J. Integr. Sci. Technol. 2021, 9(2), 92-97 Article

Journal of Integrated
SCIENCE & TECHNOLOGY

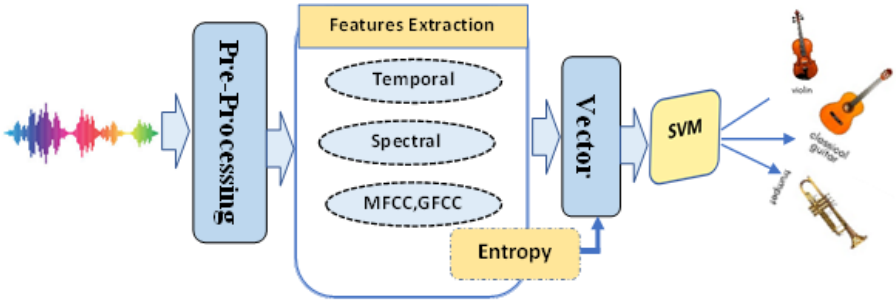
Musical instrument recognition using audio features with integrated entropy method

Seema R. Chaudhary,[1] Sangeeta N. Kakarwal,[2] R.R.Deshmukh[1]

[1]*Department of Computer Science and IT, Dr. Babasaheb Ambedkar Marathwada University, Aurangabad, Maharashtra, India*
[2]*Department of Computer Science Engineering, PES College of Engineering, Aurangabad, Maharashtra, India*

Received on: 15-Oct-2021, Accepted and Published on: 04-Dec-2021

ABSTRACT



Lots of Musical content are uploaded on social media daily. It is time-consuming to search content according to listeners' choice. Musical information retrieval is one of the evolving research fields which deals with retrieving content from audio data. Musical instrument recognition is subdomain of musical information retrieval. Previous research work has mostly focused on various western instruments belonging to distinct families, such as brass, string and woodwind are classified. The purpose of this study is to classify musical instruments using audio Features with Integrated Entropy method. Monophonic recordings of solo instrument artists are used in the experiments. Audio features have taken into account temporal, spectral, the first 13 Mel-frequency Cepstral Coefficients (MFCC) and Gammatone Frequency Cepstral Coefficients(GFCC). The proposed method generates a vector that integrates entropy with extracted features. Musical instruments are classified using generated vector. For classification, a Support Vector Machine (SVM) has been used.

Keywords: Musical Instrument Recognition, frequency range, Spectral features, Integrated Entropy, sound notes

INTRODUCTION

One of the vital components of human life is music. Music in various forms have different level of influence on human senses, like it soothes human mind and body when in soft form. Many musicians from all around the world use social media to showcase their music. Listeners may listen to a vast variety of music on the internet. Searching for a specific one based on criteria is difficult and time-consuming, depending on the performer, genre, piece of music, instrument, and so on. Some of the applications of musical information retrieval are recognizing artists, searching a song based on contents, sorting audio, classification based on music genre, music synthesis, and recognizing instruments from music.[1–5]

Each instrument has its own tone, which is determined by the material used, as well as the instrument's size and shape. Mainly four families of musical instruments are there i.e., percussion, string, brass and wind. String instruments are further divided into three categories based on how they are played: striking, plucked, and bowed. Much research work[6–8] has been done on the identification of instruments belonging to different families.[7,9–12] It is challenging to identify instruments among the instruments of the same family.

In this study, the timbre of musical instruments is analyzed by the proposed method of integration of entropy with extracted audio features. A solo dataset of Indian string instruments is created and evaluated using an Support vector machines (SVM) classifier. In addition to it, benchmark dataset TINYSOL[13] is also considered for evaluation.

This study is arranged as follows : first section contains related work in the field, next section comprises details of proposed methodology of audio features with integrated entropy method, the result and discussion is covers the results obtained from study along

Seema Chaudhary
Email: seema.chaudhary@mit.asia

Cite as: *J. Integr. Sci. Technol., 2021, 9(2), 92-97.*

 ISSN: 2321-4635 http://pubs.iscience.in/jist

with discussion on different parameters and inferences obtained from results followed by last conclusion section.

LITERATURE SURVEY

Many researchers have explored the related research work in the field of musical instrument recognition, the selected analysis of similar studies have been included here to understand the background work reported by researchers in this field.

X. Zhao and D. Wang,[14] had identified speakers using Gammatone Frequency Cepstral Coefficients (GFCC) and Mel-frequency cepstral coefficients (MFCC) features. With fully connected and recurrent neural network topologies, the efficiency of MFCC and GFCC representations is examined and evaluated over emotion and intensity categorization tasks by Gabrielle K. Liu.[15] The findings show that GFCCs outperform MFCCs when it comes to speech emotion identification.

M. Jeevan et.al.[16] has improved the performance of a text-independent speaker recognition system in a noisy environment using cross-channel utterance recordings. This study used a mix of Gammatone Frequency Cepstral Coefficients (GFCC) and i-vectors to handle noisy settings and accommodate session variations.

Mel Frequency Cepstral Coefficients (MFCC) are considered to be less fine and more strong to noise than Gammtone Frequency Cepstral Coefficients, which are less widely utilized. Over the given audio wave, adaptive whitening noise filtering is applied, and Zero Crossing Rate and Pitch were considered along with Gammatone Frequency Cepstral coefficients calculation.[17]

The performance of 13 features, including MFCCs, MPEG-7 features, and features based on perception, were compared by Peeters et. al.[18,19]. The MFCC feature scheme performed the best among the separate feature schemes in terms of categorization. Experiments in [20]showed that the MFCCs were preferred above other features. Duan et al. presented the mel-scale uniform discrete cepstrum as a characteristic to model the timbre of mixing music,[21] which was inspired by the MFCC.

In musical instrument recognition, MFCC is widely used whereas GFCC is less widely utilized.[22–25]

Most of the work using MFCC and GFCC was widely used in speaker identification, emotion detection and environment sound classification. Very few researchers have explored the use of MFCC and GFCC for the identification of musical instruments.

The work aims to propose method based on audio features in integration with entropy for the objective of musical instrument recognition, as well as to conduct an in-depth experimental investigation on their use.

METHODOLOGY

Dataset:

Two datasets have been used in this study. One is TinySOLand the second dataset is of Indian string instruments which have been generated from solo performance recordings. The TINYSOL dataset contains 1699 audio samples from the wind, brass, and string families. The wave files vary in length from 7 to 8 seconds and are monophonic.

Indian musical string instruments dataset comprises Sitar, Santoor, Sarod, Veena and Guitar. Recordings are monophonic and sampled at 44.1 KHz, MP3 files are converted to .wav format. There are 250 samples in total in the dataset. There are 50 audio samples for each musical string instrument.[26]

Each sample lasts for 5 sec. The dataset is preprocessed by erasing the silent portion of each sample at the start and end. Table 1 gives information about TINYSOL and Indian String Instruments(ISI) dataset.

Table 1: Audio Datasets

Dataset	Instruments	No. of total Samples
TINYSOL-Brass	Brass_Tuba,Horn,Trumpet	338
TINYSOL-String	Viola,Violin,ViolinCello	884
TINYSOL-Wind	Basson,Clarinet,Flute,Oboe	477
TINYSOL-ALL	Brass,String and Wind family	1699
ISI	Guitar,Santoor,Sitar,Sarod,Veena	250

In TINYSOL string family contains samples of bowed string instruments and in the ISI dataset , all are plucked string instruments.

Audio Features Extraction

Zero crossing rate (ZCR) time-domain feature is considered, it computes directly on the signal's samples without transforming the original audio signal. The ZCR is measured as how often the audio signal waveform hits the zero-amplitude level in 1 sec. interval.

For Spectral features usually obtained from the Short-Time Fourier Transform (STFT) transformation on the signal and signal spectrum is considered for the computation of features.[27,28]

The mean value of all elements in the signal is given by Mean, whereas entropy calculates the relative Shannon entropy of the signal. In information theory, Shannon entropy is used based on the following equation Eq.1.:

$$H(X) = -\sum_{i=1}^{n} p(x_i) log_b\, p(x_i)) \qquad (1)$$

where b - base of the logarithm. To acquire an entropy metric that is independent of sequence length, the relative entropy, which is calculated using Eq.2.as follows:

$$H(p) = -sum(p * \log(p))/\log(length(p)) \qquad (2)$$

The Shannon entropy provides a general description of the input curve p, as well as whether it has prominent peaks.[29]

Mel Frequency Cepstral Coefficients (MFCC):

The cepstral coefficients of the MFCC are obtained from a twisted frequency scale centered on human auditory perception. In the initial step in MFCC computation windowing is applied to the signal in order to divide it into frames. After windowing, each frame power spectrum is determined using the Fast Fourier Transform (FFT). Following that, processing of mel-scale filter bank is performed on the power spectrum. To calculate MFCC coefficients, the Discrete Cosine Transform(DCT) is applied to the

signal after taking log of power spectrum.[30] Figure 1 shows MFCC computation-

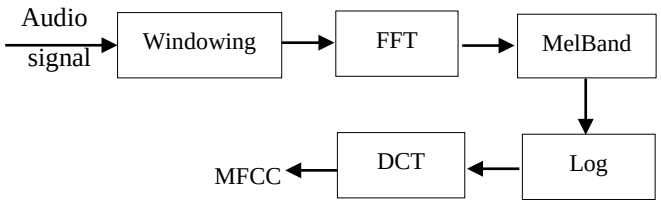


Figure 1: Computation of MFCC

Gammatone Frequency Cepstral Coefficient (GFCC) :

The Gammatone filter bank is a set of overlapping band-pass filters that simulates the human auditory system. The Fig 2 describes the process of GFCC :

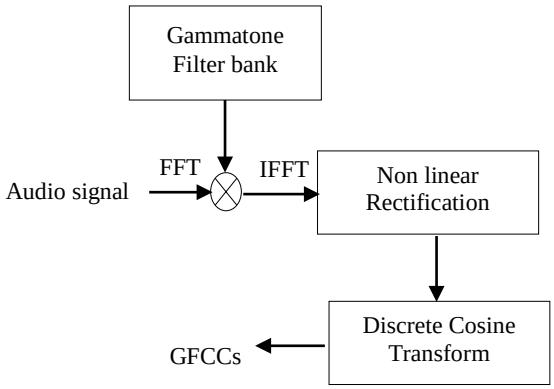


Figure 2: Computation of GFCC

In order to extract GFCC features, the speech signal is multiplied in the frequency domain by the Gammatone filter bank. As a result, the Inverse Fourier transform is used to convert this signal back to the time domain. A non-linear technique is used to take the signal's absolute value and rectify it.

In GFCC, instead of using the logarithmic operation as in MFCC, we employ the cubic root operation. Finally, to acquire GFCC features, DCT is used. [31] Table 2 gives the list of audio features.

Table 2: List of Audio Features

Type	Features	No. of Features
Temporal	Zero Crossing Rate(ZCR)	01
Spectral	Spectral Centroid, Spread and Roll-off	03
Cepstral Coefficients	Mel Frequency Cepstral Coefficient (MFCC)	13
	Gammatone Frequency Cepstral Coefficient (GFCC)	14

Methodology

The proposed system first removes the silent portion from the audio data files. The audio signal is processed to extract audio features. MFCC, GFCC, other audio features such as ZCR, Spectral, Mean and Entropy are extracted from audio samples. All extracted features are normalized.

The proposed methodology block diagram is presented in Figure 3.

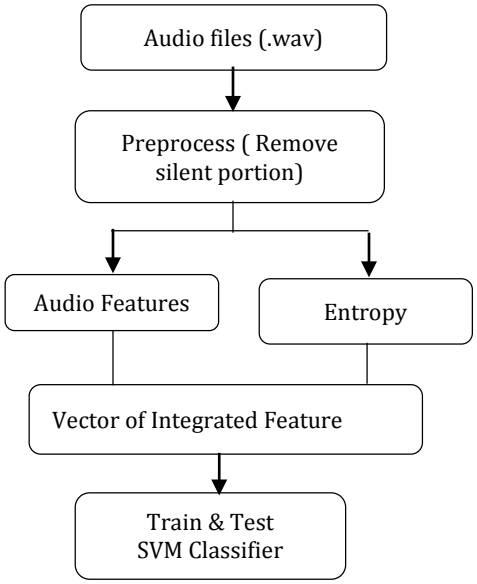


Figure 3: Block diagram of Proposed System using Audio Features with Integrated Entropy Method

Vector of audio features with integration of statistical feature Mean/Entropy is formed and given as an input to SVM[32] classifier. Five fold cross-validation is used for the evaluation of the model.

RESULTS AND DISCUSSION

Musical instruments have been classified by performing two experiments, in the first experiment, temporal and spectral features have been taken into consideration, whereas in the second experiment MFCC and GFCC features have been considered. Experiments are performed on TINYSOL and ISI datasets. The temporal, spectral, and statistical aspects of audio features are used to extract them.

The characteristics are normalized using the Max-Min method. When dealing with negative values, the absolute value of the lowest negative value is added to each value of that feature, and then Min-Max normalization[33] is used. Five fold cross-validation method is used for evaluation.

Experiment 1: Temporal, Spectral feature integrated with Mean/Entropy

In this experiment, the first vector of temporal feature i.e. ZCR is evaluated. Then it is integrated with mean/entropy features to classify instruments. Similarly spectral features such as spectral roll off, spectral spread and spectral centroid with integrated entropy vector evaluated using SVM. Table 3 shows percentage accuracy of SVM for given vector. With temporal vector integrated with

Table 3: Classification Accuracy(%) using Temporal and Spectral Features

Vector	TINYSOL-Brass	TINYSOL-String	TINYSOL-Wind	TINYSOL-All	ISI
ZCR	55.20	53.81	42.14	52.06	54.26
ZCR+Mean	56.68	54.94	44.74	53.75	73.87
ZCR+Entropy	63.74	61.29	57.06	82.40	75.87
Spectral Features	59.17	61.42	59.44	67.89	59.81
Spectral Features+Mean	59.27	63.27	61.96	69.66	66.85
Spectral Features+Entropy	64.48	66.94	59.40	84.05	67.85

Table 4: Classification Accuracy(%) using MFCC,GFCC

Vector	TINYSOL-Brass	TINYSOL-String	TINYSOL-Wind	TINYSOL-All	ISI
MFCC	98.01	94.20	86.13	84.83	97.98
MFCC+Mean	97.40	94.63	86.24	84.90	98.11
MFCC+Entropy	98.52	94.77	86.91	89.61	99
GFCC	93.69	97.43	84.47	87.92	98
GFCC+Mean	93.69	97.43	84.48	88.46	98
GFCC+Entropy	95.55	99.01	85.30	88.73	99

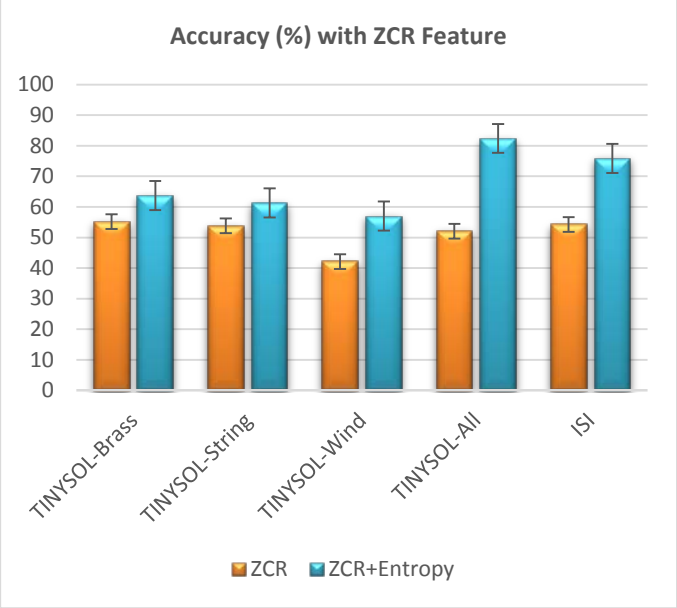
entropy within family of TINYSOL dataset of Brass- 63.74 % accuracy , String-61.29%,Wind-57.06 ,82.40 % accuracy obtained for familywise classification and 75.87% on ISI dataset. Spectral features also shown improvement in accuracy when is integrated with entropy.

Experiment 2: MFCC/GFCC integrated with Mean/Entropy

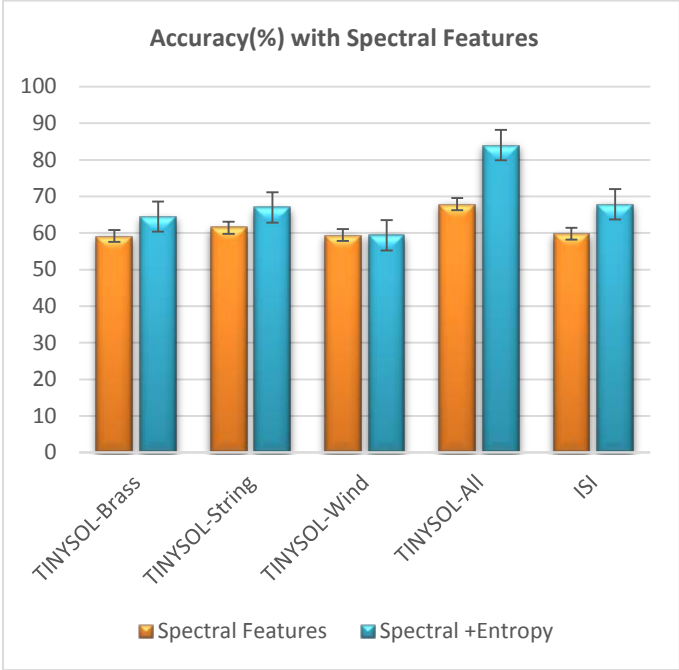
Mel frequency cepstral coefficient and Gammatone frequency cepstral coefficients are evaluated with integrated entropy. Table 4 shows Classification accuracy in percentage.

MFCC and GFCC features are based on perception of human auditory system. Within family of TINYSOL dataset highest accuracy i.e. 98.52% for brass,86.91% for wind and 89.61% for including all family has been achieved using MFCC with integrated entropy method, whereas for ISI dataset 99% accuracy obtained. For both string datasets GFCC with integrated entropy method has achieved accuracy 99%. Dataset wise results are shown in following Graphs.

Comparison of accuracy obtained by ZCR and proposed method ZCR with integrated entropy is shown in Graph 1. Due to the inclusion of entropy with ZCR substantial improvement is observed in accuracy. In TINYSOL-All dataset, which is a mix of all family i.e. brass,wind and string, proposed method has increased the accuracy by 30% .For Indian string instruments dataset -ISI, proposed method has significantly improved accuracy from 54.26% to 75.87%.

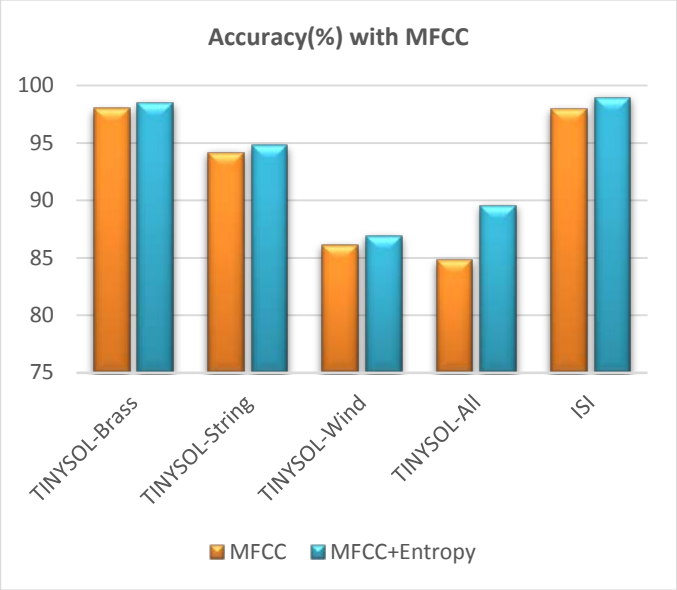


Graph 1: Classification Accuracy with ZCR Feature



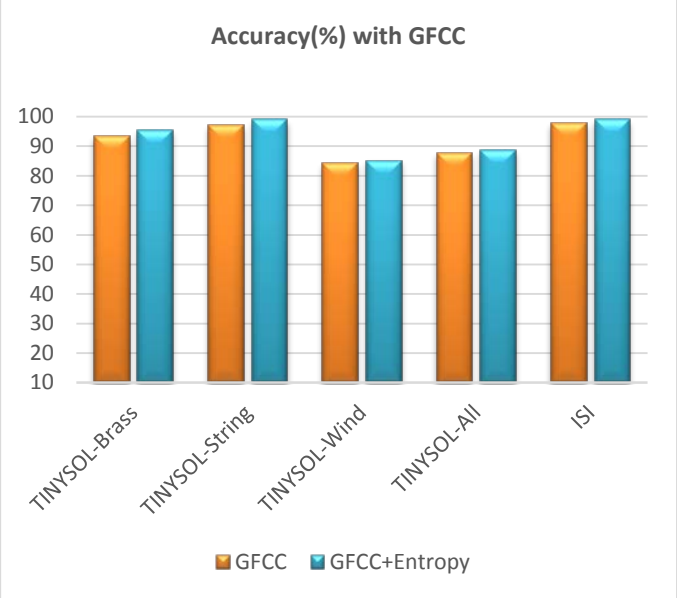
Graph 2: Classification Accuracy with Spectral Features

From Graph 2, it is observed that results have not improved in brass-wind instruments, as entropy contains less information. Average 5% improvement observed in TINYSOL-Brass and TINYSOL-String family, whereas with spectral features with integrated entropy given better result for TINYSOL-all and ISI datasets.



Graph 3: Classification Accuracy with MFCC Features

The accuracy obtained by MFCC and GFCC features are shown in Graph 3 and Graph 4.



Graph 4: Classification Accuracy with GFCC Features

Less increase in accuracy is achieved by MFCC with integrated entropy method. Average 1.5% increase is obtained by GFCC with integrated entropy.

According to the findings, audio features with integrated entropy improved the accuracy of instrument classification. In a mixed instrument data collection (TINYSOL-ALL), accuracy improves significantly as compared to individual instruments from the same family. It is also significantly improved in the ISI dataset.

CONCLUSION

In experiment 1, the average 50% accuracy by temporal ZCR and spectral features are obtained, whereas average 60% accuracy was obtained within the same family of instruments of TINYSOL with the proposed integrated entropy method. Accuracy has increased by 20% to 30% in familywise classification of TINYSOL and 10% to 20% increase in ISI dataset with ZCR and spectral features with integrated entropy method. MFCC and GFCC features with and without integration is evaluated in experiment 2. MFCC and GFCC features have classified musical instruments highly accurately than temporal and spectral features. It is seen that accuracy is slightly increased in MFCC and GFCC with the proposed integrated entropy method. Integrated entropy has given better results than the integrated mean. From experimentation, it has been observed that the proposed method of audio features with integrated entropy has improved the accuracy significantly in temporal and spectral features.

REFERENCES

1. H. Eghbal-Zadeh, M. Dorfer, G. Widmer. A Cosine-Distance Based Neural Network For Music Artist Recognition Using Raw I-Vector Features. *Proceedings of the 19thInternational Conference on Digital Audio Effects (DAFx-16), Brno, Czech Republic*. **2016**, 61-67.
2. A. Caclin, S. McAdams, B.K. Smith, S. Winsberg. Acoustic correlates of timbre space dimensions: A confirmatory study using synthetic tones. *J. Acoust. Soc. Am.* **2005**, 118 (1), 471–482.
3. P. A. Esquef, L.W. Biscainho,. Spectral-Based Analysis and Synthesis of Audio Signals. In H. Perez-Meana (Ed.), Advances in Audio and Speech Signal Processing: Technologies and Applications, **2007,** pp. 56-92. IGI Global.
4. D.F. Silva, F.V. Falcao, N. Andrade. Summarizing and comparing music data and its application on cover song identification. **2018**, 8, 732-739.
5. M. Eppe, T. Alpay, S. Wermter. Towards End-to-End Raw Audio Music Synthesis. In *Artificial Neural Networks and Machine Learning – ICANN 2018*; Kůrková, V., Manolopoulos, Y., Hammer, B., Iliadis, L., Maglogiannis, I., Eds.; Lecture Notes in Computer Science; Springer International Publishing, Cham, **2018**; Vol. 11141, pp 137–146.
6. Y. Guo, Q. Liu, A. Wang, et al. Optimized phase-space reconstruction for accurate musical-instrument signal classification. *Multimed. Tools Appl.* **2017**, 76 (20), 20719–20737.
7. A. Eronen, A. Klapuri. Musical instrument recognition using cepstral coefficients and temporal features. In *2000 IEEE International Conference on Acoustics, Speech, and Signal Processing. Proceedings (Cat. No.00CH37100)*; **2000**; Vol. 2, pp II753-II756.
8. J.S. Gomez, J. Abeßer, E. Cano. Jazz solo instrument classification with convolutional neural networks, source separation, and transfer learning. **2018**, 8, 577-584.
9. G. Agostini, M. Longari, E. Pollastri. Musical Instrument Timbres Classification with Spectral Features. *EURASIP J. Adv. Signal Process.* **2003**, 2003 (1), 1–10.
10. G. Mazarakis, P. Tzevelekos, G. Kouroupetroglou. Musical Instrument Recognition and Classification Using Time Encoded Signal Processing and Fast Artificial Neural Networks. In *Advances in Artificial Intelligence*; Antoniou, G., Potamias, G., Spyropoulos, C., Plexousakis, D., Eds.; Lecture Notes in Computer Science; Springer, Berlin, Heidelberg, **2006**; pp 246–255.

11. J.C. Brown, O. Houix, S. McAdams. Feature dependence in the automatic identification of musical woodwind instruments. *J. Acoust. Soc. Am.* **2001**, 109 (3), 1064–1072.
12. S. Ghisingh, V.K. Mittal. Classifying musical instruments using speech signal processing methods. In *2016 IEEE Annual India Conference (INDICON)*; IEEE, Bangalore, India, **2016**; pp 1–6.
13. C. Emanuele, D. Ghisi, V. Lostanlen, F. Lévy, J. Fineberg, Y. Maresz. (2020). TinySOL: an audio dataset of isolated musical notes [Data set]. Zenodo. https://doi.org/10.5281/zenodo.3632193.
14. X. Zhao, D. Wang. Analyzing noise robustness of MFCC and GFCC features in speaker identification. In *2013 IEEE International Conference on Acoustics, Speech and Signal Processing*; IEEE, Vancouver, BC, Canada, **2013**; pp 7204–7208.
15. G.K. Liu. Evaluating Gammatone Frequency Cepstral Coefficients with Neural Networks for Emotion Recognition from Speech. *ArXiv180609010 Cs Eess* **2018**.
16. M. Jeevan, A. Dhingra, M. Hanmandlu, B.K. Panigrahi. Robust Speaker Verification Using GFCC Based i-Vectors. In *Proceedings of the International Conference on Signal, Networks, Computing, and Systems*; Lobiyal, D. K., Mohapatra, D. P., Nagar, A., Sahoo, M. N., Eds.; Lecture Notes in Electrical Engineering; Springer India, New Delhi, **2017**; Vol. 395, pp 85–91.
17. K. Goli, V. Jain, J.V. Vidhya. Speaker Identification using GFCC with PITCH & ZCR. *Int. J. Adv. Sci. Technol.* **2020**, 29 (6), 26-33.
18. G. Peeters, S. McAdams, P. Herrera. Instrument Sound Description in the Context of MPEG-7. In *ICMC: International Computer Music Conference*; Berlin, Germany, **2000**; pp 166–169.
19. A. Eronen. Comparison of features for musical instrument recognition. In *Proceedings of the 2001 IEEE Workshop on the Applications of Signal Processing to Audio and Acoustics (Cat. No.01TH8575)*; **2001**; pp 19–22.
20. J.D. Deng, C. Simmermacher, S. Cranefield. A study on feature analysis for musical instrument classification. *IEEE Trans. Syst. Man Cybern. Part B Cybern. Publ. IEEE Syst. Man Cybern. Soc.* **2008**, 38 (2), 429–438.
21. Z. Duan, B.A. Pardo, L. Daudet. A novel cepstral representation for timbre modeling of sound sources in polyphonic mixtures: 2014 IEEE International Conference on Acoustics, Speech, and Signal Processing, ICASSP 2014. *2014 IEEE Int. Conf. Acoust. Speech Signal Process. ICASSP 2014* **2014**, 7495–7499.
22. D.G. Bhalke, C.B.R. Rao, D.S. Bormane. Automatic musical instrument classification using fractional fourier transform based- MFCC features and counter propagation neural network. *J. Intell. Inf. Syst.* **2016**, 46 (3), 425–446.
23. G.S. R., B.S. S., S.S. D. Cepstral (MFCC) Feature and Spectral (Timbral) Features Analysis for Musical Instrument Sounds. In *2018 IEEE Global Conference on Wireless Computing and Networking (GCWCN)*; **2018**; pp 109–113.
24. S. Chakraborty, R. Parekh. Improved Musical Instrument Classification Using Cepstral Coefficients and Neural Networks. In *Methodologies and Application Issues of Contemporary Computing Framework*; **2018**; pp 123–138.
25. T.S. Gunawan, M. Kartiwi. On the Comparison of Line Spectral Frequencies and Mel-Frequency Cepstral Coefficients Using Feedforward Neural Network for Language Identification. *Indones. J. Electr. Eng. Comput. Sci.* **2018**, 10 (1), 168.
26. S.R. Chaudhary, S.N. Kakarwal, J.V. Bagade. Feature selection and classification of indian musical string instruments using svm. *Indian J. Comput. Sci. Eng.* **2021**, 12 (4), 859–867.
27. S. Sheoran, S. Singh, A. Mann, A. Samantilleke, B. Mani, D. Singh. Novel synthesis and Optical investigation of trivalent Europium doped MGd2Si3O10 (M = Mg2+, Ca2+, Sr2+ and Ba2+) nanophosphors for full-color displays. *J. Mater. NanoSci.* **2019**, 6(2), 73–81.
28. F. Alías, J. Socoró, X. Sevillano. A Review of Physical and Perceptual Feature Extraction Techniques for Speech, Music and Environmental Sounds. *Appl. Sci.* **2016**, 6 (5), 143.
29. S. Ajibola Alim, N. Khair Alang Rashid. Some Commonly Used Speech Feature Extraction Algorithms. In *From Natural to Artificial Intelligence - Algorithms and Applications*; Lopez-Ruiz, R., Ed.; IntechOpen, **2018**.
30. Springer handbook of speech processing. Ed: J. Benesty, M. Mohan Sondhi, Y.A. Huang. **2008**. Springer.
31. J. Qi, D. Wang, J. Xu, J. Tejedor Noguerales. Bottleneck features based on gammatone frequency cepstral coefficients. *Conference: Interspeech'13*, **2013**.
32. M.A. Khan, S. Gairola, B. Jha, P. Praveen. Smart Computing: Proceedings of the 1st International Conference on Smart Machine Intelligence and Real-Time Computing (SmartCom 2020), 26-27 June 2020, Pauri, Garhwal, Uttarakhand, India; CRC Press, **2021**.
33. T. Jayalakshmi, A. Santhakumaran. Statistical Normalization and Back Propagation for Classification. *Int. J. Comput. Theory Eng.* **2011**, 89–93.