Extracting Melodic Pattern of 'Mohan Veena' from Polyphonic Audio Signal of North Indian Classical Music

Ram K. Nawasalkar
Dept. of Comp. Sci.,
Arts, Comm. and Sci. College,
Kiran Nagar, Amravati (M. S.)

Nilesh M. Shingnapure
Dept. of Comp. Sci. and Mgt.,
New Arts, Comm. and Sci.
College, Wardha (M. S.)

Pradeep K. Butey
Dept. of Computer Science,
Kamla Nehru Mahavidyalaya
Nagpur (M. S.)

ABSTRACT

There exist number of musical instruments in the world; each has its own melody and recognition techniques. The purpose of this paper is to propose a hypothesis for extraction of melodic pattern from the polyphonic audio recording of North Indian Classical Music files related to Mohan Veena. The paper discusses about the useful algorithm, application and techniques for extraction and evaluation. The five stages procedure: first, stream separation by Non-negative Matrix Factorization (NMF). Second, instrument recognition – extracting stream related to Mohan Veena using spectrogram and autocorrelation technique. Third, extracting methods which comprises of Sinusoid extraction, system analyzes the audio signal and extract spectral peaks for constructing the salience function where the spectral peaks could be used to compute representation of pitch salience over time, followed by pitch counter technique. Fourth, melody extraction from the extracted pitches counter and finally about the evaluation technique with some discussion.

Keywords

polyphonic audio, Mohan Veena, Non-negative Matrix Factorization, spectrogram, autocorrelation sinusoid extraction, salience function, pitch counter.

1. INTRODUCTION

Indian string instruments have undergone many changes throughout history. The Mohan Veena is a highly modified concord archtop1. This work is to propose a hypothesis for the extraction of melodic pitch pattern from a recorded polyphonic audio signal of Indain classical music. The work will mainly concern with instrumental music played over Mohan Veena with supporting drone instrument Tabla.

Now, what the melody is, the term melody is a musicological concept based on the judgement of human listeners [1], and we can expect to find different definition for the melody in different contexts [2] [3]. The polyphonic we refer to music in which two or more notes can sound simultaneously, be it different instruments or a single instrument capable of playing more than one note at a time [4].

The task of melody extraction involves automatically extracting a representation of the melodic line. For the reason, need for proper means of evaluation and comparing the performances of proper algorithm also come over. This initiative evolved into the Music Information Retrieval Evaluation eXchange (MIREX) [5]. The melody representation proposed by M. Goto, a sequence of fundamental frequency (F0) values that will be corresponding to the perceived pitch of the main melody [6]. To date there are various methods and systems for automatic melody extraction from polyphonic music, have been proposed to the MIREX automatic melody extraction campaign.

The complexity of the task is threefold, firstly, the polyphonic music contains the superposition of all instrument which play simultaneously thus, will be hard to attribute specific frequency bands and energy levels to specific instrument notes, further more will be in the mixing and mastering techniques.

Secondly, automatic instrument recognition, usually based on timbre-spectral models or features such as pitch, spectral centroid, energy ratios, spectral envelopes, Mel Frequency Cepstral Coefficients (MFCC) or MPEG-7 combined with statistical classifier [7][8][9]. Temporal features other than attack, duration and tremolo, are seldom are taken into account. Classification is done using k-NN classifiers, HMM, Kohonen SOM and Neural Networks[10][11]. A limitation of such methods is that in real instruments the spectral features of the sound are never constant, even when the same note is being played, the spectral components changes [12].

Thirdly, during the task of determining which pitches constitute the main melody needs to be overcome [13]. This will in turn entails three more main challenges, determining the melody is present and if not then ensuring the estimated pitches should be in the correct octave. Finally selecting the correct melody pitch when there more than one note sounding simultaneously [4].

In this paper, we describe a basic processing structure - underlying melody extraction systems, comprising three main step – multi-pitch extraction, melody identification and post processing. Whereas, alternative designs have been proposed [14], which is the predominant architecture in most current system [15, 16, 17, 18]. In section 1, imposing Non-negative Matrix Factorization (NMF) algorithm for separation of multiple audio stream[19] of Mohan Veena (prime instrument) and Tabla (drone instrument). In section 2, using the concept of recognition and extraction the audio signal of Mohan Veena will be extracted using spectrogram and autocorrelation method for instrument recognition, which will be the input to the next section of melody extraction. In section 3, three main stages of state-of-the-art melody extraction technique will be applied i.e., the extraction of the sinusoidal components, computation of a time–pitch salience function [20], and followed by pitch counter technique. In section 4, discuss about the melody selection, in which the process comprises of three steps: voicing detection, octave

[1] **http://www.vishwamohanbhatt.com/veena.htm**

error minimization/ pitch outlier removal and final melody selection. In section 5, we will discuss about evaluation method for an extensive and varied set of testing material with alternative approaches too. Finally in section 5, we conclude the paper for discussion with some proposition for future work.

2. METHOD

2.1. Non-negative matrix Factorization

The first step for moving forward would be implementation of non-negative matrix factorization (NMF) algorithm to produce separated audio stream. Non-negative Matrix Factorization, first proposed by Lee and Seung [21] is a data-adaptive linear representation method.

2.1.1. NMF for music

Initially, this method was developed for image signal processing since a 2-D image can be regarded as non-negative matrix but, the time domain audio signal are not suited for this method since they include both positive and negative values. However, the magnitude of the spectrogram meets the non-negative requirement. Smaragdis [22] showed that the components in basis matrix *W* can be individual notes and proposed an approach of polyphonic music transcription using NMF, if there are multiple instruments presented we can separate them by doing separation among those notes.

2.1.2. Decomposition of audio signal

The Non-negative Matrix Factorization algorithm [23] decomposes the input signal i.e., matrix $V \in \Re^{\geq 0, M \times N}$ (where, $\Re^{\geq 0, M \times N}$ is an *M* by *N* non-negative real value matrix) into the product of two non-negative matrices: a basis matrix W $\in \Re^{\geq 0, M \times R}$ and coefficient matrix H $\in \Re^{\geq 0, R \times N}$, in time-frequency domain. Generate time-frequency masks by comparing the energies of decomposed bases and apply those masks to the spectrogram. A spectrogram $S(t,\omega)$ is calculated by dividing the time domain signal $s(\tau)$ into small frames and performing Discrete Fourier Transform (DFT) on each frame.

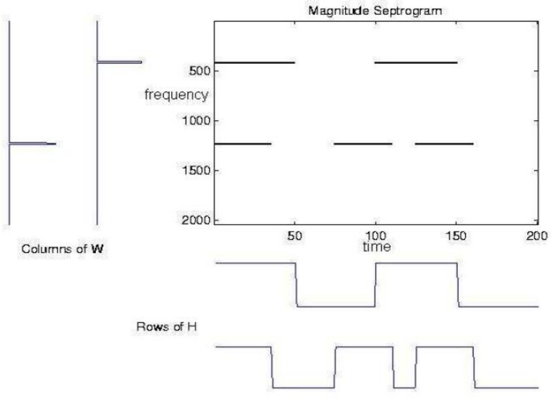

Figure 1: Decomposition of a simple spectrogram using NMF

The above figure [19] shows the result of decomposition of the spectrogram. The bases *W* matrix, are features in frequency domain which can be notes in a certain situation, and the *H* matrix records their locations along time.

2.1.3. Separation

In this part of section we can easily separate the two different frequencies by multiplying the rows of *W* with the corresponding columns of *H*, that is $w_r h_r$. Finally, grouping of bases is made in the domain to produce separated audio streams[19]. As shown in the figure 2 [19], the separated objects are the magnitudes of the spectrograms, since the whole processing are constrained to be non-negative.

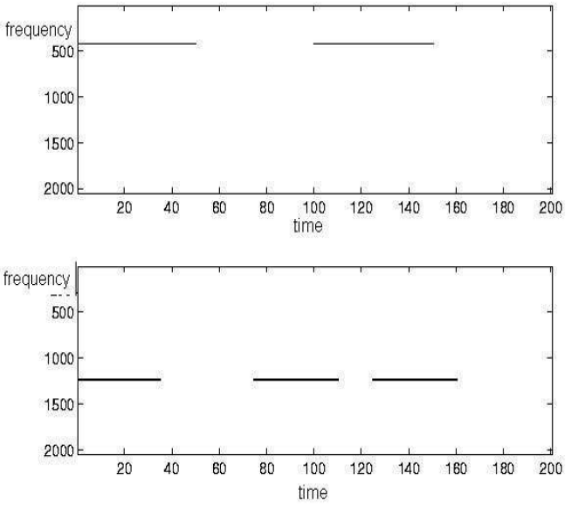

Figure 2: Separation of the spectrogram

2.2. Instrument Recognition

Musical instrument recognition can be carried out using spectrogram and autocorrelation which yield the high recognition rate [12]. The technique uses the solo recordings of a Mohan Veena and Tabla instrument. A large database can be used in order to encompass the different sound possibilities of each instrument an evaluate generalization abilities of the classification process. The basic characteristics are computed in 1 sec. interval. The resulted characteristics will be compared with every separated stream from section 1 so as to get the required stream of Mohan Veena.

2.3. Melody Pattern Recognition

This is a novel system for the automatic extraction of the main melody from polyphonic music recordings. The separated audio signal of Mohan Veena will be used as the input for the melody extraction. The approach is based on the four main blocks as depicted in figure [27].

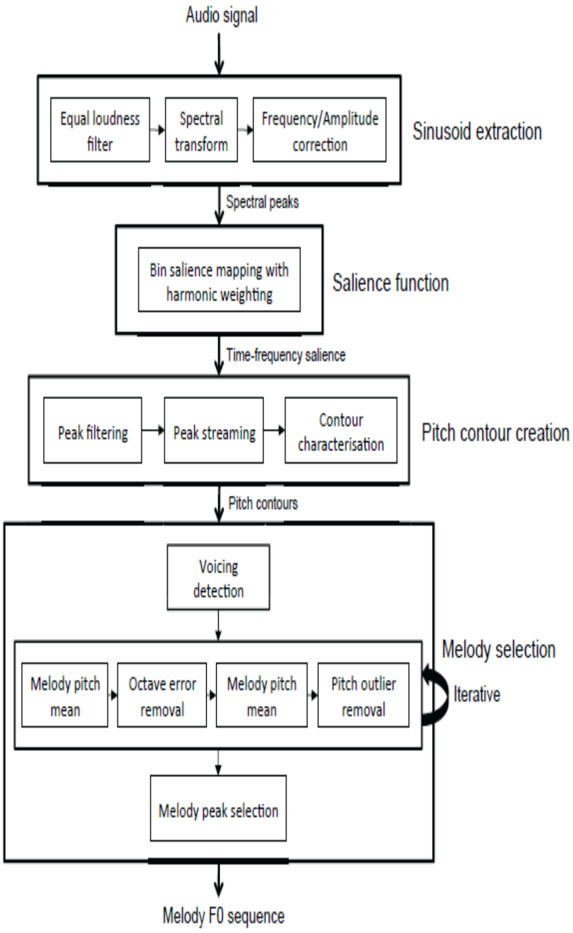

Figure 3: Block diagram of the system's four main blocks: sinusoid extraction, salience function computation, pitch contour creation and melody selection.

2.3.1.Sinosoid Extraction

In the first step, obtaining the spectral peak from the audio signal, referred to as the *front end* [6]. System will analyze the audio signal and extract spectral peaks (sinusoids). For spectral peaks there would be two common goals, Firstly, they should be as accurately as possible in terms of their frequency and amplitude. Secondly, enhance the amplitude of melody peaks whilst suppressing that of background peaks by applying some pre-filtering. For the purpose of the process can be divided into three main steps: pre-filtering, transform and frequency amplitude correction.

In the pre filtering stage apply the time-domain equal loudness filter which was shown in [20] to attenuate spectral components belonging primarily to non-melody sources. Next, we apply a spectral transform and select the peaks of the magnitude spectrum for further processing. Further, multi-resolution FFT (MRFFT) technique [19] can be used for combining spectral peaks from windows with of varying lengths. In this way, we will be able to asses the difference between a single and multi-resolution transform which will indeed significant for melody extraction. In the third step the frequency and amplitude of the selected peaks can be re-estimated by calculating the peaks' instantaneous frequency (IF) using the phase vocoder method [24][15].

2.3.2.Salience Function

Next, the spectral peaks could be used to compute representation of pitch salience over time, a salience function. Our salience function will based on harmonic summation with magnitude weighting, and spans a range of almost five octaves. Further, the parameters of the salience function were optimized for melody extraction by evaluating it directly using metrics designed to estimate the predominance of the true melody F0 compared to peaks in the salience function caused by other sources. In the results section we will examine how this optimization affects the overall performance of the complete system [25].

2.3.3. Pitch Counter:

In this part of block, the peaks of the salience function will be grouped over time using heuristics based on auditory streaming cues [26] which will be the input to the next section of melody extraction.

2.4. Melody Selection

These will be the output results of pitch counter section, contain set of pitch contours, out of which the contours belonging to the melody need would be selected. These contours will automatically analyze and a set of contour characteristics will be computed. In the final block of the system, the contour characteristics and their distributions would be used to filter out non-melody contours. First, remove contours whose features suggest that there is no melody present in this segment of the piece (voicing detection). The remaining contours would be used to iteratively calculate an overall melody pitch trajectory, which will minimize octave errors and remove pitch outliers. Finally, contour salience features can be used to select the melody F0 at each frame from the remaining contours [27].

2.5. Melody Extraction Evaluation

This procedure would be used for evaluating extracted melody.

2.5.1.Ground Truth Annotation:

The ground truth for each audio excerpt would be generated using the following procedure: first, the annotator must acquire the audio track containing just the melody of the excerpt. This can be done by using multi-track recordings for which the separate tracks are available. Given the melody track, the pitch of the melody would be estimated using a monophonic pitch tracker with a graphical user interface such as SMSTools[1] or WaveSurfer[2], producing an estimate of the fundamental frequency (F0) of the melody in every frame. This annotation can then manually inspected and corrected in cases of octave errors (double or half frequency) or when pitch get detected in frames where the melody is absent called as unvoiced frames. Finally, the estimated frequency sequence can be saved into a file with two columns - the first containing the time stamp of every frame, starting from time 0, and the second the value of the fundamental frequency in Hertz. Frames in which there is no melody present are labeled with 0 Hz.[10].

3. CONCLUSION

In proposed framework we come to the conclusion that, we can divide the polyphonic audio signal by using non negative

[1] **http://mtg.upf.edu/technologies/sms**
[2] **http://www.speech.kth.se/wavesurfer/**

matrix factorization. Further, using instrument recognition technique, method of spectrogram and autocorrelation for instrument characteristics we can recognize the instrument as well as extract the required instrument stream. The extracted stream after passing to melody pattern recognition framework we can obtain the required melody pattern for Mohan Veena successfully. Thus the framework can be very much effective for melody detection and extraction.

4. DISCUSSION

- We can analyze the variation of melody in *khayal* as well as *drut bandish* in Indian classical music.
- We can also analyze melody variation in different *thaat* present in North Indian classical music.
- We can analyze the melody of same raga in different Gharana's present in North Indian classical music.
- The salience function is vary in Khayal and drut bandish. Salience intervals are more in khayal but there intervals are less in drut bandish.
- In the current work we have proposed NMF algorithm which can be used for segregation of polyphonic music having only two musical instruments but for complex type of polyphonic music we can use LRMS separation method.
- We can recognize the instrument automatically after the stream separation from polyphonic music manually rather than technique we proposed in current system.

5. REFERENCES

[1] G. E. Poliner, D. P. W. Ellis, F. Ehmann, E. G´omez, S. Steich, and B. Ong, "Melody transcription from music audio: Approaches and evaluation". *IEEE TASLP*, vol. 15, pp. 4, 2007.

[2] R. Typke, "Music retrieval based on music similarity", *Ph. D. dissertation*, Utrecht University, Netherland 2007.

[3] A. S. Bregman, "Auditory Scene Analysis", The Perceptual Organization of Sound, *MIT Press, Fourth Edition*, 2001.

[4] J. Salamon and E. G´omez, "Melody extraction from polyphonic music signals using pitch contour characteristics". *IEEE TASLP*, vol. 20, pp. 6, 2012.

[5] J. Salamon and J. Urbano, "Current challenges in the evaluation of predominant melody extraction algorithms".

[6] M. Goto, "A real-time music-scene-description system: predominant *f0* estimation for detecting melody and base line in real world audio signals", *Speech Communication*, vol.43, pp.311-329, 2004.

[7] A. Eronen, A. Klapuri, "Musical Instrument Recognition Using Cepstral Coefficients and Temporal Features", *Proc. Of the IEEE International Conference on Acoustics, Speech and Signal Processing*, ICASSP 2000, pp. 753-756.

[8] T. Kitahara, M. Goto, H. Okuno, "Musical Instrument Identification Based on F0-Dependent Multivariate Normal Distribution", *Proc. Of the IEEE International Conference on Acoustics, Speech and Signal Processing*, ICASSP 2003, vol. V, pp. 421-424, April 2003.

[9] J. J. Bosch, J. Janer, F. Fuhrmann and P. Herrera, "A comparison of sound segregation techniques for predominant instrument recognition in musical audio signals", *13th International Society for Music Information Retrieval Conference (ISMIR 2012).*

[10] A. Eronen, "Musical instrument recognition using ICA-based transform of features and discriminatively trained HMMs", *Proc. of the Seventh International Symposium on signal Processing and its Applications*, ISSPA 2003, Paris, France, 1-4 July 2003, pp. 133-136.

[11] G. De Poli, P. Prandoni, " Sonological Models for Timbre Characterization", *Journal of New Music Research*, vol. 26 (1997), pp. 170-197, 1997.

[12] Sumit Kumar Banchhor and Arif Khan, "Musical Instrument Recognition using Spectrogram and Autocorrelation", International Journal of Soft Computing and Engineering, ISSN: 2231-2307, vol. 2, Issue-1, March 2012.

[13] J. Salamon and E. Gomez, "Melody extraction from polyphonic music signals using pitch counter characteristics", *IEEE Transaction on Audio, Speech and Language Processing*. 2010.

[14] Jean-Louis Durrieu, Gaël Richard, Bertrand David, and Cédric Févotte, "Source/filter model for unsupervised main melody extraction from polyphonic audio signals," *Trans. Audio, Speech and Lang. Proc.*, vol. 18, pp. 564 575, 2010.

[15] M. Ryynanen and A. Klapuri, "Automatic transcription of melody, bass line, and chords in polyphonic music," *Computer Music Journal*, vol. 32, no. 3, pp. 72–86, 2008.

[16] P. Cancela, "Tracking melody in polyphonic audio", *in 4th Music Information Retrieval Evaluation eXchange (MIREX)*, 2008.

[17] Karin Dressler, "Audio melody extraction for mirex 2009," *in 5th Music Information Retrieval Evaluation eXchange (MIREX)*, 2009.

[18] J. Salamon and E. Gómez, "Melody extraction from polyphonic music audio," *in 6th Music Information Retrieval Evaluation eXchange (MIREX)*, extended abstract, 2010.

[19] B. Wang and M. D. Plumbley, "Musical audio separation by Non-negative matrix Factorization".

[20] J. Salamon, E. Gomez and J. Bonada, "Sinusoid extraction and salience function design for predominant melody estimation", *Proc. Of the 14th Int. Conference on Digital Audio Effects (DAFx-11)*, Paris, France, September 19-23, 2011.

[21] D. D. Lee and H. S. Seung, "Learning the parts of objects by non-negative matrix factorization", *Nature*, 401:788–791, 1999.

[22] P. Smaragdis and J. C. Brown, "Non negative matrix factorization for polyphonic music transcription". In *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA'03)*, pp. 177–180, October 2003.

[23] Equal loudness filter, July 2011.

[24] K. Dressler, "Sinusoidal Extraction using an Efficient Implementation of a Multi-resolution FFT". *In Proc. of the Int. Conf. on Digital Audio Effects (DAFx-06)*, pages 247–252, Montreal, Quebec, Canada, Sept. 2006.

[25] J. L. Flanagan and R. M. Golden,"Phase vocoder", *Bell Systems Technical Journal*, 45:1493–1509, 1966.

[26] A. Bregman. Auditory scene analysis. MIT Press, Cambridge, Massachussetts, 1990.

[27] J. Salamon and E. G´omez, "Melody extraction from polyphonic music signals using pitch contour characteristics", *IEEE Transactions on Audio Speech, and Language Processing*, In Press (2012).