Kara1k: A karaoke dataset for cover song identification and singing voice analysis

Yann Bayle[1,2], **Ladislav Maršík**[4], Martin Rusek[3]
Matthias Robine[1,2], Pierre Hanna[1,2], Kateřina Slaninová[3], Jan Martinovič[3] and Jaroslav Pokorný[4]

1: Univ. Bordeaux, LaBRI, UMR 5800, F-33400 Talence, France
2: CNRS, LaBRI, UMR 5800, F-33400 Talence, France
3: IT4Innovations, VŠB - Technical University of Ostrava, Ostrava, Czech Republic
4: Dept. of Software Engineering, Charles University, Prague, Czech Republic

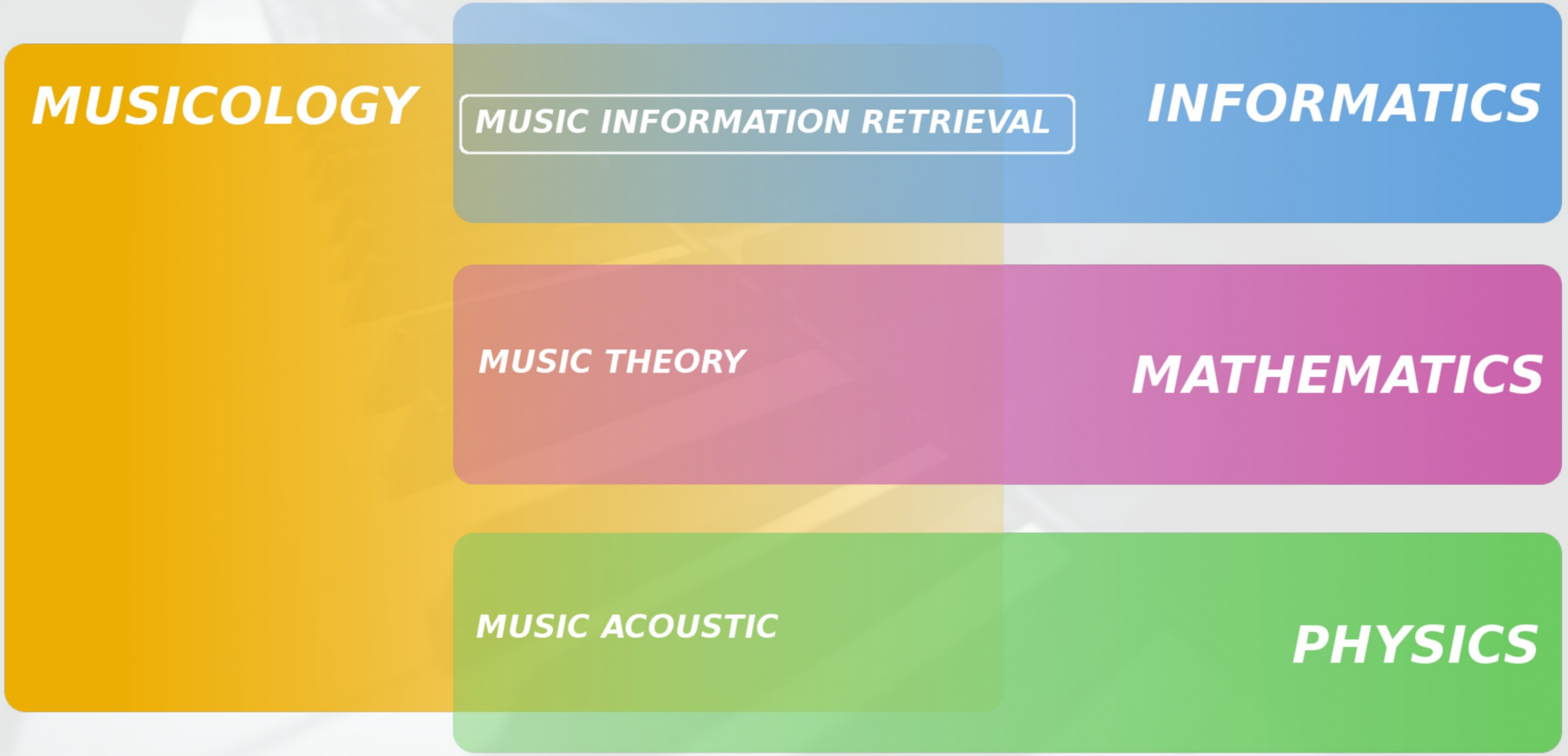
IT4Innovations
national
supercomputing
center

ISM 2017, Taichung, Taiwan

Outline

1) Introducing Kara1k, a Karaoke dataset for music research
2) Long-term project for creating new tasks and improving state-of-the-art
3) Experiments with the Kara1k dataset showing new results

Music Information Retrieval



Music Information Retrieval

Focus No.1: Cover Song Identification

- Cover song
  - An alternative version, performance, or recording of a previously published musical piece
- Input: Cover song (audio file), database of original songs;
- Output: Located original song

Queen

Queen + George Michael

Focus No.2 – Singing Voice Analysis

- Broad set of tasks: Singer Gender Classification (SGC), Singing Voice Detection, Singing Voice Separation, ...
- SGC: Input audio file, Output: gender of singer(s)

Male

Mixed

Female

Males

Females

Motivation

- For both areas, the future research has major problems
  - Stalled state-of-the-art (CSI: 2009 Serra et al.)
  - It is **NOT** a problem of techniques, rather of datasets
    - Not enough datasets to foster the research
    - Datasets do not have sufficient metadata or features
      - Datasets with > 1000 songs

        => not enough features (MillionSongDataset)

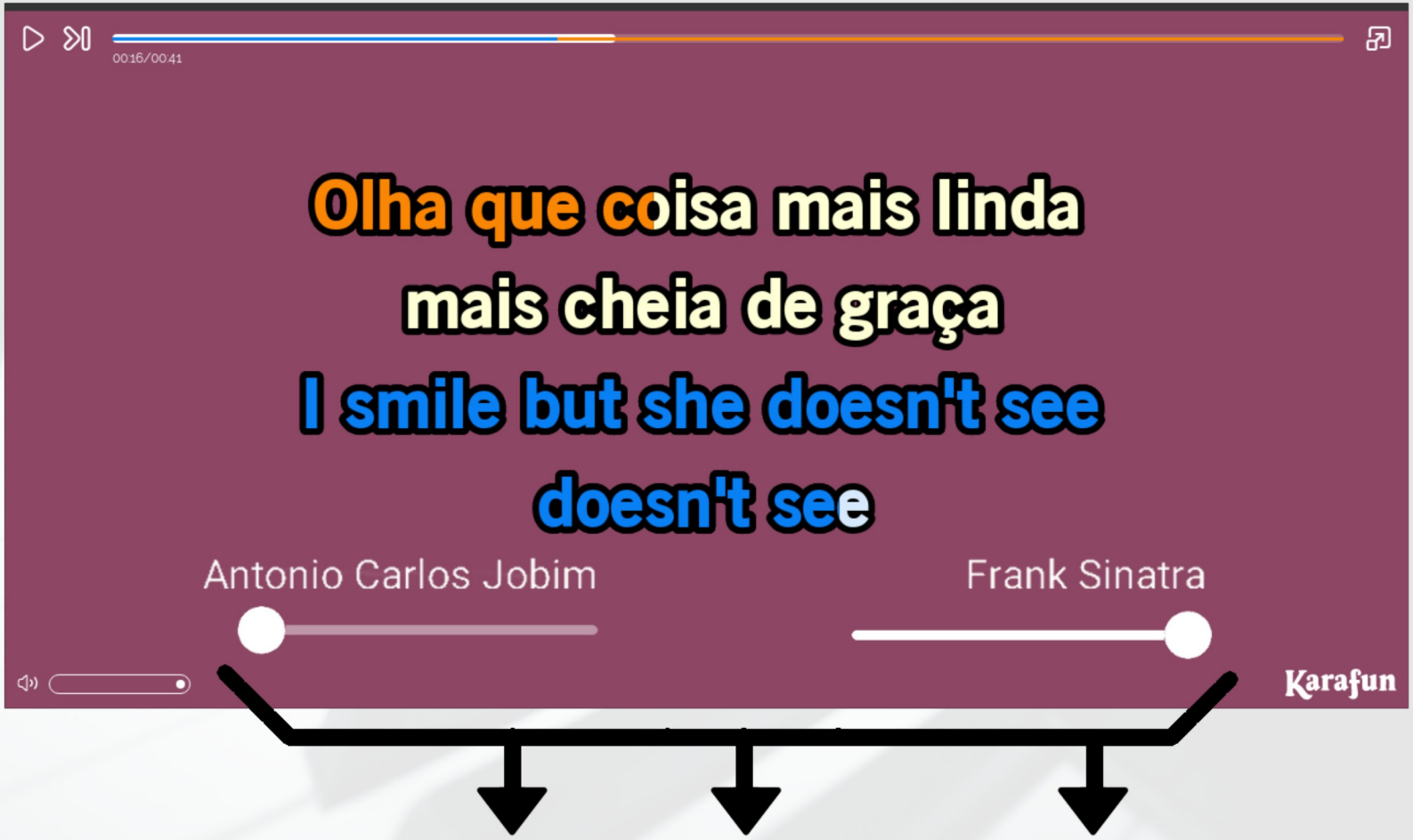
        => YouTube quality (AudioSet)
    - Audio copyright problems
  - Unable to form new tasks / new questions (e.g. how the cover song differs from original) – no datasets for that

Goals – let‘s make it better

- Cover Song Identification
  - The task would benefit **greatly** from sub-problems
    - Finding a cover song only with a different singer / players, but the same musical setup (tempo, harmony)
  - We need a **large-enough** dataset with **many** features, not breaching copyright law along the way
- Singing voice analysis
  - We need dataset with **detailed metadata** (e.g. 5 gender classes)
  - **We need separate tracks (singer, background)**
  - Versatility (gender, language, type of singing, ...)

KaraFun application by Recisio

KaraFun application by Recisio

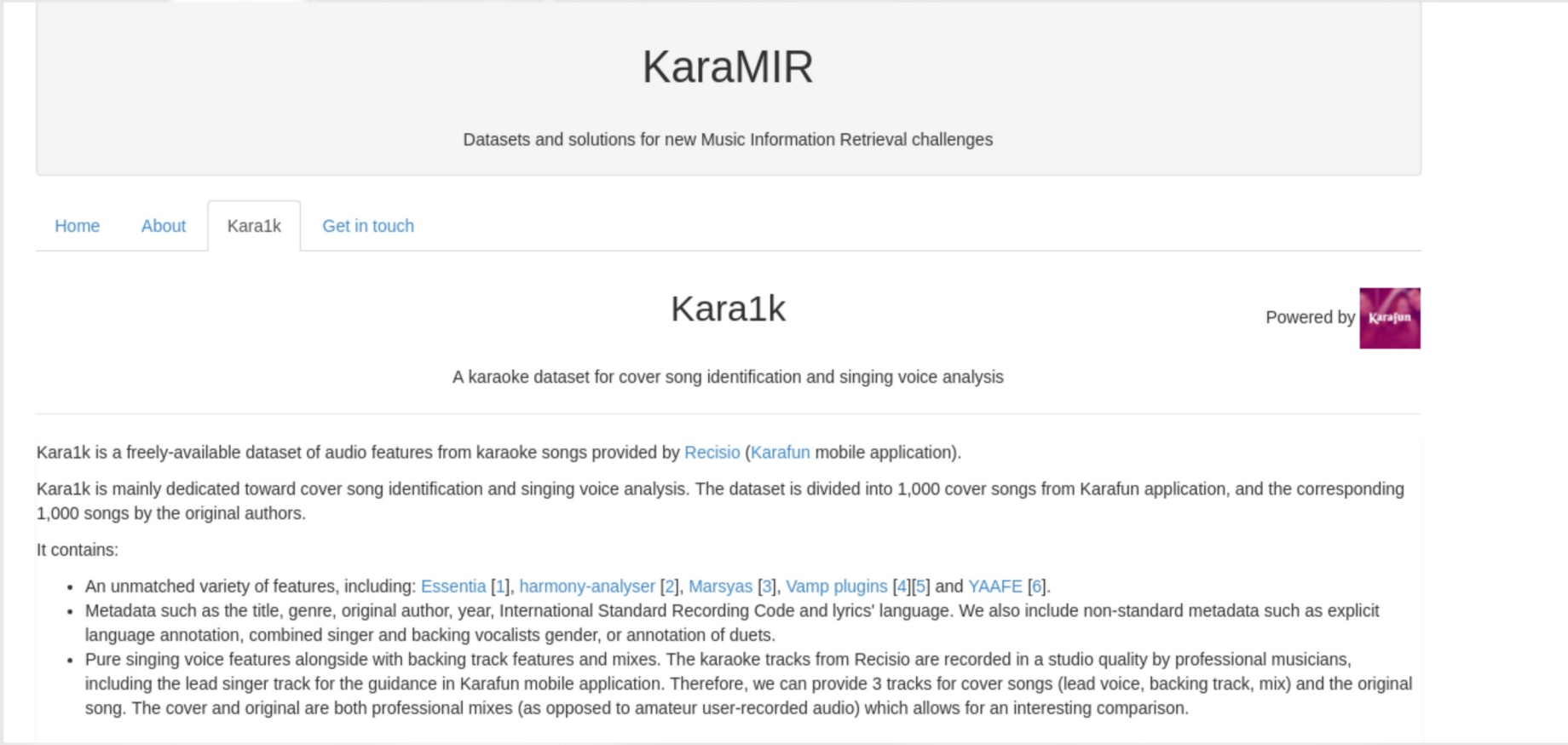


languages	genders	back_vocals	duet
English,Portuguese	male	0	1

Kara1k dataset

http://www.yannbayle.fr/karamir

Kara1k: 1000 original songs, 1000 cover songs from KaraFun

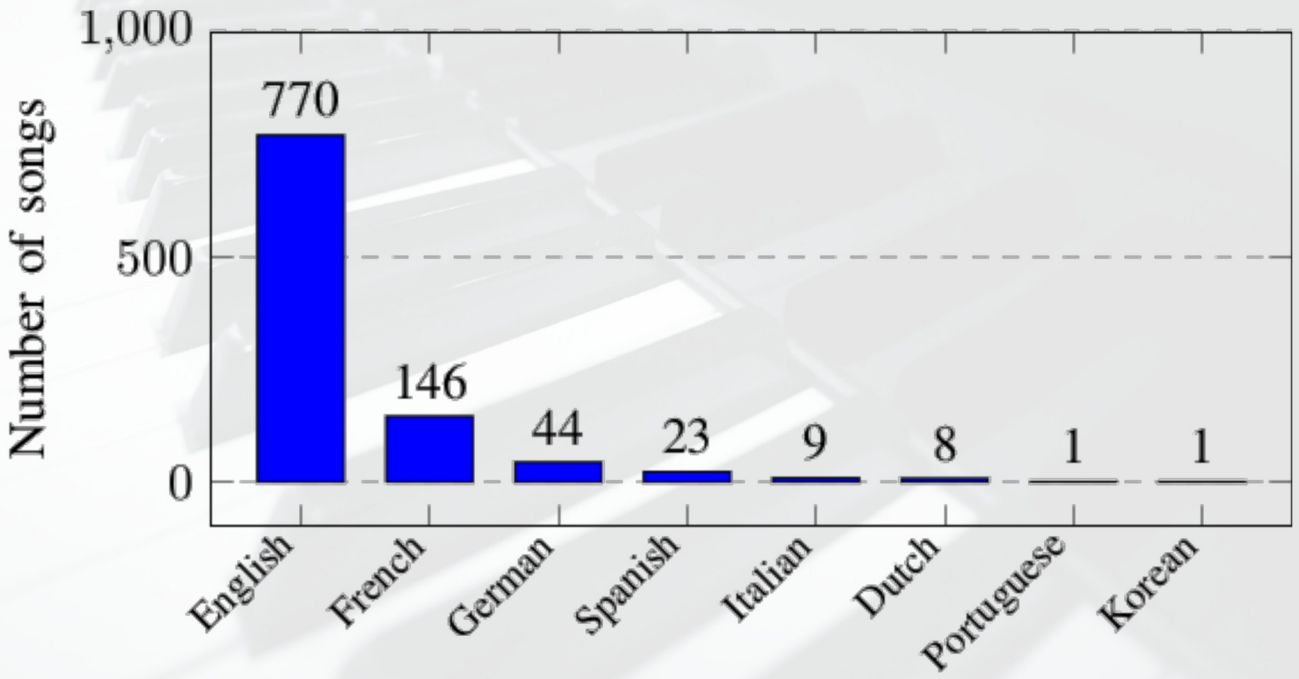


Kara1k dataset as a part of KaraMIR project

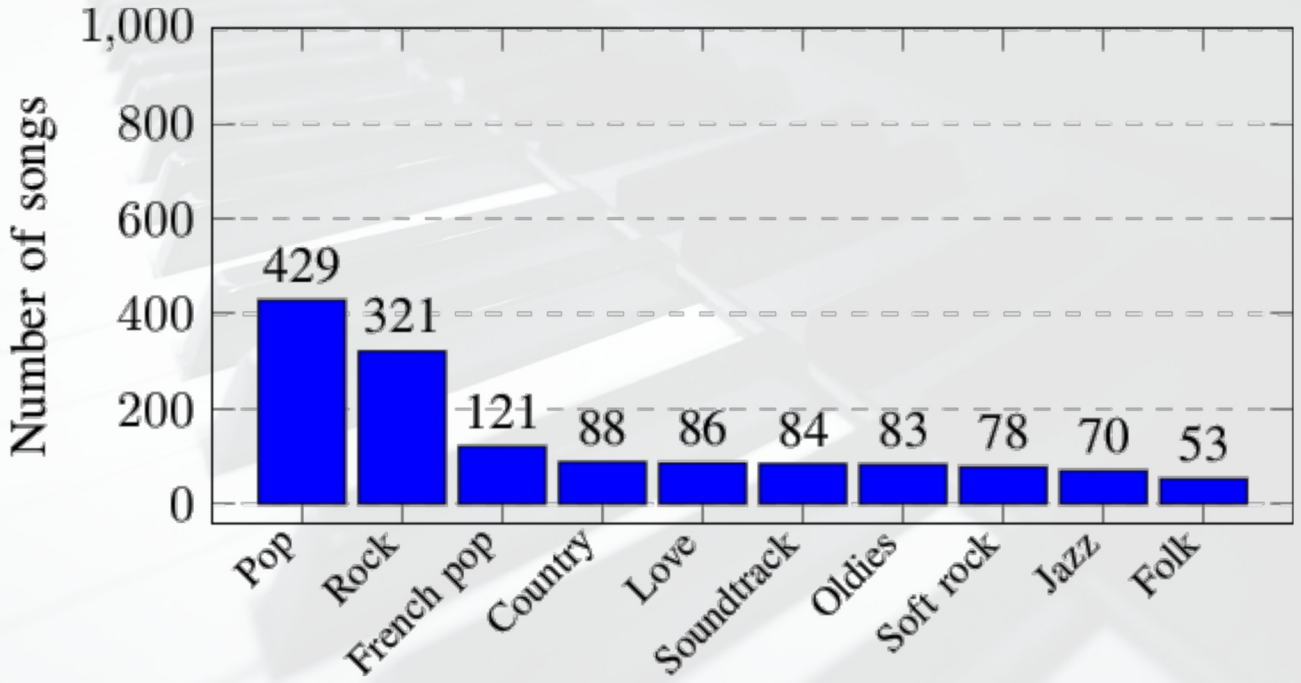
- KaraMIR project
  - Umbrella project for Kara1k, future datasets (Kara2k, …) and related research
  - Gives solutions for the stalled tasks:
    - Datasets with a variety of metadata and features
    - Partnerships with companies
    - Forming new sub-tasks

IT4Innovations
national
supercomputing
center

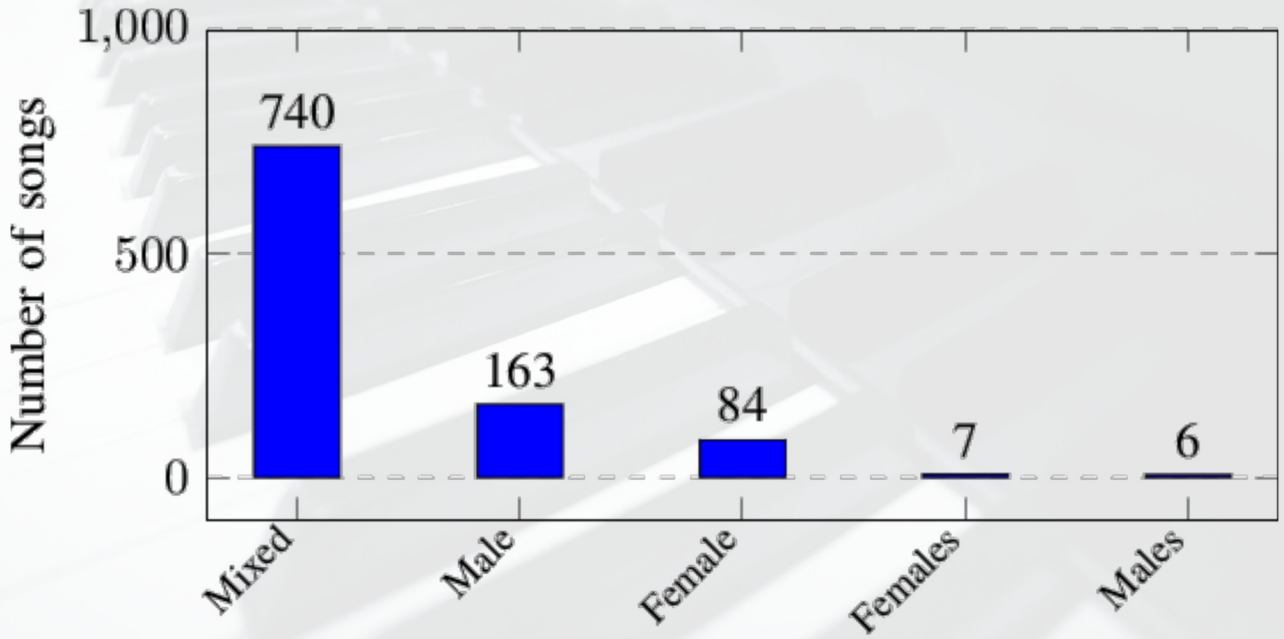
Description of Kara1k - Languages



Description of Kara1k - Genres



Description of Kara1k - Genders



Features extracted in Kara1k

- How we handled the copyright issue?
  - Extracted all features that we could …

essentia.upf.edu marsyas.info

vamp-plugins.org yaafe.sourcforge.net harmony-analyser.org

Special value of Kara1k

- For Cover Song Identification
  - 1000 x originals
  - 1000 x KaraFun cover songs

  - Studio-recorded quality! Not an amateur singing
  - Not different: harmony, tempo, structure
  - Different: singer, musicians' articulation, arrangement

Special value of Kara1k

- For Singing voice analysis
  - 1000 x KaraFun songs do provide metadata such as
    - Language**(s)**
    - Genre**(s)**
    - Gender**(s)**
    - **Backing vocals flag**
    - **Duet flag**
    - **Explicit language annotation**
    - Others (Year of recording, ISRC, …)
  - All songs: Vocal track / Instrumental track / Mix track

Easy to use / download. CSV, TXT, JSON formats

Table 2: Available features in Kara1k

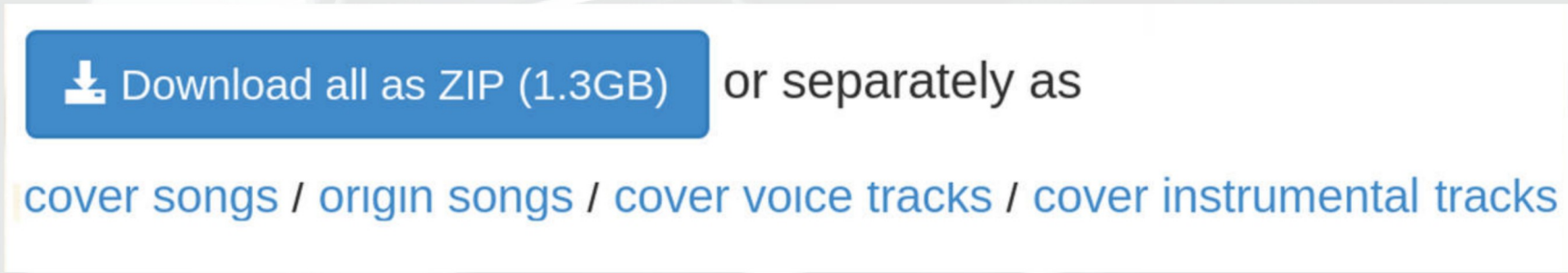
Download all as ZIP (1.3GB) or separately as Essentia, harmony-analyser, Marsyas, Vamp plugins, YAAFE.

You can also download all features for cover songs / origin songs / cover voice tracks / cover instrumental tracks, or particular features in the table below.

Please reach out to us with requests for new features or different parameters

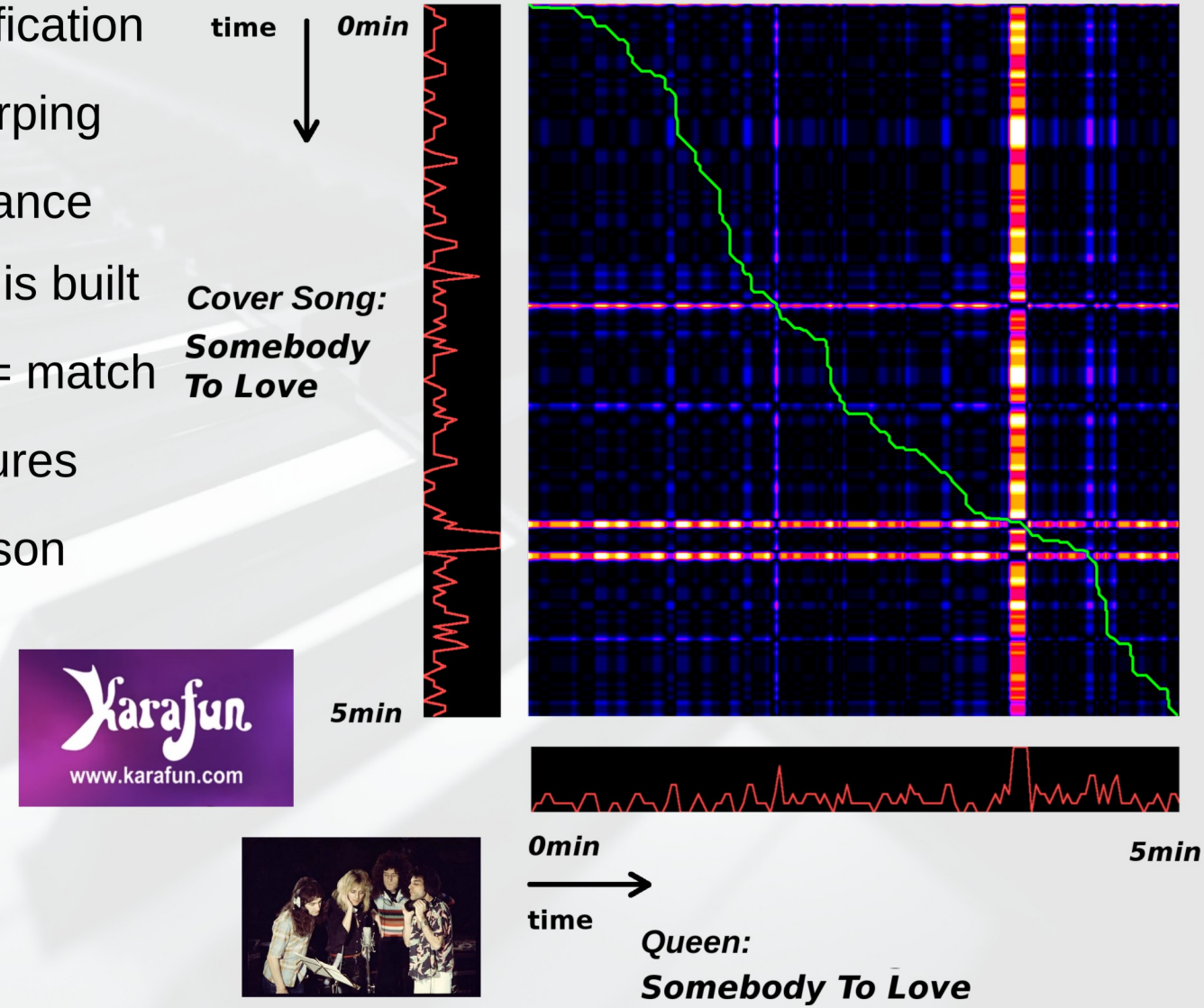
Feature extractor	Feature	Type	Description	Prerequisites	Parameters	Definition	Download
harmony-analyser [2]	Chord vectors frame	Frame	12-dimensional chord vectors, for each frame	Chordino Tones	`version: 1.2-beta`		cover/origin
	Chord vectors	Timestamp	12-dimensional chord vectors, with timestamps	Chordino Tones	`version: 1.2-beta`		cover/origin
	Key vectors frame	Frame	12-dimensional key vectors, for each frame	Key	`version: 1.2-beta; frameRate: 10 Hz`		cover/origin

Easy to use / download. CSV, TXT, JSON formats



First experiments with Kara1k – paving the way

- Cover Song Identification
- Dynamic Time Warping
  - Euclidean distance
- Matrix 1000x1000 is built
- The closest song = match
- Benchmark 6 features
  - novel comparison



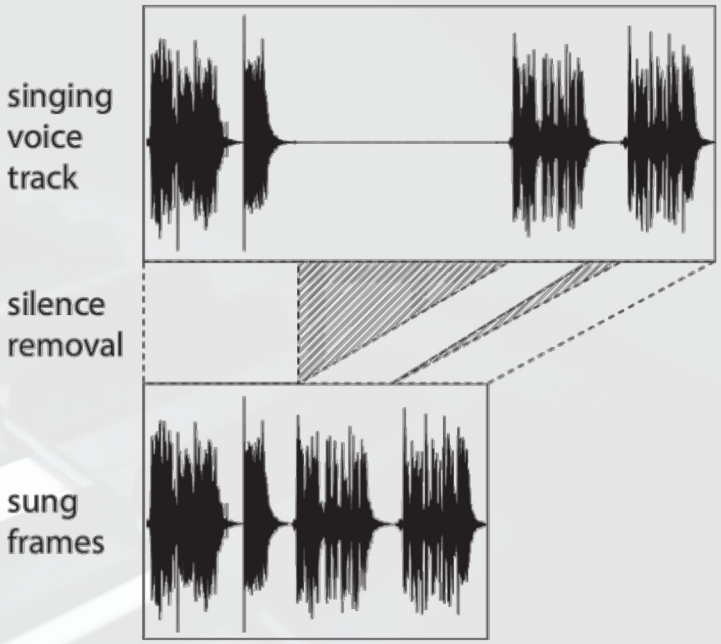
First experiments with Kara1k – paving the way

	MAP	MAR	% detected
Chroma	**0.899**	48.112	**88.9%**
MFCC	0.878	**25.161**	86.3%
Chord	0.865	43.943	84.4%
ChromaCD	0.442	70.994	36.4%
TPS	0.257	141.941	19.3%
Key	0.203	109.954	11.0%

- Novelty: Chroma features were first time compared to Chord / Key features; TPS = Tonal Pitch Space features, first comparison
- Success: Accuracy of 89% is not to be seen for CSI (our task is more simple)

First experiments with Kara1k – paving the way

- Singer Gender Classification
- Comparison: Mix vs Pure voice
- MFCC features from Mix
- MFCC features from Pure voice
- Random Forest Classifier
- 5-fold cross validation



Male

Female

First experiments with Kara1k – paving the way

	Vocal	Mix
Precision	0.721 ± 0.034	0.654 ± 0.040
Recall	0.725 ± 0.029	0.665 ± 0.035
F-score	0.722 ± 0.032	0.656 ± 0.041
Accuracy	0.725 ± 0.029	0.665 ± 0.035

- Conclusion: The performance drops with Mix condition

  MFCC features therefore are not robust against background music

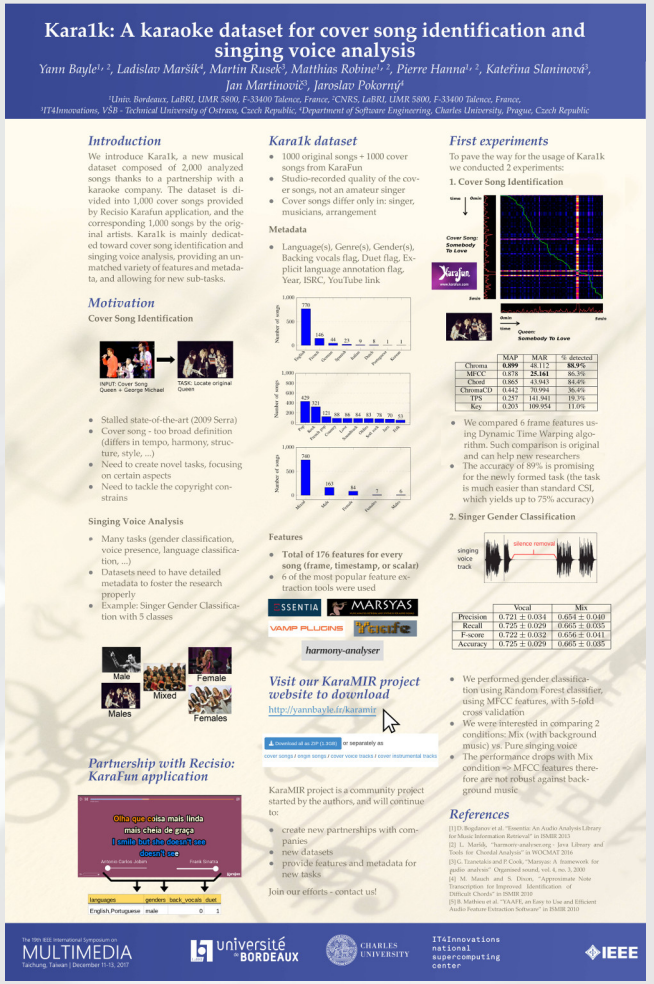
Conclusion

- We have created a new dataset, Kara1k
  - Cover Song Identification, Singing voice analysis, and more
- Freely distributable, easy to use
- Opening doors for new challenges (KaraMIR project)
  - Unmatched variety of features and metadata
  - New sub-tasks for CSI, new tasks for Singing voice analysis
  - New partnership with a company
- Paving the way – first experiments
  - Comparing new features for CSI, 89% precision
  - MFCC not robust against background music for Gender classification

http://www.yannbayle.fr/karamir

Thank you for your attention

… please come to see a poster with Demo



Kara1k: A karaoke dataset for cover song identification and singing voice analysis

Yann Bayle[1, 2], Ladislav Maršík[4], Martin Rusek[3], Matthias Robine[1, 2], Pierre Hanna[1, 2], Kateřina Slaninová[3], Jan Martinovič[3], Jaroslav Pokorný[4]

[1]Univ. Bordeaux, LaBRI, UMR 5800, F-33400 Talence, France. [2]CNRS, LaBRI, UMR 5800, F-33400 Talence, France. [3]IT4Innovations, VŠB - Technical University of Ostrava, Czech Republic, [4]Department of Software Engineering, Charles University, Prague, Czech Republic

Introduction

We introduce Kara1k, a new musical dataset composed of 2,000 analyzed songs thanks to a partnership with a karaoke company. The dataset is divided into 1,000 cover songs provided by Recisio Karafun application, and the corresponding 1,000 songs by the original artists. Kara1k is mainly dedicated toward cover song identification and singing voice analysis, providing an unmatched variety of features and metadata, and allowing for new sub-tasks.

Motivation

Cover Song Identification

- Stalled state-of-the-art (2009 Serra)
- Cover song - too broad definition (differs in tempo, harmony, structure, style, ...)
- Need to create novel tasks, focusing on certain aspects
- Need to tackle the copyright constrains

Singing Voice Analysis

- Many tasks (gender classification, voice presence, language classification, ...)
- Datasets need to have detailed metadata to foster the research properly
- Example: Singer Gender Classification with 5 classes

Partnership with Recisio: KaraFun application

Kara1k dataset

- 1000 original songs + 1000 cover songs from KaraFun
- Studio-recorded quality of the cover songs, not an amateur singer
- Cover songs differ only in: singer, musicians, arrangement

Metadata

- Language(s), Genre(s), Gender(s), Backing vocals flag, Duet flag, Explicit language annotation flag, Year, ISRC, YouTube link

Features

- **Total of 176 features for every song (frame, timestamp, or scalar)**
- 6 of the most popular feature extraction tools were used

ESSENTIA MARSYAS VAMP PLUGINS Aubio

harmony-analyser

Visit our KaraMIR project website to download

http://yannbayle.fr/karamir

KaraMIR project is a community project started by the authors, and will continue to:

- create new partnerships with companies
- new datasets
- provide features and metadata for new tasks

Join our efforts - contact us!

First experiments

To pave the way for the usage of Kara1k we conducted 2 experiments:

1. Cover Song Identification

	MAP	MAR	% detected
Chroma	**0.109**	48.112	**88.9%**
MFCC	0.078	**25.161**	86.3%
Chord	0.065	43.943	84.4%
ChromaCH	0.442	70.994	36.4%
TPS	0.257	141.941	19.3%
Key	0.203	100.954	11.0%

- We compared 6 frame features using Dynamic Time Warping algorithm. Such comparison is original and can help new researchers
- The accuracy of 89% is promising for the newly formed task (the task is much easier than standard CSI, which yields up to 75% accuracy)

2. Singer Gender Classification

	Vocal	Mix
Precision	0.721 ± 0.034	0.654 ± 0.040
Recall	0.725 ± 0.029	0.665 ± 0.035
F-score	0.722 ± 0.032	0.656 ± 0.041
Accuracy	0.725 ± 0.029	0.665 ± 0.035

- We performed gender classification using Random Forest classifier, using MFCC features, with 5-fold cross validation
- We were interested in comparing 2 conditions: Mix (with background music) vs. Pure singing voice
- The performance drops with Mix condition so MFCC features therefore are not robust against background music

References

[1] D. Bogdanov et al. "Essentia: An Audio Analysis Library for Music Information Retrieval" in ISMIR 2013

[2] L. Maršík, "harmony-analyser.org - Java Library and Tools for Chordal Analysis" in WOCMAT 2016

[3] G. Tzanetakis and P. Cook, "Marsyas: A framework for audio analysis" Organised sound, vol. 4, no. 3, 2000

[4] M. Mauch and S. Dixon, "Approximate Note Transcriptions for Improved Identification of Difficult Chords" in ISMIR 2010

[5] B. Mathieu et al. "YAAFE, an Easy to Use and Efficient Audio Feature Extraction Software" in ISMIR 2010

The 19th IEEE International Symposium on MULTIMEDIA
Taichung, Taiwan | December 11-13, 2017

université de BORDEAUX · CHARLES UNIVERSITY · IT4Innovations national supercomputing center · IEEE