UNCOVERING THE MICROTONES IN A RAAG FROM NOTE TRANSCRIPTIONS

Neeraja Abhyankar
Unaffiliated
neeraja.abhyankar@gmail.com

ABSTRACT

The presentation of Indian Classical Music admits the use of microtones ("shruti"s or frequency ratios) that are characteristic of each Raag which can be described as a set of notes (a scale) plus a "chalan" (melody patterns and rules regarding permissible/emphasized combinations of notes) meant to create a certain mood [1, 2]. Often, the Raag's definition as appearing in pedagogy does not include microtonal information. Hindustani Classical Music is transcribed and spoken of with a 12-note system. Despite this, most advanced vocalists or musicians with a continuous-frequency-domain instrument will tend to use appropriate microtones during rendition. This suggests that the microtonal information may be consistently derived based on the given description of the Raag. In this study, we attempt to use empirical analysis and MIR techniques to discover rules, if any, that may guide this derivation algorithmically.

1. INTRODUCTION

It is well-known [3, 4] that the vocal practice (or that on a continuously-tunable instrument) of Hindustani Classical Music uses Just Intonation, i.e., "shrutis" or microtones that are simple rational multiples of the chosen root node or tonic frequency. The selection of these tones is dependent on the Raag being presented, and hence, is influenced by other characteristics of the Raag, namely, the weightage being placed on each note, the motifs that frequently appear in the Raag, the glides/ornamentation that is commonly applied to a note, the rules that encourage/prohibit certain combinations of consecutive notes, etc. As an example (see Table 1), Raag Bhoopali (a night-time Raag that centers around the Ga uses different tone frequencies than the morning Raag Deshkar that centers around the Dha) even though both appear to use the same *notes* [1].

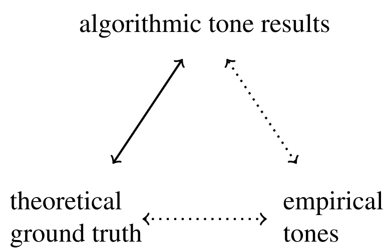
[1] Notation: for the rest of this text, I will be using the word *note* to refer to an approximate position in the scale – one of a set of 12: referred to as Sa, re, Re, ga, Ga, ma, Ma, Pa, dha, Dha, ni, Ni in Hindustani Music, whereas the word *tone* to refer to a precise ratio w.r.t. the tonic.

	Sa	**Re**	**Ga**	**Ma**	**Pa**	**Dha**	**Ni**
Bhoopali	1	10/9	5/4	–	3/2	5/3	–
Deshkar	1	9/8	81/64	–	3/2	27/16	–
Yaman	1	9/8	5/4	45/32	3/2	27/16	15/8

Table 1. Selection of *tones* in Raags that have *notes* in common [5].

While this microtonal information is prescribed in certain music theory texts for some Raags, there is disagreement among experts for some others. More importantly, insights behind *why* certain tones are appropriate are rarely encountered. In this work, we would like to understand the process by which this selection of tones takes place. Given an instance of a rendition done in a Raag, if we encode the melody using a 12-note octave system, can we arrive at the correct choices of tones used as a maximization of some objective function? If so, we may be able to gain some insight into axioms underlying the emergence of Raags in Hindustani Classical Music, and be able to quantify the landscape of possibilities on which existing Raags are more favorable configurations. Other potential applications are error correction or compression: Given a piece performed by a skilled musician on an equal-tempered instrument such as the harmonium (analogous to expressing a tone with $log_2(12)$ bits per octave or a MIDI format), it may be possible to reconstruct from first principles, a piece with a rich microtonal structure.

In the following section, we will formulate a *learnt, algorithmic* way to arrive at the microtones used in a piece starting from a 12-note octave transcription. This will be compared with the ground-truth labels seen in Hindustani Classical Music Theory texts [1,5]. In future work, we will use highly accurate pitch-detection methods to obtain empirical ground truth labels that match those in theory, and use these to finetune our algorithm further.



2. FORMULATION

To establish the most general subspace of rational ratios of the tonic from which tones may be selected, we assume a maximum allowable *prime-limit* p and a corresponding power limit α_{p_i} for each $2 < p_i \leq p$. Powers of two are ignored since they merely correspond to octave shifts. See Figure 1 as an example of a Tonnetz representation of $p = 5; \alpha_3 = 4; \alpha_5 = 2$. The tones are labeled with the closest approximate notes from the 12-note octave.

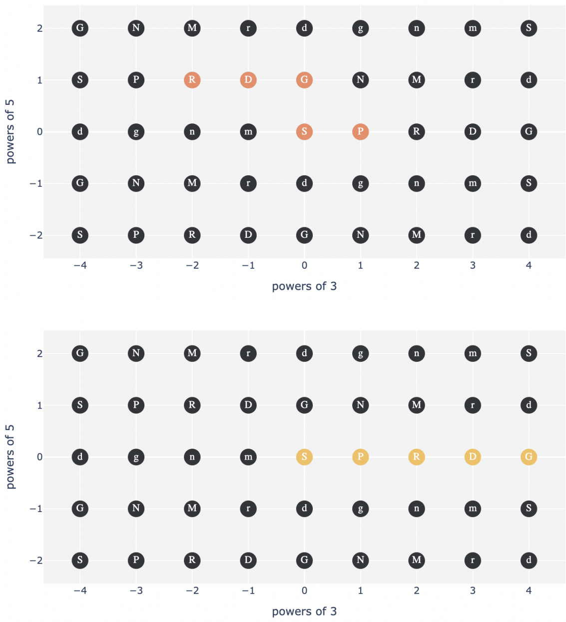


Figure 1. Tonnetz net (5-prime-limit) with tones of Raags Bhoopali (above) and Deshkar (below) highlighted in yellow. The tonic is the Sa ("S") in the center.

Given a piece of music represented as a **continuous time process** x over a discrete set of notes (see Eqn (1)), we wish to arrive at an algorithm or model that outputs a subset $\hat{\mathcal{R}}$ (standing for Raag) of the allowable Tonnetz net nodes $\mathcal{T}$.

$$
\begin{aligned}
& x(t) \in \mathcal{S}^K \qquad (1) \\
\text{where } \mathcal{S} = & \text{ the set of 12 notes} \\
\text{and } K = & \text{ the number of octaves} \\
& \text{ that the process is allowed to span}
\end{aligned}
$$

One possible approach was to learn a metric M_θ (see Eqn (2)) that can evaluate a selection of tones from a given Tonnetz net $\mathcal{T}$, and predict a set $\hat{\mathcal{R}}$ that maximizes this metric.

$$
\begin{aligned}
\hat{\mathcal{R}} = & \underset{\text{permissible } \mathcal{R} \subset \mathcal{T}}{\arg\max} \; M_\theta(\mathcal{R}, x) \qquad (2) \\
\text{where } \mathcal{R} \subset T & \text{ is a chosen set of tones,} \\
& \text{one for each note occurring in } x
\end{aligned}
$$

However, given the exponential number of evaluations necessary, an attempt was made to instead featurize x as a size $|\mathcal{T}|$ multi-hot encoded vector (with ones at all the nodes

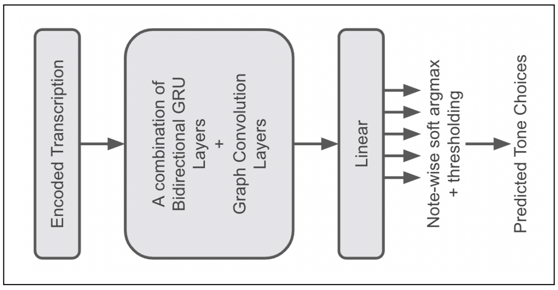


Figure 2. Temporal Graph Neural Network

corresponding to $x(t)$) and have a Temporal Graph Neural Network (TGNN) output a probability on each of the nodes being included in the final set $\mathcal{R}$. Taking a note-wise softargmax on each of the possible tones corresponding to a note will result in the predicted $\hat{\mathcal{R}}$ [2].

3. TGNN METHODOLOGY

The input to the TGNN is extracted from recordings of professional vocalists. A simple pitch-tracking algorithm was used (since we only require a 12-note transcription), and the found notes were normalized relative to a manually labeled tonic. Certain other works [6, 7] have found that octave-folding and duration-based discretization (as opposed to note count based) of the tracked pitch fairly capture the properties of an Indian Classical Music piece. Hence, and also to reduce the method's complexity, an octave folded sequence $x(t) \in \mathcal{S}$ is obtained. A skeleton of a TGNN architecture is proposed in Figure 2. Since this is an ongoing work, this architecture has *not* been proven to work and should undergo refinement. Also worth exploring are various other architectures for GNNs on the Tonnetz net for other tasks [8–10]. The crux of this method is to aggregate bidirectional sequential information as well as spatial information w.r.t. the Tonnetz lattice (using libraries such as [11]) to arrive at tone embeddings, which are further used to output likelihoods of tones being chosen in a Raag.

The code for these experiments is in-progress and can be found here [12].

4. ONGOING AND FUTURE WORK

As mentioned above, leveraging highly accurate pitch detection methods, either as a source for labels or to get a richer transcription as a starting point, will unlock more accuracy. Once a reliable model is trained, understanding its learnt weights and making connections with known as well as new music theory will be one of the main goals of this work. Future work also includes the possibility of allowing multiple microtones for the same note in a given piece. Along with this, the joint estimation of Raag and Microtone will also be possible using methods similar to [6].

[2] There are differing opinions among music theorists on whether each note is associated with a unique tone in a given Raag. For this study, we will assume this is the case. In future work, we shall allow for this possibility.

5. REFERENCES

[1] V. N. Bhatkhande, *Hindustani Sangeet Paddhati (Marathi/Hindi), Vol. 1–4*. Popular Prakashan, Mumbai, 1995.

[2] S. Rao and P. Rao, "An overview of hindustani music in the context of computational musicology," *Journal of new music research*, vol. 43, no. 1, pp. 24–33, 2014.

[3] J. Stephens, "The microtones of bharata's natyashastra," *Analytical Approaches to World Music*, vol. 6, no. 1, pp. 1–47, 2017.

[4] D. S. Thakur, "The notion of twenty-two shrutis: Frequency ratios in hindustani classical music," *Resonance*, vol. 20, pp. 515–531, 2015.

[5] V. Oke, *22 Shrutis and Melodium*. Sanskar Prakashan, Mumbai, 2007.

[6] 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.

[7] M. Clayton, P. Rao, N. N. Shikarpur, S. Roychowdhury, and J. Li, "Raga classification from vocal performances using multimodal analysis." in *ISMIR*, 2022, pp. 283–290.

[8] M. Aminian, E. Kehoe, X. Ma, A. Peterson, and M. Kirby, "Exploring musical structure using tonnetz lattice geometry and lstms," in *Computational Science–ICCS 2020: 20th International Conference, Amsterdam, The Netherlands, June 3–5, 2020, Proceedings, Part II 20*. Springer, 2020, pp. 414–424.

[9] E. J. Humphrey, T. Cho, and J. P. Bello, "Learning a robust tonnetz-space transform for automatic chord recognition," in *2012 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2012, pp. 453–456.

[10] G. Magai, A. Polevoi, I. Shanin, and K. Sorokin, "Chordal embeddings based on topology of the tonal space," in *Artificial Intelligence in Music, Sound, Art and Design: 12th International Conference, EvoMUSART 2023, Held as Part of EvoStar 2023, Brno, Czech Republic, April 12–14, 2023, Proceedings*, vol. 13988. Springer Nature, 2023, p. 20.

[11] B. Rozemberczki, P. Scherer, Y. He, G. Panagopoulos, A. Riedel, M. Astefanoaei, O. Kiss, F. Beres, G. Lopez, N. Collignon, and R. Sarkar, "PyTorch Geometric Temporal: Spatiotemporal Signal Processing with Neural Machine Learning Models," in *Proceedings of the 30th ACM International Conference on Information and Knowledge Management*, 2021, p. 4564–4573.

[12] N. Abhyankar, "Ismir 2024 demo code," https://github.com/neerajaabhyankar/libmogra-sandbox/tree/main/ISMIR2024, 2024.