Research Article

Role of Temporal Fine Structure and Envelope in Identifying Ragas: A Prospective Study for Temporal Based Raga Classifier

Sridhar Sampath[1], Udhayakumar Ravirose[2], Devi Neelamegarajan[3], Kamalakannan Karupaiah[4*], Rashmi Eraiah[3]

[1] *Department of Audiology, School of Rehabilitation and Behavioral Sciences, Vinayaka Mission's Research Foundation, Arupadai Veedu Medical College Campus, Pondicherry, India*

[2] *Department of Audiology Speech and Language Pathology, SRM Medical College Hospital and Research Centre, SRMIST, Chennai, India*

[3] *Department of Audiology, All India Institute of Speech and Hearing, Mysuru, India*

[4] *Department of Audiology and Speech-Language Pathology, Holy Cross College, (Autonomous), Tiruchirappalli, India*

Use your device to scan and read the article online

Citation: Sampath S, Ravirose U, Neelamegarajan D, Karupaiah K, Eraiah R. Role of Temporal Fine Structure and Envelope in Identifying Ragas: A Prospective Study for Temporal Based Raga Classifier. Aud Vestib Res. 2024;33(4):294-302.

https://doi.org/10.18502/avr.v33i4.16647

Highlights

- Identifying ragas using temporal fine structure and envelope cues by musicians
- Overall, both the temporal cues are essential in a trade-off manner
- This knowledge can be prospectively applied in a temporal-based raga classifier

Article info:

Received: 23 Oct 2023

Revised: 13 Dec 2023

Accepted: 17 Jan 2024

*** Corresponding Author:**

Department of Audiology and Speech-Language Pathology, Holy Cross College, (Autonomous), Tiruchirappalli, India.

kamal.audiology@gmail.com

ABSTRACT

Background and Aim: A raga is characterized by its distinctive melodic shape. The ability to perceive intricate melodic and pitch patterns depends on the Temporal Envelope (ENV) and Fine Structure (TFS). The present study aimed to understand the importance of temporal envelope and temporal fine structure cues in identifying ragas in Indian music.

Methods: Twenty‐one adult's musicians were included in the study. In experiment 1, professional musicians were involved in a raga identification task using music chimaeras. In experiment 2, the chimaeras were subjected to acoustic analysis using the envelope difference index, to understand better how the ENV and TFS changed depending on how many frequency bands were used to create the chimaeras. The subjective impression of a new group of trained musicians was then compared to these results. Friedman's test and Wilcoxon tests were carried out.

Results: Results showed that both cues are crucial in a trade-off manner; when TFS are not significantly accessible, ENV aids in raga identification. It was reflected in experiment 1 as an increase in ENV scores and a decrease in TFS scores as the number of frequency bands increased. In experiment 2, the envelope difference index for ENV increases with a number of frequency bands, and it correlates with perceptual scores for ENV.

Conclusion: The current study highlights the perceptual role of temporal cues in raga identification and directs future work for a temporal-based raga classifier.

Keywords: Ragas; envelope; fine structure; Indian music; music information retrieval

Introduction

The human auditory system interprets complex acoustic stimuli like music in terms of spectral and temporal properties. Spectral properties are analyzed with tonotopic place information along the cochlea's basilar membrane, containing a finite number of bandpass filters. Temporal properties are transformed into magnitude and phase information with the help of the cochlea and distal end of the cochlear nerve. The time signal at a specific location in the basilar membrane is decomposed into the slowly varying temporal Envelope (ENV) and rapidly oscillating Temporal Fine Structures (TFS). TFS is represented in the cochlea by synchronizing nerve spikes to a specific carrier phase (phase locking) [1]. Processing TFS cues yield perceptual benefits in pitch and timbre appreciation [2].

Indian music is governed by ragas, a notion encompassing western music's scale and pitch. Ragas are distinguished by their notes, hierarchy, and distinct phrases. The distinguishing phrase describes the melodic shape of the note sequence [3]. Raga identification is key to appreciating, comparing, and learning Indian music [4]. Scale in music is a set of notes ordered by the f0 or pitch. A scale in the order of increasing pitch is an ascending scale, and a scale in the order of decreasing pitch is a descending scale. They are denoted as arohanam and avarohanam in Indian music, respectively. Computationally, ragas can be identified using note intonation, scale, note progression and characteristic phrases. The note progression is basically the envelope or the melodic shape. The scale and other components can be obtained from the fine structures of the melody [5].

Music Informational Retrieval (MIR) involves the application of knowledge in mathematics, statistics, signal processing, machine learning, human physiology, and musicology. The function of a MIR system is to identify and classify music. Several MIR approaches identify and classify ragas. However, most are based on time-frequency trajectories, which are spectral-based [6]. The raga classifiers used in MIR involve onset-offset detection and feature extraction [7]. The arohanam and avarohanam are the onset and offset cues of ragas. Approaches to onset detection include looking for the changes like increases in spectral energy, changes in spectral energy distribution or phase changes in detected pitch, etc. [8, 9]. Swaras or notes are important in feature extractions from ragas and are based on ratios of fundamental frequencies with the frequency of the initial note [9]. The MIRs based on temporal cues such as ENV and TFS are rare. These temporal cues might play a vital role in a trained individual's identification of ragas because their characteristic melodic contours usually identify them [10]. ENV and TFS cues are essential for perceiving complex pitch and melodic contours. Previous studies have highlighted the relative importance of both these cues in the perception of western melodies [11]. Unlike western melodies, which are based on composition and composed based on consistent rules of scale and meter, Indian ragas are improvisational in nature, making them dynamic. The improvisation can be reflected in the changes in the melodic shape of the ragas. Melodic shapes are physically represented as ENV and TFS. Hence, understanding the importance of these cues in raga recognition is vital. Also, to develop a computation algorithm for MIR based on temporal cues, it is essential to explore the importance of these cues in raga identification by a human model. Hence, the current study was carried out.

The Hilbert transform provides a mathematically rigorous definition of the envelope and fine structure free of arbitrary parameters [11]. This elegant mathematical method decomposes the auditory signals into a slow component (ENV) and a fast component (TFS) [12]. The ENV and TFS extracted using the Hilbert transform were chimerised (generating signals by using two sound waveforms as input to an algorithm here, the ENV from a particular signal and TFS from another signal serve as the input) to form a hybrid signal called chimaera with ENV and TFS of different sources. Varying the number of frequency bands during the Hilbert transform alters the number of ENV and TFS cues in the chimaeras. A music chimaera is a hybrid sound signal consisting of ENV and TFS from two musical pieces. Our study used musical pieces with two different ragas to construct a music chimaera. Therefore, chimaeras of music signals from different ragas constructed with a different number of frequency bands can be valuable in studying the role of ENV and TFS cues in identifying ragas using Indian music chimaeras.

This study aimed to understand the importance of temporal ENV and TFS cues in identifying ragas in

Indian music. The study's primary objective was to analyze the role of these temporal cues in identifying ragas by trained Indian musicians. This objective was accomplished by experiment 1. The secondary objective was to verify how changes in these cues affect the identification of ragas using objective acoustic analysis and correlate the acoustic analysis results with subjective perceptual performance. This objective was accomplished in experiment 2. The method and results of these experiments are described in the following sections.

Methods

Experiment 1

Twenty-one adult (mean age: 25.28, SD: 9.64) musicians trained in Indian classical music (either in Carnatic or Hindustani) for at least five years and practicing music for at least one hour a day for five days a week were included in the study. These were the participants who were familiar with and able to identify the list of 18 ragas (including Hamsadhwani, Mohana, Shankarabhrana, Kalyani, Mayamalava Gowla, Abhogi, Desh, Khapi, Bhairavi, Yaman, Anandabhairavi, Hindola, Durga, Bhupali, Poorvi, Bhimpalas, Karaharapriya and Jog) correctly and passed the cut-off criteria for questionaire on music perception ability [13] and the Mini Profile of Music Perception Skills (Mini-PROMS) [14] that assessed their music perception ability.

Arohanam (melody with ascending scale) and Avarohanam (melody with descending scale) of 18 different ragas in Indian classical music were sung by a trained female singer and recorded using a Computerized Speech Lab (CSL™). The pairs of ragas with the same number of notes (Swaras) and length were chosen to chimerize. ENV and TFS of these pairs were chimerised using Hilbert transform implemented in MATLAB software with varying numbers of frequency bands (including 2, 4, 8, 16, and 32) with frequencies up to the bandwidth of the original signal. The frequency bands span different frequency ranges in such a manner that the width of each frequency band is approximately linearly spaced according to the receptivity of the human cochlea. In the process of chimerasing the ragas, the original signals (sung ragas) were subjected to the Hilbert transform. The product of the Hilbert transform provides the ENV and TFS of the original signal [11]. The ENV of one raga and TFS of another raga in a pair were mixed to generate a music chimaera. Ten pairs of ragas were used to prepare five sets of chimaeras in a particular number of bands. Totally 25 sets of chimaeras were designed from 18 different ragas.

The chimaeras were randomized and played at a comfortable presentation level through headphones. The participants were asked to identify the raga on an open-ended task. For each chimaera, if the subject identifies the fine structure-based raga, a score of 1 was given for TFS. If the subject identifies the envelope-based raga, a score of 1 was given for ENV. So, for every chimaera, a score of 1 will be given for either ENV or TFS. The total score for ENV and TFS for each number of frequency bands was calculated. This total score for ENV or TFS reflects the perception of these cues at each number of frequency bands.

Experiment 2

The second experiment involves an acoustic analysis of the Indian music chimaeras. For this purpose, a new set of chimaeras were prepared since the set of chimaeras prepared in the previous experiment used different ragas at different number of frequency bands. Using different ragas at the different number of frequency bands adds another variable (inherent variation in pitch contours of ragas) to the experiment, other than the number of frequency bands, which is an objective of the experiment (to study the effect of the number of frequency bands). So, while comparing across the different number of frequency bands, using the same set of ragas to prepare chimaeras (which was not followed in experiment 1) helps to eliminate the effect of inherent variation in pitch contours of ragas and study the effect of the number of frequency bands. This experiment used ten ragas (like Abhogi, Hamshandwani, Bhimpalas, Poorvi, etc.) to prepare 12 chimaeras at each number of frequency bands. The same set of raga combinations were used in all levels of frequency bands. The chimaera preparation technique and the number of frequency bands were similar to that followed in experiment 1.

The Envelope Difference Index (EDI) is an acoustic measure that quantifies the temporal envelope contrast between the two sound signals [15]. This technique renders the precise difference between the two envelopes of a signal. The EDI of sources (originally recorded ragas) and the chimaeras were calculated. At each number of frequency bands, the source signals (originally recorded raga) that contributed to TFS and ENV were compared with the chimaeras prepared with that many number of frequency bands to calculate EDI. Then the EDI for TFS and ENV signals were tabulated and analyzed across the different frequency bands. This will help to understand how the changes in the amount of envelope information in chimaeras prepared with the different number of frequency bands correlate with the envelope of the original ragas used to prepare those chimaeras.

A group of twelve adults (mean age: 23.28, SD: 8.53) meeting the same criteria used in experiment 1 participated in an open set raga identification task similar to that of experiment 1, using this new set of chimaeras. The scores for TFS and ENV were calculated and correlated with the EDI.

Statistical analysis

Statistical Package for the Social Sciences (IBM SPSS) version 21.0 was used for data analysis. In experiment 1, Scores were calculated separately for ENV and TFS across different bands. The Shapiro-Wilks test for normality was administered to the raw data to assess the distribution. Since it did not follow normality ($p<0.05$), Friedman's test and Wilcoxon pairwise comparison (with a level of significance $p<0.05$) were carried out to evaluate the effect of the number of frequency bands on TFS and ENV scores.

In experiment 2, Shapiro Wilks test for normality was administered in the raw data to assess the distribution since it did not follow normality ($p<0.05$). Spearman correlation was carried out to compare the median EDI for TFS and ENV across the different frequency bands to correlate it with perceptual scores of TFS and ENV.

Results

Experiment 1

Descriptive statistics show that median TFS scores decrease as the frequency band number increases, and the negative scores in the higher number of frequency bands indicate that the subject identifies more ENV cues. (Figure 1) However, the ENV scores increase as the frequency band increases, and the negative scores in the lower number of frequency bands indicate that the subject identifies more TFS cues (Figure 2). Friedman's test for the effect of the number of frequency bands on TFS ($\chi2(4)=37.51$, $p<0.001$) and ENV ($\chi2(4)=27.24$, $p\leq0.001$) scores were significant. Wilcoxon pairwise comparison showed that TFS and ENV scores differed significantly in 2 bands from all other bands. The Z scores for this Wilcoxon pairwise comparison, and significant pairs are described in Tables 1 and 2. Wilcoxon Signed Rank test was carried out to compare the scores between TFS and ENV across the different frequency bands, and it revealed a significant difference only in 2 bands ($Z=–3.83$, $p\leq0.001$) and 32 bands ($Z=–2.64$, $p<0.001$) (Figure 3). The overall scores for TFS were also significantly higher than scores for ENV ($Z=—2.72$, $p\leq0.001$).

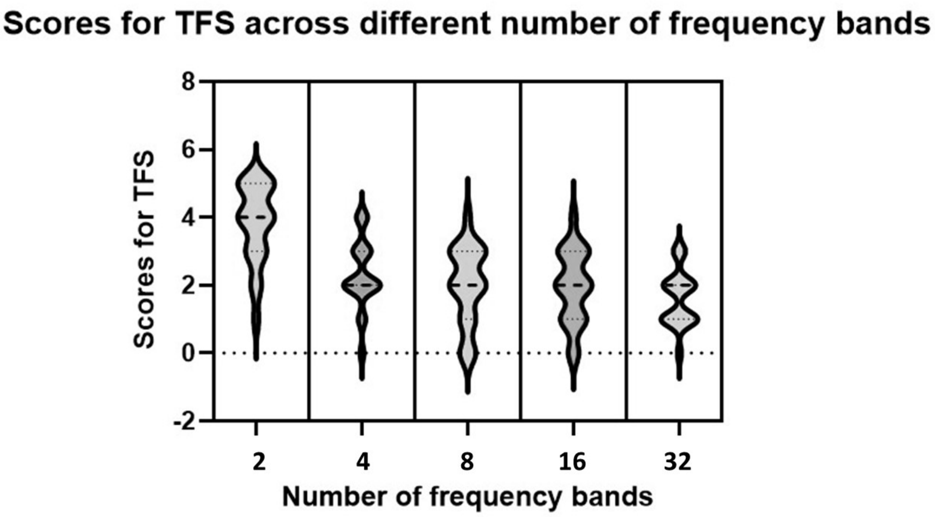

Figure 1. Violin plot showing temporal fine structures score decreasing as the number of frequency bands increases. TFS; temporal fine structures

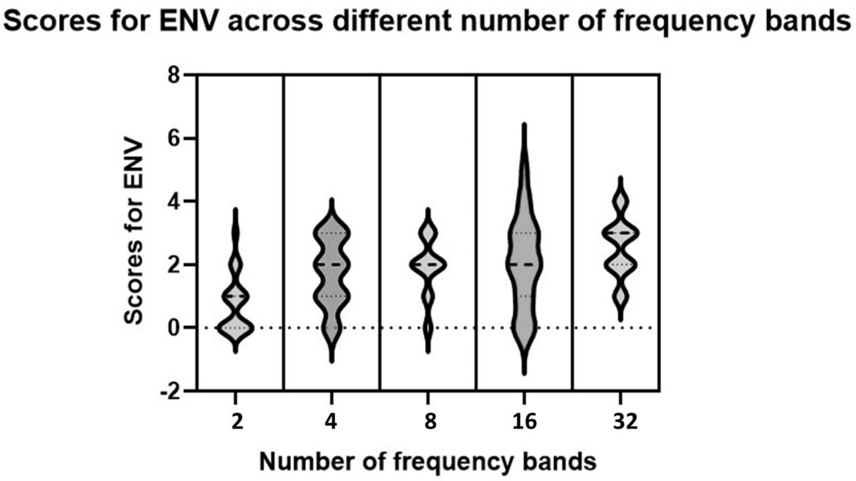

Figure 2. Violin plot showing temporal envelope score increasing as the number of frequency bands increases. ENV; envelope

Table 1. Wilcoxon pairwise comparison of temporal fine structures scores across the different number of frequency bands

No. of frequency bands	2	4	8	16	32
2		3.37**	3.60**	3.43**	3.85**
4			1.46	1.25	2.72*
8				0.16	1.55
16					0.99
32					

** Indicates p<0.01, * indicates p<0.05

Table 2. Wilcoxon pairwise comparison of temporal envelope scores across the different number of frequency bands

No. of frequency bands	2	4	8	16	32
2		2.91**	3.33**	2.99**	3.77**
4			0.96	1.03	2.28*
8				0.46	2.21*
16					1.06
32					

** Indicates p<0.01, * Indicates p<0.05

Experiment 2

Descriptive statistics show that the EDI calculated for TFS did not vary across the different number of frequency bands, while the EDI calculated for ENV increased as the number of frequency bands increased (Figure 4). The perceptual scores follow a similar trend to that of experiment 1.

Friedman’s test for the effect of the number of frequency bands in EDI for TFS ($\chi2(4)=2.19$, p=0.70) was negative, whereas EDI for ENV was positive ($\chi2(4)=44.45$, $p\leq0.001$). Wilcoxon pairwise comparison showed that EDI for ENV significantly increases at every level of frequency bands, that is, at 2, 4, 8, 16 and 32 (Table 3).

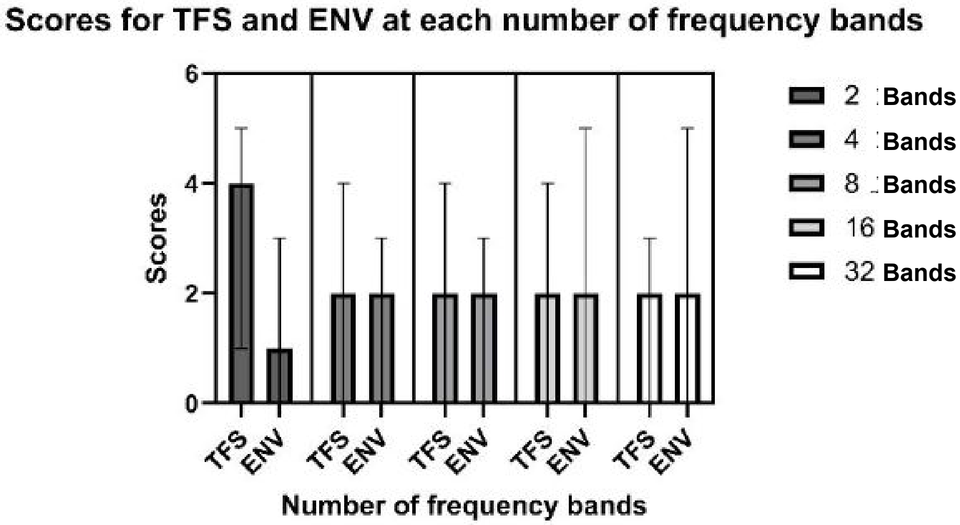

Figure 3. Bar graph showing temporal fine structures and temporal envelope scores in each number of frequency band. TFS; temporal fine structures, ENV; envelope

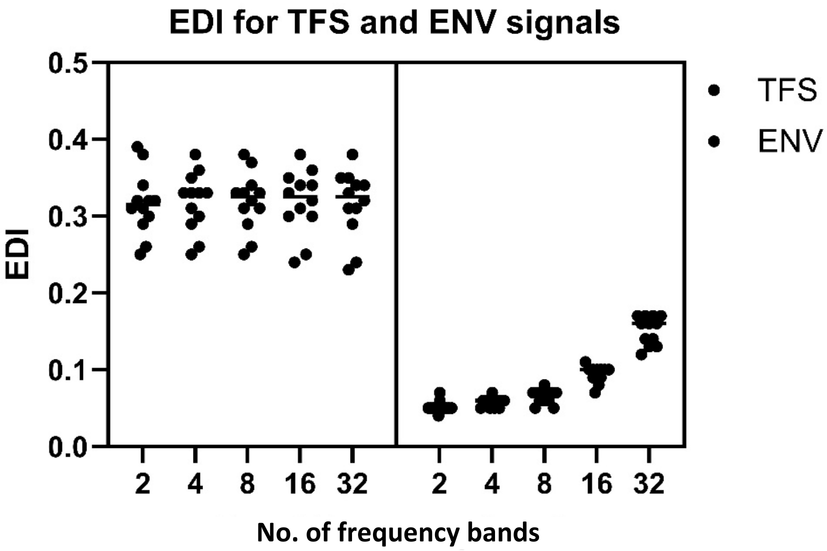

Figure 4. Scatter plot showing envelope difference index for temporal fine structures and temporal envelope signals in each number of the frequency band. EDI; Envelope Difference Index, TFS; temporal fine structures, ENV; envelope

Table 3. Wilcoxon pairwise comparison of envelope difference index for temporal envelope across the different number of frequency bands

	2	4	8	16	32
2		2.64**	2.72**	3.10**	3.07**
4			2.12*	3.08**	3.07**
8				3.10**	3.07**
16					3.09**
32					

** Indicates p<0.01, * Indicates p<0.05

Table 4. Spearman rho correlation coefficient of envelope difference index and perceptual temporal envelope scores across different frequency bands

Spearman's rho correlation coefficient		Envelope difference index				
		2	4	8	16	32
Perceptual scores	2	0.49				
	4		0.12			
	8			0.36		
	16				0.12	
	32					0.52

Spearman rho correlation showed that the EDI and the perceptual scores have a mild to moderate correlation at a marginal significance (Table 4).

Discussion

Increasing the number of frequency bands alters the temporal cues being reproduced in chimaeras. The lower the number of frequency bands, the better the TFS cues are represented, which explains the better scores for TFS at two bands compared to other number of bands. In contrast, the lower the number of frequency bands, the poorer the ENV cues, thus explaining poorer scores for ENV at two bands compared to the other number of frequency bands. At the lower number of frequency bands (like 2), TFS information was predominant over ENV and vice versa at a higher number of frequency bands (like 32). Hence, TFS were the cue at two bands, and ENV was the cue at 32 bands to identify ragas in Indian classical music. This finding is in line with the results of Smith et al. [11] that whenever TFS cues are available, they will be employed in music perception, and when they are not available, ENV cues will be employed. When both the cues are substantially available, importance could be shared among TFS and ENV [11]. In the context of raga identification, the importance of ENV and TFS can be understood psychophysically. Fine structure sensitivity is essential for the perception of pitch. Ragas encompasses notes arranged in the order of scale. The scale in perceived by the pitch of the notes and its relative distance in terms of pitch. Hence, TFS is crucial in learning a raga's notes and hierarchal scale. Similarly, ENV helps to learn the melodic shape of the ragas, which are characterised by the contours of the notes and their relative position in the frequency and time domain [16]. At lower frequency bands like 2, there are possible chances of natural recovery of the temporal envelope by the auditory filters of the cochlea [17]. In the current study, the scores of TFS at 2 bands are relatively higher than the scores of ENV at 32 bands. This can be interpreted as follows. In a condition where substantial envelope cues are not reproduced (such as in 2 bands), the available TFS cues would dominate the perception, which envelope recovery could have contributed. However, at 32 bands, where the required ENV cues are present, the ENV scores are still not higher than the TFS. Hence, envelope recovery cannot be the sole reason for such an outcome at lower bands. The current study did not attempt to address the effect of digital filter ringing, which occurs at higher frequency bands like 32. It could have affected the results at 32 bands, where the ENV and TFS did not very much. The overall scores for TFS were also significantly higher than scores for ENV, implying that fine structure sensitivity is more important for identifying ragas using Indian music chimaeras.

The outcome of experiment 2 supported these findings. Looking at Figure 4, it is evident that EDI was higher for the TFS signal than the ENV signal, implying that music chimaeras better represent the TFS signal. Also, there is no effect of the number of frequency bands on the EDI for the TFS signal, while the EDI for the ENV signal increases with the number of frequency bands. This could be because EDI is better for the ENV than the TFS.

The positive effect of the number of frequency bands on EDI for ENV signals implies that the ENV cue varies with the number of frequency bands. This finding supports the outcome of perceptual measures in experiment 1. Furthermore, the EDI correlates with the perceptual measure of experiment 2. Though the correlation is mild, it could indicate the possible reason for a trend in the perceptual scores of ENV. Also, EDI is widely used for speech signals, but its application in music signals is still unexplored. Also, a smaller sample size of musicians in experiment 2 could be attributed to the mild correlation.

Conclusion

The current study's findings strongly support the role of the temporal cues, temporal fine structures and envelope, in identifying ragas using Indian music chimaeras. Among the two temporal cues, though fine structures dominate their role more than the envelope, it should be noted that both cues are essential in a trade-off manner; when temporal fine structures are not substantially available, envelope also helps in the identification of ragas using Indian music chimaeras. These outcomes can be implicated in developing music informational retrieval algorithms for classifying Indian ragas based on temporal cues. Hilbert algorithm, used in this study, is being applied in biomedical engineering to construct signal processors in auditory prostheses like cochlear implants, which indicates its usefulness in analyzing temporal counterparts of complex signals, especially fine structure. Hence future studies are warranted to test the applicability of the Hilbert transform in music classifiers, which could be a temporal-based algorithm.

Limitations and future direction

The current study was carried out with a prospective goal of developing a computational tool for raga identification based on temporal characteristics such as envelope and fine structure. Only the psychophysical experiment was conducted to understand the importance these cues in raga identification was carried out currently. A prototype model implementing the outcome of this study was not built and tested. Hence future studies are warranted in these directions.

Ethical Considerations

Compliance with ethical guidelines

All of the testing procedures were accomplished using a non-invasive technique in the current study and adhered to the conditions of the institutional ethical approval committee. This study was approved by the institute Ethical Board (ref: DOR.9.1/915/2020-21 with effect from January 9, 2023). The test procedures were clearly explained to the participants before testing. Prior informed consent was taken from the participants for their willingness to participate in the study.

Funding

The authors declared that this study has received no financial support.

Authors' contributions

SS, UR: Study design, acquisition of data, drafting the manuscript, interpretation of the results and statistical analysis; DN: Study design, supervision, interpretation of the results, critical revision of the manuscript; KK, RE: Acquisition of data, drafting the manuscript.

Conflict of interest

The authors report no conflicts of interest.

Acknowledgments

The authors acknowledge the Director of All India Institute of Speech and Hearing for allowing to carry out this study. The authors acknowledge the participant for their cooperation.

References

1. Moon IJ, Hong SH. What is temporal fine structure and why is it important? Korean J Audiol. 2014;18(1):1-7. [DOI:10.7874/kja.2014.18.1.1]

2. Moore BCJ. The roles of temporal envelope and fine structure information in auditory perception. Acoust Sci Technol. 2019;40(2):61-83. [DOI:10.1250/ast.40.61]

3. Ganguli KK, Rao P. On the perception of raga motifs by trained musicians. J Acoust Soc Am. 2019;145(4):2418. [DOI:10.1121/1.5097588]

4. Kumar V, Pandya H, Jawahar CV. Identifying ragas in indian music. In2014 22nd International Conference on Pattern Recognition 2014 Aug 24 (pp. 767-772). IEEE. [DOI:10.1109/ICPR.2014.142]

5. Shetty S, Achary KK. Raga Mining of Indian Music by Extracting Arohana-Avarohana Pattern. Int J Recent Trends Eng. 2009;1(1):362-6.

6. Gajjar K, Patel M. Computational Musicology for Raga Analysis in Indian Classical Music: A Critical Review. Int J Comput Appl. 2017;172(9):42-7.

7. Kirthika P, Chattamvelli R. A review of raga-based music classification and music information retrieval (MIR). In2012 IEEE International Conference on Engineering Education: Innovative Practices and Future Trends (AICERA) 2012 Jul 19 (pp. 1-5). IEEE. [DOI:10.1109/AICERA.2012.6306752]

8. Sridhar R, Geetha TV. Swara indentification for south indian classical music. In9th International Conference on Information Technology (ICIT'06) 2006 Dec 18 (pp. 143-144). IEEE. [DOI:10.1109/ICIT.2006.83]

9. Sridhar R, Geetha TV. Raga Identification of Carnatic music for Music Information Retrieval. Int J Recent Trends Eng. 2009;1(1):571-4.

10. Velankar M, Deshpande A, Kulkarni P. Melodic pattern recognition in Indian classical music for raga identification. Int J Inf Tecnol. 2021;13:251-8. [DOI:10.1007/s41870-018-0245-6]

11. Smith ZM, Delgutte B, Oxenham AJ. Chimaeric sounds reveal dichotomies in auditory perception. Nature. 2002;416(6876):87-90. [DOI:10.1038/416087a]

12. Shukla M. Dichotomies in the perception of speech. J Biosci. 2002;27(3):189-90. [DOI:10.1007/BF02704907]

13. Neelamegarajan D, Uttappa AK, Arpitha V, Khyathi G. Development and Standardization of 'Questionnaire on Music Perception Ability'. Sangeet Galaxy. 2017;6(1):3-13.

14. Zentner M, Strauss H. Assessing musical ability quickly and objectively: development and validation of the Short-PROMS and the Mini-PROMS. Ann N Y Acad Sci. 2017;1400(1):33-45. [DOI:10.1111/nyas.13410]

15. Fortune TW, Woodruff BD, Preves DA. A new technique for quantifying temporal envelope contrasts. Ear Hear. 1994;15(1):93-9. [DOI:10.1097/00003446-199402000-00011]

16. Moore BC. The role of temporal fine structure processing in pitch perception, masking, and speech perception for normal-hearing and hearing-impaired people. J Assoc Res Otolaryngol. 2008;9(4):399-406. [DOI:10.1007/s10162-008-0143-x]

17. Ghitza O. On the upper cutoff frequency of the auditory critical-band envelope detectors in the context of speech perception. J Acoust Soc Am. 2001;110(3 Pt 1):1628-40. [DOI:10.1121/1.1396325]