Frequency Estimation for Monophonical Music by Using a Modified VMD Method

Berrak Ozturk Simsek
Dept. of Electrical-Electronics Engineering
Istanbul Kultur University
Istanbul, Turkey
Email: bozturk@iku.edu.tr

Aydin Akan
Dept. of Electrical-Electronics Engineering
Istanbul University
Istanbul, Turkey
Email: akan@istanbul.edu.tr

***Abstract*—In this paper, a new Variational Mode Decomposition (VMD) is introduced, and applied to the fundamental frequency estimation of monophonical Turkish maqam music. VMD is a method to decompose an input signal into an ensemble of sub-signals (modes) which is entirely non-recursive. It determines the relevant bands adaptively, and estimates the corresponding modes concurrently. In order to optimally decompose a given signal, VMD seeks an ensemble of modes with narrow-band properties corresponding to the Intrinsic Mode Function (IMF) definition used in Empirical Mode Decomposition (EMD). In our proposed modified VMD approach, in order to obtain the bandwidth of a mode, each mode is shifted to baseband by mixing an exponential that is adjusted to the respective center frequency. The bandwidth is estimated through elastic net method that linearly combines penalties of the Lasso and Ridge Regression methods. Simulation results on fundamental frequency estimation of real music and synthetic test data show better performance compared to classical VMD based approach, and other common methods used for music signals, such as YIN and MELODIA based methods.**

I. Introduction

Because of developing technology and Music Information Retrieval-MIR, the automatic music transcription has become an important and frequently studied in recent years [1]. Automatic transcription methods for western music cannot be applied to Turkish music directly, because of the many differences between Western and Turkish music such as sound system, pitch, tonality, rhythm and maqam [2], [3]. In addition, the embellishing of Turkish maqam music performance and the instruments with microtonal pitch frequency complicates the fundamental frequency analysis.

Western music has standard chord frequency, unlike Turkish music [4]. In Western music, the frequency space for fundamental frequency that is referred to note name is divided into 12 equal pieces; in case of Turkish maqam music however, it is divided into 53, 106, or 159 equal pieces. However, there is no evidence of which is appropriate for performance and theory [4], [5]. Because of differences between these two music types, the tools and methods that are used to analysis of western music should be redeveloped for Turkish music. As for today, computational methods aiming at automatic transcription of Turkish music is very limited in number.

Notion of Maqam is the cornerstone of Turkish music and has some similar features with the traditional music of Asia, Middle East and North Africa. That's why; the automatic transcription of Turkish music will contribute to the understanding of music belonging to a very large multicultural area [6].

Pitch frequency in music is a perceptual feature of sound. It is important in terms of hearing and understanding of sound. Physical f_0 term corresponds to periodicity [5], [7]. It is difficult to determine the starting point of the pitch period. Hence, it is the general approach to define a note within an interval around a center frequency. In order to determine this center frequency, estimation of the physical f_0 is required. Turkish maqam music is performed monophonically and heterophonically unlike Western music. The frequent use of embellishment in performance of Turkish music creates a challenge in the estimation of fundamental frequency. In this study, we propose a new approach for fundamental frequency analysis in monophonical Turkish maqam music recordings by using a modified VMD method.

II. Fundamental Frequency Estimation Methods

In the literature, automatic transcription problem is mostly studied for western music. As such, studies are focused on the decomposition of polyphonical music in general.

Spectrogram is the most basic method for time-frequency representation of signals that is based on Short Time Fourier Transform (STFT) [8] and it is widely used for comparison and accuracy testing.

EMD has been proposed at the end of 1990's by Huang and used to obtain Hilbert Huang transform (HHT) [9]. Aim of EMD is to decompose a signal into sub-signals which have different spectral bands. HHT/EMD is widely used for the analysis and classification of non-stationary signal in many applications.

Some of fundamental frequency estimation methods adopt Wavelet Transform. The core assumption of this approach is based on choosing of wavelet scale. Other similar studies use Empirical Wavelet Transform to decompose the signal into adaptive wavelets in adaptive sub-bands [10]. This model makes a spectral separation from obtained maximum points and obtains the proper wavelet filter bank.

YIN is one of the most commonly used method, which employs autocorrelation parameter in the fundamental frequency estimation [4]. In this method first the autocorrelation function

ISBN 978-0-9928626-7-1 © EURASIP 2017

is calculated, then the local peaks of the autocorrelation function is detected, and the distance between peaks is found as fundamental period of the signal.

MELODIA has proven to be a popular method for the melody estimation. It groups pitch frequency candidates as a continuous sequence (pitch contour). A number of characteristics are defined from pitch contours and rules are defined in order to differentiate melodical and non-melodical pitch contour by examining the characteristics distribution [11]. An equal-loudness filter is employed which increases mid-band frequencies, that is assumed to contain melody, while decreasing low-band frequencies. After the filtering, Short Time Fourier Transform (STFT) is applied to find local maxima and consequently spectral peaks. Finally, the saliency function is obtained from the spectral peaks that give the f_0 candidates [5].

III. Variational Mode Decomposition

Empirical Mode Decomposition (EMD) is widely used in speech processing. However, due to its lack of theoretical definition and recursive structure, backward error correction is not possible. Popular methods for fundamental frequency estimation have numerous limitations such as recursiveness, incompetency in noise handling and strict band boundaries [12]. VMD uses variational model to determine relevant bands adaptively and concurrently estimates corresponding modes. Narrow-band properties representing intrinsic mode function definition in EMD are also used in VMD. Intrinsic mode functions (u_k) are Amplitude Modulated-Frequency Modulated (AM-FM) signals [5], [12]:

$$u_k(t) = A_k(t)\cos(\phi_k(t)) \tag{1}$$

where $\phi_k(t)$ is the phase that is a non-decreasing function, $A_k(t) \geq 0$ is a non-negative envelope, and, $\omega_k(t) := \phi_k'(t)$ is the instantaneous frequency [12].

A series of optimally reconstructed modes in input signal are sought. These modes are band limited around the center frequency. Variational model is still robust in the presence of noise, contrary to EMD. Its tight relationship with Wiener Filter is an indication of this advantage. Essentially, VMD is a generalized form of Wiener Filter as an adaptive and multiple band method. Variational model defines bandwidth of modes in the form of H1-norm, to be more precise after shifting the Hilbert-complemented, analytic signal into baseband by multiplying with the complex harmonic. Total bandwidth of a mode is given by the sum of maximum frequency deviation in frequency modulation Δf, bandwidth of the envelope $A_k(t)$, and FM signal bandwidth [5], [12].

$$BW_{AM-FM} = 2(\Delta f + f_{AM} + f_{FM}) \tag{2}$$

The resulting optimization is straightforward. Modes u_k are iteratively updated in the frequency domain, then the center frequency estimation is repeated to find center-of-gravity of the mode's power spectrum [12], [13]. This method produces better results than EMD in tone separation and identification without requiring harmonic frequencies. Moreover, it performs favorably for synthetic and real music data and it is more robust in the presence of noise [5], [12], [13].

$$x_0 = x + \eta \tag{3}$$

where x is original input signal, x_0 is observed signal and η is zero mean Gaussian noise.

In order to recover the input signal x, VMD uses the Tikhonov Regularization as a solution of the minimization problem as follows:

$$\min_x \{ \|x - x_0\|_2^2 + \alpha \|\partial_t x\| \tag{4}$$

from which the Euler-Lagrange equations are easily obtained and typically solved in the frequency domain:

$$\hat{x}(\omega) = \frac{\hat{x}_0}{1 + \alpha\omega^2} \tag{5}$$

where $\hat{x}(\omega)$ is the Fourier Transform of the signal x. The recovered signal x is a narrow-band version of the input signal x_0 around $\omega = 0$, where α represents the variance of the noise.

The results of variational problem is the following:

$$\min_{u_k, \omega_k} \left\{ \sum_k \left\| \partial_t \left[\left(\delta(t) + \frac{j}{\pi t} \right) * u_k \right] \mathrm{e}^{-j\omega_k t} \right\|_2^2 \right\}$$
$$\sum_k u_k = x \tag{6}$$

The method use both a quadratic penalty term, and Lagrangian multipliers in order to render the problem unconstrained. Final augmented Lagrangian is given by

$$\begin{aligned}\mathcal{L}(u_k, \omega_k, \lambda) = \alpha \sum_k \left\| \partial_t \left[\left(\delta(t) + \frac{j}{\pi t} \right) * u_k \right] \mathrm{e}^{-j\omega_k t} \right\|_2^2 \\ + \left\| x - \sum u_k \right\|_2^2 + \left\langle \lambda, x - \sum_k u_k \right\rangle \end{aligned} \tag{7}$$

Now, the solution of minimization problem is saddle point of the Lagrangian $\mathcal{L}$.

We show an example of fundamental frequency estimation of a synthetic signal $f(t) = 6t + 5\cos(2\pi 20t) + 10\cos(2\pi 25t)$ by using VMD in Fig. 1.

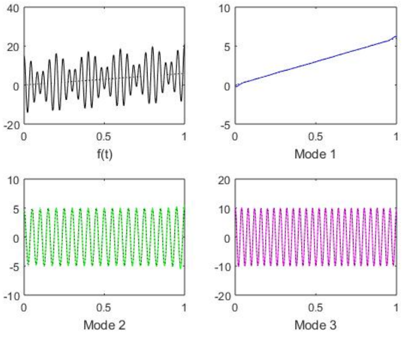

Fig. 1. Input signal f(t) and its modes

IV. Proposed Approach

VMD enables estimation of fundamental frequency of the signal, composed of similar frequency and/or amplitude signals, with high accuracy. However, it has been observed that using auto-correlation function instead of given signal in VMD based frequency estimation method has proved to be more successful. Hence, in the proposed method, frequency estimation was performed by decomposing the auto-correlation function $R_{xx}(k)$ of signal $x(n)$ into the modes with VMD [5], [13].

$$R_{xx}(k) = \sum_{n \in Z} x(n)x^*(n-k) \tag{8}$$

A. Modified VMD Method

While, the original VMD algorithm makes use of Ridge Regression (Tikhonov Regularization) in order to obtain the reconstructed input signal x, our proposed method introduces a different type of regression that is ElasticNet to obtain the bandwidth. Elasticnet Regression [15], [16] method that combines penalties of the Lasso and Ridge Regression methods linearly. This type of regression is useful when there are many properties which are correlated. This method use a quadratic penalty term to overcome the limitation of Lasso regression that has ℓ_1 term. The minimization problem which is originate in Equation 3 can then be written as,

$$\min_x \{\|x - x_0\|_2^2 + \alpha\|\partial_t x\|_2^2 + \beta\|\partial_t x\|_1\} \tag{9}$$

The solution of the minimization problem:

$$\hat{x}(\omega) = \frac{\hat{x}_0}{1 + \alpha\omega^2 + \beta j\omega} \tag{10}$$

A quadratic penalty term make the cost function strictly convex, as such it provides a single, global minimum.

$$\begin{aligned}\mathcal{L}(u_k, \omega_k, \lambda) = \alpha \sum_k \left\| \partial_t \left[\left(\delta(t) + \frac{j}{\pi t} \right) * u_k \right] \mathrm{e}^{-j\omega_k t} \right\|_2^2 \\ + \beta \sum_k \left\| \partial_t \left[\left(\delta(t) + \frac{j}{\pi t} \right) * u_k \right] \mathrm{e}^{-j\omega_k t} \right\|_1 \\ + \left\| x - \sum u_k \right\|_2^2 + \left\langle \lambda, x - \sum_k u_k \right\rangle \end{aligned} \tag{11}$$

B. Enhancement of Pitch Contour

In some situations fundamental frequency estimation methods yield faulty outcomes. It is possible to enhance some of these outcomes by post-processing.

Using a filter helps improve the octave errors which are occured when the continuity is lost. To obtain continuous pitch contour with filtering, following assumptions are taken into account as unlikely in Turkish maqam music: the size of a short duration pitch chunk is larger than five, and a shorter pitch chunk is compared to the adjacent longer pitch chunks in upper and lower octaves, and then appended to the front or end of them to correct the octave errors. Another assumption is that the melodic dynamic range is larger than four octaves [13], [14]. In fig. 2, we give the original and enhanced pitch contour of the signal shown in Fig. 1.

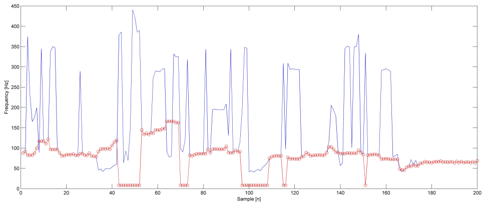

Fig. 2. The pitch contour (solid line) and enhanced version (dotted line).

In our experimental studies, fundamental frequency of four different maqam music recordings are estimated using the proposed method. We evaluated the success of the results for fundamental frequency, whose octave errors have been corrected by the the filtering approach.

V. Simulation Results and Discussion

We used four different maqam music pieces performed by 8 different instruments including wind, percussive and stringed to test the performance of our proposed fundamental frequency estimation. All instruments were recorded separately. The number of modes in the signal to be decomposed should be determined beforehand. In our study the number of modes was determined experimentally, and seven modes were used in the analysis.

It is seen in Figure 3 that pitch contour obtained by the proposed method in monophonic Turkish maqam music records match well with their spectrograms.

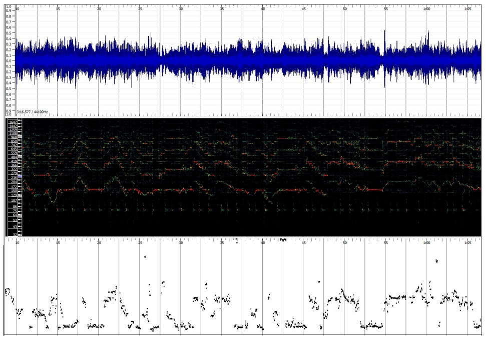

Fig. 3. On the top music signal record, middle section is peak spectrogram of record and on the bottom proposed method result

The proposed method has been evaluated with MIREX 2008 (Audio Melody Extraction) results [17]. The evaluation has been conducted with 31 different 30-second sections that were chosen from the about three minute recording. This comparison has been performed upon Voicing Detection (VD), Voicing False Alarm Rate (VFA), Raw Pitch Accuracy (RPA), Raw Chroma Accuracy (RCA) and Overall Accuracy (OA) [17], [18]. Since true fundamental frequency values of this data set are not known, these values have been compared with the values obtained from spectrogram.

ISBN 978-0-9928626-7-1 © EURASIP 2017

Voicing Detection: the proportion of frames labeled voiced in the ground truth that are estimated as voiced by the algorithm [17], [18].

Voicing False Alarm Rate: The proportion of frames labeled as unvoiced in the ground truth that are mistakenly estimated as voiced by the algorithm.

Raw Pitch Accuracy: the proportion of voiced frames in the ground truth for which the F0 estimated by the algorithm is within $\pm 1/4$ tone ($\pm\%3$) of the ground truth f_0.

Raw Chroma Accuracy: same as the raw pitch accuracy except that both the estimated and ground truth f_0s are mapped into a single octave. This ignores errors where the pitch is wrong by an exact multiple of an octave (octave errors).

Overall Accuracy: this measure combines the performance of the pitch estimation and voicing detection tasks to give an overall performance score for the system.

The performance of the fundamental frequency estimation in 31 monophonic Turkish music records with different maqam are given in Table I.

TABLE I
COMPARISON OF THE PROPOSED METHOD, ORIGINAL VMD, YIN AND MELODIA

Method	VD	VFA	RPA	RCA	OA
Proposed Method	59.21%	**27.05**%	**61.12**%	**42.42** %	22.08%
Original VMD	64.37%	48.32%	53.25%	38.14%	18.48%
MELODIA	74.58%	34.67%	57.26%	42.40%	33.12%
YIN	**99.56**%	92.11%	39.45%	23.32%	**38.34**%

VI. CONCLUSIONS

We proposed a modified VMD based method for fundamental frequency estimation and obtaining the pitch contour in monophonic Turkish maqam music records.

It is shown that estimated fundamental frequencies by the proposed method are in agreement with correct fundamental frequency values in the recordings. Results have shown that the method can be successfully utilized for monophonic Turkish maqam music records, which is a novel contribution to the literature. This study is a step forward for heterophonic music analysis, automatic transcription systems, as it opens the way for new studies on melodic analysis and rhythm analysis [5].

It has been observed that the results obtained in real monophonic music with the proposed method are comparable with polyphonic music oriented methods such as YIN and MELODIA in terms of their accuracy. In our future works new algorithms, such as time-frequency reassignment, synchrosqueezing and empirical wavelet transform, will be introduced in order to further increase the performance of the proposed method.

REFERENCES

[1] Matti P. Ryynänen, A. Klapuri, "Automatic transcription of melody, bass line, and chords in polyphonic music," Computer Music Journal, 32.3, 72-86, 2008.

[2] B. Bozkurt, R. Ayangil, A. Holzapfel, "Computational analysis of turkish makam music: Review of state of the art and challenges," Journal of New Music Research, 43(1), 3-23, 2014.

[3] S. Sentürk, "Computational modeling of improvisation in Turkish folk music using variablelength Markov models," Ph.D. dissertation, Georgia Institute of Technology, Atlanta, 2011.

[4] B. Bozkurt, A. C. Gedik, A. Savacı, M. K. Karaosmanoğlu, M. E. Özbek, *Klasik Türk müziği kayıtlarının otomatik olarak notaya dökülmesi ve otomatik makam tanıma*, Tübitak Projesi, İzmir Yüksek Teknoloji Enstitüsü, İzmir, 2007-2010.

[5] B. Ozturk Simsek, A. Akan, B. Bozkurt, "Fundamental frequency estimation for monophonical Turkish music by using VMD," *Signal Processing and Communications Applications Conference (SIU)*, 16-19 May 2015 23th, pp.1022-1025.

[6] E. Benetos, A. Holzapfel, "Automatic Transcription of Turkish Makam Music," *In Proceedings of ISMIR - International Conference on Music Information Retrieval*, November. 4–8th, Curitiba, Brazil, 2013.

[7] A. Klapuri, "Automatic Music Transcription as We Know it Today," Journal of New Music Research, Vol. 33, No. 3, pp. 269-282, 2004.

[8] L. Cohen, "Time-Frequency Analysis", Vol. 778. Englewood Cliffs, NJ:Pretice Hall PTR, 1995.

[9] N. E. Huang, S. S. Samuel, "Hilbert-Huang Transform and Its Applications", Vol. 5. World Scientific, 2005

[10] J. Gilles, "Emprical Wavelet Transform", *Signal Processing*, IEEE Transactions on 61.16, 3999-4010-, 2013

[11] J. Salamon, G. Emilia, "Melody extraction from polyphonic music signals using pitch contour characteristics," *Audio, Speech, and Language Processing*, IEEE Transactions on 20.6 (2012): 1759-1770, 2012.

[12] K. Dragomiretskiy, Z. Dominique, "Variational Mode Decomposition," *IEEE Transactions on Signal Processing*, vol. 62, No. 3, 2014.

[13] B. Öztürk Şimşek, B. Bozkurt, A. Akan, "Fundamental frequency estimation for heterophonical Turkish music by using VMD," Signal Processing and Communication Application Conference (SIU), 2016 24th. IEEE, Zonguldak, 2016.

[14] B. Bozkurt, A. C. Gedik, M. K. Karaosmanoğlu, "An automatic transcription system for Turkish music," Signal Processing and Communications Applications (SIU), 2011 IEEE 19th Conference on. IEEE, 2011.

[15] H. Zou, T. Hastie, "Regularization and variable selection via the elastic net," Journal of the Royal Statistical Society: Series B (Statistical Methodology), 67(2), 301-320, 2005.

[16] C. Hans, "Elastic net regression modeling with the orthant normal prior," Journal of the American Statistical Association, 106(496), 1383-1393, 2011.

[17] MIREX Wiki: Audio Melody Extraction, http:www.music-ir.orgmirexwiki2008:MIREX2008_Results

[18] J. Salamon, E. Gómez, "Melody extraction from polyphonic music signals using pitch contour characteristics," Audio, Speech, and Language Processing, IEEE Transactions on 20.6, 1759-1770, 2012.