International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 4 Issue 06, June-2015

Melody Extraction from Polyphonic Music Signal using STFT and Fanchirp Transform

Sridevi S. H.
Department of E&TC
DYPSOEA
Ambi, Talegaon, Pune,India

Prof. S. R. Gulhane
Department of E&TC
DYPCOE
Ambi, Talegaon,Pune,India

***Abstract*— Music is an art form whose medium is sound. It includes various attributes like rhythm, melody, timber etc. The term melody is a musicological concept based on the judgment of human listener's .Melody extraction from polyphonic music is a difficult task in music information retrieval. In melody identification stage, the main job is to find the vocal melody. In a polyphonic music 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. The main aim of melody extraction is to produce a sequence of frequency values corresponding to the pitch of the dominant melody present in the given musical recording. In this paper the melody is extracted from polyphonic music signal by using Short Time Fourier Transform and Fan Chirp Transform.**

Keywords—MIR, STFT, Fan chirp Transform(FCht), Melody, multiptch, Poyphonic,FFT

I.INTRODUCTION

The development in the field of music information retrieval (MIR) has created a need for indexing systems that automatically extract semantic descriptions from music signals. This description would typically include melodic, tonal, timbral, and rhythmic information. So far, the scientific community has mostly focused on the extraction of melodic and tonal information (multipitch estimation, melody transcription, chords, and tonality recognition) but also to a lesser extent on the estimation of the main rhythmic structure. Most of the time the concept of melody is associated to a sequence of pitch notes. This definition can be found: "A combination of a pitch series and a rhythm having a clearly defined shape" (Solomon 1996), and on Grove Music: "Pitched sounds arranged in musical time in accordance with given cultural conventions and constraints" (Grove's online 2002). Multiple fundamental frequency (f0) estimation is one of the most important problems in music signal analysis and constitutes a fundamental step in several applications such as melody extraction.In this paper an effort has been made to extract melody from the polyphonic music signal using Fan chirp transform.

II. RELATED WORKS ON MELODY EXTRACTION.

Pitch detection algorithms (PDAs) in audio signal processing, especially in speech processing, have been an active topic of research since the late twentieth century. A comprehensive review of the early approaches to pitch detection in speech signals is provided in (Hess, 1983) and a comparative evaluation of pitch detection algorithms in speech signals is provided in (Rabiner, Cheng, Rosenberg, & Mc Gonegal, 1976). A more recent review of previous approaches to pitch detection in speech and music signals is provided in (Hess, 2004). The general recent consensus is that pitch detection or tracking for monophonic signals (speech or music) is practically a solved problem and most state-of-the-art approaches yield high quality and acceptable solutions (Hess, 2004; Klapuri, 2004). The problem of melody extraction from polyphony is different from the monophonic speech pitch detection problem in two major aspects:

1. Multiple sound sources (pitched and unpitched) are usually simultaneously present.
2. The characteristics of the target source (here the singing voice) are a larger pitch range, more dynamic variation, and more expressive content than normal speech.

Table 1. Principle melody transcription algorithms.

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

The second column, "Front end", concerns the initial signal processing applied to input audio to reveal the pitch content. The most popular technique is to take the magnitude of the short-time Fourier transform (STFT) – the Fourier transform of successive, windowed, snippets of the original waveform– denoted |STFT| in the table, and commonly visualized as the spectrogram. |STFT| is invariant to relative or absolute time or phase shifts in the harmonics because the STFT phase is discarded. Frequency resolution of the STFT improves with temporal window length, these systems tend to use long windows, from 46 ms for Dressler, to 128 ms for Poliner. Goto uses a hierarchy of STFTs to achieve a multiresolution Fourier analysis, downsampling his original 16 kHz audio through 4 factor-of-2 stages to have a 512 ms window at his lowest 1 kHz sampling rate.

The final column, "Voicing", considers how, specifically, the systems distinguish between intervals where the melody is present and those where it is silent (gaps between melodies).Goto reports his best pitch estimate at every frame and do not admit gaps. Poliner's basic pitch extraction engine is also continuous, but this is then gated by a separate melody detector; a simple global energy threshold over an appropriate frequency range was reported to work as well as a more complex scheme based on a trained classifier. As discussed above, the selection of notes or fragments in Dressler naturally leads to gaps where no suitable element is selected; Dressler augments this with a local threshold to discount low-energy notes.

III. PROPOSED METHODOLOGY.

3.1 Terminologies of melody extraction

Melody extraction can be termed as

- Audio melody extraction.
- Predominant melody extraction/estimation.
- Predominant fundamental frequency (f0) estimation.

The aim is to obtain a sequence of frequency values representing the pitch of the dominant melodic line.

Melody line tends to have the most predominant harmonic structure in middle and high-frequency regions. The F0 of the most predominant harmonic structure the most predominant F0 corresponding to the melody line within an intentionally limited frequency range of the input sound mixture is estimated.

Polyphonic music signal

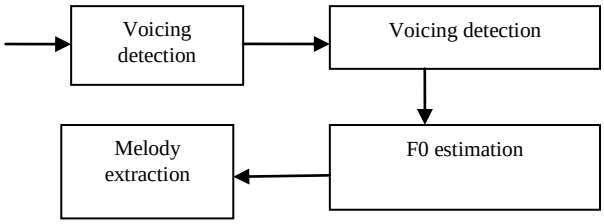

Fig 3.1.1 Block diagram of the proposed method.

The fundamental approach with all the above methods mentioned is that they work with above is that they work under assumptions that tonal energy manifests in the short time spectrum as a distinct peak, allowing a simple detection. Such an assumption hardly holds in case of instruments with free inotation [2].Music is a non-stationary signal by nature. The STFT is the standard method for the time-frequency analysis. This representation holds well under the assumption that the signal is stationary within the analysis frame. Among the chirp based transforms, the Fan-Chirp transform is better suited since it provides time-frequency localization in fan geometry.FCht can be considered as time-wraping followed by a Fourier transform, which leads to efficient implementation using FFT.

3.2 Pitch Extraction

The perceptual counterpart of fundamental frequency is pitch, which is a subjective quality often described as highness or lowness. In the context of melody identification, the problem is to decide which candidate pitches belong to the melody, and to detect whether the melody is present or not at each frame. In this paper FCht is used for the analysis of pitch content in the polyphonic music signal. It allows us to reveal hidden spectral peaks related to non-stationary high frequency partials. Non-salient pitch candidates are filtered out to minimize the creation of contours belong to instrument or noise. The remaining problem is to choose the correct contours which belong to the vocal melody.

3.3 Short Time Fourier Transform (STFT)

Music is not a stationary signal, i.e. it has the properties that vary with time. Thus a single representation based on all the samples of a polyphonic music signal, for the most part, has no meaning. Instead, we define a time-dependent Fourier transform (TDFT or STFT) that changes periodically as the polyphonic music signal's properties change over the time. 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 allows controlling the trade-off between frequency resolution and side-lobe suppression (i.e. how sharp a peak in frequency is versus how high are the side lobes).

3.4 FanChirp Transform(FCht)

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[2]. In this paper FCht is applied to the analysis of pitch content in polyphonic music signal. A F0gram is calculated based on collecting harmonically related peaks of the FCht.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.

IV. RESULT AND DISCUSSION

Melody extraction has become an increasing research topic area.In this paper a novel way of extracting melody from the polyphonic music signal is described. The technique is based on STFT and FCht. The FCht provides salient information about the non stationary signal like music signal. This technique is based on current pitch salience representation called f0gram. The result obtained is shown in Fig 4.1, 4.2, 4.3. Grouping the F0gram peaks into contours involves the determination of where does a contour starts and when does it ends, necessarily leaving some time intervals without melody estimation. This is avoided when isolated F0gram peaks are considered as main melody estimates, since for every melody labeled frame there is always pitch estimation. Therefore, this performance measure can be considered as a best possible reference.

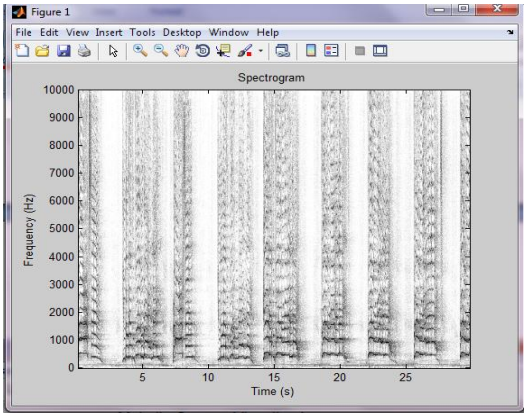

Fig 4.1 Spectrogram of the polyphonic music signal considered.

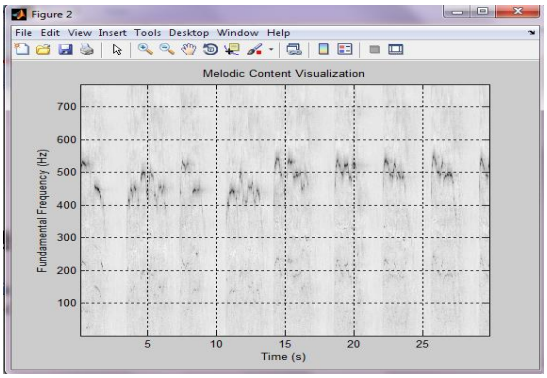

Fig 4.2 Melodic content visulasation

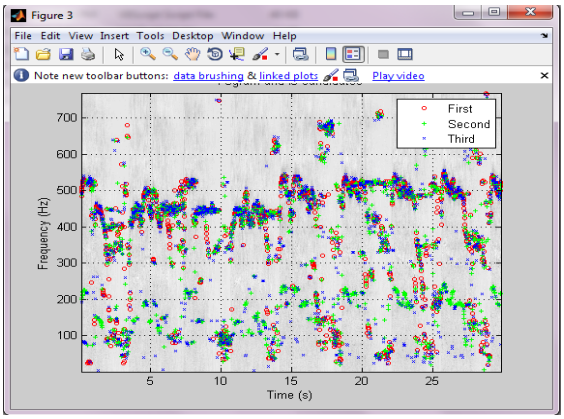

Fig 4.3 Colour representation of melodic content visulasation

V.CONCLUSION

At an abstract level, the benefits of common, standardized evaluation are clearly shown by this effort and analysis. In this paper a system for automatically extracting the main melody of a polyphonic piece of music from its audio signal is described.Melody extraction has many advantages, it can be used in Query by humming, music de-soloing, music retrieval , music classification and so on.

VI. ACKNOWLEDGEMENT

I take this opportunity to express my deep heartfelt gratitude to all those people who have helped me in the successful completion of the paper. First and foremost, I would like to express my sincere gratitude towards 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 for his 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.

VII. 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.