Learning Unsupervised Hierarchies Of Audio Concepts

Darius Afchar[1,2], Romain Hennequin[1], Vincent Guigue[2]

[1]Deezer Research, Paris
[2]MLIA, ISIR - Sorbonne Université - CNRS, France

Introduction

Raw music signals are challenging to interpret
→ concept learning was proposed in *computer vision*

However, music has specificities not well handled:
→ high number of concepts
→ strong correlation of concepts (eg. Funk-Rock)

We propose a **novel task** more suited to MIR:

Learning and hierarchising concepts

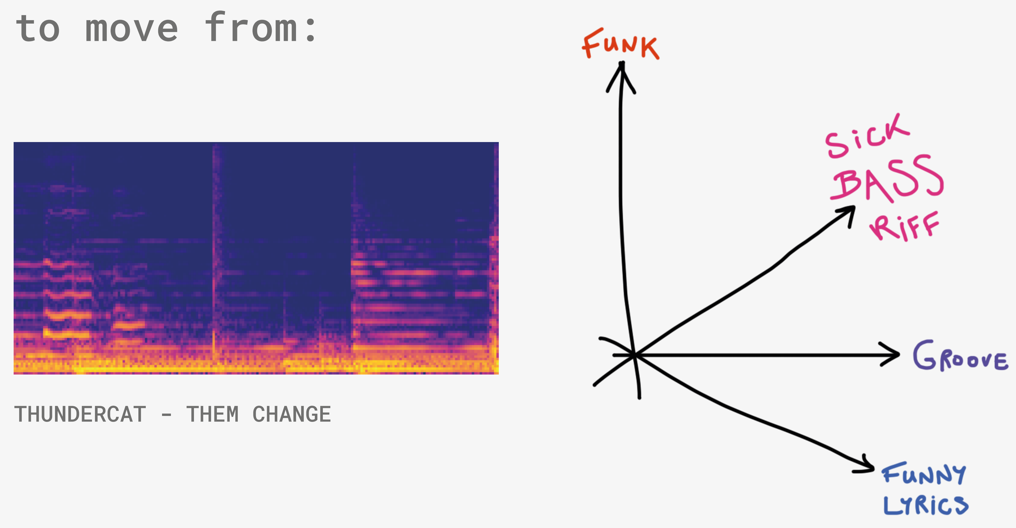
→ open the doors to new music descriptors (eg. folksonomy)
→ better understand how music is organised
→ solution to a new task in concept learning

Explainability goals

Informed by XAI papers, we define three explanation goals to ensure the relevance of our description:

1. **Attribution**
Identify music concept within a music track
2. **Transferability**
Invariance of concepts to setting and task
3. **Generality**
Invariance of hierarchy, make sense in general

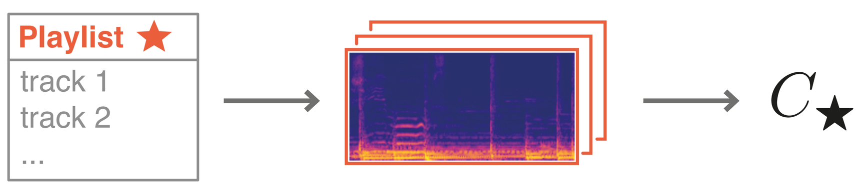
The general spirit of concept learning is to move from: ... to:

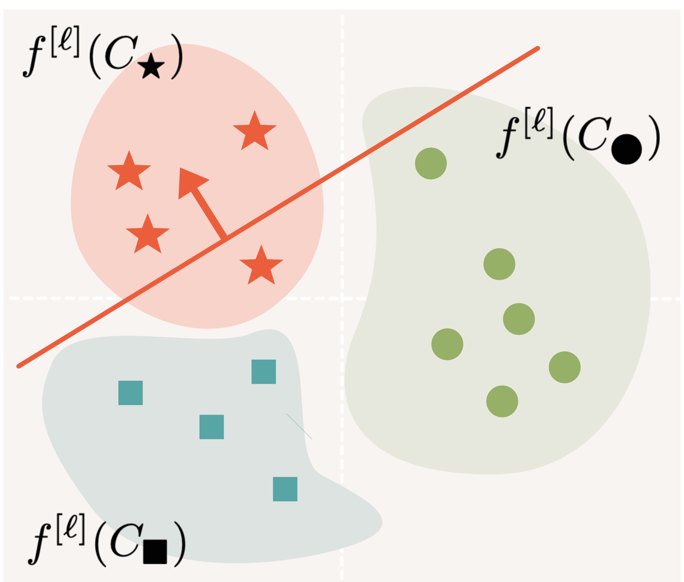


We explore this idea purely for audio signals
→ experiment with playlists as concepts
→ (co)-first paper on concept learning in MIR

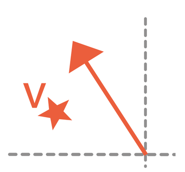
Learning audio concepts

We **adapt TCAV** to music tracks spectrograms
→ consider playlists as source of concepts





→ we obtain a concept vector



Background on TCAV [Kim2018]

→ was proposed in *computer vision*
→ we learn to linearly discriminate:

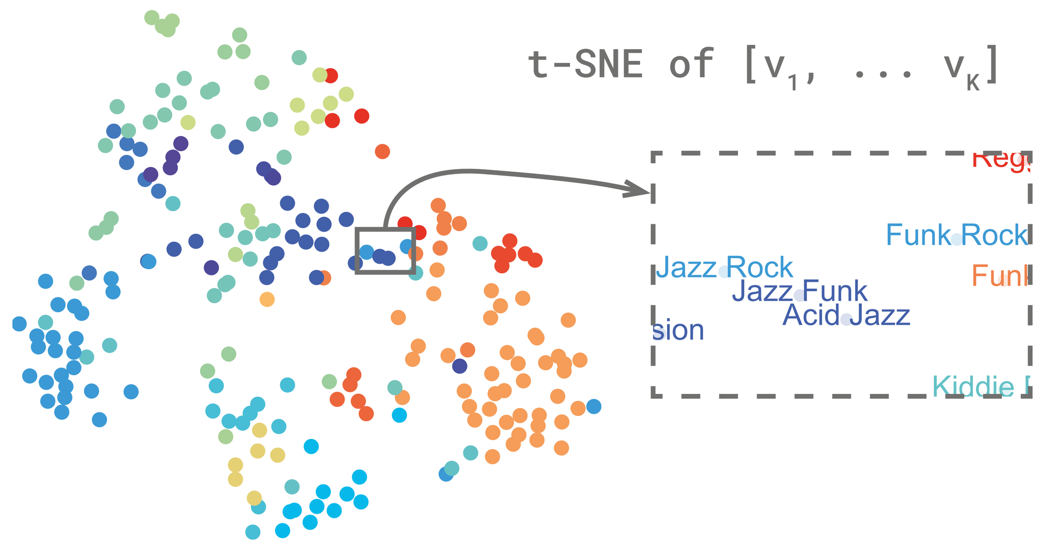
$$\{\vec{f}^{[\ell]}(x) \mid x \in C\} \quad \text{versus} \quad \{\cdots \mid x \notin C\}$$

backbone — few-shot examples

→ solved in practice with a logistic regression
→ learned parameter: *concept activation vector*

Hierarchising concepts

Useful observation: similarities make sense
→ vector similarity coincides with concept similarity



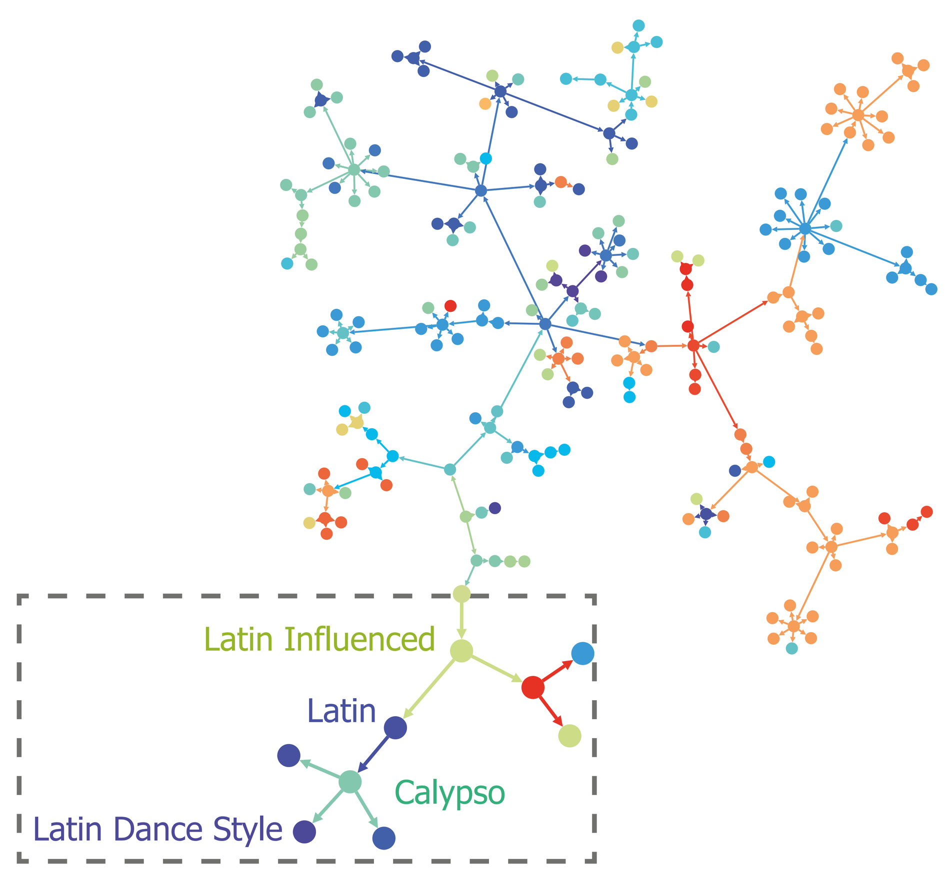
→ we propose a fast similarity estimation:

$$S_{i,j}(\mathcal{V}) = \mathbb{E}_{x \sim C_i}\left[\sigma(\langle \vec{v}_j, \vec{f}^{[\ell]}(x) \rangle)\right]$$

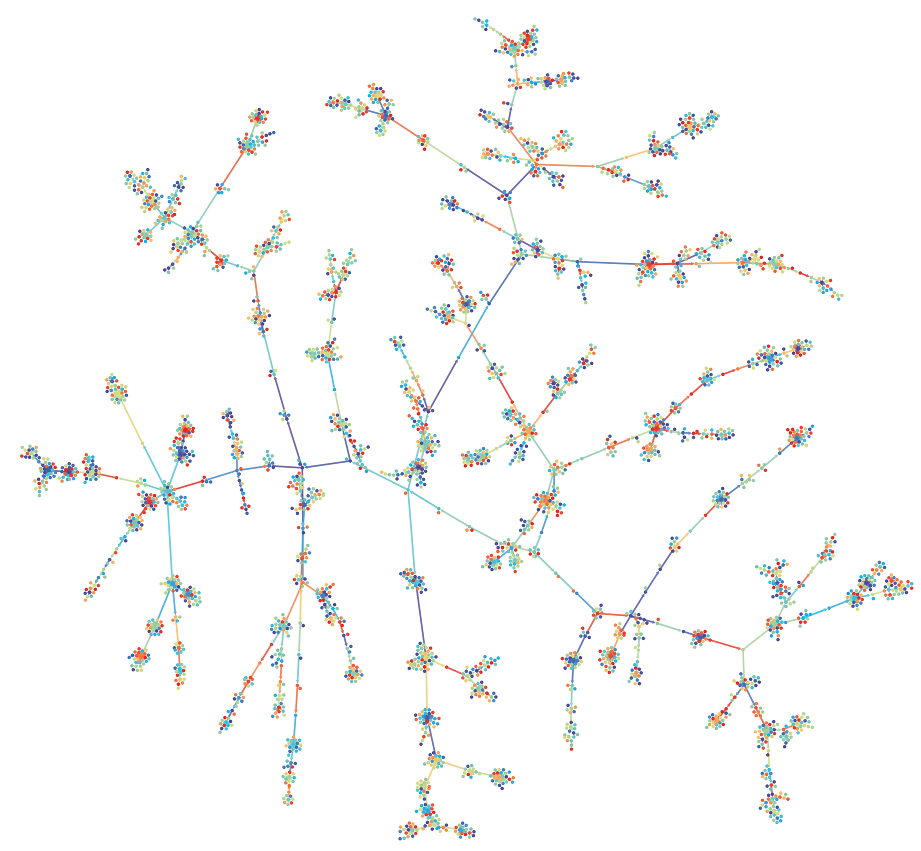
Useful algorithm: centrality-based hierarchy
→ proposed in NLP to folksonomies from tags
→ we show that the hypotheses work for music

Betweeness hierarchy [Heymann2006]

```
* INPUT: similarity graph S
* compute betweeness-centrality of S
* for nodes of S ordered by centrality:
*    add new node to hierarchy H
*    edge to most similar node in H
* OUTPUT: tree H
```



Resulting hierarchy of APM's genres

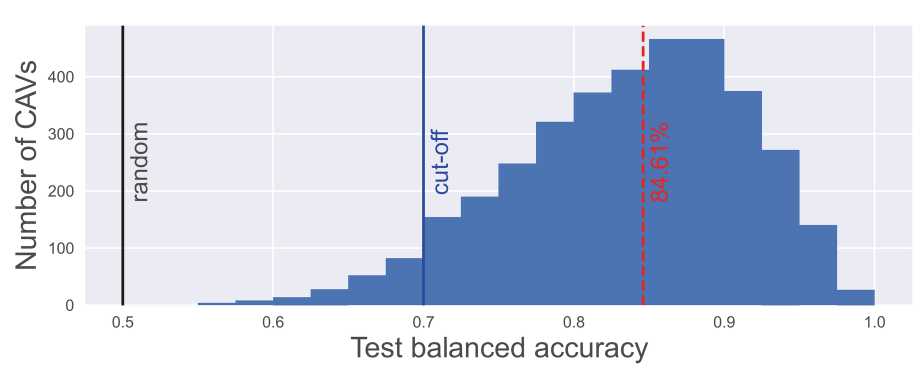


Resulting hierarchy of 3443 Deezer playlists

Experiments

Leverage three datasets:
→ APM Music's genres (219)
→ APM Music's moods (165)
→ Deezer playlists (3443) ie. 3443^{3441} trees

1. **Attribution**



2. **Transferability**
→ related to the design, choice of backbone, concepts

3. **Generality**
→ discuss and analysis of priors and expected hierarchy
→ ground-truth evaluation for genre and mood
→ proxy evaluation for concept of mixed types
→ quantitative discussion

Source	Audio (↓)	CF (↓)	BERT (↑)	$W2V_1$ (↑)	$W2V_2$ (↑)
$H(S(\mathcal{V}_D))$	**2.449 ± 0.022**	0.845 ± 0.013	0.345 ± 0.007	0.286 ± 0.009	0.542 ± 0.007
$H(S_{CF})$	**2.413 ± 0.021**	**0.345 ± 0.007**	0.416 ± 0.008	0.336 ± 0.010	0.601 ± 0.008
$H(S_{BERT})$	2.858 ± 0.028	0.868 ± 0.013	**0.726 ± 0.005**	0.505 ± 0.011	0.652 ± 0.008
$H(S_{W2V\text{-}1})$	2.952 ± 0.028	0.932 ± 0.012	0.523 ± 0.008	**0.804 ± 0.005**	0.721 ± 0.007
$H(S_{W2V\text{-}2})$	2.843 ± 0.026	0.847 ± 0.012	0.531 ± 0.008	0.596 ± 0.009	**0.836 ± 0.004**
Random	3.388 ± 0.027	1.104 ± 0.006	0.239 ± 0.004	0.142 ± 0.006	0.452 ± 0.006

Gist of the table: the hierarchy found from audio compared to one we can find through collaborative filtering or specialised w2v.

Take-away

→ First paper of concept learning on music spectrograms
→ Novel idea to hierarchise learned concepts
→ Framed through XAI to be as relevant as possible
→ Complementary to expert ontologies, our method enables seeing how music is organically described by users and editors

Future work

→ Better performances, curb biases, save humanity
→ Dynamic recommendations
→ Play with concepts evolution through time

@dafchar
research@deezer.com
research.deezer.com/concept_hierarchy/

23rd International Society of Music Information Retrieval Conference (ISMIR 2022)