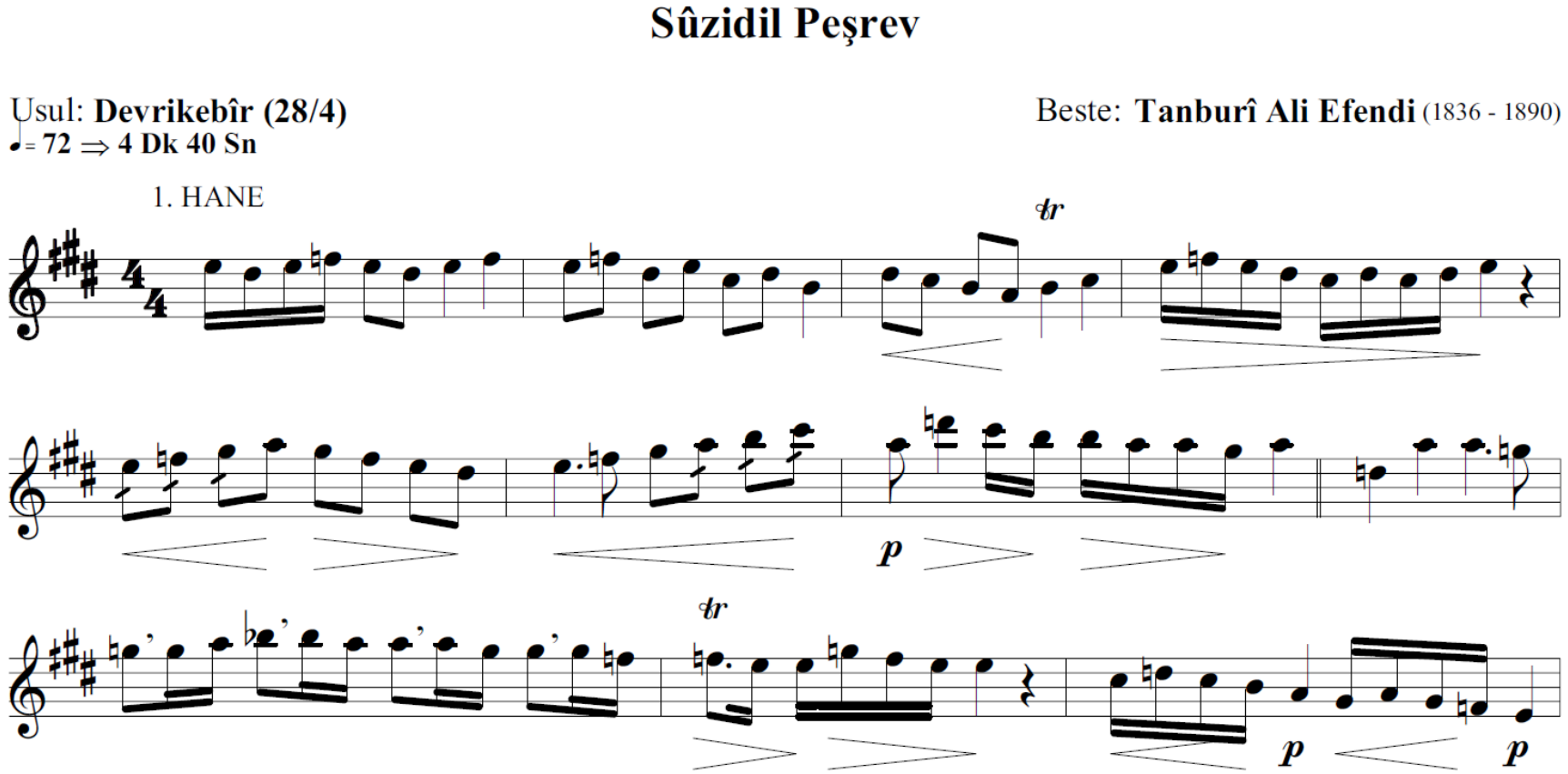
Outline

- Turkish Makam Music
- Research Corpus
- Test Datasets



Turkish Makam Music

- *Makam* ~ melodic structure

  Makams are modal structures

  Melodies typically revolve around an initial tone and a final tone.

  The final tone is synonymous to *tonic*.

  500+ *makams*, ~150 of them are still alive

- *Usul* ~ rhythmic structure

  A group of strokes with different velocities, e.g. the beats and downbeats

- *Form*

  Instrumental, vocal compositions, religious compositions...

<u>Tuning</u>

Arel-Ezgi-Uzdilek is the mainstream theory

Uses 24 tones from 53-TET system

Corpus & Dataset

CompMusic

- Mainly focuses on *melodic* and *rhythmic* characteristics

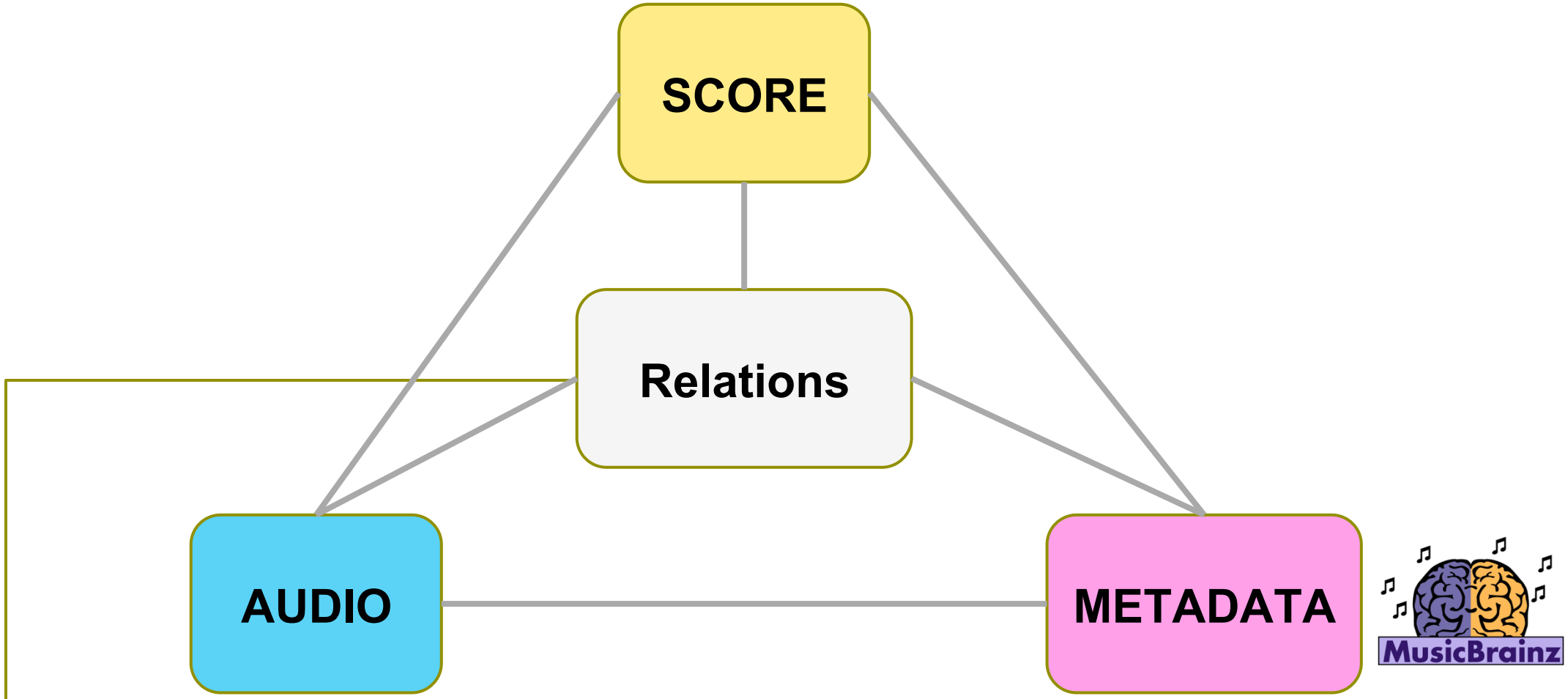
Corpus

- Collection of music documents, representing a specific tradition
- Audio recordings, music scores, lyrics and editorial metadata...

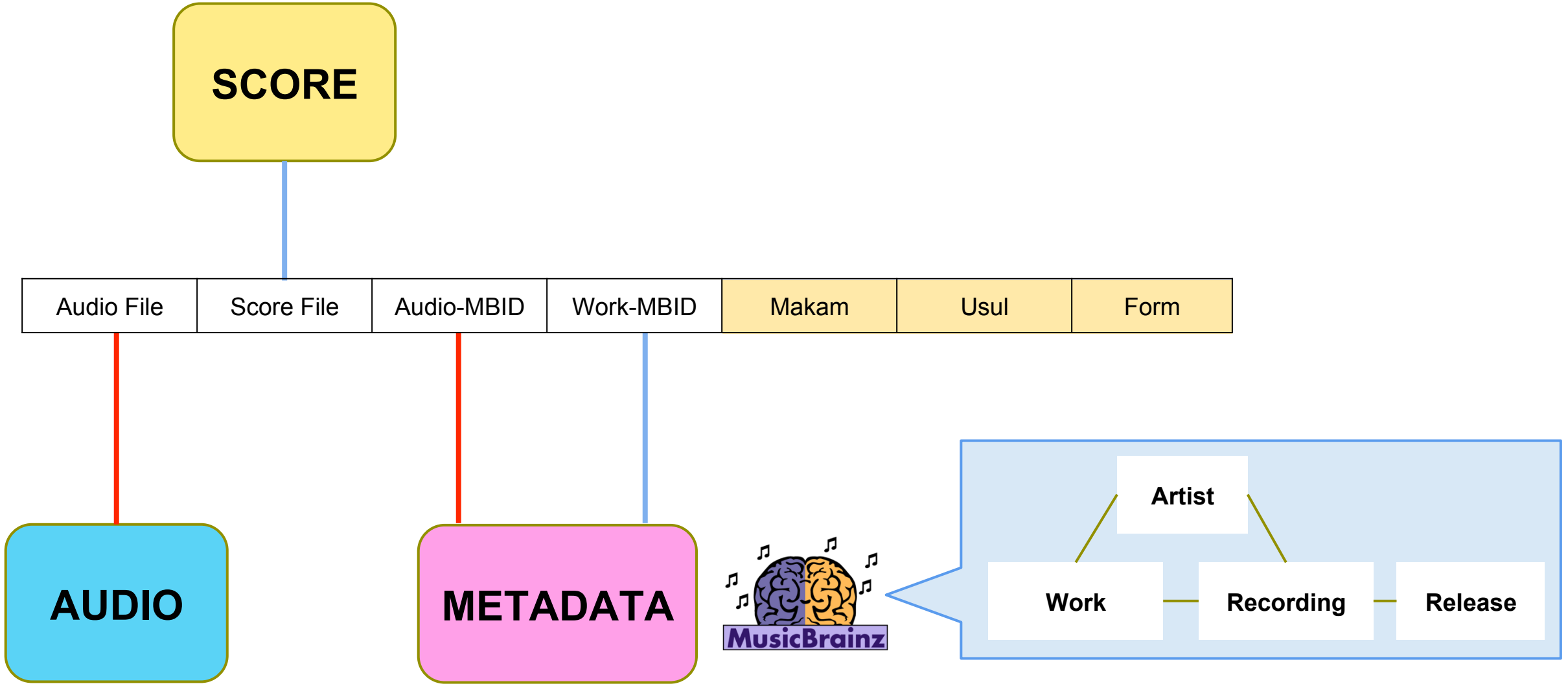
Test Dataset

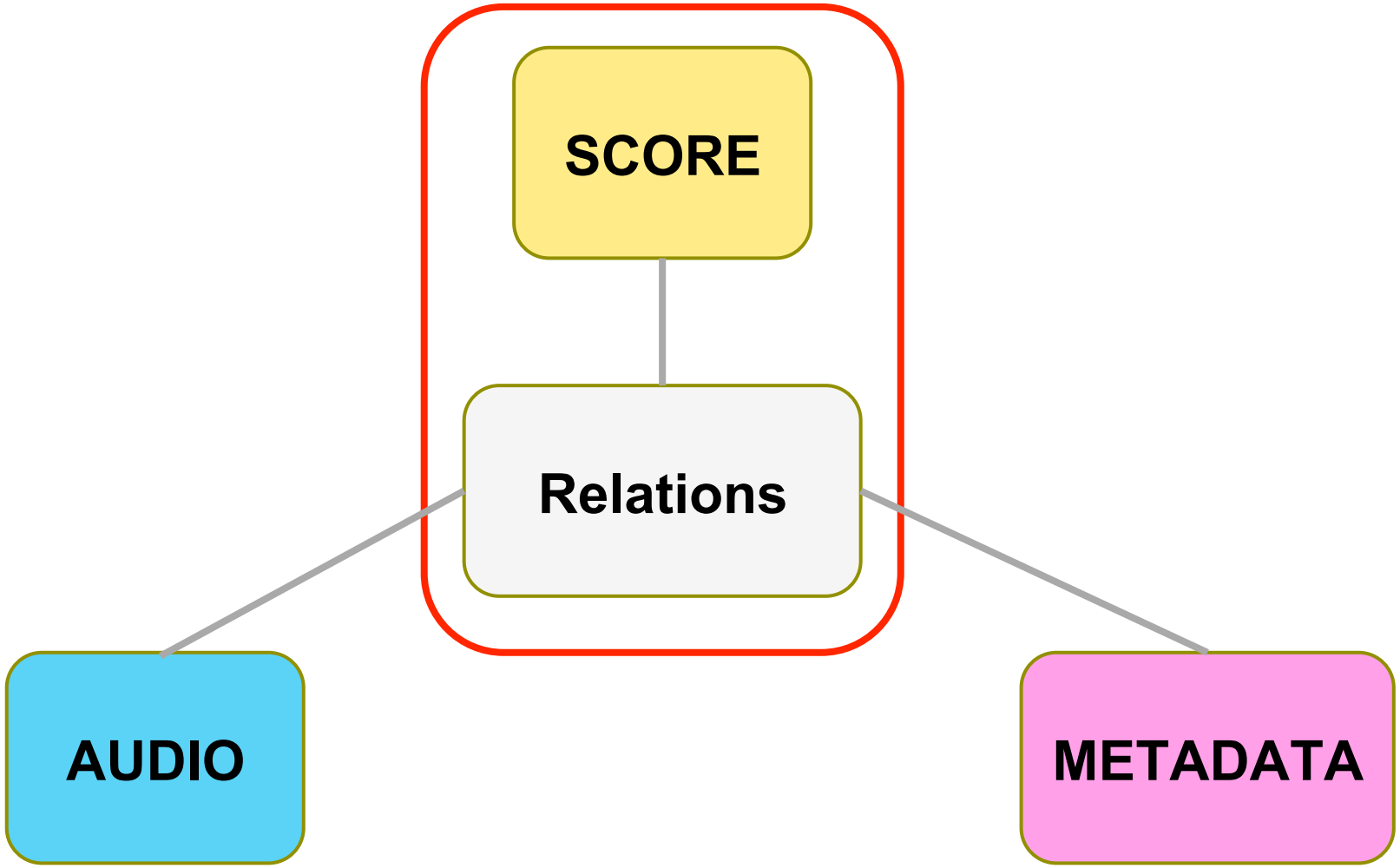
- Collection for a specific research task
- To *test*, *calibrate* and *evaluate* methodologies
- Can be a subset of *corpus* or include data which is not in *corpus*

Research Corpus



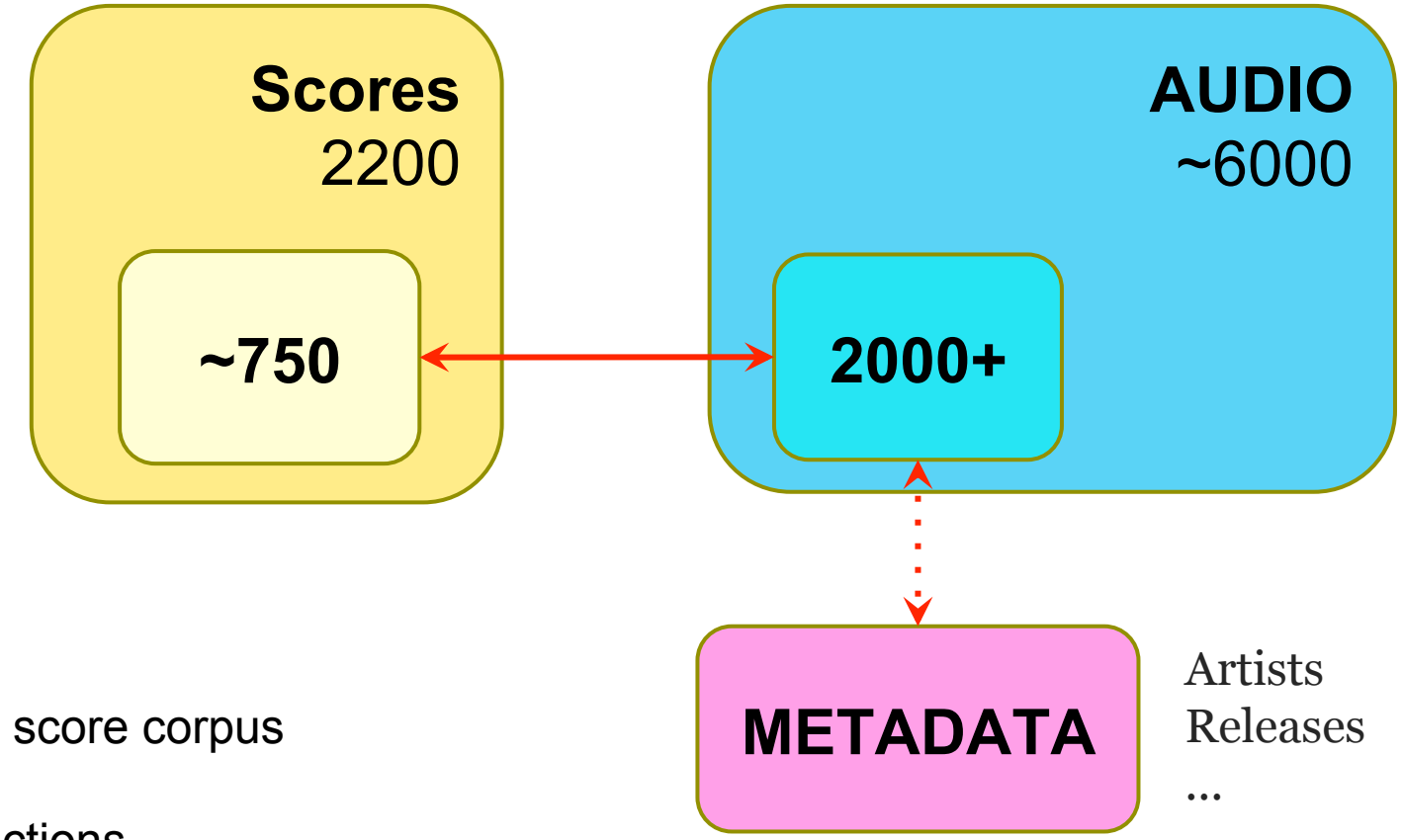
Research Corpus




SCORE
Relations
AUDIO
METADATA

Score Corpus

	#
scores	2200
makams	150+
usuls	80+
forms	60+

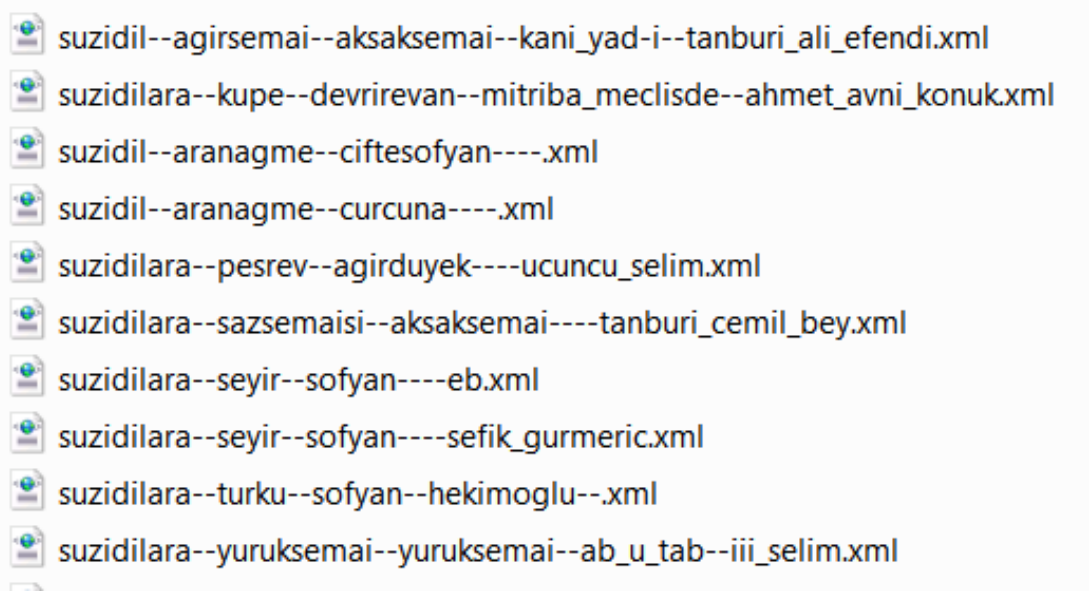


- The biggest machine-readable score corpus for Turkish makam music.
- Covers the existing score collections

Score Corpus

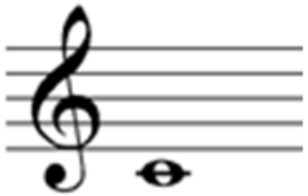
- Conventional naming
- Notes in 53-TET
- Lyrics
- Nominal tempo
- Section
- Measure
- Machine Readable -> MusicXML

makam	form	usul	title	composer



Score Corpus

What is MusicXML?



```
<note>
  <pitch>
    <step>C</step>
    <octave>4</octave>
  </pitch>
  <type>whole</type>
</note>
```

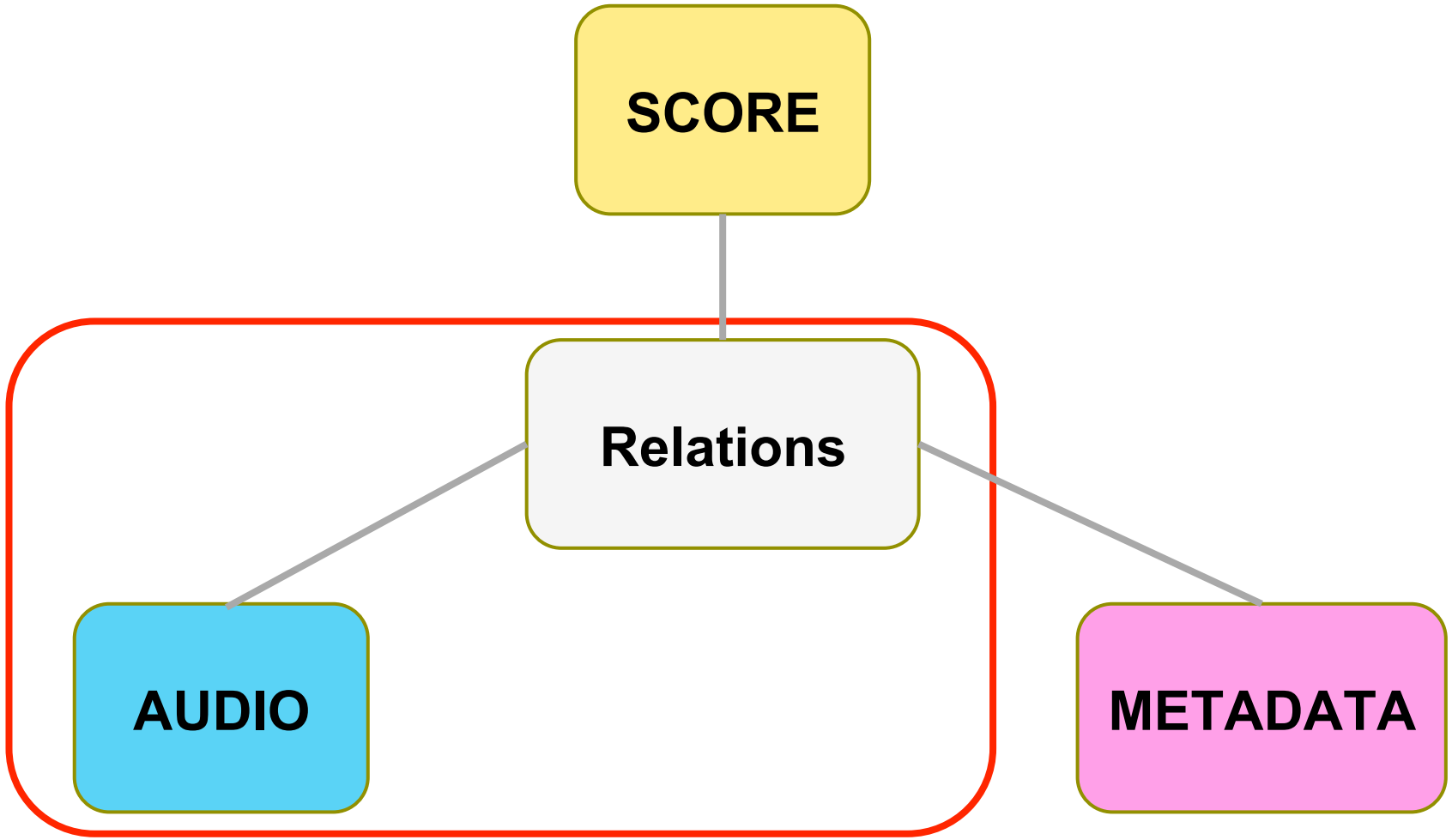
Advantages of MusicXML in our case

- Supports makam music symbols
- Compatible with well-known software: MuseScore, Finale, Sibelius
- Abstract sheets for users
- Ability to import and modify
- Easy to share new sheet-music
- Good for community-contribution
- Appropriate for end-users and researchers

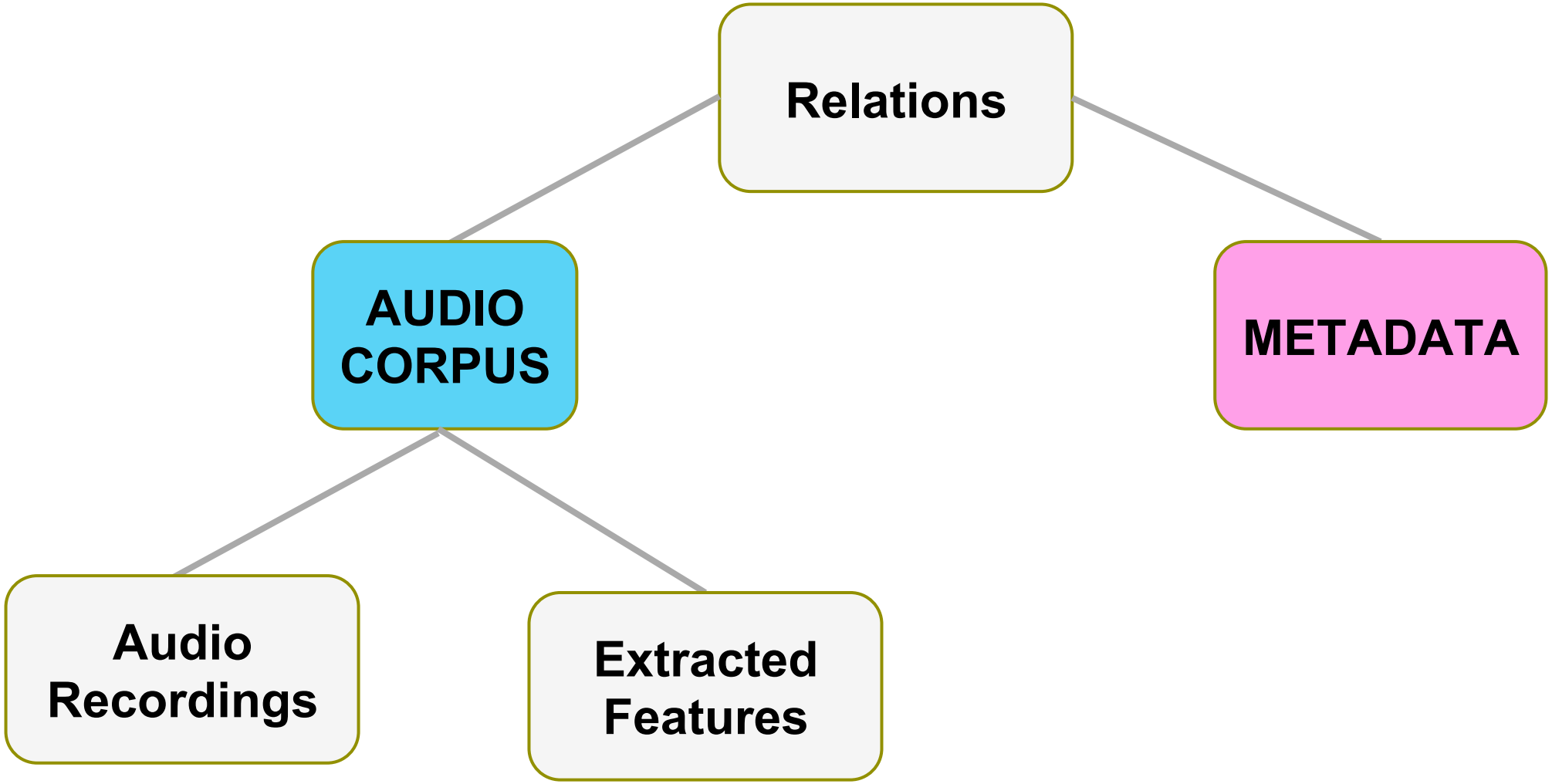
Score Corpus

Publicly available on:

https://github.com/MTG/SymbTr/releases/tag/v2.0.0

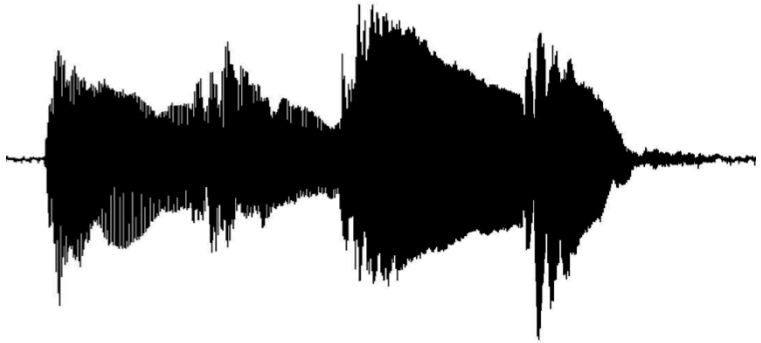

SCORE
Relations
AUDIO
METADATA

Audio Corpus



Audio Corpus

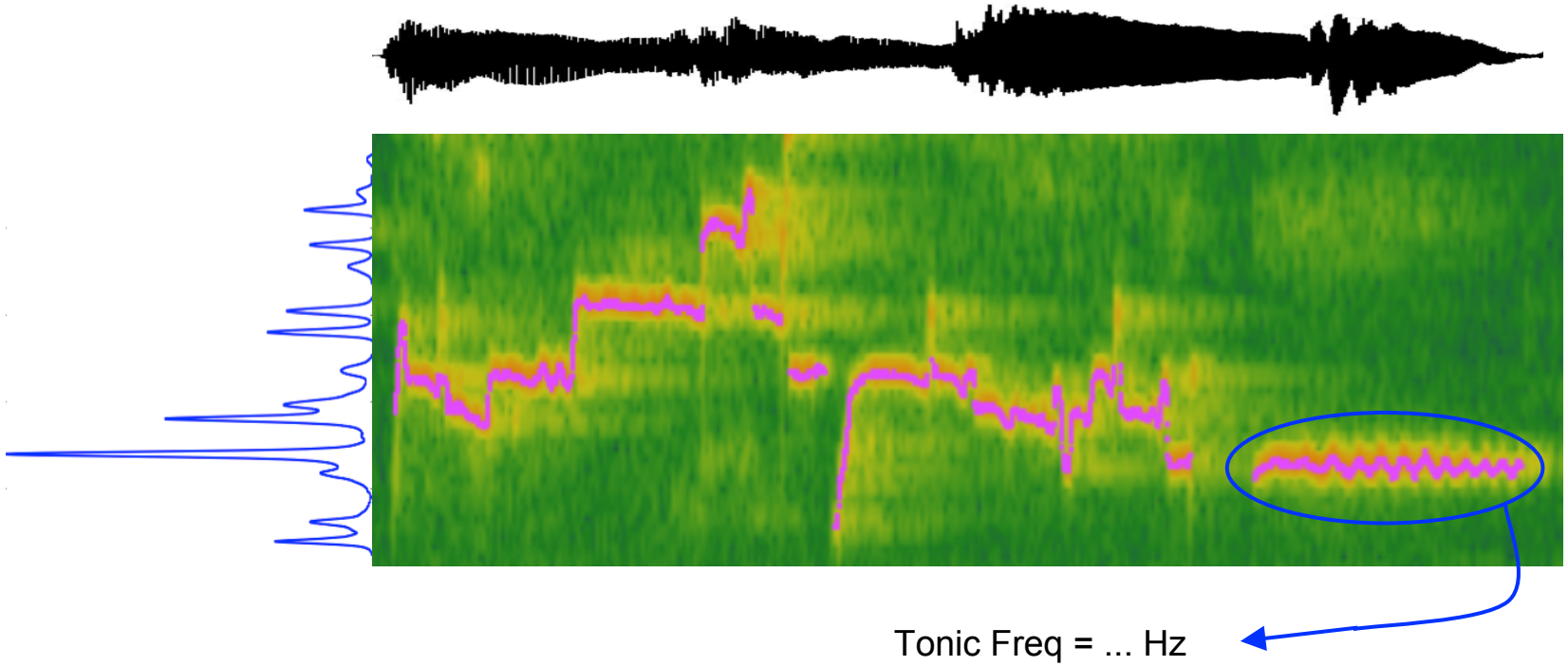
- Audio Recordings
  - ➢ 5953 Audio Recordings ~ 375 hours of play time
  - ➢ Commercial & Non-commercial
  - ➢ Solo & Ensemble recordings
  - ➢ Classical, Folk & Religious repertoire



Audio Corpus

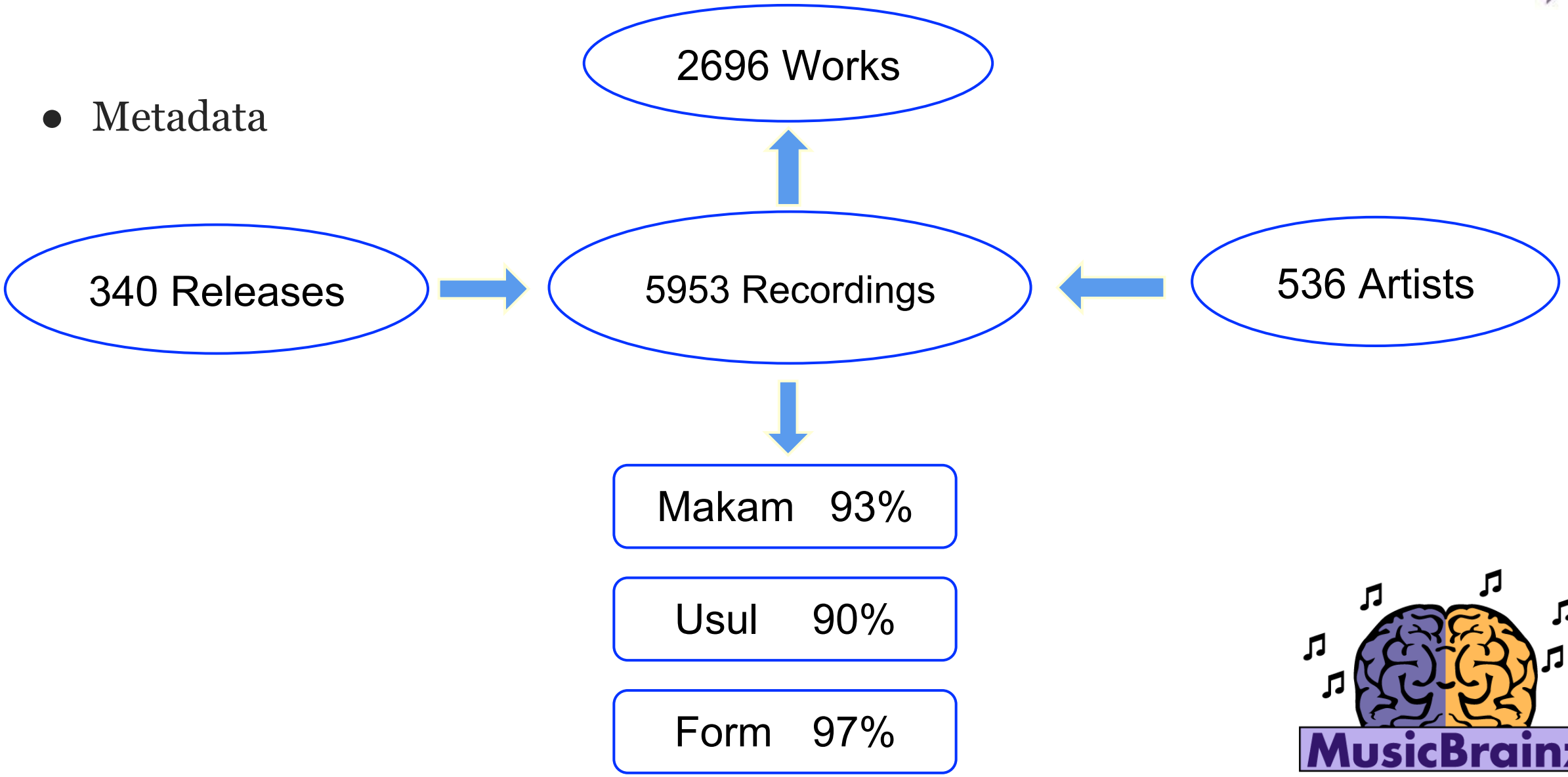
- Audio Features

  - ➢ Predominant Melody
  - ➢ Pitch Histogram
  - ➢ Tonic Frequency



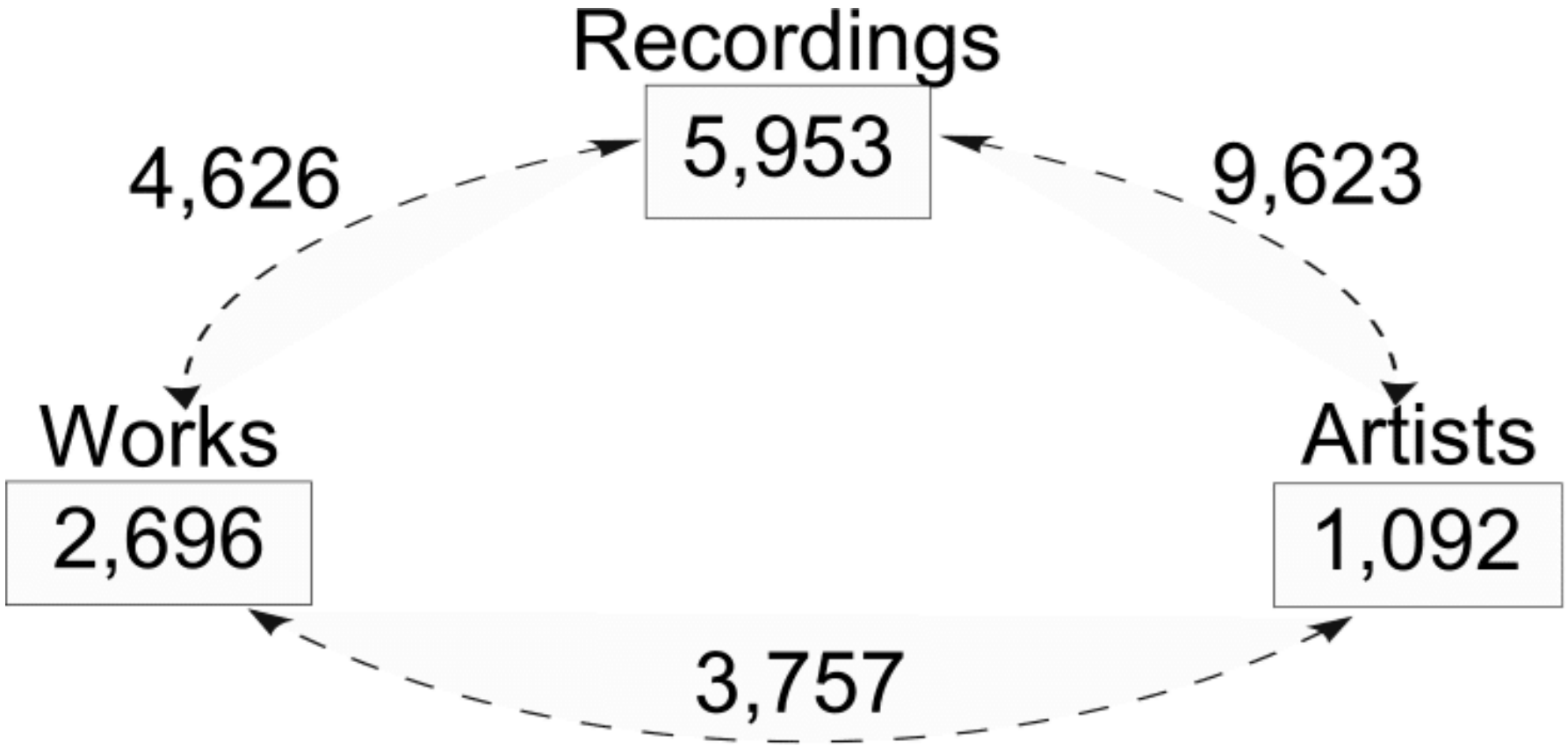
Audio Corpus

- Metadata



Audio Corpus

- Metadata



Test Datasets

- Melodic and Symbolic Phrase Test Dataset
- Tonic Test Dataset
- Section Test Dataset
- Audio - Score Alignment Test Dataset
- Audio - Lyrics Alignment Test Dataset

http://compmusic.upf.edu/datasets

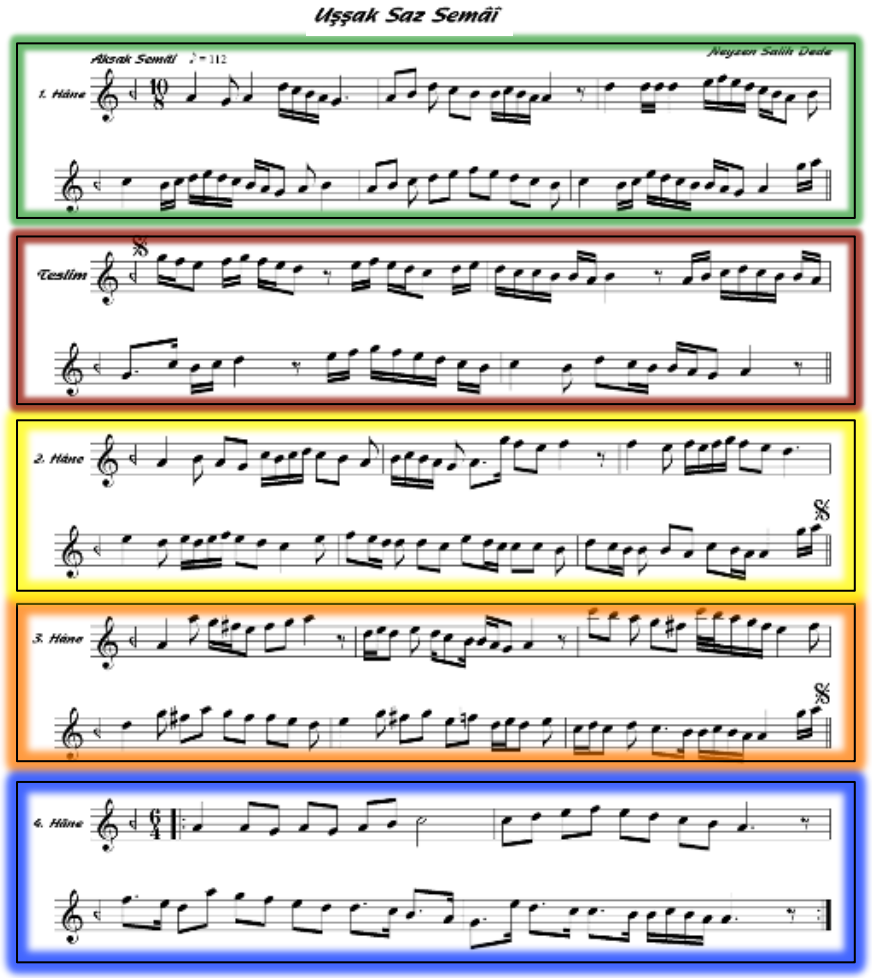
Test Datasets

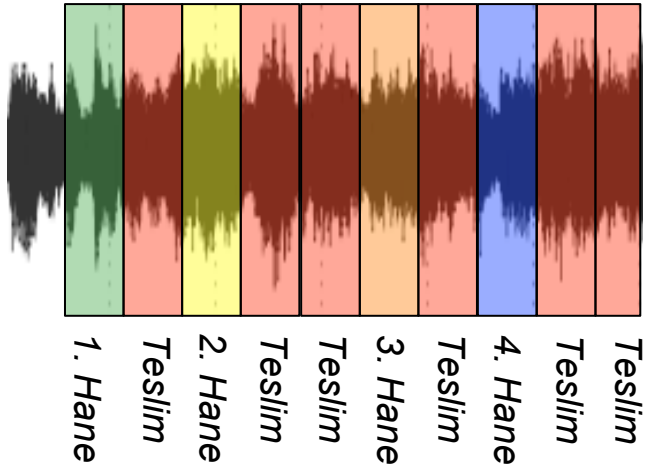
- Melodic and Symbolic Phrase Test Dataset
  - 3 experts
  - 31362 manual annotation , 899 scores

Test Datasets

Section Test Dataset

- 2095 sections annotated in 257 audio recording of 58 compositions.

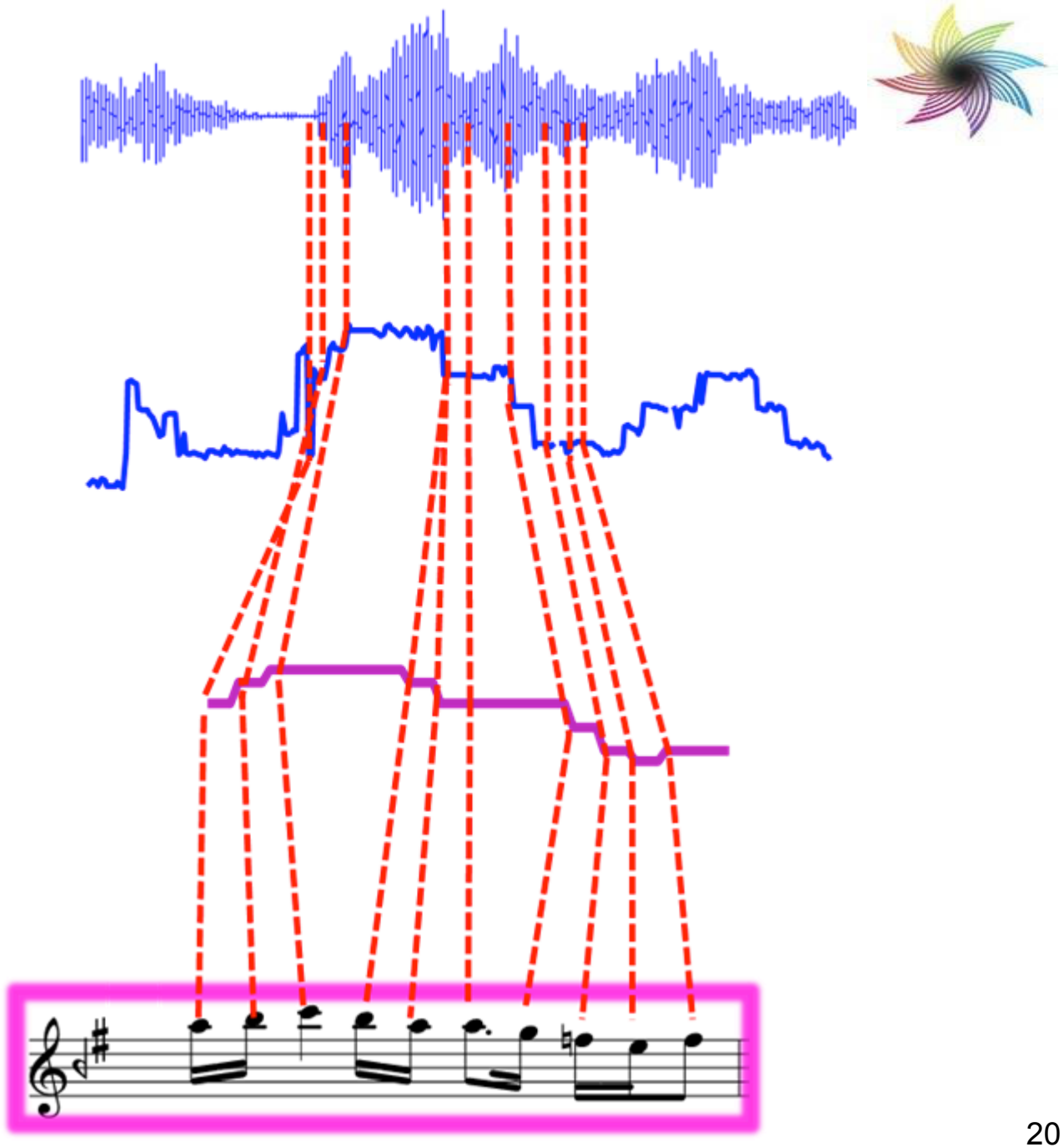




Test Datasets

Audio - Score Alignment Test Dataset

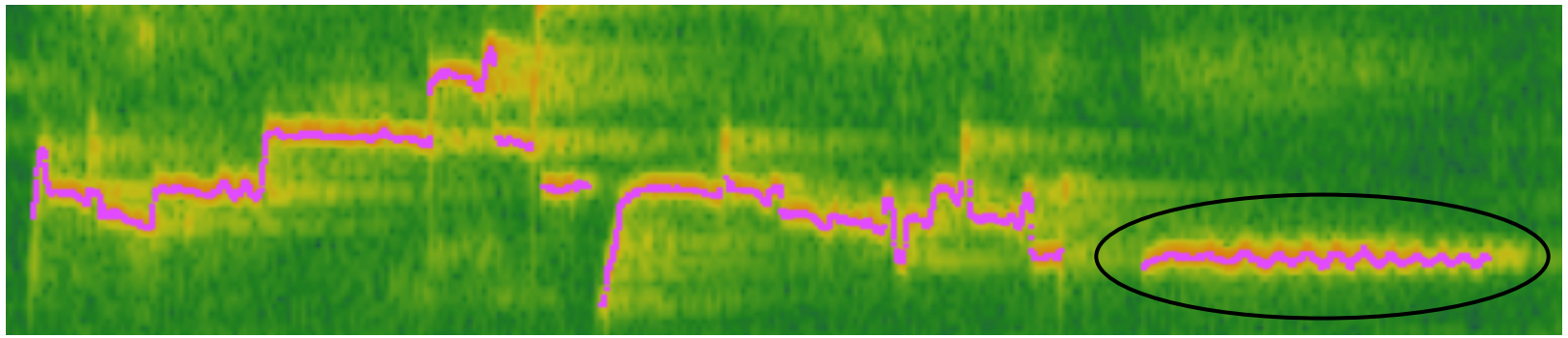
- 6 audio recordings of 4 compositions
- 51 sections in total
- 3896 note annotation



Test Datasets

Tonic Test Dataset

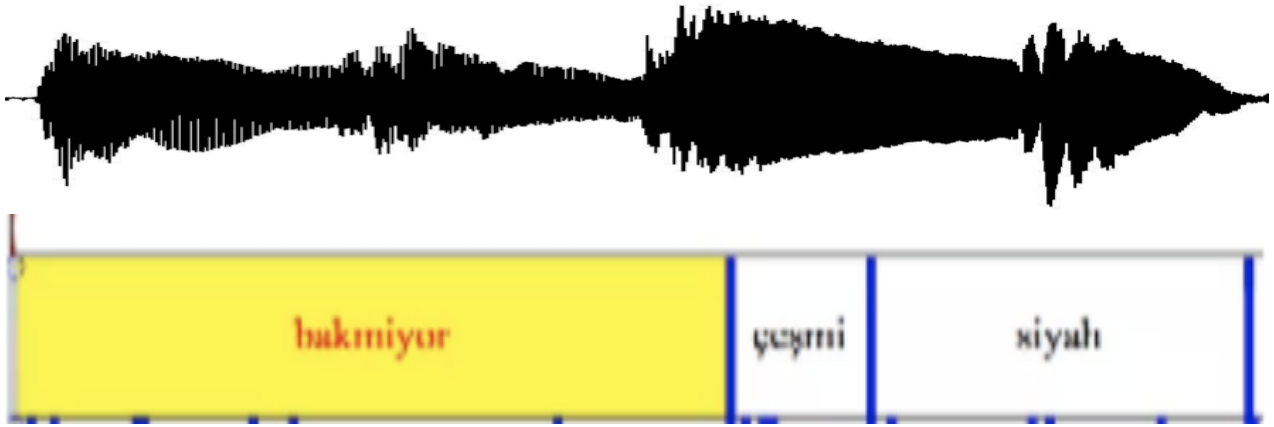
- Tonic frequency of 257 audio recordings associated with 57 compositions
- Tonic frequency of 3384 audio recordings manually annotated.



Test Datasets

Audio - Lyrics Alignment Test Dataset

- Lyrics of 10 songs manually divided and aligned.



Conclusion

- Music Information Retrieval and Computational Musicology

- Possible Research Task,
  - ➢ Structure Analysis
  - ➢ Automatic Tonic Identification
  - ➢ Automatic Ornamentation Segmentation
  - ➢ Melodic Phrase Segmentation

- We are open to contributions!

THANK YOU