

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

Yann Bayle^{1, 2}, Ladislav Maršík⁴, Martin Rusek³, Matthias Robine^{1, 2}, Pierre Hanna^{1, 2}, Katerina Slaninová³, Jan Martinovič³, Jaroslav Pokorný⁴

¹Univ. Bordeaux, LaBRI, UMR 5800, F-33400 Talence, France, ²CNRS, LaBRI, UMR 5800, F-33400 Talence, France,

³IT4Innovations, VŠB - Technical University of Ostrava, Czech Republic, ⁴Department 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

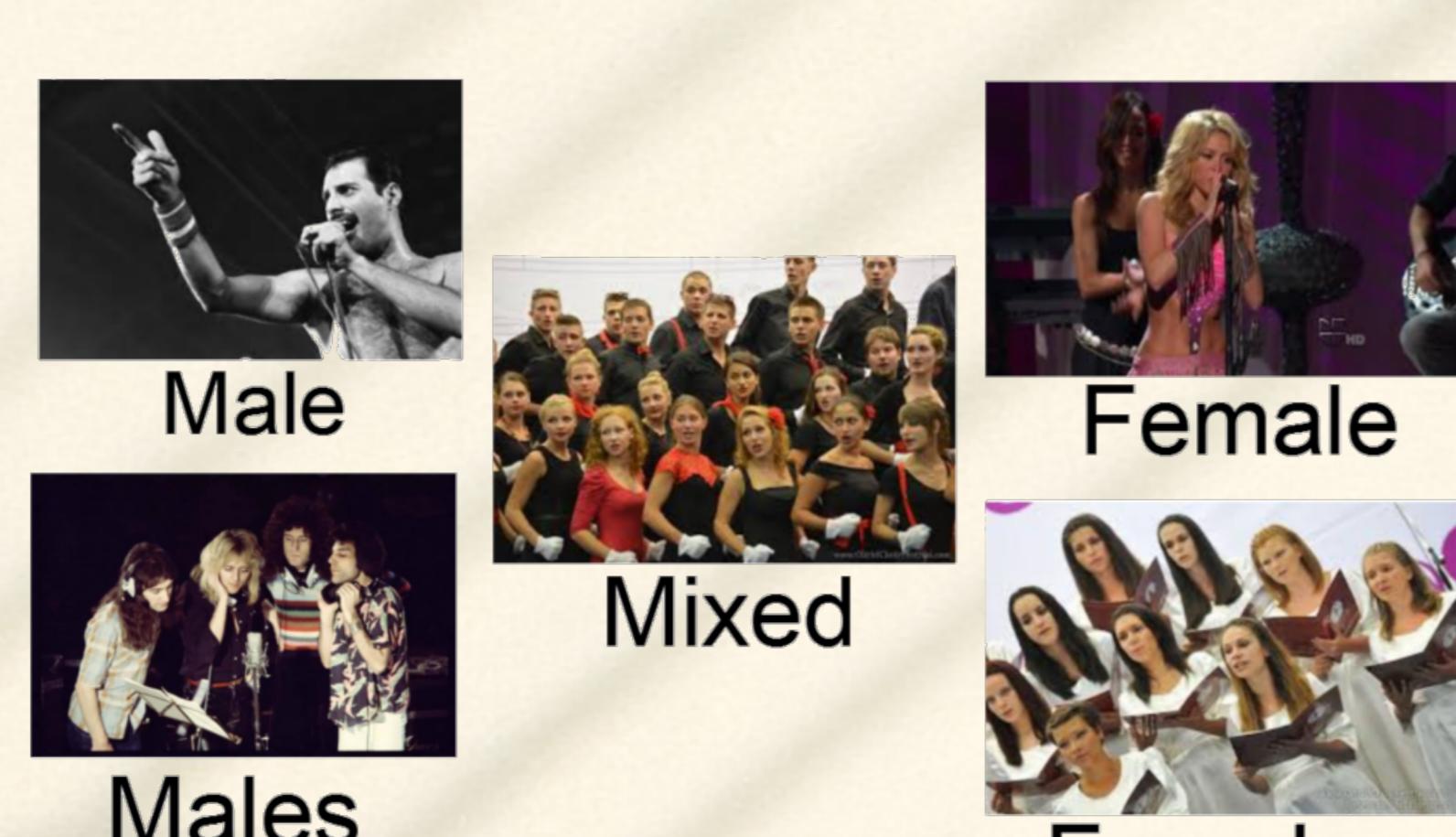
Cover Song Identification



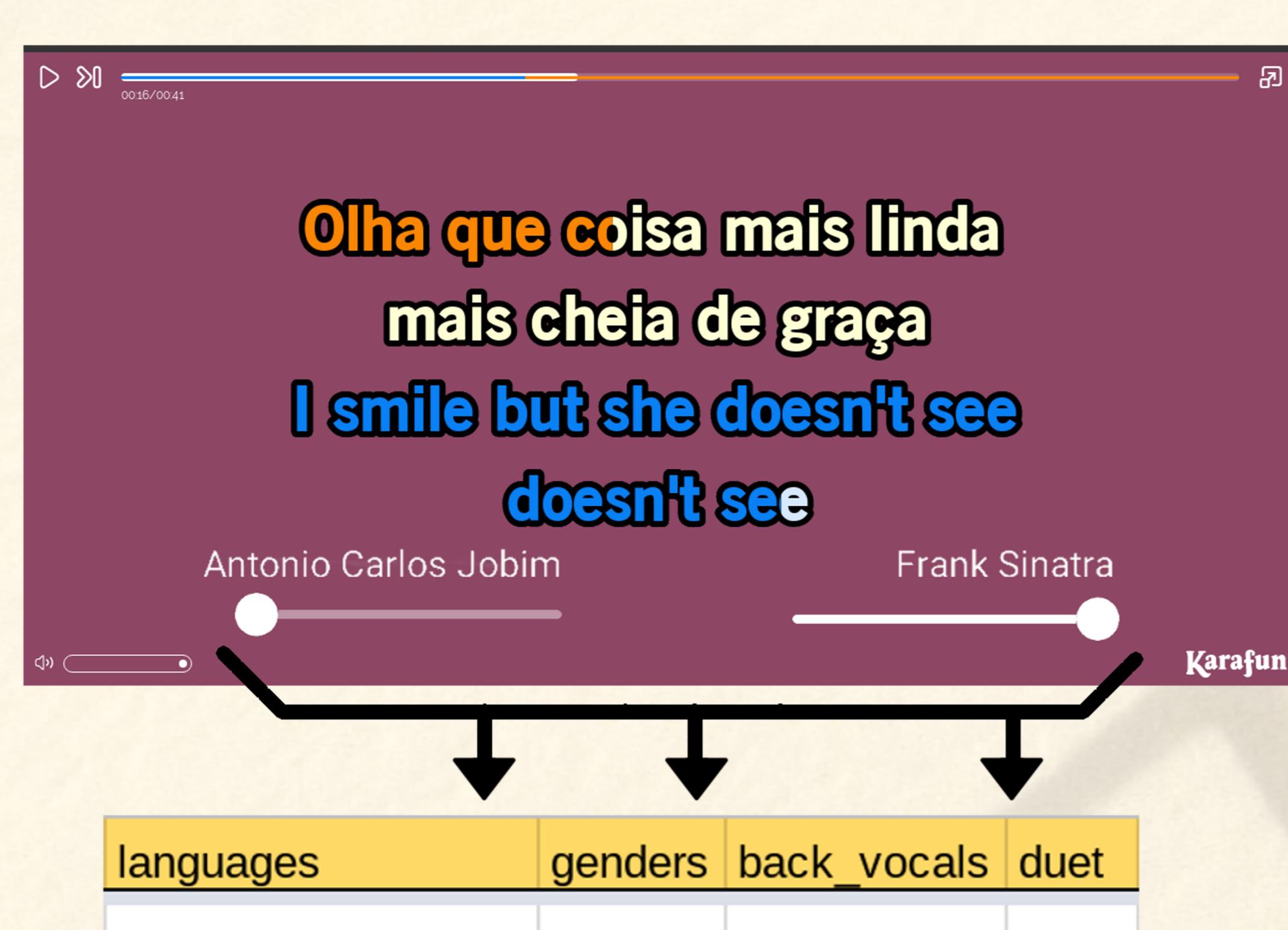
- Stalled state-of-the-art (2009 Serra)
- Cover song - too broad definition (differs in tempo, harmony, structure, style, ...)
- Need to create novel tasks, focusing on certain aspects
- Need to tackle the copyright constraints

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

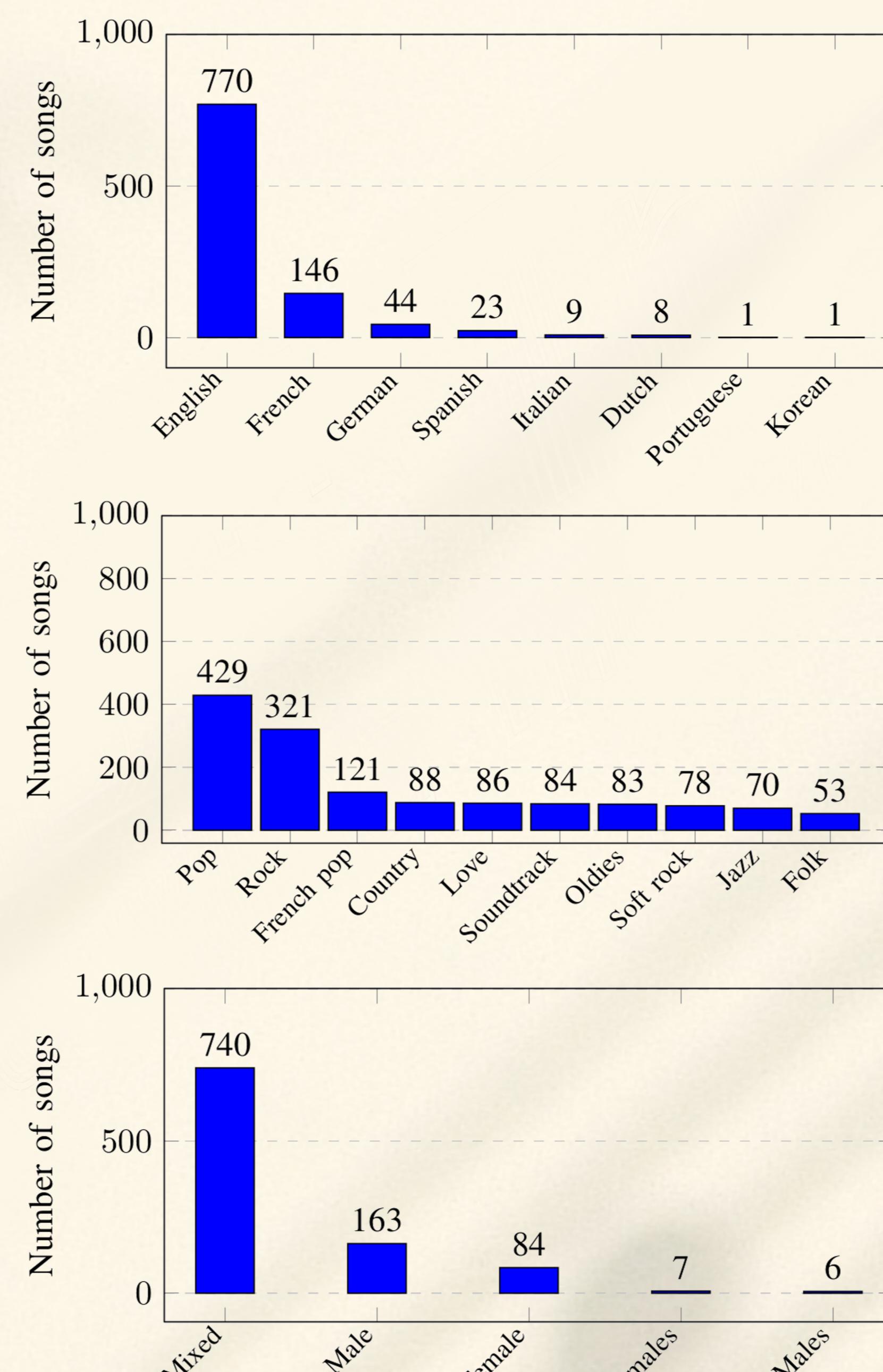


Kara1k dataset

- 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

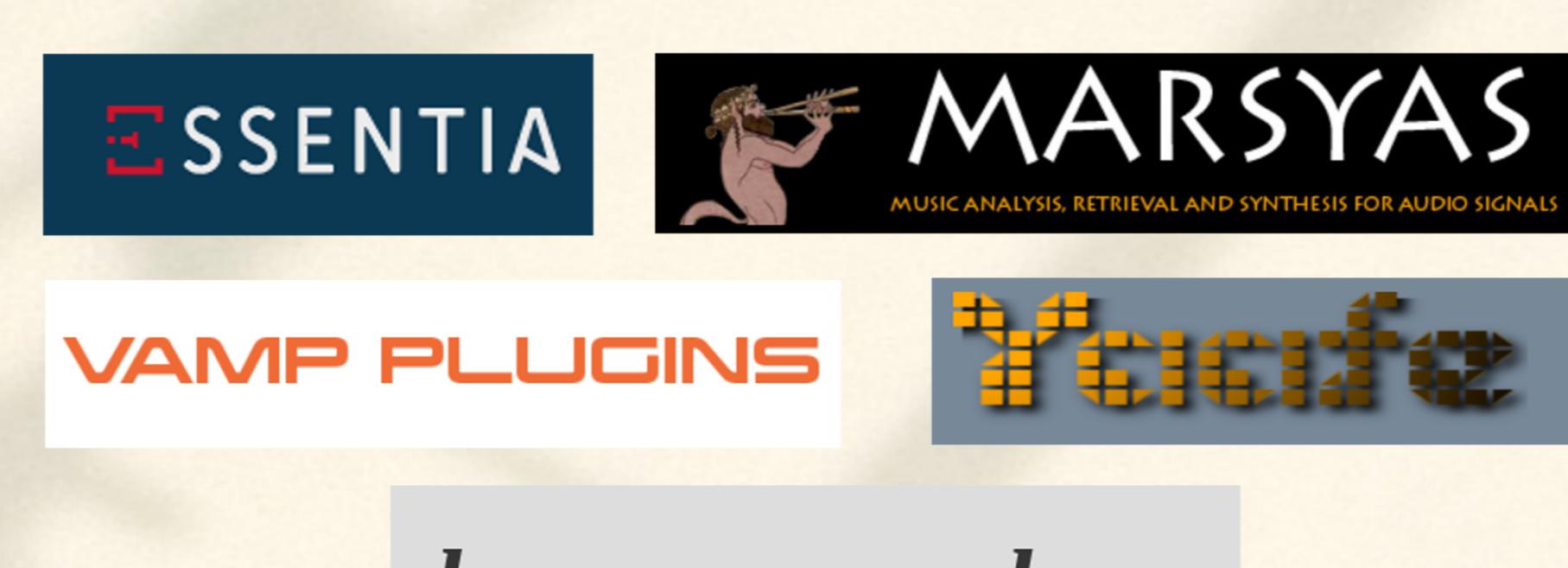
Metadata

- 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



Visit our KaraMIR project website to download

<http://yannbayle.fr/karamir>

Download all as ZIP (1.3GB) or separately as
cover songs / origin songs / cover voice tracks / cover instrumental tracks

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

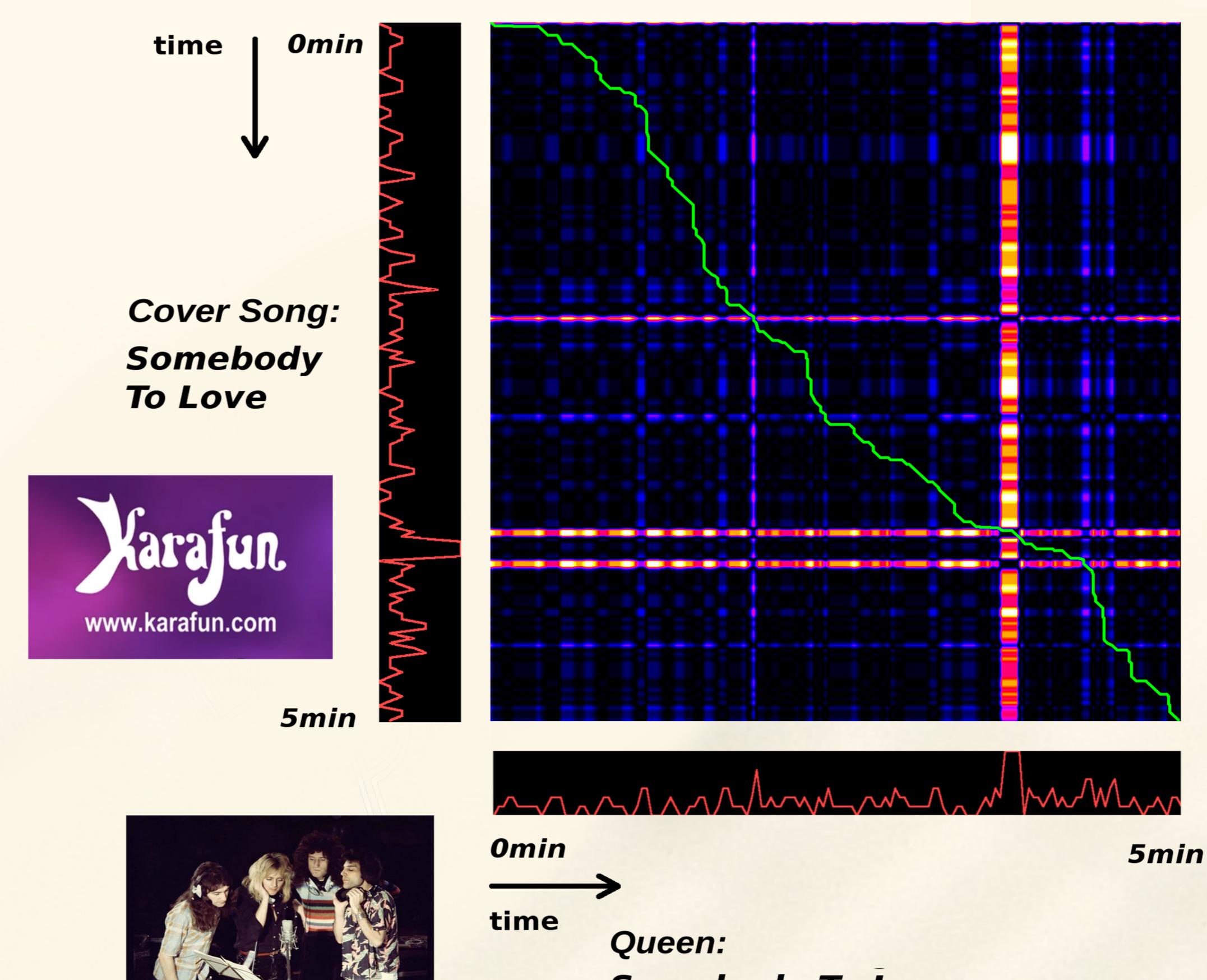
- create new partnerships with companies
- new datasets
- provide features and metadata for new tasks

Join our efforts - contact us!

First experiments

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

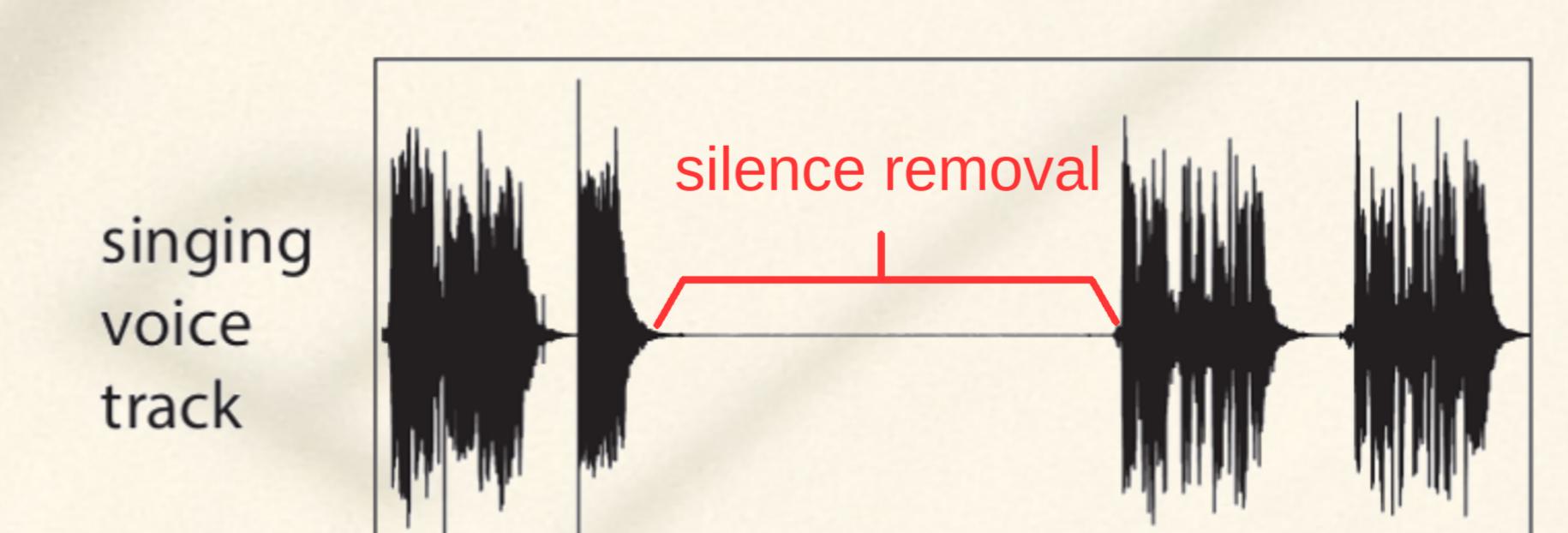
1. Cover Song Identification



	MAP	MAR	% detected
Chroma	0.899	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%

- 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 => 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 Transcription 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