International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 05 Issue: 11 | Nov 2018 www.irjet.net p-ISSN: 2395-0072

# Chord Classification of an Audio Signal using Artificial Neural Network

**Ronesh Shrestha**

*Student, Department of Electrical and Electronic Engineering, Kathmandu University, Dhulikhel, Nepal*

---

**Abstract -** *The variations that may arise in different chords played at different time creates a challenging problem while performing chord classification. Hence, this project proposes an effective machine learning based supervised learning method using the two-layer feed-forward network which is trained with scaled conjugate gradient backpropagation in MATLAB for chord classification. In this project, logarithmic compression techniques are used to extract the Chroma DCT-Reduced Log Pitch (CRP) feature from an audio signal. This chroma feature is extracted from the training set, which is a database containing 2,000 recordings of 10 guitar chords. For each chord, there are 200 '.wav' files sampled at 44.100 KHz and quantized at 16 bits. The CRP features of all the 2,000 samples were extracted and this data was used as the training set for the artificial neural network. Each sample for each chord was truncated to 12x10 matrix. The neural network was modeled with 5 hidden layers. This trained neural network was then used to classify the input chord. The result of this method had an overall accuracy of 89.3%.*

***Key Words***: **Chord classification, machine learning, artificial neural network, chroma DCT-Reduced Log Pitch (CRP), chroma feature.**

## 1. INTRODUCTION

A chord is defined as a harmonic set of two or more musical notes that are heard as if they were simultaneously sounding [1]. These are considered to be one of the best characterizations of music. The expansive production of digital music by many artists has made it very difficult to process the data manually but opened the door to automate information retrieval of music. Although, many researches and algorithms have been devised and applied to extract information from a musical signal, this research focuses mainly on chord and its classification.

A musical note is a single tone of a specified pitch that is sustained for a given duration [2]. Since, musical note is the building block of music, it is important to identify the notes present in it. Further analysis of these notes can then result in classifying a chord successfully.

Chord classification is a difficult task to perform due to the dynamic variations of different chords that are played differently. Although, there is a mathematical relationship between the chords, it is very difficult to model it. Hence, in order to model this complex relationship and not impose any restriction in the possibility of input variations, this research makes use of artificial neural network.

## 2. RELATED WORK

There are different ways to identify a chord. Pitch class Profile (PCP) is one of the methods to identify a chord, which was first proposed by Fujishima in [1]. The PCP introduced by Fujishima is a twelve-dimension vector that represents the intensities of the twelve semitone pitch classes [1]. Also, Hidden Markov Model (HMM) proposed by Sheh and Ellis (Sheh and Ellis, 2003) has been notable in the area of chord recognition which uses probabilistic chord template as in [3]. Harte and Sandler have also proposed a method using the Constant Q-Transform (CQT) for chord recognition in [4]. Harte and Sandler derived a 12-bin semitone quantized chromogram in order to automatically identify the chord.

However, this project uses the Chroma DCT (Discrete Cosine Transform)-Reduced Log Pitch (CRP) introduced in [5] as the feature to train the artificial neural network (ANN) in order to develop a system model capable of chord identification.

## 3. PROPOSED ALGORITHM

The methodology involved with the chord recognition technique is mainly based on two steps: chroma feature extraction and pattern matching. For the feature extraction, this project has used the CRP feature extracted from waveform-based audio signals. For the pattern matching process, this project has used ANN where all of the inputs and the standard audio signals (chords) are compared on the basis of their chroma features and the output is displayed on the basis of the comparison. The project uses the same dataset as used by Osmalskyj, Julien & Embrechts, Jean Jacques & Droogenbroeck, Marc & Piérard, Sebastian in [6]. Also, it is important to note that the dataset introduced in [6] are limited to the most frequent chords which are a subset of 10 chords as: A, Am, Bm, C, D, Dm, E, Em, F, G. This project uses the first subset of dataset introduced in [6] which are produced with an acoustic guitar for extracting the CRP feature.

### A. *Chroma Feature Extraction*

In this step, the harmonic features are extracted from the audio signal. A chroma feature vector, also referred to as pitch class profile (PCP), represents the energy distribution of a signal's frequency content across the 12 pitch classes of the equal-tempered scale. A temporal sequence of these chroma vectors is often called a chromagram [7]. However, in this project, the chroma feature extracted is the CRP

feature that has been introduced in [5]. The CRP feature helps to boost the degree of timbre invariance. The general idea is to discard timbre-related information similar to that expressed by certain mel-frequency cepstral coefficients (MFCCs) [8].

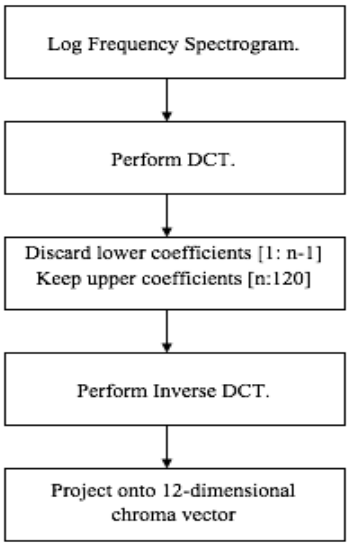



**Fig - 1:** Steps involved in calculating CRP Feature.

In the first phase, the nonlinear mel-scale is replaced with a nonlinear pitch scale and then DCT is applied on the logarithmized pitch representation to obtain pitch-frequency cepstral coefficients (PFCCs). Then only the upper coefficients are kept, and an inverse DCT is applied, and finally the resulting pitch vectors is projected onto 12-dimensional chroma vectors [9]. These vectors are referred to as CRP features [9]. The flowchart of calculating CRP feature is shown in Fig - 1.

B. *Training the Artificial Neural Network for pattern matching*

After extracting the CRP feature of 2000 samples of guitar chords, i.e. 200 samples of 10 chords (A, Am, Bm, C, D, Dm, E, Em, F and G), the stored data is converted into a 'csv' file. The CRP feature extracted vary in size from 12x10 to 12x50 depending upon the length of the respective '.wav' file. In order to prepare a uniform dataset of the training chord, all the random length of feature is truncated into 12x10 matrix. This matrix signifies the 10-feature value for each sample.

The 'csv' file is then named '*training.csv*'. The size of this dataset is 12x20000. A target dataset is then prepared that is equivalent in size corresponding to the training dataset. This target dataset represents the chord respectively as A, Am, Bm, C, D, Dm, E, Em, F and G. The target dataset is stored in 'target.csv' file.

The training dataset and target dataset is fed into the neural network pattern recognition tool in MATLAB in order to train the dataset. The tool is a two-layer feed-forward network with sigmoid output neurons. The hidden layer is selected to be 5. The network is established for 70% training data, 15% validation data and 15% testing data. After running the whole network, the neural network was able to classify the input dataset. The steps involved in training ANN is shown in Fig - 2.

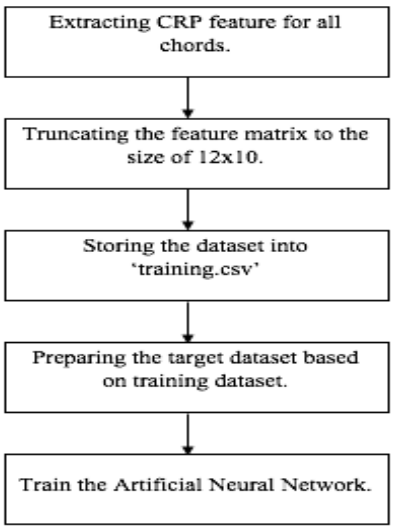



**Fig - 2:** Steps involved in training ANN.

## 4. SIMULATION RESULTS

The evaluation was done via simulation in MATLAB. The feature vector was extracted as explained in Section III and the neural network was trained using the neural network toolbox [10]. Chord-C from the dataset of chords was retrieved and it was implemented in the program. The following are the results after running the input signal through it.

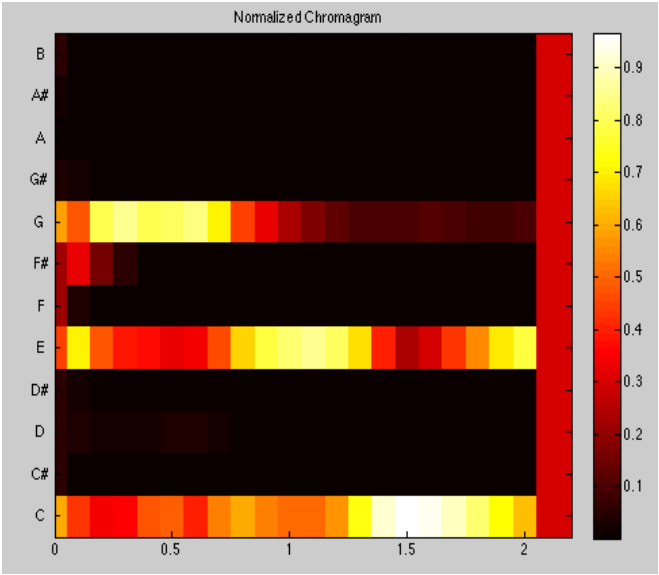



**Fig - 3:** Representation of Normalized Chromogram performed on Chord-C.

In Fig - 3, the normalized chromagram of guitar chord C is shown. The chord C consists of 3 notes A3 (pitch=57), C4 (pitch=60) and E4 (pitch=64) [11]. It can be clearly seen in Fig – 3 that the signal's energy is contained in chroma A, C and E. The smaller amount of energy seen in band G comes from G5, which is the third harmonic of C4 [11].

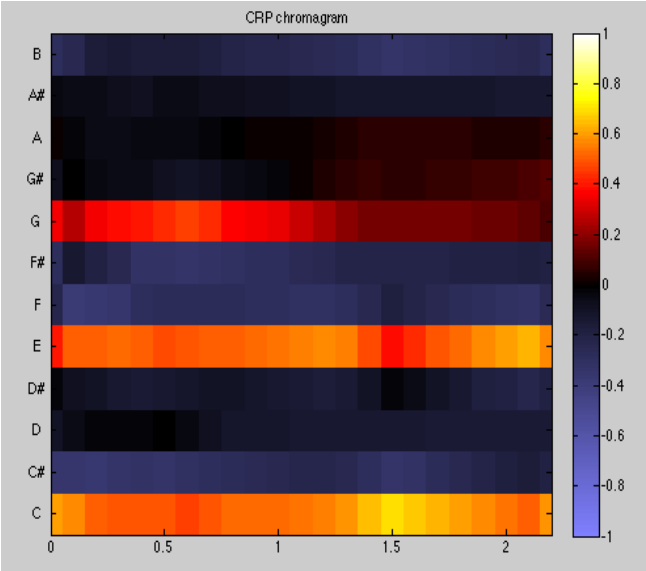



**Fig - 4:** Representation of CRP Chromogram, performed on Chord-C.

In Fig - 4, there is a boost in the degree of timbre invariance [9]. It can be seen that the timbre-related information has been discarded and the non-linear pitch scale had been applied with DCT on the logarithmized pitch presentation as explained in [9]. Then inverse DCT for upper coefficients had been performed and plotted to give a smoothed CRP chromagram.

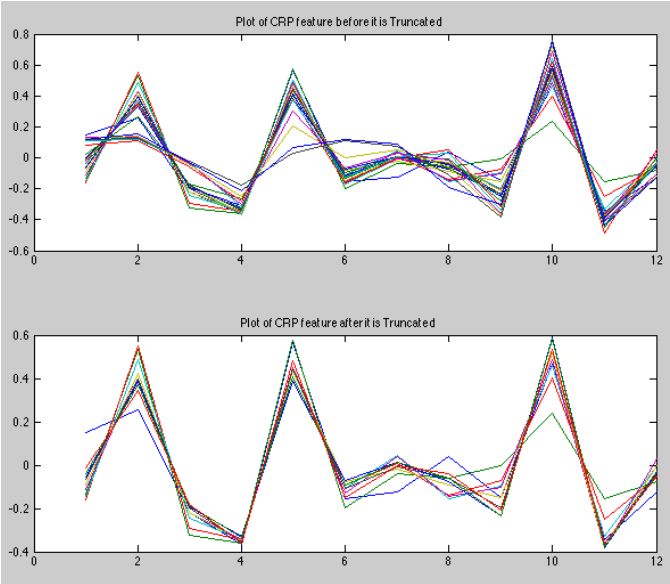
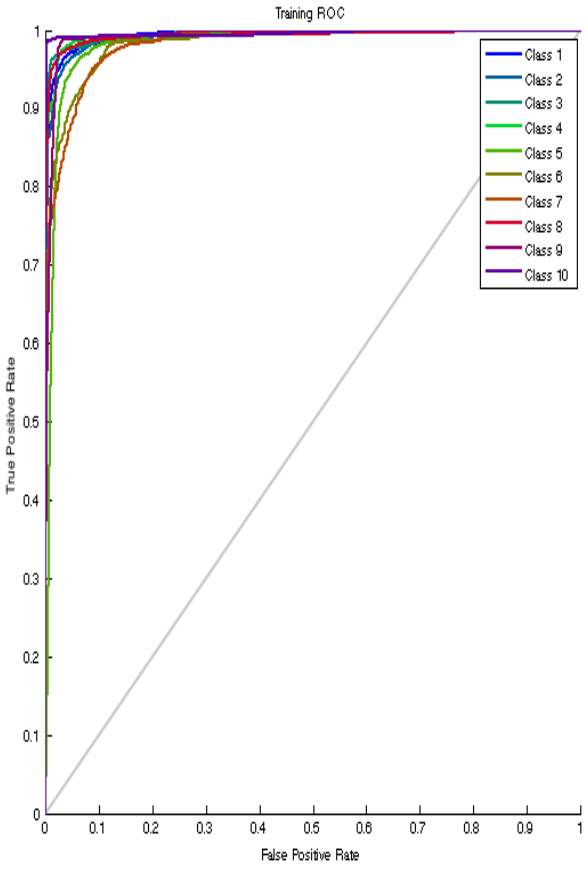



**Fig - 5:** Effect of truncating the feature vector.

The upper plot in Fig - 5 shows the graph representing the CRP feature before it is truncated for chord class 'A'. It consisted of 12x13 data values that were plotted. Hence, the 12 data points had 13 elements for each and after truncating the feature to 12x10, the 12 data points had 10 elements for each, which is shown on the lower plot of Fig - 5.

In this way, it can be observed that with decrease in the columns of each CRP feature, there is reduction in its sample element that is redundant. Hence, the truncating of the data feature does not severely damage the output, and likewise, the training data prepared is also not affected much as it consists of 200 samples for each class of chord. Each sample also has 10 elements to signify its feature. In totality, for each chord sample there are 2000 samples that the ANN is trained with. This sums up to 20000 datasets for 10 chords.

After training the given sequences and plotting the Receiver Operating Characteristic (ROC) for targets, it was observed that the system behaved very effectively as the ROC plot suggested in Fig - 6. The ROC plot in Fig - 6 shows the percentage of true positive class predictions as a function of how many false positive class predictions that the system is willing to accept. It can be seen that the line follows towards the top and left of the plot which attributes to better result.



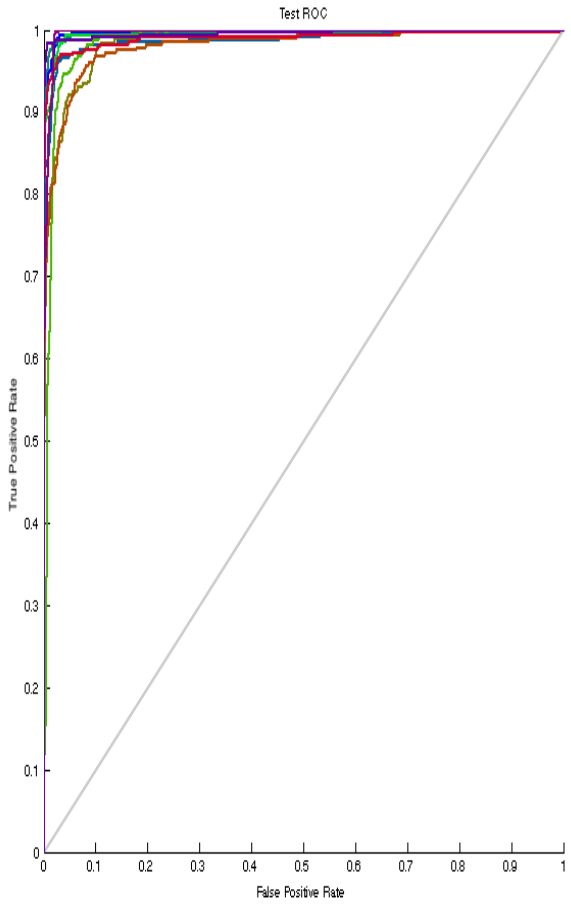



**Fig - 6:** ROC plot showing that ANN performs well with given inputs and target dataset.

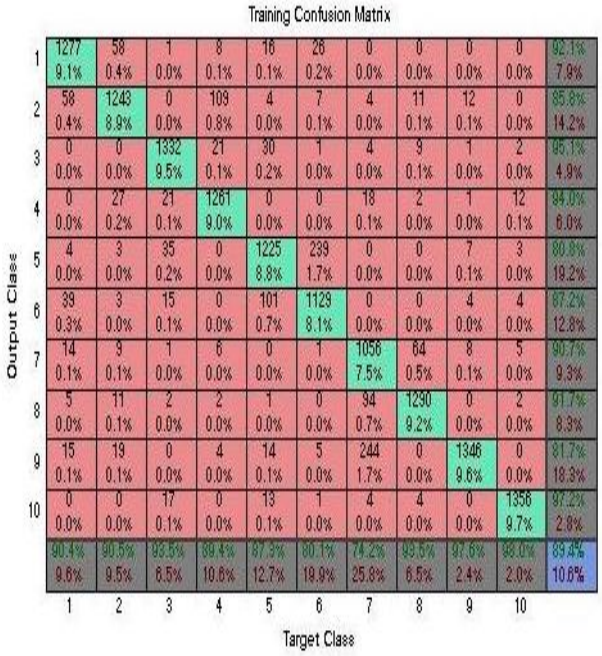



**Fig - 7:** Training confusion matrix.

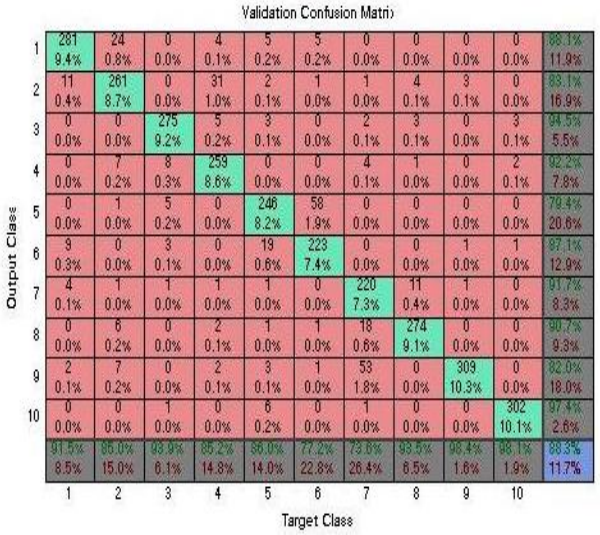



**Fig - 8:** Validation confusion matrix.

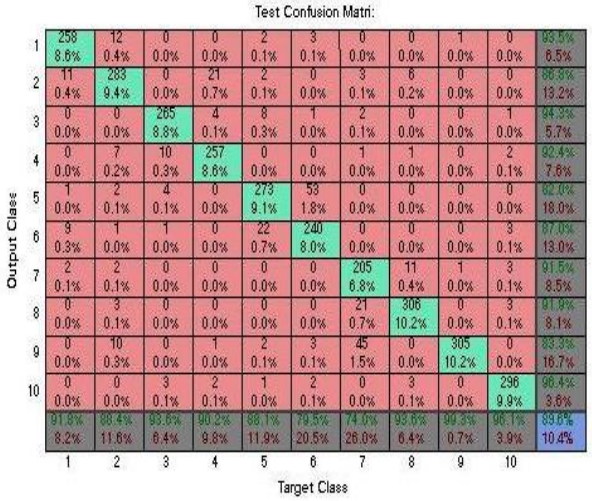



**Fig - 9:** Test confusion matrix.

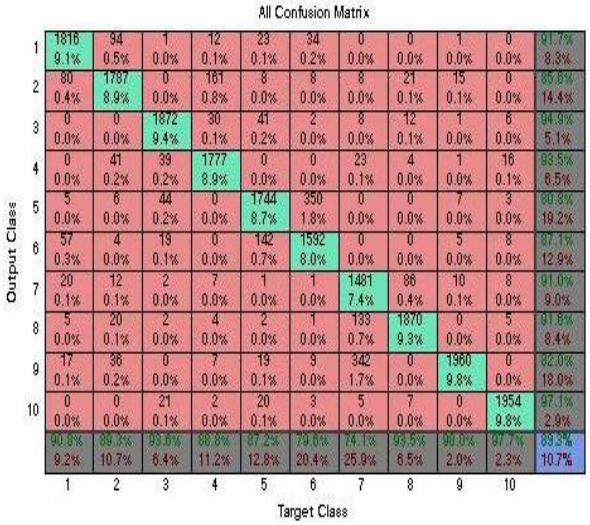



**Fig - 10:** All confusion matrix.

In order to model the Artificial Neural Networks (ANN), as suggested in proposed methodology, 70 %, 15 %, 15 % of data were used for training, validating, and testing, respectively. In the confusion matrix shown in Fig - 7, the data set is divided into 10 sets, which is given by the number 1-10 in the matrix. The first matrix i.e. 1 represents a note and subsequently other row follows the order as A, Am, Bm, C, D, Dm, E, Em, F and G. The result obtained for the training, validation and test are shown in Fig - 7, Fig - 8 and Fig - 9 respectively. The final accuracy of the system model is found to be 89.3% which is the correctly classified dataset and 10.7% of the dataset is incorrectly classified.

## 5. CONCLUSION AND FUTURE WORK

In this project, a wide variety of state-of-the-art chord recognition techniques were investigated, and several novel methods were discussed with the aim of improving chord recognition performance. However, this project has proposed to devise a chord recognition system using ANN. This project has tended to focus the attention on only one instrument namely guitar for the time being and take a live recording of a guitar chord for it to be examined. This project proposes a model to classify a chord with accuracy of 89.3%. With the incorporation of machine learning, the scale to which the end result must be satisfied has risen to a greater extent. However, the project has been in the right track to accomplish its goals.

The future work will be applying machine-learning mechanism to full extent to transcribe the chords by supplying the greater dataset. The dataset required to train the neural network needs to be larger and it must have more variations in order to perform with increased accuracy in real-time world.

## REFERENCES

[1] T. Fujishima. 'Realtime chord recognition of musical sound: A system using common Lisp music.' Proceedings International Computer Music Conference (ICMC), Beijing, China, 1999.

[2] L. Coffey. "Elpin – What is a note?", April 2, 2010. [Online]. Available: http://www.elpin.com/tutorials/musicalnote.php. [Accessed: Nov. 3, 2018].

[3] A. Sheh and D. Ellis. 'Chord segmentation and recognition using EM-trained hidden Markov models.' Proceedings 4th International Society for Music Information Retrieval Conference (ISMIR), pp. 185–191, 2003.

[4] C. Harte and M. Sandler. 'Automatic chord identification using a quantised chromagram.' Proceedings of the 118th Audio Engineering Society (AES), Barcelona, Spain, 2005.

[5] M. Müller and S. Ewert. 'Chroma Toolbox: MATLAB implementations for extracting variants of chroma-based audio features.' Proceedings of the 12th International Society for Music Information Retrieval Conference (ISMIR), Miami, Florida, USA, pp. 215-220, 2012.

[6] J. Osmalskyj, J. J. Embrechts, S. Piérard and M. Van Droogenbroeck. 'Neural Networks for Musical Chords Recognition.' Journees D'Informatiotique Musicale, Mons, Belgium, 2012.

[7] G. Wakefield. 'Mathematical representation of joint time-chroma distributions.' Proceedings SPIE Int. Symp. Opt. Sci., Eng., Instrum., vol. 99, pp. 18–23, 1999.

[8] M. Müller, S. Ewert, and S. Kreuzer. 'Making chroma features more robust to timbre changes.' Proceedings International Conference on Acoustics, Speech, and Signal Processing (ICASSP), Taipei, Taiwan, pp. 1869–1872, 2009.

[9] M. Müller and S. Ewert. 'Towards timbre-invariant audio features for harmony-based music.' IEEE Transactions on Audio, Speech and Language Processing, vol. 18, no. 3, pp. 649-662, 2010.

[10] MATLAB and Neural Network Toolbox Release 2008a, The MathWorks, Inc., Natick, Massachusetts, United States.

[11] M. Müller. 'Fundamentals of Music Processing: Audio, Analysis, Algorithms, Applications.' Springer International Publishing, 2015, pp. 123-125. Accessed on: Nov. 3, 2018. [Online]. doi: 10.1007/978-3-319-21945-5.