Musical Instrument Recognition and Transcription using Neural Network

V. S. Shelar, D. G. Bhalke
JSPM's Rajarshi Shahu College of Engineering, Tathwade, Pune, (M.S), India

ABSTRACT
In this paper musical instrument recognition and transcription for piano, guitar, violin is discussed. The system is implementing in two stages; first stage is musical instrument recognised using spectral features after recognising instrument musical note is recognised using different frequency estimation methods. Feed forward Neural Network has been used as classifier. The system is implemented for Single Instrument Single Note (SISN), Single Instrument Multiple Note (SIMN) and Multiple Instrument Multiple Note (MIMN). The average accuracy is achieved for three instruments is recorded 80%.

Keywords
Feature, Feature extraction and Music Transcription.

1. INTRODUCTION
Music information retrieval has many practical applications including structured coding, database retrieval systems and automatic musical signal annotation. A subtask of this, automatic musical instrument identification is of significant importance in solving these problems. Therefore there is necessity to develop efficient musical instrument identification and transcription system. In Monophonic Music notes of single instrument are played one-by-one and in polyphonic music two or several notes of one instrument or more than one instrument can be played simultaneously. Transcription can be used in Audio Watermarking & a visualization of media players, which displays the music score during the playback. It can also be used to monitor students playing musical instruments, by transcribing the music played by the student and evaluating it against the standard score. Transcription also plays an Indispensable role in content based music retrieval, such as query by humming. This transcription system is implemented with the help of feed forward neural network and the instruments we cover in this system are piano, guitar and violin.

(Matija Marolt, 3) first compare the performance of several neural network models on the task of recognizing tones from time-frequency representation of a musical signal. Oscillator networks improve the accuracy of transcription with neural networks. When the system was not specifically tuned for the piano sample used, it correctly found 90% of all notes. (Kenichi Miyamoto, 4) introduces automatic music transcription system in which audio input signal to music score by integrating probabilistic approaches to multi pitch spectral analysis, rhythm recognition and tempo estimation. In spectral analysis, acoustic energies in spectrogram are clustered into acoustic objects with our method called Harmonic-Temporal-structured Clustering (HTC) utilizing EM algorithm over a structured Gaussian mixture and tempo are simultaneously recognized and estimated in terms of maximum posterior probability given a probabilistic note duration models with HMM. Here, nearly - correct score was estimated successfully for piano. From this Literature survey we conclude that, Automatic music transcription (AMT) is basically consist of Time-Frequency analysis from that we can extract features. These extracted features give feature vector. Finally with the help of Neural Network we are able to obtain a note sequence.

Acknowledging the challenges inherent to designing good features, Pachet et al pioneered work in automatic feature optimization [5], and more recently deep learning methods have been employed to produce robust Tonnetz features [6]. Alternatively, some work leverages the repetitive structure of music to smooth a chroma features prior to classification [7]. Various classification strategies have been investigated to a lesser extent [8], but Gaussian Mixture Models (GMM) are conventionally preferred for the probabilistic interpretation. The choice of post filtering methods has been shown to significantly impact classification accuracy, and much research has focused on properly tuning HMMs [9], in addition to exploring other post-filtering methods such as Dynamic Bayesian Networks (DBNs) [10]. Hence the significant information that characterizes the music signal need to be extracted. Our task is to find out most important information of the signal in time domain as well as in frequency domain. Also it is proposed that to work on different feature vectors of signal; so that we will find the best feature to identify the instrument. These features then utilized in neural networks for transcription. In this project system is developed for monophonic music Transcription.

Introduction is discussed in this section. Proposed method is for musical instrument transcription is discussed in section 2. Result is discussed in next section 3.

2. PROPOSED METHOD
The music samples are collected from McGill University Master Samples collection, a fabulous set of DVDs of instruments playing every note in their range, recorded in studio conditions. One important reason to use .WAV files of arbitrary size. The signals are samples at 44.1 KHz. The music signal is pre-processed before going into the feature extraction block. The pre-processing is simply a kind of normalization by scaling the sampled sound file data to fall within the range of -1 to 1.

In this paper we have discussed the system for three type of signal namely as follows:

i. Signal consist of Single Instrument Single Note (SISN)

ii. Signal consist of Single Instrument Multiple Note (SIMN)

iii. Signal consist of Multiple Instrument Multiple Note (MIMN)

Following is the block diagram for musical instrument recognition; after recognising the musical instrument transcription is done with the help of fundamental frequency estimation.

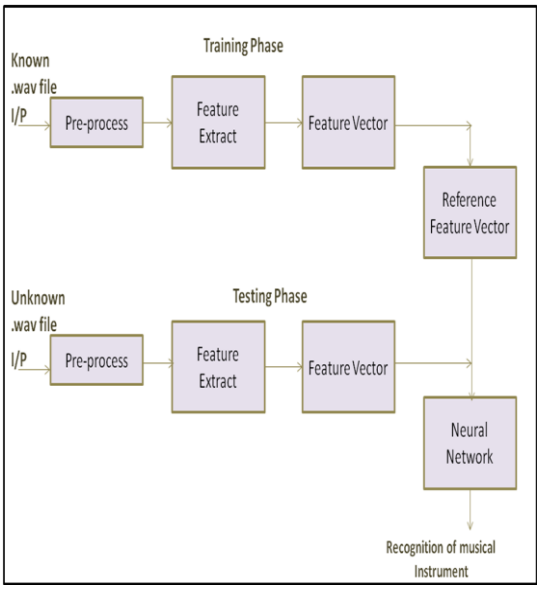


Fig 1 Block diagram of Musical Instrument Recognition

Feature Extraction: The purpose of feature extraction is to obtain the relevant information from the input data to execute certain task using desired set of features. The feature extraction methods discussed here covers the temporal features and spectral features. Due to large numbers of audio features are available, how to chose or combine them to achieve higher transcription accuracy is studied in this report. Desired feature selection is necessary for best performance. Features extraction procedure is very important to find feature vector. In feature extraction procedure, first silence part of music signal is detected and removed from signal. The music signal is pre-processed before going into the feature extraction block.

Energy: Energy is simply the sum of the amplitudes present in a frame, and is defined as in

$$Energy = \frac{1}{N}\sum_{n-1}^{N-1}(x[n])^2 \qquad (1)$$

Where, x[n] is the amplitude of the sample.

Spectral Features: Spectral features are obtained from the samples in the frequency domain of the musical signal. These features are extracted by considering Fourier transform of signal frame in our system.

Spectral Centroid (Brightness): It measures the average frequency weighted by amplitude of a spectrum. This is the amplitude-weighted average, or centroid, of the frequency spectrum, which can be related to a human perception of 'brightness'. It is calculated by multiplying the value of each frequency by its magnitude in the spectrum, then taking the sum of all these. The value is then normalized by dividing it by the sum of all the magnitudes. It is defined as in

$$\text{Brightness} = \left(\frac{\sum mag[i]\times freq[i]}{\sum mag[i]}\right) \qquad (2)$$

Where: mag = the magnitude spectrum.

Freq = the frequency corresponding to each magnitude element.

Spectral Flux: This is a measure of the amount of local spectral change. This is defined as the squared difference between the normalized magnitude spectra of successive frames. It is defined as in

$$\text{Flux} = \sum(norm_f[i]-(norm_{f-1}[i])^2 \qquad (3)$$

Where norm is the magnitude spectrum of the current frame scaled to the range 0….1, and norm f-1 is the normalized magnitude spectrum of the previous frame.

Spectral Spread: The spectral spread is a measure of variance (or spread) of the spectrum around the mean value μ. It is defined as in

$$\text{Spectral spread} = \sqrt{\frac{\sum_{k=0}^{N/2}(freq_k - SC)^2 mag^2}{\sum_{k=0}^{N/2} mag^2}} \qquad (4)$$

Where:

mag = the magnitude

spectrum.Freq = the frequency corresponding to each magnitude element.

SC = Spectral centroid

Spectral Skewness: The skewness is a measure of the asymmetry of the distribution around the mean value. The skewness is calculated from the 3rd order moment. It is defined as in

$$\text{Spectral skewness} = \frac{\sum(freq-SC)^3 \times mag}{\sum mag} \qquad (5)$$

Where: mag = the magnitude

Spectrum Freq = the frequency corresponding to each magnitude element.

SC = Spectral centroid.

Spectral roll-off: Spectral Roll-off is defined as the frequency bin M below which 85% of the magnitude distributions concentrated. This is one more measure of Spectral Shape.

$$\text{Spectral Roll off} = \sum_{n=0}^{M} f(n) = 0.85 * \sum_{n=0}^{N} f(n) \qquad (6)$$

Feed Forward Neural Network: A feed forward neural network begins with an input layer. This input layer must be connected to a hidden layer. This hidden layer can then be connected to another hidden layer or directly to the output layer.

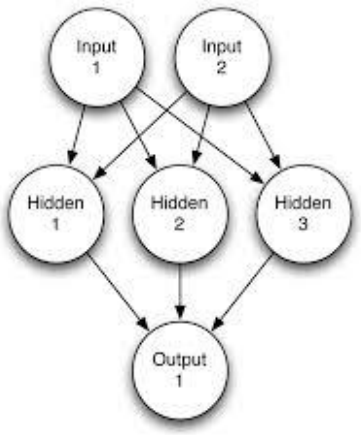


Fig 1: Feed forward neural network

To implement the robust system, it is necessary to increase the performance of system. The performance of system is improved by using the selected features we are boosting some features. The input to classifier is all features as feature vector. The System is used to build for following three different signals which is consist of SISN, SIMN and MIMN.

Fundamental Frequency F0: The dominant Fundamental Frequency F0 of a signal is detected using a technique called Autocorrelation. The technique is to multiply the frame by a time lagged copy of itself, then to measure the amplitude of the new signal. Where the amplitude reaches its peak will be where the peak (s) of the original signal are multiplied by the peak(s) of its copy, i.e. where the first period of the signal has been completed. The value of the time-lag where this peak occurs can then be considered the Fundamental Frequency F0 of the signal. The autocorrelation function is defined as in

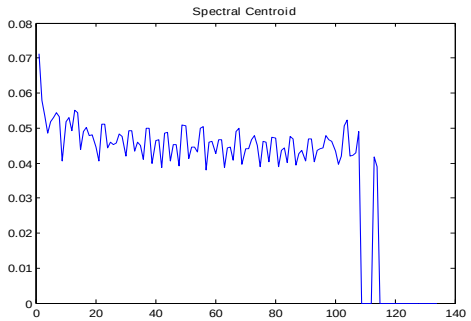
$$\text{Autocorrelation (k)} = \sum_{t=1}^{N} x(t)x(t-k) \qquad (7)$$

i. e. the signal X(t) multiplied by time-lagged copy of itself x (t-k).

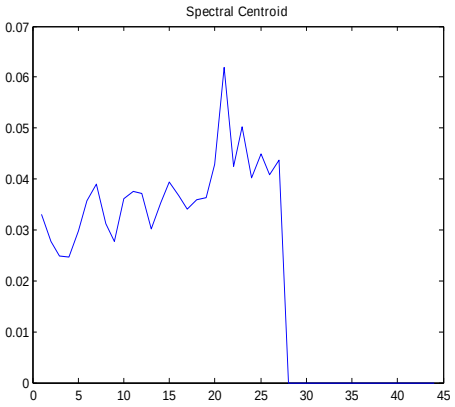
3. RESULT

The Transcription of musical instrument using Feed Forward Neural Network is done. We use spectral features for FFNN classifier. Following are the result of spectral features, which are shown in fig.

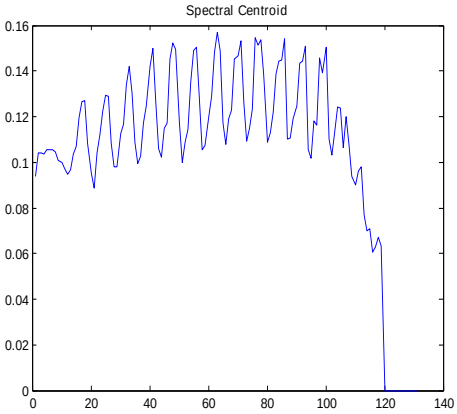
Spectral Centroid of Piano C5, Guitar C5 and Violin C5 music signal are shown in fig 1 (a,b,c respectively).



(a)



(b)



(c)

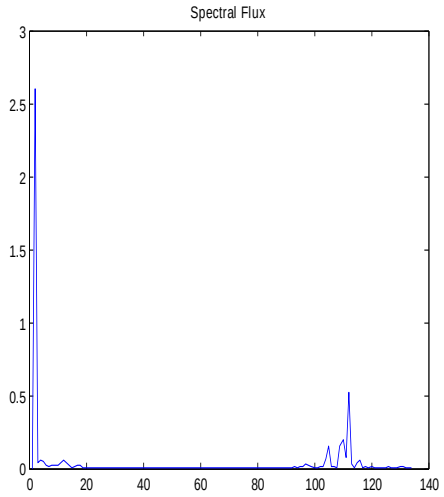
Fig1 (a). Spectral Centroid of Piano C5,
(b). Spectral Centroid of Guitar C5,
(c). Spectral Centroid of Violin C5

Table 1
Spectral Centroid Values

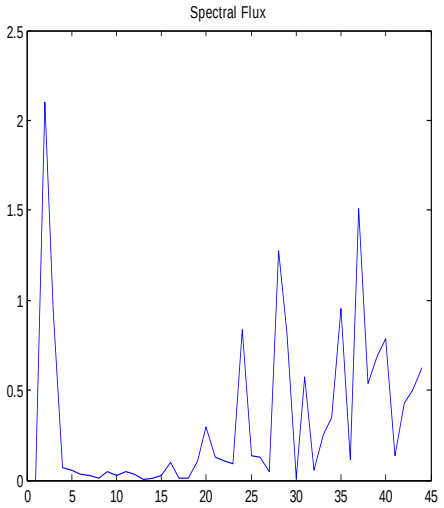
Note	Piano	Guitar	Violin
C1	0.0352	-	-
C2	0.0333	-	-
C3	0.0317	0.0380	0.0136
C4	0.0455	0.0434	0.0522
C5	0.0378	0.0638	0.1142
C6	0.0351	-	-
C7	0.0561	-	-

Table 1 Shows Spectral Centroid Values of Piano, Guitar and Violin Instrument for C-Notation.

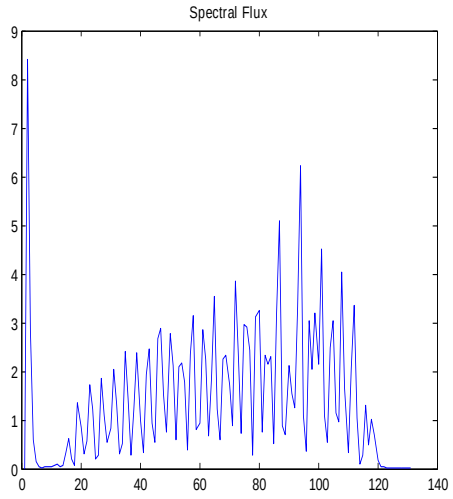
Spectral Flux of Piano C5, Guitar C5 and Violin C5 music signal are shown in fig 2 (a,b,c respectively).



(a)



(b)



(c)

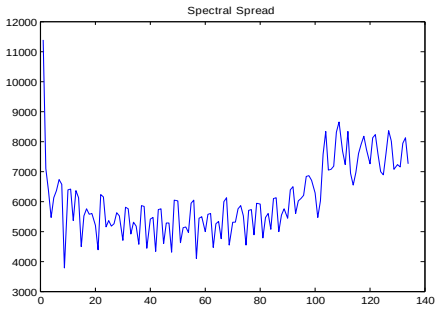
Fig2(a). Spectral Flux of Piano C5,
(b). Spectral Flux of Guitar C5,
(c). Spectral Flux of Violin C5

Table 2
Spectral Flux Values

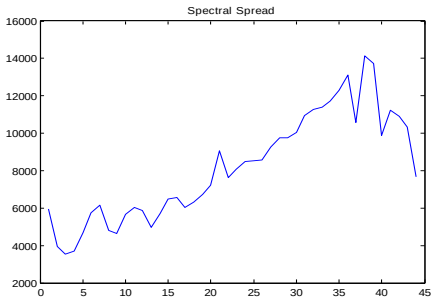
Note	Piano	Guitar	Violin
C1	2.3838	-	-
C2	1.4700	-	-
C3	0.0422	0.0924	
C4	0.0287	0.0379	0.1388
C5	0.0356	0.3751	0.3681
C6	0.0604	-	-
C7	0.0502	-	-

Table 2 Shows Spectral Flux Values of Piano, Guitar and Violin Instrument for C-Notation.

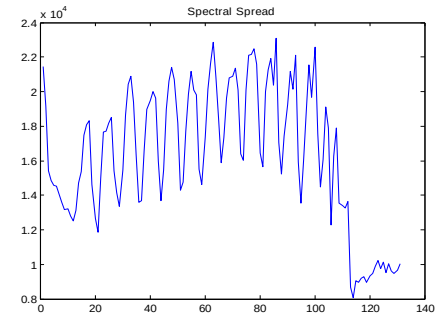
Spectral Spread of Piano C5, Guitar C5 and Violin C5 music signal are shown in fig 3 (a,b,c respectively).



(a)



(b)



(c)

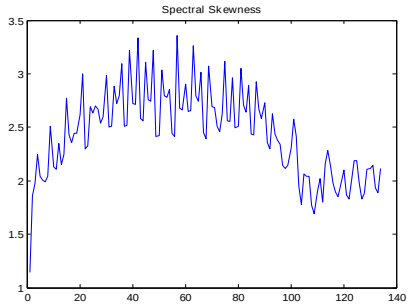
Fig3 (a). Spectral Spread of Piano C5,
(b). Spectral Spread of Guitar C5,
(c). Spectral Spread of Violin C5

Table 3
Spectral Spread Values

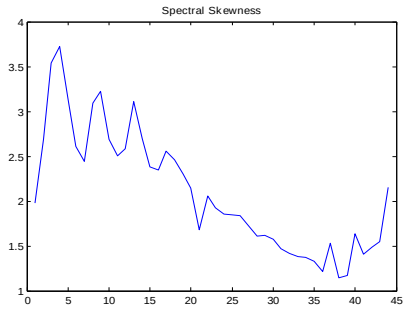
Note	Piano	Guitar	Violin
C1	0.0175	-	-
C2	0.0453	-	-
C3	0.0357	0.0230	
C4	0.0615	0.0514	0.0122
C5	0.0428	0.0418	0.1002
C6	0.0451	-	-
C7	0.0864	-	-

Table 3 Shows Spectral Spread Values of Piano, Guitar and Violin Instrument for C-Notation.

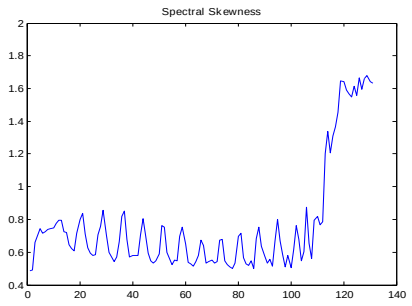
Spectral Skewness of Piano C5, Guitar C5 and Violin C5 music signal are shown in fig 4 (a,b,c respectively).



(a)



(b)



(c)

Fig4 (a). Spectral Skewness of Piano C5,
(b). Spectral Skewness of Guitar C5,
(c). Spectral Skewness of Violin C5

Table 4
Spectral Skewness Values

Note	Piano	Guitar	Violin
C1	0.0642	-	-
C2	0.0333	-	-
C3	0.0317	0.0380	-
C4	0.0455	0.0434	0.0522
C5	0.0378	0.0638	0.1142
C6	0.0351	-	-
C7	0.0561	-	-

Table 4 Shows Spectral Skewness Values of Piano, Guitar and Violin Instrument for C-Notation.
Following is the table 5 which shows the spectral features (Spectral Centroid, Spectral Flux, Spectral Spread, and Spectral Skewness) values of Piano C5, Guitar C5, and Violin C5 Note.

Table 5
Transcription accuracy for Instrument

Instrumen	Transcription accuracy (%)
Piano	84
Guitar	80
Violin	92

Feature vector is formed using these entire features. This feature vector is used as input to FFNN classifier, to classify the musical instruments. The table 5 shows transcription accuracy for different musical instrument. The transcription accuracy is achieved using all features are for Piano is 84 %, Guitar is 80 % and Violin is 92 %. Following is the table 6 which shows transcription accuracy for different type of signal. Signal Consist of Single Instrument Single Note (SISN), Single Instrument Multiple Note (SIMN) and Multiple Instrument Multiple Note (MIMN). The average transcription accuracy is achieved using all features are for SISN is 91 %, SIMN is 84% and MIMN is 83%.

Table 6
Transcription accuracy for different type of signal

Instrument	SISN (%)	SIMN (%)	MIMN (%)
Piano	90	76	86
Guitar	81	84	74
Violin	97	92	87

Following figures shows output of musical instrument transcription system in MATLAB with GUI Model.

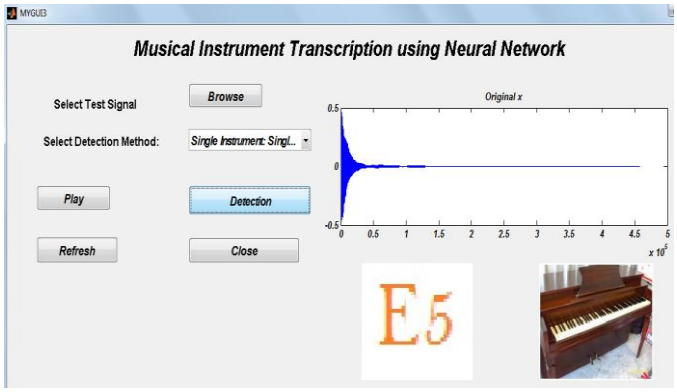


Fig 5: GUI model for Musical instrument transcription using NN for SISN

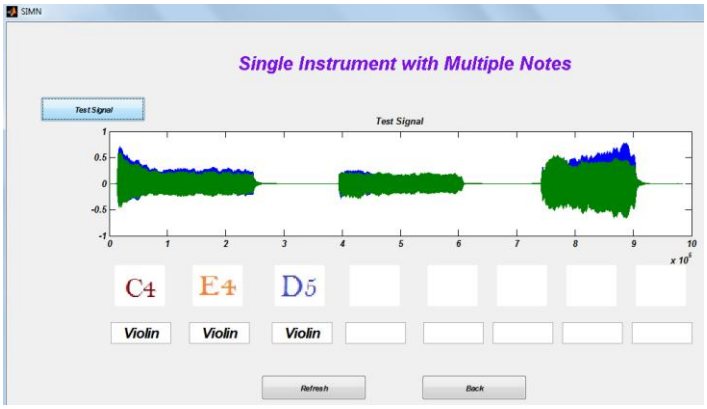


Fig 6: GUI model for Musical instrument transcription using NN for SIMN

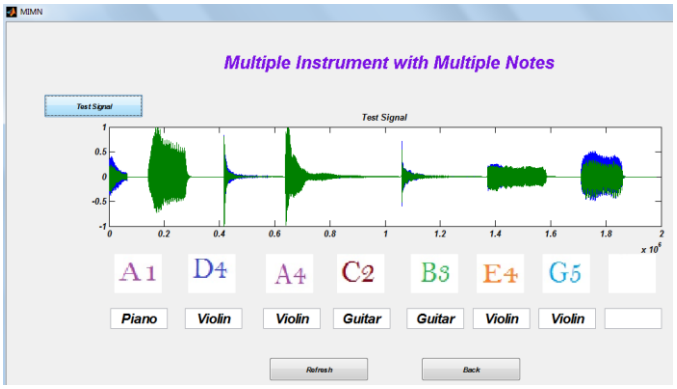


Fig 7: GUI model for Musical instrument transcription using NN for MIMN

4. CONCLUSION

Classification of musical instrument is done using FFNN Classifier. The musical instrument transcription system is developed by combining temporal, Spectral features. The transcription is done on Single Instrument Single Note (SISN), Single Instrument Multiple Note (SIMN) and Multiple Instrument Multiple Note (MIMN). for Piano, Guitar and Violin Musical Instrument. The Transcription rate achieved using all features are for SISN is 91 %, SIMN is 84% and MIMN is 83%

5. REFERENCE

[1] Md. Omar Faruqe, S Ahmad, Md. Al-Mehedi Hasan & Farazul H Bhuiyan), "Template Music Transcription for Different types of Musical Instruments", IEEE, Computer and Automation Engineering (ICCAE), 2010 The 2nd International Conference, Volume 5, (2010), pp 737-742

[2] Mahdi, Triki, Dirk T.M. and Slock, "Perceptually motivated quasi-periodic signal selection for polyphonic music transcription", ICASSP International conference on Acoustics, Speech and Signal Processing, Taipei, Taiwan April 19-24, (2009), pp 305-308.

[3] Matija Marol), "A Connectionist Approach to Automatic Transcription of Polyphonic Piano Music", IEEE Conference, (2004).

[4] Signalskenichi miyamoto, hirokazu kameoka, haruto takeda, takuya Nishimoto and Shigeki Sagayama, "Probabilistic approach to automatic music transcription from audio", IEEE Conference, (2007).

[5] A. Zils and F. Pachet. "Automatic extraction of music descriptors from acoustic signals using EDS." In *Proc. AES*, 2004.

[6] E. J. Humphrey, T. Cho, and J. P. Bello, Learning a Robust Tonnetz-space Representation for Chord Recognition," In *Proc.ICASSP*, 2011.

[7] T. Cho and J. P. Bello, "A Feature Smoothing Method For Chord Recognition Using Recurrence Plots." In *Proc. ISMIR*, 2011

[8] A. Sheh and D. P. W. Ellis, "Chord segmentation and recognition using EM-trained Hidden Markov Models," In *Proc.*

[9] *ISMIR*, 2003.

[10] T. Cho, R. J. Weiss and J. P. Bello, "Exploring Common Variations in State of the Art Chord Recognition Systems." In *Proc. SMC*, 2010.

[11] M. Mauch and S. Dixon, "Approximate Note Transcription For The Improved Identification Of Difficult Chords." In *Proc. ISMIR*, 2010.