Comparative study of classification of Indian Classical Music Ragas based on features from wavelet transforms

Problem Description:

Audio content analysis has been at the forefront of music information retrieval. Devising efficient ways to catalog and analyze audio has been of great interest for commercial reasons and the need for complex digital signal processing required in analyzing the music has piqued the interest of the scientific community in order to simplify the process. Efficient music categorization is an ongoing endeavor. The music being analyzed varies vastly and with that, the algorithms required to process them efficiently. As one moves outside the realm of western music, into music of other cultures, a lot of the techniques need modifications as per the grammar and conventions of that music.

One such genre of “non-conventional” music is the Indian Classical Music; specifically the Indian Ragas. Indian ragas are a complex set of melodies and art pieces that have a defined set of rules, but a huge margin of improvisation during the performance. So long as the set of rules defining the raga are followed, it is not classified as a different raga, despite an incredible amount of variations in the performance. To add to the complexity, various ragas have similar tonal structures and are often very difficult to differentiate. It is reasonable to say that the Indian ragas cannot be analyzed based on the algorithms and features developed for western music.

This leads researchers to develop algorithms that can mimic the trained ear of an Indian classical musician. Hence, a very good point to begin the analysis and identification of ragas is to figure out the Mel Frequency Cepstral Coefficients and develop on them. These coefficients fall short when we take into account the transitions between notes within the raga. So, a need for a better and more robust feature set arises, perhaps beginning from the emulation of the human hearing system, i.e. the cochlea. Once the emulation of cochlea and logarithmic representation of frequency is done, we move on to develop some more complex audio features which can give us the contour of the raga and help develop that into a feature vector.

The problem I am going to be tackling in this project is to determine a robust set of audio features that can take into account the spatial complexity of Indian ragas and their note transitions. I will proceed to then classify and recognise them. I will also be comparing the features derived from wavelet transform and try to contrast them with STFT and simple MEL-frequency cepstral coefficients.

Current methodologies:

Papers [14][18] discuss the applications of DWT in the extraction and classification of musical features and contrast them with Mel coefficient-based classification. A similar paper [3] discusses how to use wavelet transform to derive coefficients that are close to log based frequency distribution and use them to identify a speaker. This is an interesting application of DWT since it aims to determine a speaker and use that methodology for biometric identification, similar to how we identify individuals over a phone call. There has also been some research on this subject by Richard Kronland-Martinet [19][20], who has used DWT to analyze audio in great detail. He has used this process in the analysis of diverse signals, however the ones that stand out the most are in speech recognition and music analysis. Papers [10][14] also talk about how to use audio wavelet transform and processing to figure out the phonetics and vowels in speech. The interesting thing about all the papers that have been cited here, and more, have been invested in analyzing audio based on log scale representation of frequencies which is how humans perceive pitch.

If we move to the domain of audio processing and commercial applications, people are still using the STFT based approach. As is clear in Prof. Alexander Lerch's book "An Introduction to Audio Content Analysis", the wavelet transforms based filter banks have not been successfully incorporated in music analysis because 1. a filterbank with reasonable frequency resolution is computationally intensive, and 2. the lack of perfect reconstruction ability that can convert the frequency analyzed audio back to time domain. Which is why, audio applications still use STFT based analysis methods. Numerous papers by Dr. Lerch and others [4][5][8] and the field of Music information retrieval is based on STFT analysis.

Moving on to Indian Classical Music analysis, the conventional characterization and identification methods are using Dynamic Time Warping, distance measure and statistical modelling [7][21]. The analysis is still heavily reliant on deep learning models and the accuracy still lacks compared to popular music. People have also tried to use n-grams for analyzing the ragas [21].

Advantage of Filter Bank based Approach:

Since Indian Classical music, specifically, the Indian ragas, are extensively dependent on the contextual proceedings of the notes and transitions used, it is imperative to realize their classification using a wavelet transform rather than an STFT based approach. This approach has not been used with Indian Classical music before and it would be interesting to see the result of this assessment. This doesn't conflict with Lerch's idea of the high complexity of assessment since we can afford to do so for a small portion of audio signals.

Proposed approach:

I propose to use discrete wavelet transforms to derive log based cepstral coefficients for frequency representation which can mimic the cochlea and how we perceive pitch. This will involve using various wavelet functions that are in use and mentioned in current methodologies. I would also like to explore the derivatives of these coefficients that can give the change in frequency with time; a desired feature for exploring the pitch contour for characteristic ragas in Indian classical music. Since a huge dataset for Indian Ragas is not readily available, I will also use music from other music datasets.

Once the features using discrete wavelet transforms are extracted, I will contrast them with STFT and MEL coefficient based features. I will try to compare the performance in terms of wavelet choice (orthogonal vs. biorthogonal/linear-phase), filter orders and number of levels. My hypothesis is that wavelet transform will outperform the other feature sets in case of identification and classification of Indian classical music. Since the merit of the classifier is not the focus, I will use a simple classifier model to test the features.

References:

1. V. Ricordel, A. Drémeau, C. Guillemot, and M. Cagnazzo, “State of the art in 2D content representation and compression,” p. 33.
2. T. Lambrou, P. Kudumakis, R. Speller, M. Sandler, and A. Linney, “Classification of audio signals using statistical features on time and wavelet transform domains,” in *Proceedings of the 1998 IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP ’98 (Cat. No.98CH36181)*, Seattle, WA, USA, 1998, vol. 6, pp. 3621–3624. doi: 10.1109/ICASSP.1998.679665.
3. M. I. Abdalla and H. S. Ali, “Wavelet-Based Mel-Frequency Cepstral Coefficients for Speaker Identification using Hidden Markov Models,” vol. 1, no. 2, p. 6, 2010.
4. L.-C. Yang and A. Lerch, “On the evaluation of generative models in music,” *Neural Comput & Applic*, vol. 32, no. 9, pp. 4773–4784, May 2020, doi: 10.1007/s00521-018-3849-7.
5. P. Seshadri and A. Lerch, “Improving Music Performance Assessment with Contrastive Learning,” *arXiv:2108.01711 [cs, eess]*, Aug. 2021, Accessed: Nov. 03, 2021. [Online]. Available: http://arxiv.org/abs/2108.01711.
6. R. F. Lyon, A. G. Katsiamis, and E. M. Drakakis, “History and future of auditory filter models,” in *Proceedings of 2010 IEEE International Symposium on Circuits and Systems*, Paris, France, May 2010, pp. 3809–3812. doi: 10.1109/ISCAS.2010.5537724.
7. A. K. Sharma *et al.*, “Classification of Indian Classical Music With Time-Series Matching Deep Learning Approach,” *IEEE Access*, vol. 9, pp. 102041–102052, 2021, doi: 10.1109/ACCESS.2021.3093911.
8. S. Gururani and A. Lerch, “AUTOMATIC SAMPLE DETECTION IN POLYPHONIC MUSIC,” p. 8.

9. A. Biswas, E. Wennekes, T.-P. Hong, and A. Wieczorkowska, Eds., *Advances in Speech and Music Technology: Proceedings of FRSM 2020*, vol. 1320. Singapore: Springer Singapore, 2021. doi: 10.1007/978-981-33-6881-1.
10. G. F. Choueiter and J. R. Glass, "A Wavelet and Filter Bank Framework For Phonetic Classification," in *Proceedings. (ICASSP '05). IEEE International Conference on Acoustics, Speech, and Signal Processing, 2005.*, Philadelphia, Pennsylvania, USA, 2005, vol. 1, pp. 933–936. doi: 10.1109/ICASSP.2005.1415268.
11. Chen, Ying-Jui. "Wavelets and multirate filter banks : theory, structure, design, and applications." (2004).
12. M. S. Sinith and K. V. V. Murthy, "Automatic Music Composition based on HMM and identified wavelets in musical instruments," in *2011 International Conference on Signal Processing, Communication, Computing and Networking Technologies*, Thuckalay, Tamil Nadu, India, Jul. 2011, pp. 163–166. doi: 10.1109/ICSCCN.2011.6024535.
13. M. S. Sinith, M. N. Nair, N. P. Nair, and S. Parvathy, "Identification of wavelets and filter bank coefficients in musical instruments," in *2010 International Conference on Audio, Language and Image Processing*, Shanghai, China, Nov. 2010, pp. 727–731. doi: 10.1109/ICALIP.2010.5685071.
14. M. Gupta and A. Gilbert, "Robust speech recognition using wavelet coefficient features," in *IEEE Workshop on Automatic Speech Recognition and Understanding, 2001. ASRU '01.*, Madonna di Campiglio, Italy, 2001, pp. 445–448. doi: 10.1109/ASRU.2001.1034680.
15. F. Ahmed, "Wavelet transform-based correlator for the recognition of rotationally distorted images," *Opt. Eng*, vol. 34, no. 11, p. 3187, Nov. 1995, doi: 10.1117/12.213613.
16. Chapa, Joseph O., and Raghuveer M. Rao. "Algorithms for designing wavelets to match a specified signal." *IEEE transactions on signal processing* 48, no. 12 (2000): 3395-3406.
17. Strang, Gilbert, and Truong Nguyen. *Wavelets and filter banks*. SIAM, 1996.
18. G. Tzanetakis, G. Essl, and P. Cook, "Audio Analysis using the Discrete Wavelet Transform," p. 6.
19. Kronland-Martinet, Richard. "The wavelet transforms for analysis, synthesis, and processing of speech and music sounds." Computer Music Journal 12, no. 4 (1988): 11-20.
20. Kronland-Martinet, Richard, Jean Morlet, and Alexander Grossmann. "Analysis of sound patterns through wavelet transforms." *International journal of pattern recognition and artificial intelligence* 1, no. 02 (1987): 273-302.
21. Dandawate, Yogesh H., Prabha Kumari, and Anagha Bidkar. "Indian instrumental music: Raga analysis and classification." In *2015 1st international conference on next generation computing technologies (NGCT)*, pp. 725-729. IEEE, 2015.