Extraction of Melody from Polyphonic Music using Modified Morlet Wavelet

Article *in* Microprocessors and Microsystems · December 2020

DOI: 10.1016/j.micpro.2020.103612

Prabira kumar Sethy
Sambalpur University

Neeraj Kumar
Indian Institute of Technology (ISM) Dhanbad

Microprocessors and Microsystems 80 (2021) 103612

Contents lists available at ScienceDirect

Microprocessors and Microsystems

journal homepage: www.elsevier.com/locate/micpro

Extraction of melody from polyphonic music using modified morlet wavelet

Neeraj Kumar [a], Raubin Kumar [b], Govind Murmu [a], Prabira Kumar Sethy [c,*]

[a] Department of Electronics Engineering, IIT(ISM) Dhanbad, Jharkhand 826004, India
[b] Talentica Software (1) Pvt. Limited, Pune, India
[c] Department of Electronics, Sambalpur University, Sambalpur, India

ARTICLE INFO

Keywords:
GPE
Melody
Music
MW
MMW
MRFFT
Pitch extraction
STFT

ABSTRACT

In this paper, the extraction of melody for polyphonic music is discussed. The method proposes a simple but effective modification in the Morlet wavelet as a modified Morlet wavelet (MMW). The pitch detection concept is applied to the transformed wavelet music signal for primary melody estimation and extraction. The proposed method is then compared with three other methods reported in the literature (the method focused on Short Time Fourier Transform, multi-resolution Fast Fourier Transform, and Morlet Wavelet). The percentage gross pitch error (GPE) parameter is considered for purposes of comparison. In comparison, it is observed that the proposed approach has the lowest GPE. The simulation results also present a lower percentage error in the predominant pitch frequency estimation for the proposed method.

1. Introduction

An artist can record his or her own version, of the composition of another artist, in popular music. For example, a cover song can retain the original melody but modify the harmonization or maintain the same inherent progression of the chord but devise a new melody. “A promising approach to cover version identification is indeed to independently isolate the main melody, accompaniment and search in both parts [1]”. Apart from its use in detecting cover versions, main melody extraction finds use in a wide variety of music-related applications. Several fundamental components and the accompanying harmonics form a polyphonic music signal. However, the inclusion of multiple harmonic sounds makes it difficult to derive the exact melody from polyphonic music. Y. Chien and others [2] used accompanied singing recordings to remove the vocal melodies. M.G Reddy and K.S. Rao [3] implemented the combined Spectro-temporal dependent process to extract the predominant melody from vocal polyphonic music signals. “A multistage approach follows the problem of melody detection in polyphonic audio, inspired by the concepts of perceptual theory and musical practice presented in [4]”. Pitch is a perceptual property that enables sounds to be organized on a frequency-related scale that ranges from high to low [5]. “Using frequency analysis techniques such as Fast Fourier Transforms (FFT), many of the available methods for music transcription are based on extracting simple frequencies [6-8]”. These strategies are also computationally costly, but in time-critical applications, the FFT remains the tool of choice. Long FFT windows have to be implemented in need of good frequency resolution, accepting a skewed spectrum for faster shifting signal components.

Yoon et al. have suggested a method based on the harmonic structure to extract the predominant melody of polyphonic music [9]. “Zhang et al. pointed out that the extraction of Melody from polyphonic music is necessary as well as a tough challenge in the music information retrieval community [10]”. Salamon et al. suggested that the melody extraction algorithms generate from a musical recording a series of frequency values corresponding to the pitch of the dominant melody [11].

Some researchers propose to use the melody extraction approach based on particle filter and dynamic programming. A melody extraction technique using the efficient computation of STFT spectra at different time-frequency resolutions was proposed by Dressler [12]. A comparison between the STFT and the MRFFT was provided by Salamon et al. [13]. The statistical difference between the transformations used for melody extraction was shown.

A novel transform domain pitch detection method based on wavelet transform is proposed in this paper. Several wavelets have been tested, and it was found that wavelets with a bandpass nature and lower bandwidths on a smaller scale are better for pitch detection. The Morlet wavelet transform meets these criteria and can, therefore, be used for pitch detection. Hence for music transcription, the Morlet wavelet transform method is applied here. It produces an effective result that could not be obtained with the Fourier transformation technique [11].

* Corresponding author.
E-mail address: prabirsethy.05@gmail.com (P.K. Sethy).

https://doi.org/10.1016/j.micpro.2020.103612
Received 15 October 2020; Received in revised form 19 November 2020; Accepted 3 December 2020
Available online 6 December 2020
0141-9331/

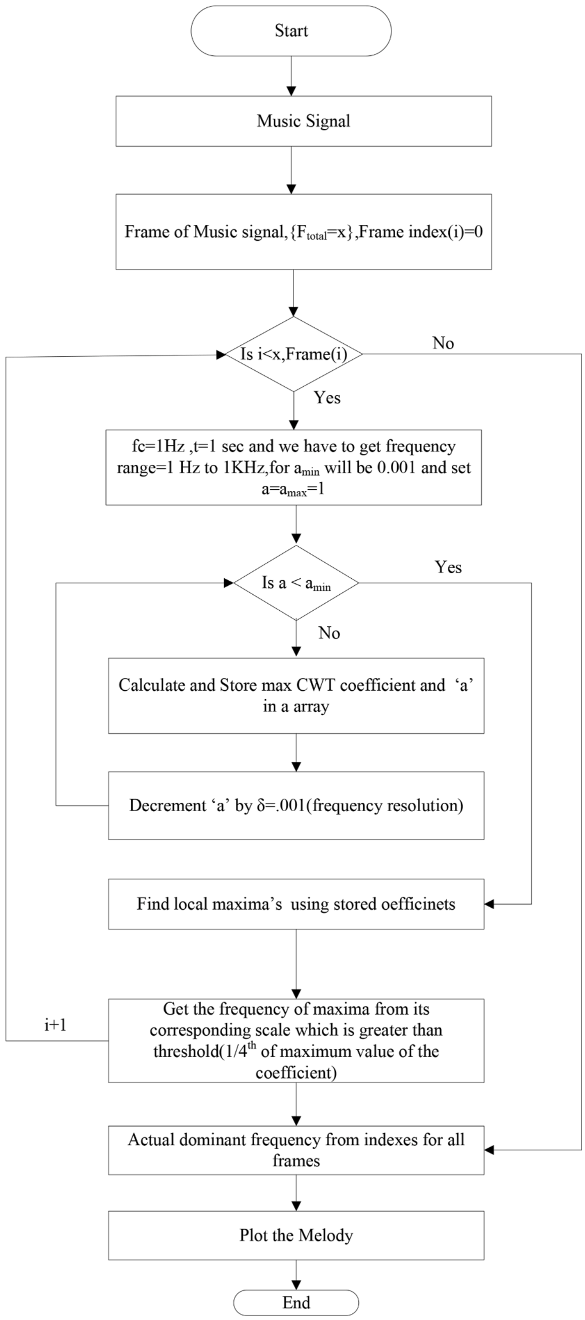

Fig. 1. flowchart of Melody Extraction.

Some improvements were needed in the exponential portion of the Morlet wavelet to improve frequency detection accuracy. In section II, these are discussed in depth. The simulation and outcomes are discussed in Section III, along with the discussions on them. Finally, the work is concluded in Section IV.

2. Melody extraction from polyphonic music using MMW

The following subsection details the methodology and frequency response of MMW with its basics.

2.1. Methodology

The melody extraction flow map from polyphonic music is shown in Fig. 1. The algorithm given in Fig. 1 uses the determination of the highest dominant frequencies from its corresponding scale that is greater than the threshold. The threshold value is chosen as 1/4th of the maximum value of the wavelet transformed data. As we are interested in the dominant frequency, so we have only considered the peaks whose strength is more than 75% of the peak value. The music signal is divided into many frames. One by one, the frames of data are observed and processed. The pseudo frequency has been calculated for optimum performance with $f_c = 1Hz$ and $T = 1\ sec$. So from each frame, the maximum CWT coefficient is calculated and stored. The frequency resolution is chosen as 0.001, and the local maxima are obtained from the stored coefficient. Once local maxima are obtained, only those peaks are considered whose maxima are included from the dominant frequency calculation, which is more than 75% of the peak value. The dominant frequencies and their corresponding index values are calculated and observed. In polyphonic music, a series of such frequency values with their indices, describing the pitch of frames, construct the audio signal melody. The dominant frequency line series is thus called the Melody line or Melody.

2.2. MWM and its frequency response

With a simple start and end time for each note, the Morlet wavelet transform is capable of capturing short bursts of repeated and alternating music notes [14]. The time-domain of wavelet and wavelet magnitude response at scale $a = 0.1$ are presented in Fig. 2. The waveform has a 44.1 kHz sampling frequency.

The general relationship between scale and its frequency is [15]

$$f_a = \frac{f_c}{a \times T} \tag{1}$$

where, f_a is the pseudo frequency related to scale 'a', f_c is the center frequency which is explained below, 'a' is scale and 'T' is the sampling time.

The wavelet function Ψ(t) for Morlet can be expressed as [16].

$$\Psi(t) = \exp\{-(at)^2\}\cos\left(\sqrt{\pi\frac{2}{ln2}}\ t\right)$$

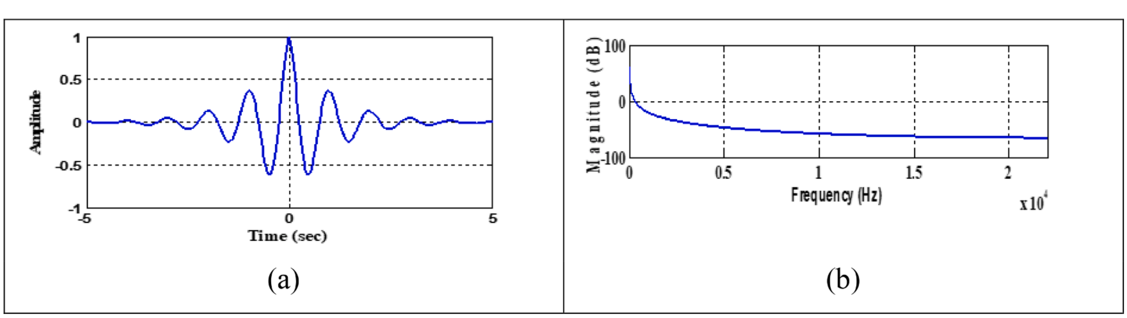

Fig. 2. wavelet and it's response (a) Amplitude in time-domain (b) magnitude response at scale $a = 0.1$ with center frequency 10 Hz.

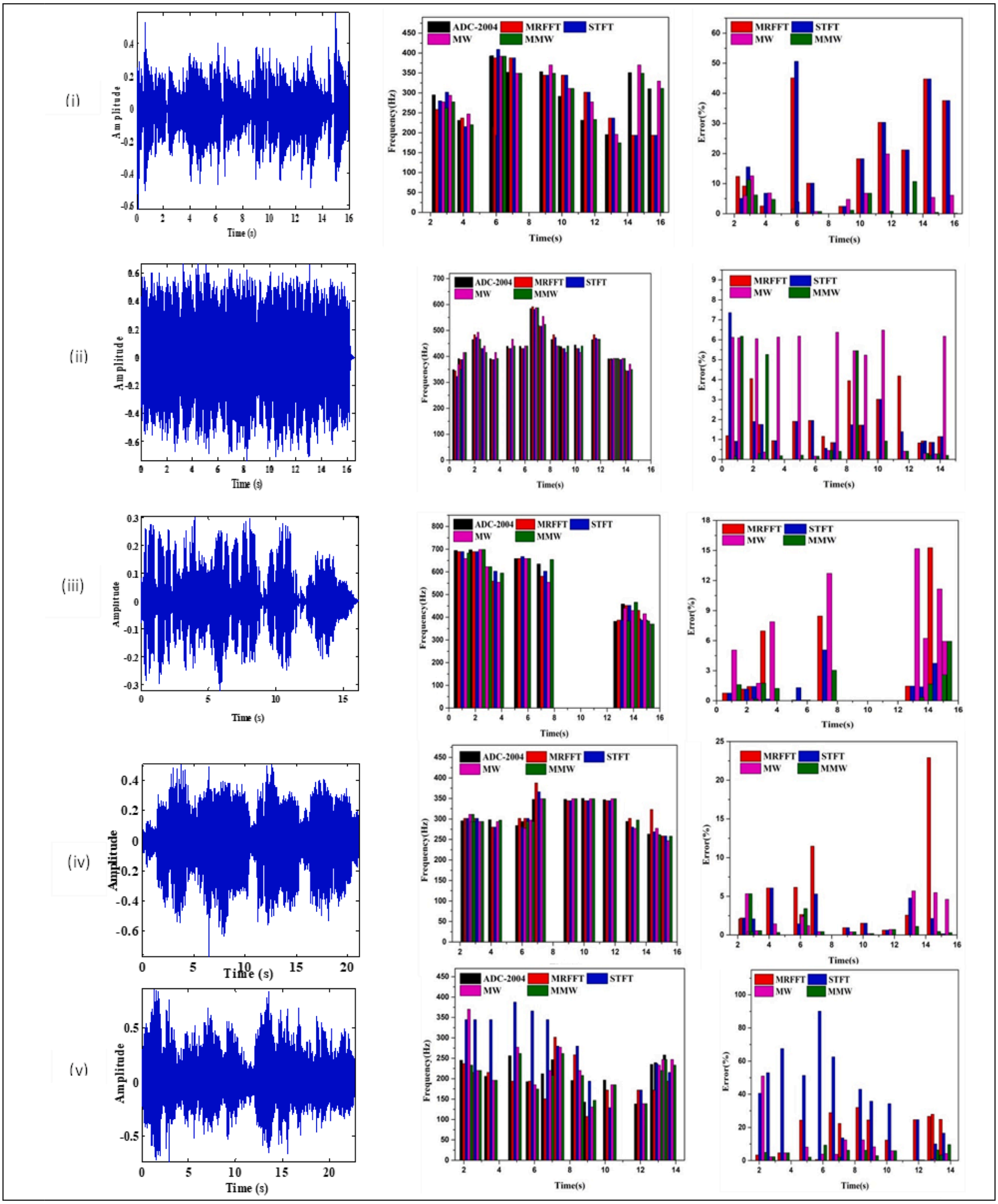

Fig. 3. Simulation results for music signal (i) jazz1 (ii) midi2 (iii) opera_fem2 (iv) daisy1 (v) pop1. (left column: Audio waveform, Middle column: extracted melody frequency, Right column: Average percentage pitch error of melody extraction).

It can be simplified as:

$$\Psi(\mathrm{t}) = \exp\left\{-(at)^2\right\}\cos(2\pi \mathrm{f_i t}) \tag{2}$$

where 'α' is the inverse of scale and $\mathrm{f_i} = \sqrt{\frac{1}{2\ln 2}}$ is a constant. The Fourier transform of the wavelet function (exponential part of the wavelet) is,

$$\mathrm{F}\left[e^{-(\alpha t)^2}\right] = \frac{1}{\alpha\sqrt{2}}\exp\left\{-(f/2\alpha)^2\right\} \tag{3}$$

$$F[\cos(2\pi \mathrm{f_i t})] = \sqrt{\frac{\pi}{2}}[\delta(f-2\pi f_i) + \delta(f+2\pi f_i)] \tag{4}$$

Therefore, Fourier transform of the wavelet function $\psi(\mathrm{t})$,

$$\mathrm{F}[\psi(\mathrm{t})] = \frac{1}{2\alpha}\left[\exp\left\{-\frac{(f+2\pi \mathrm{f_i})^2}{4\alpha^2}\right\} + \exp\left\{-\frac{(f-2\pi \mathrm{f_i})^2}{4\alpha^2}\right\}\right] \tag{5}$$

Thus, in the frequency domain, the Morlet wavelet will appear as a Gaussian at different frequency points for different scales. However, on a

Table 1
Average gross pitch error of melody extraction method.

S. No.	Audio File	Average Gross Pitch Error of Melody Detection Method MRFFT	STFT	MW	The proposed method based on MMW
1	jazz1	17%	18%	6%	3%
2	midi2	2%	2%	4%	1%
3	opera_fem2	4%	2%	7%	2%
4	daisy1	6%	2%	2%	1%
5	pop1	18%	37%	8%	5%

particular scale, all the wavelets represent a particular frequency. Again, it is needed to provide wavelets with smaller bandwidth at lower scales for pitch detection applications. Since the time domain signal contracts are lower in scale and contribute to frequency domain expansion, the Morlet wavelet can be used for multiple pitch detection. A modification of the Morlet wavelet is necessary to detect the closest pitch frequency in polyphonic music. The exponential section leading to the shift in the wavelet on each scale is updated here. The adjustment was made in such a fashion that the wavelet could appear as $e^{\frac{-t^2}{a}}$, with scale 'a', with time. The real Morlet wavelet has a greater frequency bandwidth because of the opposite of the scale-square present in its Fourier transform. Due to such a relationship, the closely separated frequency signals are unable to be detected. The new wavelet uses $e^{-|t|}$ to restrict the wavelet into the time domain. Here the period 't' is correlated with the scale 'a'. The proposed wavelet is given as,

$$\Psi(t) = \exp\{-|t|\}cos(2\pi t) \quad (6)$$

Its Fourier domain representation can be given as

$$F[\Psi(t)] = \frac{1}{4\pi^2 f^2 + 1} * [\delta(f - 2\pi) + \delta(f + 2\pi)] \quad (7)$$

Considering a constant 'k' ($k>1$) in the exponential part of the wavelet as

$$\Psi(t) = \exp\{-|t/k|\}cos(2\pi t) \quad (8)$$

Now the Fourier transform of the wavelet Ψ(t) is,

$$\mathrm{F}[\Psi(\mathrm{t})] = \frac{\mathrm{k}}{(2\pi\mathrm{fk})^2 + 1} * [\delta(f - 2\pi) + \delta(f + 2\pi)] \quad (9)$$

Here, '*' denotes the convolution. The time-domain wavelet will expand with the factor 'k', but it leads to the contraction of the frequency domain. When we consider detecting multiple pitches, it is critical because the bandwidth will increase as the scale decreases. Therefore, we need to provide a 'k' factor so that the wavelet can monitor the individual notes present in the signal. However, as we switch from lower notes to higher notes, the frequency separation between two consecutive notes increases steadily. Still, the bandwidth extends at far lower scales with the inclusion of scale square (k^2) in the transformed domain expression and will not track the signals that are closely separated by frequency. Therefore, the value of 'k' should be greater than 1, so it will help to provide a smaller bandwidth at lower scales.

3. Experiments and results

The algorithm for melody extraction from polyphonic music was presented in Fig. 1.

In 2004, the first Audio Description Contest (ADC) was hosted by the Music Technology Group at Universitat Pompeu Fabra in Barcelona, Spain. For the Audio Description Contest (ADC) held in conjunction with the 2004 ISMIR conference in Barcelona by Cano et al., [17]. Audio files used in the simulation are taken from *the ADC2004 database*, which is available online. The results for five excerpts in the genres files (*jazz1.wav, midi2.wav, opera_fem2.wav, daisy1.wav, and pop1.wav*) are presented here. Each audio file is played for about 20 s, and its sampling frequency is 44.1 kHz. The proposed algorithm is used for melody extraction from polyphonic music. The presence of predominant frequency at a particular time frame was detected and extracted. The output of the algorithm for five excerpts in the genres (i) jazz1, (ii) midi2, (iii) opera_fem2, (iv) daisy1, (v) pop1 is shown in Fig. 3. Each plot has three columns: Left column is a waveform, middle column display: (frequency vs time) and the right column display the percentage error of the melody extraction for different methods (MRFFT, STFT, MW, and MMW). Table 1 shows that the pitch detection error remains less than 3% that satisfies the acceptable pitch detection error up to 5% [18]. The calculation of Gross Pitch Error (GPE) using the following formula

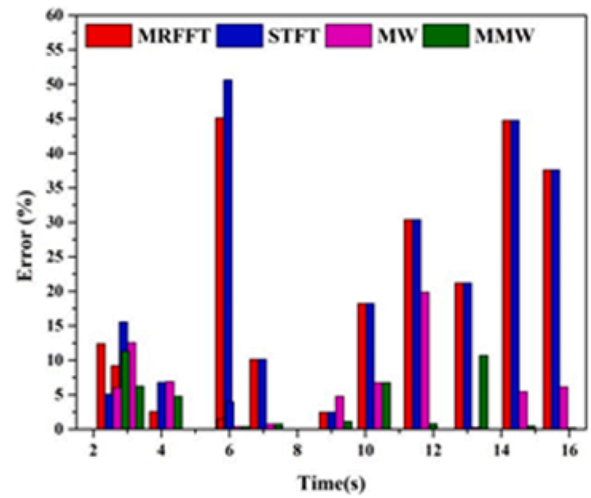

Fig. 4. Average gross pitch error of melody extraction method.

$$GPE = \frac{(P_{Ground\ truth} - P_{estimate})}{P_{ground\ truth}} \quad (10)$$

The middle column of Fig. 3 shows the extracted melody frequency and observation of these figures indicate that predominant fundamental frequency is not present at all point of time.

It can be seen in Fig. 3(iii) opera_fem2 music file, where the predominant frequency is not present from 8 s to 12 s. While similar observations are present in the entire center figure, it is observable in Fig. 3 (iii). The rightmost column presents the average pitch error for all the methods, shown in Table 1. The related calculations GPE are obtained from Eq. (10).

4. Conclusion

In this paper, the extraction of melody is tried for polyphonic music. For this purpose morlet wavelet is modified for detection of closed frequency. The results obtained from MMW scheme is then compared with the method available in the literature - MRFFT, STFT, and MW. The various audio files (like jazz1, midi2, opera_fem2, daisy1, and pop1) have been used in this melody extraction experiment (Fig. 4).

The GPE is obtained for all the methods using these audio files. The proposed MMW is found to have better GPE performance. Thus the proposed method is higher accuracy and can be considered over these algorithms. However, this improves performance comes at the cost of increased computation due to the involvement of the wavelet transform. Thus the future work of the current experiment is to reduce the complexity by maintaining the GPE performance

Declaration of Competing Interest

The authors have no conflict of interests towards publication of manuscript in this journal.

References

[1] R. Foucard, J.-.L. Durrieu, M. Lagrange, G. Richard, Multimodal similarity between musical streams for cover version detection, in: Proc. ICASSP, Dallas, TX, Mar. 2010, pp. 5514–5517.
[2] Y. Chien, H. Wang, S. Jeng, An acoustic-phonetic model of F0 likelihood for vocal melody extraction, IEEE/ACM Trans Audio Speech Lang Process 23 (9) (2015) 1457–1468, https://doi.org/10.1109/TASLP.2015.2436345.
[3] M.G. Reddy, K.S. Rao, Predominant melody extraction from vocal polyphonic music signal by combined spectro-temporal method, in: 2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Shanghai, 2016, pp. 455–459, https://doi.org/10.1109/ICASSP.2016.7471716.
[4] R.P. Paiva, T. Mendes, A. Cardoso, *Melody detection in polyphonic musical signals: exploiting perceptual rules, note salience, and melodic smoothness*, Comput. Music, J. 30 (4) (2006) 80–98.
[5] A. Klapuri, M. Davy (Eds.), Signal Processing Methods for Music Transcription, Springer, New York, 2006.
[6] J.A. Moorer, Signal processing aspects of computer music: a survey, in: Proceedings of IEEE, 1977, pp. 1108–1137.
[7] W.J. Pielemeiier, G.H. Wakefie2014ld, A high- resolution time-frequency representation of musical instrument signals, J. Acoust. Soc. Am. 99 (4) (1996) 2382–2396.
[8] S. Foster, J. Rockmore, W. Schloss, Toward an intelligent editor of digital audio: signal processing methods, Comput. Music J. 6 (1) (1982) 142–161.
[9] J.Y. Yoon, C.J. Song, S.P. Lee, H. Park, Extracting predominant melody of polyphonic music based on harmonic structure, in: 7th Music Information Retrieval Evaluation Exchange (MIREX), 2011 extended abstract.
[10] W. Zhang, Z. Chen, F. Yin, Q. Zhang, Melody extraction from polyphonic music using particle filter and dynamic programming, IEEE/ACM Trans Audio Speech Lang Process 26 (9) (Sept. 2018) 1620–1632, https://doi.org/10.1109/TASLP.2018.2834722.
[11] J. Salamon, E. Gomez, D.P.W. Ellis, G. Richard, *Melody extraction from polyphonic music signals: approaches, applications and challenges*, IEEE Signal Process. Mag. 31 (2) (Mar. 2014) 118–134.
[12] K. Dressler, Sinusoidal extraction using an efficient implementation of a multi-resolution FFT, in: Proceeding 9th International Conference on Digital Audio Effects. (DAFx-06), Montreal, Canada, 2006, pp. 247–252, 33, 57, 58, 59.
[13] Salamon, E. Gómez, Melody extraction from polyphonic music signals using pitch contour characteristics, IEEE Trans. Audio, Speech, Lang. Processing 20 (6) (2012) 1759–1770.
[14] https://en.wikipedia.org/wiki/Morlet_wavelet.
[15] Michel Misiti, Yves Misiti, Georges Oppenheim, and Jean-Michel Poggi. "Wavelet toolbox, 15, The MathWorks Inc, Natick, MA, 1996, p. 21.
[16] Raghuveer M. Rao, A.jit S. Bopardikar, Wavelet Transforms: Introduction to Theory and Applications, Addison-Wesley, 1998, p. 3. Chapter-1.
[17] P. Cano, E. G_omez, F. Gouyon, P. Herrera, M. Koppenberger, B. Ong, X. Serra, S. Streich, N. Wack, ISMIR 2004 Audio Description Contest, Tech. rep., Music Technology Group, Universitat Pompeu Fabra, Barcelona, Spain, 2006.
[18] A.P. Klapuri, Multiple fundamental frequency estimation by summing harmonic amplitudes, in: Proc. Int. Conf. Music Inf. Retrieval (ISMIR), 2006, pp. 216–221.

Neeraj Kumar received the B.Tech.in Electronics & Communication Engineering from MIT Purnea, Bihar, M. Tech. degree in Electrical Engineering (Specialization in Control System) from BIT Sindri Jharkhand and pursuing Ph.D. in Electronics Engineering from IIT(ISM) Dhanbad, Jharkhand. His-research interests include Signal Processing, Musical Signal Processing, and Biomedical Signal Processing.

Raubin Kumar received the B.Tech. in Electronics & Communication Engineering from School of Technology, North Eastern Hill University, Shillong, Megahlaya and M.Tech. degree in Electronics Engineering from IIT(ISM) Dhanbad, Jharkhand. He is currently working as software development Engineer with Talentica software(I) Pvt. Ltd. His-research interests include Signal Processing, Musical Signal Processing, and Biomedical Signal Processing.

Dr. Govind Murmu received his B.Tech. degree in ECE and M. Tech. degree in Signal Processing from NIT, Hamirpur, India, in 2002 and 2010 respectively. He is currently Assistant Professor in Electronics Engineering Dept, at IIT(ISM) Dhanbad, Jharkhand. His-research interests include signal processing applications in communications and control Engineering.

Prabira Kumar Sethy has received his Master of Technology degree in Communication Engineering from IIT (ISM) Dhanbad. He is working as Assistant Professor in Electronics, Sambalpur University, Odisha. His interest of research area is Image Processing. He has published more than 60 number of research article in International Journal and Conferences. He has received "InSc Young Achiever Award" for the research paper "Detection of coronavirus (COVID-19) based on Deep Features and Support Vector Machine, organised by Institute of Scholars, Ministry of MSME, Government of India.

He also received Best Research Paper presented award in IEEE International Conference on Communication and Signal Processing (ICCSP-2017), Chennai, India.

He qualified UGC NET in Electronic Science on June 2014 & December 2015 and MHRD GATE Scholarship during MTech. He selected as the Best Participant in Refresher Course in Mathematics (MD) with theme of "Mathematical Modeling & Simulation" Conducted by UGC Academic Staff College, Sambalpur University,15th December to 5th January 2015. He is the Editorial Board Members in International Journal of Electrical and Computer Engineering (IJECE). ISSN: 2088–8708. Scopus Indexed (Q2). He also served as TPC in various International Conferences and Symposium.