Phrase-based Rāga Recognition using Vector Space Modeling

Sankalp Gulati*, Joan Serrà^, Vignesh Ishwar*, Sertan Şentürk* and Xavier Serra*

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

^Telefonica Research, Barcelona, Spain

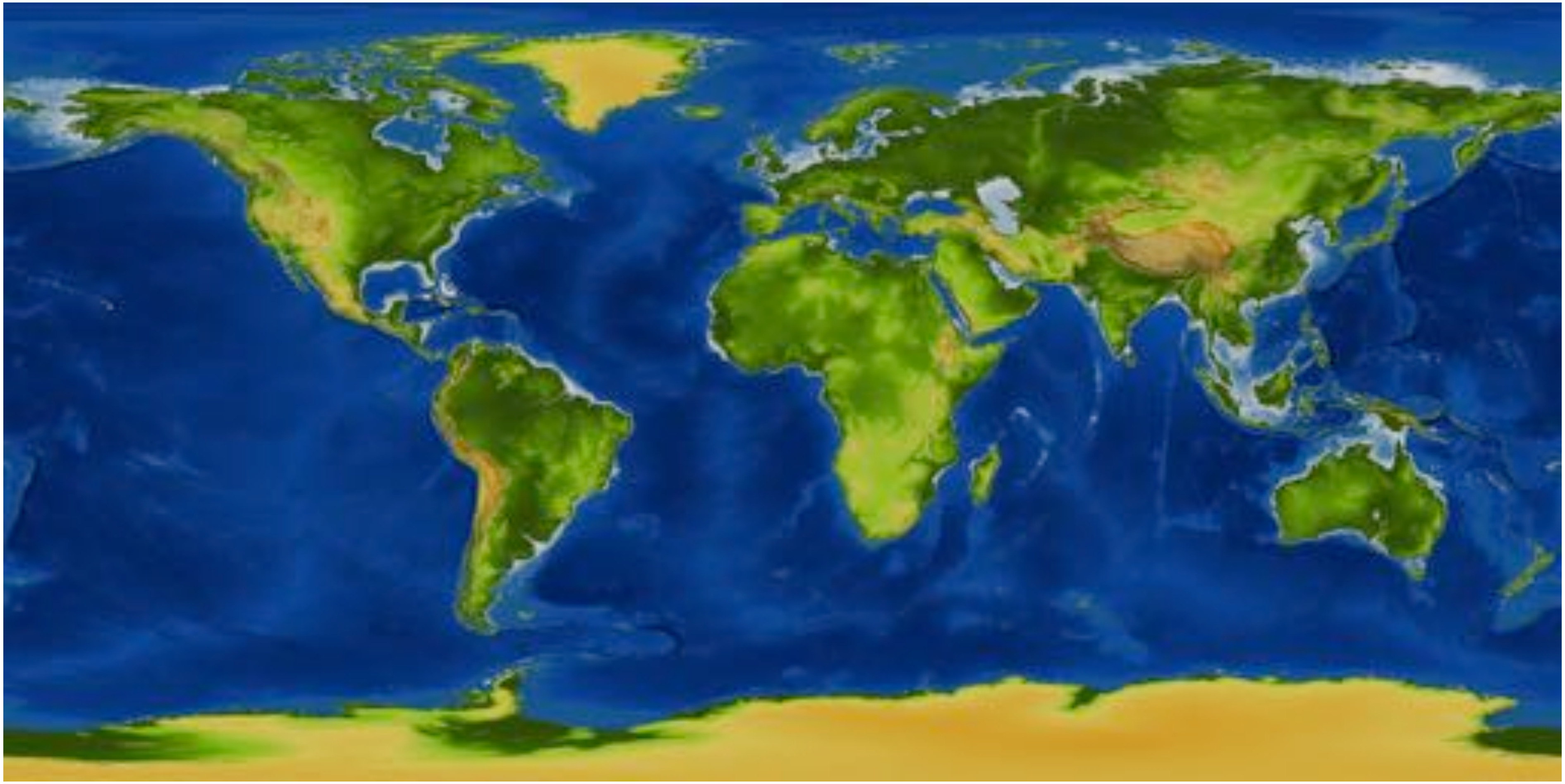
The 41st IEEE International Conference on Acoustics, Speech and Signal Processing

Shanghai, China, 2016

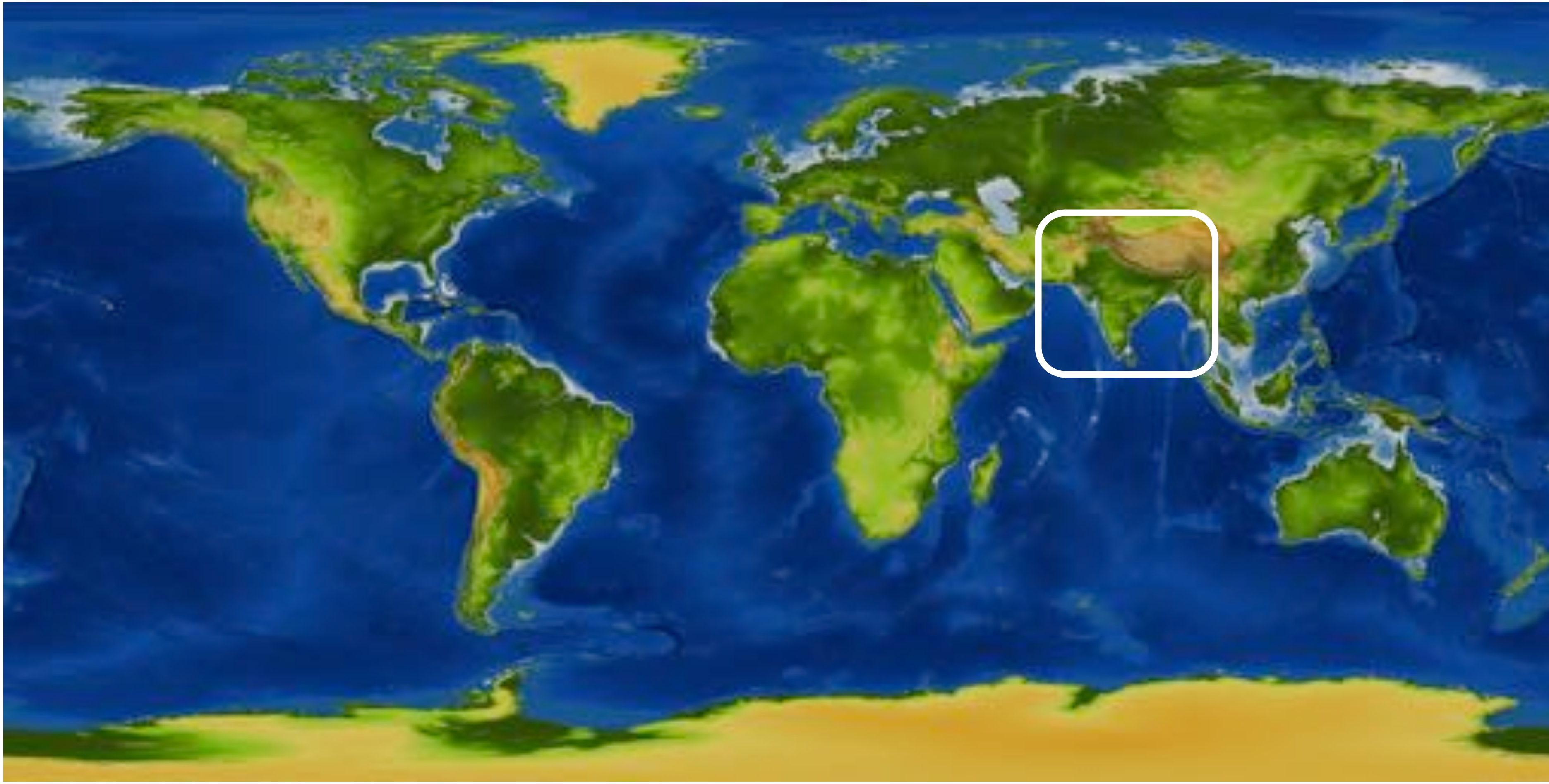
upf. Music Technology Group

compmusic

Indian Art Music

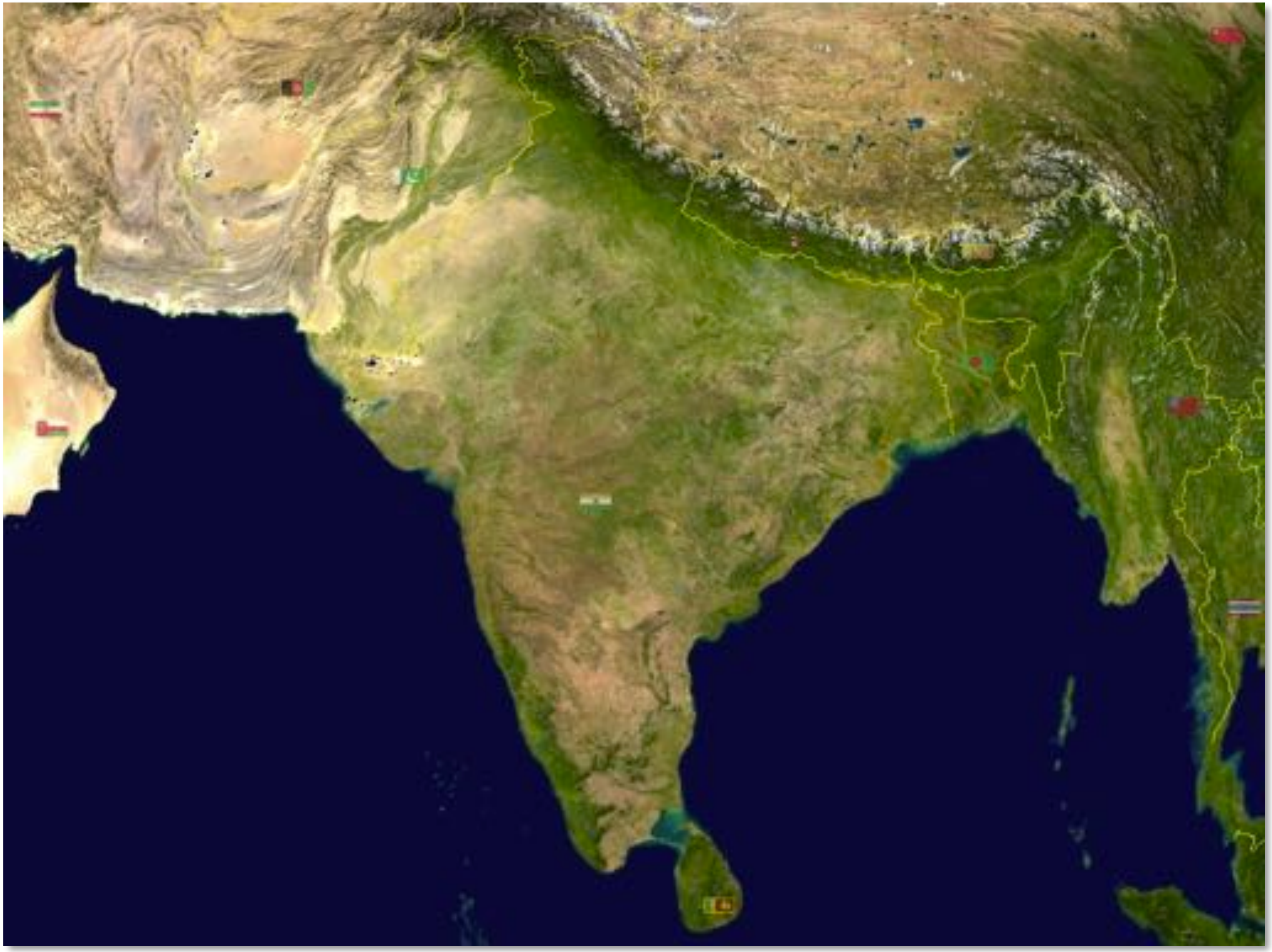


Indian Art Music

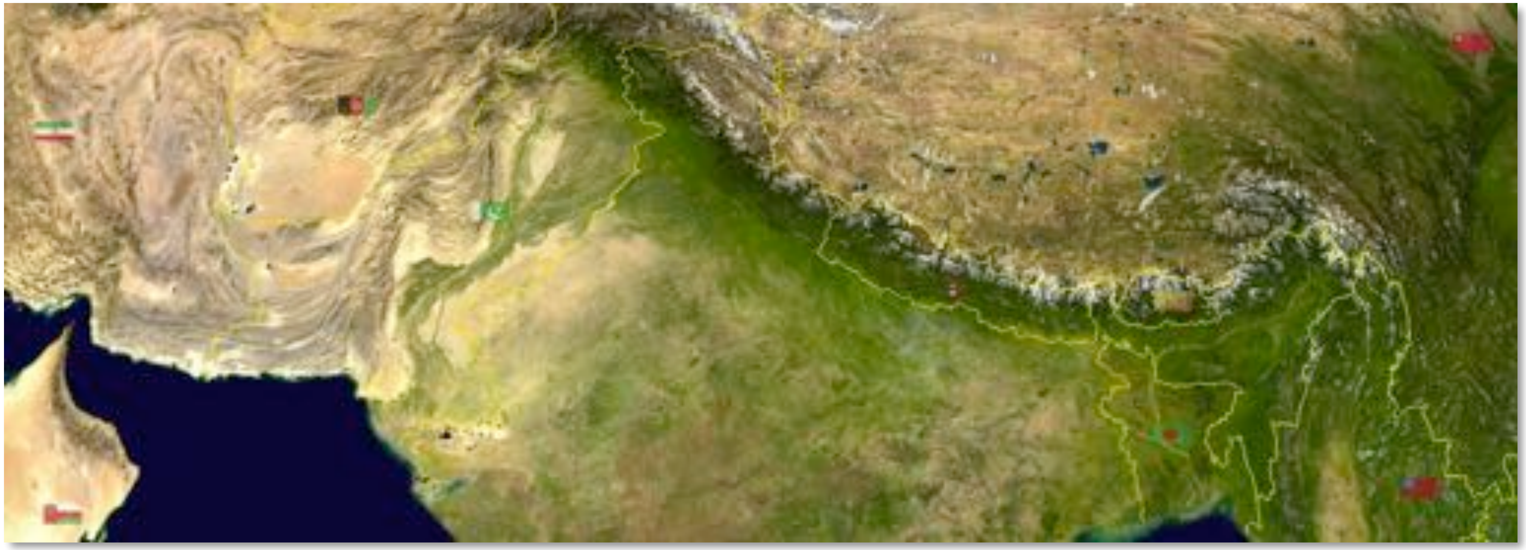


Indian subcontinent: India, Pakistan, Bangladesh, Srilanka, Nepal and Bhutan

Indian Art Music



Indian Art Music: Hindustani music



Indian Art Music: Carnatic music

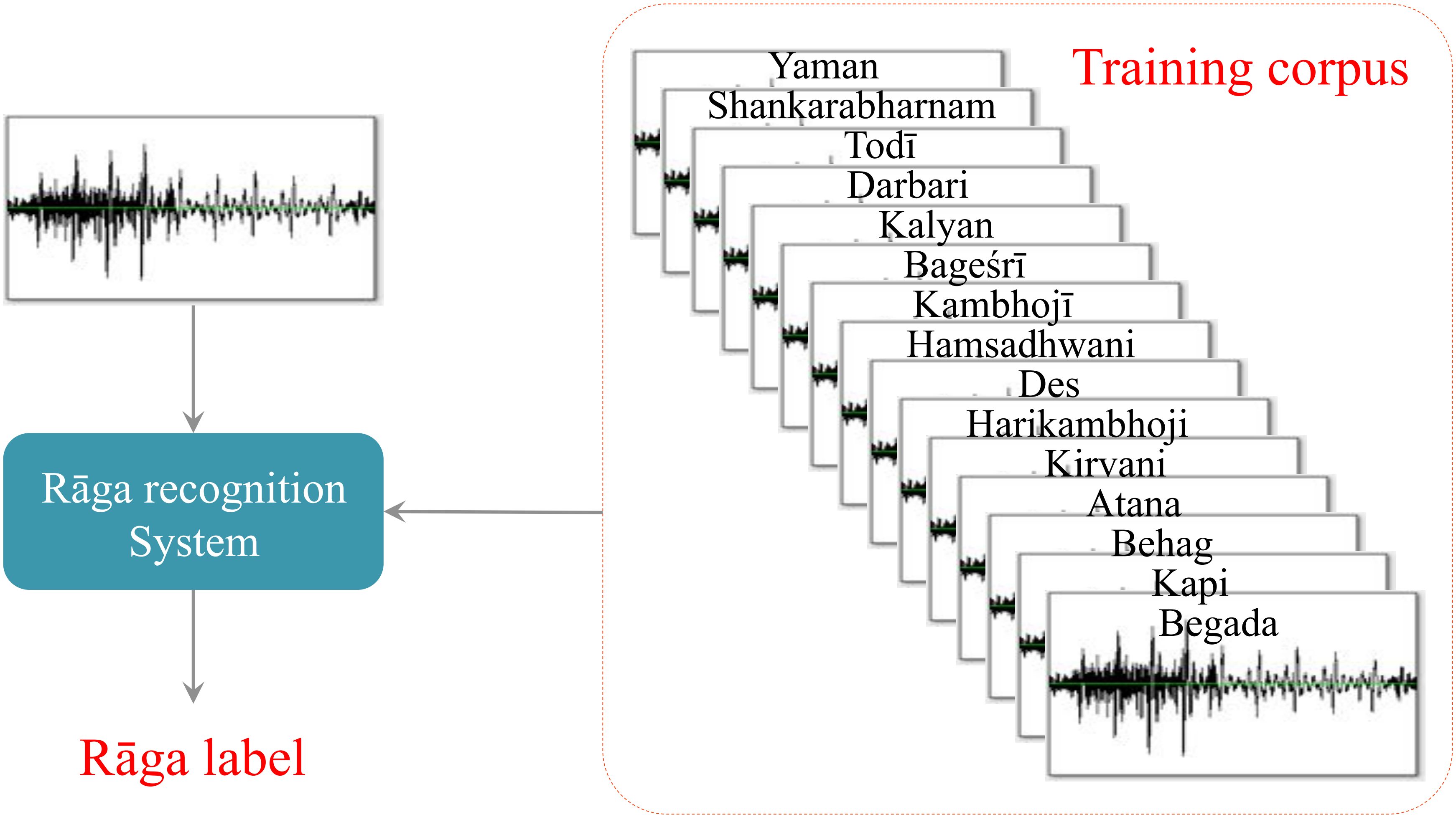
Music Technology Group

Rāga: melodic framework

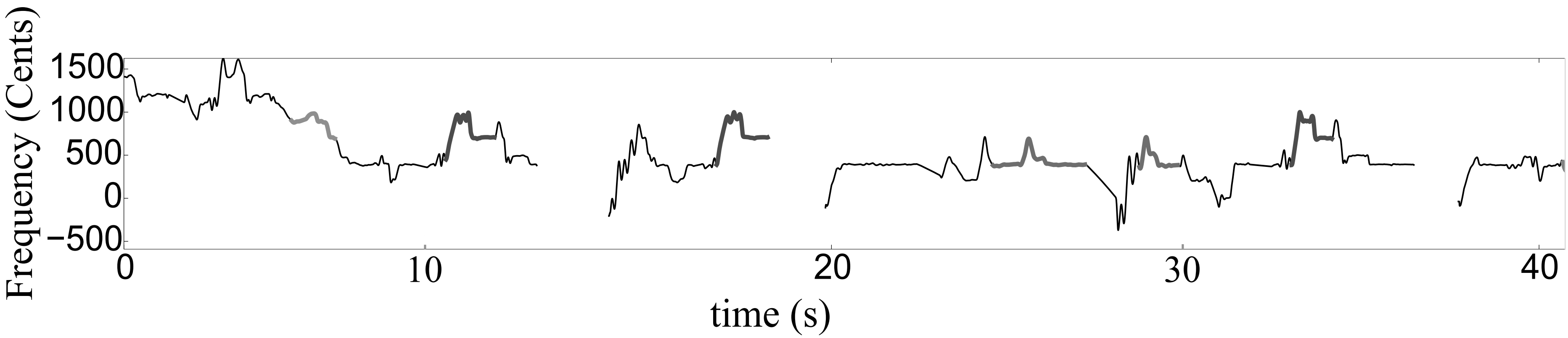
- Grammar for improvisation and composition

upf. Music Technology Group

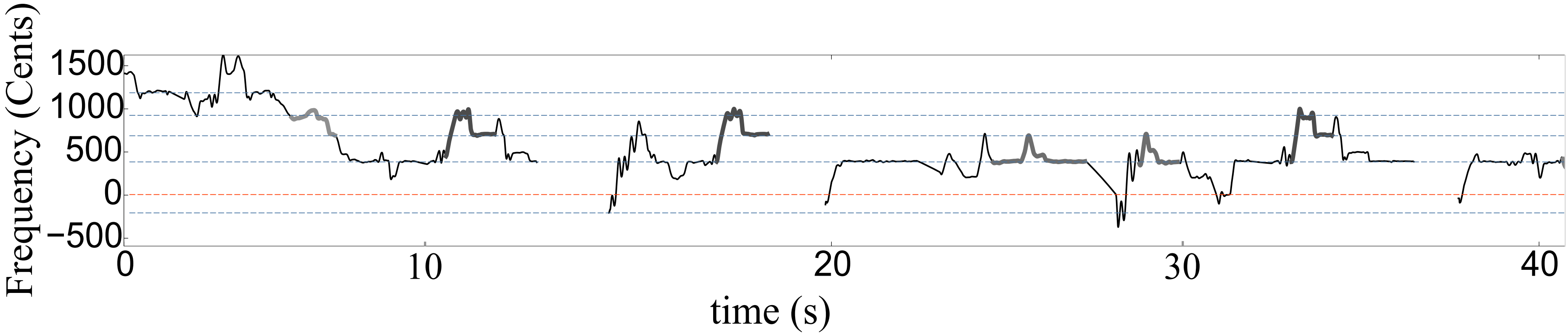
Automatic Rāga Recognition



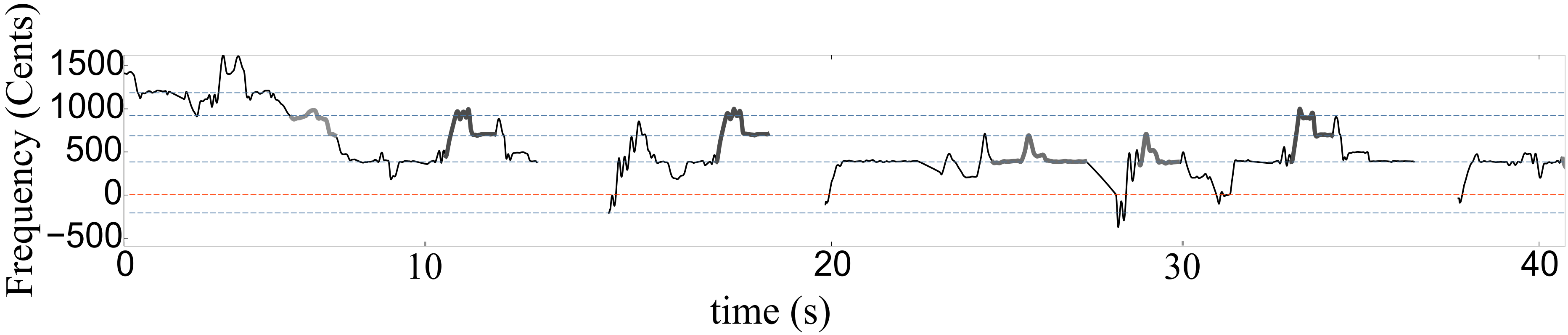
Rāga Characterization: Svaras

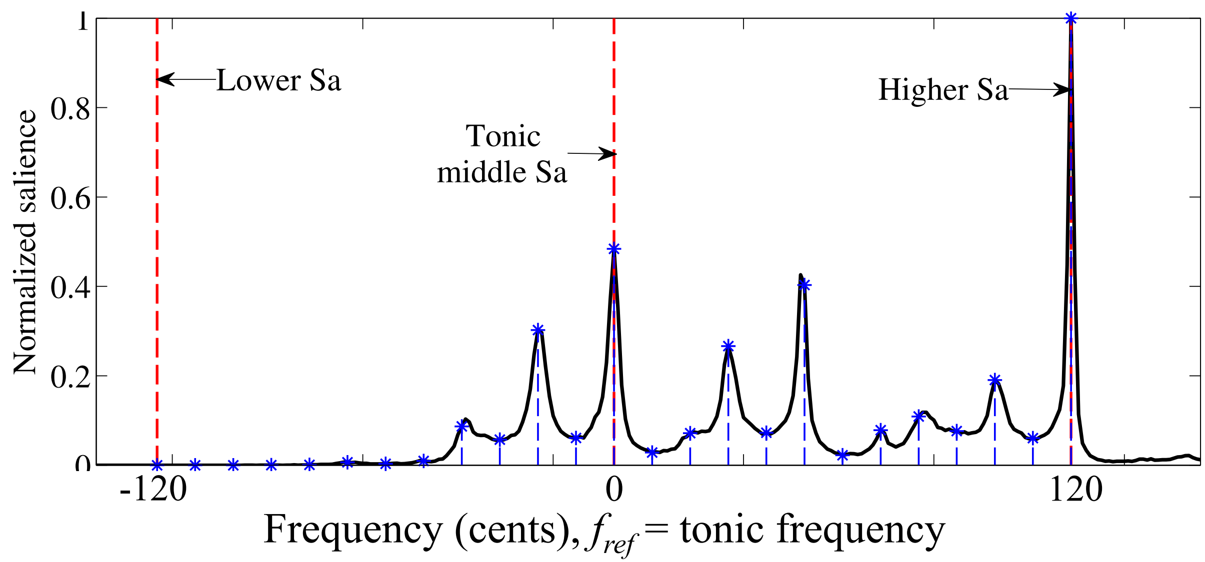


Rāga Characterization: Svaras



Rāga Characterization: Svaras

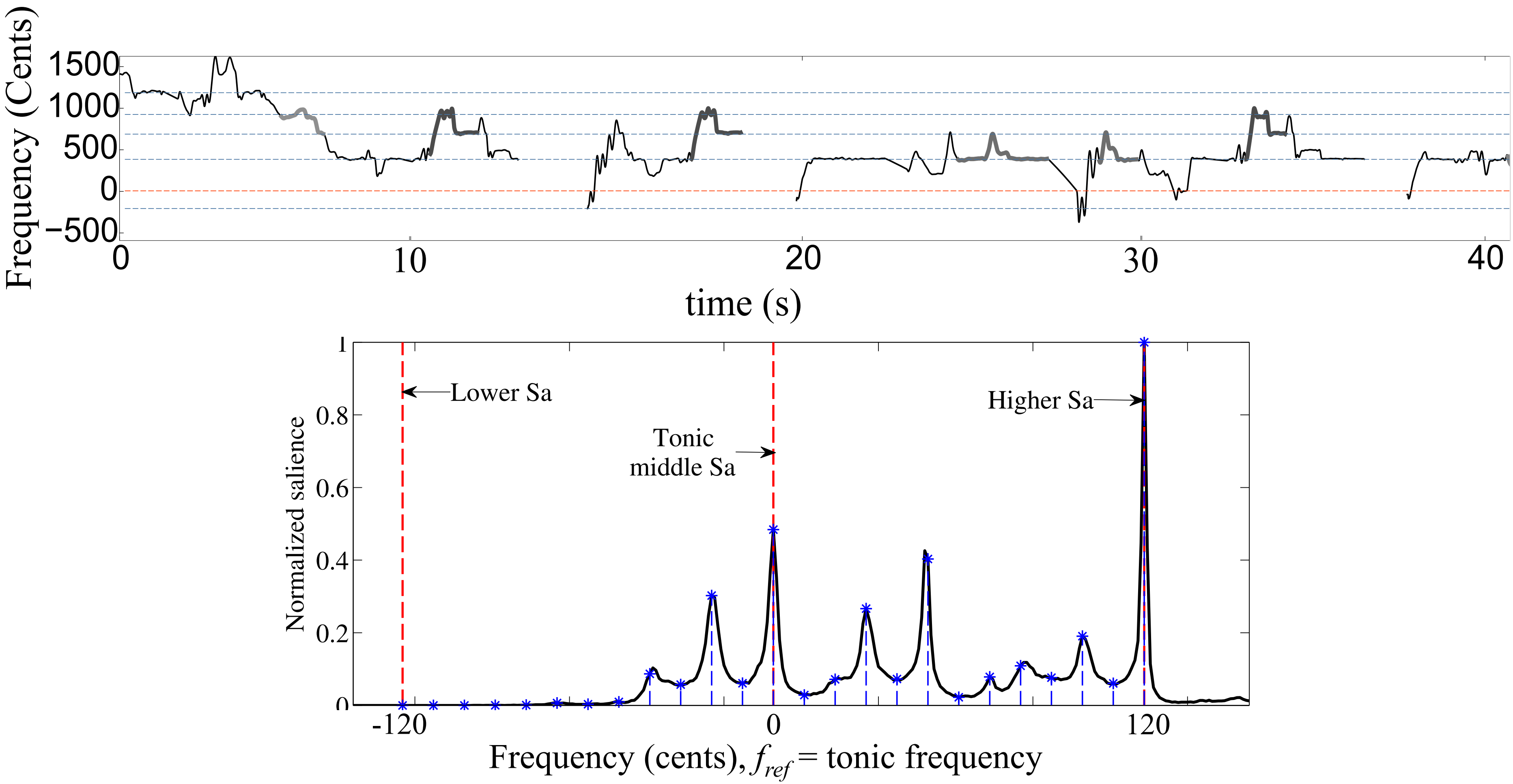




- P. Chordia and S. Şentürk, "Joint recognition of raag and tonic in North Indian music," Computer Music Journal, vol. 37, no. 3, pp. 82–98, 2013.
- G. K. Koduri, S. Gulati, P. Rao, and X. Serra, "Rāga recognition based on pitch distribution methods," Journal of New Music Research, vol. 41, no. 4, pp. 337–350, 2012.

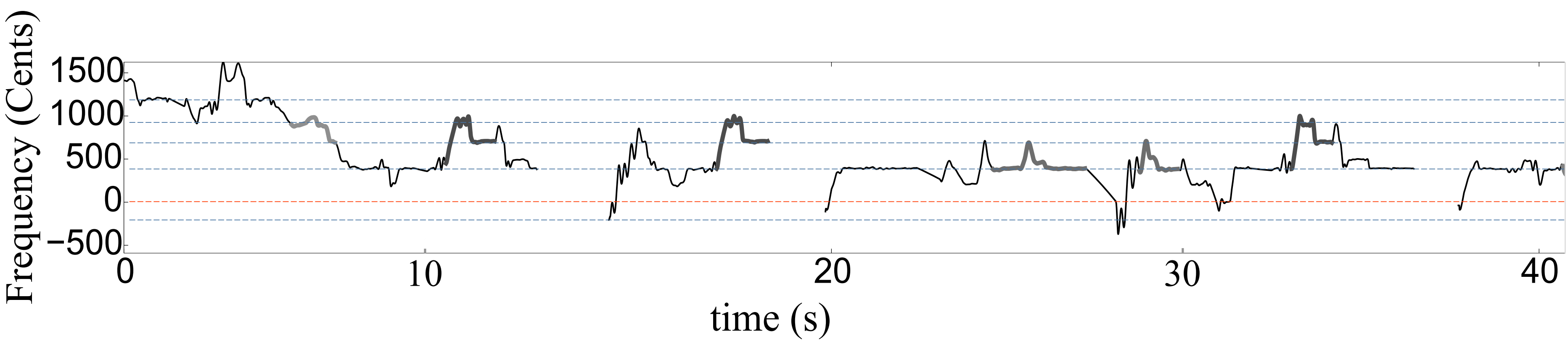
upf. Music Technology Group

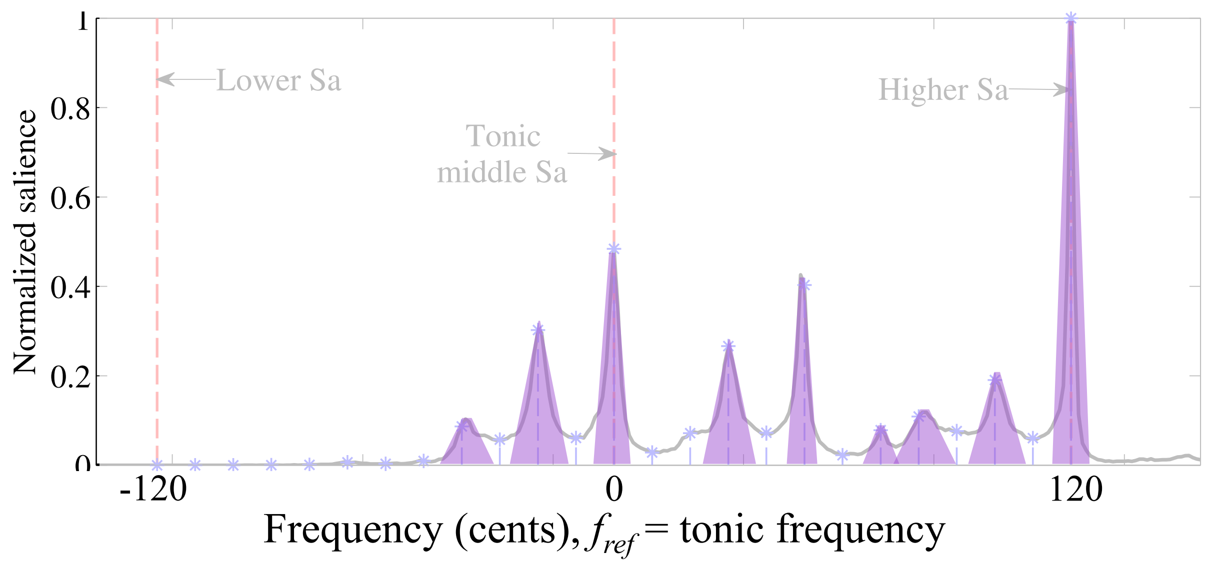
Rāga Characterization: Intonation



Music
Technology
Group

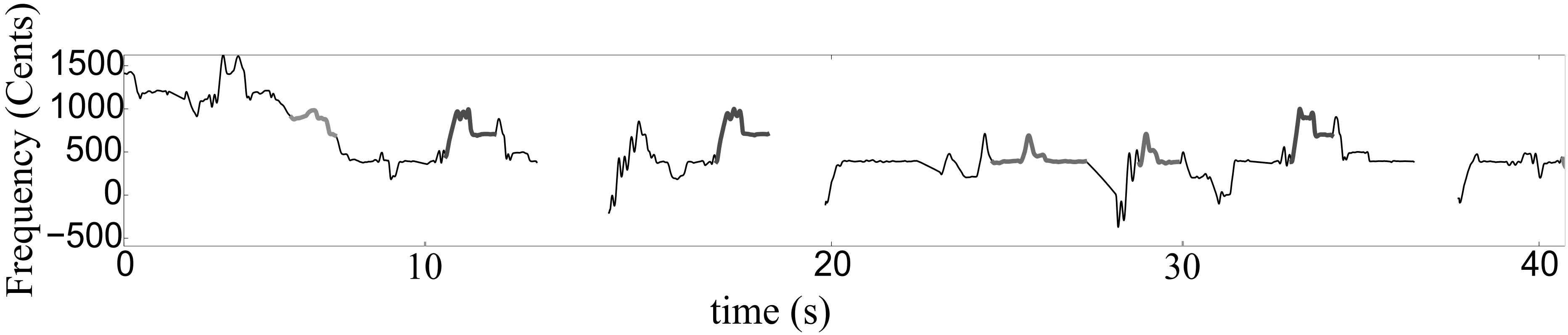
Rāga Characterization: Intonation





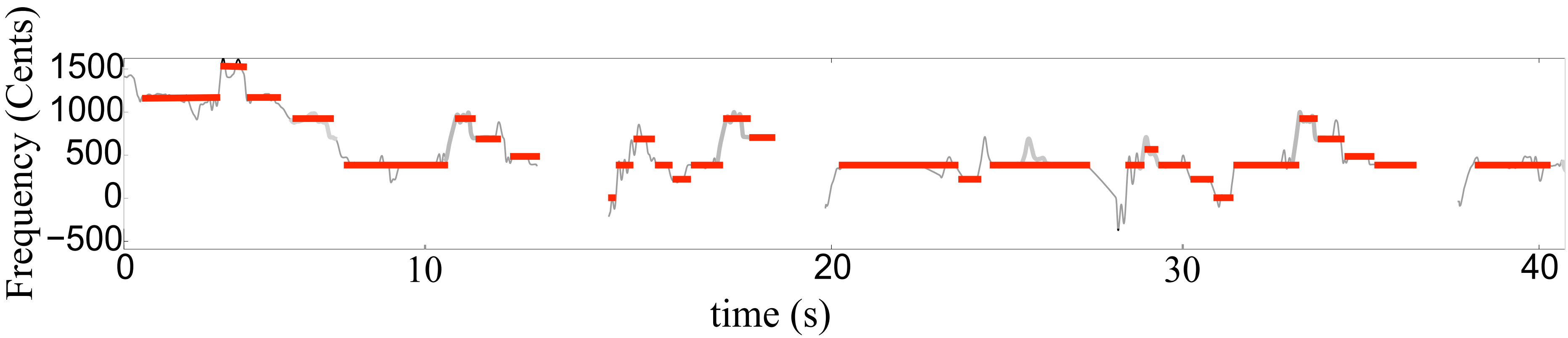
- G.K.Koduri,V.Ishwar,J.Serrà,andX.Serra,"Intonation analysis of rāgas in Carnatic music," Journal of New Music Research, vol. 43, no. 1, pp. 72–93, 2014.
- H. G. Ranjani, S. Arthi, and T. V. Sreenivas, "Carnatic music analysis: Shadja, swara identification and raga verification in alapana using stochastic models," in IEEE WASPAA, 2011, pp. 29–32.

Rāga Characterization: Ārōh-Avrōh



- Ascending-descending svara pattern; melodic progression

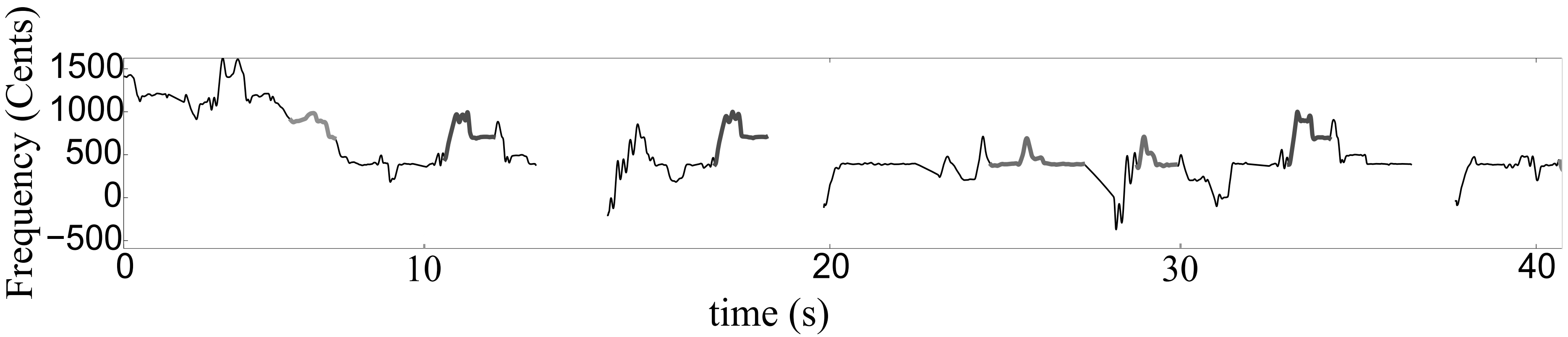
Rāga Characterization: Ārōh-Avrōh



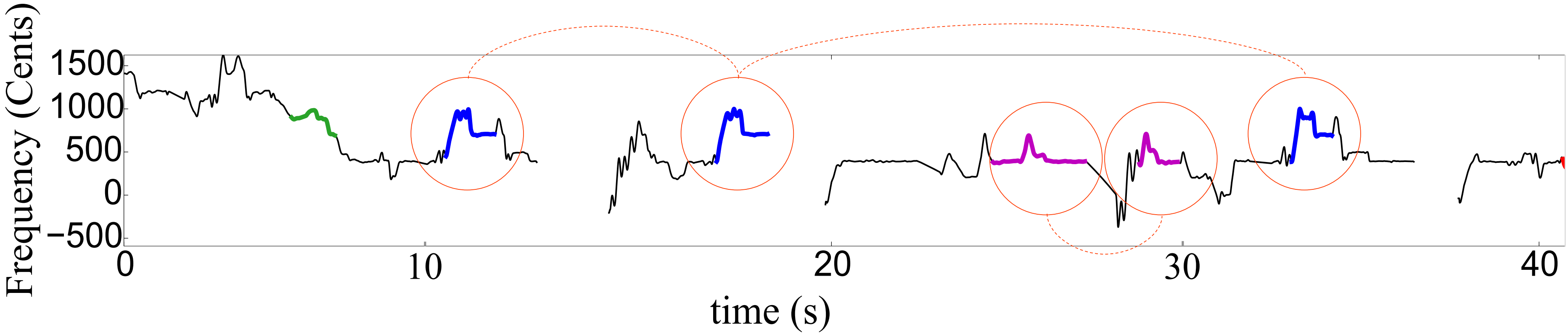
Melodic Progression Templates
N-gram Distribution
Hidden Markov Model

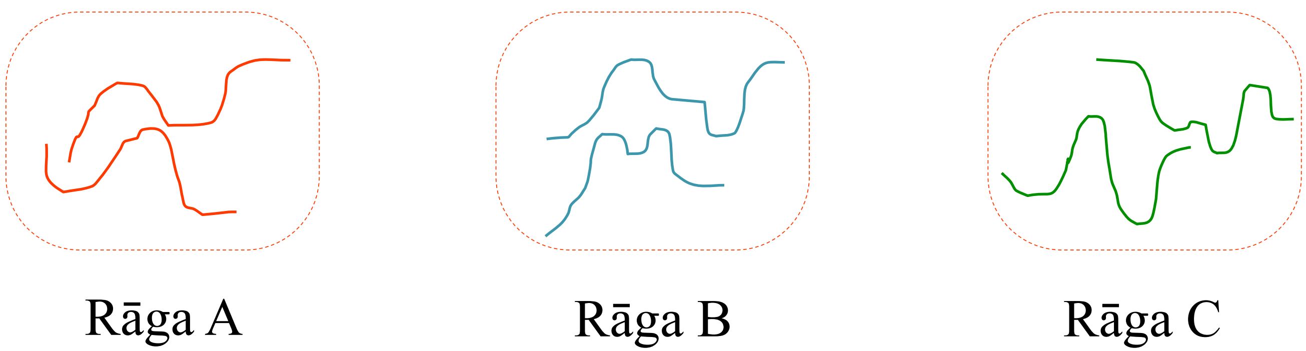
- S. Shetty and K. K. Achary, "Raga mining of indian music by extracting arohana-avarohana pattern," Int. Journal of Recent Trends in Engineering, vol. 1, no. 1, pp. 362–366, 2009.
- V. Kumar, H Pandya, and C. V. Jawahar, "Identifying ragas in indian music," in 22nd Int. Conf. on Pattern Recognition (ICPR), 2014, pp. 767–772.
- P. V. Rajkumar, K. P. Saishankar, and M. John, "Identification of Carnatic raagas using hidden markov models," in IEEE 9th Int. Symposium on Applied Machine Intelligence and Informatics (SAMI), 2011, pp. 107–110.

Rāga Characterization: Melodic motifs



Rāga Characterization: Melodic motifs

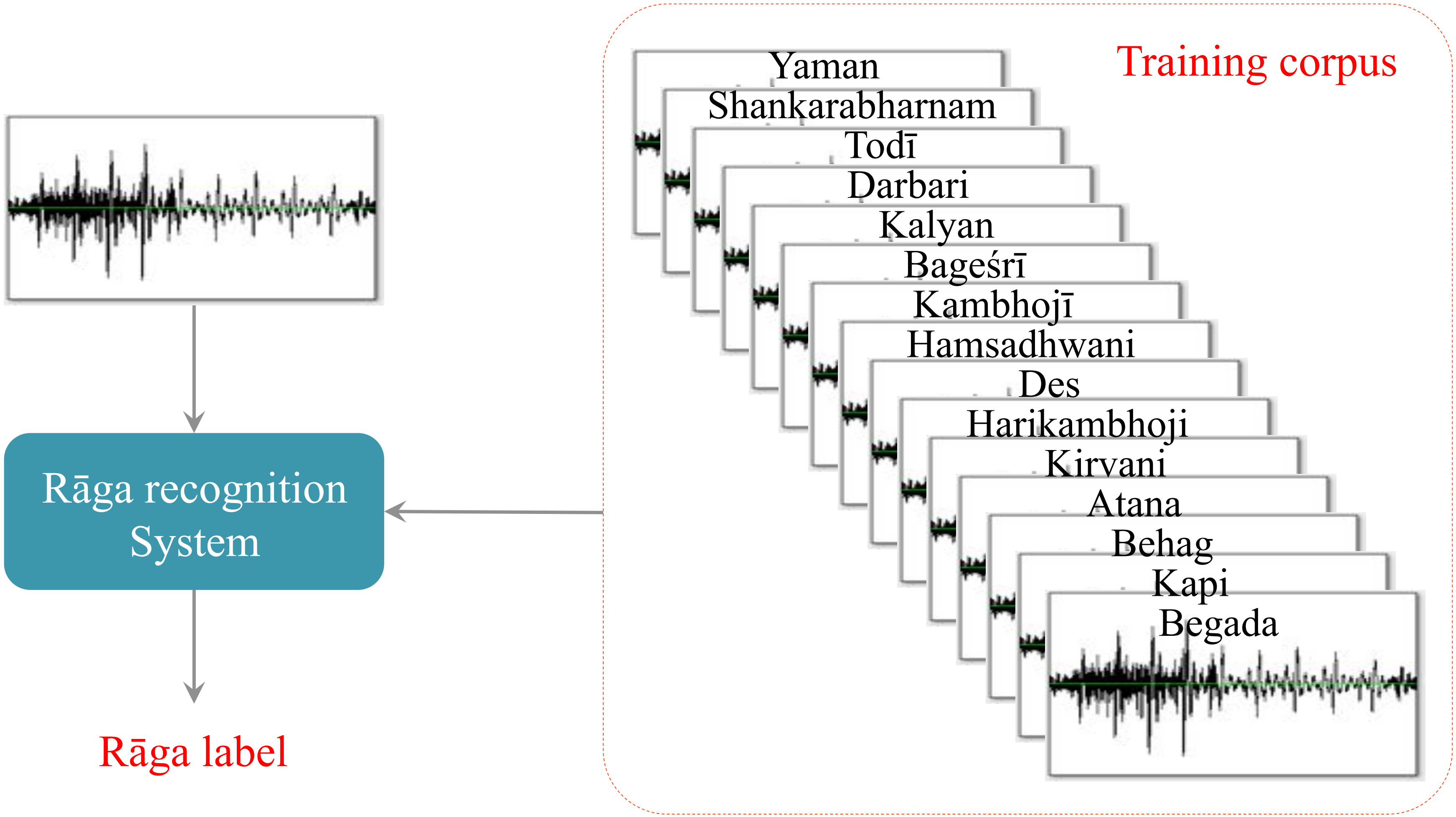




- R. Sridhar and T. V. Geetha, "Raga identification of carnatic music for music information retrieval," International Journal of Recent Trends in Engineering, vol. 1, no. 1, pp. 571–574, 2009.
- S. Dutta, S. PV Krishnaraj, and H. A. Murthy, "Raga verification in carnatic music using longest common segment set," in Int. Soc. for Music Information Retrieval Conf. (ISMIR), pp. 605-611,2015

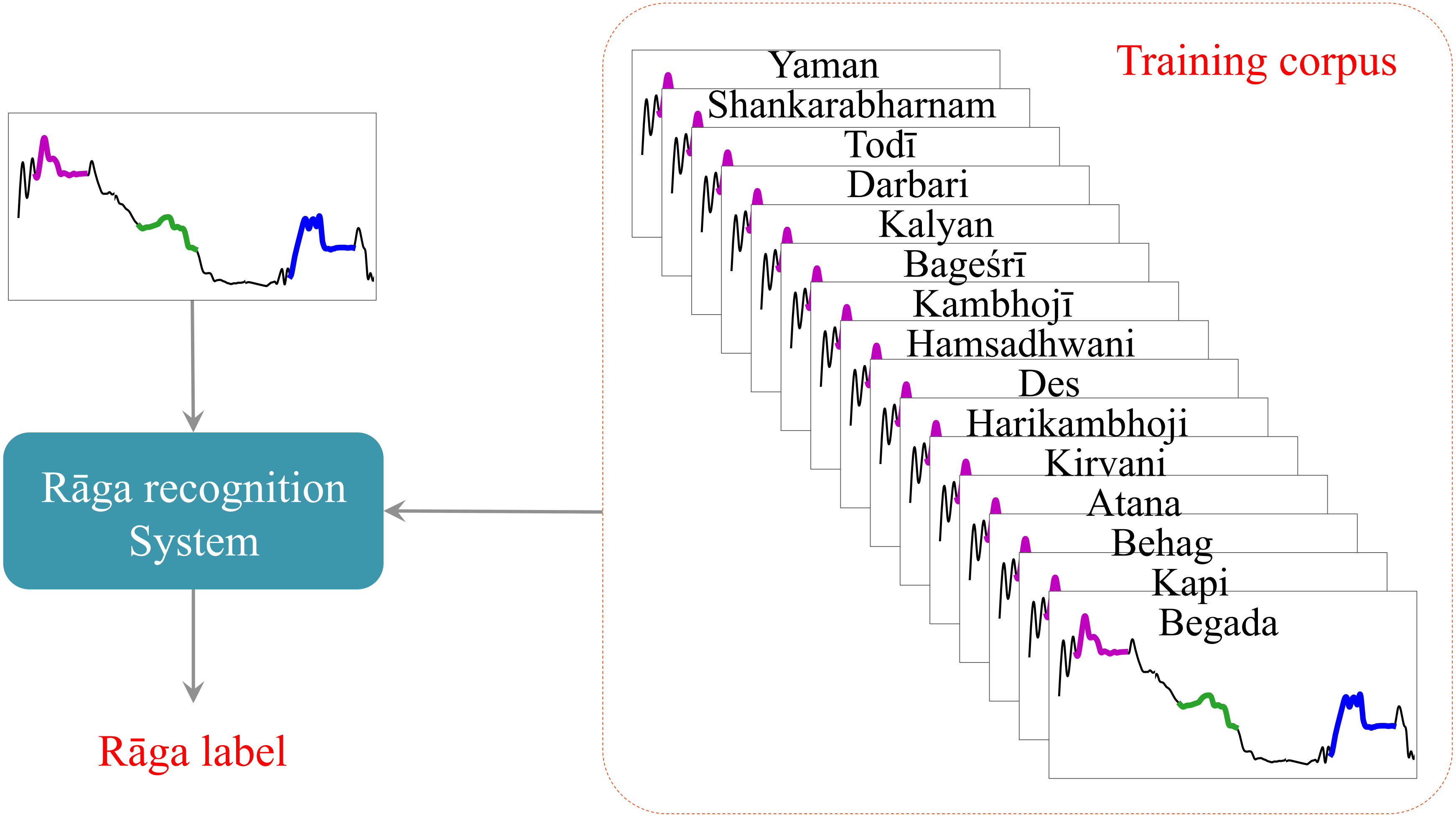
Goal

- Automatic rāga recognition

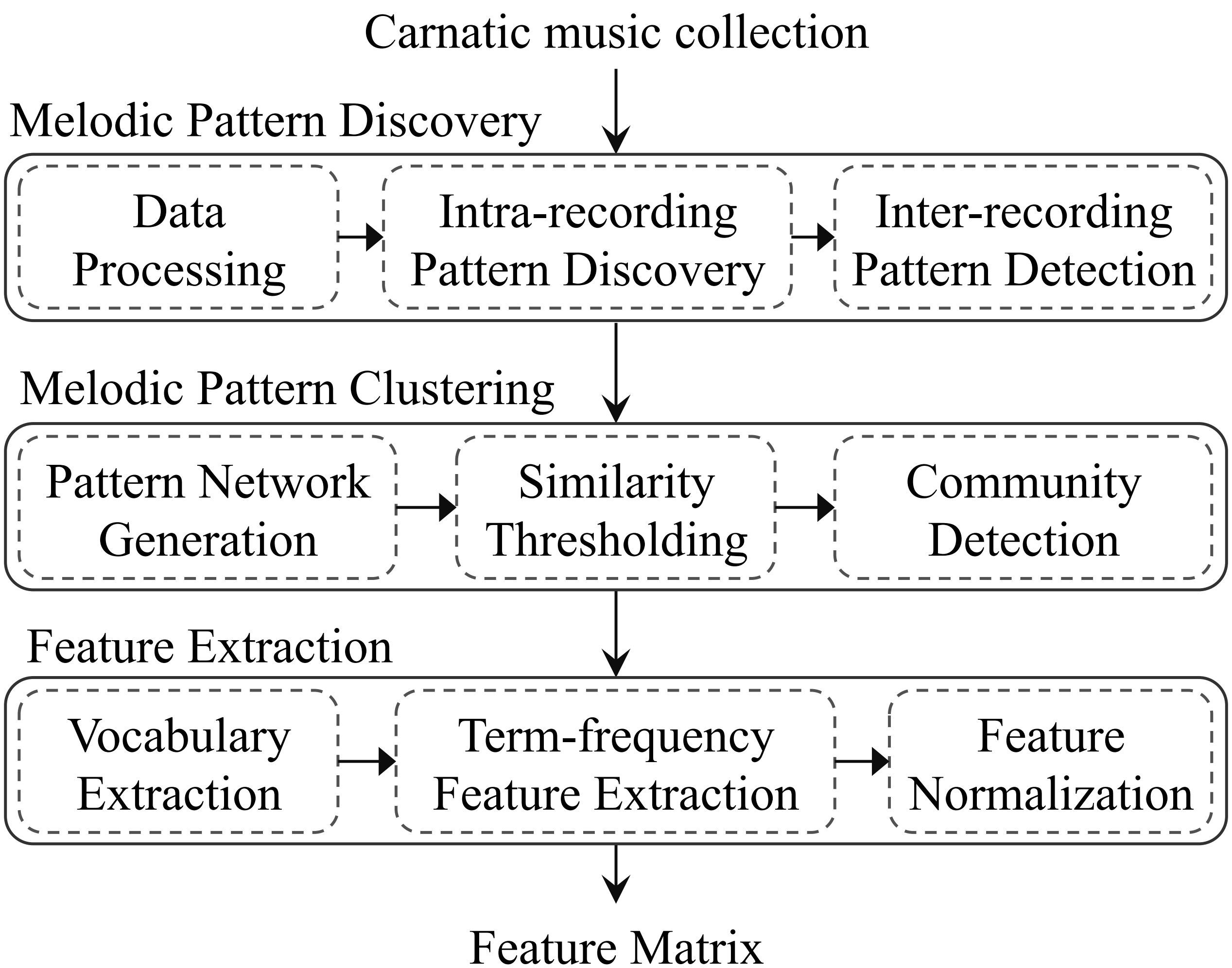


Goal

- Automatic rāga recognition

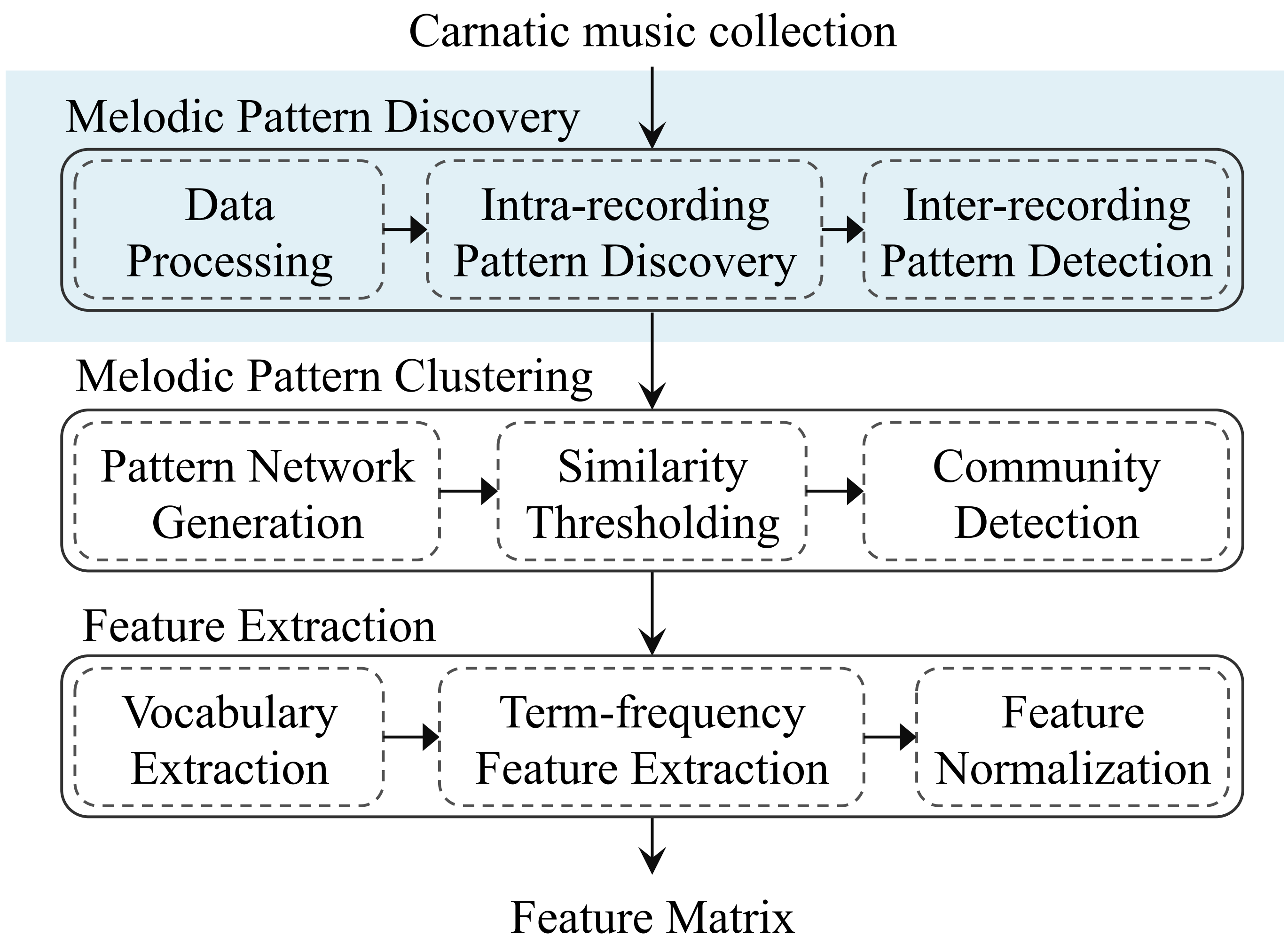


Block Diagram: Proposed Approach

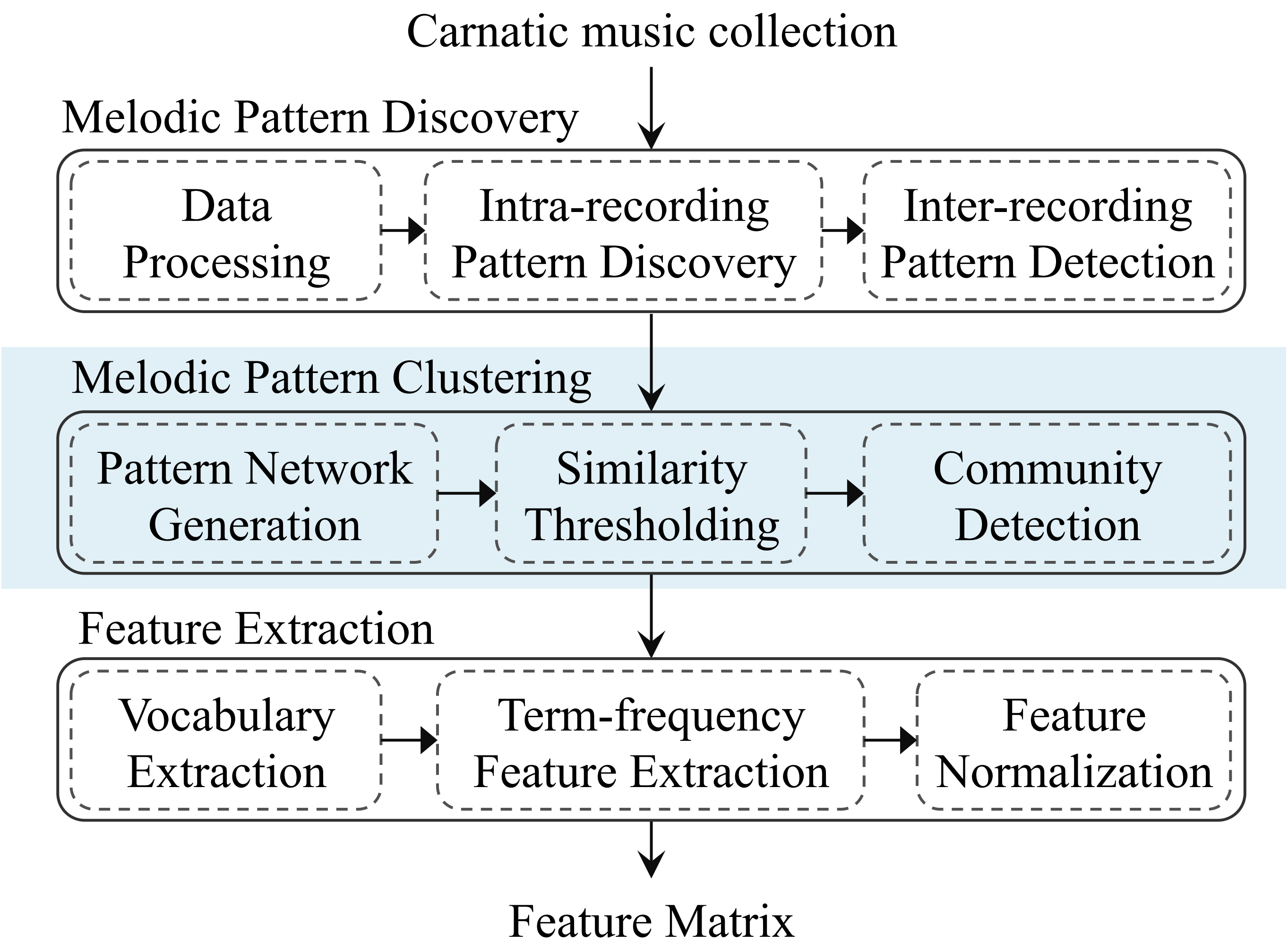


upf. Music Technology Group

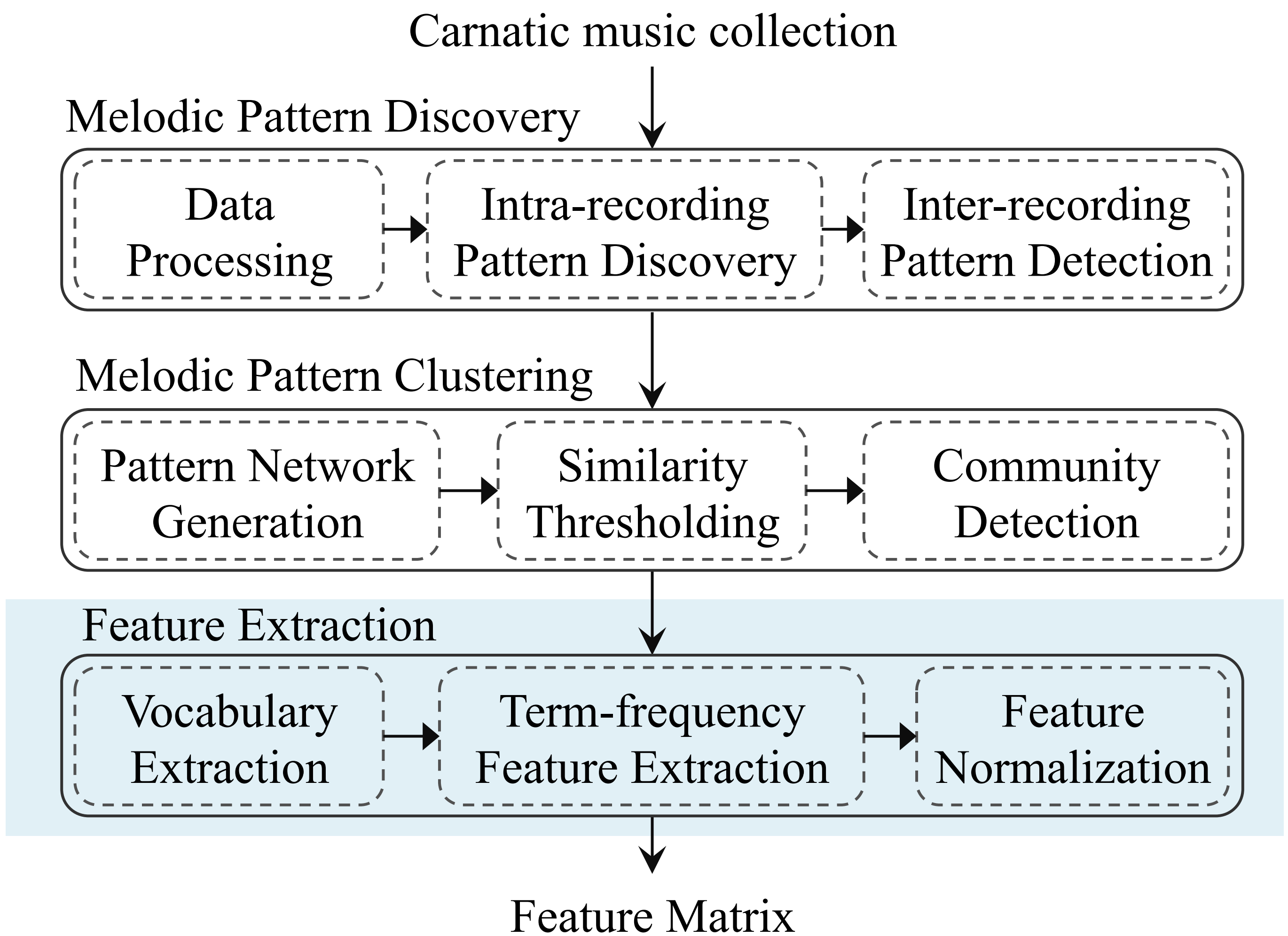
Block Diagram: Pattern Discovery



Block Diagram: Pattern Clustering

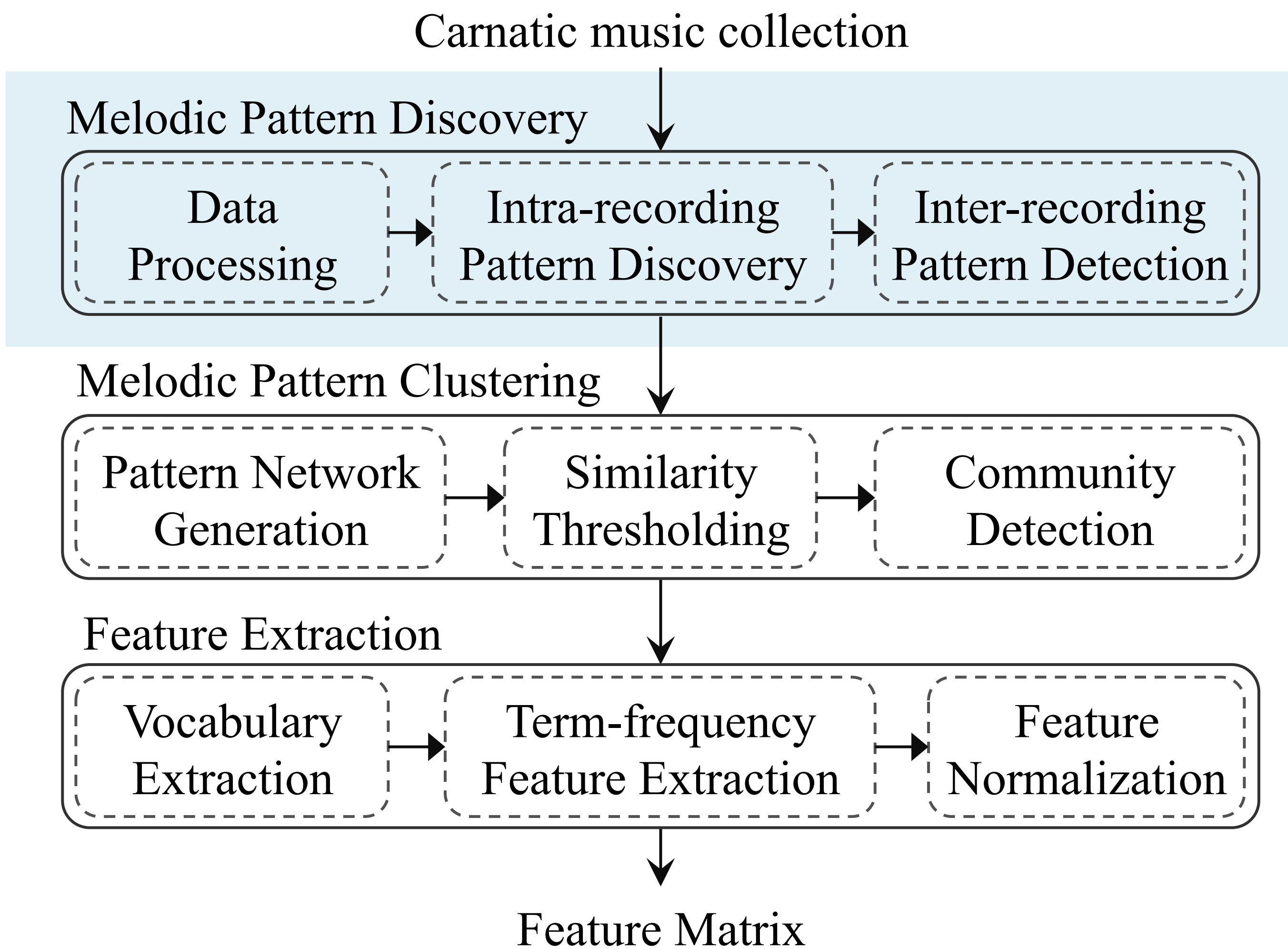


Block Diagram: Feature Extraction

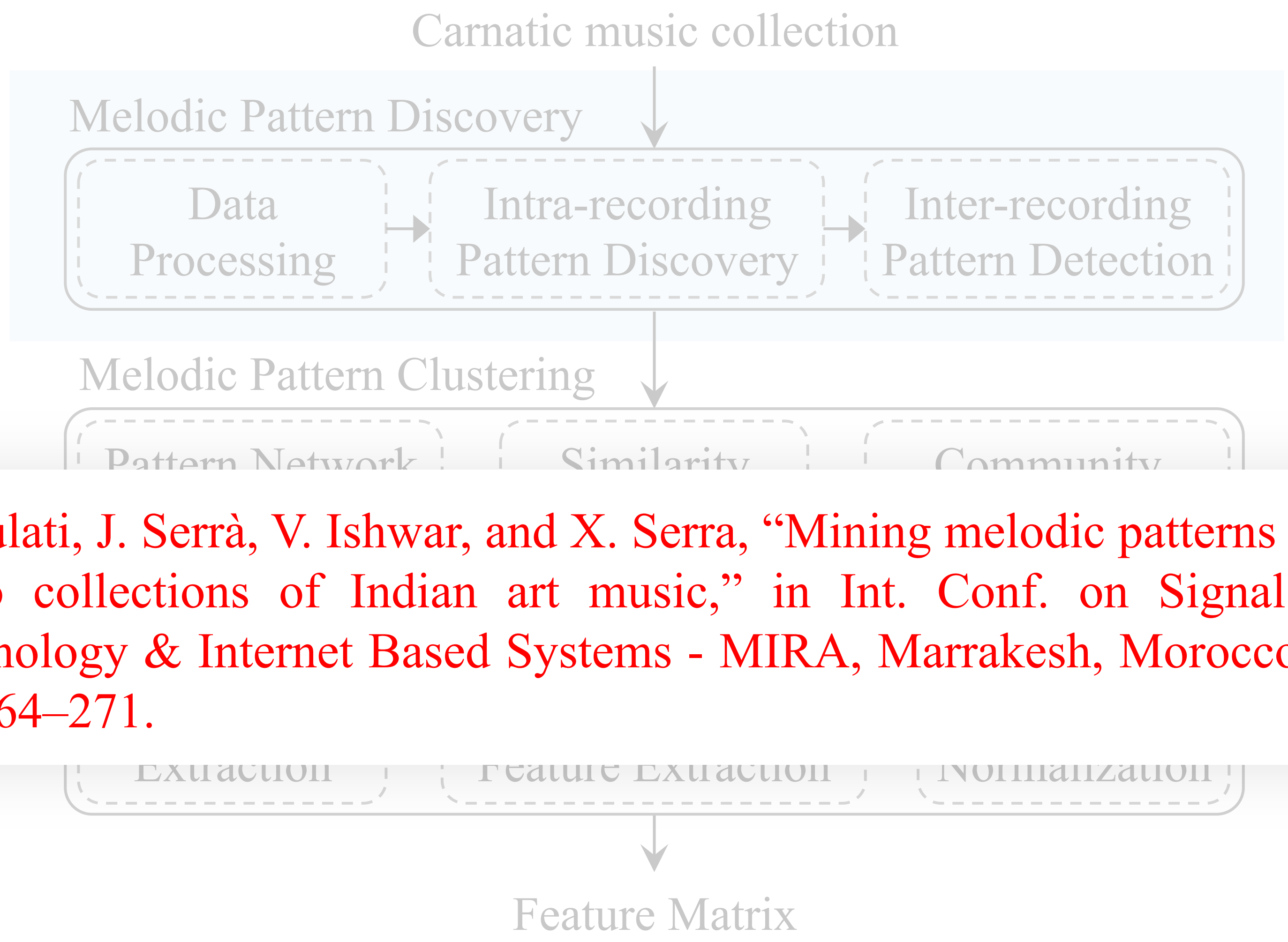


upf. Music Technology Group

Block Diagram: Pattern Discovery

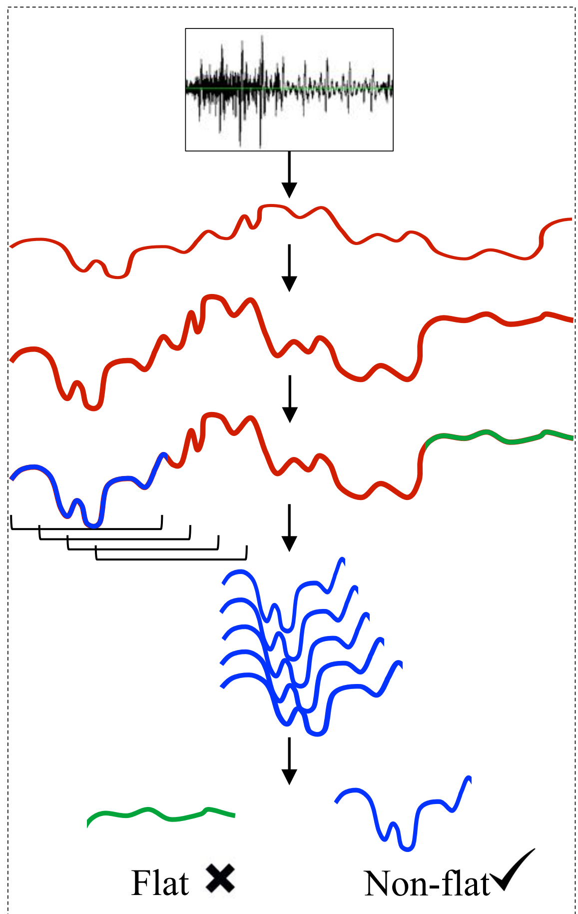


Block Diagram: Pattern Discovery



S. Gulati, J. Serrà, V. Ishwar, and X. Serra, "Mining melodic patterns in large audio collections of Indian art music," in Int. Conf. on Signal Image Technology & Internet Based Systems - MIRA, Marrakesh, Morocco, 2014, pp. 264–271.

Proposed Approach: Pattern Discovery



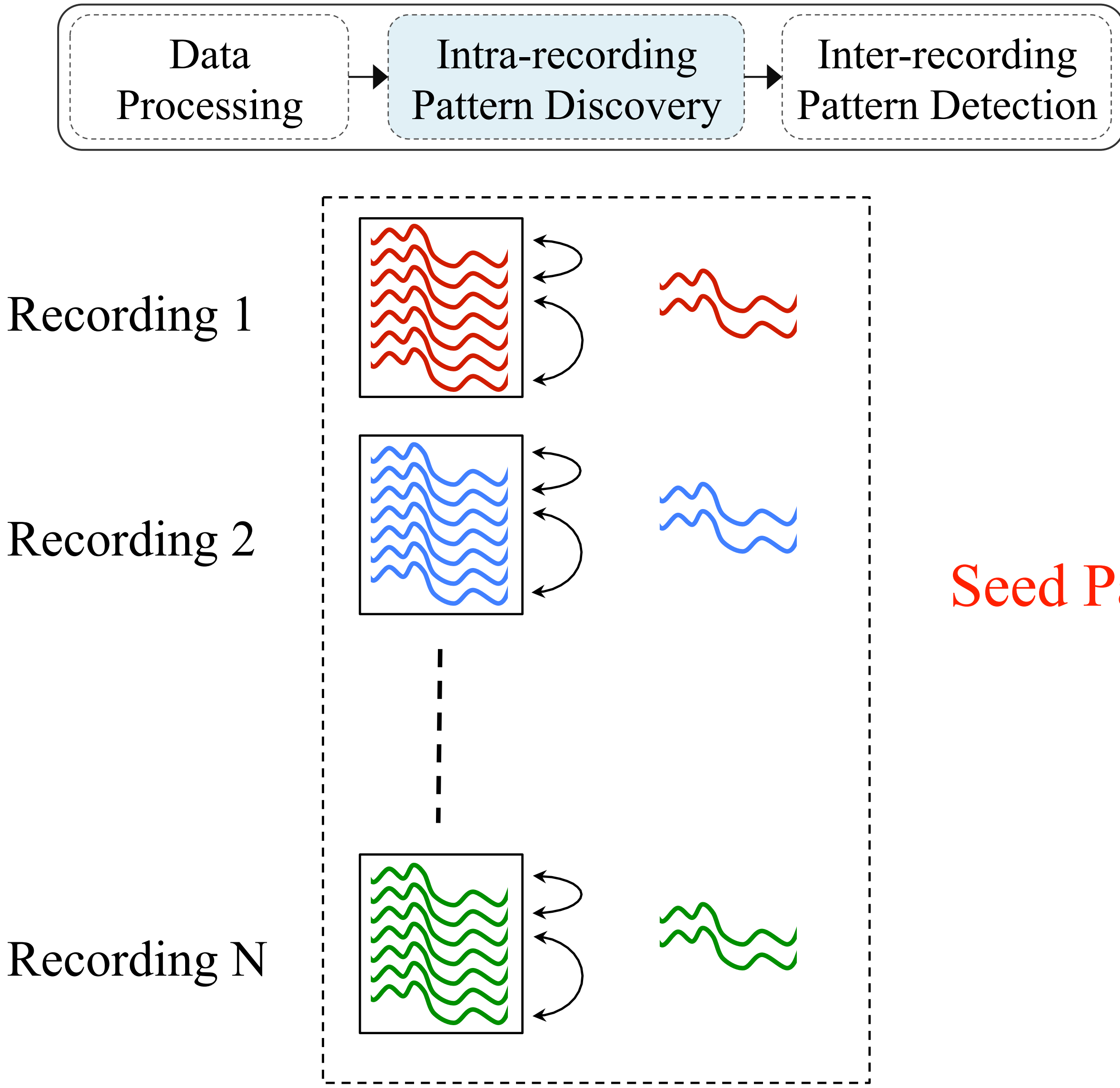
-Predominant pitch estimation

-Downsampling
-Hz to Cents
-Tonic normalization
-Brute-force segmentation

-Uniform Time-scaling

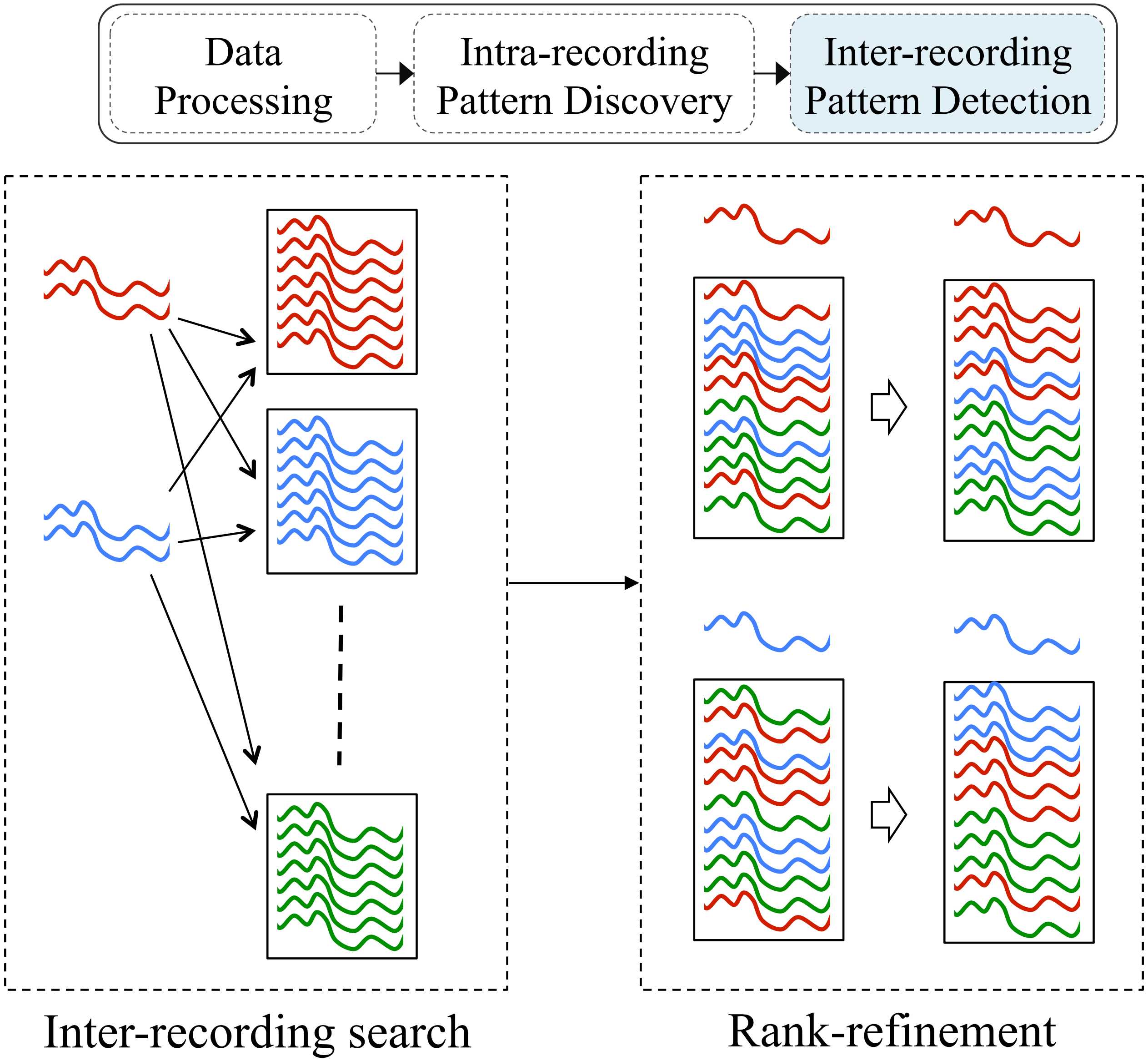
-Segment filtering

Proposed Approach: Pattern Discovery

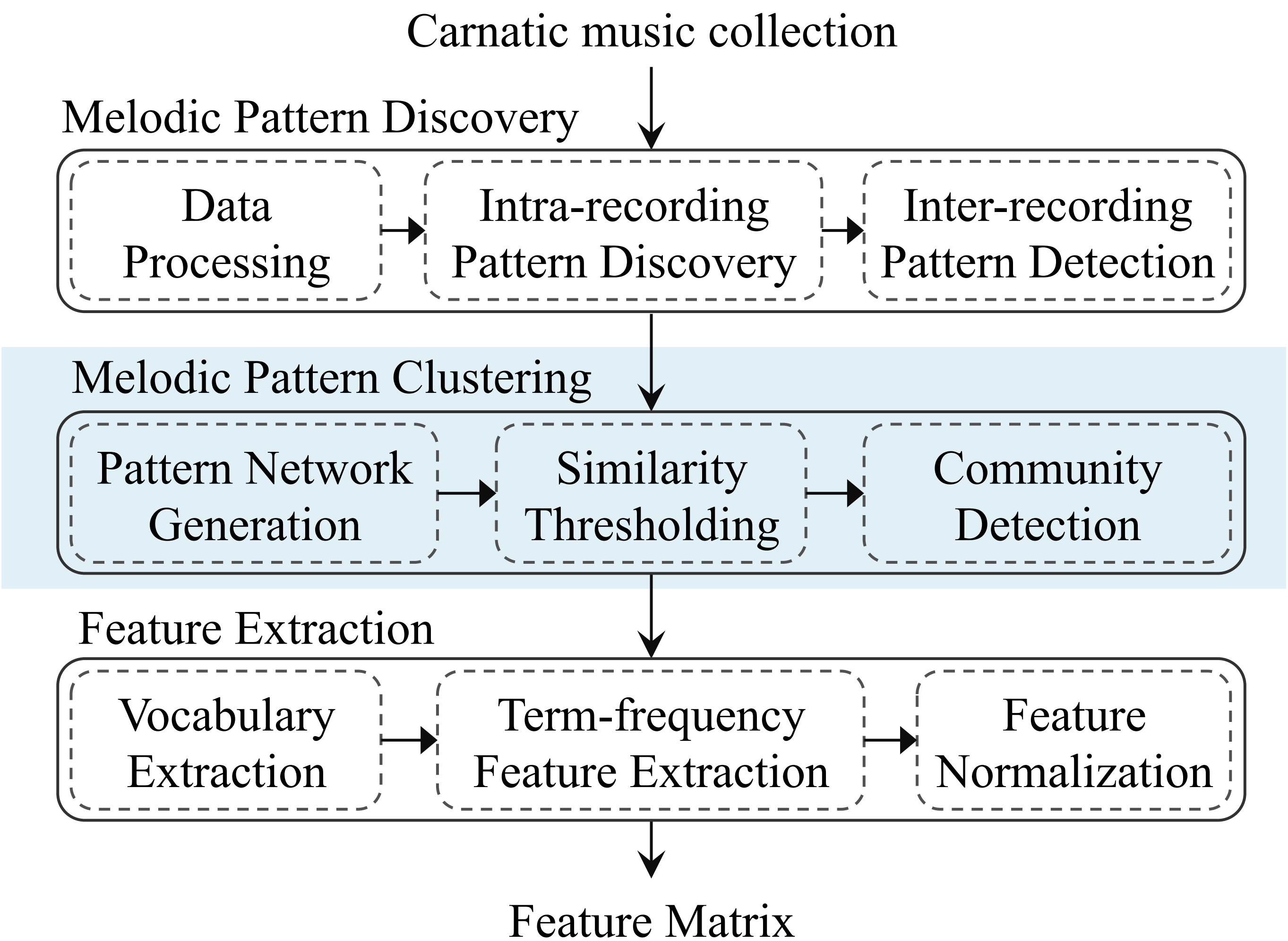


Seed Patterns

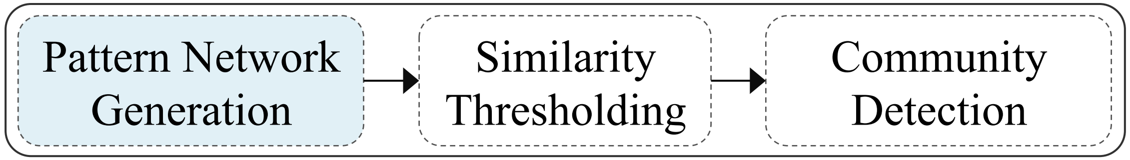
Proposed Approach: Pattern Discovery

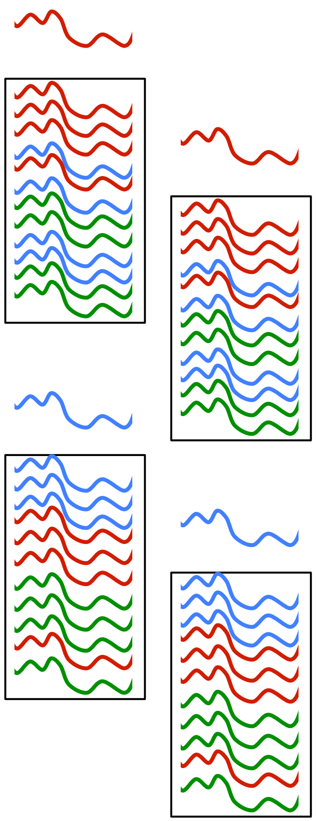


Block Diagram: Pattern Clustering



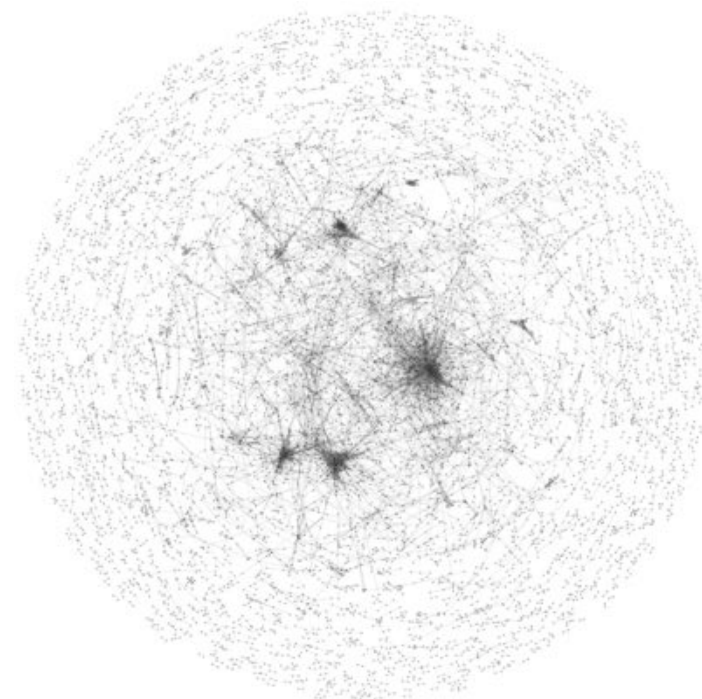
Proposed Approach: Pattern Clustering





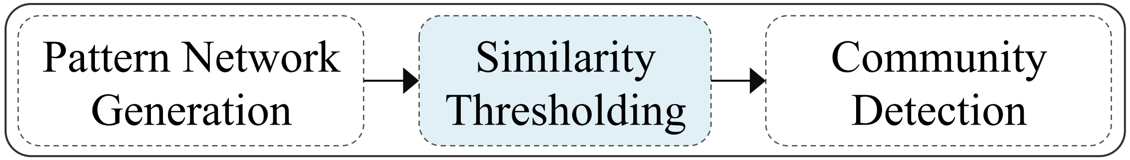
$$W_{ij} = e^{-D_{ij}/\overline{D}}$$

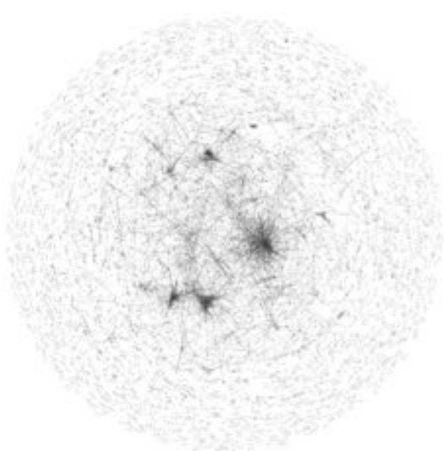
Undirectional

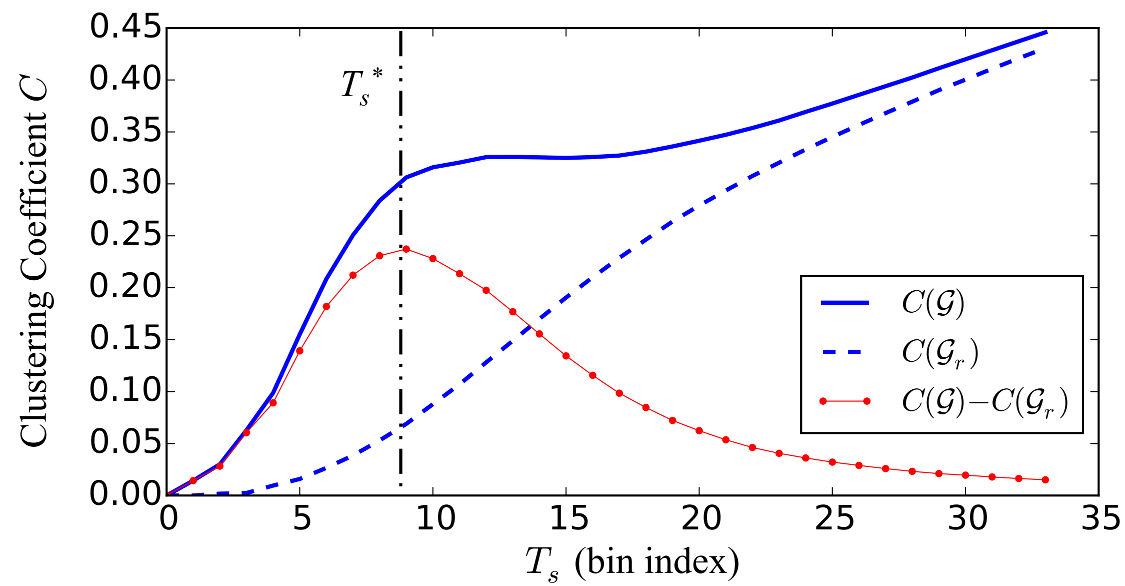


- M. EJ Newman, "The structure and function of complex networks," Society for Industrial and Applied Mathematics (SIAM) review, vol. 45, no. 2, pp. 167–256, 2003.

Proposed Approach: Pattern Clustering



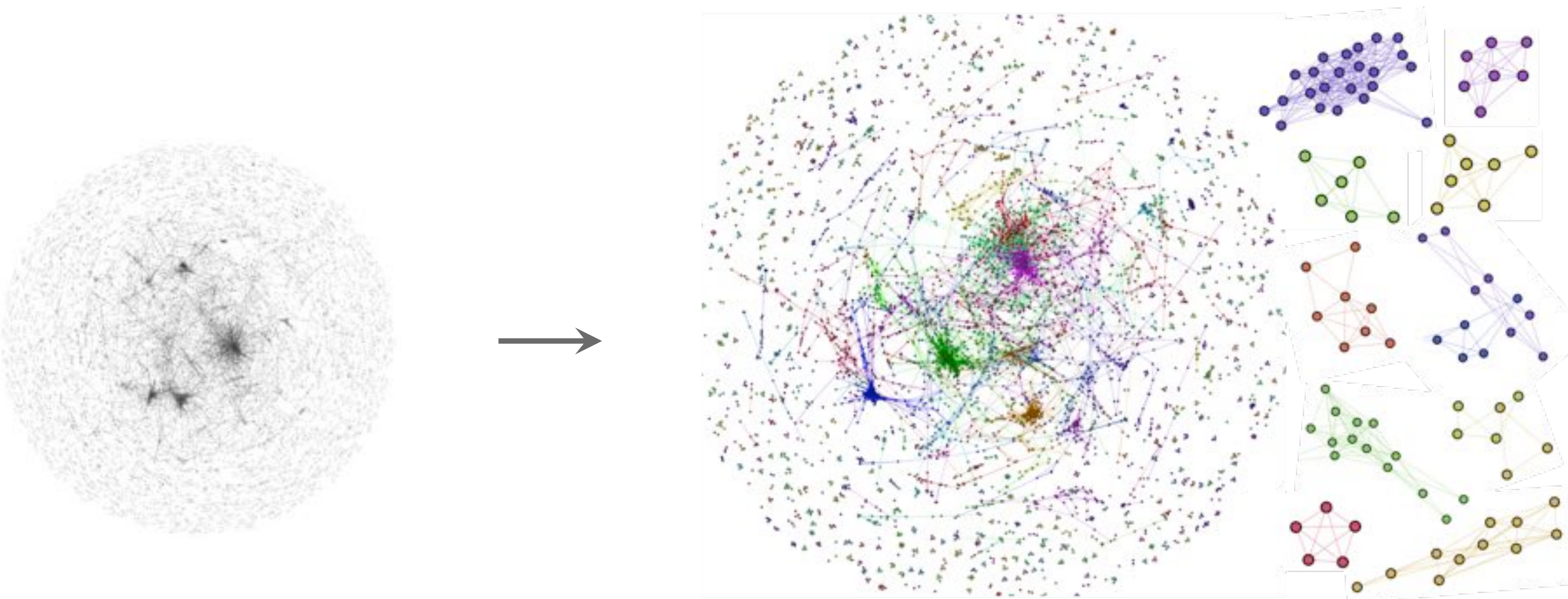




- M. EJ Newman, "The structure and function of complex networks," Society for Industrial and Applied Mathematics (SIAM) review, vol. 45, no. 2, pp. 167–256, 2003.
- S. Maslov and K. Sneppen, "Specificity and stability in topology of protein networks," Science, vol. 296, no. 5569, pp. 910– 913, 2002.

Proposed Approach: Pattern Clustering

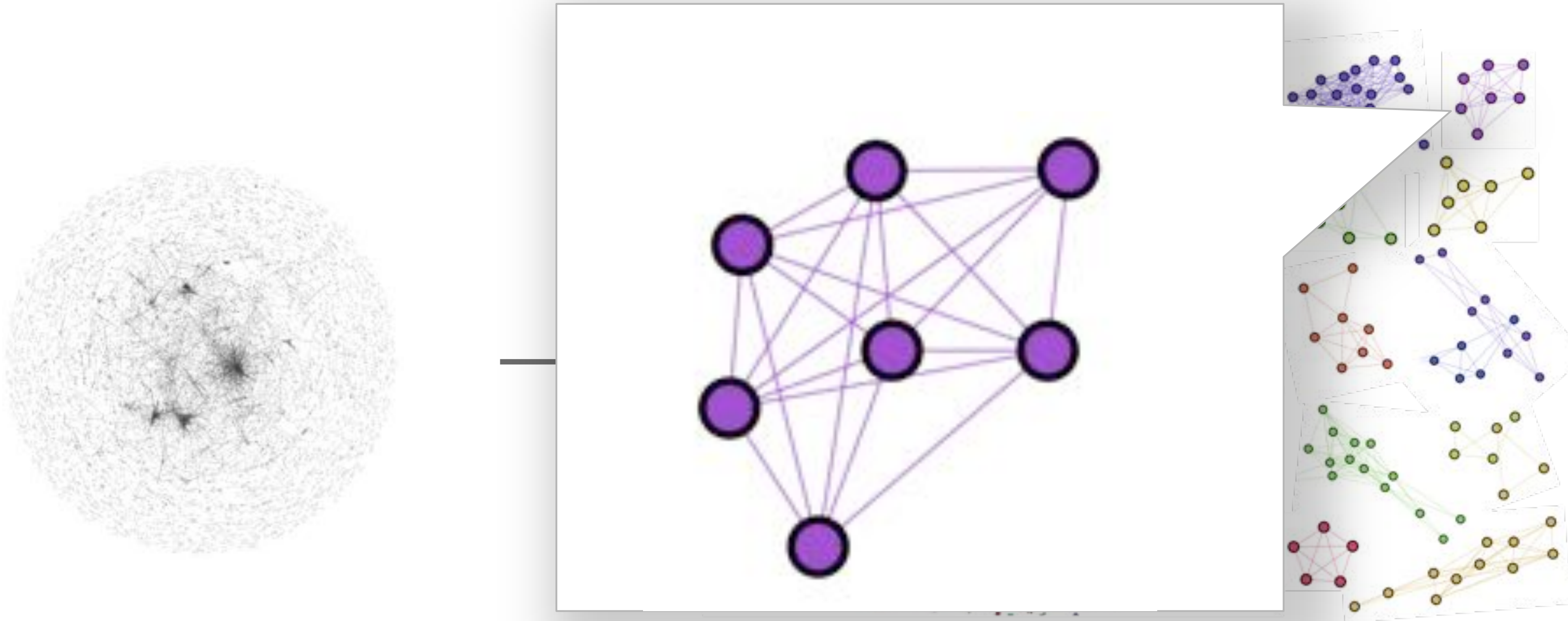
Pattern Network Generation → Similarity Thresholding → Community Detection



- V. D. Blondel, J. L. Guillaume, R. Lambiotte, and E. Lefebvre, "Fast unfolding of communities in large networks," Journal of Statistical Mechanics: Theory and Experiment, vol. 2008, no. 10, pp. P10008, 2008.
- S Fortunato, "Community detection in graphs," *Physics Reports*, vol. 486, no. 3, pp. 75–174, 2010.

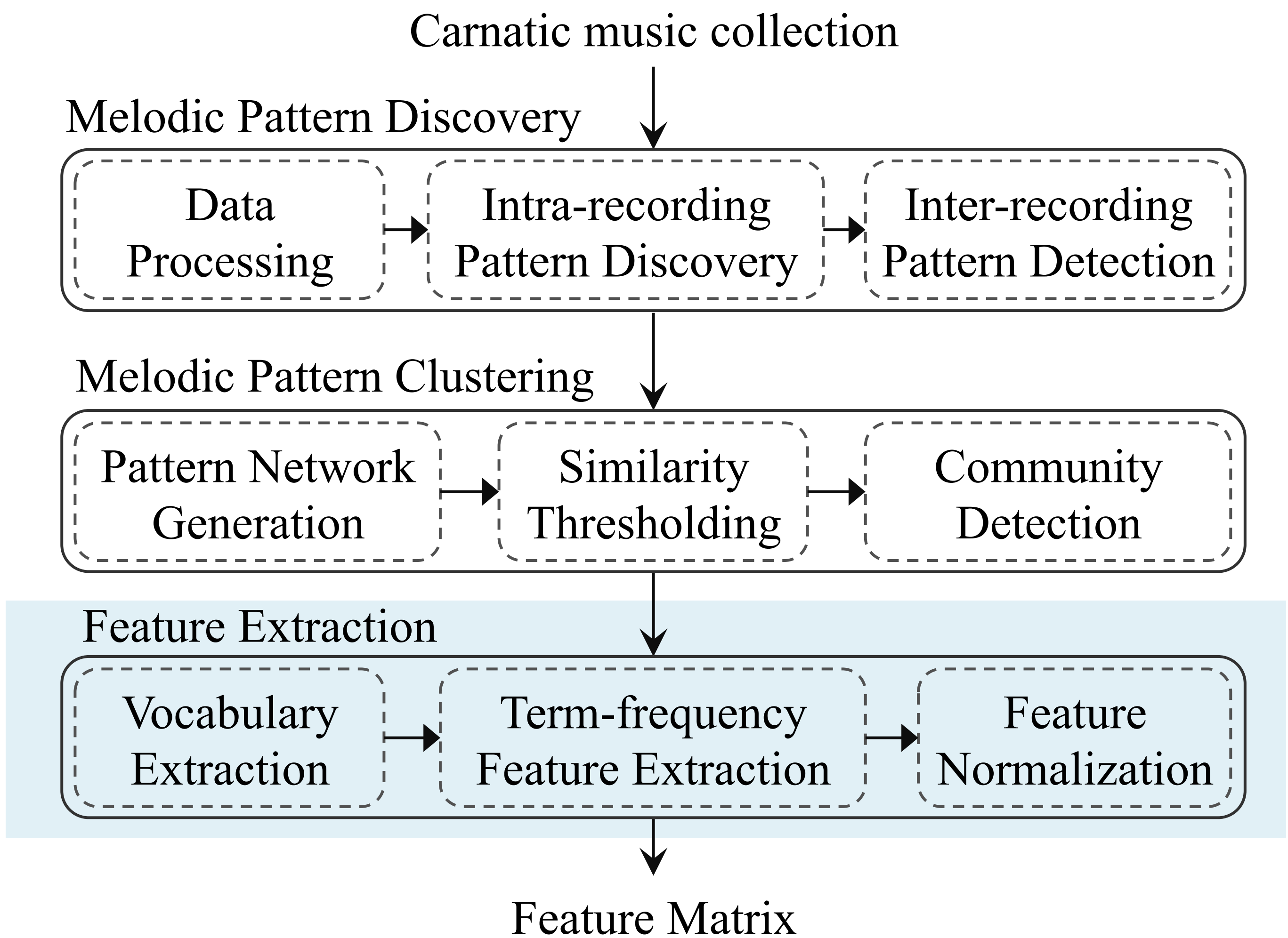
Proposed Approach: Pattern Clustering

Pattern Network Generation → Similarity Thresholding → Community Detection

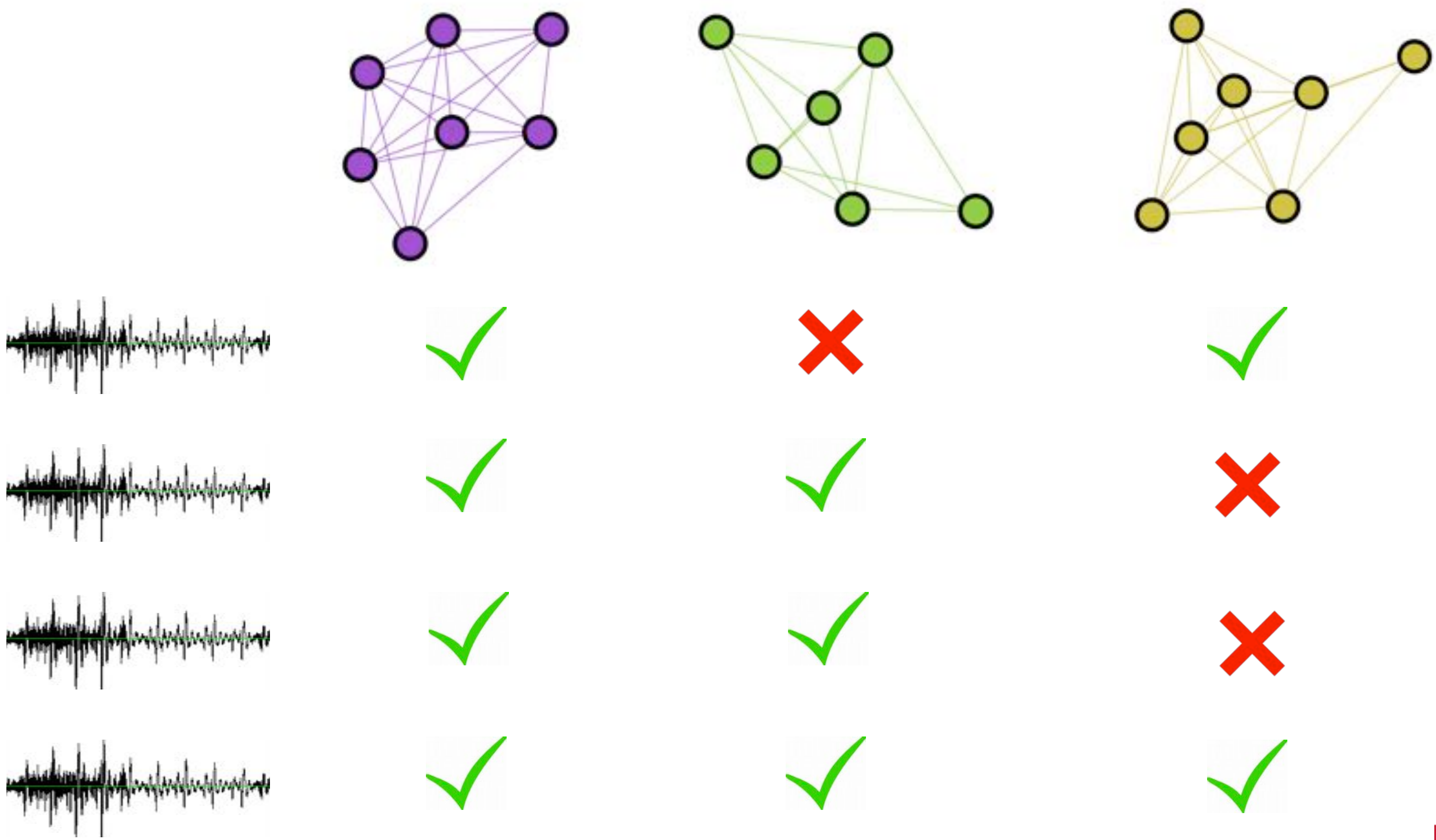


- V. D. Blondel, J. L. Guillaume, R. Lambiotte, and E. Lefebvre, "Fast unfolding of communities in large networks," Journal of Statistical Mechanics: Theory and Experiment, vol. 2008, no. 10, pp. P10008, 2008.
- S Fortunato, "Community detection in graphs," *Physics Reports*, vol. 486, no. 3, pp. 75–174, 2010.

Block Diagram: Feature Extraction



Proposed Approach: Feature Extraction



Proposed Approach: Feature Extraction

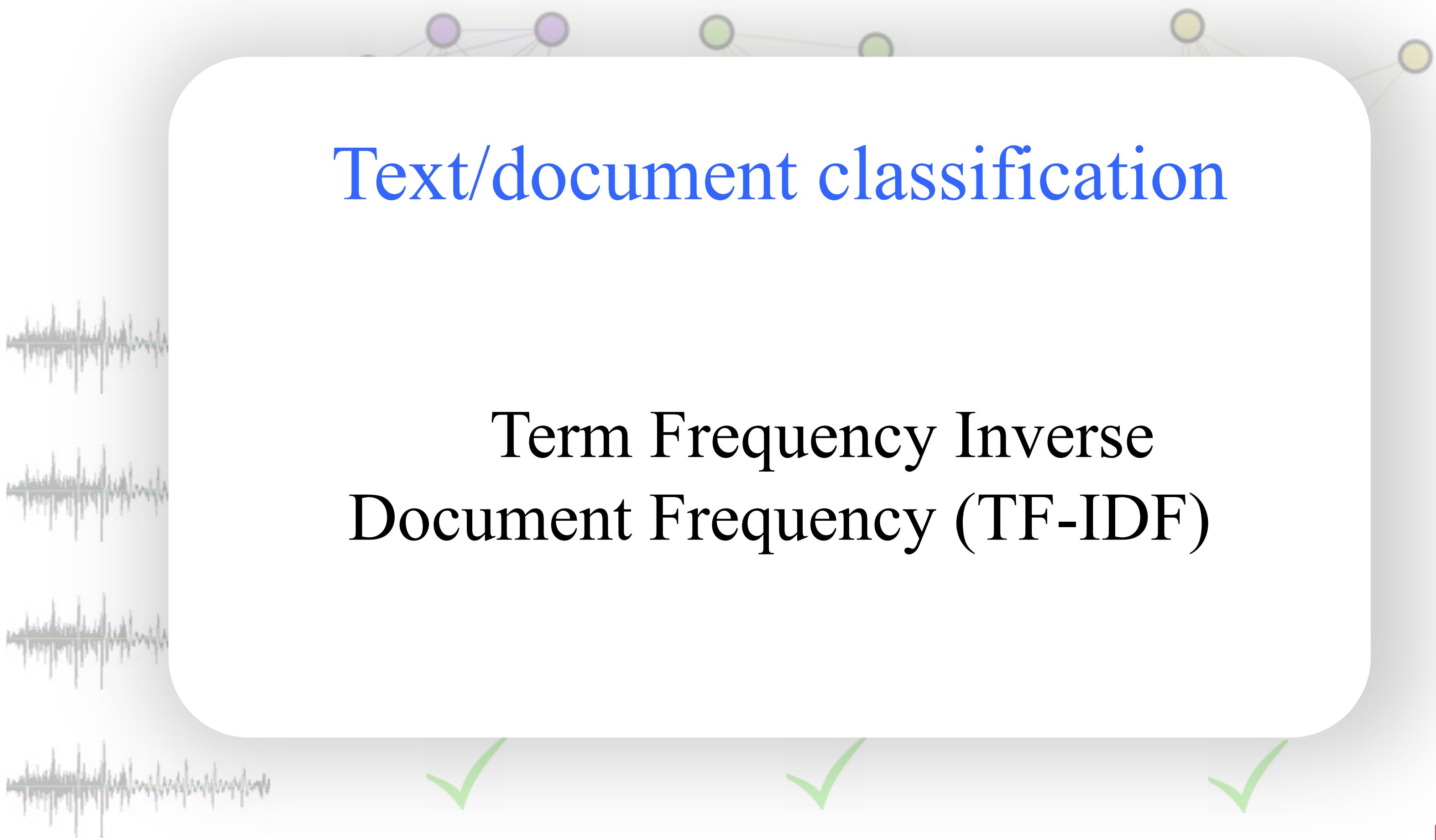
Text/document classification

Phrases ⟷ Words

Rāga ⟷ Topic

Recordings ⟷ Documents

Proposed Approach: Feature Extraction

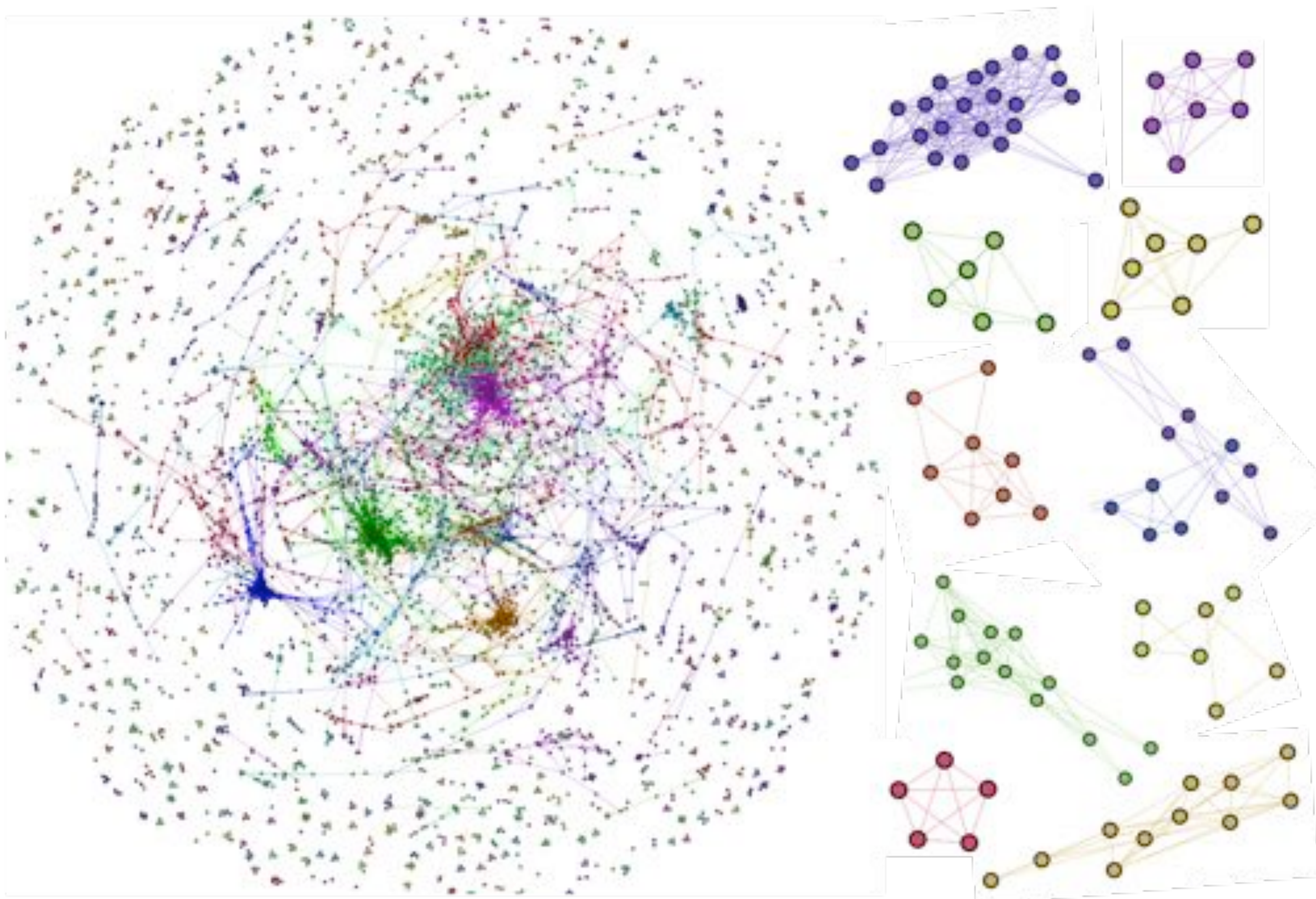


Text/document classification

Term Frequency Inverse Document Frequency (TF-IDF)

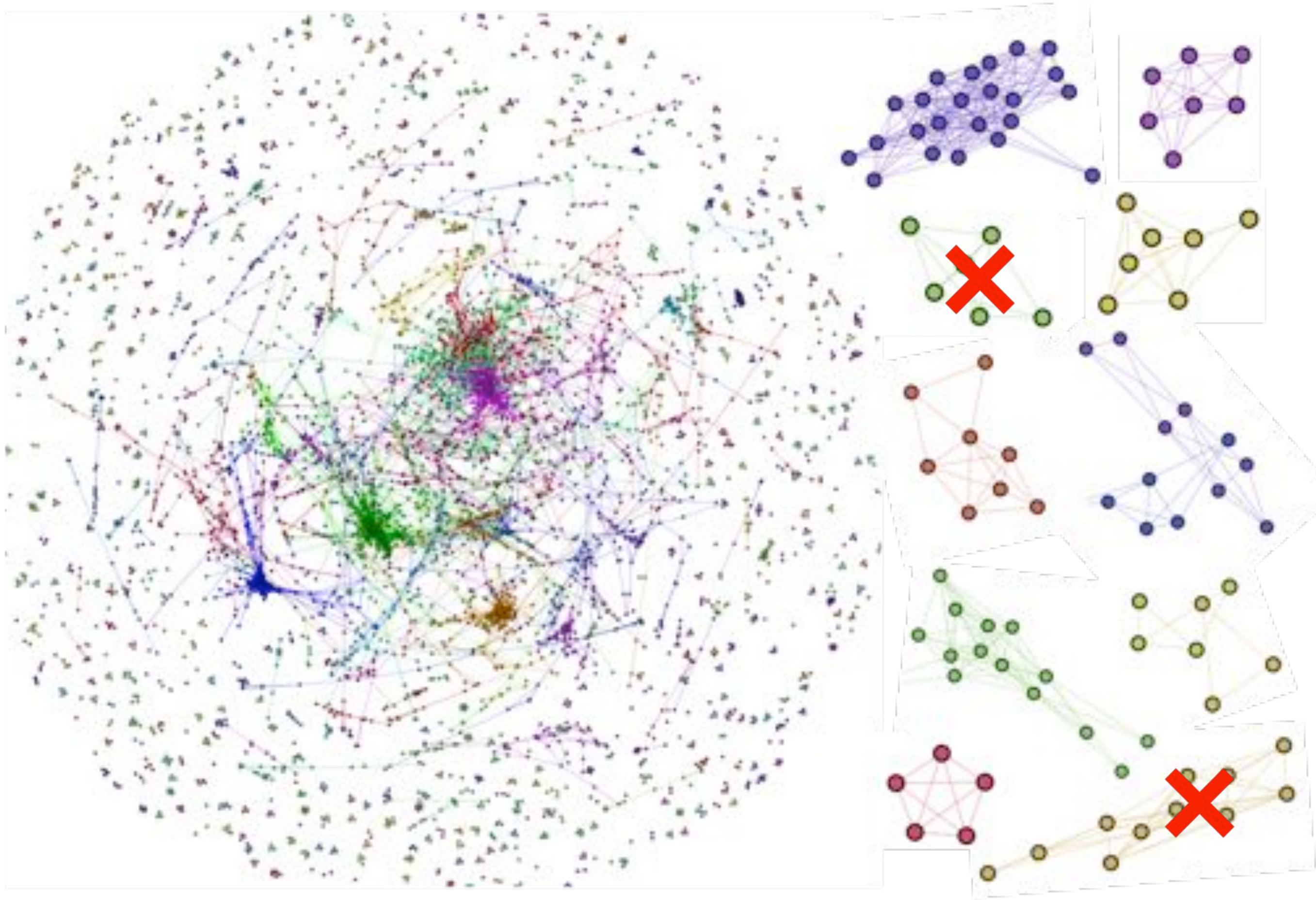
Proposed Approach: Feature Extraction

Vocabulary Extraction → Term-frequency Feature Extraction → Feature Normalization



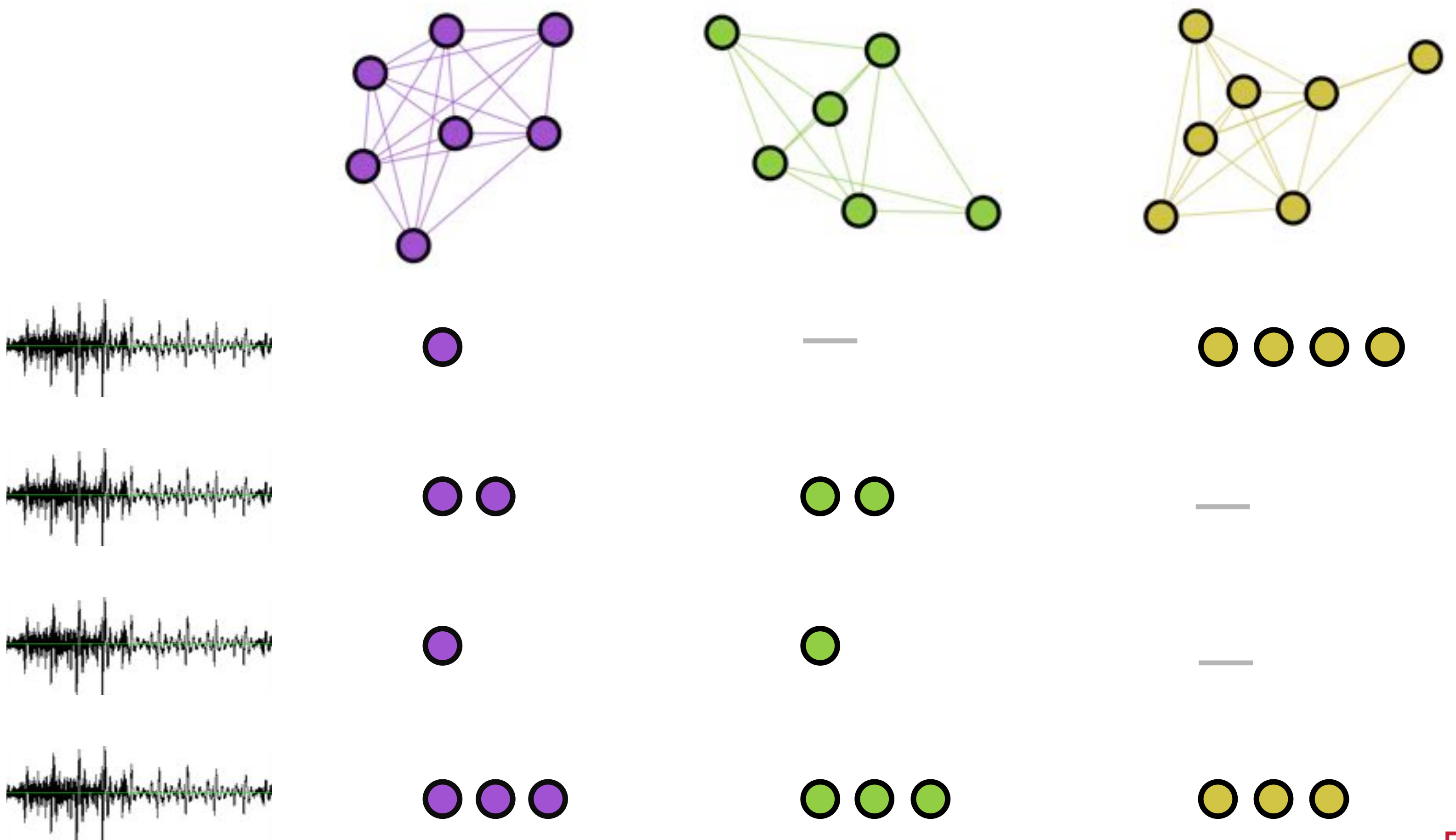
Proposed Approach: Feature Extraction

Vocabulary Extraction → Term-frequency Feature Extraction → Feature Normalization



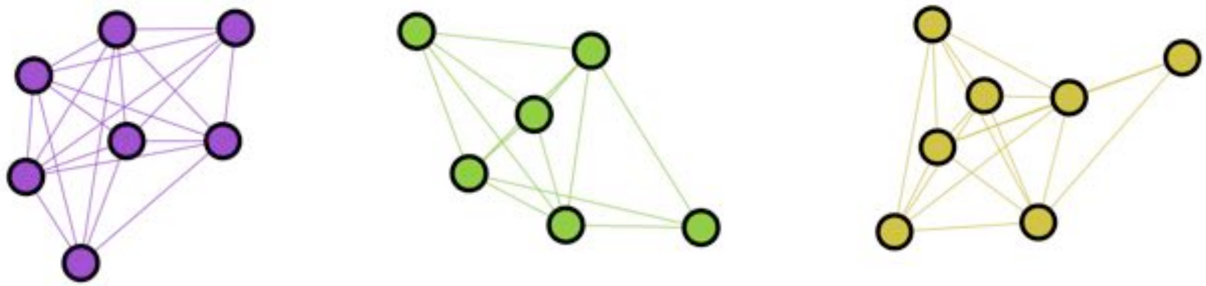
Proposed Approach: Feature Extraction

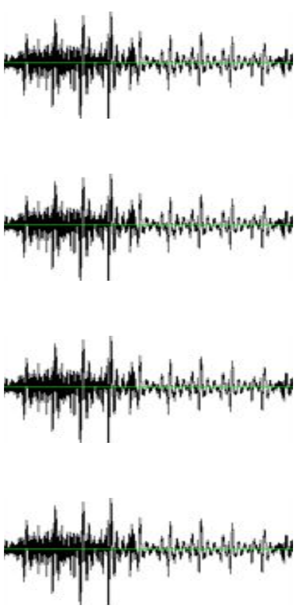
Vocabulary Extraction → Term-frequency Feature Extraction → Feature Normalization



Proposed Approach: Feature Extraction

Vocabulary Extraction → Term-frequency Feature Extraction → Feature Normalization

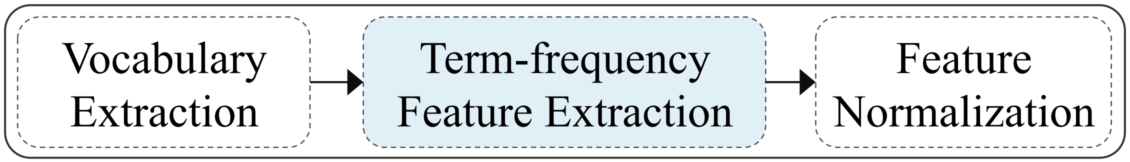




1	0	4
2	2	0
1	1	0
3	3	3

upf. Music Technology Group

Proposed Approach: Feature Extraction

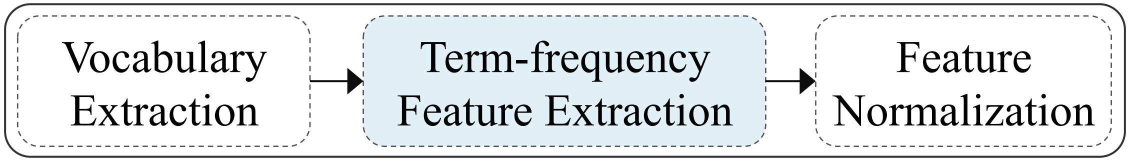


Proposed Approach: Feature Extraction

Vocabulary Extraction → Term-frequency Feature Extraction → Feature Normalization

$$F_1(p, r) = \begin{cases} 1, & \text{if} \quad f(p, r) > 0 \\ 0, & \text{otherwise} \end{cases}$$

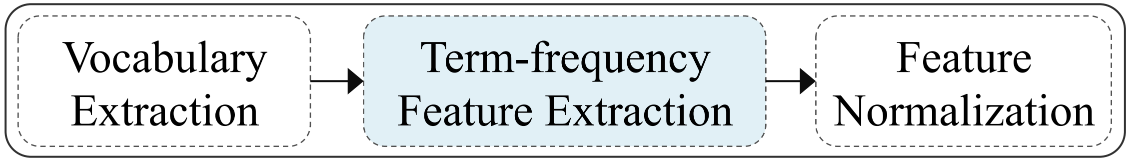
Proposed Approach: Feature Extraction



- $$F_1(p, r) = \begin{cases} 1, & \text{if} \quad f(p, r) > 0 \\ 0, & \text{otherwise} \end{cases}$$

- $$F_2(p, r) = f(p, r)$$

Proposed Approach: Feature Extraction



- $$\mathrm{F}_1(p, r) = \begin{cases} 1, & \text{if} \quad f(p, r) > 0 \\ 0, & \text{otherwise} \end{cases}$$

- $$\mathrm{F}_2(p, r) = f(p, r)$$

- $$\mathrm{F}_3(p, r) = f(p, r) \times \mathrm{irf}(p, R)$$

$$\mathrm{irf}(p, R) = \log\left(\frac{N}{|\{r \in R : p \in r\}|}\right)$$

Evaluation: Music Collection

Evaluation: Music Collection

- Corpus: CompMusic Carnatic music
  - Commercial released music (~325) CDs
  - Metadata available in Musicbrainz

Evaluation: Music Collection

- Corpus: CompMusic Carnatic music
  - Commercial released music (~325) CDs
  - Metadata available in Musicbrainz
- Datasets: subsets of corpus
  - DB40rāga
    - 480 audio recordings, 124 hours of music
    - 40 diverse set of rāgas
    - 310 compositions, 62 unique artists
  - DB10rāga
    - 10 rāga subset of DB40rāga

compmusic

MusicBrainz

Evaluation: Music Collection

- Corpus: CompMusic Carnatic music
  - Commercial released music (~325) CDs
  - Metadata available in Musicbrainz
- Datasets: subsets of corpus
  - DB40rāga
    - 480 audio recordings, 124 hours of music
    - 40 diverse set of rāgas
    - 310 compositions, 62 unique artists
  - DB10rāga
    - 10 rāga subset of DB40rāga

compmusic

MusicBrainz

http://compmusic.upf.edu/node/278

Evaluation: Classification methodology

- Experimental setup
  - Stratified 12-fold cross validation (balanced)
  - Repeat experiment 20 times
  - Evaluation measure: mean classification accuracy

Evaluation: Classification methodology

- Experimental setup
  - Stratified 12-fold cross validation (balanced)
  - Repeat experiment 20 times
  - Evaluation measure: mean classification accuracy
- Classifiers
  - Multinomial, Gaussian and Bernoulli naive Bayes (NBM, NBG and NBB)
  - SVM with a linear and RBF-kernel, and with a SGD learning (SVML, SVMR and SGD)
  - logistic regression (LR) and random forest (RF)

Evaluation: Classification methodology

- Statistical significance
  - Mann-Whitney U test ($p < 0.01$)
  - Multiple comparisons: Holm Bonferroni method

- 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 procedure,” Scandinavian journal of statistics, vol. 6, no. 2, pp. 65–70, 1979.

Evaluation: Classification methodology

- Statistical significance
  - Mann-Whitney U test ($p < 0.01$)
  - Multiple comparisons: Holm Bonferroni method
- Comparison with the state-of-the-art
  - S_1: Pitch-class-distribution (PCD)-based method (PCD_{120}, PCD_{full})
  - S_2: Parameterized (PCD)-based method (PCD_{param})

- 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 procedure," Scandinavian journal of statistics, vol. 6, no. 2, pp. 65–70, 1979.
- P. Chordia and S. Şentürk, "Joint recognition of raag and tonic in North Indian music," Computer Music Journal, vol. 37, no. 3, pp. 82–98, 2013.
- G.K.Koduri,V.Ishwar,J.Serrà,andX.Serra,"Intonation analysis of rāgas in Carnatic music," Journal of New Music Research, vol. 43, no. 1, pp. 72–93, 2014.

Results

db	Mtd	Ftr	NBM	NBB	LR	SVML	1NN
DB10rāga	M	F_1	90.6	74	84.1	81.2	-
		F_2	**91.7**	73.8	84.8	81.2	-
		F_3	90.5	74.5	84.3	80.7	-
	S_1	PCD_{120}	-	-	-	-	82.2
		PCD_{full}	-	-	-	-	**89.5**
	S_2	PD_{param}	37.9	11.2	**70.1**	65.7	-
DB40rāga	M	F_1	69.6	61.3	55.9	54.6	-
		F_2	**69.6**	61.7	55.7	54.3	-
		F_3	69.5	61.5	55.9	54.5	-
	S_1	PCD_{120}	-	-	-	-	66.4
		PCD_{full}	-	-	-	-	**74.1**
	S_2	PD_{param}	20.8	2.6	**51.4**	44.2	-

Results

db	Mtd	Ftr	NBM	NBB	LR	SVML	1NN
DB10rāga	M	F_1	90.6	74	84.1	81.2	-
		F_2	**91.7**	73.8	84.8	81.2	-
		F_3	90.5	74.5	84.3	80.7	-
	S_1	PCD_{120}	-	-	-	-	82.2
		PCD_{full}	-	-	-	-	**89.5**
	S_2	PD_{param}	37.9	11.2	**70.1**	65.7	-
DB40rāga	M	F_1	69.6	61.3	55.9	54.6	-
		F_2	**69.6**	61.7	55.7	54.3	-
		F_3	69.5	61.5	55.9	54.5	-
	S_1	PCD_{120}	-	-	-	-	66.4
		PCD_{full}	-	-	-	-	**74.1**
	S_2	PD_{param}	20.8	2.6	**51.4**	44.2	-

upf. Music Technology Group

Results

db	Mtd	Ftr	NBM	NBB	LR	SVML	1NN
DB10rāga	M	F_1	90.6	74	84.1	81.2	-
		F_2	**91.7**	73.8	84.8	81.2	-
		F_3	90.5	74.5	84.3	80.7	-
	S_1	PCD_{120}	-	-	-	-	82.2
		PCD_{full}	-	-	-	-	**89.5**
	S_2	PD_{param}	37.9	11.2	**70.1**	65.7	-
DB40rāga	M	F_1	69.6	61.3	55.9	54.6	-
		F_2	**69.6**	61.7	55.7	54.3	-
		F_3	69.5	61.5	55.9	54.5	-
	S_1	PCD_{120}	-	-	-	-	66.4
		PCD_{full}	-	-	-	-	**74.1**
	S_2	PD_{param}	20.8	2.6	**51.4**	44.2	-

Results

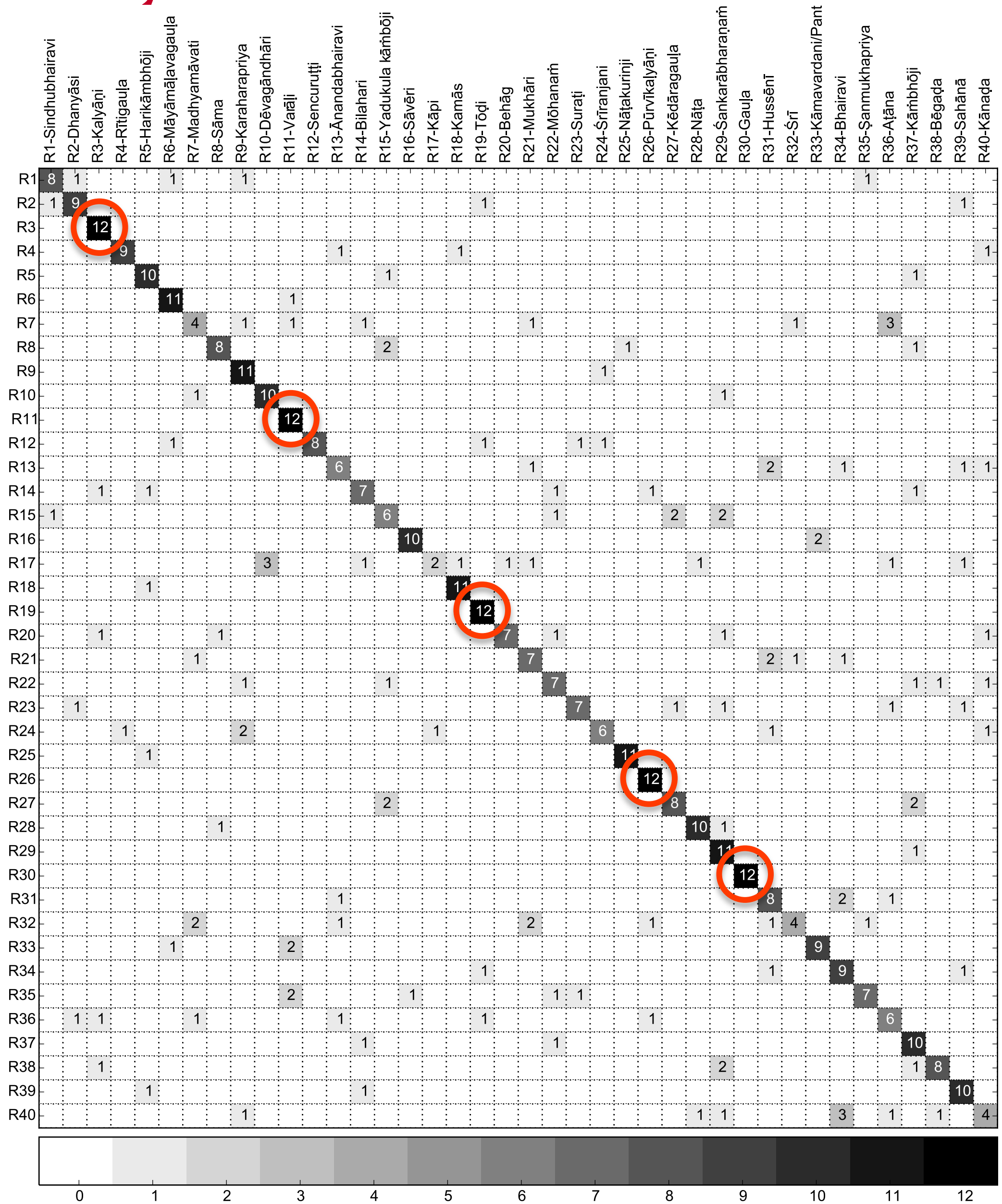
db	Mtd	Ftr	NBM	NBB	LR	SVML	1NN
DB10rāga	M	F_1	90.6	74	84.1	81.2	-
		F_2	**91.7**	73.8	84.8	81.2	-
		F_3	90.5	74.5	84.3	80.7	-
	S_1	PCD_{120}	-	-	-	-	82.2
		PCD_{full}	-	-	-	-	**89.5**
	S_2	PD_{param}	37.9	11.2	**70.1**	65.7	-
DB40rāga	M	F_1	69.6	61.3	55.9	54.6	-
		F_2	**69.6**	61.7	55.7	54.3	-
		F_3	69.5	61.5	55.9	54.5	-
	S_1	PCD_{120}	-	-	-	-	66.4
		PCD_{full}	-	-	-	-	**74.1**
	S_2	PD_{param}	20.8	2.6	**51.4**	44.2	-

upf. Music Technology Group

Results

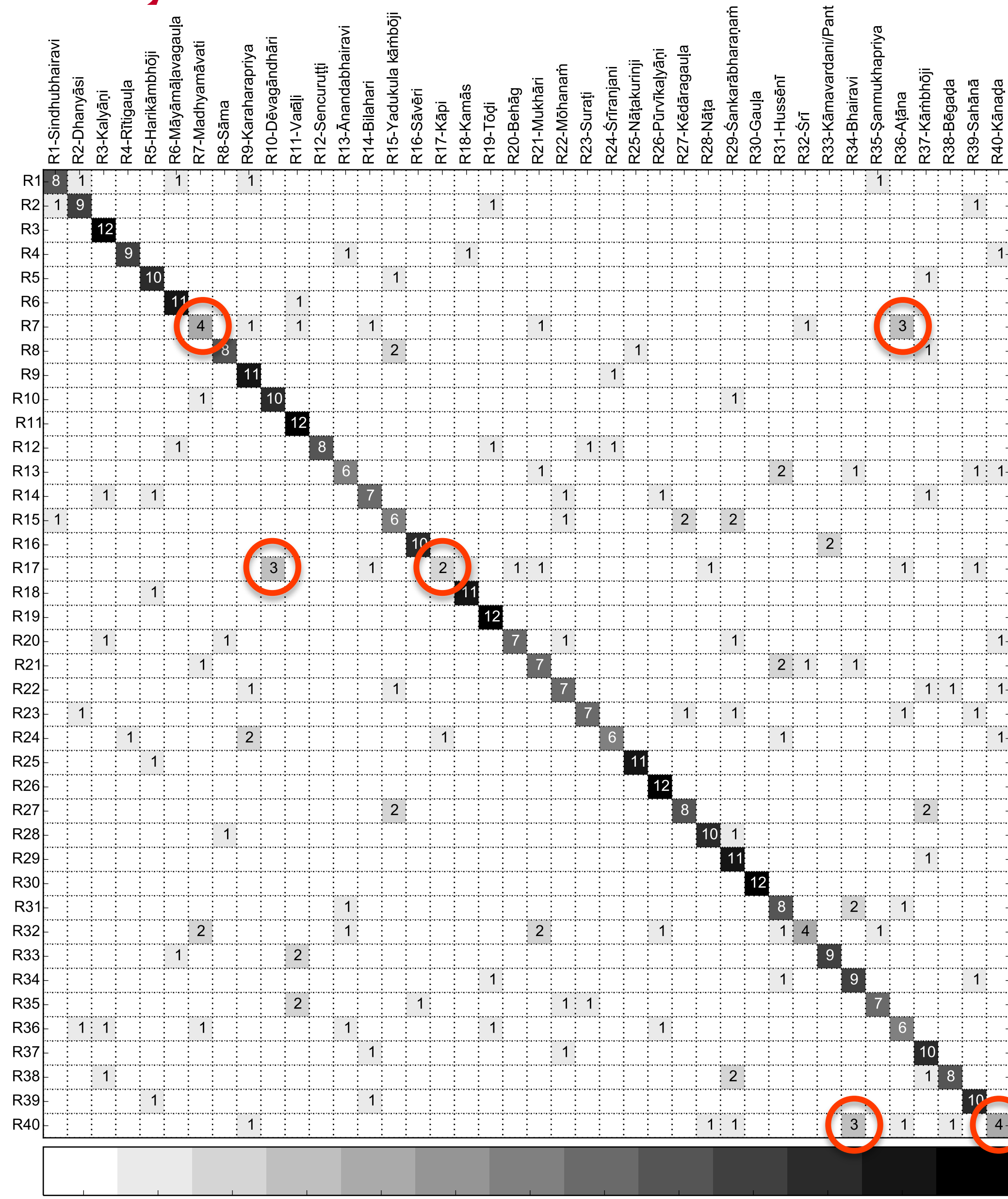
db	Mtd	Ftr	NBM	NBB	LR	SVML	1NN
DB10rāga	M	F_1	90.6	74	84.1	81.2	-
		F_2	**91.7**	73.8	84.8	81.2	-
		F_3	90.5	74.5	84.3	80.7	-
	S_1	PCD_{120}	-	-	-	-	82.2
		PCD_{full}	-	-	-	-	**89.5**
	S_2	PD_{param}	37.9	11.2	**70.1**	65.7	-
DB40rāga	M	F_1	69.6	61.3	55.9	54.6	-
		F_2	**69.6**	61.7	55.7	54.3	-
		F_3	69.5	61.5	55.9	54.5	-
	S_1	PCD_{120}	-	-	-	-	66.4
		PCD_{full}	-	-	-	-	**74.1**
	S_2	PD_{param}	20.8	2.6	**51.4**	44.2	-

Error Analysis



Kalyāṇi
Varāḷi
Tōḍi
Pūrvikalyāṇi

Error Analysis

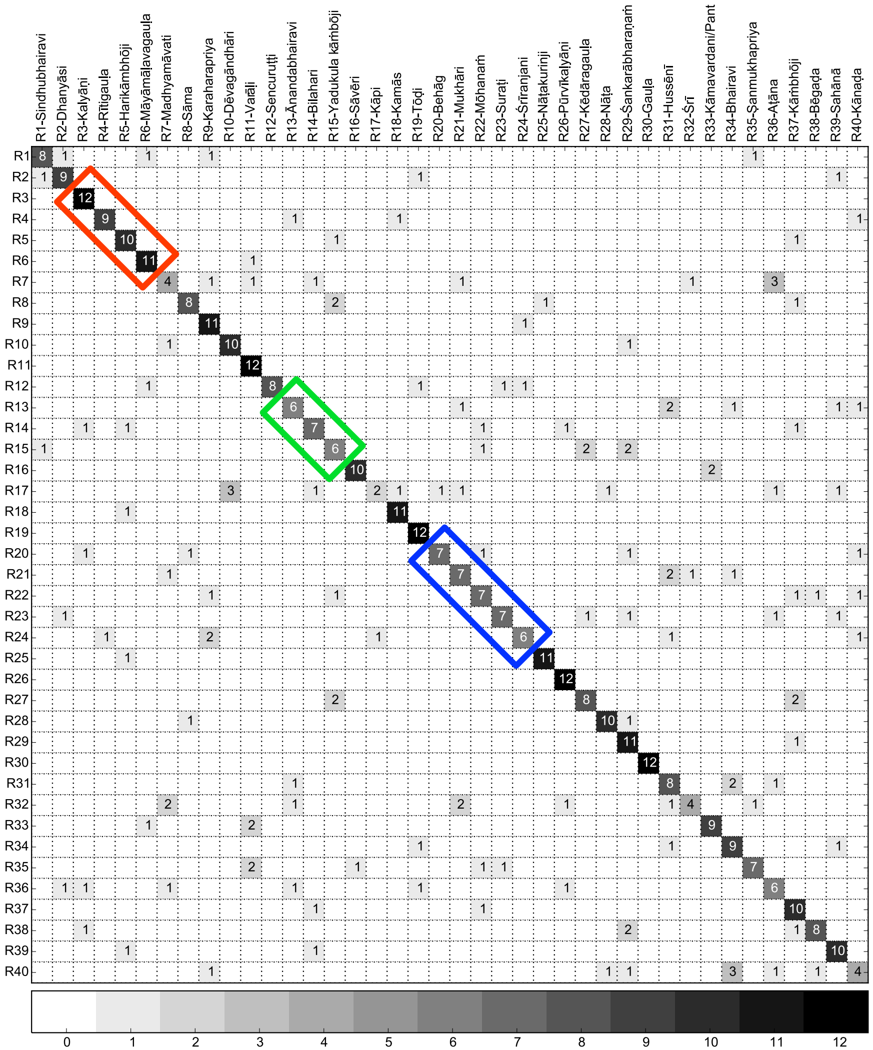


Allied rāgas

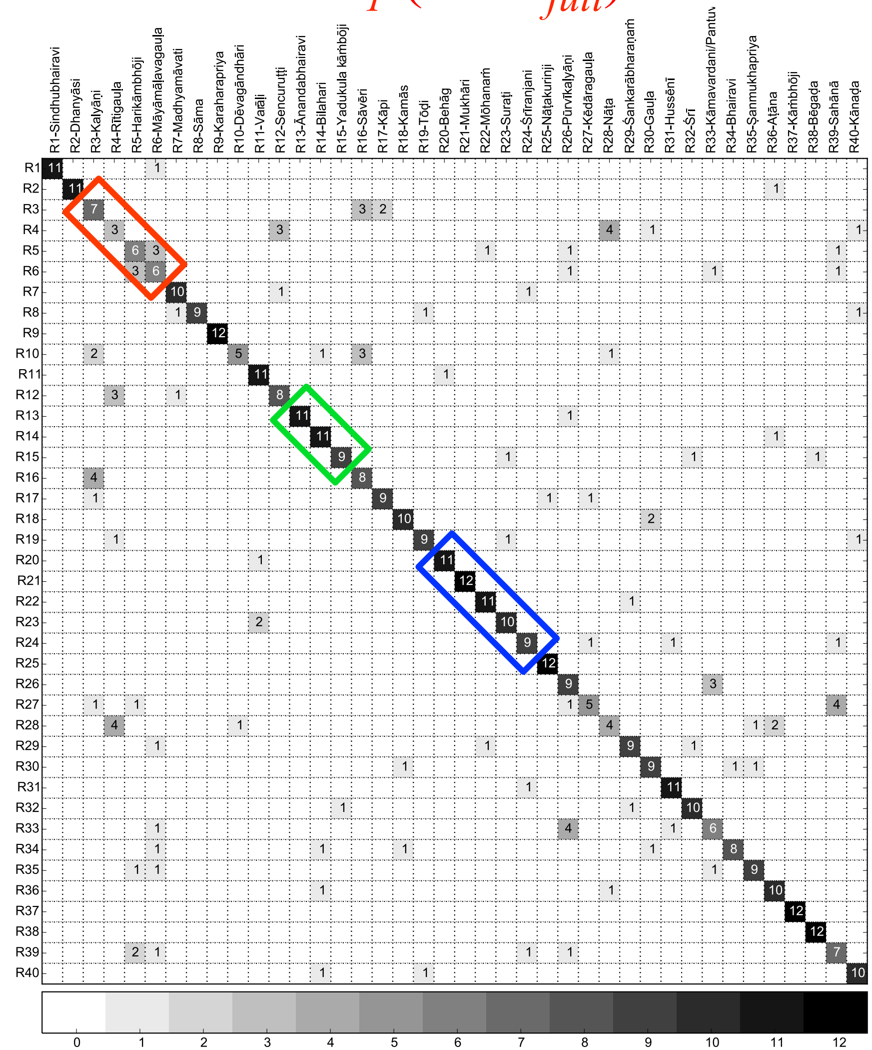
Music Technology Group

Error Analysis: complementary with S_1

M (F_1)



S_1 (PCD_{full})



Summary

Summary

- Phrase-based rāga recognition using VSM is a successful strategy

Summary

- Phrase-based rāga recognition using VSM is a successful strategy
- Mere presence/absence of melodic phrases is enough to recognize rāga

Summary

- Phrase-based rāga recognition using VSM is a successful strategy
- Mere presence/absence of melodic phrases enough to recognize rāga
- Multinomial Naive Bayes classifier outperforms the rest

Summary

- Phrase-based rāga recognition using VSM is a successful strategy
- Mere presence/absence of melodic phrases enough to recognize rāga
- Multinomial Naive Bayes classifier outperforms the rest
- Topological properties of network - similarity threshold

Summary

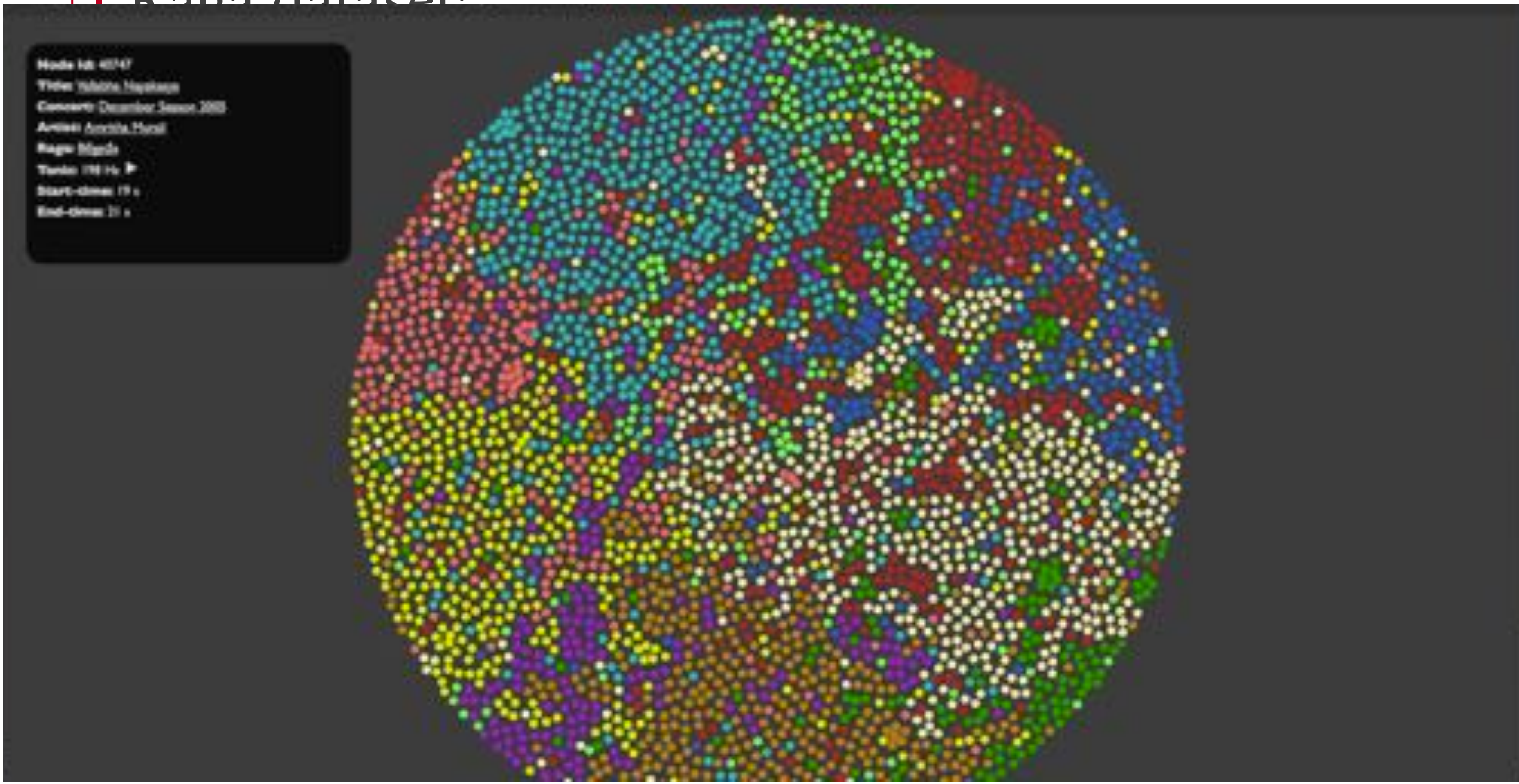
- Phrase-based rāga recognition using VSM is a successful strategy
- Mere presence/absence of melodic phrases enough to recognize rāga
- Multinomial Naive Bayes classifier outperforms the rest
- Topological properties of network - similarity threshold
- Complementary errors in prediction compared to PCD-based methods (Future Work)

Resources

- Rāga dataset:
  - http://compmusic.upf.edu/node/278
- Demo:
  - http://dunya.compmusic.upf.edu/pattern_network/
- CompMusic:
  - http://compmusic.upf.edu/
- Related datasets:
  - http://compmusic.upf.edu/datasets

Resources

- Rāga dataset:



  - http://compmusic.upf.edu/datasets

Resources

- Rāga dataset:
  - http://compmusic.upf.edu/node/278
- Demo:
  - http://dunya.compmusic.upf.edu/pattern_network/
- CompMusic (Project):
  - http://compmusic.upf.edu/
- Related datasets:
  - http://compmusic.upf.edu/datasets

Music
Technology
Group

Phrase-based Rāga Recognition using Vector Space Modeling

Sankalp Gulati*, Joan Serrà^, Vignesh Ishwar*, Sertan Şentürk* and Xavier Serra*

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

^Telefonica Research, Barcelona, Spain

The 41st IEEE International Conference on Acoustics, Speech and Signal Processing

Shanghai, China, 2016