Contents lists available at Mapta Publishing Group

Vol. 1, No. 2, August, 2018

Mapta Journal of Electrical and Computer Engineering (MJECE)

journal homepage: http://maptapublishing.com/index.php/mjece

ISSN: 2631-5718

The Effect of Sound Characteristics in the Turkish Musical Instruments Classification

Sina Mojtahedi*

Biomedical Science and Engineering Department, Faculty of Engineering, KOÇ University, Istanbul, Turkey

Article Info

Article history:

Received May 18th, 2018
Revised Jul 19th, 2018
Accepted Aug 4th, 2018

Keyword:

Feature Extraction
Audio Signal Processing
Musical Instrument Recognition

Abstract

This study aims to investigate the effect of length features of Turkish musical instruments on the quality of their recognition and classification, and to explore the ways in which using exclusively the general features of voice recognition or using exclusively the short duration or long and medium length features can provide acceptable results in the field of Turkish musical instruments. For this purpose, after collecting musical datasets in four instrumental categories, including bow, plectrum, percussion, and woodwind, three experiments with similar conditions by using the neural network on this database are conducted to assess the effectiveness of the use of general features as well as the long, short, and medium lengths features. The result indicates that the use of the general properties and characteristics of sound, which have been used in several previous authoritative types of research, leads to the categorization with the accuracy of 59%, and the use of short-term features and long and medium length features lead to the categorization with the accuracy of 63% and 81% respectively. However, the selection of features from all four categories will lead to the categorization of 98% accuracy.

1. Introduction

There is a sustained interest in the musical instrument classification schemes of local communities. The attention toward these schemes which reflects culturally specific concepts and values is also increasing day by day. Based on the need of accounting for the embedded nature of musical instruments and their conceptualization, previous studies [1, 2], notes how much is lost while separating the physical object from local meanings, contexts, associations, and experiences". They argue the difference among instrument, not primarily because of what vibrates in or on the instrument (as the standard classification would lead us to believe), rather, each grows from and into human lives and worlds differently. Mrázek's well-crafted statement implies that the differences which define musical instruments are not stable, unchanging categories. Instead, the understanding of what distinguishes instruments from each other is the result of continous production of differences that is shaped through interactions, including the production of embodied meanings in musical performance and instrument construction within formal written texts and informal speech in different regions and societies. Recently, there is a huge attention to the direct/indirect impacts that musical instruments have on people, where, understanding this effect on human and non-human entities considered as a form of agency [3, 4]. These approaches have produced new insights through a focus on material culture as well as a recognition of the potential influence of non-human actors on social networks. Either, their efficacy is limited when dealing with phenomena that are unique to the human experience. This influence may include social norms and taxonomic concepts.

Each audio signal in terms of extracted frame length is categorized into short, medium, and long. If the extracted frames length is 30 milliseconds, the feature extracted will be a short one. Features such as a total number of Mel Frequency Cepstral Coefficients (MFCC) or zero-crossing detector are introduced as short frame features. If several consecutive short frames are connected by using the aggregation or implementation of other operators, medium frames will be created. Operators such as standard deviation (SD), mean (M) and derivative for aggregation are used in the

* Corresponding author: sna.mojtahedi@gmail.com

middle frame. The next classification of characteristics is a feature with a long frame length that is created by aggregation of features with medium frame lengths. Features with a medium frame length are one second in length and features with long frame lengths are frames of 10 seconds in length [5-8].

Table 1. The length features are separated from each other and described as:

The strength of Strongest Beat, Compactness, Strongest Beat, Beat sum, Spectral Rolloff Point, Spectral Flux, Spectral Centroid, The fraction of Low Energy frames, Spectrum, LPC, MFCC, Zero Crossing	Short Features
Standard Deviation of attributes Derivative of attributes Mean of attributes	Medium Features
Multiplication Hystegram	Long Features

2. Basic concepts

2.1. Neural networks

The neural network is a clustering method of data independence based on distance measurement. Neural Networks use biological concepts to recognize patterns. The output of this idea is the creation of Artificial Neural Networks (ANN) inspired by intuitive physical inference of human mind. The neural network is a combination of different units. In addition, neural network algorithms are used as statistical methods to improve performance and optimize results in neural networks [6, 7].

2.2. Extraction of features

The extraction of a feature is actually an operation that takes place in order to find the indexes, points, and determinants on a specific data. A large body of data concerning the issue of identification of the instrument has been carried out and reported that several features have been used to process audio and speech signals.

2.3. Evaluation criteria

Sensitivity: In the test performed, the sensitivity is the ability to recognize all the properties of each group of instruments. For instance, if the test can correctly identify all the sounds that belong to the woodwind group, the sensitivity of the test in the woodwind instrument group is 100 percent. The method of calculating it is equal to the sum of the division of the negative integers into the sum of the positive integers and the false countermeasures [9-11].

Exclusivity: In fact, this criterion is the probability that for an audacity of an instrument, the algorithm can correctly determine which vocal mechanism is not related to which group, whose method of calculating is by dividing the negative integers into the sum of positive integers and false positive countermeasures.

Accuracy: This criterion calculates the total accuracy of an algorithm. In fact, this criterion is the most popular criterion for calculating the efficiency of classifier algorithms, which shows that the algorithm correctly categorizes several percents of the entire set of experimental records.

Verification: This criterion calculates the integrity of an algorithm.

Error: Classification error criterion and classification accuracy criterion are two opposite extremes. The best performance occurs when the lowest value is zero, and similarly, the least efficiency occurs when the maximum value is equal to one.

2.4. An overview of the pattern recognition

Pattern recognition is applied in a variety of domains such as data classification, image processing, data mining, industrial automation, handwriting analysis, voice recognition, etc. The similarity of all the programs that are used in all domains in order to obtain a solution requires the recognition, extraction, and the analysis of the features. It is common in the identification or classification for all purposes [12-15].

There are three common steps in all of the pattern recognition processes:

A. Data extraction: A process for converting primary data into a form that is acceptable by the calculator and capable of processing.

B. Data analysis: At this stage, data analyst training data is analyzed in different situations and events.

C. Classification: At the analysis stage, new data is decided to be assigned to the known categories and classifications.

3. Database introduction

The musical database of the Turkish musical instruments was used in order to carry out this experiment. For this purpose, Turkish musical instruments were divided into four categories: bow, plectrum, percussion, and woodwind. The groups identified by index 1, 2, 3, and 4 are plectrum, bow, woodwind and percussion groups respectively. In each category, the music files are in different styles and from different instruments according to the type of structure. A total of 400 music files were used in this research. Then the preprocessing steps, such as noise cancellation and voice enhancement, were done by using the FFT algorithm [16-19].

4. Methods

First, an information database was created and then the preprocessing stages were performed on it in order to evaluate the role of sound characteristics in the quality of the identification and classification of Turkish instruments. First and foremost, a complete set of features including general, longitude, and long and medium length features were extracted from all data in the original database as described in table 1. Afterward, three experiments were designed and implemented. In the first experiment, the general characteristics of the information database were extracted and the results were compared with that of the extraction of all features by the evaluation criteria. In the second experiment, the features of short length, and in the third experiment, the long and medium length were extracted from the information database. Like general features, the results were compared with that of the whole test of the extracted properties. The results are described in the following sections.

5. Performance evaluation

5.1. Performance evaluation by using general features

Most of the research-related features of the previous researches have been extracted from this dataset in order to evaluate the efficiency and effect of the general features of the dataset presented in this study. In general, the neural network algorithm was taught and tested using a dataset containing predominantly introduced features. The extracted features are as follows:

Spectral Centroid, Spectral Flux, Spectral Roll-Off Point, Spectral Variability, Low Fraction of Windows, Zero Crossings, MIFCC, Constanta, LPC

Evaluation criteria on the results of computing were set in table 2, as well as comparing the results of using these features and the results of doing them with a test in which all their general features, short and long length, were used. In group I, the results of the dominant extraction of the features that were introduced in the previous researches are shown in the dataset used in this study and the grouping of the instruments was performed using the neural network; and in group II, the results of the experiment were described using all the described features [20, 21].

Table 2. Evaluation criteria on the results of computing by using general features

Group	I				II			
	1	2	3	4	1	2	3	4
Sensitivity	10%	50%	59.25%	12.50%	80%	90%	100%	100%
Exclusivity	73.68%	34.48%	73.68%	100%	100%	95%	94%	100%
Accuracy	51.72%	38.46%	65.22%	69.57%	95.06%	93.90%	96.25%	100%
Verification	16.67%	20.83%	76.19%	100%	100%	85.71%	90%	100%
Error	48.28%	61.54%	34.78%	30.43%	4.94%	6.10%	3.75%	0

5.2. Performance evaluation by using short-length features

The dataset presented in this study was used, and features with a short frame length were extracted from that dataset in order to evaluate the efficiency and impact of short-term characteristics. In fact, the neural network algorithm was taught and tested using a dataset containing features of the short frame length. The extracted features are as follows:

The strength of Strongest Beat, Compactness, Strongest Beat, Bcat sum, Spectral Rolloff Point, Spectral Flux, Spectral Centroid, Fraction of Low Energy frames, Magnitude Spectrum, Power Spectrum, LPC, MFCC, Zero Crossing

Evaluation criteria are based on the results of the calculation and comparison of the results of the experiment with a test in which all general features, short and long lengths, were arranged in the table 3. In a group, I, the results of the extraction of features with a short frame length of the dataset that is used in this study and the categories of instruments are shown using the neural network, and in group II, the results of the experiment are shown using all the described features.

Table 3. Evaluation criteria on the results of computing by using short-length features

Group	I				II			
	1	2	3	4	1	2	3	4
Sensitivity	40%	90%	60.71%	6.25%	80%	90%	100%	100%
Exclusivity	83.72%	49.60%	81.82%	100%	100%	95%	94%	100%
Accuracy	69.84%	60.27%	72.13%	74%	95.06%	93.90%	96.25%	100%
Verification	53.33%	40%	73.91%	100%	100%	85.71%	90%	100%
Error	30.16%	39.73%	27%	25.42%	4.94%	6.10%	3.75%	0

6. Performance appraisal by using medium and long length features

In order to evaluate the efficiency and effect of features with a long and medium length, the datasets presented in this study were used and the long and medium length features were extracted. In fact, the neural network algorithm was taught and tested using a set of data including features that are medium and high frame rates. Features with moderate frame lengths come from multiple consecutive short frames. Therefore, operators such as deviations from the metric for aggregation in the middle frame phase are used. Features with long frame lengths are created from the combination of features with medium frame lengths. Features with an average frame length of 1 and features with a long frame length of 10 seconds. In order to carry out this test, the standard deviation and the mean introduced in the general features section were used. The corresponding features were extracted using a standard deviation of the dataset and introduced as an input to the neural network. The proposed evaluation criteria that are based on the results of the calculation and the comparison of the results of this experiment in which all the general features were used, short and long lengths, were arranged in table 4. In a group, I, the results of the features extraction with a long frame length from the dataset used in this research and the categorization of instruments by using the neural network are shown. And in group II, the results of the experiment are shown using all the described features [22, 23].

Table 4. Performance appraisal by using medium and long length features

Group	I				II			
	1	2	3	4	1	2	3	4
Sensitivity	100%	30%	25.93%	89.47%	80%	90%	100%	100%
Exclusivity	46.88%	95.65%	100%	100%	100%	95%	94%	100%
Accuracy	59.92%	75.76%	71.43%	96.25%	95.06%	93.90%	96.25%	100%
Verification	37.40%	75%	100%	100%	100%	85.71%	90%	100%
Error	40.48%	24.24%	28.57%	3.85%	4.94%	6.10%	3.75%	0

7. Conclusions

By examining the experiments and comparing the results, it can be understood that music is a combination of different traits. Therefore, these attributes should be taken into account in the extraction of the features, and acceptable results cannot be yielded only by the extraction of specific and general features that have been used in most of the previous studies. Thus, in the dataset derived from Turkish solo instruments for categorizing instruments, the use of features that are often specific to speech recognition cannot result in an acceptable and appropriate outcome. Moreover, using only features that are short or long, and medium in length, cannot provide an efficient and acceptable result. As a result of the experiments, the identification of Turkish musical instruments with medium and long frame lengths in identifying bow and percussion groups would provide a more acceptable result than that of others [21, 23]. Furthermore, short lengths features in the identification of woodwind instruments will provide more acceptable results than that of the rest of the categories. However, the use of only one of these two groups of features cannot provide effective and acceptable results [11, 24].

References

[1] K. Portorab, M. Alizadeh, H. Fatah. Theoretical and Constructive Foundations of Iranian Music. Printing and Publishing Company of Iranian Textbooks 2012.

[2] P. Mansoori. Theory of Music. in: Workshop, (Ed.).1995.

[3] D. Giannoulis, E. Benetos, A. Klapuri, M.D. Plumbley. Improving instrument recognition in polyphonic music through system integration. Acoustics, Speech and Signal Processing (ICASSP), 2014 IEEE International Conference on. IEEE2014. pp. 5222-6.

[4] S.K. Banchhor, A. Khan. Musical instrument recognition using zero crossing rate and short-time energy. Musical Instrument. 1 (2012).

[5] A. Eftekhari. Hearing Aids, Shahid Beheshti University, 2010.

[6] B. Vojdani. A Comprehensive Dictionary of Iranian Music. in: Mirran, (Ed.).2005.

[7] A. Atrabi, M.R. Darvishi. Iranian Systematics. Publishing Company of Iran Textbooks2013.

[8] M. Lotfi, B. Misganaw, M. Vidyasagar. Prediction of Time to Tumor Recurrence in Ovarian Cancer: Comparison of Three Sparse Regression Methods. International Symposium on Bioinformatics Research and Applications. Springer2017. pp. 1-11.

[9] A. Perrakis, W. Hohenberger, T. Horbach. Integrated operation systems and voice recognition in minimally invasive surgery: comparison of two systems. Surgical endoscopy. 27 (2013) 575-9.

[10] N. Kubota, Y. Toda. Multimodal communication for human-friendly robot partners in informationally structured space. IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews). 42 (2012) 1142-51.

[11] R. Kothe, D. Bhalke, P. Gutal. Musical instrument recognition using k-nearest neighbour and Support Vector Machine. Advances in Electronics, Communication and Computer Technology (ICAECCT), 2016 IEEE International Conference on. IEEE2016. pp. 308-13.

[12] M.A. LAYEGH, S. HAGHIPOUR, Y.N. SAREM. Classification of the Radif of Mirza Abdollah a canonic repertoire of Persian music using SVM method. Gazi University Journal of Science Part A: Engineering and Innovation. 1 (2013) 57-66.

[13] P. Heydarian. Automatic recognition of Persian musical modes in audio musical signals. London Metropolitan University2016.

[14] S. Abdoli. Iranian Traditional Music Dastgah Classification. ISMIR2011. pp. 275-80.

[15] M. Lotfi, M. Vidyasagar. Exact recovery of sparse signals from binary measurements. Indian Control Conference (ICC) 2018. IEEE2018. pp. 83-8.

[16] O.P. Prabhakar, N.K. Sahu. A survey on: Voice command recognition technique. International Journal of Advanced Research in Computer Science and Software Engineering. 3 (2013).

[17] S. Asht. Pattern recognition techniques: A review. International Journal of Computer Science and Telecommunications. 3 (2012).

[18] Y. Han, J. Kim, K. Lee, Y. Han, J. Kim, K. Lee. Deep convolutional neural networks for predominant instrument recognition in polyphonic music. IEEE/ACM Transactions on Audio, Speech and Language Processing (TASLP). 25 (2017) 208-21.

[19] W. Chen, Z.-J. Fu, C.-S. Chen. Recent advances in radial basis function collocation methods. Springer2014.

[20] A.K. Datta, S.S. Solanki, R. Sengupta, S. Chakraborty, K. Mahto, A. Patranabis. Automatic Musical Instrument Recognition. Signal Analysis of Hindustani Classical Music. Springer2017. pp. 167-232.

[21] Y. Han, S. Lee, J. Nam, K. Lee. Sparse feature learning for instrument identification: Effects of sampling and pooling methods. The Journal of the Acoustical Society of America. 139 (2016) 2290-8.

[22] D. Bhalke, C.R. Rao, D. Bormane. Hybridization of Fractional Fourier Transform and Acoustic Features for Musical Instrument Recognition. International Journal of Signal Processing, Image Processing and Pattern Recognition. 7 (2014) 275-82.

[23] K. Patil, M. Elhilali. Biomimetic spectro-temporal features for music instrument recognition in isolated notes and solo phrases. EURASIP Journal on Audio, Speech, and Music Processing. 2015 (2015) 27.

[24] I. Vatolkin, M. Preuß, G. Rudolph, M. Eichhoff, C. Weihs. Multi-objective evolutionary feature selection for instrument recognition in polyphonic audio mixtures. Soft Computing. 16 (2012) 2027-47.