



Universitat  
Pompeu Fabra  
*Barcelona*

**MTG**  
Music Technology  
Group



---

# Signal Processing and Machine Learning Challenges in Sound and Music Computing

Xavier Serra

Music Technology Group

Universitat Pompeu Fabra, Barcelona

# CV

PhD



Stanford  
University

Post-doc



Since 1994



Universitat  
Pompeu Fabra  
*Barcelona*

**MTG**  
Music Technology  
Group

# Disclaimer

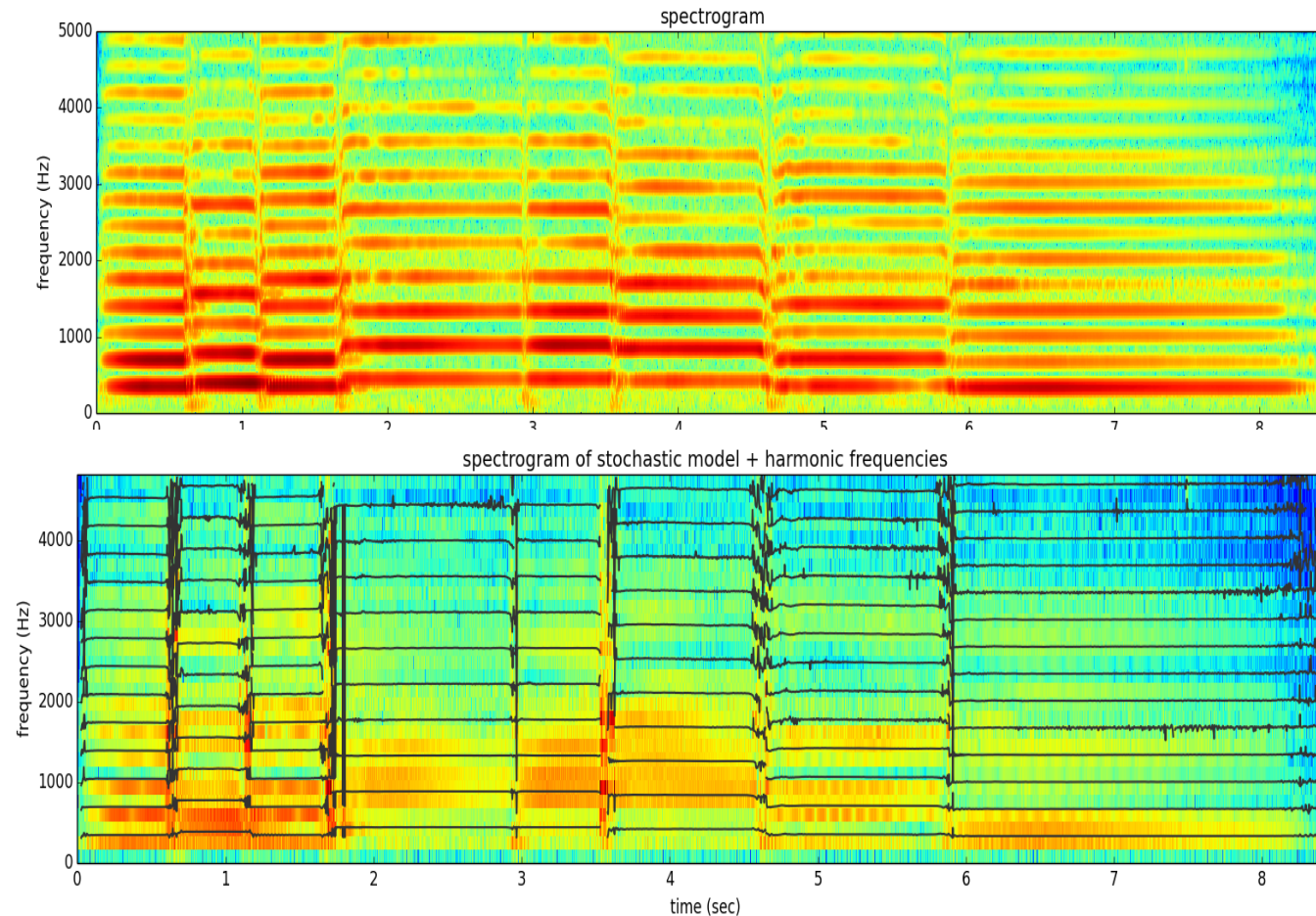
Biased towards  
our research !!!

# Index

- Personal research highlights
- The field of Sound and Music Computing
- Current research challenges
- Conclusions

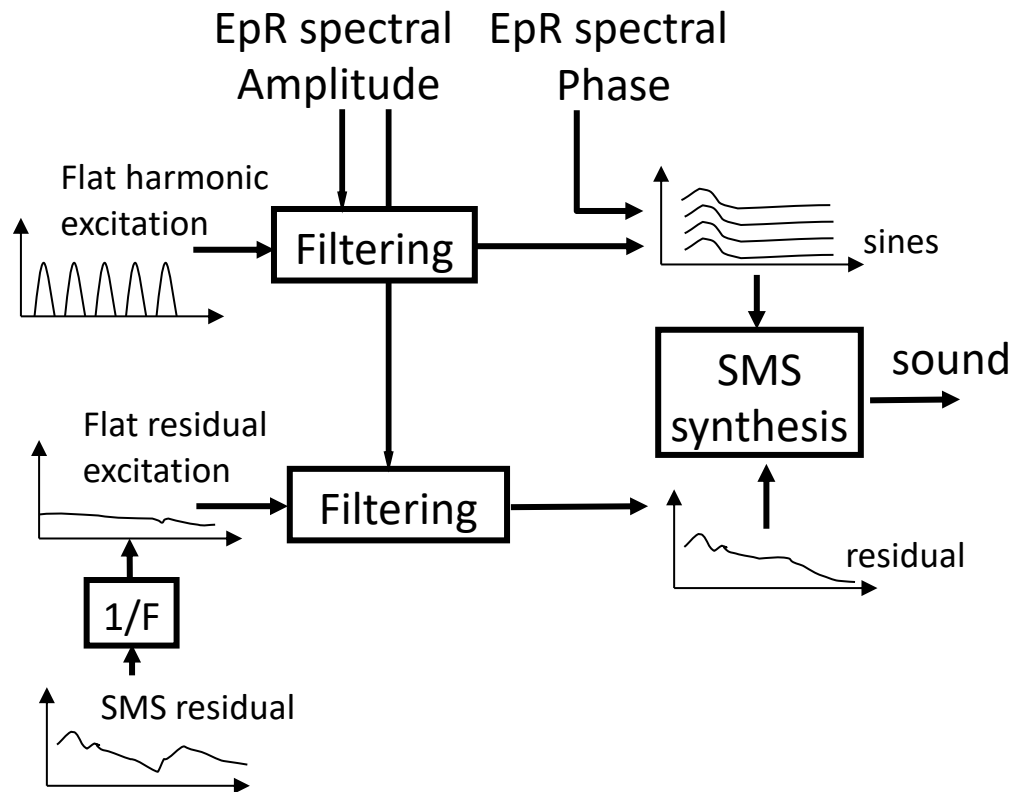


# Personal research trajectory (1 of 4)



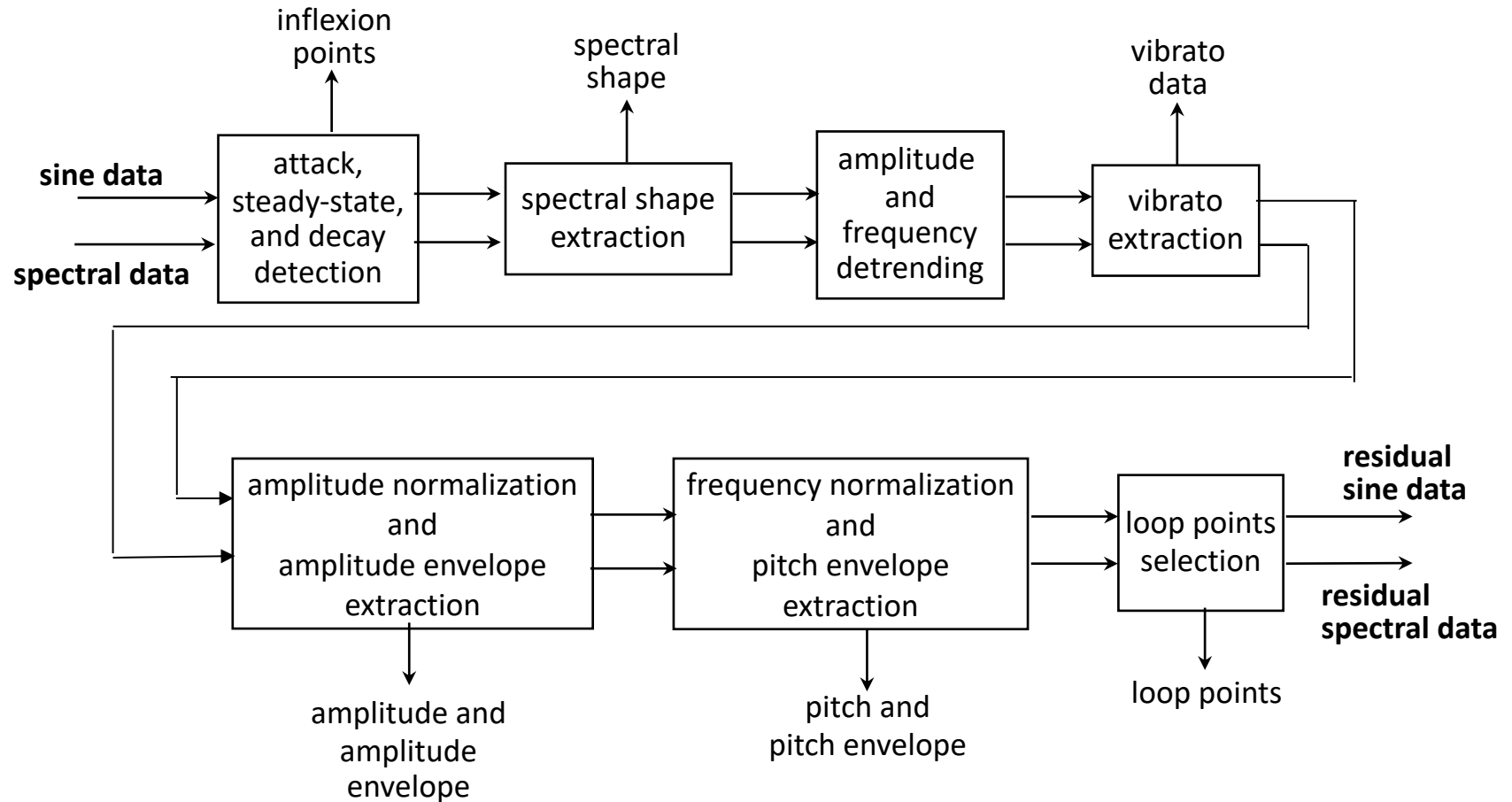
**X. Serra.** 1989. *A System for Sound Analysis/Transformation/Synthesis based on a Deterministic plus Stochastic Decomposition.* PhD Thesis.

# Personal research trajectory (2 of 4)



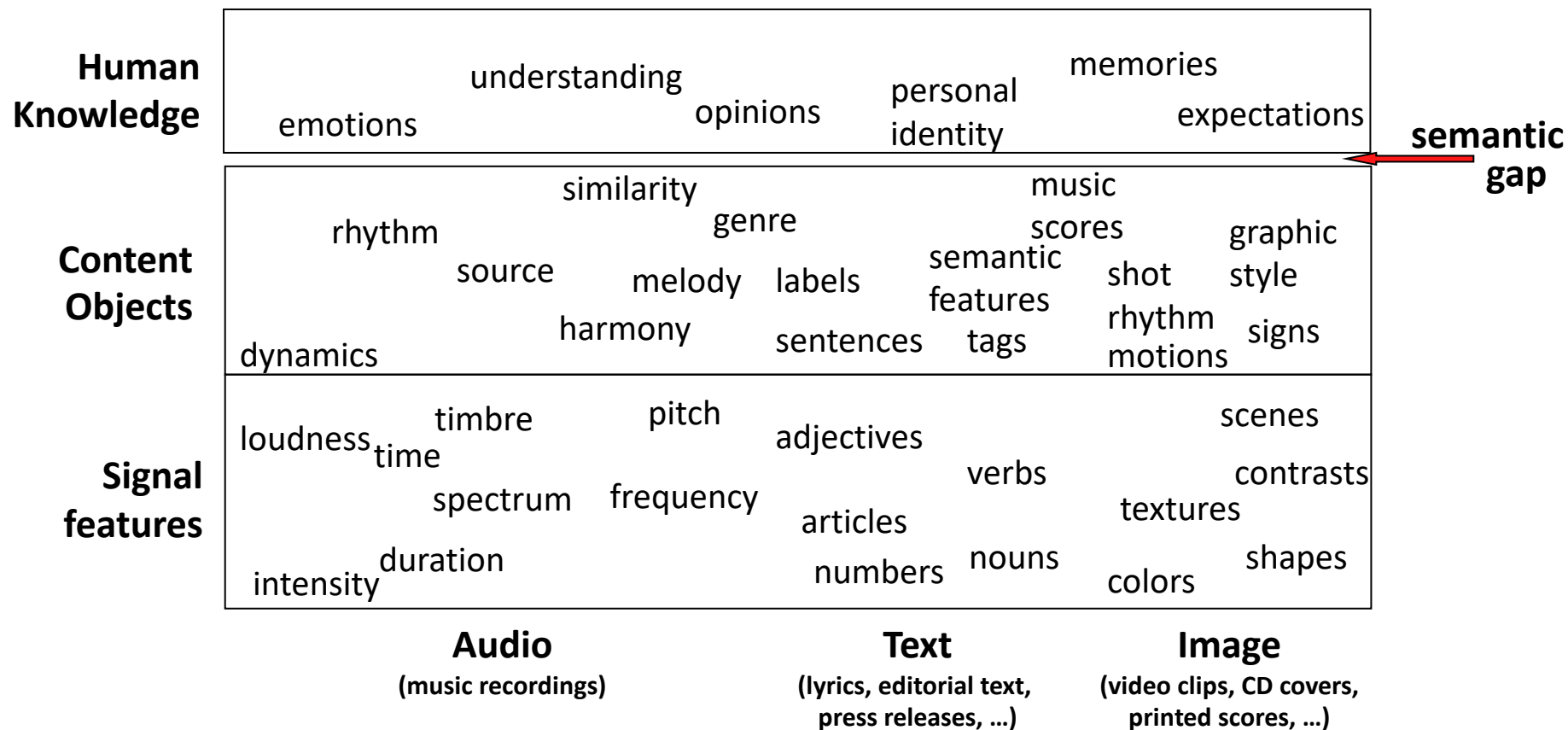
J. Bonada, O. Celma, A. Loscos, J. Ortolà, X. Serra. 2001. "Singing Voice Synthesis Combining Excitation plus Resonance and Sinusoidal plus Residual Models." *ICMC*.

# Personal research trajectory (3 of 4)



P. Herrera, X. Serra, G. Peeters. 1999. "Audio Descriptors and Descriptor Schemes in the Context of MPEG-7." *ICMC*.

# Personal research trajectory (4 of 4)



P. Cano, M. Koppenberger, N. Wack, J. G. Mahedero, T. Aussenac, R. Marxer, J. Masip, O. Celma, D. Garcia, E. Gómez, F. Gouyon, E. Guaus, P. Herrera, J. Massaguer, B. Ong, M. Ramírez, S. Streich, X. Serra. 2005. "Content-based Music Audio Recommendation." *ACM Multimedia*.

# The field of Sound and Music Computing (1 of 2)

“By combining scientific, technological and artistic methodologies it aims at **understanding**, **modelling** and **generating** sound and music through computational approaches.”

X. Serra, G. Widmer, M. Leman. 2007. *A Roadmap for Sound and Music Computing*. S2S Consortium.

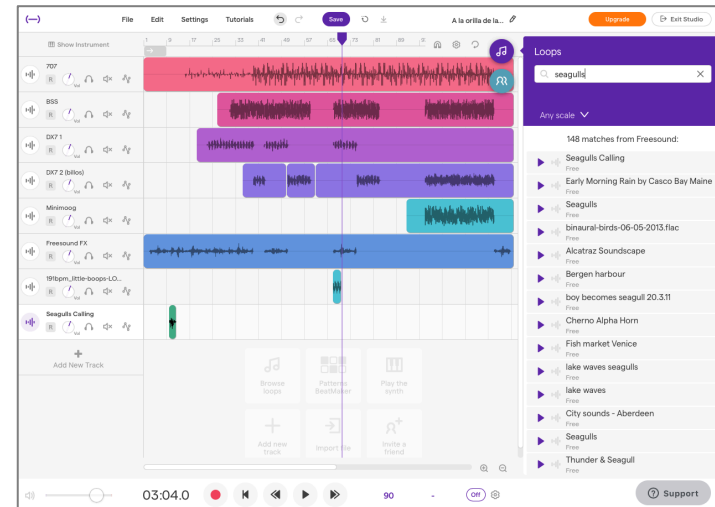
# The field of Sound and Music Computing (2 of 2)

- Digital music instruments



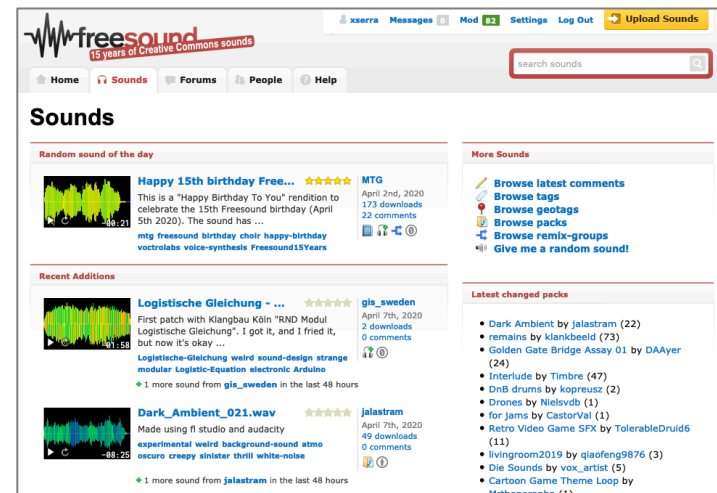
# The field of Sound and Music Computing (2 of 2)

- Digital music instruments
- **Sound and music production**



# The field of Sound and Music Computing (2 of 2)

- Digital music instruments
- Sound and music production
- Sound and music archiving





# The field of Sound and Music Computing (2 of 2)

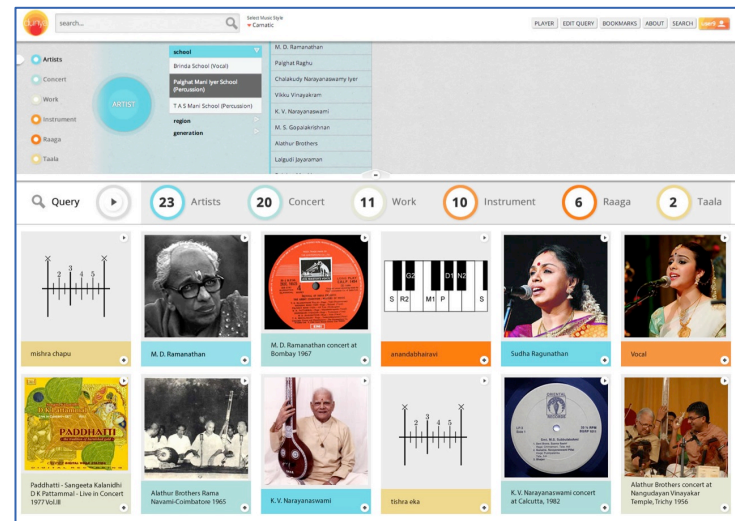
- Digital music instruments
- Sound and music production
- Sound and music archiving
- **Sound and music retrieval**



The screenshot displays the JAMENDO search interface. On the left, there is a 'REFINE YOUR SEARCH' sidebar with various filters: 'Kind Of Audio' (Music, Sound sample), 'Creation Date' (Start/End Date), 'License Type' (CC BY, CC BY-SA, CC BY-NC), 'Genre' (Pop, Electronic, Dance, Hip-Hop, Country, Jazz, Classical, Indie, Punk, Reggae, Latin, World, Ambient, Blues, R&B, Funk, Folk, Soul, New Age, Film, Experimental, Ambient, Synth, Latin, March, Spoken Word, Cabaret, Protest), 'Moods' (Happy, Energetic, Sad), and 'Themes'. The main area shows '1037 results found' from the 'JAMENDO' provider, sorted by 'Relevance High To Low'. The results table includes columns for Title, waveform, Author, Creation date, Duration, License type, Download, and Provider Link. The first few results are: 'afoperawo - Equip...' by afoperawo (4 Feb 2015, 04:46), 'Messes' by Seth Power (3 Oct 2017, 04:19), 'corazón temeroso' by Aursonika (12 Aug 2018, 05:31), 'Camaleão' by Safári de Safarino (29 Jan 2018, 02:13), 'Sergio de 3' by Safári de Safarino (29 Jan 2018, 03:34), 'Vicio' by Safári de Safarino (29 Jan 2018, 03:44), 'Memorizar' by Kurita Desiranta (9 Feb 2018, 01:40), 'Cobertes' by Safári de Safarino (10 Jan 2017, 04:07), 'BeBass' by Kurita Desiranta (9 Feb 2018, 01:39), and 'Historia de alguém' by Safári de Safarino (29 Jan 2018, 02:56).

# The field of Sound and Music Computing (2 of 2)

- Digital music instruments
- Sound and music production
- Sound and music archiving
- Sound and music retrieval
- **Computational musicology**



# The field of Sound and Music Computing (2 of 2)

- Digital music instruments
- Sound and music production
- Sound and music archiving
- Sound and music retrieval
- Computational musicology
- ...

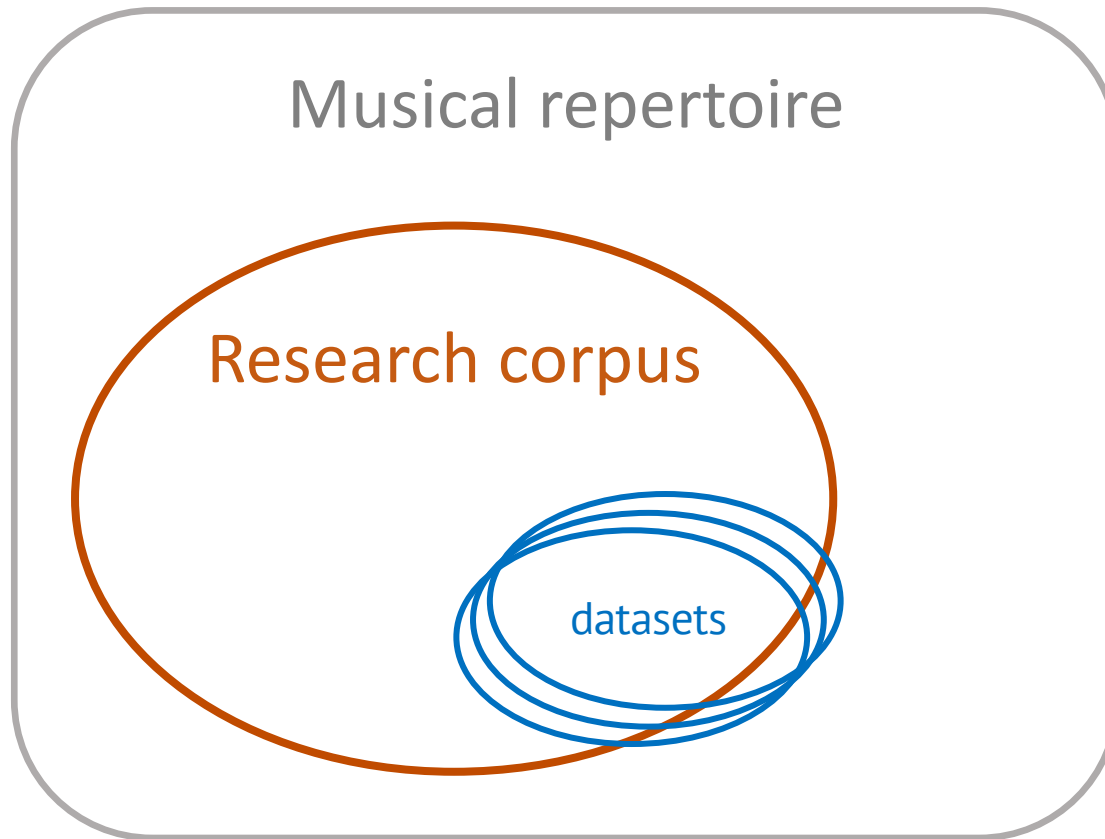
# Current research challenges

- Corpora and datasets
- Musical understanding
- Neural Networks for musical tasks
- Musical sound synthesis
- Separation of musical audio sources
- Sound and music classification
- Applications
- ...

# Current research challenges

- **Corpora and datasets**
- Musical understanding
- Neural Networks for musical tasks
- Musical sound synthesis
- Separation of musical audio sources
- Sound and music classification
- Applications
- ...

# Challenge: Corpora and Datasets (1 of 2)



- Purpose
- Coverage
- Completeness
- Quality
- Reusability

X. Serra. 2014. "Creating Research Corpora for the Computational Study of Music: the case of the CompMusic Project." *AES*.

# Challenge: Corpora and Datasets (2 of 2)

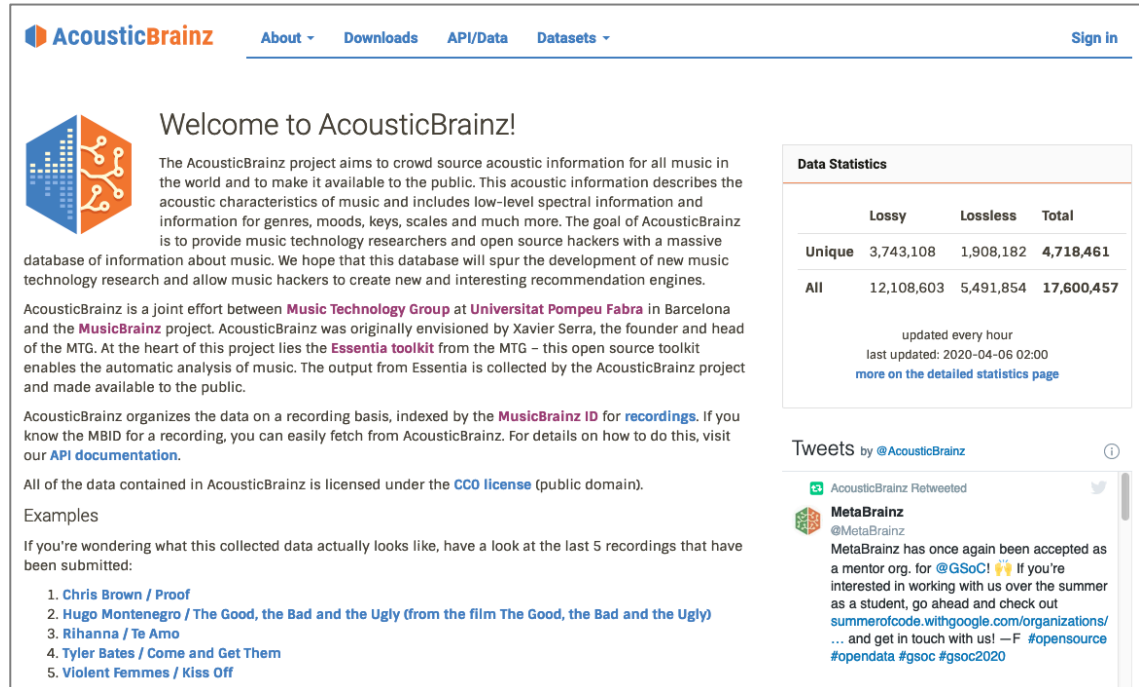
- **freesound.org**

The screenshot shows the homepage of freesound.org. At the top, there is a navigation bar with links for "Sounds", "Forums", "People", and "Help". A search bar is located on the right side of the header. The main content area is divided into several sections:

- Random sound of the day:** Features a sound titled "jumps from reason.3.Rev.wav" by user "martian". It includes a waveform visualization, a 5-star rating, and statistics: "May 26th, 2006", "1769 downloads", and "5 comments".
- Freesound Blog:** Contains three recent posts:
  - 15 years of Freesound!** (April 4th, 2020 by frederic.font)
  - Sustainability Report 2019** (March 31st, 2020 by frederic.font)
  - Recording Foley for Beginners on a Budget** (March 20th, 2020 by Joao\_Janz)
- Support Us. Get your Freesound T-Shirt!** A section promoting merchandise with buttons for "Europe" and "US Canada and Asia".
- Love Freesound?** A section with a "Donate Now" button.
- Active Forum Threads:** A section listing active discussions, including "Site is operating slowly" and "drum pad practicing or rehearsing beats and breaks".

# Challenge: Corpora and Datasets (2 of 2)

- freesound.org
- acousticbrainz.org



The screenshot shows the AcousticBrainz website. The header includes the logo and navigation links: About, Downloads, API/Data, Datasets, and Sign in. The main content area features a welcome message and a detailed description of the project's goals and mission. A 'Data Statistics' table is prominently displayed, showing the number of recordings in Lossy, Lossless, and Total formats, categorized by Unique and All. Below the table, it notes that the statistics are updated every hour and provides the last update time. A section for 'Tweets by @AcousticBrainz' is also visible, featuring a tweet from MetaBrainz.

**AcousticBrainz** About Downloads API/Data Datasets Sign in

## Welcome to AcousticBrainz!

The AcousticBrainz project aims to crowd source acoustic information for all music in the world and to make it available to the public. This acoustic information describes the acoustic characteristics of music and includes low-level spectral information and information for genres, moods, keys, scales and much more. The goal of AcousticBrainz is to provide music technology researchers and open source hackers with a massive database of information about music. We hope that this database will spur the development of new music technology research and allow music hackers to create new and interesting recommendation engines.

AcousticBrainz is a joint effort between **Music Technology Group at Universitat Pompeu Fabra** in Barcelona and the **MusicBrainz** project. AcousticBrainz was originally envisioned by Xavier Serra, the founder and head of the MTG. At the heart of this project lies the **Essentia toolkit** from the MTG – this open source toolkit enables the automatic analysis of music. The output from Essentia is collected by the AcousticBrainz project and made available to the public.

AcousticBrainz organizes the data on a recording basis, indexed by the **MusicBrainz ID** for **recordings**. If you know the MBID for a recording, you can easily fetch from AcousticBrainz. For details on how to do this, visit our [API documentation](#).

All of the data contained in AcousticBrainz is licensed under the [CCO license](#) (public domain).

### Examples

If you're wondering what this collected data actually looks like, have a look at the last 5 recordings that have been submitted:

1. [Chris Brown / Proof](#)
2. [Hugo Montenegro / The Good, the Bad and the Ugly \(from the film The Good, the Bad and the Ugly\)](#)
3. [Rihanna / Te Amo](#)
4. [Tyler Bates / Come and Get Them](#)
5. [Violent Femmes / Kiss Off](#)

### Data Statistics

|        | Lossy      | Lossless  | Total             |
|--------|------------|-----------|-------------------|
| Unique | 3,743,108  | 1,908,182 | <b>4,718,461</b>  |
| All    | 12,108,603 | 5,491,854 | <b>17,600,457</b> |

updated every hour  
last updated: 2020-04-06 02:00  
[more on the detailed statistics page](#)

### Tweets by @AcousticBrainz

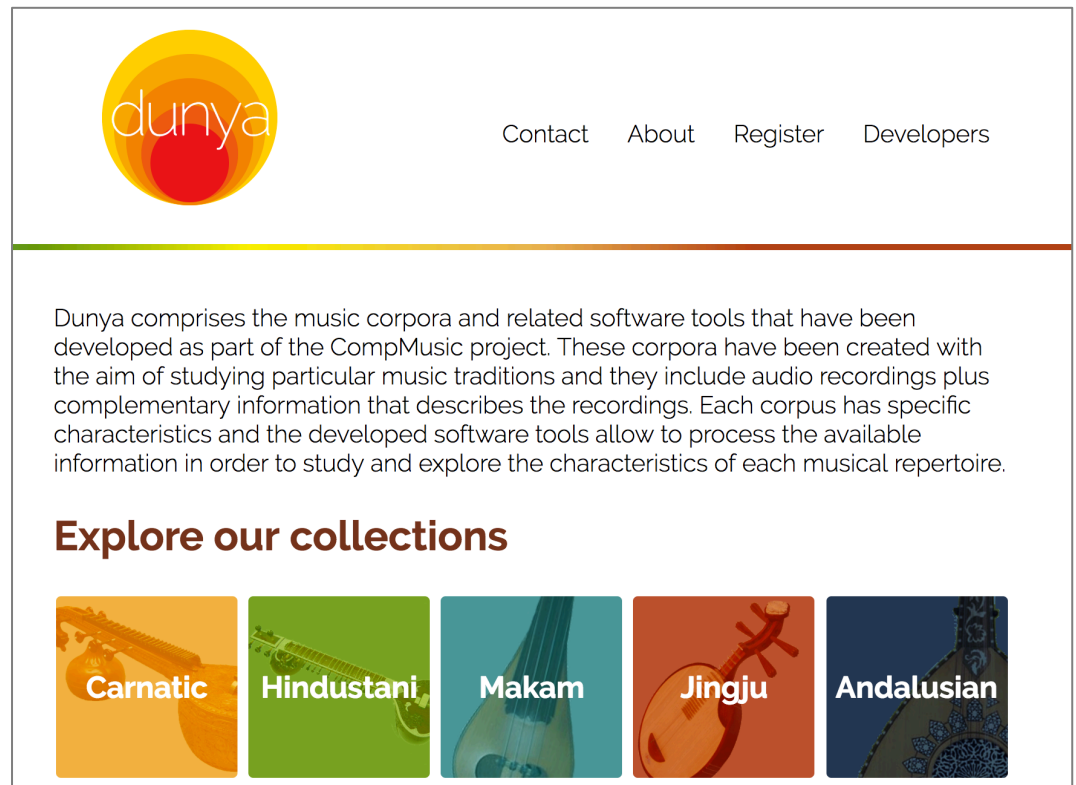
AcousticBrainz Retweeted

**MetaBrainz** @MetaBrainz  
MetaBrainz has once again been accepted as a mentor org. for @GSoc! 🙌 If you're interested in working with us over the summer as a student, go ahead and check out [summerofcode.withgoogle.com/organizations/...](#) and get in touch with us! —F #opensource #opendata #gsoc #gsoc2020



# Challenge: Corpora and Datasets (2 of 2)

- [freesound.org](https://freesound.org)
- [acousticbrainz.org](https://acousticbrainz.org)
- [dunya.upf.edu](https://dunya.upf.edu)



The screenshot shows the Dunya website interface. At the top left is the Dunya logo, a stylized sun with the word "dunya" written across it. To the right of the logo are navigation links: "Contact", "About", "Register", and "Developers". Below the navigation bar is a horizontal line. The main content area contains a paragraph of text describing the project: "Dunya comprises the music corpora and related software tools that have been developed as part of the CompMusic project. These corpora have been created with the aim of studying particular music traditions and they include audio recordings plus complementary information that describes the recordings. Each corpus has specific characteristics and the developed software tools allow to process the available information in order to study and explore the characteristics of each musical repertoire." Below this text is a section titled "Explore our collections" with five colored buttons representing different musical traditions: Carnatic (orange), Hindustani (green), Makam (teal), Jingju (red), and Andalusian (dark blue).

**dunya**      Contact    About    Register    Developers

Dunya comprises the music corpora and related software tools that have been developed as part of the CompMusic project. These corpora have been created with the aim of studying particular music traditions and they include audio recordings plus complementary information that describes the recordings. Each corpus has specific characteristics and the developed software tools allow to process the available information in order to study and explore the characteristics of each musical repertoire.

### Explore our collections

- Carnatic
- Hindustani
- Makam
- Jingju
- Andalusian

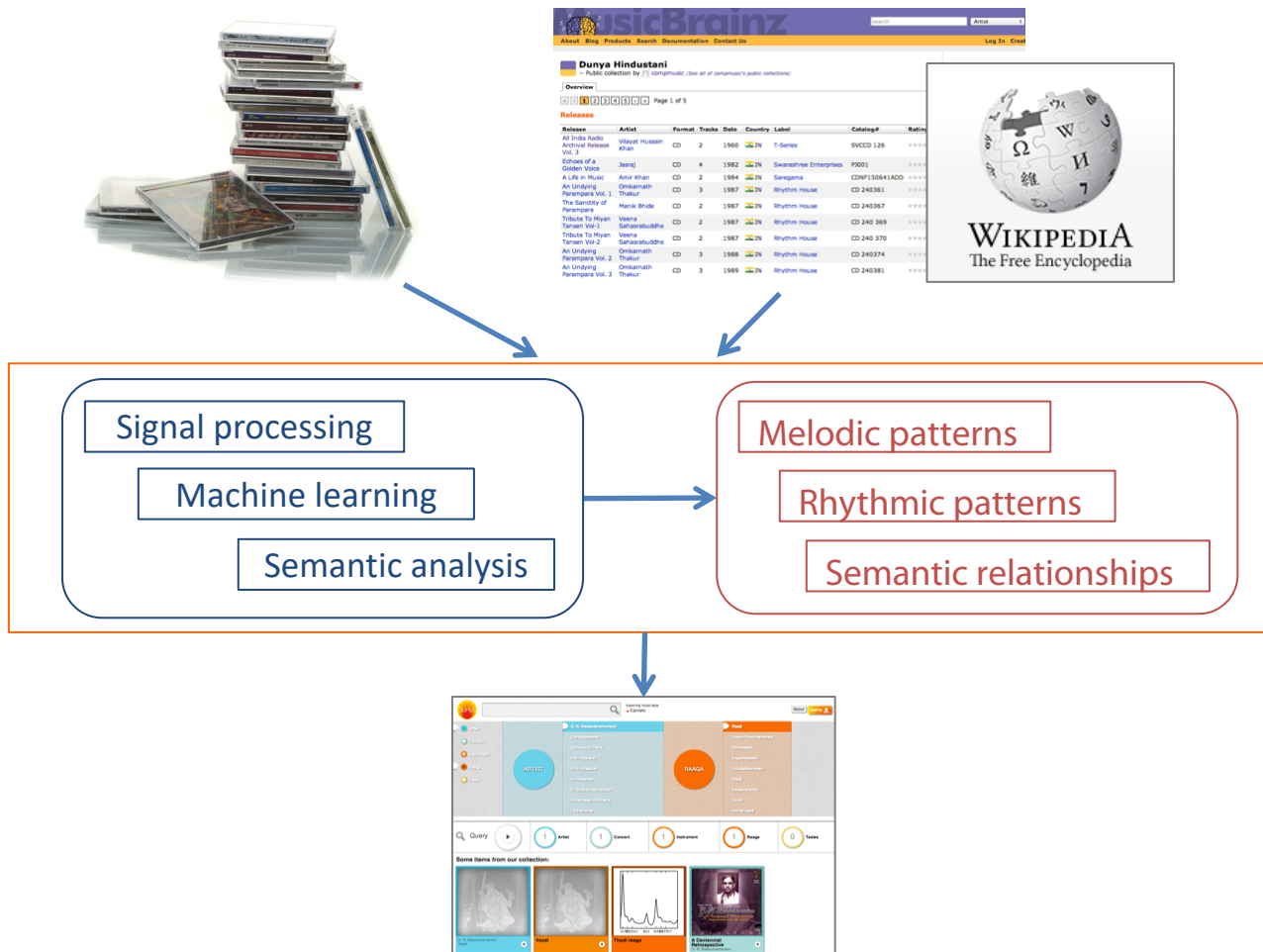
# Challenge: Corpora and Datasets (2 of 2)

- [freesound.org](https://freesound.org)
- [acousticbrainz.org](https://acousticbrainz.org)
- [dunya.upf.edu](https://dunya.upf.edu)
- ...

# Current research challenges

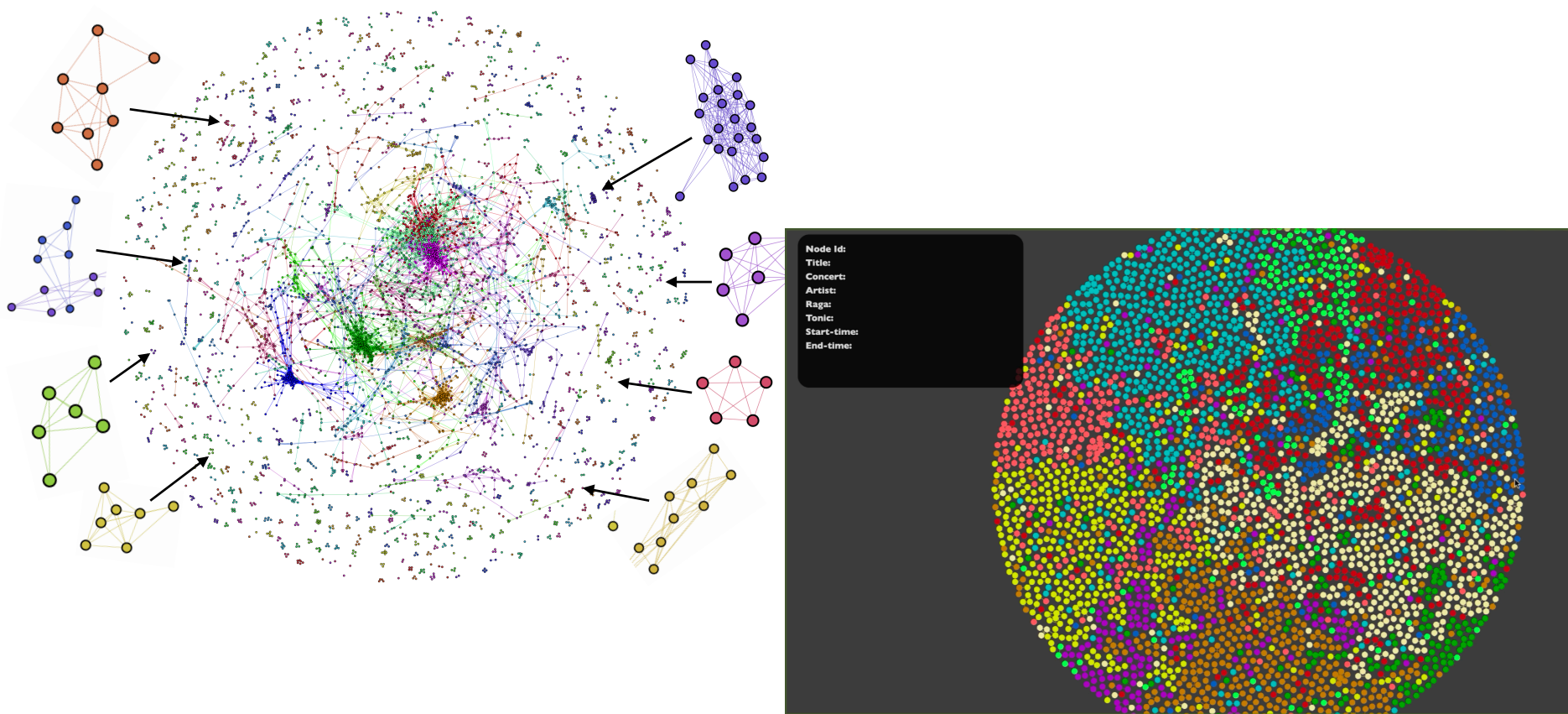
- Corpora and datasets
- **Musical understanding**
- Neural Networks for musical tasks
- Musical sound synthesis
- Separation of musical audio sources
- Sound and music classification
- Applications
- ...

# Challenge: Musical understanding (1 of 2)



X. Serra. 2017. "The computational study of a musical culture through its digital traces."  
*Acta Musicologica.*

# Challenge: Musical understanding (2 of 2)

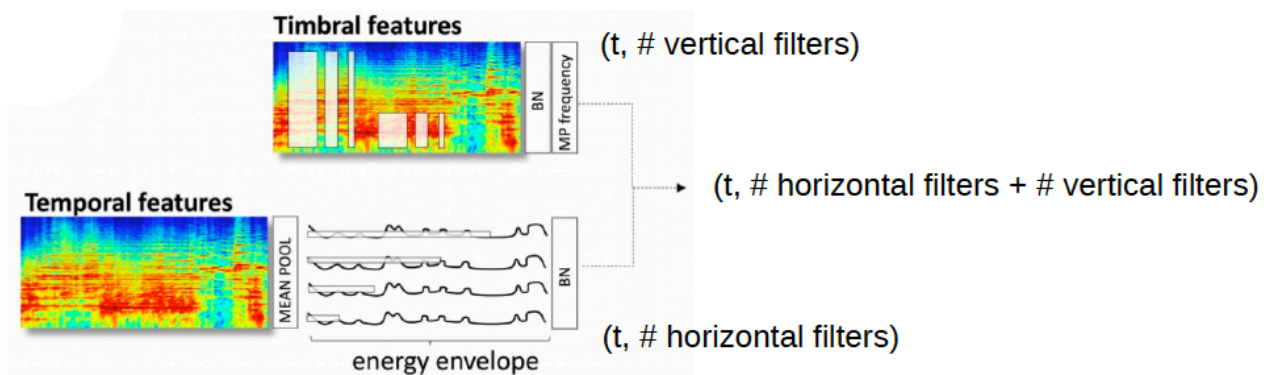


S. Gulati. 2016. *Computational Approaches for Melodic Description in Indian Art Music Corpora*.  
PhD thesis.

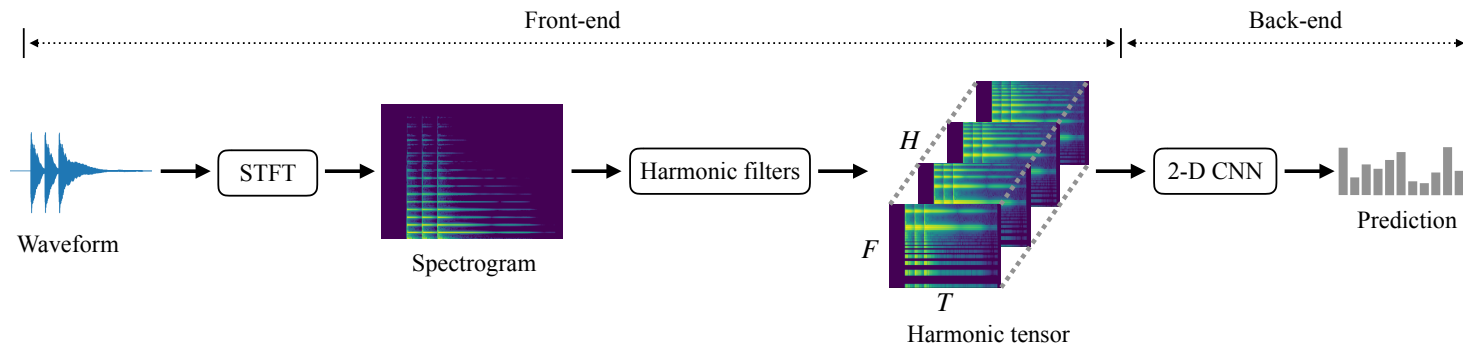
# Current research challenges

- Corpora and datasets
- Musical understanding
- **Neural Networks for musical tasks**
- Musical sound synthesis
- Separation of musical audio sources
- Sound and music classification
- Applications
- ...

# Challenge: Neural Networks for musical tasks (1 of 3)



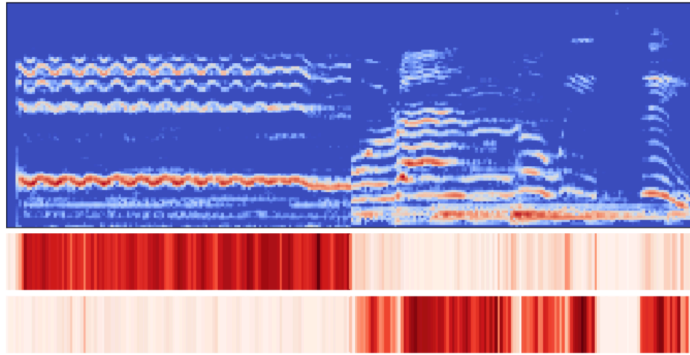
J. Pons. 2019. *Deep neural Networks for Music and Audio Tagging*. PhD Thesis.



M. Won, S. Chun, O. Nieto, X. Serra. 2020. "Data-driven Harmonic Filters for Audio Representation Learning." **ICASSP2020**.

# Challenge: Neural Networks for musical tasks (2 of 3)

Female Male

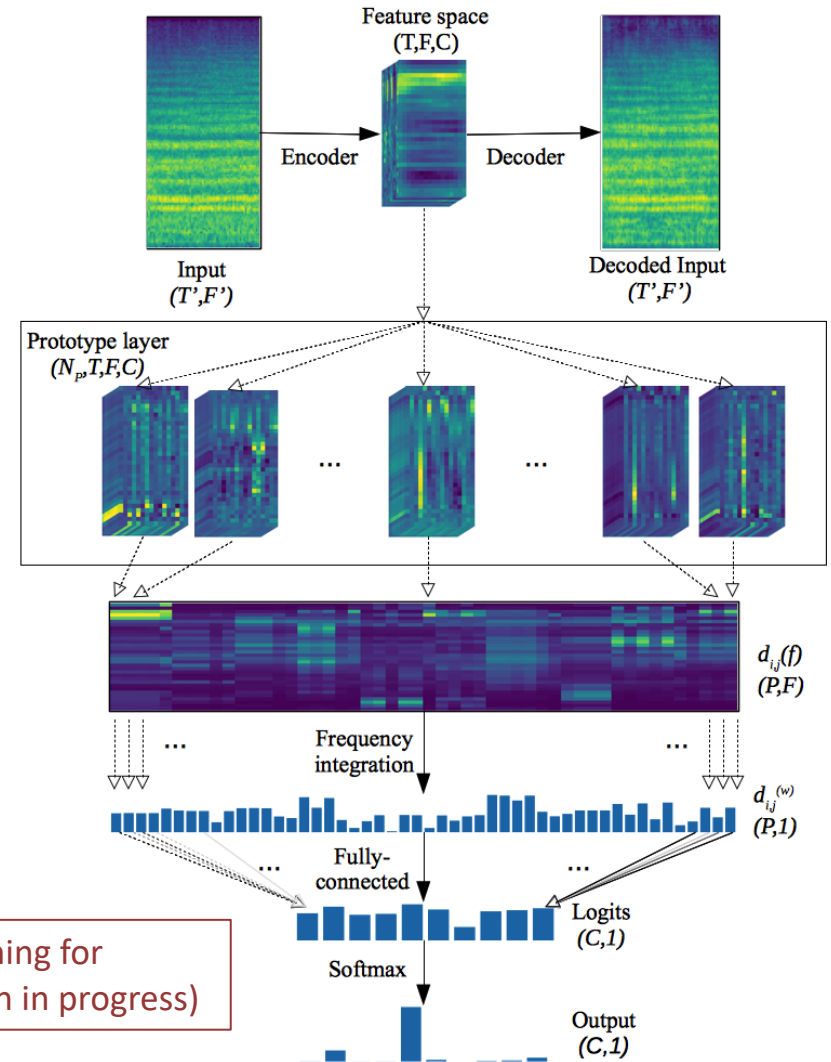


Female  
Male



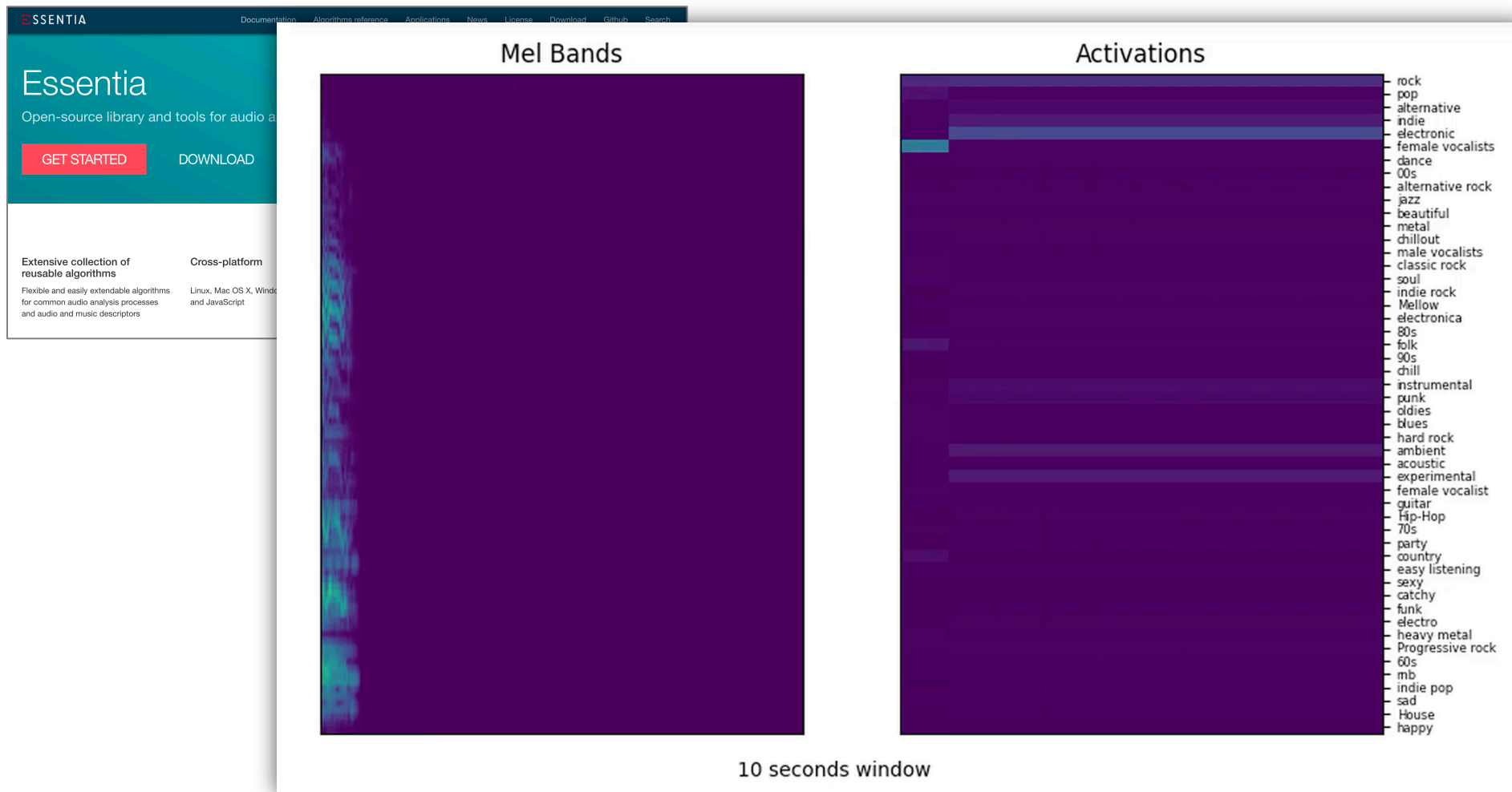
M. Won, S. Chun, X. Serra. 2019. "Toward Interpretable Music Tagging with Self-Attention." *arXiv*.

P. Zinemanas. "Explainable Deep Learning for Environmental Sound Classification." (research in progress)





# Challenge: Neural Networks for musical tasks (3 of 3)

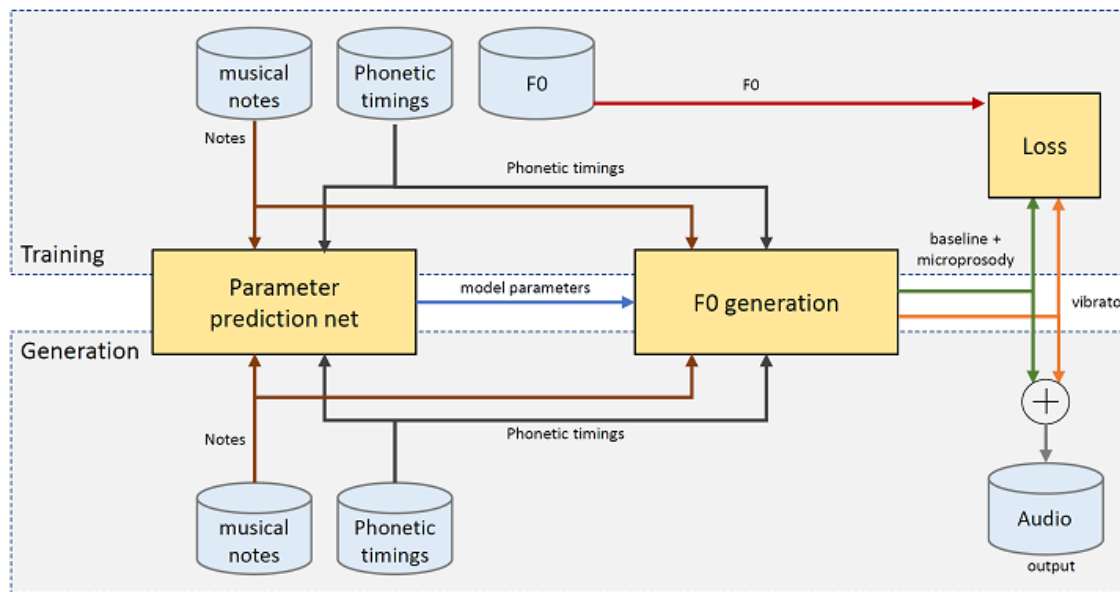


P. Alonso-Jiménez, D. Bogdanov, J. Pons, X. Serra. 2020. "Tensorflow Audio Models in Essentia." *ICASSP2020*.

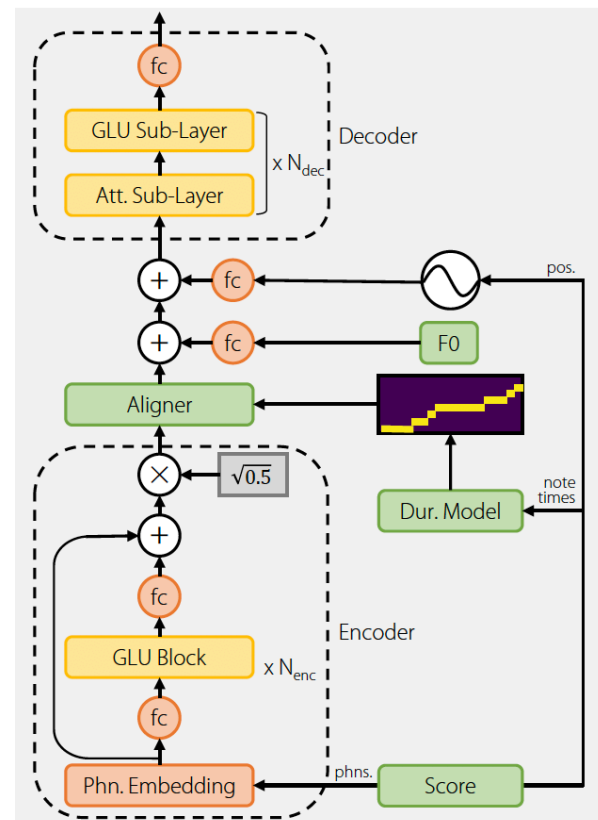
# Current research challenges

- Corpora and datasets
- Musical understanding
- Neural Networks for musical tasks
- **Musical sound synthesis**
- Separation of musical audio sources
- Sound and music classification
- Applications
- ...

# Challenge: Musical sound synthesis



**J. Bonada, M. Blaauw. 2020. "Hybrid Neural-Parametric F0 Model for Singing Synthesis." ICASSP2020.**

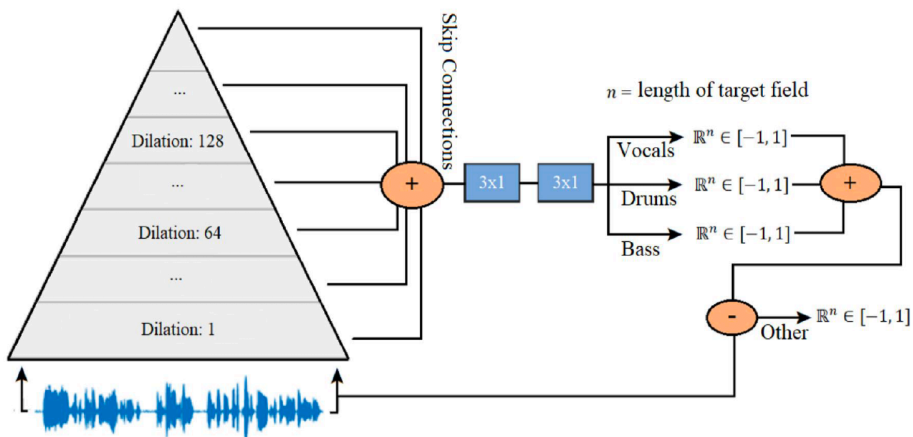


**M. Blaauw, J. Bonada. 2020. "Sequence-to-sequence Singing Synthesis Using the Feed-forward Transformer." ICASSP2020.**

# Current research challenges

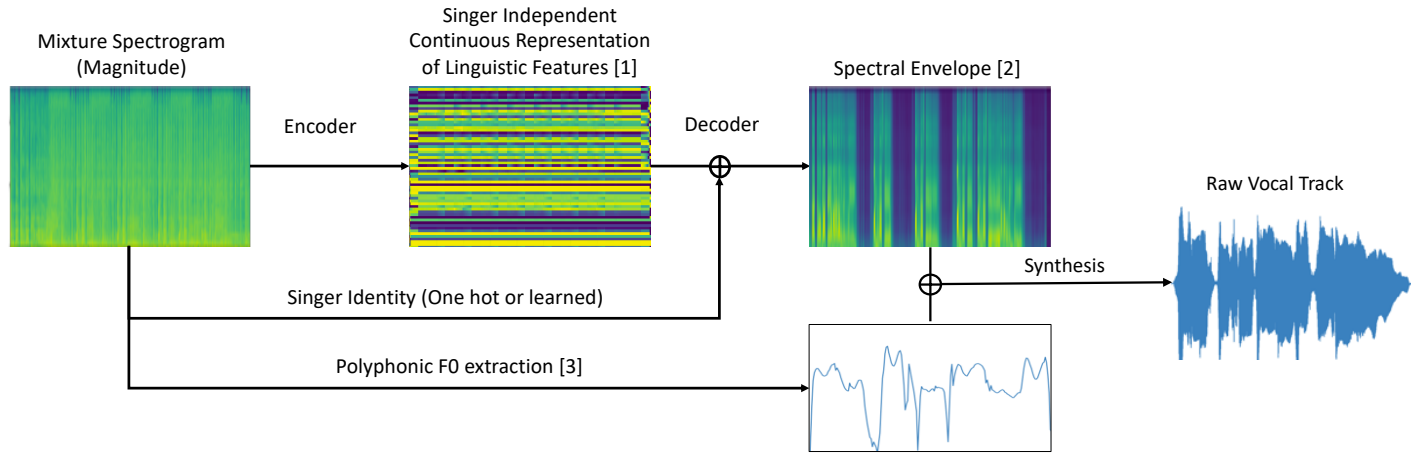
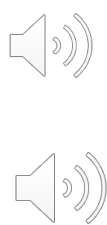
- Corpora and datasets
- Musical understanding
- Neural Networks for musical tasks
- Musical sound synthesis
- **Separation of musical audio sources**
- Sound and music classification
- Applications
- ...

# Challenge: Separation of musical audio sources



F. Lluís, J. Pons, X. Serra. 2019. "End-to-End Music Source Separation: Is it Possible in the Waveform Domain?." *Interspeech*.

Mixture Fragment  
Non-causal



P. Chandna, M. Blaauw, J. Bonada, E. Gomez. 2020. "Content Based Singing Voice Extraction From a Musical Mixture." *ICASSP2020*.

# Current research challenges

- Corpora and datasets
- Musical understanding
- Neural Networks for musical tasks
- Musical sound synthesis
- Separation of musical audio sources
- **Sound and music classification**
- Applications
- ...

# Challenge: Sound and music classification (1 of 3)

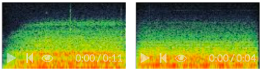
FSD → Air conditioning

Hierarchy: Sounds of things > Mechanisms > Air conditioning

Description: The sound of a device that removes heat from the air inside a building or vehicle, thus lowering the air temperature.

URI: [http://en.wikipedia.org/wiki/Air\\_conditioning](http://en.wikipedia.org/wiki/Air_conditioning)

Siblings: Mechanical fan, Clock, Cash register, Ratchet, pawl, Sewing machine, Printer, Camera, Pulleys, Gears

Examples: 

# audio samples: 131


# ground truth: 48 (propagated from children: 0) Not present

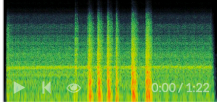
42.27% 10.65% 4.12% 42.96%

| # categories | # audio samples | # ground truth annotations | # candidate annotations | # votes |
|--------------|-----------------|----------------------------|-------------------------|---------|
| 632          | 297144          | 245692                     | 685384                  | 371690  |

Is **Engine starting** present in the following sounds?

Help ? Try other sounds ↻

#1 see in 




0:00 / 1:22

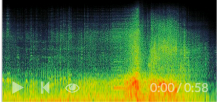
Present and predominant

Present but not predominant

Not present

Unsure

#2 see in 




0:00 / 0:58

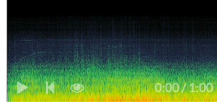
Present and predominant

Present but not predominant

Not present

Unsure

#3 see in 




0:00 / 1:00

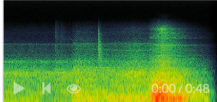
Present and predominant

Present but not predominant

Not present

Unsure

#4 see in 




0:00 / 0:48

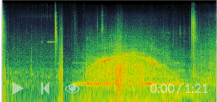
Present and predominant

Present but not predominant

Not present

Unsure

#5 see in 




0:00 / 1:21

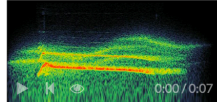
Present and predominant

Present but not predominant

Not present

Unsure

#6 see in 



0:00 / 0:07

Present and predominant

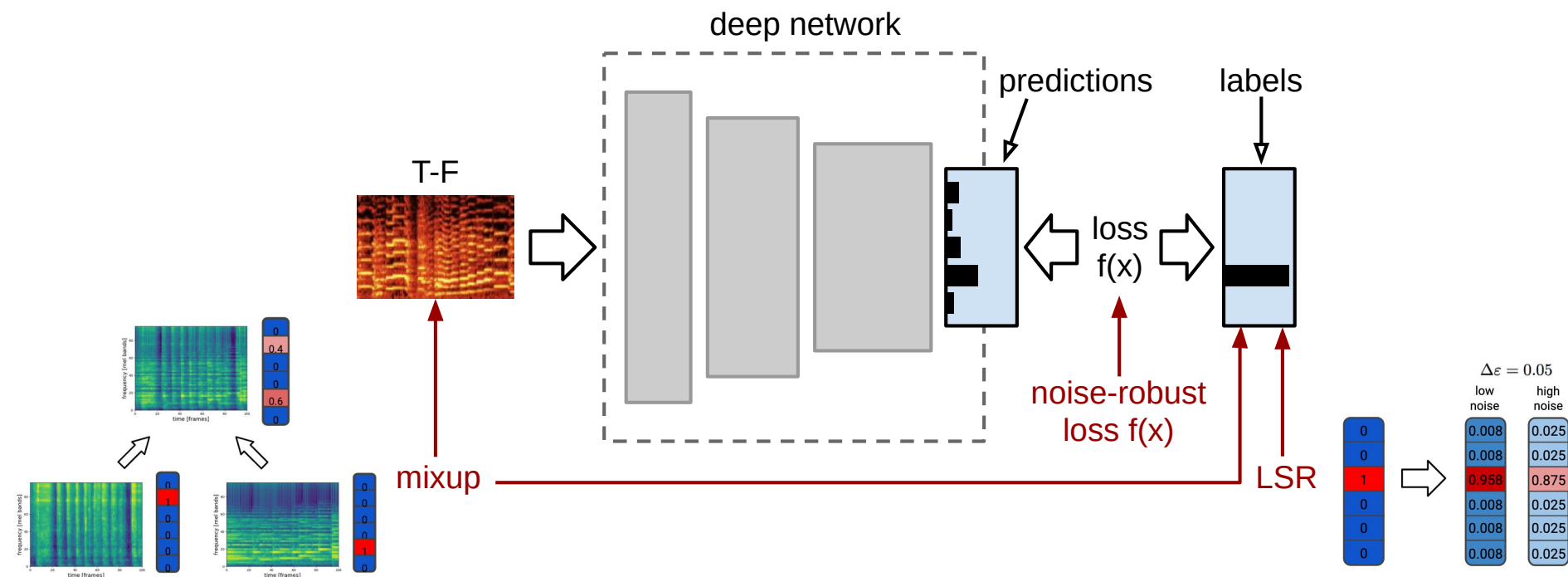
Present but not predominant

Not present

Unsure

E. Fonseca, J. Pons, X. Favory, F. Font, D. Bogdanov, A. Ferraro, S. Oramas, A. Porter, X. Serra. 2017. "Freesound Datasets: A Platform for the Creation of Open Audio Datasets." *ISMIR*.

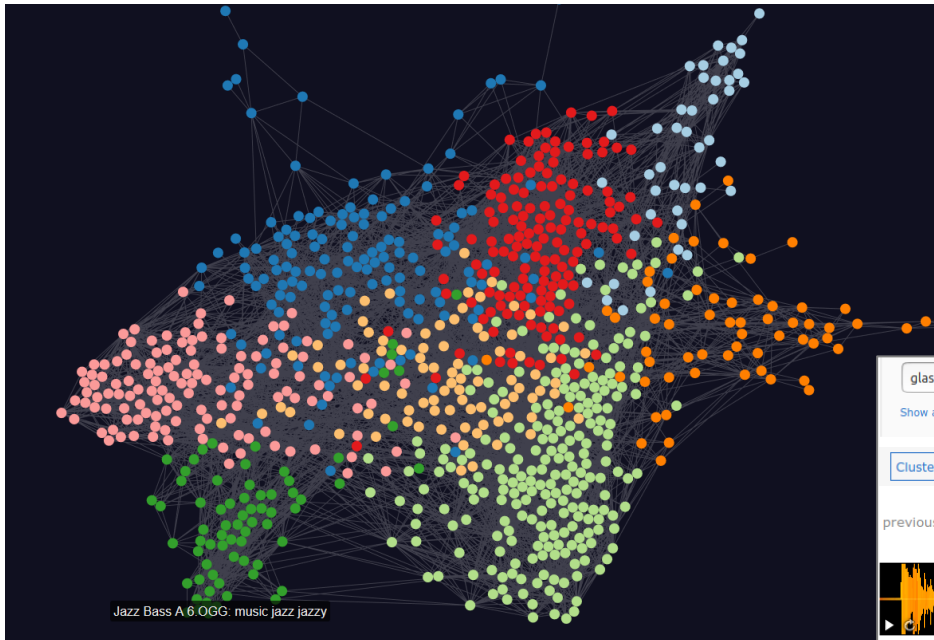
# Challenge: Sound and music classification (2 of 3)



E. Fonseca, F. Font, X. Serra. 2019. "Model-agnostic Approaches to Handling Noisy Labels When Training Sound Event Classifiers." *WASPAA*.



# Challenge: Sound and music classification (3 of 3)



glass Automatic by relevance search

Show advanced search options

Cluster #1

previous next 1 2 3 4 5 6 7 ... 15 | 217 sounds

|  |  |  |
|--|--|--|
|  | <b>Glass Smash, Bottle, C.wav</b> ★★★★★<br>Raw audio of dropping a <b>glass</b> bottle on a tiled floor. This sound is NOT in the public domain. You ...<br>Break Breaking crash bottle Glass Shatter breaking shatter Tile Smashing break glass | <b>InspectorJ</b><br>April 26th, 2016<br>3854 downloads<br>11 comments |
|  | <b>Glass Smash, Bottle, A.wav</b> ★★★★★<br>Raw audio of dropping a <b>glass</b> bottle on a tiled floor. This sound is NOT in the public domain. You ...<br>Break Breaking crash bottle Glass Shatter breaking shatter Tile Smashing break glass | <b>InspectorJ</b><br>April 26th, 2016<br>2754 downloads<br>3 comments  |
|  | <b>Glass Smash, Bottle, G.wav</b> ★★★★★<br>Raw audio of dropping a <b>glass</b> bottle on a tiled floor. This sound is NOT in the public domain. You ...<br>Break Breaking crash bottle Glass Shatter breaking shatter Tile Smashing break glass | <b>InspectorJ</b><br>April 26th, 2016<br>4354 downloads<br>8 comments  |
|  | <b>glass19.flac</b> ★★★★★<br>glass   | <b>Craxic</b><br>October 31st,   |

**clusters**

- Cluster #1 (break shatter smash, 217)
- Cluster #2 (water tone wine-glass, 280)
- Cluster #3 (break smash breaking, 62)
- Cluster #4 (broken break jar, 186)
- Cluster #5 (hit ring bowl, 242)

**licenses**

- Attribution (120)
- Attribution Noncommercial (16)
- Creative Commons 0 (80)
- Sampling+ (1)

**tags**

24-bit 48k 48khz bottle break  
breaking bris broken chest clinks  
crack crash crunch drop dropping falling field-  
recording floor foley game **glass** glasses  
gold hit percussion sample shards  
shatter smash verre

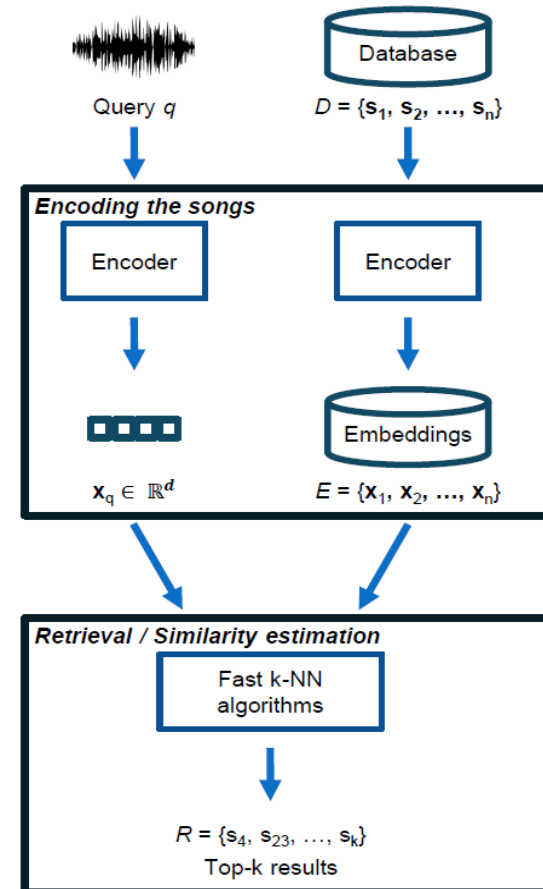
X. Favory, F. Font, X. Serra. 2020. "Search Result Clustering in Collaborative Sound Collections." ICMR.

# Current research challenges

- Corpora and datasets
- Musical understanding
- Neural Networks for musical tasks
- Musical sound synthesis
- Separation of musical audio sources
- Sound and music classification
- **Applications**
- ...

# Challenge: Applications

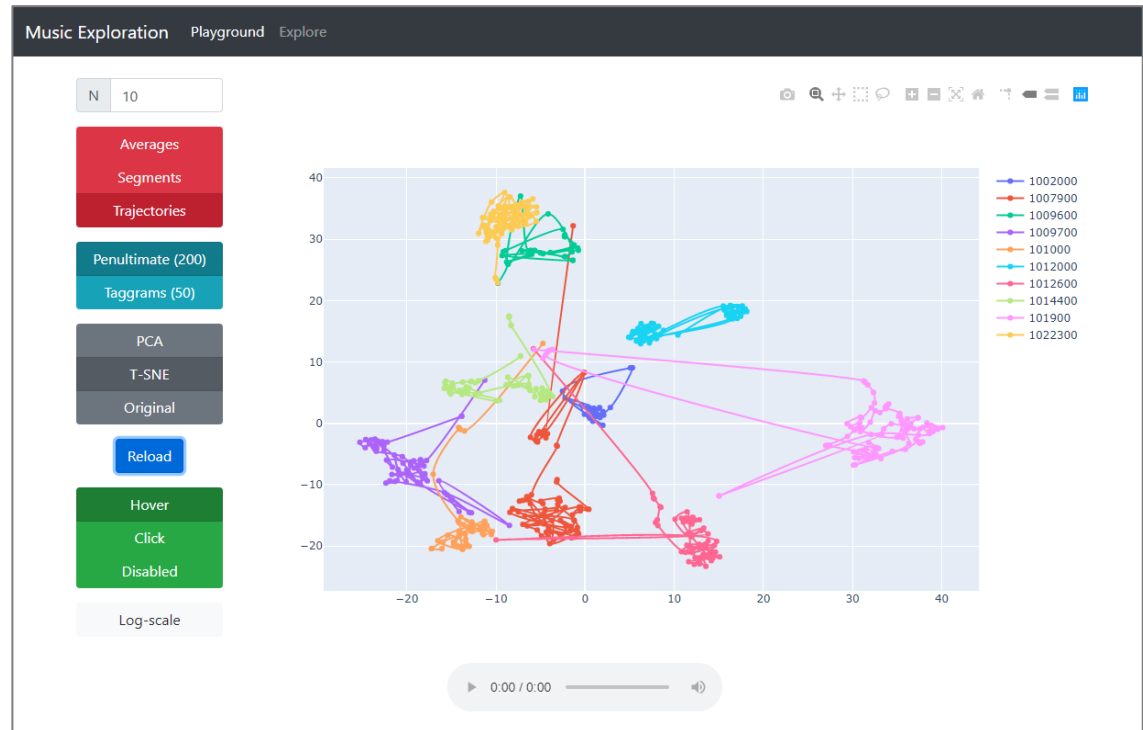
- Music identification



**F. Yesiler, J. Serrà, E. Gómez. 2020.**  
“Accurate and Scalable Version  
Identification Using Musically-Motivated  
Embeddings.” **ICASSP2020.**

# Challenge: Applications

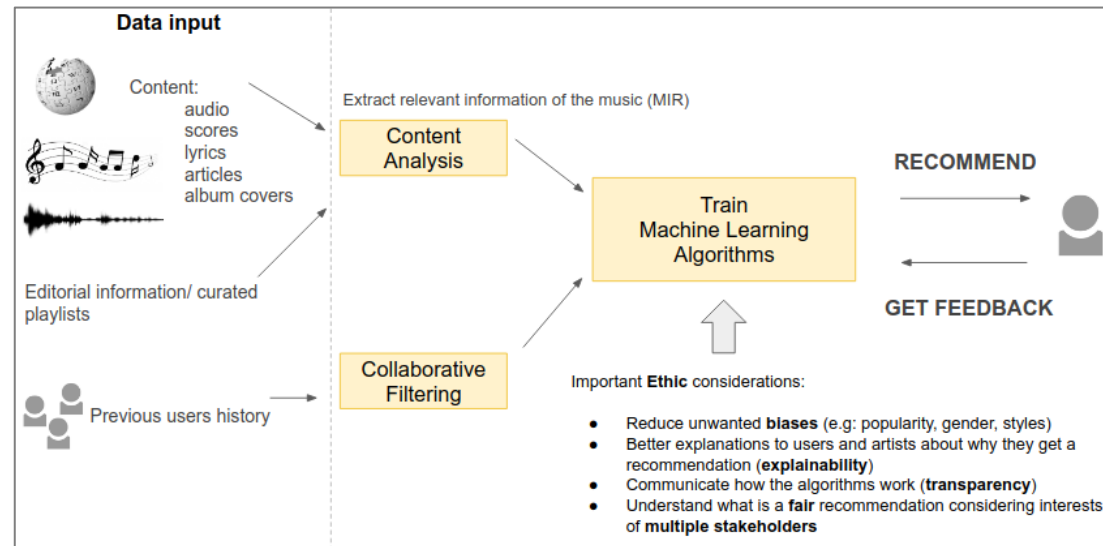
- Music identification
- Music exploration



**P. Tovstogan.** "Facilitating Interactive Music Exploration." (research in progress)

# Challenge: Applications

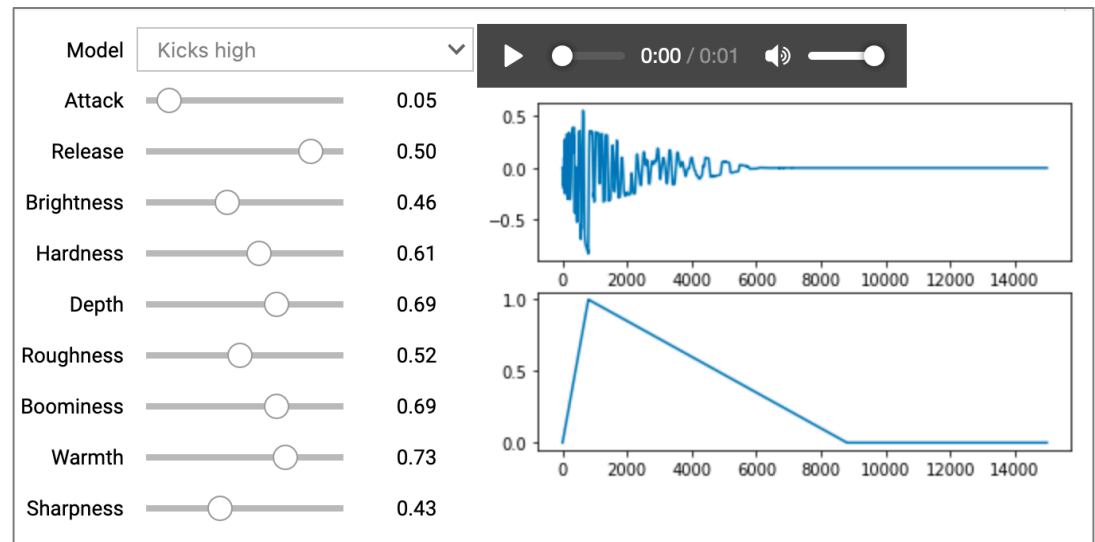
- Music identification
- Music exploration
- **Music recommendation**



**A. Ferraro, D. Bogdanov, X. Serra, J. Yoon. 2019. "Artists and Style Exposure Bias in Collaborative Filtering Based Music Recommendation." wsHCMIR.**

# Challenge: Applications

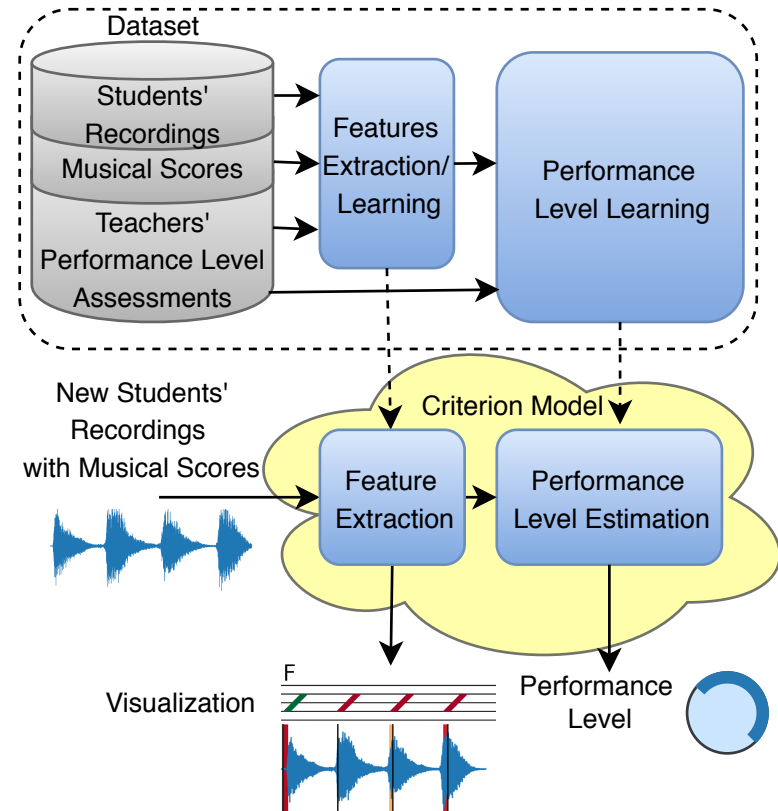
- Music identification
- Music exploration
- Music recommendation
- **Music creation**



**A. Ramires, P. Chandna, X. Favory, E. Gómez, X. Serra. 2020. "Neural Percussive Synthesis Parameterised by High-Level Timbral Features." ICASSP2020.**

# Challenge: Applications

- Music identification
- Music exploration
- Music recommendation
- Music creation
- **Music education**



V. Eremenko, A. Morsi, J. Narang, X. Serra. 2020. "Performance Assessment Technologies for the Support of Musical Instrument Learning." *CSME*.

# Challenge: Applications

- Music identification
- Music exploration
- Music recommendation
- Music creation
- Music education
- ...



# Conclusions

- Many things left out
- Specificity of sound and music research and applications
- Multidisciplinarity
- Publications:
  - <https://www.upf.edu/web/mtg/research/publications>
- Software & Datasets:
  - <https://www.upf.edu/web/mtg/software-datasets>

Thanks !!

xavier.serra@upf.edu