FEATURES FOR COMPARING TUNE SIMILARITY OF SONGS ACROSS DIFFERENT LANGUAGES

Naveen Kumar, Andreas Tsiartas, Shrikanth Narayanan
Signal Analysis and Interpretation Laboratory (SAIL), University of Southern California, Los Angeles, CA-900089
{komathnk,tsiartas}@usc.edu, shri@sipi.usc.edu

Problem

- Cross Lingual tune matching
- Uses in indexing to study global music influences
- Plagiarism detection for digital media rights management

Song Database

- 24 parallel songs obtained from Youtube in these languages

Hindi	17	Tamil	11
Bengali	6	Korean	6
Japanese	3	English	2
Others	3	Total	48

- Explicit cross lingual covers or dubs chosen
- Songs from different quality, sources

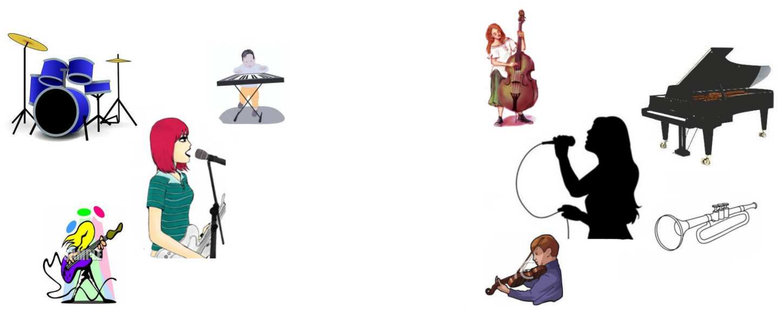
Existing Approaches

Popular features for Music Information Retrieval include:

- Chroma Features
- MFCC Features
- Melody Features on vocal segments

Motivation

- Main melody in **vocals** is an **invariant** feature in a song.
- By nature of the problem a vocal track is always guaranteed.
- Capture mutliple band information, instead of tracking vocals.



Features

Consider a music signal $x[n]$ and its autocorrelation

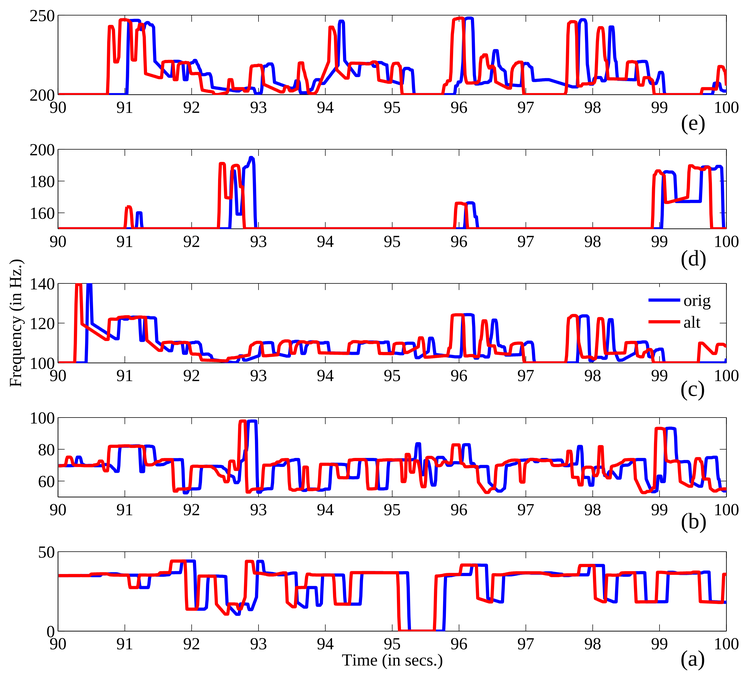
$$R[k] = \sum_{m=2}^{L-1-k} x[m]x[m+k]; \forall k \in \left(\frac{F_s}{f_{max}}, \frac{F_s}{f_{min}}\right)$$

Computed frequencies f_i corresponding to peaks in $R[k]$ for the bands

i	f_{min}(Hz)	f_{max}(Hz)
1	5	50
2	50	100
3	100	150
4	150	200
5	200	250
6	250	350
7	350	450
8	450	550

Performed feature interpolation, median filtering and normalization.

Multi Band Autocorrelation Peaks (MBAP)



Figs. 1a-1e show MBAP for bands 1-5 for a song segment in `orig` and `alt`.

Algorithm

- Problem defined as 24-way classification task for each song.
- Used **Dynamic Time Warping** (DTW) for song alignment.
- Classification metric : total alignment cost
- Distance metric : **Mahalanobis distance**
- **Downsampling** to reduce the time complexity.

Results

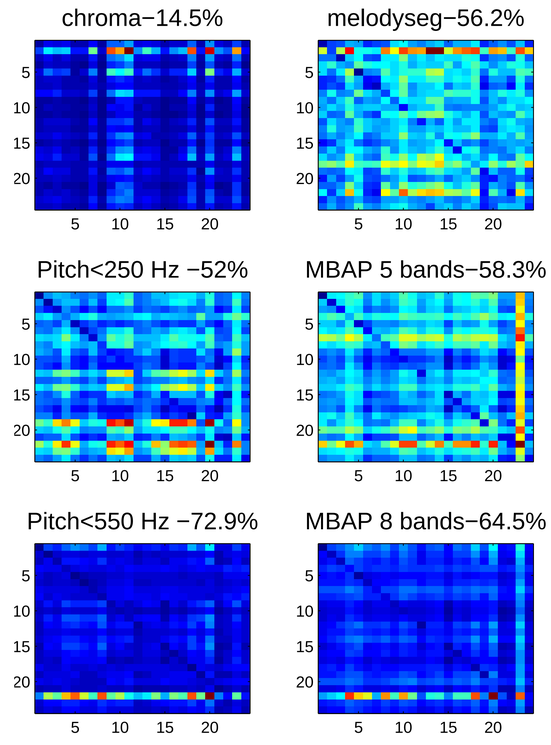


Fig 2. 24 × 24 DTW alignment cost matrices for the features: Chroma, Melody, Pitch and MBAP. *A strong diagonal suggests high accuracy.*

- Improvements are significant (p=0.1)
- Fundamental frequencies in multiple bands increase robustness
- Performs better on vocal unsegmented songs
- MBAP superior to Pitch for non-transposed songs (81% vs 77%)

Future Work

- Deal with the effect of key/scale transposes in higher MBAP bands
- Explore faster time series similarity methods
- Experiment on larger and more extensive datasets