Volume 5, Issue 6, June 2015 ISSN: 2277 128X
International Journal of Advanced Research in
Computer Science and Software Engineering
Research Paper
Available online at: www.ijarcsse.com

Melodic Content Extraction from Polyphonic Music Using STFT, Fanchirp and Constant Q-Transform

Sridevi S. H.*, Prof. S. R. Gulhane, Prof. Vaishali V. Thorat
E&TC, Savitribai Phule Pune University
Maharashtra, India

Abstract— Increase in the demand of music has lead to research in its field. Music is an art form whose medium is sound. Music was the first mass industry which made great advancement in the digital industry. Music has several attributes like timber, rhythm, melody, raga harmony, form etc. In this paper a novel approach has been made to identify the melodic content present in the given polyphonic music signal. The melodic content is extracted by making using of different transforms like STFT,Fanchirp transform and constant Q-transform.

Keywords— STFT, FanChirp, Melody, CQ transform, Polyphony, Pitch

I. INTRODUCTION

Music is an art that contains various attributes. A listener can extract various information while listening to a polyphonic music. It may be beats of the music, melody of the music, rhythm, raga etc. Melody extraction from polyphonic music signal aims at extracting the notes that represent the main melodic content. The term melody has different definitions in different context as given by different authors and researchers. Over the last decade years there has been a remarkable progress in the area of transcription of polyphonic music - a topic that is closely related to melody extraction. The pitch sequence is usually manifested as the fundamental frequency (F0) contour of the singing voice in the polyphonic mixture fields like music retrieval, singer identification , cover song identification, query by humming etc. In melody extraction, the main job is to find the vocal melody.

II. RELATED WORKS ON MELODY EXTRACTION

Audio signal can be converted into a description of all the notes being played is a task that can usually be achieved by a trained student of music and has long been a topic of computational research. But this task is proved to be very difficult due to the complexity and overlapping of spectrum of the music signal. A comprehensive review of the early approaches to pitch detection in speech signals is provided by Hess by in 1983 and a comparative evaluation of pitch detection algorithms in speech signals is provided in 1976. A more recent review of previous approaches to pitch detection in speech and music signals is provided in Hess, 2004. The recent researches states that pitch detection or tracking for monophonic signals (speech or music) is practically solved. But the latest publications show some promising results in the transcription of polyphonic audio pieces with some constraints like use of special instrument or it may be on the limitation on the number of instruments being used in the polyphonic music signal that is considered. Main attention has been received from the publications of Masataka Goto who introduced his Predominant F0 estimation algorithm (Goto and Hayamizu, 1999; Goto, 2000). Goto was the first person to achieve an extraction of the melody and bass line from real world CD recordings.

TABLE 1. Related works on melody extraction

System	Front end	No. of Pitches	Voicing
Dressler[6]	|STFT|+sines	5	Melody + local threshold
Marolt [23]	|STFT|+sines	> 2	Melody grouping
Goto [14]	Hier.|STFT|+sines	> 2	Continuous
Poliner[27]	|STFT|	1	Global Threshold

III. PROPOSED METHODOLY

The proposed method consists of three steps where in the pitch content is extracted first followed by the formation of pitch contours and singing voice identification wherein the fundamental frequency is estimated.

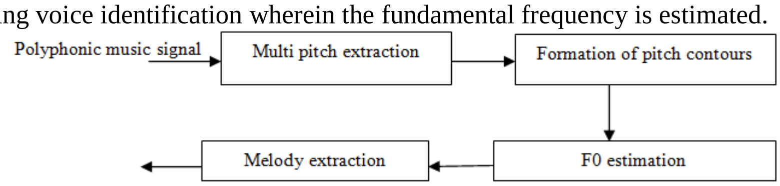

Fig 1 Block diagram of the proposed methodology

A) Pitch detection

For a given frame of polyphonic music signal Short Time Fourier Transform (STFT) is applied. The STFT is the standard method used for the time-frequency analysis. This transforms holds good under the assumption that the signal is stationary in the analysis frame. It is used to determine the sinusoidal frequency and phase content of local sections of a signal as it changes over time. So, a time-dependent Fourier transform that changes periodically as the polyphonic music signal's properties change over the time can be adopted. Short Time Fourier Transform performs FFT analysis on short windows in time. The results of the FFT represent the contents of the audio signal in terms of time-frequency information. The window used in STFT controls the trade-off between frequency resolution and side-lobe suppression. Some melody extraction systems use a multi-resolution transform instead of the STFT which has a fixed time-frequency resolution.

B) Detection of spectral filters

After spectrum transform, the signal is now transformed from temporal domain to the spectral domain. The spectral peaks that have originated either from vocal or instrumental accompaniment signals or the noise are used to calculate the pitch salience. Usually, the number of original peaks is large because of the signals that have accompanied. When it comes to a vocal frame, there exist some peaks with salient magnitude. It is prominent as they are the candidate pitches. But there is the possibility that these pitches may be the pitches of the instruments used in the polyphonic music signal. On the other hand, the number is much larger for a silent frame and at the same time the peaks magnitude is smaller compared to vocal frame, as the peaks are all originated from noise with low energy. Spectral peaks are often disposed using the highest spectral peaks. Peaks with a magnitude more than 80 dB below the highest spectral peak in a frame are not considered

C) Fan Chirp transform

Wavelet is a piece of a wave, and a chirplet, similarly, is a piece of a chirp. In other words a chirplet can be defined as a windowed portion of a chirp function, where the window provides some time localization property. The Fan Chirp transform provides an insight representation of harmonically related linear chirp signals. It can be considered as time wrapping followed by a Fourier transform In this paper Fan chirp transform is applied to the analysis of pitch content in polyphonic music signal. A F0gram is calculated based on collecting harmonically related peaks of the Fan chirp. The number of valid f0 values in the frame is calculated. Considering a masking function given by the valid pitches a correct estimate of near boundaries are estimated. The f0 parameters are chosen as ;the minimum fundamental frequency to be 80Hertz, the number of octaves to be equal to 4 and the number of f0's per octave is taken to be 192.The 3 most salient f0gram peaks are selected as pitch candidates to form pitch contours are considered as main melody. The Fan Chirp transform provides enhanced spectrograms for capturing pitch or formant modulation information, removing the spectral illusions.

D) Constant Q transform

In general, Constant Q transform (CQ transform) transform is well suited to musical data, and this can be seen in some of its advantages compared to the Fast Fourier Transform. In this transform fewer frequency bins are required to cover a given range effectively, and this proves useful where frequencies span several octaves. As the range of human hearing covers approximately ten octaves from 20 Hz to around 20 kHz, this reduction in output data is significant. The transform exhibits a reduction in frequency resolution with higher frequency bins—which is desirable for auditory applications. At the bottom of the piano scale (about 30 Hz), a difference of 1 semitone is a difference of approximately 1.5 Hz, whereas at the top of the musical scale (about 5 kHz), a difference of 1 semitone is a difference of approximately 200 Hz. Till now for the musical data the exponential frequency resolution of Constant Q is ideal.Moreover the harmonics of musical notes form a pattern characteristic of the timbre of the instrument in this transform. Assuming the same relative strengths of each harmonic, as the fundamental frequency changes, the relative position of these harmonics remains constant. This can make identification of instruments much easier.

IV. RESULTS OBTAINED

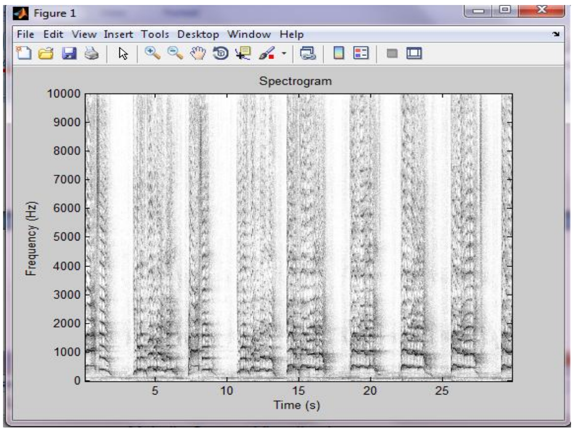

Fig 2 Spectrogram of the polyphonic music audio considered.

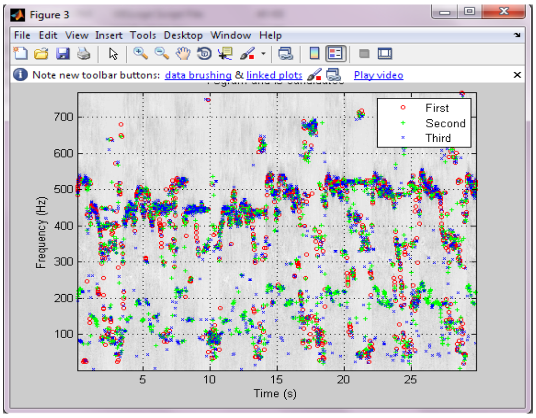

Fig 3 Colour representation of melodic content visulasation

V. CONCLUSION

In this an effort has been made to extract the melody from the polyphonic music signal by using STFT which is mainly used for a non-stationary signal. And music is also a non –stationary signal which varies with time. Hence a combination of STFT, constant Q-transform and Fan chirp transform provides better results. At an abstract level, the benefits of common, standardized evaluation are clearly shown by this effort and analysis.Melody extraction has very great future in music industry and is relatively a young research.

ACKNOWLEDGEMENT

I take this opportunity to express my deep heartfelt gratitude to all those people who have helped me in the successful completion of this paper. First and foremost, I would like to express my sincere gratitude towards my parents who have constantly encouraged me bringing out a paper. I wish to thank my guide Prof .S.R. Gulhane for providing excellent guidance, encouragement. Without his valuable guidance, this work would never have been a successful one. I would like to express my sincere gratitude to our Head of the Department of Electronics & Communication Engineering, Prof. Santosh G. Bari and our P.G. coordinator Prof. Vaishali.Thorat for their guidance and inspiration. I would like to thank our Principal Dr.V.N. Nitnaware for providing all the facilities and a proper environment to work in the college campus.

REFERENCES

[1] Justin Salamon, Emilia Gómez, Daniel P.W. Ellis, and Gaël Richard, “Melody extraction from polyphonic music signal”, IEEE signal processing magazine, March 2014, date of publication ,Feburary 12.

[2] Pablo Cancela , Ernesto López, Martín Rocamora, “Fan chirp transform for music representation”, Proc. of the 13th Int. Conference on Digital Audio Effects (DAFx-10), Graz, Austria , September 6-10, 2010.

[3] Olivier Gillet Sand Gaël Richard,” Transcription and Separation of Drum Signals From Polyphonic Music”, IEEE transactions on audio, speech, and language processing, vol. 16, no. 3, march 2008.

[4] G. Poliner, D. Ellis, A. Ehmann, E. G´omez, S. Streich, and B. Ong. Melody transcription from music audio: Approaches and evaluation. IEEE Tr. Audio, Speech, Lang. Proc., 14(4):1247–1256, May 2007.

[5] Gael Richard, “Melody Extraction from Polyphonic Music Signals”, International Workshop on Acoustic Signal Enhancement (IWAENC 2014) Sept. 11th, 2014.

[6] Jean-Louis Durrieu, Gaël Richard, Bertrand David, and Cédric Févotte, “Source/Filter Model for Unsupervised Main Melody Extraction From Polyphonic Audio Signals”, IEEE transactions on audio, speech, and language processing, vol. 18, no. 3, march 2010.

[7] Salamon, J.; Gomez, E, “Melody Extraction From Polyphonic Music Signals Using Pitch Contour Characteristics," Audio, Speech, and Language Processing, IEEE Transactions on, vol.20, no.6, pp.1759-1770, Aug. 2012.

