) reveals the importance of spectrotemporal dynamics.

- **Consensus clustering** can be interpreted post hoc:

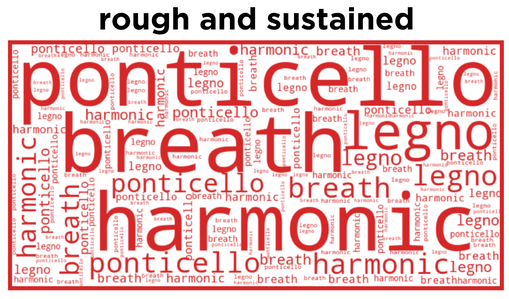
soft and percussive



rough and percussive

soft and sustained

rough and sustained



- The well-known **percussive/sustained dichotomy** is maintained beyond the perception of *ordinario* samples.

Towards a chiromusicology of timbre

- Timbre is not only about **organology** (instruments as inert objects) but also **chiromusicology** (gestures).
- Work in progress: discussing consensus and idiosyncrasy.
- **Open-source** software libraries:
  - in MATLAB: `github.com/lostanlen/scattering.m`
  - in Python: `pip install kymatio`

NEW YORK UNIVERSITY
FLATIRON INSTITUTE
Division of Simons Foundation
cnrs
CENTRALE NANTES