Evaluation of Chord and Chroma Features and Dynamic Time Warping Scores on Cover Song Identification Task

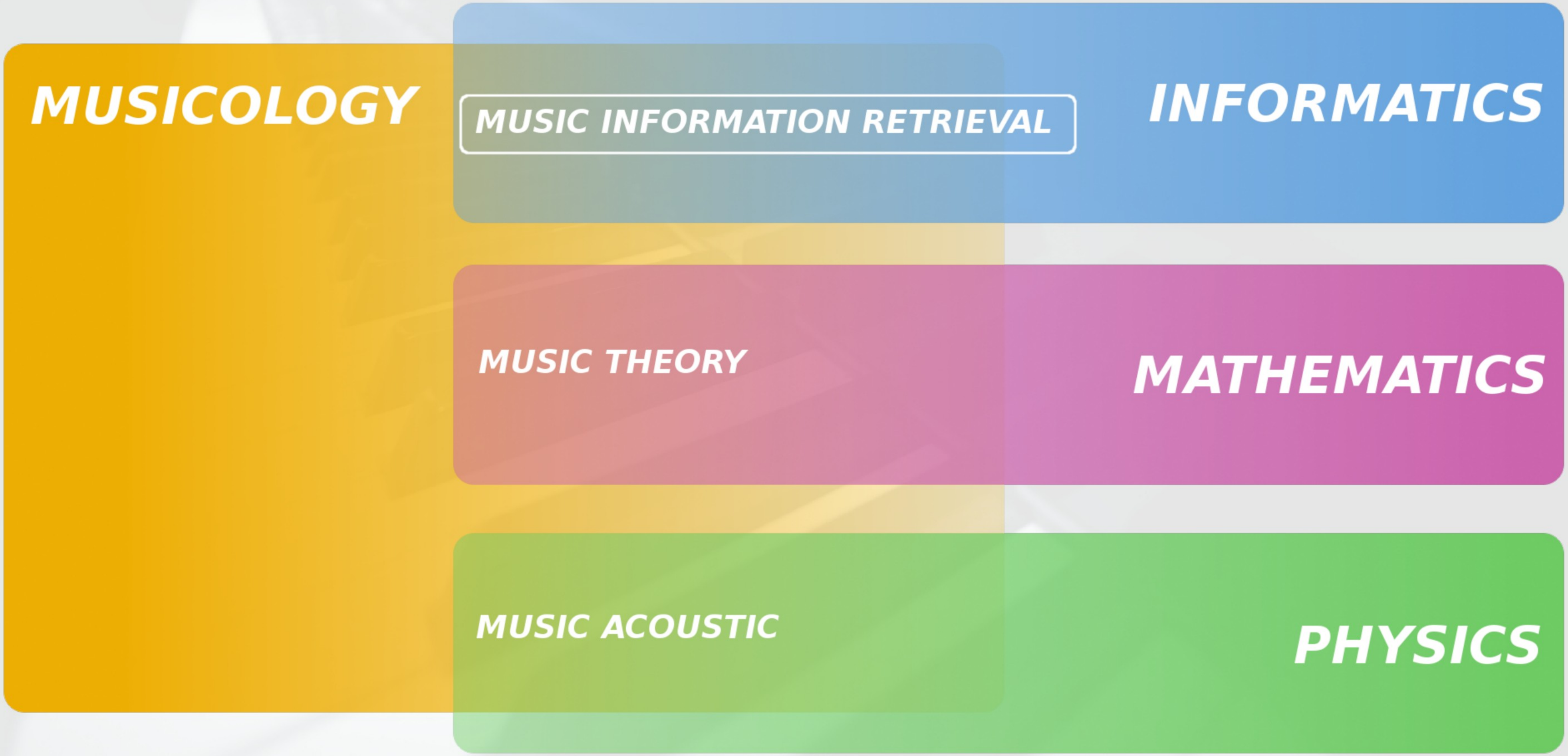
Ladislav Maršík[3], Martin Rusek[1]
Kateřina Slaninová[1,2], Jan Martinovič[1,2], Jaroslav Pokorný[3]

1: IT4Innovations, VŠB - Technical University of Ostrava, Ostrava, Czech Republic
2: Department of Computer Science, FEECS, VŠB - Technical University of Ostrava, Ostrava, Czech Republic
3: Faculty of Mathematics and Physics, Charles University, Prague, Czech Republic

IT4Innovations national supercomputing center

CISIM 2017, Bialystok, Poland

Music Information Retrieval



Music Information Retrieval

MUSICOLOGY

INFORMATICS

MUSIC TECHNOLOGY
AUDIO SIGNAL PROCESSING
MUSIC INFORMATION RETRIEVAL
OPTICAL MUSIC SCORE RECOGNITION
MUSIC COGNITION
MUSIC SYNTHESIS & COMPUTER MUSIC

Task – Cover Song Identification

- Cover song
  - An alternative version, performance, or recording of a previously published musical piece

Queen

Somebody To Love

Queen + George Michael

Somebody To Love

Motivation

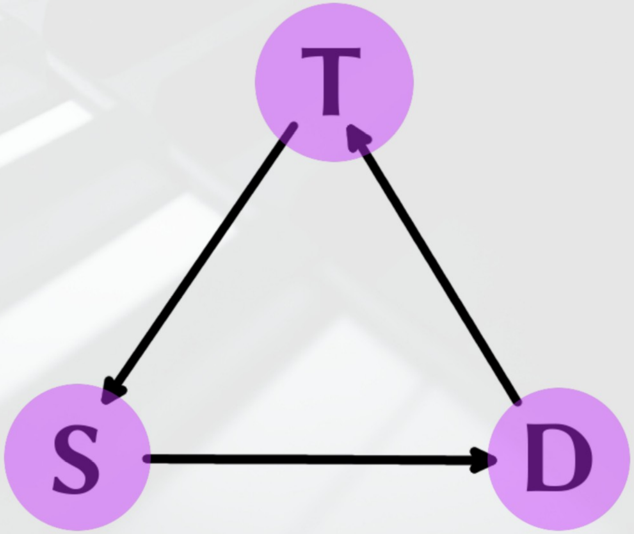
- Contrary to Audio Fingerprinting (e.g. Shazam), **Cover Song Identification** is not widely known and used yet
  - Complexity of the problem, cover song variability
  - State-of-the-art (Serra et al., MIREX 2007, 73.5 % acc.)
  - New impuls needed for CSI
    1. Features comparison
       - We compare harmony features for CSI
    2. Algorithms comparison
       - We compare (Dynamic Time Warping) DTW scores
    3. Large-scale research
       - We form harmony „fingerprints“
    4. New data sets
       - Future work (ISMIR)

Why Music Harmony?

„Most important in music is its harmony.“

Ilja Zeljenka, Slovak Composer

- Tonic
- Subdominant
- Dominant

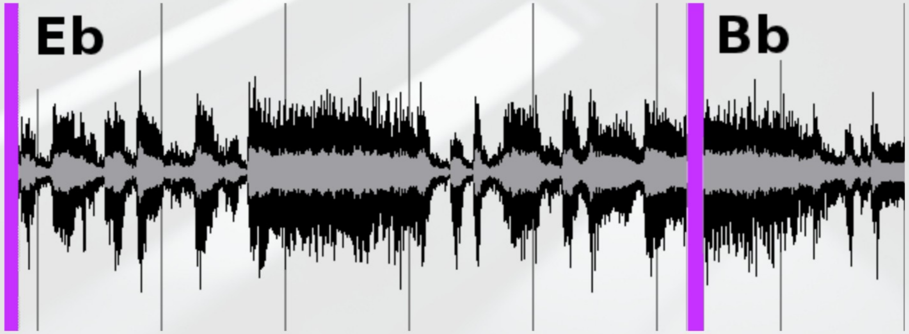


Demo 1: Simple Harmony Movement

Folk song: Slovenské mamičky

Basic harmonic functions

T – D – T – D



Demo 2: Complex Harmony Movement

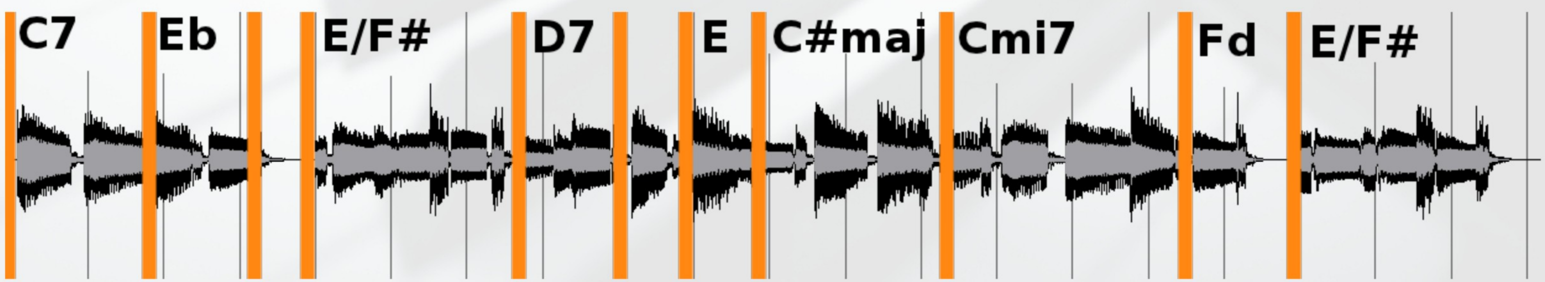
Hiromi: 010101 (Binary System)

Modifications of basic harmonic functions

$D_p - T_p - S_p - D_p$

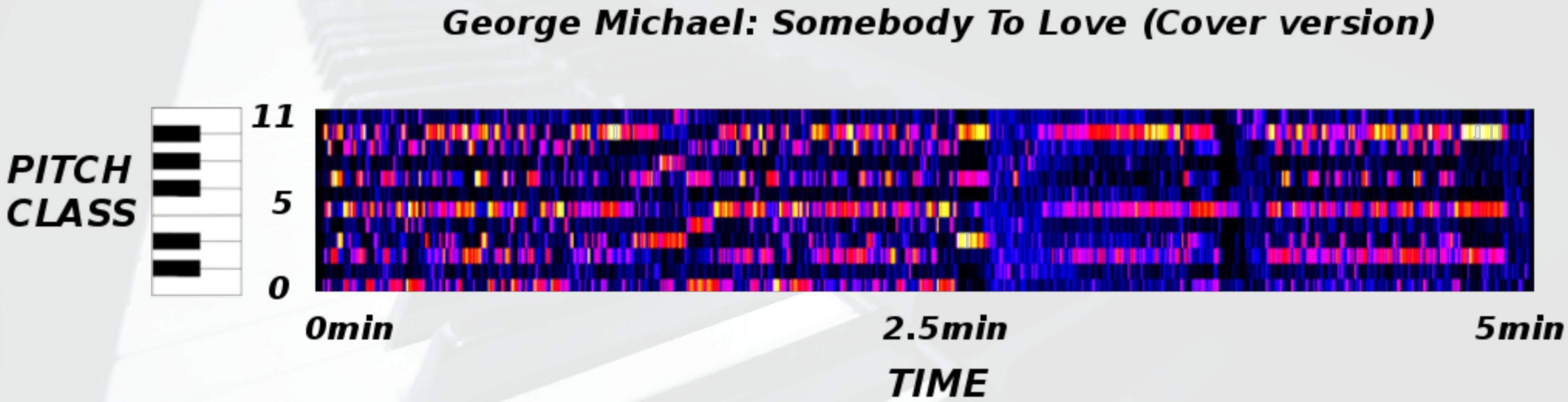
$T_7 - D_p - D_7$ $S_p - D_{mi}$ $T_p - S_{dim} - D_p$

$S_p - D_p - S_7$



Harmony feature 1: Chroma Vectors „raw feature“

Raw Chroma Vectors



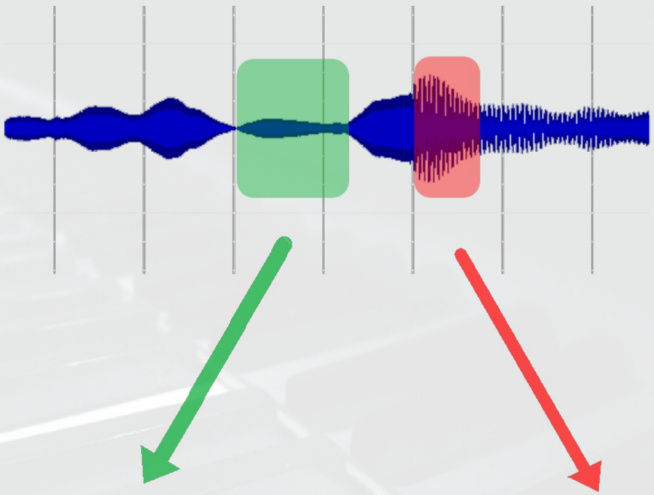
- Frequently used in MIR / Cover Song Identification
- **12-dimensional feature**, fixed sample rate (22Hz)
- Advantage: Proved to contain enough harmony information
- Drawback: Shifting needed (Cover song can be played in another key)

Harmony feature 2: Chroma Distances „fingerprint“

Chroma Complexity Distance (ChromaCD)

- Idea: A distance derived from a pair of subsequent Chroma vectors
- **1 dimension (scalar)**, fixed sample rate (22Hz)
- Advantage: Reduced size, abstraction from the key, still can detect specific song parts (instrument solos, rhytmic patterns)
- Drawback: Loss of information

In depth: Feature 2 (Chroma Complexity Distance)

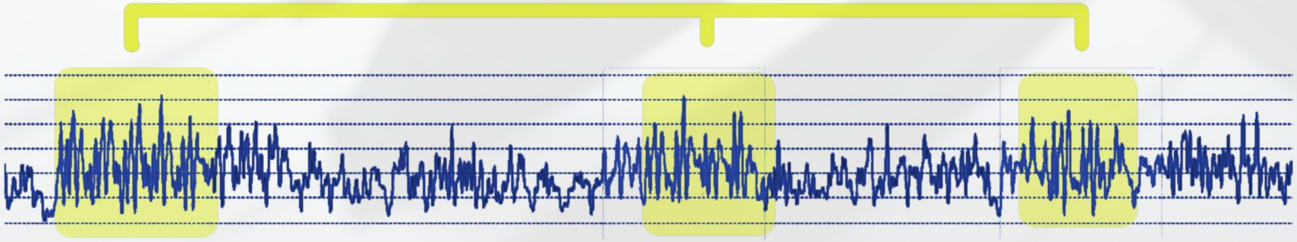


There are no tonal changes

The tonality changes rapidly

Michael Jackson: The Smooth Criminal

The same guitar riffs



Harmony feature 3: Chord Distances „fingerprint“

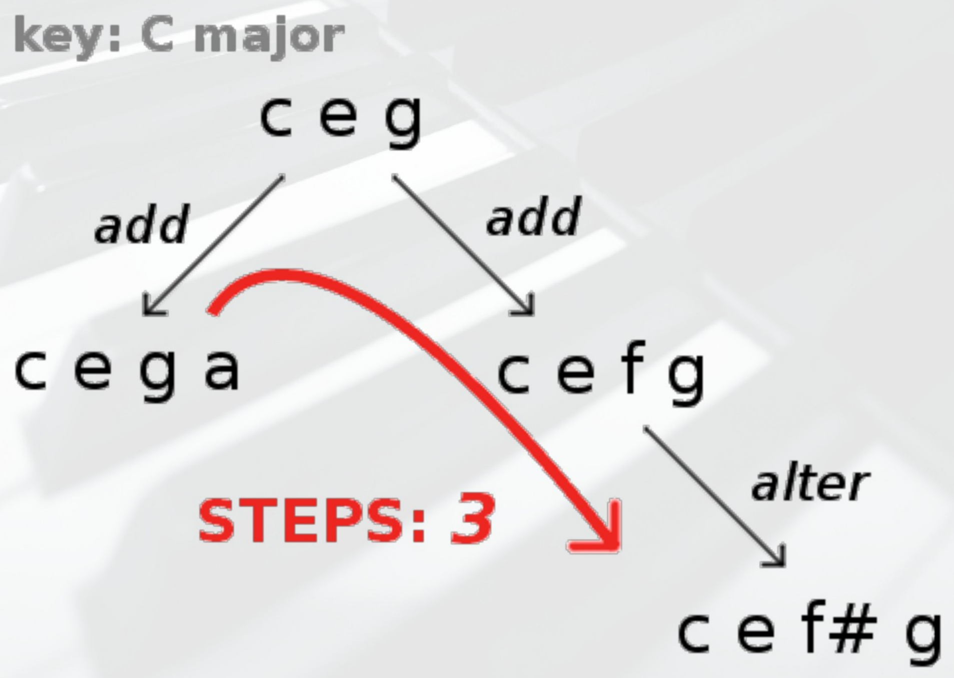
Chord Complexity Distance (ChordCD)

- Idea: A distance derived from a pair of subsequent Chords
- **1 dimension (scalar)**, **chord segmentation detects the new block**
- Advantage: Greatly reduced size, abstraction from the key, abstraction from the specific song parts (only considers harmony)
- Drawback: Great loss of information

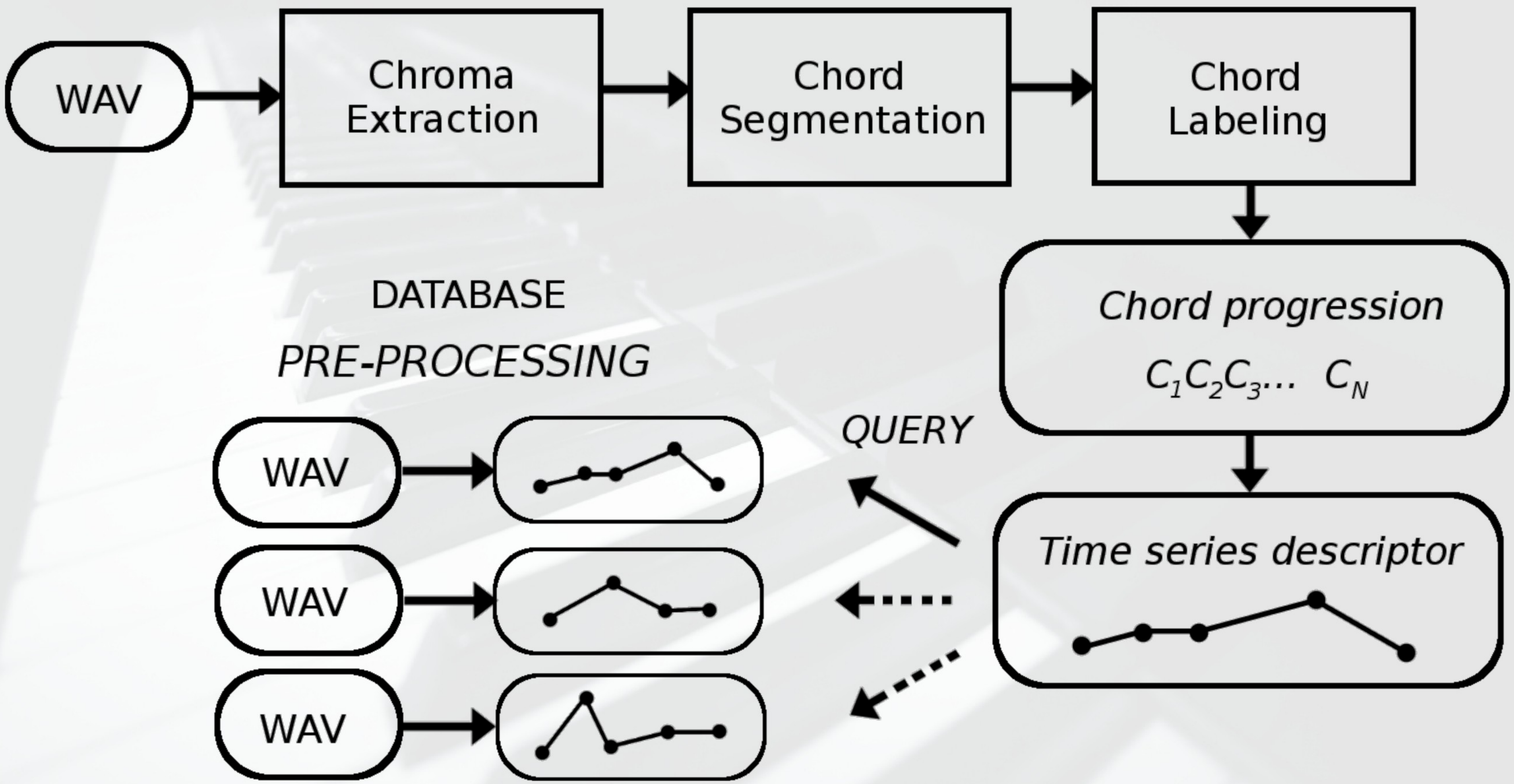
In depth: Feature 3 (Chord Complexity Distance)

Chord Complexity Distance (ChordCD)

(a variation of Edit Distance)



Experiments

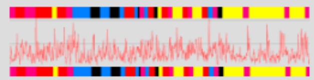


All computation were performed on ANSELM supercomputer in IT4Innovations national supercomputing center

harmony-analyser.org

harmony-analyser.org Repository Releases Screenshots Documentation Publications Contact

harmony-analyser



harmony-analyser is a set of visual tools for music harmony analysis of WAV/MIDI input, powered by JHarmonyAnalyser library

The difference we bring is the approach based on music theory, chord and chroma distances. JHarmonyAnalyser uses recent music theory models to extract musical meaning and distances between chords and chroma vectors. We aim to develop open-source music player, which is musician / musicologist-friendly and aid recent music information retrieval tasks.

harmony-analyser tools and JHarmonyAnalyser library are licenced under the GNU GPL License.

Tools are compatible with GPL Licensed Vamp plugins which can be used for additional analysis.

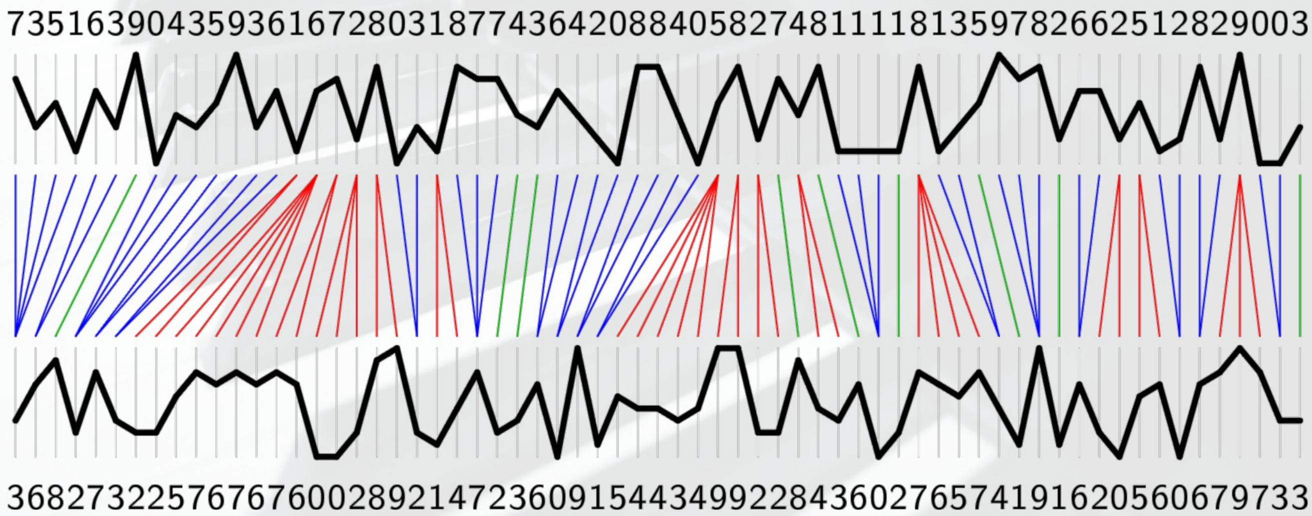
To contribute, please follow our guideline in GitHub repository.

Releases

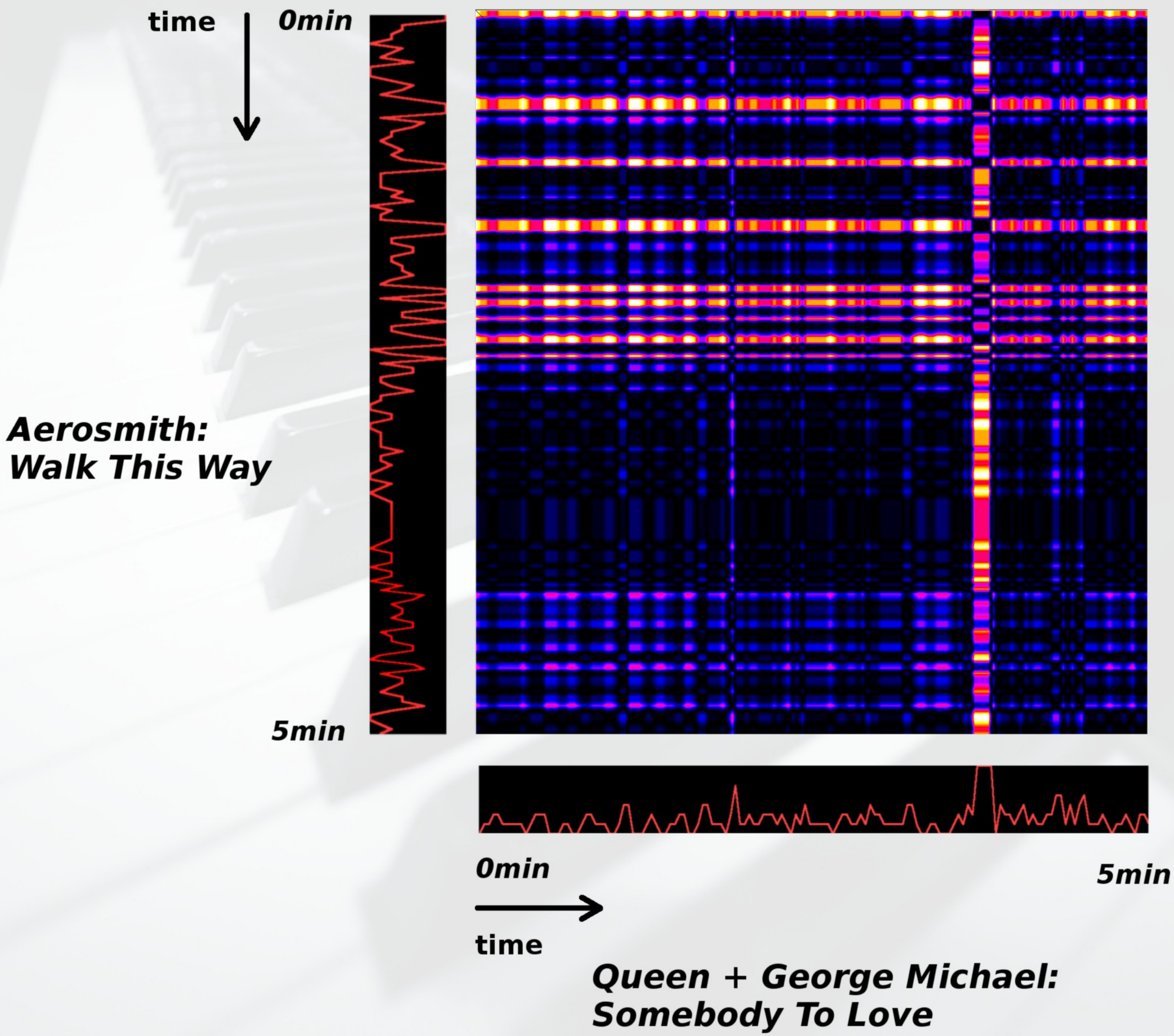
Please choose from the releases below:

Dynamic Time Warping – Idea

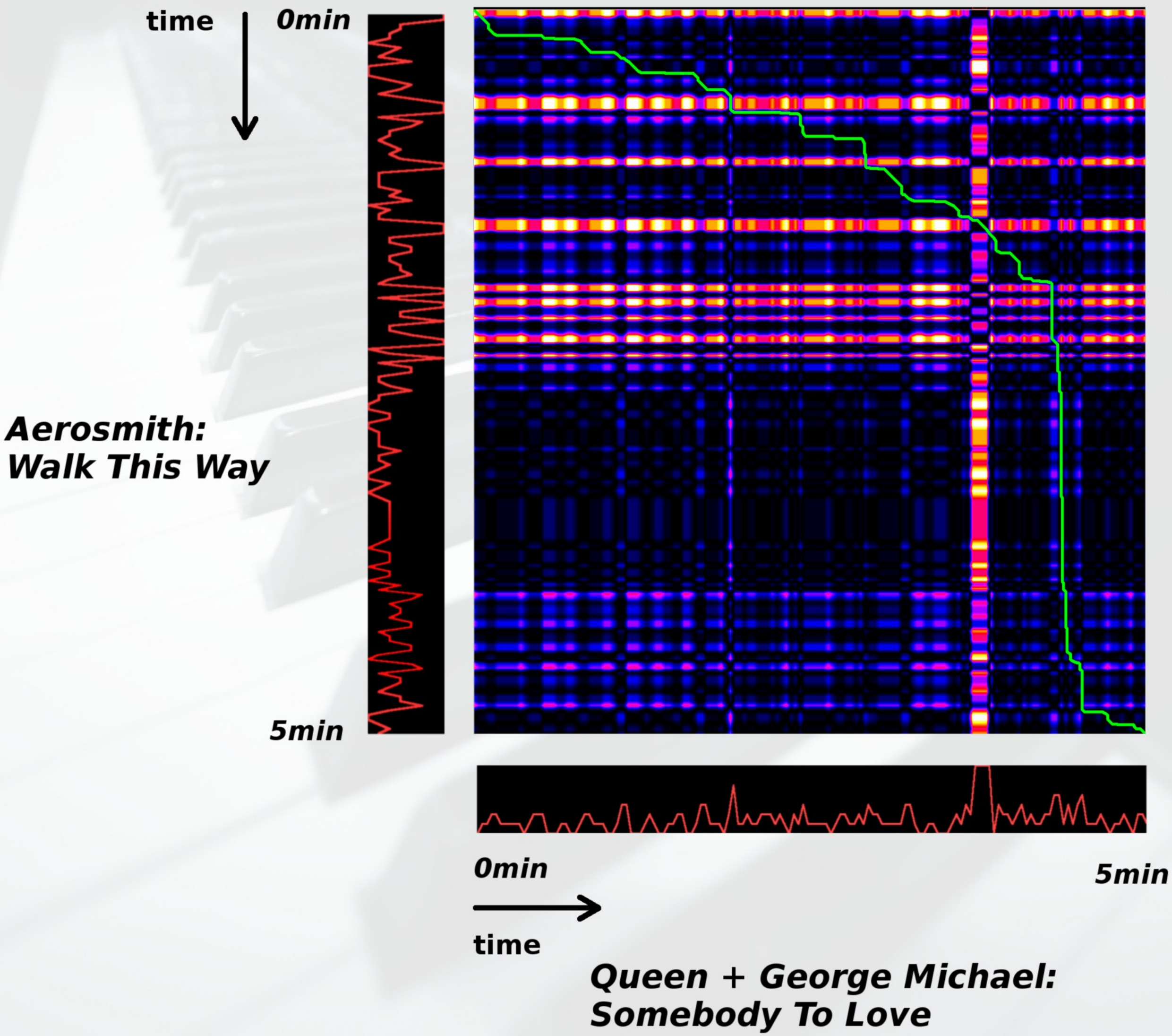
Dynamic Time Warping as alignment of 2 time series



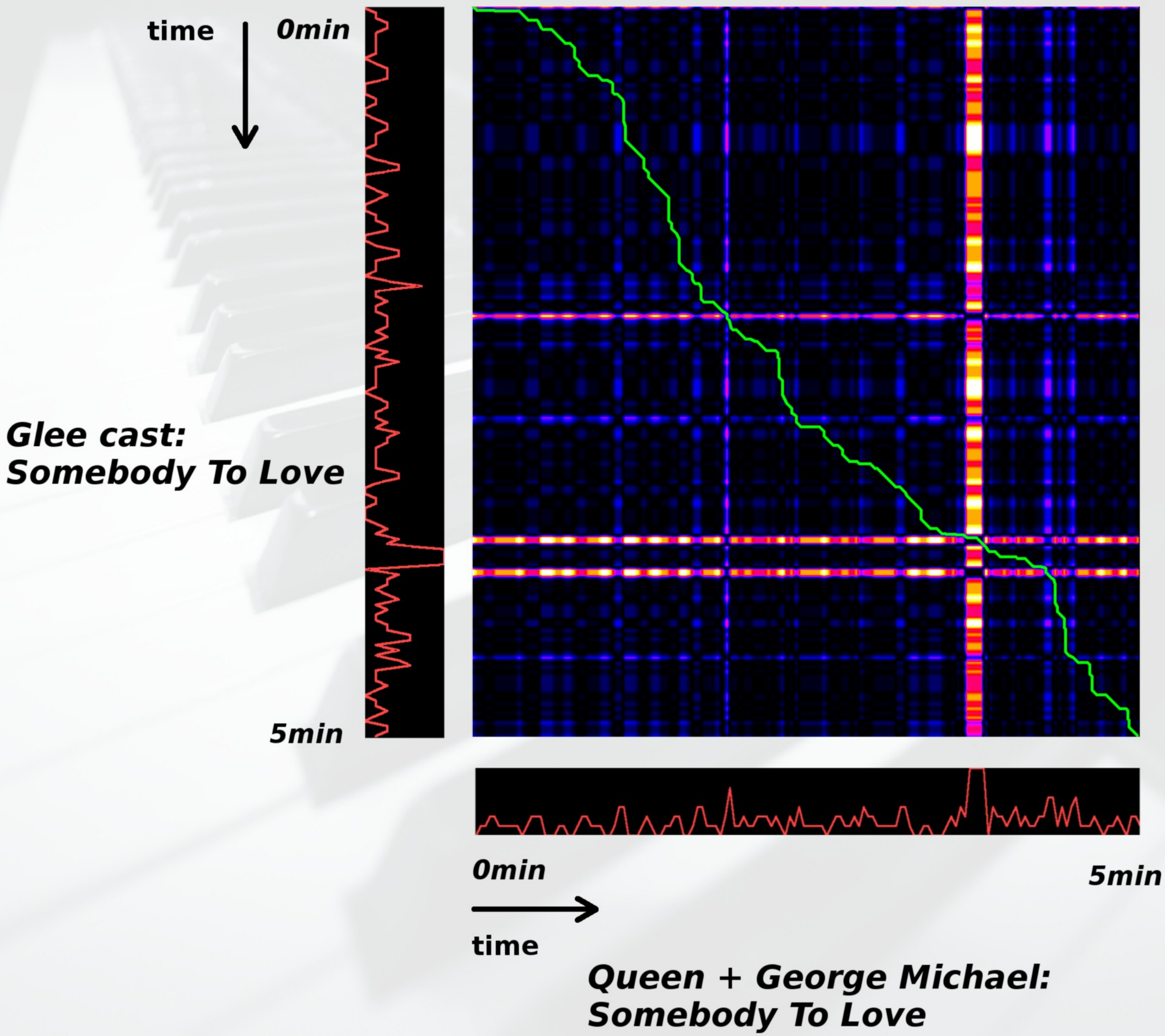
Dynamic Time Warping – Step 1: Distance matrix



Dynamic Time Warping – Step 2: Warping path



Dynamic Time Warping – Step 2: Warping path



Dynamic Time Warping Scores

Score 1 (accumulated distance)

$$DTW(S,T) = \min\left\{\sqrt{\sum_{l=1}^{L} w_l}\right.$$

Score 2 (accumulated distance / length)

$$DTW(S,T) = \min\left\{\frac{1}{L}\sqrt{\sum_{l=1}^{L} w_l}\right.$$

Score 3 (horizontal/vertical movements)

$$DTW_{sim}(S,T) = \frac{|S| + |T|}{|S| + l + |T| + u}$$

Score 4 (ratio to the maximal DTW)

$$DTW_{sim}(S,T) = 1 - \frac{\sqrt{DTW(S,T)}}{\sqrt{DTW_{max}(S,T)}}$$

Score 5 (ratio to the maximal DTW x differences)

$$DTW_{sim}(S,T) = \left(1 - \frac{\sqrt{DTW(S,T)}}{\sqrt{DTW_{max}(S,T)}}\right) \cdot \frac{min(|S|,|T|)}{max(|S|,|T|)}$$

Results of the comparison

	Raw chroma vectors	ChromaCD	ChordCD
AVERAGE RANK	**321 / 1000**	341 / 1000	403 / 1000
COMPUTATION TIME	550s	411s	**100s**

Comparison of mean average rank score on SecondHandSongs dataset.

	Raw chroma vectors	ChromaCD	ChordCD
AVERAGE RANK	**13.963 / 80**	21.538 / 80	26.688 / 80
COMPUTATION TIME	56s	51s	**25ms**

Comparison of mean average rank score on covers80 dataset.

Results of the comparison

Score	Raw chroma vectors
score (4.1)	**0.482**
score (4.2)	0.103
score (4.3)	0.417
score (4.4)	0.454
score (4.5)	0.082

MAP results for each DTW score and feature for covers80 dataset.

Score 1 (accumulated distance)

$$DTW(S,T) = \min\left\{\sqrt{\sum_{l=1}^{L} w_l}\right.$$

Score 2 (accumulated distance / length)

$$DTW(S,T) = \min\left\{\frac{1}{L}\sqrt{\sum_{l=1}^{L} w_l}\right.$$

Score 3 (horizontal/vertical movements)

$$DTW_{sim}(S,T) = \frac{|S| + |T|}{|S| + l + |T| + u}$$

Score 4 (ratio to the maximal DTW)

$$DTW_{sim}(S,T) = 1 - \frac{\sqrt{DTW(S,T)}}{\sqrt{DTW_{max}(S,T)}}$$

Score 5 (ratio to the maximal DTW x differences)

$$DTW_{sim}(S,T) = \left(1 - \frac{\sqrt{DTW(S,T)}}{\sqrt{DTW_{max}(S,T)}}\right) \cdot \frac{min(|S|,|T|)}{max(|S|,|T|)}$$

Results of the comparison

Score	Raw chroma vectors
score (4.1)	**0.482**
score (4.2)	0.103
score (4.3)	0.417
score (4.4)	0.454
score (4.5)	0.082

MAP results for each DTW score and feature for covers80 dataset.

Score 1 (accumulated distance)

$$DTW(S,T) = \min\left\{\sqrt{\sum_{l=1}^{L} w_l}\right.$$

Score 2 (accumulated distance / length)

$$DTW(S,T) = \min\left\{\frac{1}{L}\sqrt{\sum_{l=1}^{L} w_l}\right.$$

Score 3 (horizontal/vertical movements)

$$DTW_{sim}(S,T) = \frac{|S| + |T|}{|S| + l + |T| + u}$$

Score 4 (ratio to the maximal DTW)

$$DTW_{sim}(S,T) = 1 - \frac{\sqrt{DTW(S,T)}}{\sqrt{DTW_{max}(S,T)}}$$

Score 5 (ratio to the maximal DTW x differences)

$$DTW_{sim}(S,T) = \left(1 - \frac{\sqrt{DTW(S,T)}}{\sqrt{DTW_{max}(S,T)}}\right) \cdot \frac{min(|S|,|T|)}{max(|S|,|T|)}$$

Results of the comparison

Score	Raw chroma vectors
score (4.1)	**0.482**
score (4.2)	0.103
score (4.3)	0.417
score (4.4)	0.454
score (4.5)	0.082

MAP results for each DTW score and feature for covers80 dataset.

Score 1 (accumulated distance)

$$DTW(S,T) = \min\left\{\sqrt{\sum_{l=1}^{L} w_l}\right.$$

✓

Score 2 (accumulated distance / length)

$$DTW(S,T) = \min\left\{\frac{1}{L}\sqrt{\sum_{l=1}^{L} w_l}\right.$$

✗

Score 3 (horizontal/vertical movements)

$$DTW_{sim}(S,T) = \frac{|S| + |T|}{|S| + l + |T| + u}$$

Score 4 (ratio to the maximal DTW)

$$DTW_{sim}(S,T) = 1 - \frac{\sqrt{DTW(S,T)}}{\sqrt{DTW_{max}(S,T)}}$$

Score 5 (ratio to the maximal DTW x differences)

$$DTW_{sim}(S,T) = \left(1 - \frac{\sqrt{DTW(S,T)}}{\sqrt{DTW_{max}(S,T)}}\right) \cdot \frac{min(|S|,|T|)}{max(|S|,|T|)}$$

Results of the comparison

Score	Raw chroma vectors
score (4.1)	**0.482**
score (4.2)	0.103
score (4.3)	0.417
score (4.4)	0.454
score (4.5)	0.082

MAP results for each DTW score and feature for covers80 dataset.

Score 1 (accumulated distance)

$$DTW(S,T) = \min\left\{\sqrt{\sum_{l=1}^{L} w_l}\right.$$

Score 2 (accumulated distance / length)

$$DTW(S,T) = \min\left\{\frac{1}{L}\sqrt{\sum_{l=1}^{L} w_l}\right.$$

Score 3 (horizontal/vertical movements)

$$DTW_{sim}(S,T) = \frac{|S|+|T|}{|S|+l+|T|+u}$$

Score 4 (ratio to the maximal DTW)

$$DTW_{sim}(S,T) = 1 - \frac{\sqrt{DTW(S,T)}}{\sqrt{DTW_{max}(S,T)}}$$

Score 5 (ratio to the maximal DTW x differences)

$$DTW_{sim}(S,T) = \left(1 - \frac{\sqrt{DTW(S,T)}}{\sqrt{DTW_{max}(S,T)}}\right) \cdot \frac{min(|S|,|T|)}{max(|S|,|T|)}$$

Results of the comparison

Score	Raw chroma vectors
score (4.1)	**0.482**
score (4.2)	0.103
score (4.3)	0.417
score (4.4)	0.454
score (4.5)	0.082

MAP results for each DTW score and feature for covers80 dataset.

Score 1 (accumulated distance)

$$DTW(S,T) = \min\left\{\sqrt{\sum_{l=1}^{L} w_l}\right.$$

Score 2 (accumulated distance / length)

$$DTW(S,T) = \min\left\{\frac{1}{L}\sqrt{\sum_{l=1}^{L} w_l}\right.$$

Score 3 (horizontal/vertical movements)

$$DTW_{sim}(S,T) = \frac{|S| + |T|}{|S| + l + |T| + u}$$

Score 4 (ratio to the maximal DTW)

$$DTW_{sim}(S,T) = 1 - \frac{\sqrt{DTW(S,T)}}{\sqrt{DTW_{max}(S,T)}}$$

Score 5 (ratio to the maximal DTW x differences)

$$DTW_{sim}(S,T) = \left(1 - \frac{\sqrt{DTW(S,T)}}{\sqrt{DTW_{max}(S,T)}}\right) \cdot \frac{min(|S|,|T|)}{max(|S|,|T|)}$$

Results of the comparison

Score	Raw chroma vectors
score (4.1)	**0.482**
score (4.2)	0.103
score (4.3)	0.417
score (4.4)	0.454
score (4.5)	0.082

MAP results for each DTW score and feature for covers80 dataset.

Score 1 (accumulated distance)

$$DTW(S,T) = \min\left\{\sqrt{\sum_{l=1}^{L} w_l}\right.$$

Score 2 (accumulated distance / length)

$$DTW(S,T) = \min\left\{\frac{1}{L}\sqrt{\sum_{l=1}^{L} w_l}\right.$$

Score 3 (horizontal/vertical movements)

$$DTW_{sim}(S,T) = \frac{|S|+|T|}{|S|+l+|T|+u}$$

Score 4 (ratio to the maximal DTW)

$$DTW_{sim}(S,T) = 1 - \frac{\sqrt{DTW(S,T)}}{\sqrt{DTW_{max}(S,T)}}$$

Score 5 (ratio to the maximal DTW x differences)

$$DTW_{sim}(S,T) = \left(1 - \frac{\sqrt{DTW(S,T)}}{\sqrt{DTW_{max}(S,T)}}\right) \cdot \frac{min(|S|,|T|)}{max(|S|,|T|)}$$

Conclusion and Future work

1. We have compared harmony features for CSI
    - For future work: „raw“ chroma features show better accuracy
2. We have compared 5 DTW scores for CSI
    - 3 scores stand strong for CSI, 2 lack the accuracy
3. We have proposed harmony fingerprints
    - Can save time for the big data set pre-processing
4. New Data sets for CSI
    - Future work (ISMIR) – karaoke dataset

Thank you for your attention

In depth: Feature 2 (Chroma Complexity Distance)

Chroma Complexity Distance (ChromaCD)

$$\delta(x, y) = \sum_{i=1}^{12} |w(x)_i x_i - w(y)_i y_i|$$

- *x, y* : 12-dimensional chroma vectors
- *w(x), w(y)*: 12-dimensional weight vectors (the context of *x* to *y* transition)
- The weight vectors are dependent on the local chords and key

- Represents how complex is the chroma vector distance in the tonal context

In depth: Feature 2 (Chroma Complexity Distance)

$$\delta(x,y) = \sum_{i=1}^{12} |w(x)_i x_i - w(y)_i y_i|$$

WEIGHT VECTORS

key: C major

PITCH CLASS		chord: C major w(x)	chord: C major w(y)	chord: G major w(z)
11	B	1	1	0
10	A#	2	2	2
9	A	1	1	1
8	G#	2	2	2
7	G	0	0	0
6	F#	2	2	2
5	F	1	1	1
4	E	0	0	1
3	D#	2	2	2
2	D	1	1	0
1	C#	2	2	2
0	C	0	0	1

Model of Chord Complexity Distance

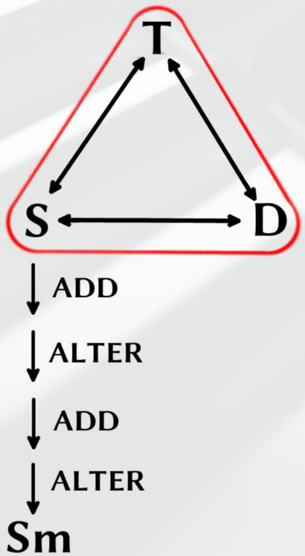
- Similar to formal grammars
- 2 successive chords in a musical piece

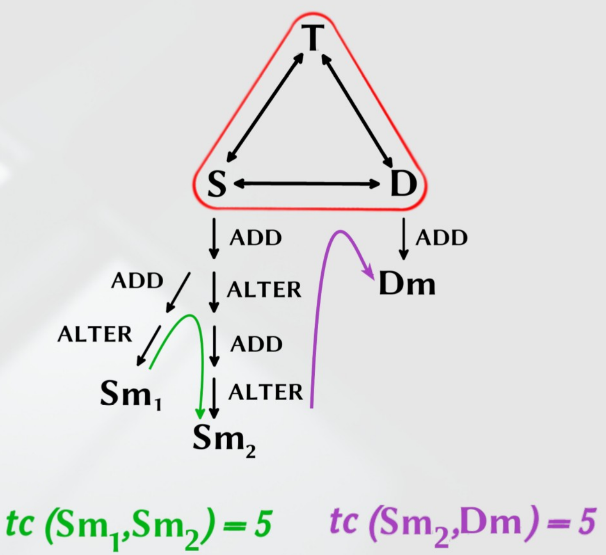
  =

  start strign & end string
- Example:

$$\text{CEG} \xrightarrow{\text{ADD}} \text{CEFG} \xrightarrow{\text{ALTER}} \text{CEF\#G} \xrightarrow{\text{ALTER}} \text{CEF\#G\#}$$

- 2 rules applicable on sentential form:
  - ADD – adds a new tone
  - ALTER – alters the tone

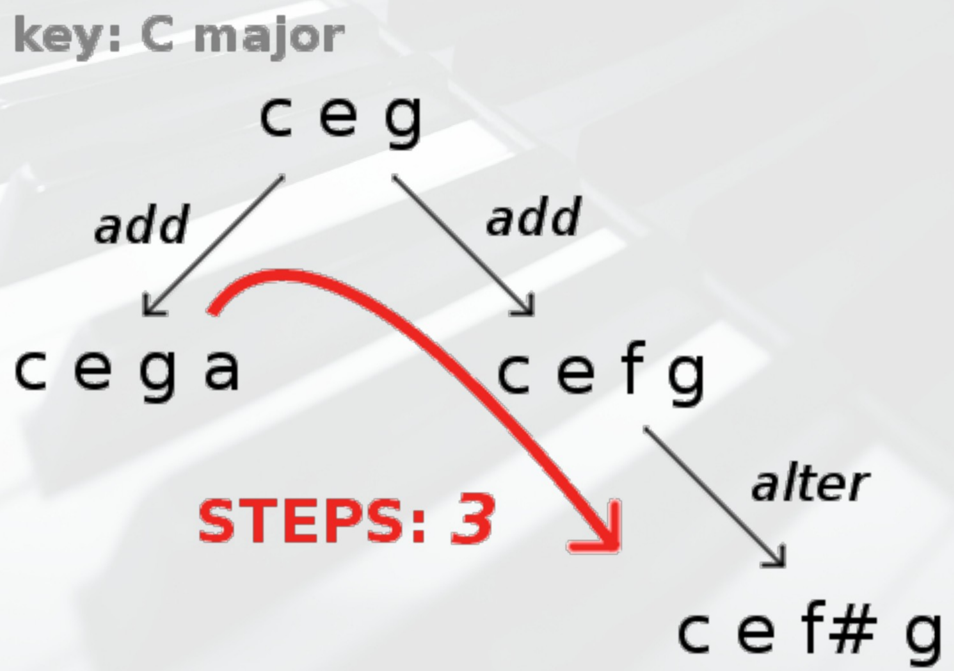




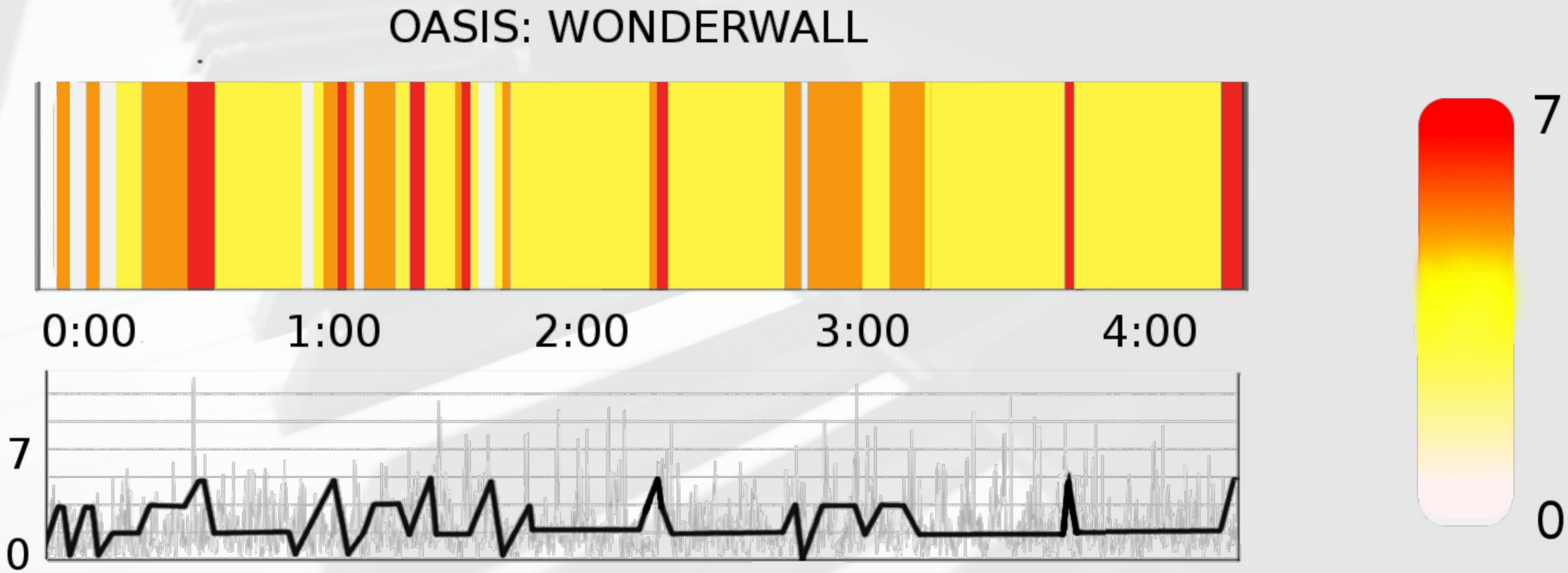
In depth: Feature 3 (Chord Complexity Distance)

Chord Complexity Distance (ChordCD)

(a variation of Edit Distance)



In depth: Feature 3 (Chord Complexity Distance)



In depth: Feature 3 (Chord Complexity Distance)

There are many different Chord Distances, ChordCD is only one possibilty

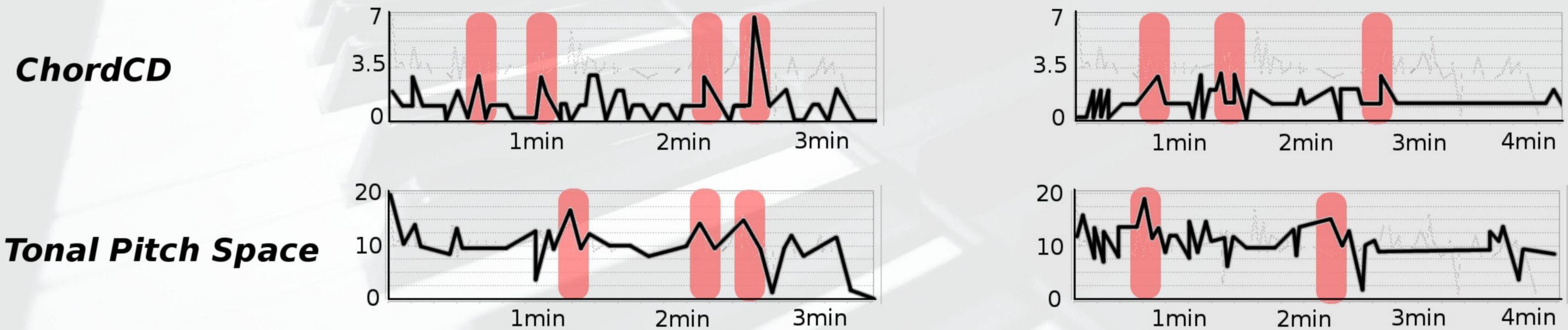
The most common one is: **Tonal Pitch Space** (Fred Lerdahl)

(a)	0												(0)
(b)	0							7					(0)
(c)	0				4			7					(0)
(d)	0		2		4	5		7		9		11	(0)
(e)	0	1	2	3	4	5	6	7	8	9	10	11	(0)

TPS of C major chord in a C major key

In depth: Feature 3 (Chord Complexity Distance)

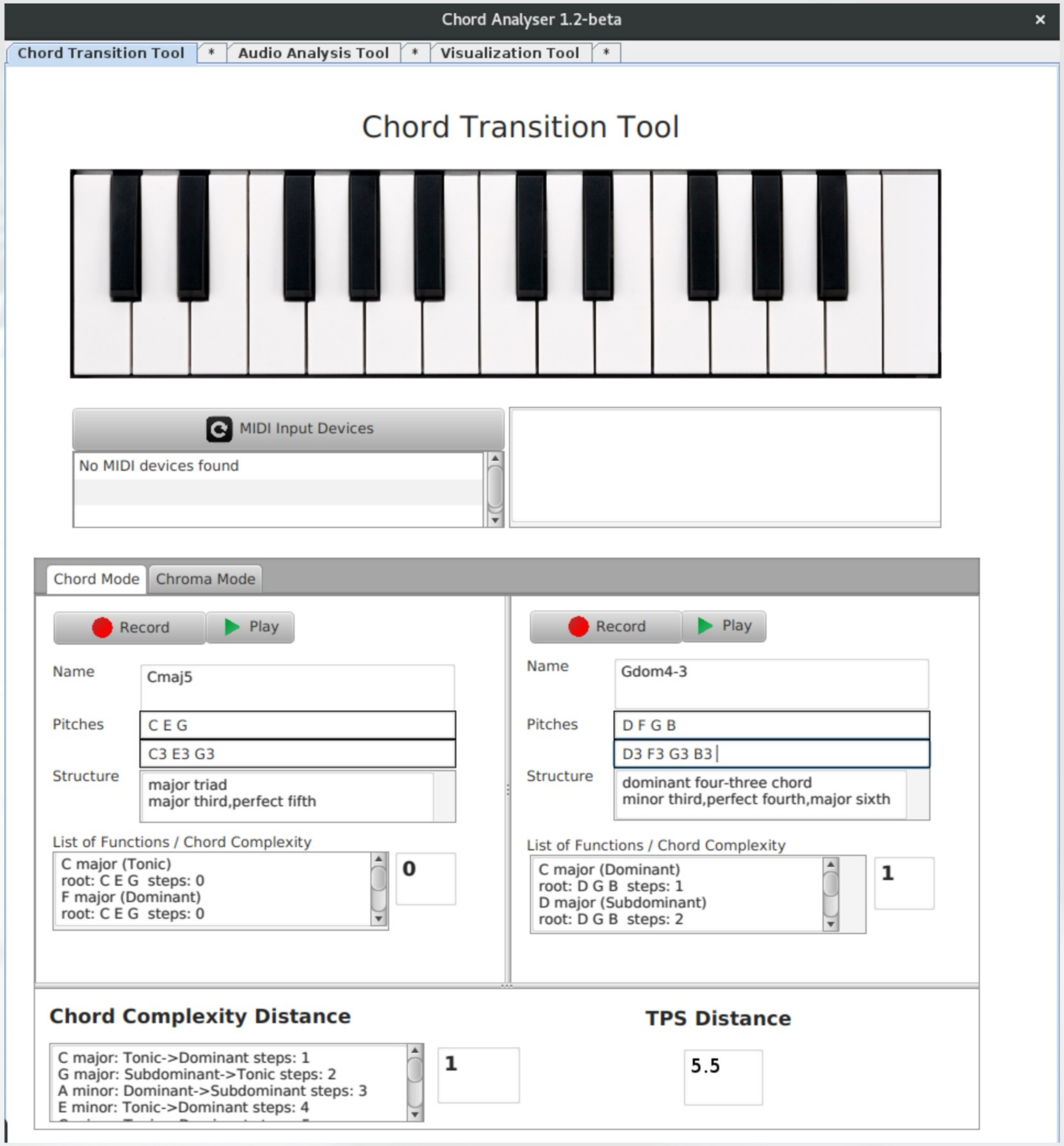
ChordCD vs Tonal Pitch Space



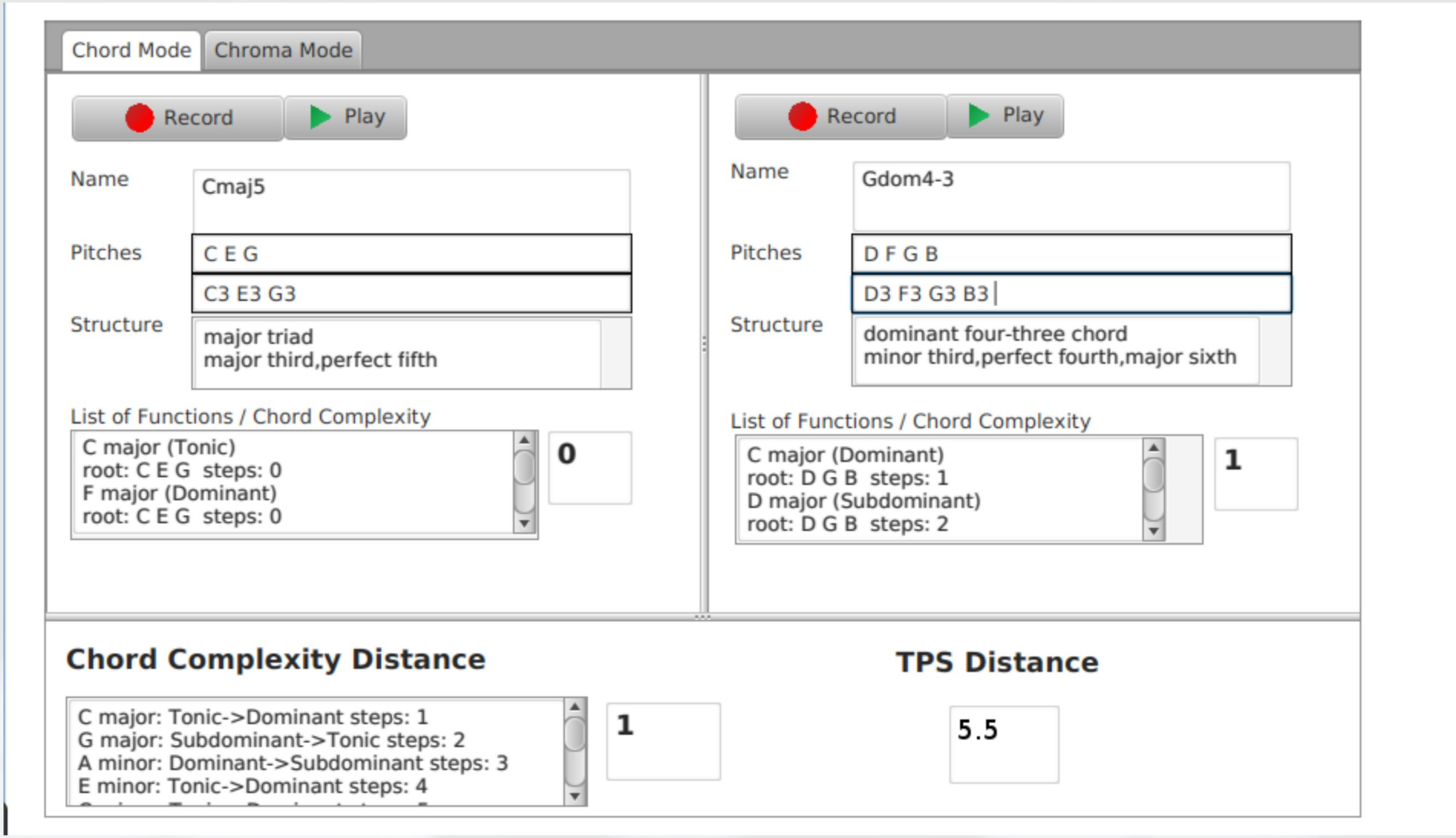
Bastian Baker: Hallelujah

Oasis: Wonderwall

harmony-analyser.org

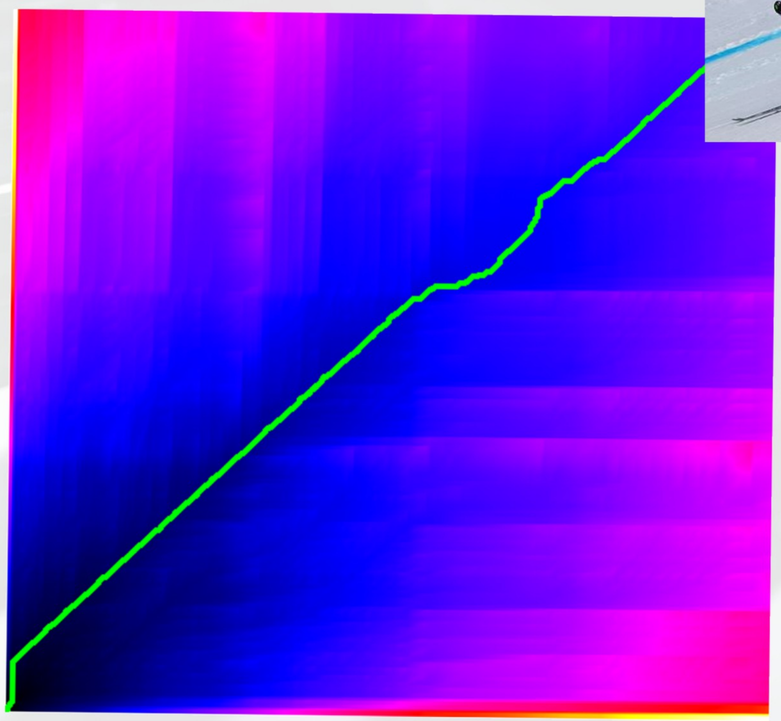


harmony-analyser.org



Dynamic Time Warping – Step 2: Warping path

Analogy – „Downhill skiing“

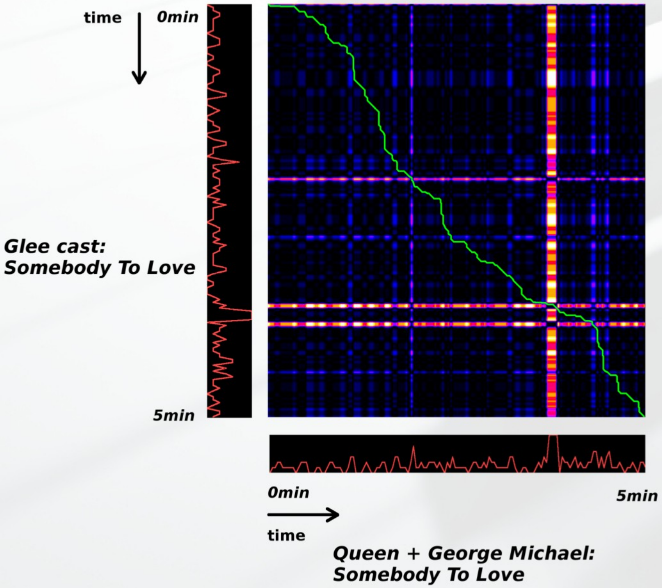


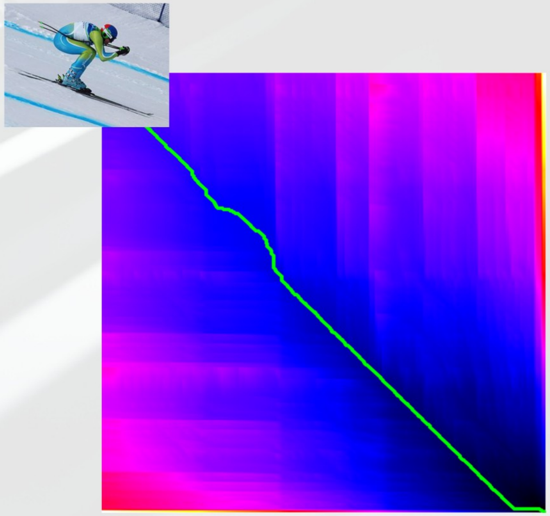
Dynamic Time Warping Scores

Score 1 (accumulated distance)

$$DTW(S,T) = \min\left\{\sqrt{\sum_{l=1}^{L} w_l}\right.$$

L = length of the path; w_l = distance of the step l



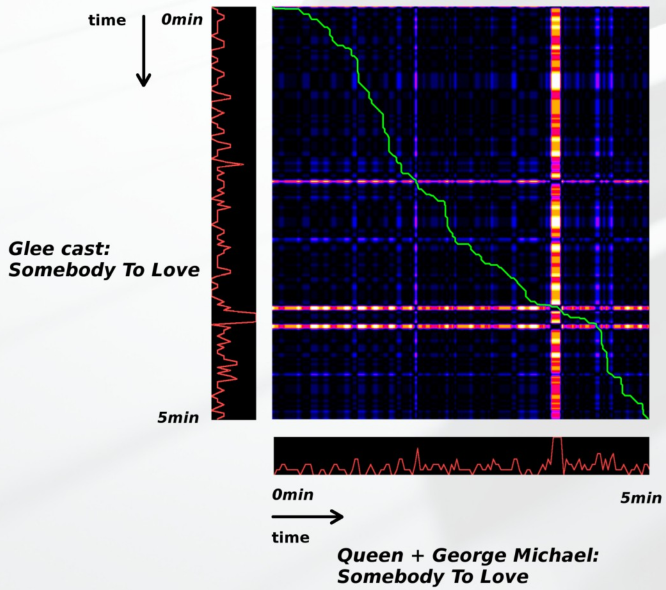


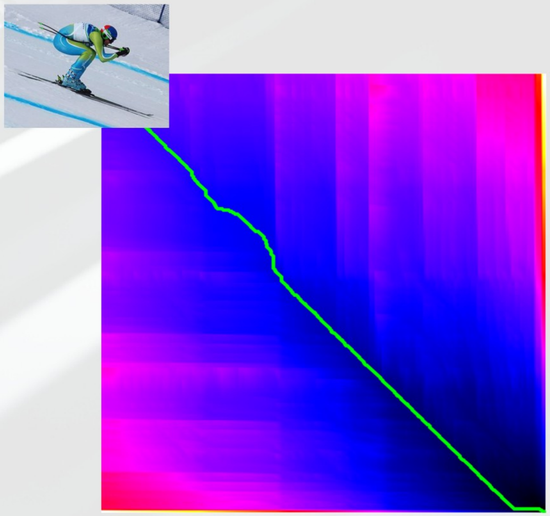
Dynamic Time Warping Scores

Score 2 (accumulated distance / path length)

$$DTW(S,T) = \min \left\{ \frac{1}{L} \sqrt{\sum_{l=1}^{L} w_l} \right.$$

L = length of the path; w_l = distance of the step l



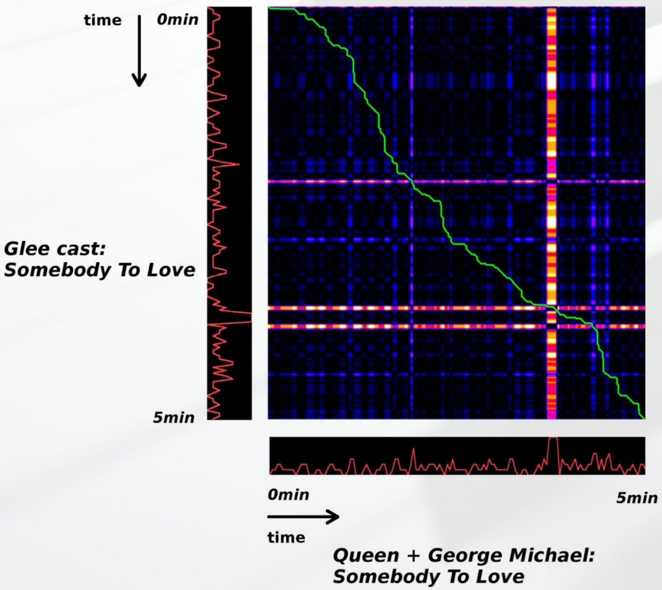


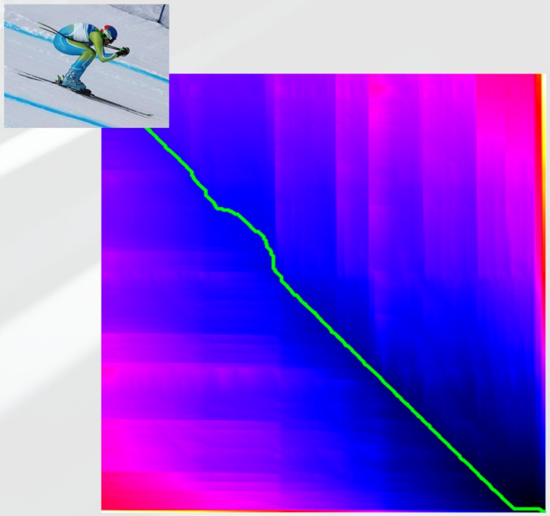
Dynamic Time Warping Scores

Score 3 (horizontal / vertical movements)

$$DTW_{sim}(S,T) = \frac{|S| + |T|}{|S| + l + |T| + u}$$

|S|, |T| = length of series; l, u = number of horizontal / vertical movements



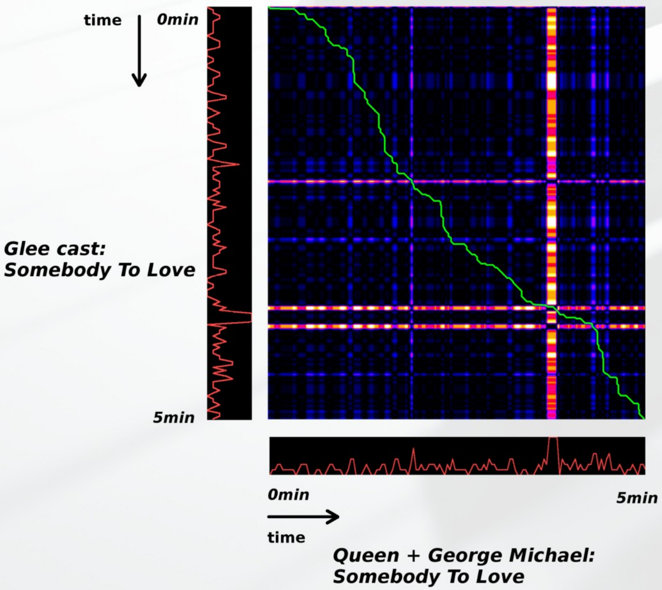


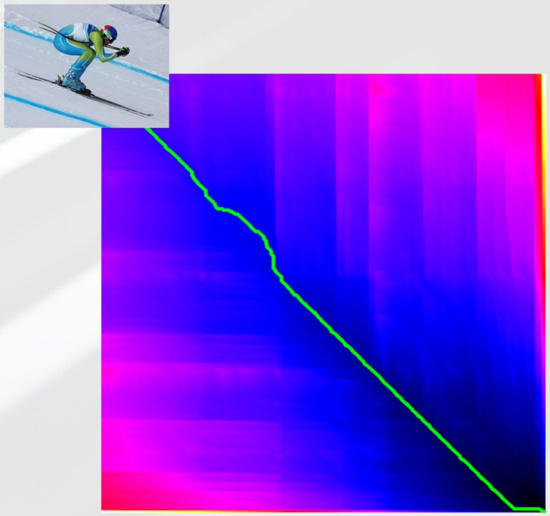
Dynamic Time Warping Scores

Score 4 (ratio to the maximal DTW)

$$DTW_{sim}(S,T) = 1 - \frac{\sqrt{DTW(S,T)}}{\sqrt{DTW_{max}(S,T)}}$$

$DTW_{max}(S,T)$ = maximal theoreticly possible DTW, when S and T are different in all points





Dynamic Time Warping Scores

Score 5 (ratio to the maximal DTW x differences ratio)

$$DTW_{sim}(S,T) = \left(1 - \frac{\sqrt{DTW(S,T)}}{\sqrt{DTW_{max}(S,T)}}\right) \cdot \frac{min(|S|,|T|)}{max(|S|,|T|)}$$

$DTW_{max}(S,T)$ = maximal theoreticly possible DTW, when S and T are different in all points