

# A Recommender System for Music Less Singing Voice Signals

G. M. Nazmus Salehin and Md. Shahjahan

Dept. of Electrical and Electronic Engineering, Khulna University of Engineering & Technology (KUET)

**Abstract**—With widespread access to internet and huge availability of music players like iPhone, Smart phone songs are now available anytime and anywhere to any people. The problem now is not the availability of songs, but to find right song for right person. Song recommendations are grouped by different aspects like melody, rhythm etc. This paper proposes a recommender system by melodic similarity of songs. Fundamental frequency of song has been extracted and a dynamic programming approach named dynamic time warping has been used to find total fundamental frequency deviation of two songs and this process is continued for all song tracks in a playlist and finally recommendations are given as ascending order of overall fundamental frequency deviation percentage. This system also can give a rhythmic rating based on how a target song deviates with respect to a reference song.

**Keywords**— voice signal, signal processing, frequency estimation, time warping, recommendation system

## I. INTRODUCTION

The essence of evaluating pure voice signal without musical instrument is that one can understand the original characteristic of one's voice singing signal. With the advancements of technology, internet access and file sharing options are easier than ever before. Anyone can substantially build a huge collections of digital music. The task of finding proper song for user is called song recommendation. As singing voice is very complex in nature and have different aspects like melody, rhythm, pitch etc. This process is not so easy. The underlying goal of the recommendation system is to personalize content and identify relevant data for our audiences. Song recommender systems which are built based on user listening history, popularity are successfully implemented already. But building a recommender system which focuses on audio content of the song is a challenging task. In our proposed system we have implemented a content based recommender system which can recommend new songs to user on the basis of fundamental frequency deviation.

It is easy to differentiate between male and female voice signal because they have a differentiable tone and fundamental frequency deviation. However, it is a challenging task to differentiate among same gender voice signal because they have same set of frequency band with many similarities. This intelligent system attempts to introduce a technique to find differentiable features and recommend which voice signal is relatively finer than others signals.

A primary aim is to find out similar songs which have identical melodic content in a playlist. As a result, we are focusing on quality of melodic representation of the singing voice. We extract fundamental frequency of the song and so every song in the playlist is represented by its fundamental frequency. After selecting property of the song, the next approach is how to recommend songs to user. For this approach we have

selected a powerful dynamic programming algorithm named dynamic time warping which is used to find non-linear similarity between two time series signals. Songs in a playlist are compared by dynamic time warping. The total cost in the dynamic time warping path is the total deviation of the target song with respect to reference song. This process is continued by making every song as reference and finding out the total cost of remaining songs. By placing them in a matrix one can visualize each song's fundamental frequency deviation with respect to other. The total sum of cost is then computed by summing along row/column for each song. Finally, we calculate a deviation percentage for each song and recommend songs by ascending order of deviation percentage.

This paper is organized as follows: Section II provides the overall approach of the recommender system. Section III describes on the data set. Section IV provides an algorithm for frequency estimation. Section V describes the principles of dynamic time warping. Section VI describes all the calculation for frequency and time warping. Section VII shows the results on recommendation and finally the paper draws conclusion at section VIII.

## II. THE APPROACH

A block diagram is shown of the selected approach in Fig. 1.

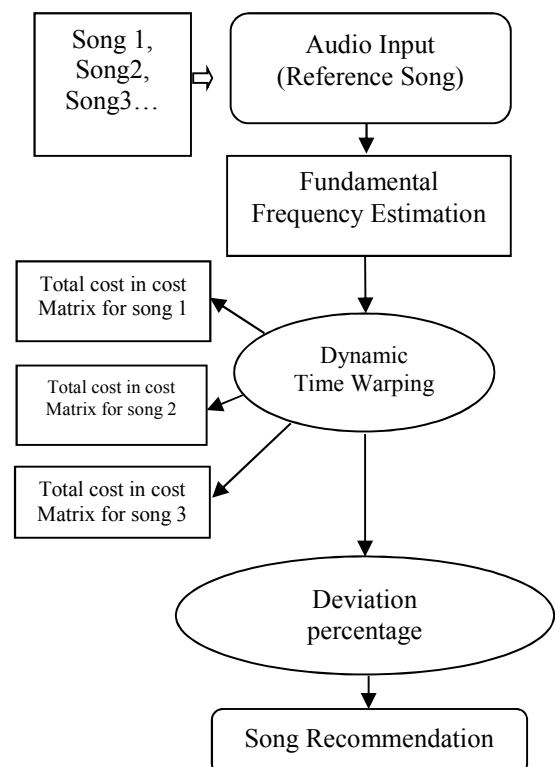


Fig. 1: Block diagram of the approach

Our recommendation mechanism works by comparing melodic similarity of a target song with respect to a reference melody. The audio input is represented by fundamental frequency of the singing voice. The following sections provides step by step procedure of our selected approach. Finally we provide song recommendations by calculating a deviation percentage of fundamental frequency.

### III. THE DATA SET

TABLE I: Characteristics of Data

Category	Short name of songs	Possible age of singer	Frequency band (min-max)
Islamic song	1. Buke Amar Kabar Sobi (BAK), 2. Shuni Muhammad (S) Naam (SMN), 3.He Khoda Doya Moy (HKD), 4. Ma aaaj keno amay (MKA)	12-50	225-655 Hz
Rock song (without music)	1. Fix you (FY), 2. Lithium (LI), 3. Broken (BR) 4. With or without you (WY), 5. No more lies (NL)	25-45	270-615 Hz

Table 1 depicts the nature of data set. Two types of data are used to recommend – Islamic song (Gazal) and rock song. Islamic song category has four songs and other has 5 songs. Every song is designated by abbreviated characters for simplicity sake such as “Buke Amar Kabar Sobi” is shortly named as BAK. All song are male voices with different ages and identical frequency band. The source of Islamic and rock songs are available at web link [http://almodina.com/site\\_bangla\\_islamic\\_song.xhtml](http://almodina.com/site_bangla_islamic_song.xhtml) and <https://www.acapellas4u.co.uk/> respectively. However, the rock songs are extracted as without music and Islamic songs are originally instrumental music free.

### IV. FUNDAMENTAL FREQUENCY ESTIMATION

There are many approaches for estimating fundamental frequency of audio signal. For monophonic signals there are several approaches like YIN algorithm [1] and two way mismatch method[2]. For polyphonic signals, fundamental frequency estimation is quite difficult since several sound sources alongside voice like bass, guitar and drums etc are available. Salamon and Gomez developed an algorithm to extract melody from polyphonic music[3]. As our main goal is to recommend songs only taking account of voice portion and without any musical content we have selected YIN algorithm for fundamental frequency estimation as this algorithm provides good results for monophonic signals. This algorithm is an autocorrelation based approach and works successfully for monophonic signals.

### V. DYNAMIC TIME WARPING

Dynamic time warping is employed to find optimal match between two given sequences. It performs non-linear series comparison between input and template. In this paper, we select a reference song and a target song as input. The first of

this algorithm is to construct a cost matrix which is constructed by calculating squared difference of fundamental frequency at any instant between 2 songs (reference and target). The cost matrix M is defined as follows:

$$M_{ij} = \{f_{0T}(i) - f_{0R}(j)\}^2 \quad (1)$$

Where i and j is time instant,  $f_{0T}$  is fundamental frequency of target song and  $f_{0R}$  is fundamental frequency of reference song. The cost matrix is shown in Fig. 2.

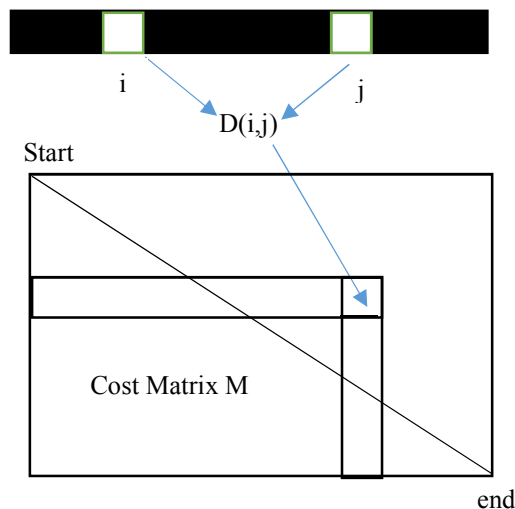


Fig. 2: Cost Matrix

The warping path which is represented by  $W = w_1, w_2, \dots, w_k$ , which begin with  $w_1(1,1)$  and end with  $w_k(m, n)$ . The distance can be found as [4].

$$D(i, j) = Dist(i, j) + \min[D(i, j - 1), D(i - 1, j), D(i - 1, j - 1)] \dots \dots (2)$$

This equation recursively fills up the cost matrix. Finally we get the shortest distance which is denoted by 2<sup>nd</sup> equation between fundamental frequencies of the target and reference song.

### VI. THE CALCULATION

#### A. Fundamental frequency deviation

The total distance of the optimal cost path in the cost matrix which is denoted by D (m,n) gives us fundamental frequency deviation of the target song with respect to the reference song. This value is mainly used for our recommender system to recommend new songs to user. The fundamental frequency deviation (FFD) is defined as Eq. (3)

$$FFD = \sum_{k=1}^K M_{i_k j_k} \dots \dots (3)$$

Where k=0, 1, 2 .... is index of cost path.

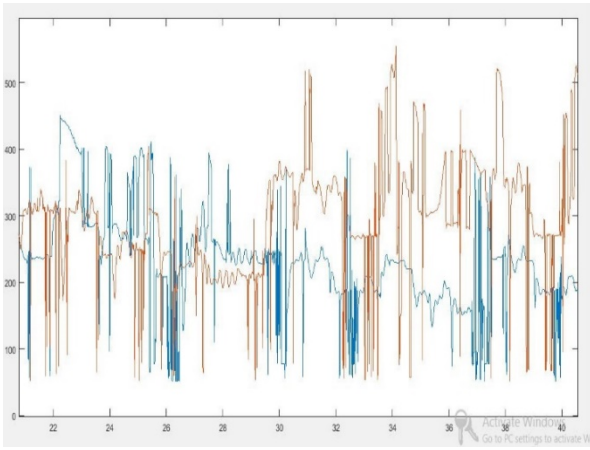


Fig. 3. Fundamental frequency deviation within 1<sup>st</sup> 60 seconds of HKD and SMN songs.

### B. Rhythmic deviation

For any target songs we can use DTW to find rhythmic deviations of a target song with respect to a reference song [4]. By analyzing the shape of the optimal cost path we can get an idea if a target song is in rhythmically stable condition with respect to the reference song. A perfect rhythmically similar song would produce an optimal cost path which is 45° straight line. An unstable target song with respect to the reference would produce deviations in the optimal cost path. Deviation location of the cost path can be extracted precisely. Total rhythmic deviation can be calculated by fitting a 1<sup>st</sup> order linear regression fit to the cost path and computing the root mean squared error. We can get the total rhythmic deviation (RD) in Eq. (4) as follows.

$$RD = \sqrt{\frac{1}{N} \sum_{k=1}^K \varepsilon_k^2} \dots \dots (4)$$

### C. Time warping

We can easily calculate the dynamic time warping between two signals as shown in Fig. 5. It reveals the hypothetical picture of Fig. 4.

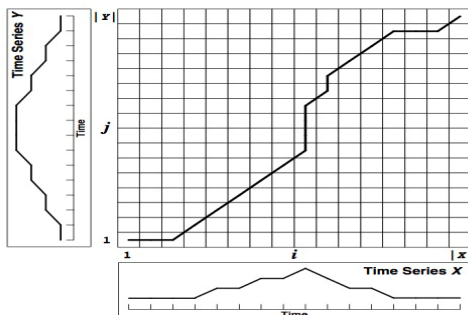


Fig. 4: Hypothetical dynamic time warping of two time series

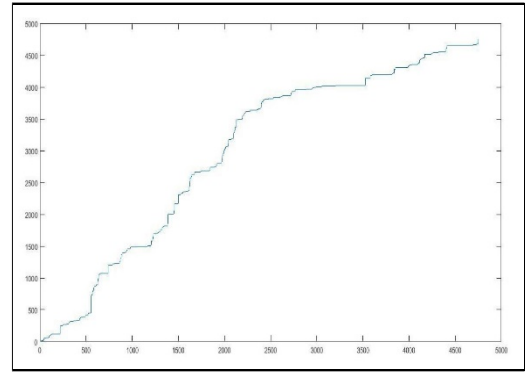


Fig. 5: Real DTW between reference song (HKD) and target song SMN within 1<sup>st</sup> 60 seconds.

## VII. RESULTS FOR RECOMMENDATION

In song recommendation process we have considered fundamental frequency deviation only to provide recommendation to user as this process correlates highly with human judge score in [6]. Song ranking process works as follows. We consider a database of songs of any category. We consider only the voice portion of every song. We consider every song in the database as reference song and find fundamental frequency deviation of each song with respect to that reference song. Overall results are stored in a matrix where each row corresponds to a reference song and each column corresponds to fundamental frequency deviation with respect to the reference song. When any reference song is compared with the same song, deviation becomes zero. Overall sum taken along row or column represents a song's total fundamental frequency deviation in a database. We sum the values along row or column which refers to different song's contribution in the playlist. A sample deviation between two songs is shown in Fig. 3.

	Song1	Song2	Song3	Overall Deviation
Song1	0	$\sum M_{ij}$	$\sum M_{ij}$	$\sum \sum M_{ij}$
Song2	$\sum M_{ij}$	0	$\sum M_{ij}$	$\sum \sum M_{ij}$
Song3	$\sum M_{ij}$	$\sum M_{ij}$	0	$\sum \sum M_{ij}$

Fig. 6: Song recommendation matrix

Then we get a deviation percentage by taking a ratio by a song's total fundamental frequency deviation and dividing it with all song's total fundamental frequency deviation. This percentage is stored finally from low value to high value and ranked in ascending order.

Our system is organized in the following way:

- 1) Construct a dataset of Songs (Bangla/English).
- 2) Preprocess all songs and save them as 44.1 kHz/128 kbps and mono MP3 format.

- 3) Find fundamental frequency for each song and represent each song with their corresponding fundamental frequency.
- 4) Apply DTW algorithm to every song in the dataset.
- 5) Using song recommendation matrix extract overall fundamental frequency deviation of each song.
- 6) Calculate deviation percentage and rank songs from lowest percentage to highest percentage.

Our database consists of Islamic Songs, Rock songs and Pop songs. We have considered only the voice portion and did not consider instrument/music of the songs. Islamic songs on our database only contains voice and does not contain instruments. For other category we have experimented with **Acapella's** (voice part without instrumental accompaniment). Instrumental and music information are filtered out for having only voice part. All songs are full length and sampled at 44100 Hz.

For the 1<sup>st</sup> experiment we consider Islamic songs. The songs names are given below. Total fundamental frequency deviations (FFD) are arranged in Table 2 for the Islamic songs. The ranking of each song is shown in Table 3 as percentage deviation of fundamental frequency (PDFF).

TABLE II: Experimental results on Islamic Songs ( $\times 10^5$ )

	BAK	SMN	HKD	MKA	Total FFD
BAK	0	3.73	3.8239	5.1601	12.714
SMN	3.739	0	3.1847	6.2152	13.139
HKD	3.8239	3.18	0	6.55	13.553
MKA	5.16	6.21	6.55	3.73	17.92

TABLE III: Ranking of selected Islamic songs

Song Name	Rank	PDFF
BAK	1	0.2217
SMN	2	0.2291
HKD	3	0.2361
MKA	4	0.3127

The same process has been carried out for Rock category (5 songs). Results are given as follows:

TABLE IV: Experimental results on Rock songs ( $\times 10^5$ )

	FY	LI	BR	WY	NL	Total FFD
FY	0	7.204	3.034	3.450	4.111	17.8
LI	7.204	0	3.132	3.562	2.141	16
BR	3.034	3.132	0	3.021	5.367	14.6
WY	3.450	3.562	3.021	0	3.044	13.1
NL	4.111	2.141	5.367	3.044	0	14.7

TABLE V: Ranking of rock songs

Rank	Song Name	PDFF
1	WY	0.1717
2	BR	0.1911
3	NL	0.1926
4	LI	0.2106
5	FY	0.2338

There are several advantages of the proposed algorithm. Firstly, comparison between two time series was made using DTW which is more reliable than Euclidian distance method. This is because DTW scan several nearby points with respect to reference point. Secondly, fundamental frequency deviation calculation gives us main sound of the singer. Thirdly, one can use it in the evaluation of the quality of voice in a competition arranged by different sectors. In addition, national investigation department can use it easily to identify voice quality of a criminal detection program.

The finding obtained here is a primary target. The method can be improved by applying more modern density estimation instead of distance calculation.

## VIII. CONCLUSIONS

Our proposed recommendation system presents a new scheme for comparing melodic similarity and can recommend new songs to user. Dynamic time warping algorithm is simple but effective algorithm for automatic music assessment previously and such approach has not been much focused on prior work. Rhythmic deviations calculated by this system can be used for recommendation also as well as automatic singing voice assessment. Though this system has been tested on small scale, it can easily be incorporated into large datasets. Recommended songs in our playlist can be tested by human judges so that we may know how these songs are melodically similar with each other. For future improvements we can consider other aspects like rhythmic information directly in this system or use supervised learning methods.

## REFERENCES

- [1] C. A. de and H. Kawahara, "YIN, a fundamental frequency estimator for speech and music," *The Journal of the Acoustical Society of America* 111, no. 12002874, pp. 1917-1930, 2002.
- [2] C. M. Robert and B. W., "Fundamental frequency estimation of musical signals using a Two-Way Mismatch procedure," vol. 95, pp. 2254-2263, 1994.
- [3] S. J and E. Gomez, "Melody Extraction from Polyphonic Music Signals using Pitch Contour Characteristics," pp. 1759-1770, August 2012.
- [4] P. Senin, "Dynamic time warping algorithm review," Department of Information and Computer Sciences, University of Hawaii, 2008.
- [5] D. Ellis and R. Turetsky, October 2003. [Online]. Available: <http://www.ee.columbia.edu/~dpwe/resources/matlab/dtw/>.
- [6] E. Molina, E. Gomez and I. Barbancho, Automatic Scoring of singing voice based on melodic similarity measures, Barcelona: Universitat Pompeu Fabra, Music Technology Group, 2012.