Mining Melodic Patterns in Large Audio Collections of Indian Art Music

Sankalp Gulati[*], Joan Serrà[†], Vignesh Ishwar[*] and Xavier Serra[*]

[*]*Music Technology Group, Universitat Pompeu Fabra, Barcelona, Spain*

[†]*Artificial Intelligence Research Institute (IIIA-CSIC), Bellaterra, Barcelona, Spain*

SITIS – 2014 (MIRA), Marrakech, Morocco

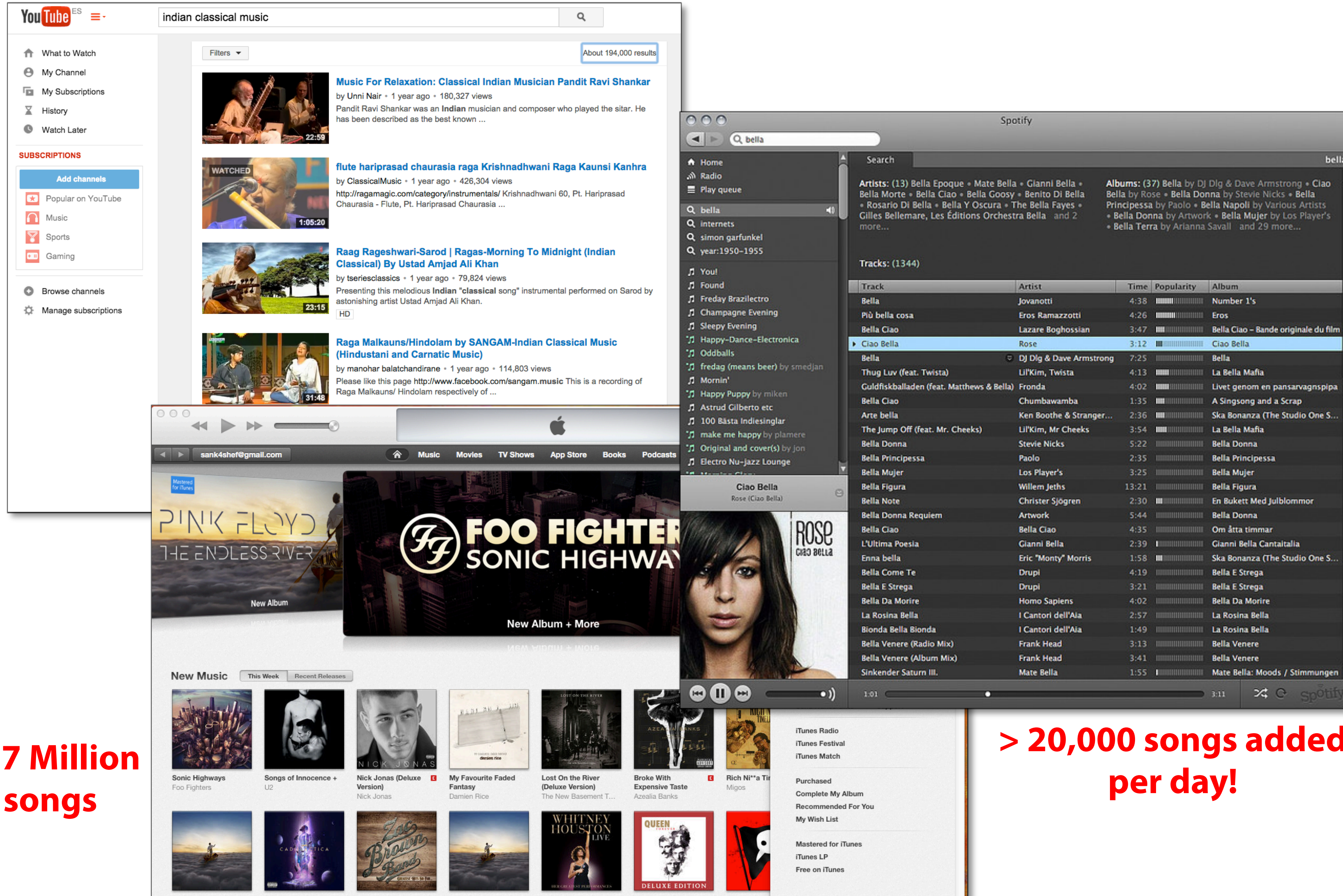
upf. Music Technology Group

compmusic

INSTITUT D'INVESTIGACIÓ EN INTEL.LIGÈNCIA ARTIFICIAL - IIIA

Multimedia Content – Music Recordings

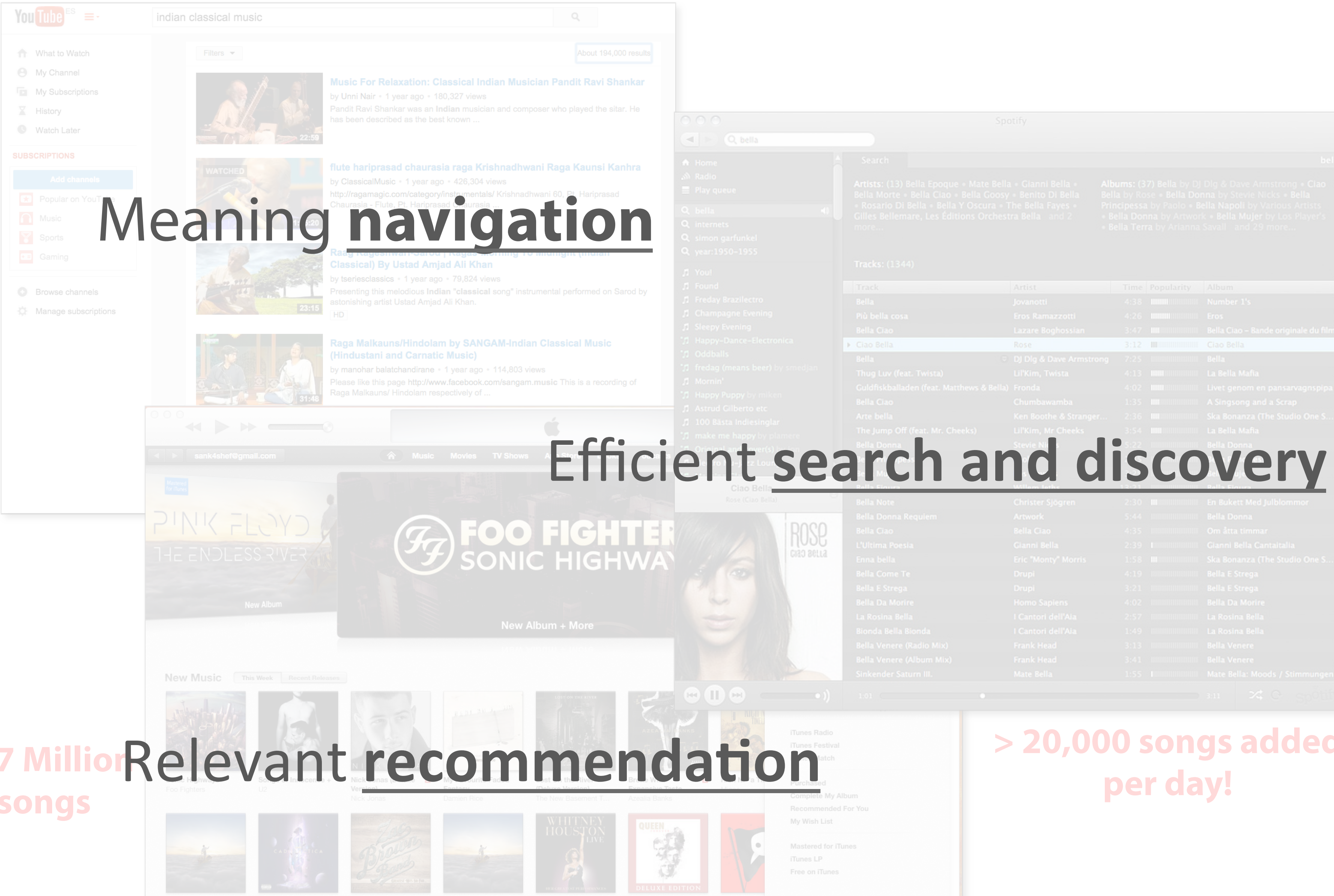
compmusic



>37 Million songs

> 20,000 songs added per day!

Multimedia Content – Audio Music



Meaning **navigation**

Efficient **search and discovery**

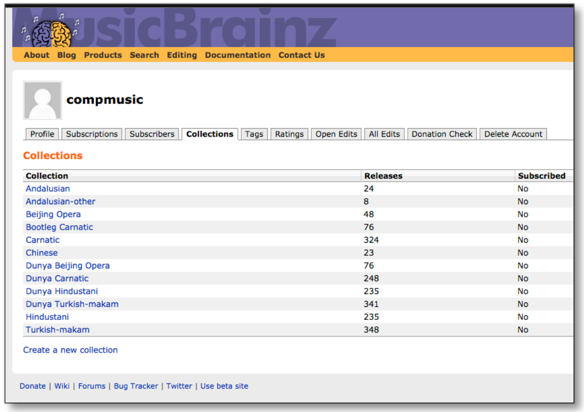
Relevant **recommendation**

>37 Million songs

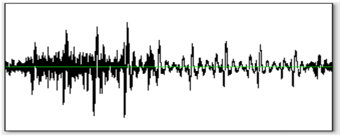
> 20,000 songs added per day!

Utilizing Available Data

- **Editorial Metadata**

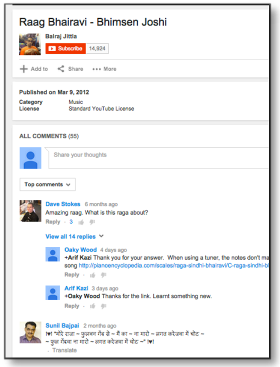


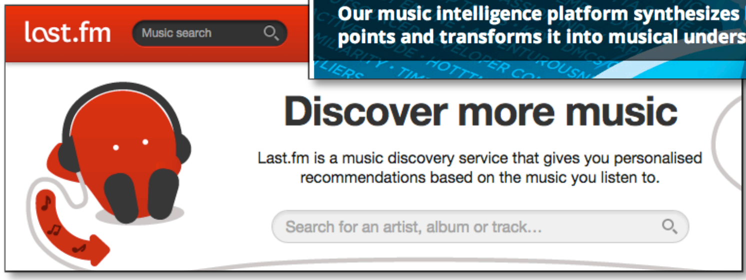
- **Audio Content**



ESSENTIA

- **Context**

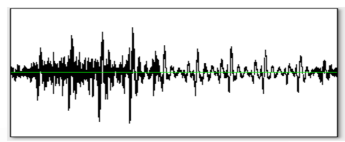




Utilizing Available Data

- Editorial Metadata

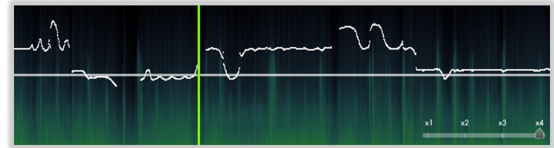
- **Audio Content**

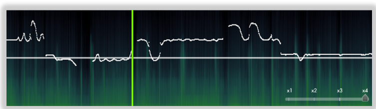


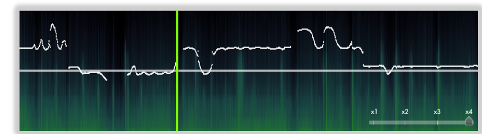
ESSENTIA

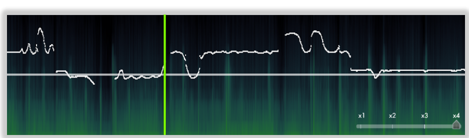
- Context

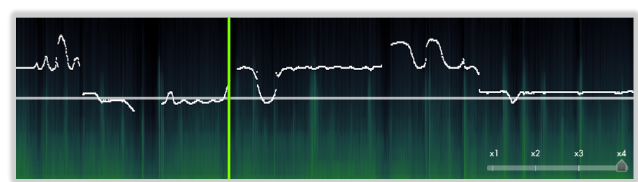
Objective



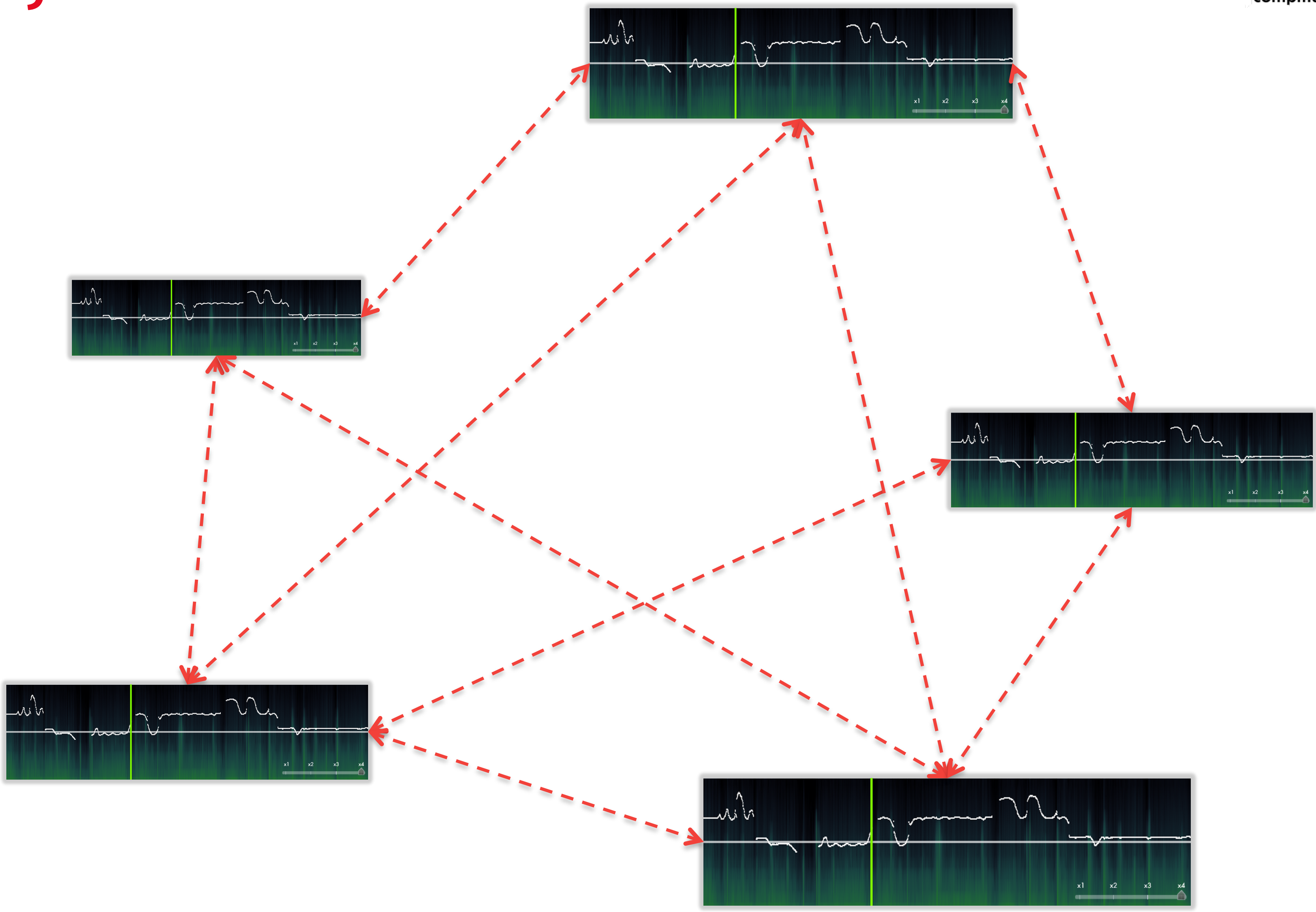




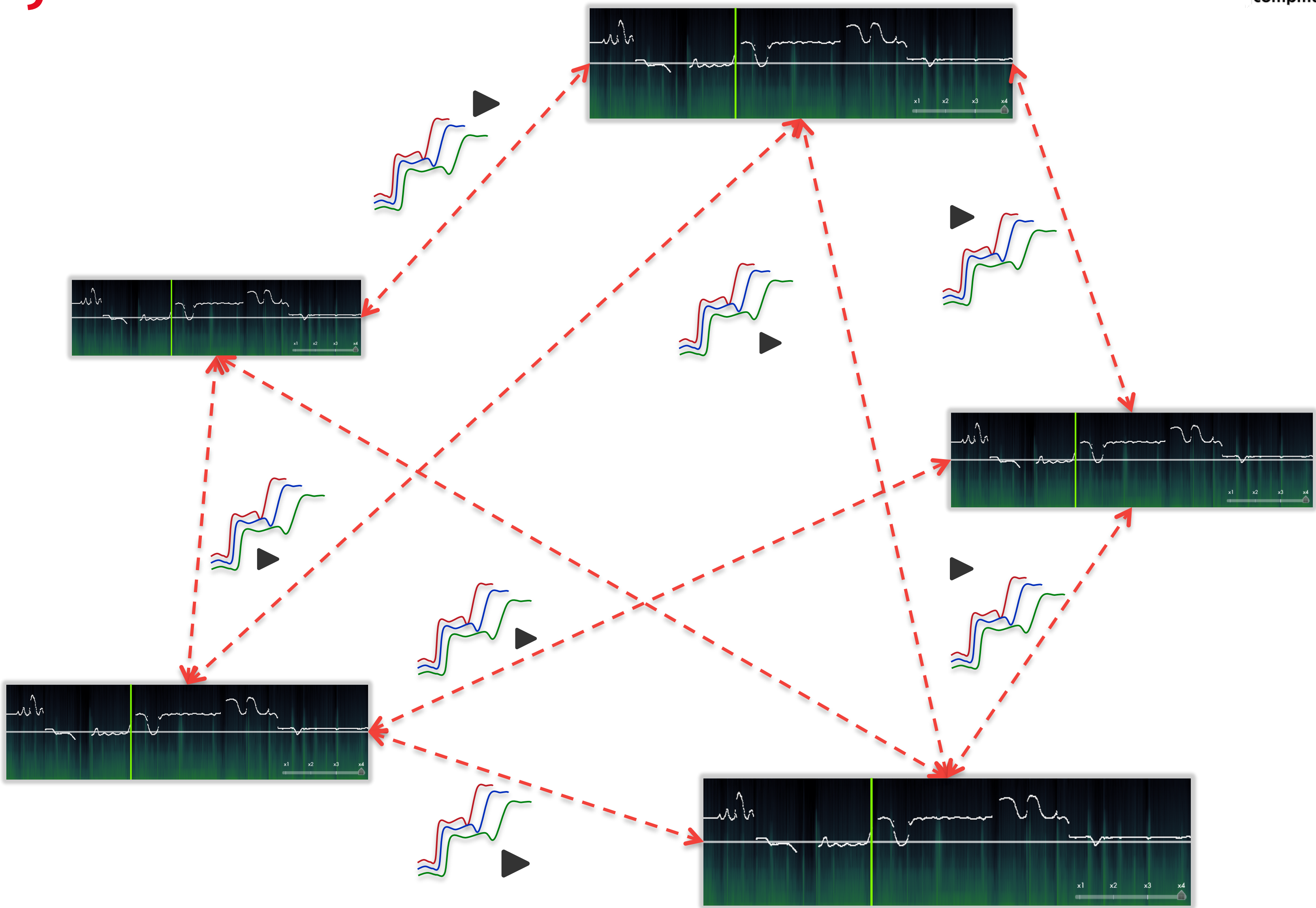




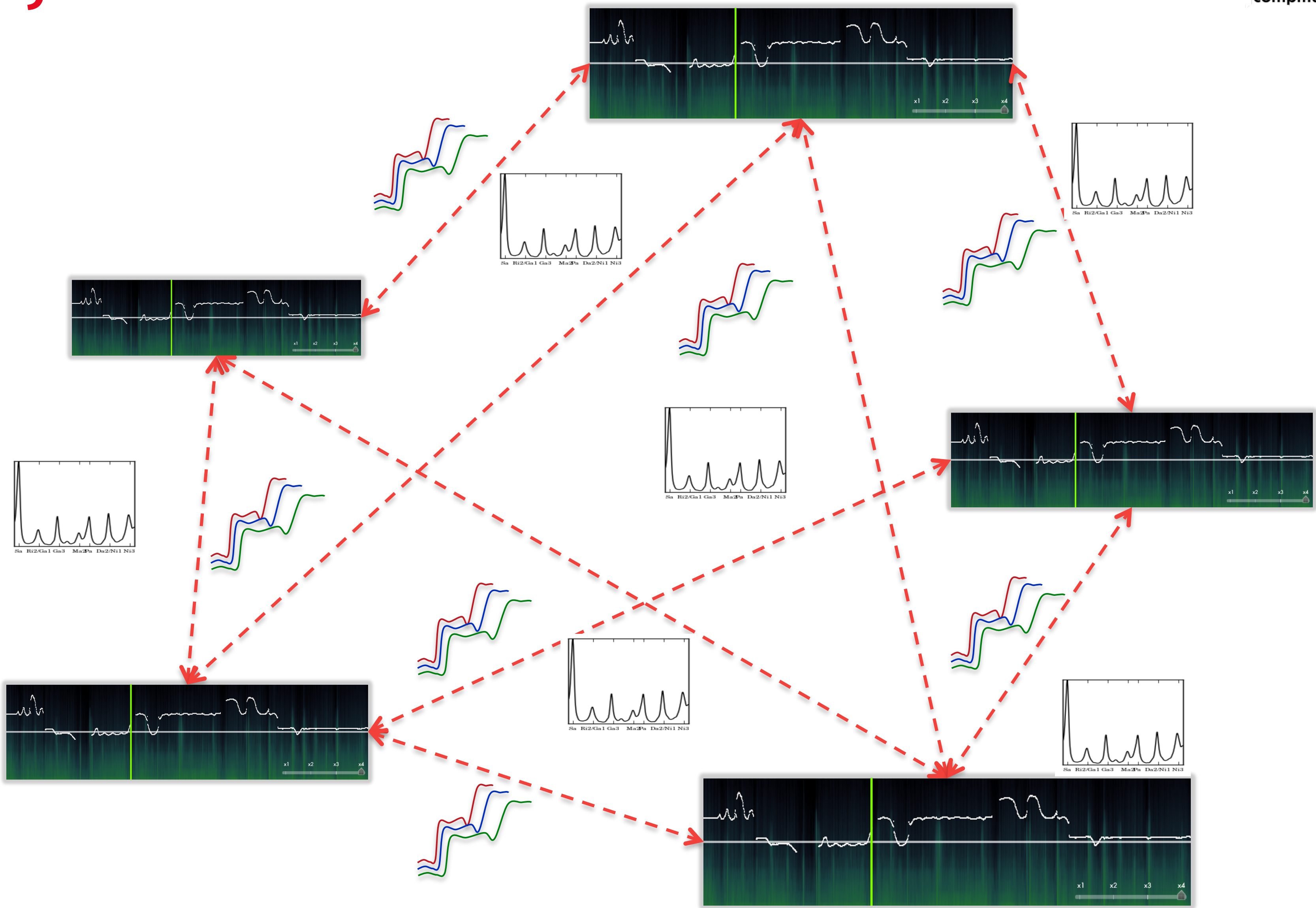
Objective



Objective



Objective

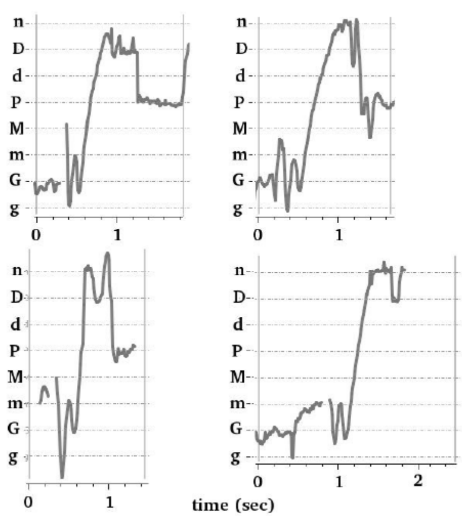


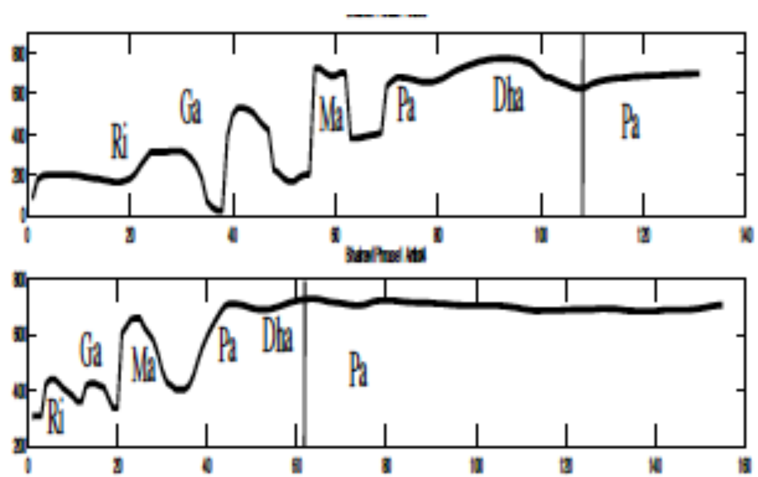
Music Collection

- Indian Art Music
  - Carnatic music
  - Hindustani music
- Rāga: melodic framework
  - Swar (note)
  - Swar prominence/role (Vādi, Samvādi, Nyās, Grah)
  - Pakads (characteristic melodic patterns)

Music Collection – Carnatic music

- Why this music tradition?
  - Signal processing steps relatively easier
  - Main challenge due to melodic characteristics and improvisation
  - Melodic patterns cues to rāga identification





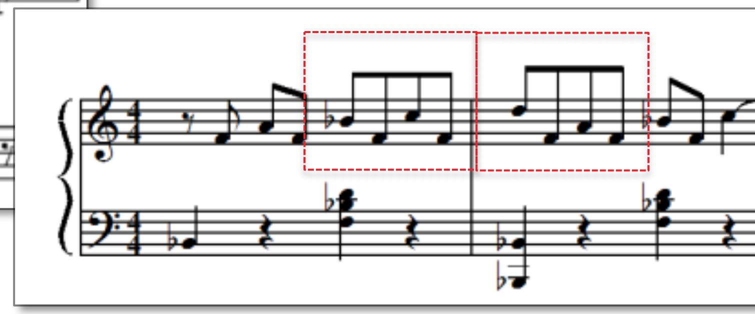
Music Collection – Carnatic music

- Dataset details
  - **1764** commercially available **polyphonic audio** recordings (subset of CompMusic collection)
  - **365** hours of music (> 50 billion audio samples)
  - Diverse dataset – gender, #ragas, #compositions…

compmusic

X. Serra, "Creating research corpora for the computational study of music: the case of the Compmusic project," in Proc. of the 53rd AES International Conference on Semantic Audio, London, Jan. 2014.

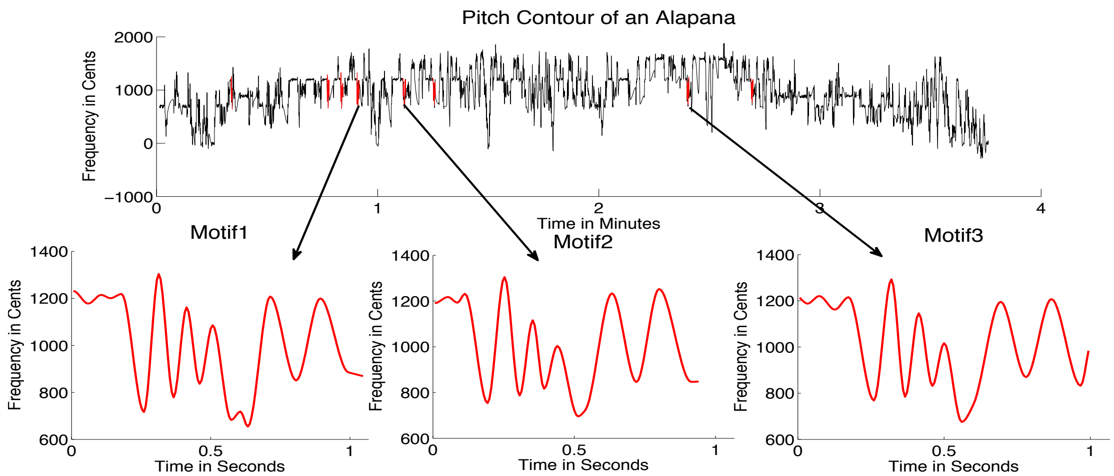
Previous Work – Symbolic Data



- Hungarian, Slovak, French, Sicilian, Bulgarian and Appalachian Folk Melodies - (Juhász, 2006)
- Cretan, Nova scotia and Essen Folk Melodies – (Conklin and Anagnostopoulou, 2010, 2006)
- Tunisian modal music -(Lartillot & Ayari, 2006).

Juhász, Z. (2006, June). A systematic comparison of different European folk music traditions using self-organizing maps. Journal of New Music Research, 35(2), 95–112.
Conklin, D., & Anagnostopoulou, C. (2006). Segmental pattern discovery in music. INFORMS Journal on Computing, 18(3), 285–293.
Lartillot, O., & Ayari, M. (2006). Motivic pattern extraction in music, and application to the study of Tunisian modal music. South African Computer Journal, 36, 16–28.

Previous Work – Audio Data (IAM)



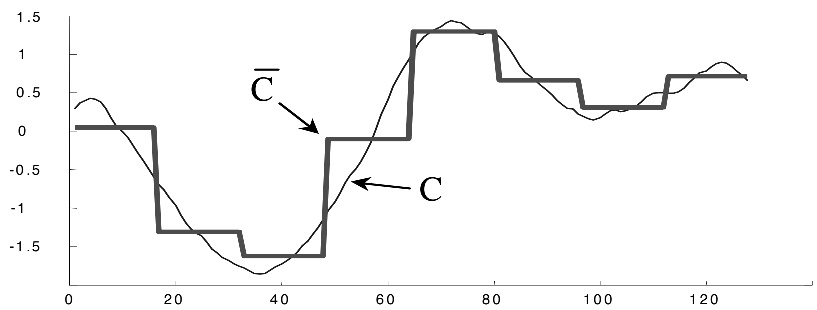
- Spotting motifs in Carnatic Music
- Detecting melodic motifs in Hindustani music
- Classification of melodic motifs
- Discovering typical melodic motifs from one-liners

V. Ishwar, S. Dutta, A. Bellur, and H. Murthy, "Motif spotting in an Alapana in Carnatic music," in *Proc. of Int. Conf. on Music Information Retrieval (ISMIR)*, 2013, pp. 499–504.

J. C. Ross, T. P. Vinutha, and P. Rao, "Detecting melodic motifs from audio for Hindustani classical music," in *Proc. of Int. Conf. on Music Information Retrieval (ISMIR)*, 2012, pp. 193–198.

P. Rao, J. C. Ross, K. K. Ganguli, V. Pandit, V. Ishwar, A. Bellur, and H. A. Murthy, "Classification of melodic motifs in raga music with time-series matching," *Journal of New Music Research*, vol. 43, no. 1, pp. 115–131, Jan. 2014.

Previous Work – Time Series Analysis



- Motif Discovery approaches
  - Exact, Online, Probabilistic discovery
- Lower bounding (indexing techniques)
  - Symbolic representation - SAX
  - Lower bounds on dynamic time warping (DTW)

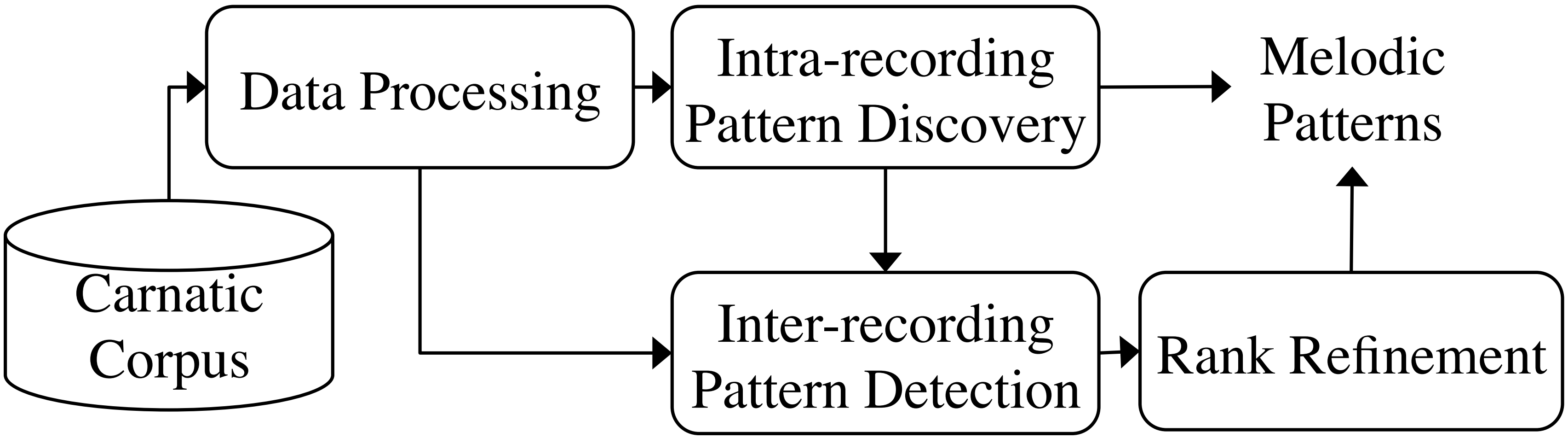
A. Mueen, E. Keogh, Q. Zhu, S. Cash, and B. Westover, "Exact discovery of time series motifs," in Proc. of SIAM Int. Con. on Data Mining (SDM), 2009, pp. 1–12.

B. Chiu, E. Keogh, and S. Lonardi, "Probabilistic discovery of time series motifs," Proc. ninth ACM SIGKDD Int. Conf. Knowl. Discov. data Min. - KDD '03, p. 493, 2003.

J. Lin, E. Keogh, S. Lonardi, and B. Chiu, "A symbolic representation of time series, with implications for streaming algorithms," Proc. 8th ACM SIGMOD Work. Res. issues data Min. Knowl. Discov. - DMKD '03, p. 2, 2003.

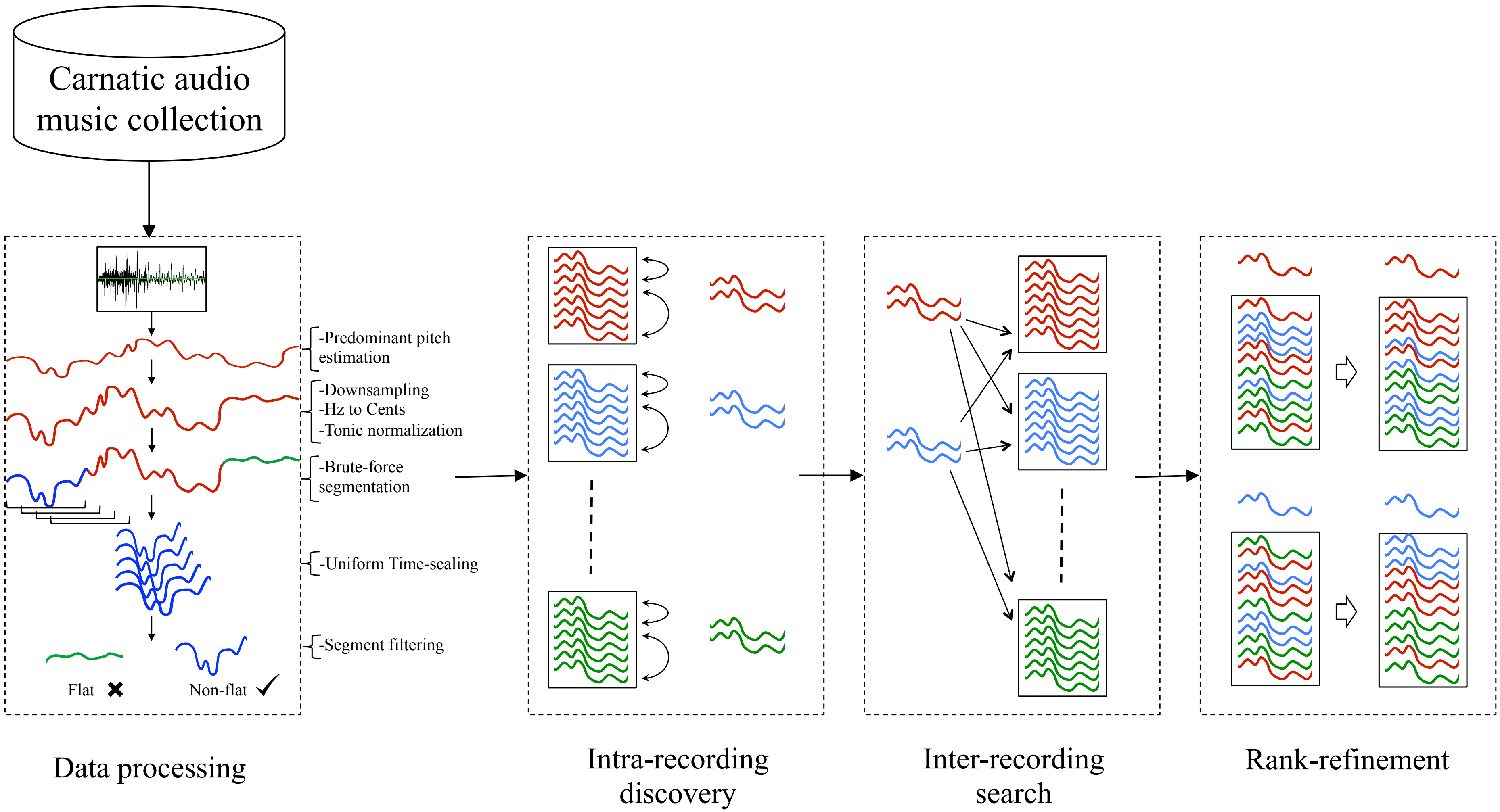
T. Rakthanmanon, B. Campana, A. Mueen, G. Batista, B. Westover, Q. Zhu, J. Zakaria, and E. Keogh, "Addressing big data time series: mining trillions of time series subsequences under dynamic time warping," ACM Transactions on Knowledge Discovery from Data (TKDD), vol. 7, no. 3, pp. 10:1–10:31, Sep. 2013.

Proposed Methodology

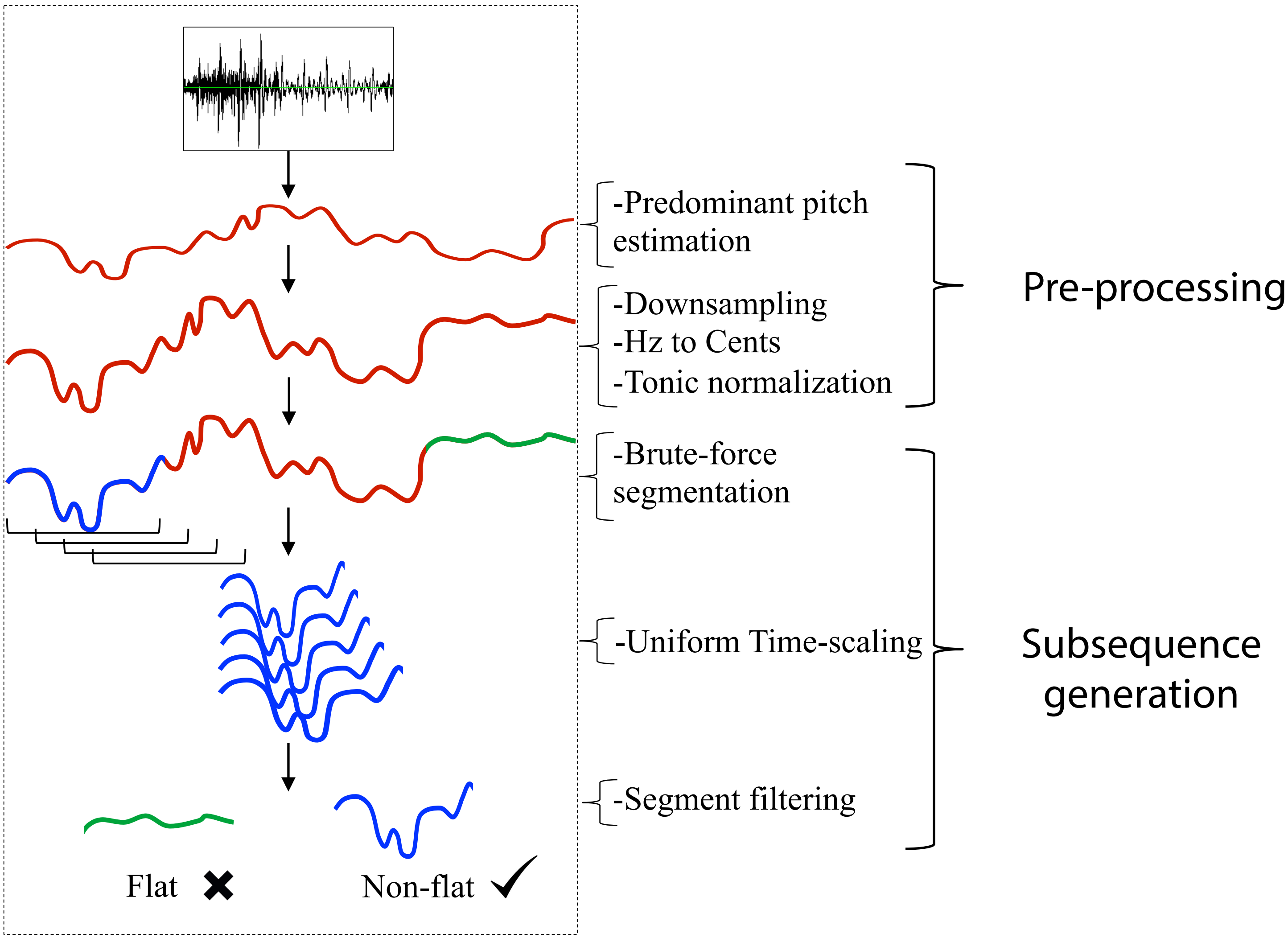


Block diagram of the proposed approach

Proposed Methodology

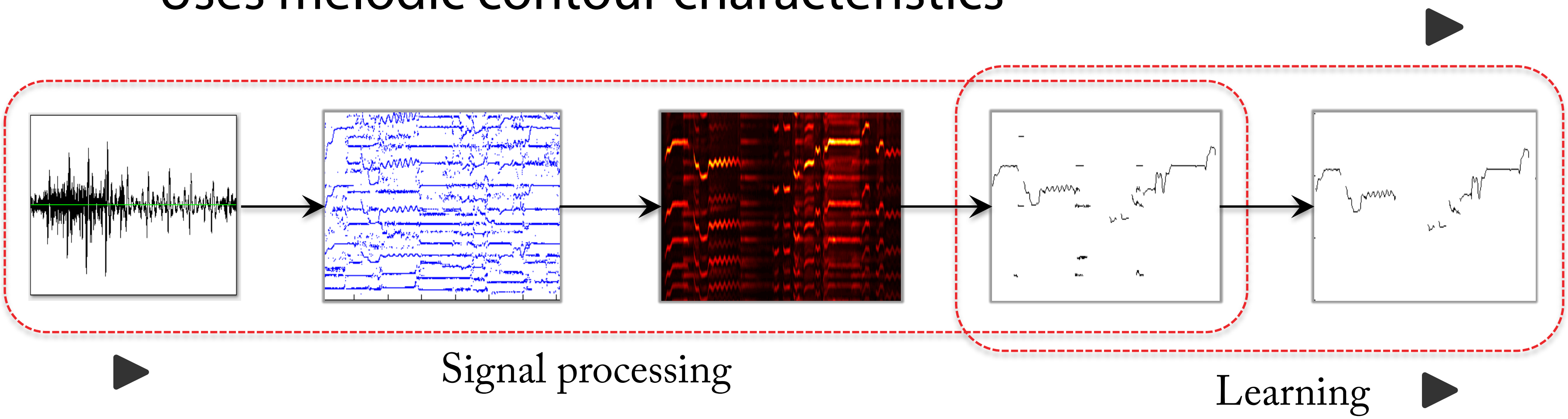


Data processing



Data processing: Pre-processing

- Predominant pitch estimation – Melodia
  - Designed for polyphonic music audio
  - Uses melodic contour characteristics

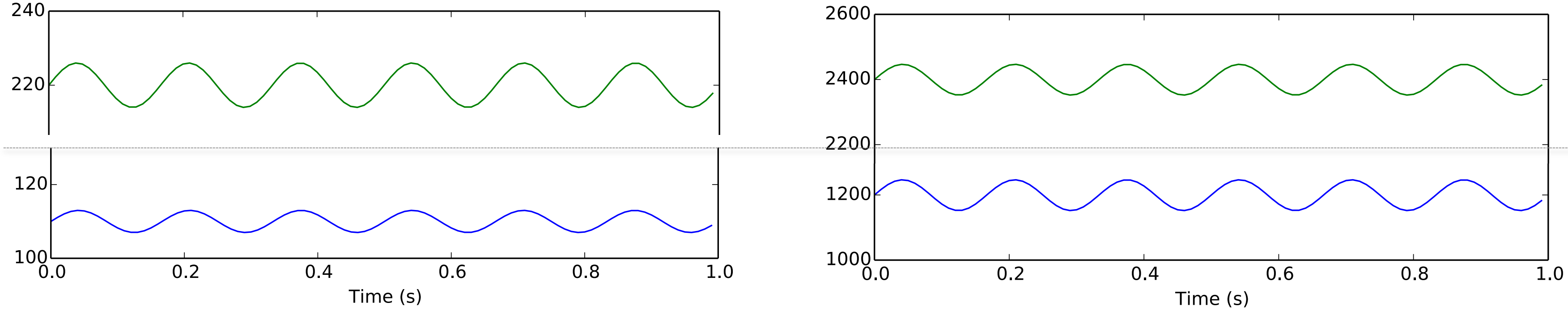


- Essentia implementation of Melodia
- Use default parameters in Essentia
- Performed well for Indian art music dataset in MIREX'11[1]

Salamon, Justin, and Emilia Gómez. "Melody extraction from polyphonic music signals using pitch contour characteristics." Audio, Speech, and Language Processing, IEEE Transactions on 20.6 (2012): 1759-1770.

[1]http://nema.lis.illinois.edu/nema out/mirex2011/results/ame/indian08/ summary.html

Data processing: Pre-processing



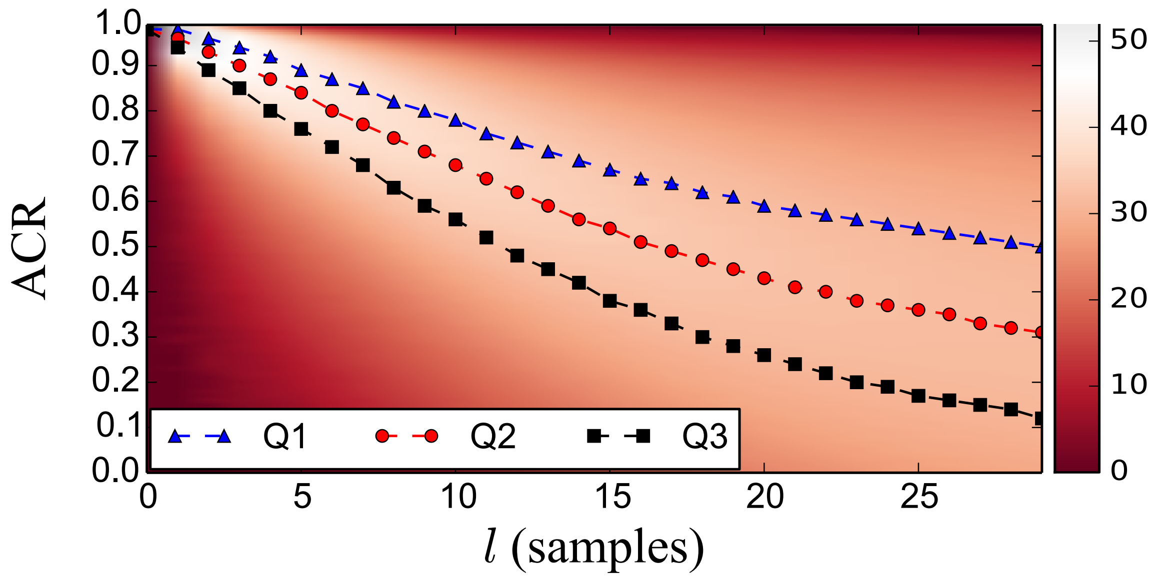
- Hertz to Cents conversion
  - Musically meaningful (logarithmic) scale
- Tonic normalization
  - Robustness against different tonic pitches of the lead artists
  - Automatic tonic identification (Essentia implementation)

$$P_{cents} = 1200 log_2(P_{Hz}/f_{tonic})$$

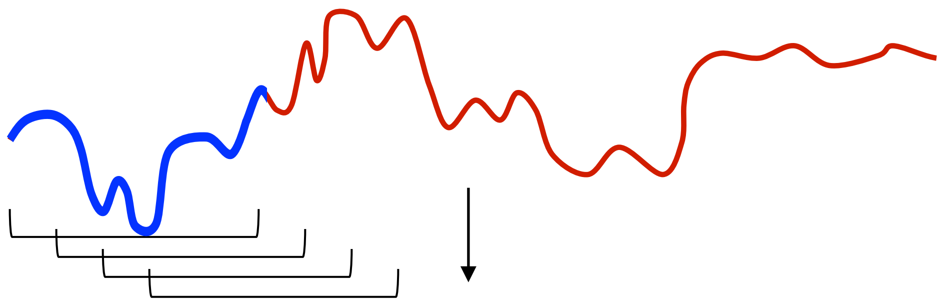
S. Gulati, A. Bellur, J. Salamon, H. Ranjani, V. Ishwar, H. A. Murthy, and X. Serra, "Automatic Tonic Identi- fication in Indian Art Music: Approaches and Evalua- tion," Journal of New Music Research, vol. 43, no. 01, pp. 55–73, 2014.

Data processing: Pre-processing

- Down-sampling
  - Histogram of Auto-correlation (ACF) at each lag value
  - Segments of 2 seconds
  - Significant drop in ACF for sampling rate of the pitch contour more than 22.2 ms



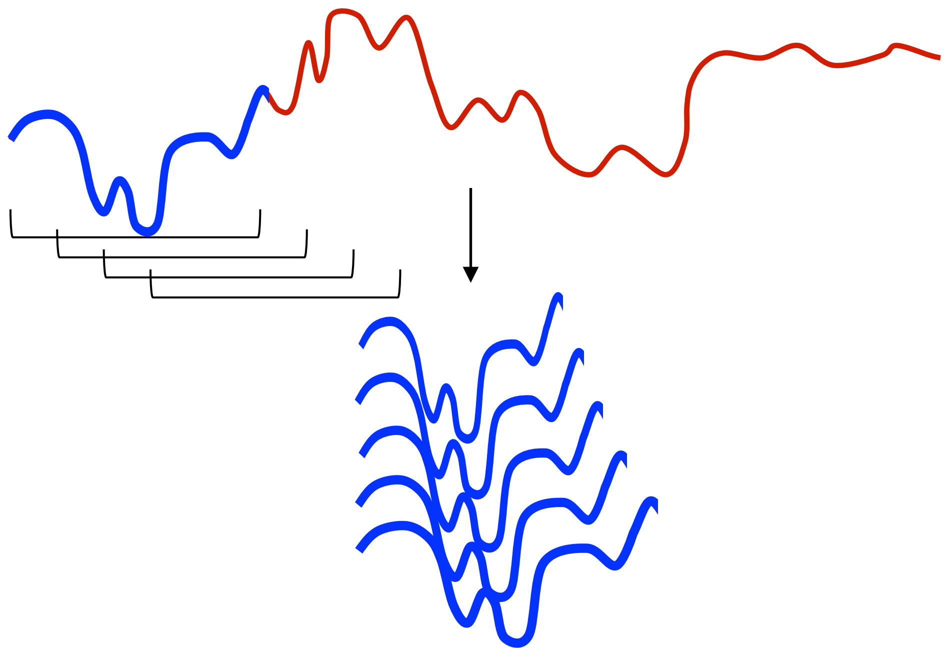
Data processing: Subsequence generation



- Melody segmentation – hard task
  - Nyās based segmentation
- Brute-force segmentation
  - Sliding window with constant hop
  - 2 second window
  - Remove segments across silence regions (> 0.5 seconds)

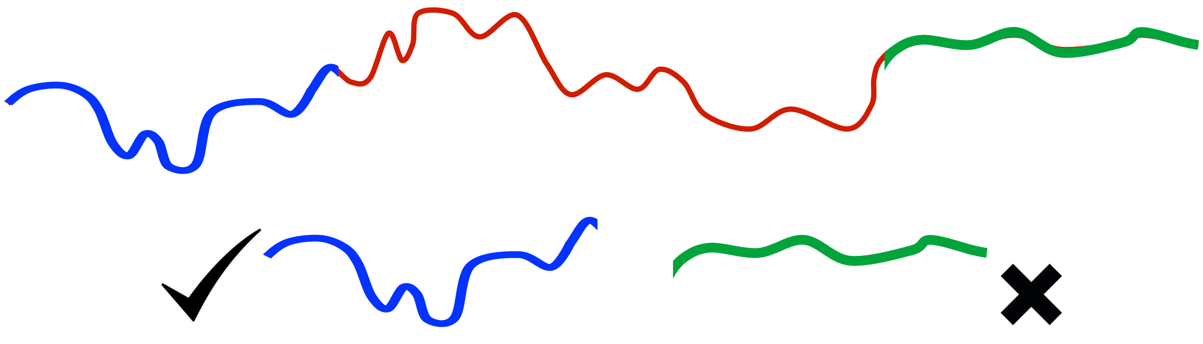
S. Gulati, J. Serrà, K. K. Ganguli, and X. Serra, "Landmark detection in hindustani music melodies," in Proc. of Int. Computer Music Conf., Sound and Music Computing Conf., Athens, Greece, 2014, pp. 1062–1068.

Data processing: Subsequence generation

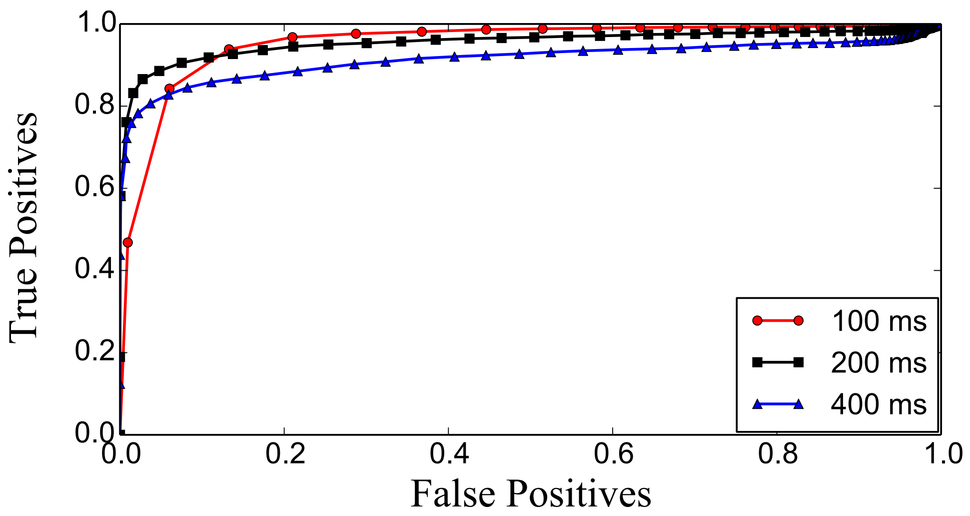


- Uniform time-scaling
  - 5 scale factors{0.9, 0.95, 1.0, 1.05, 1.1}
  - Similarity computation of 16 out of 25 combinations saved!! ($S_{1.0\rightarrow 1.05} = S_{1.05\rightarrow 1.1}$)

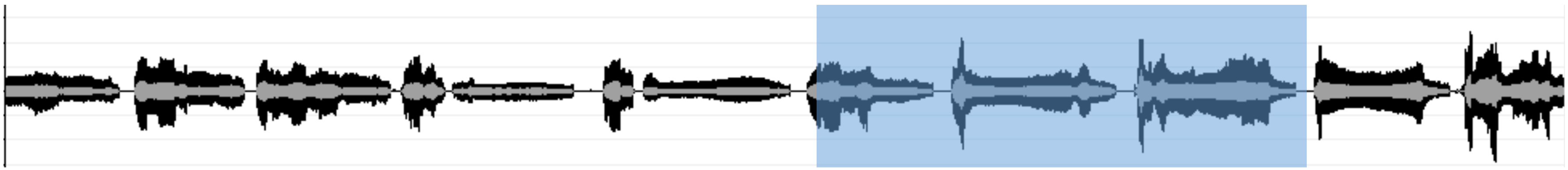
Data processing: Subsequence generation



- Segment filtering
  - Remove flat segments
  - Local Variance
    - Window length
    - Variance threshold



Data processing: Mridangam Segments



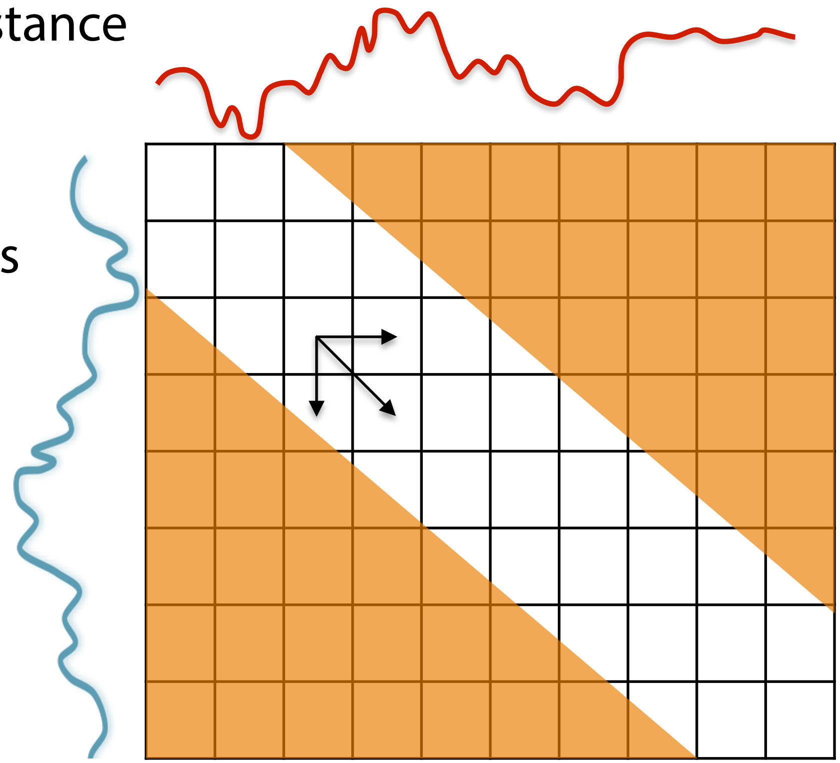
- Model based filtering
  - MFCC, spectral centroid & flatness
  - 46 ms frame size
  - Aggregate duration 2 seconds
  - Classifiers: Tree, KNN, NB, LR, SVM
  - Median filtering 20 seconds

D. Bogdanov, N. Wack, E. Go´mez, S. Gulati, P. Herrera, O. Mayor, G. Roma, J. Salamon, J. Zapata, and X. Serra, "Essentia: an audio analysis library for music information retrieval," in Proc. of Int. Society for Music Information Retrieval Conf. (ISMIR), 2013, pp. 493–498

F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, "Scikit-learn: machine learning in Python," Journal of Machine Learning Research, vol. 12, pp. 2825–2830, 2011.

Intra-recording Discovery

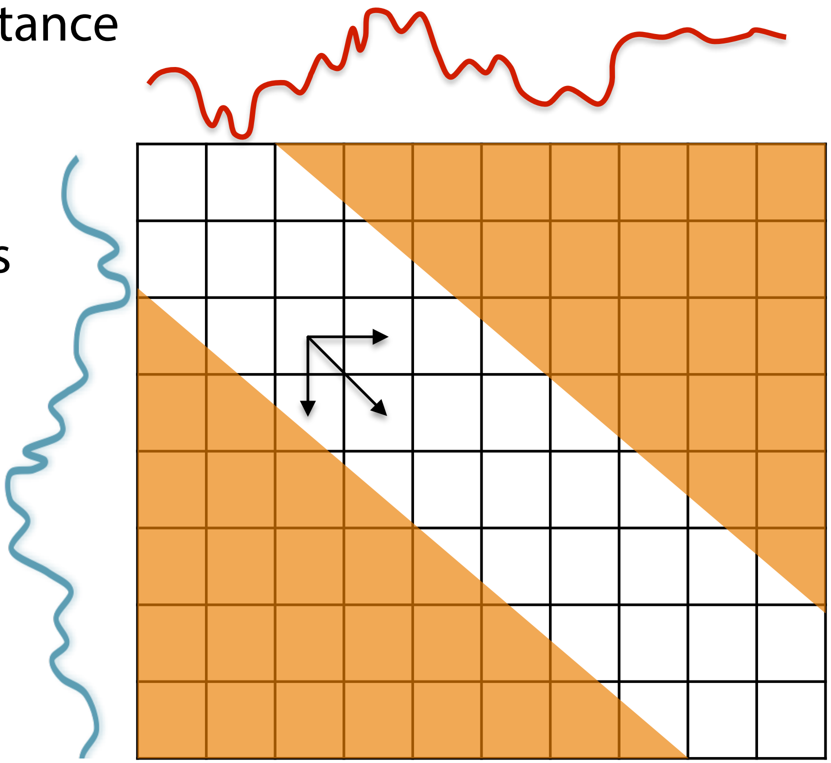
- Melodic Similarity
  - Dynamic time warping (DTW)
    - Cost matrix – Sq. Euclidean distance
    - 10% Sakoe-Chiba band
    - Step size [(1,0),(1,1),(0,1)]
    - No local constraint or penalties
- Lower bounds
  - FL bound
  - LB_Keogh_EC / EQ
- Statistics
  - 25 patterns per song
  - 79,000 total melodic patterns
  - 1.43 trillion similarity computations
  - 76 % computations avoided!!



H. Sakoe and S. Chiba, “Dynamic programming algorithm optimization for spoken word recognition,” IEEE Trans. on Acoustics, Speech, and Language Processing, vol. 26, no. 1, pp. 43–50, 1978.

Inter-recording Search

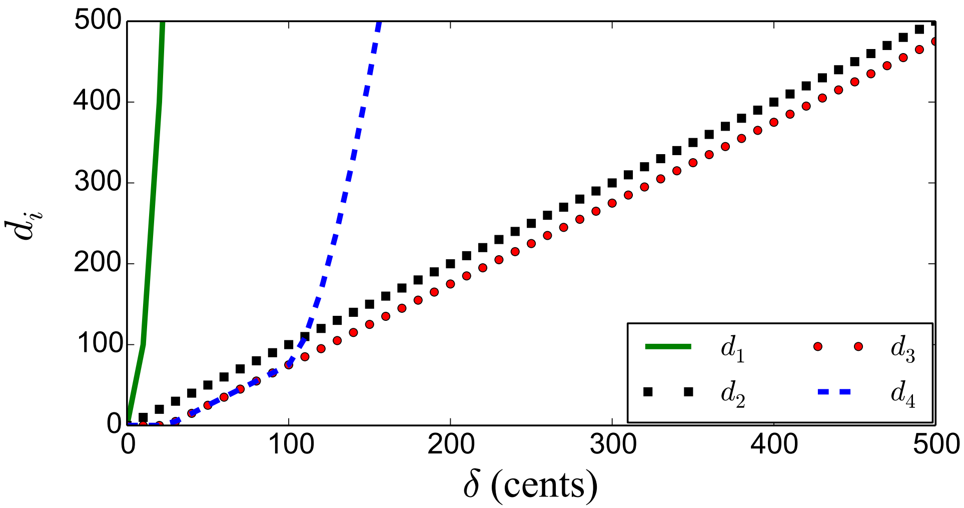
- Melodic Similarity
  - Dynamic time warping (DTW)
    - Cost matrix – Sq. Euclidean distance
    - 10% Sakoe-Chiba band
    - Step size [(1,0),(1,1),(0,1)]
    - No local constraint or penalties
- Lower bounds
  - FL bound
  - LB_Keogh_EC / EQ
- Statistics
  - 200 nearest neighbors
  - 15.8 million melodic patterns
  - 12.4 trillion similarity computations
  - 99 % computations avoided!!

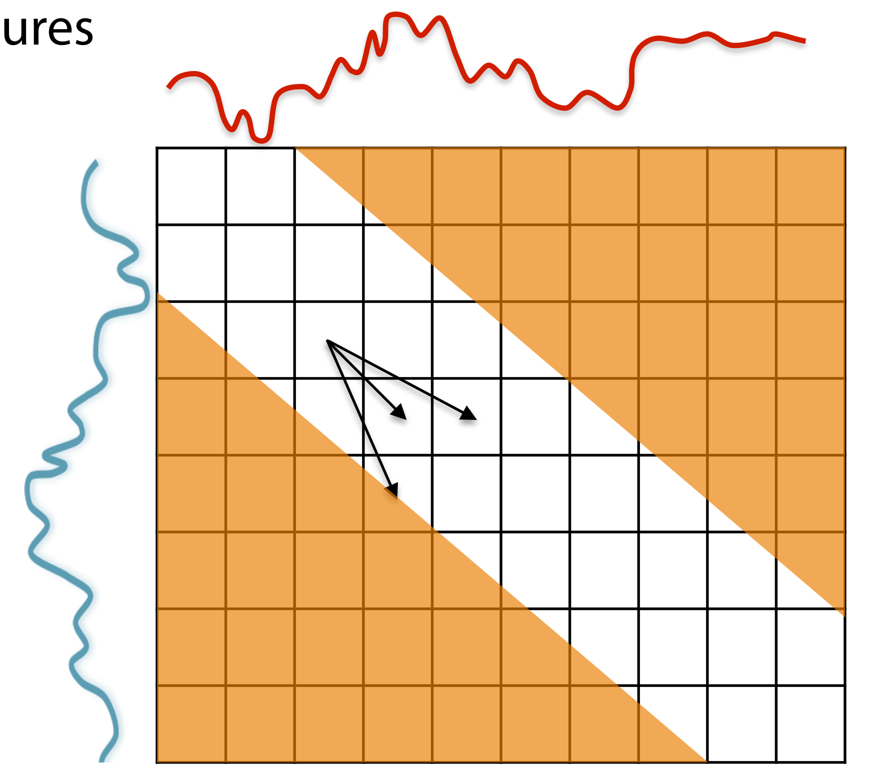


H. Sakoe and S. Chiba, "Dynamic programming algorithm optimization for spoken word recognition," IEEE Trans. on Acoustics, Speech, and Language Processing, vol. 26, no. 1, pp. 43–50, 1978.

Rank Refinement

- Melodic Similarity
  - Dynamic time warping (DTW)
    - Cost matrix – 4 distance measures
    - 10% Sakoe-Chiba band
    - Step size [(2,1),(1,1),(1,2)]
    - Local constraint, no penalties
- No lower bounds

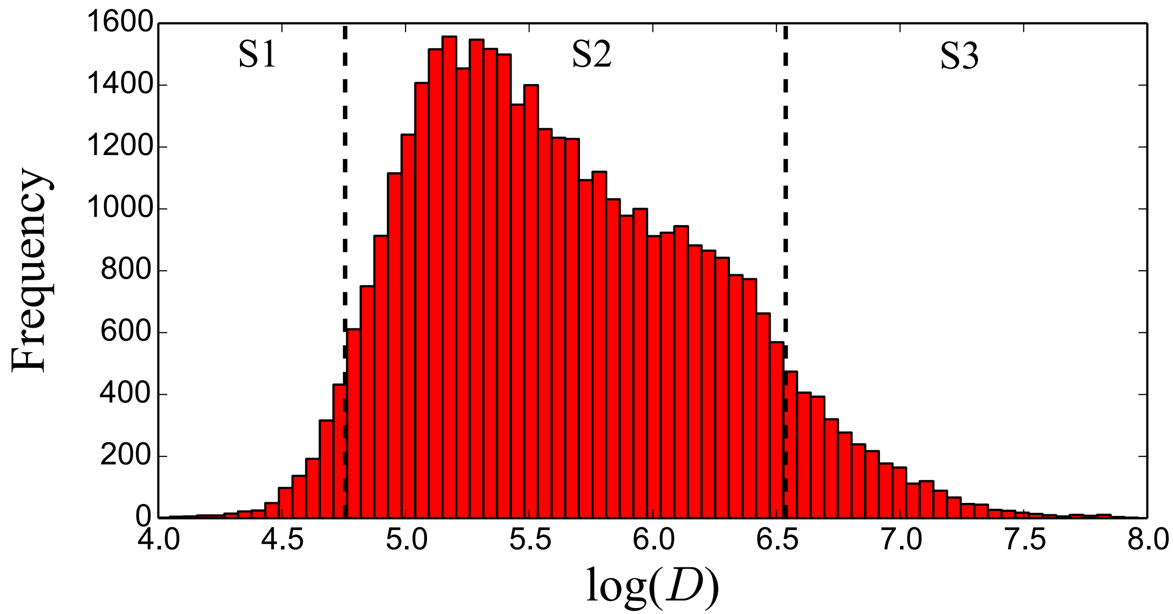




H. Sakoe and S. Chiba, "Dynamic programming algorithm optimization for spoken word recognition," IEEE Trans. on Acoustics, Speech, and Language Processing, vol. 26, no. 1, pp. 43–50, 1978.

Evaluation

- 79,000 seed patterns, 15.8 million searched patterns
- 4 different distance measure for rank refinement
- 200 seed pattern pairs
  - Top 10 searched patterns for 4 methods
- Total of 8000 patterns (200*10*4)



Evaluation - Annotations

- Professional musician with over 20 years of formal training.
- Listening short audio fragments (melodic patterns)
- Listening Melodically similar: 1 (Good)
- Melodically dissimilar: 0 (Bad)

You are rating motif Id = 1, Version = 1

Listen to the Seed Motif id = 1

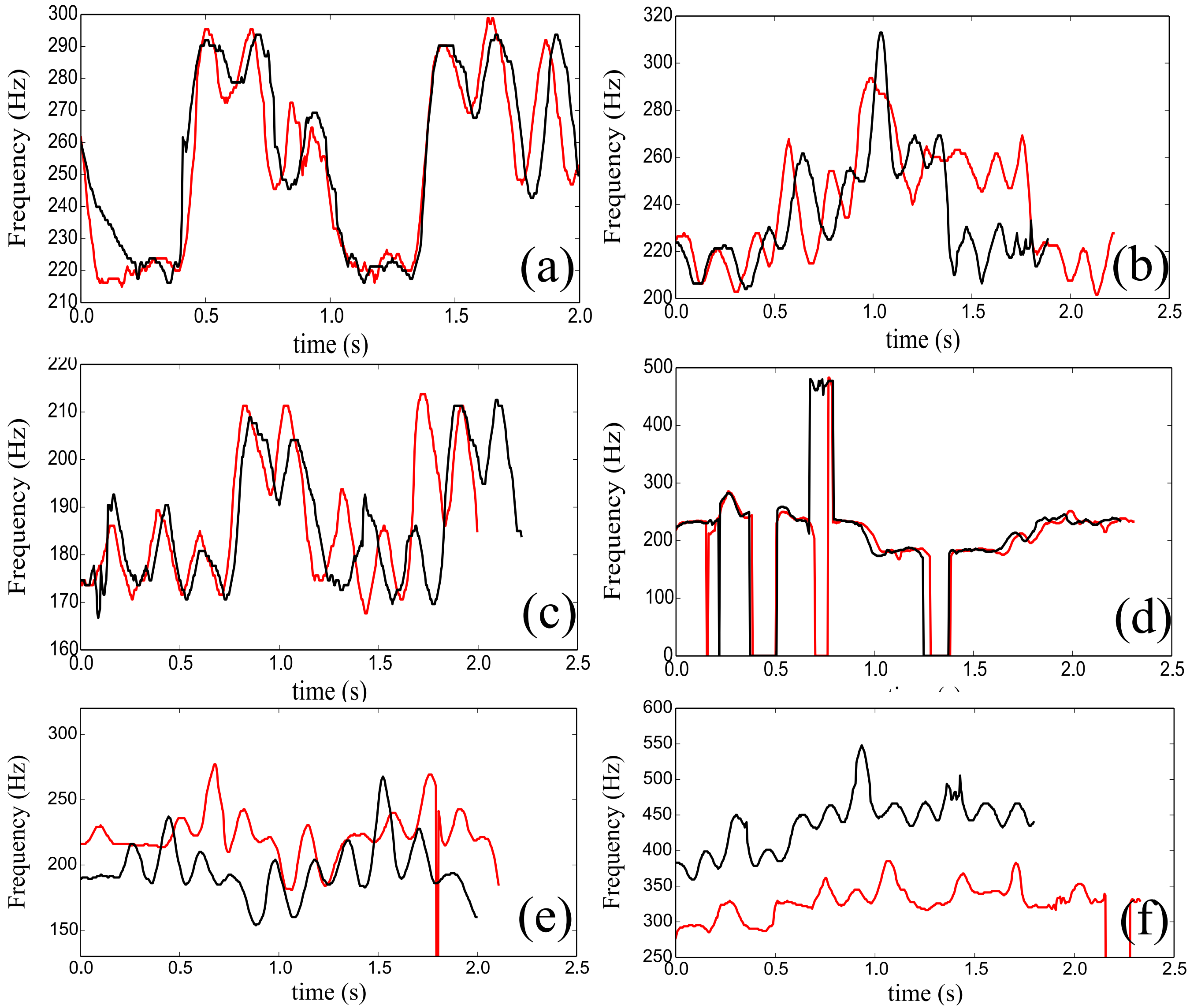
Search pattern	Status	Bad	OK	Good	Submit
1	1	○	○	○	Submit
2	1	○	○	○	Submit
3	1	○	○	○	Submit
4	1	○	○	○	Submit
5	1	○	○	○	Submit
6	1	○	○	○	Submit
7	1	○	○	○	Submit
8	1	○	○	○	Submit
9	1	○	○	○	Submit
10	1	○	○	○	Submit

Evaluation - Measures

- Mean Average Precision (MAP)
- Statistical significance
  - Mann-Whitney U test ($P < 0.05$)
- Multiple comparison compensation
  - Holm-Bonferroni method

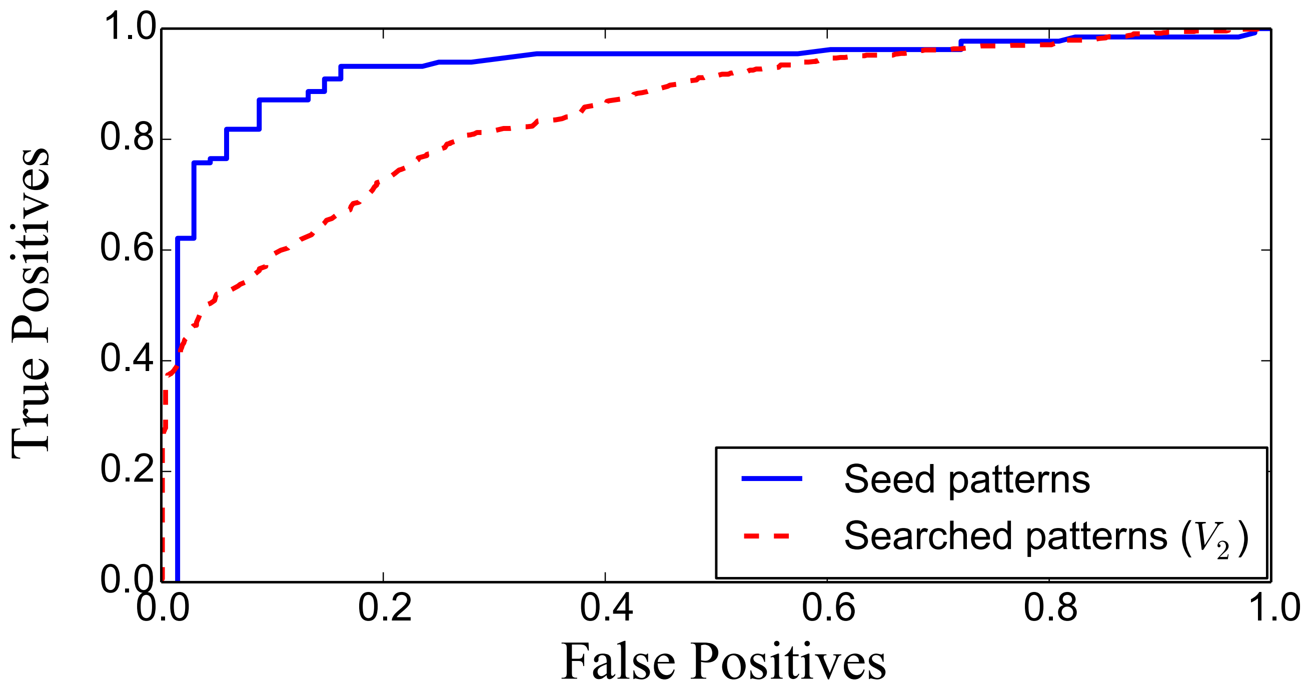
C. D. Manning, P. Raghavan, and H. Schütze, Introduction to information retrieval. Cambridge university press Cam- bridge, 2008, vol. 1.
H. B. Mann and D. R. Whitney, "On a test of whether one of two random variables is stochastically larger than the other," The annals of mathematical statistics, vol. 18, no. 1, pp. 50– 60, 1947.
S. Holm, "A simple sequentially rejective multiple test pro- cedure," Scandinavian journal of statistics, vol. 6, no. 2, pp. 65–70, 1979.

Results - Qualitative



Results – Intra recording patterns

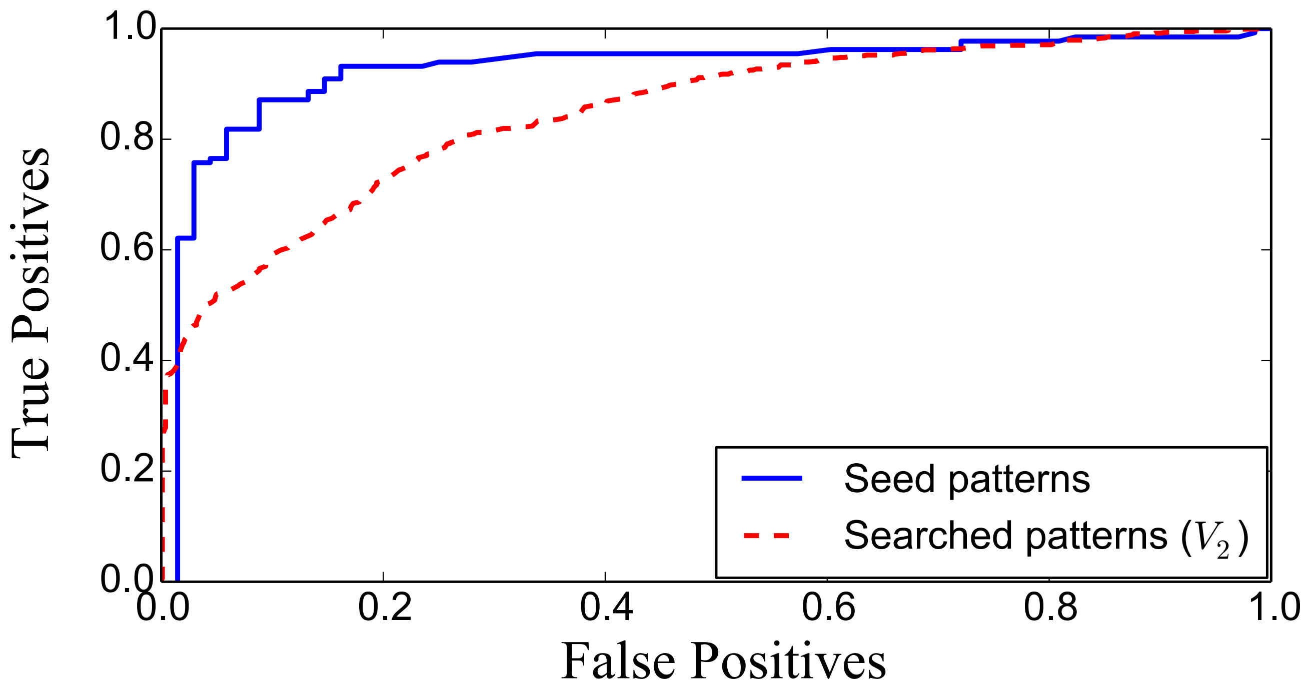
- Fraction of melodically similar seed patterns
  - S1 (0.98), S2(0.67) and S3(0.31)
- Well separated distance distributions



Results – Inter recording patterns

MAP SCORES FOR FOUR VARIANTS OF RANK REFINEMENT METHOD (V_i) FOR EACH SEED CATEGORY (S1, S2 AND S3).

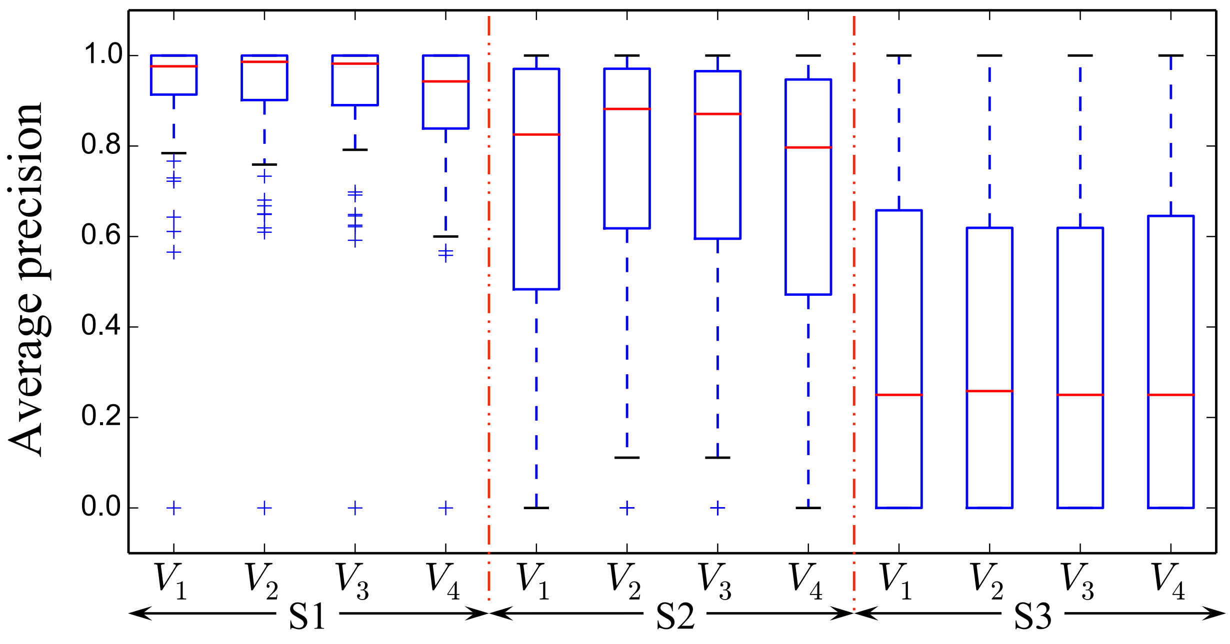
Seed Category	V_1	V_2	V_3	V_4
S1	0.92	0.92	0.91	0.89
S2	0.68	0.73	0.73	0.66
S3	0.35	0.34	0.35	0.35



Results – Inter recording patterns

MAP SCORES FOR FOUR VARIANTS OF RANK REFINEMENT METHOD (V_i) FOR EACH SEED CATEGORY (S1, S2 AND S3).

Seed Category	V_1	V_2	V_3	V_4
S1	0.92	0.92	0.91	0.89
S2	0.68	0.73	0.73	0.66
S3	0.35	0.34	0.35	0.35

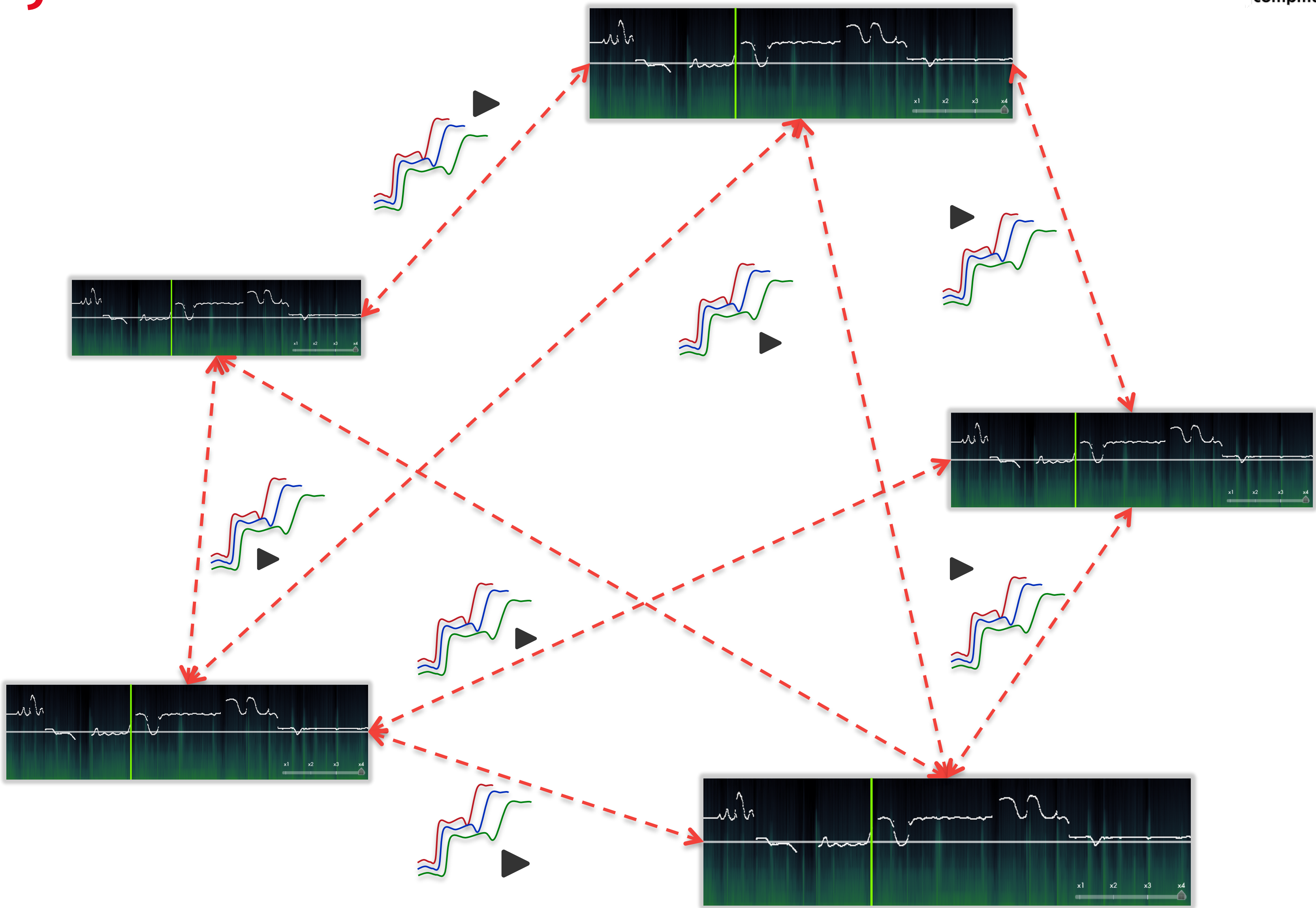


Conclusions and Future work

- Data driven unsupervised approach – melodic pattern discovery
- DTW based distance measure is good for intra recording discovery
- Need informed distance measures for inter song pattern search
- DTW using Cityblock distance performs little better than the rest
- Closer seed pattern pairs have higher MAP scores → higher number of repetitions

- Future Work
  - Similar analysis on Hindustani music
  - Transposition invariance
  - Network analysis from mined patterns

Objective



Demo:

compmusic

http://dunya.compmusic.upf.edu/motifdiscovery/

Navigation:

- Browsing options
- Artist list
- Releases
- Recordings
- Seeds

Stop Audio

Search results using chosen seed from Nadamadi

Similarity	Seed_Id	Start(s)	End(s)	Pair_Id	Start(s)	End(s)	Musicbrainz ID (searched file)	Distance
✓	15506039	95.6	97.5	15506885	372.8	375.0	2bade8d8-1cfa-4076-9329-98f7cacc65a0	1318.68
✓	15506039	95.6	97.5	15506914	57.9	60.2	70761911-9f70-436c-97ed-d23ea74e7ed9	2416.83
✓	15506039	95.6	97.5	15506904	223.7	225.9	5c342c56-c07a-4905-89cc-bd5d1151d20a	2507.13
✓	15506039	95.6	97.5	15506891	2448.1	2450.1	1d99a413-bc0a-430d-9587-410932113eaf	2554.72
✓	15506039	95.6	97.5	15506925	98.7	101.0	b6af2720-6beb-454b-ba8c-f912ea8ab27b	2573.89
✓	15506039	95.6	97.5	15506921	77.7	80.1	b6af2720-6beb-454b-ba8c-f912ea8ab27b	2615.10
✓	15506039	95.6	97.5	15506888	132.3	134.5	70761911-9f70-436c-97ed-d23ea74e7ed9	2615.37
✓	15506039	95.6	97.5	15506886	538.7	541.2	170970da-a19a-462d-8dae-4ece614f2780	2617.46
✓	15506039	95.6	97.5	15506887	538.7	541.2	170970da-a19a-462d-8dae-4ece614f2780	2617.46
✓	15506039	95.6	97.5	15506890	10.7	12.5	367f884a-5de9-4f45-a139-82a067c13865	2630.05
✓	15506039	95.6	97.5	15506901	69.8	72.1	bec3b237-0a03-4011-9d8b-394415b0a6b2	2635.77
✓	15506039	95.6	97.5	15506990	927.7	929.9	5269b678-e274-4732-a906-4b17607df9c3	2658.49
✓	15506039	95.6	97.5	15506910	261.2	263.2	5c342c56-c07a-4905-89cc-bd5d1151d20a	2708.06
✓	15506039	95.6	97.5	15506899	175.1	177.1	5c342c56-c07a-4905-89cc-bd5d1151d20a	2740.02
✓	15506039	95.6	97.5	15506894	272.5	274.5	0298a06d-ffe9-4d83-922d-dedbc3bfde21	2768.00
✓	15506039	95.6	97.5	15506889	663.2	665.2	829df365-78bc-4157-9346-5a3b39bf12a5	2770.42
✓	15506039	95.6	97.5	15506919	260.4	262.4	2f9b5ddc-f253-46be-a316-36f9ce111b9e	2801.36
✓	15506039	95.6	97.5	15506902	219.2	221.3	5c24dc68-51e2-4ce5-a7c7-74f160482e2b	2827.50
✓	15506039	95.6	97.5	15506976	32.9	35.1	d7112257-77c5-4f52-a284-c73226cad4d0	2833.71
✓	15506039	95.6	97.5	15506897	206.3	209.0	bedc82f2-d42c-4062-9fc1-832f7f1bfd62	2836.99
✓	15506039	95.6	97.5	15507021	33.0	34.9	8e31cd33-0143-4357-83cd-31c87443055d	2852.56
✓	15506039	95.6	97.5	15506893	144.6	147.0	097411d7-bd64-41b7-a604-56bdcb584886	2869.28
✓	15506039	95.6	97.5	15506913	49.8	52.0	e00a3860-8ae2-400b-8300-4d72204969b3	2878.99

> 16 million melodic patterns

Mining Melodic Patterns in Large Audio Collections of Indian Art Music

Sankalp Gulati*, Joan Serrà†, Vignesh Ishwar* and Xavier Serra*

**Music Technology Group, Universitat Pompeu Fabra, Barcelona, Spain*

†Artificial Intelligence Research Institute (IIIA-CSIC), Bellaterra, Barcelona, Spain

SITIS – 2014 (MIRA), Marrakech, Morocco

upf. Music Technology Group

compmusic

INSTITUT D'INVESTIGACIÓ EN INTEL.LIGÈNCIA ARTIFICIAL - IIIA