Analysis and Pattern Recognition of Woodwind Musical Tones Applied to Query-by-Playing Melody Retrieval

Majid A. Al-Taee, Mohammed T. Al-Ghawanmeh, Fadi M. Al-Ghawanmeh and Baha O. Abu Al-Own

Abstract— **Unlike occidental flutes, Arabian flutes have gained only limited attention in literature over the past years. Moreover, the least tone-to-tone distance in Arabian music is only half of that of occidental music and therefore the analysis and features extraction process of Arabian music is more challenging. This paper investigates the tone-to-tone features of two traditional Arabian flutes; Al-Nay and Shabbaba. These two kinds of flute gain wide popularity in the Middle Eastern and North African countries. A previously reported analysis and features extraction package is adapted and used for query-by –playing melody retrieval. Experimental tone-to-tone investigations based on the Discrete Fourier Transform (DFT) are presented for both flutes. The obtained results showed that the analysis approach is robust and successfully performed the task of query-by –playing melody retrieval. Comparison results for Al-Nay and Shabbaba were obtained and evaluated using the Arabian musical Scale "Bayat", as a reference.**

Index Terms— **Arabian flute, automatic music transcription, features extraction, music information retrieval, time-frequency analysis.**

I. INTRODUCTION

Digital music is one of the most important data types distributed by the Internet. However, it is still difficult for a computer to automatically analyze music content, especially to automatically classify and recognize music content [1]. Music Information Retrieval (MIR) is a relatively new multi-disciplinary area of applications emerged in the late 1990s. It views problems by bringing various fields of paradigms and inquiries together. Nowadays, providing a dynamic MIR system is one of the most challenging issues. It should be done by enriching past experiences and traditions with new paradigms and techniques [2].

A number of methods have been proposed to discriminate music, speech, silence, and environment sound. The most successful achievement in this area is speech/music discrimination, because speech and music are quite different in spectral distribution and temporal change pattern. Saunders [3] used the average zero-crossing rate and the short time energy as features and applied a simple thresholding method to discriminate speech and music from the radio broadcast. Scheirer and Slaney [4] used thirteen features in time, frequency and spectrum domains, and different classification methods to achieve robust performance. However, most of the previous studies and analyses were performed with reference to the occidental music which differs from Arabian music in the tone-to-tone interval [5].

Arabian music theories identify tone-to-tone interval as a multiple of quarter tones rather than half tones as in occidental music [6], [7]. The Shabbaba which is a woodwind Arabian musical instrument can be considered the base upon which the classical Arabian flute (Nay) was developed. However, the sound domain of the Nay is broader and more accurate and therefore it is qualified to be among the ensemble of Arabian music or Arabian orchestras. The limited sound domain of Shabbaba is found adequate for Jordanian traditional songs which are simple and mostly composed on the Arabian scale "Bayat" [8].

The work reported in this paper focuses on the time-frequency analysis and pattern recognition of musical tones for two Arabian woodwind instruments (Nay and Shabbaba) with the ultimate goal of developing MIR system for Arabian music. The proposed analysis method allows for accurate pitch detection and an effective removal of undesirable noise. The "query-by-playing" application is then experimented successfully for both the Nay and Shabbaba.

II. NAY AND SHABBABA FEATURES

The Nay and Shabbaba, which are traditional Arabian woodwinds, are common in several Arabian countries despite the innovation and evolution of music in terms of composition and industry [9]. These instruments are typically made of reeds. In some regions with limited water resources, the reeds-made flutes are replaced by metal- or plastic-made flutes. The Nay has six holes grouped into threes in its top side and one hole in its bottom. Nay reed stick is composed of nine segments separated by eight nodes. In contrast, Shabbaba has five or six equally separated holes with no restrictions on segments number [10].

Unlike the Shabbaba, the Nay cavity has a narrow waist at the first node; it is done by leaving this node not fully unfilled when making the instrument. The Nay waist makes it possible

Manuscript received March 22, 2009. (The authors acknowledge financial support of this work by Abdulhameed Shoman fund to Support Scientific Research and the EU fund to Support Research and Technological Development (SRTD) in Jordan).

M. A. Al-Taee is with the Computer Engineering Department, University of Jordan, Amman 11942, Jordan (phone: +962-6-5355000; fax: +962-6-5355588; e-mail: altaeem@ ju.edu.jo).

M. T. Al-Ghawanmeh is with the Music Department, Yarmouk University, Irbid, Jordan (e-mail: Ghawanmeh_mohd@yahoo.com).

F. M. Al-Ghawanmeh is with the Music Department, University of Jordan (e-mail: fadighawanmeh@yahoo.com).

B. O. Abu Al-Own is with the Computer Engineering Department, University of Jordan (e-mail: baha_omar@hotmail.com).

for the instrument to perform the third octave tones when the player enforces blowing into the instrument. The Nay player usually has a set of seven instruments and the choice among them is entirely dependent on the level of the required musical scale [5].

The tones that can be played by the Nay Nawa and Shabbaba are shown in Table 1 along with their corresponding frequencies. shabbaba tones which are only a subset of the Nay Nawa tones are shaded in this table. Half flat tones (Arabian tones) are distinguished by the superscript "d" and are written in underlined bold italic font, as illustrated. Some of the tones ($F^\#$, A, $C^\#$, and E) needs special playing skills due to the necessity of closing half holes in the Nay Nawa rather than fully closing holes. The tone E^b of the Shabbaba can only be performed when the instrument has six holes rather than five.

Table 1. The Nay Nawa and Shabbaba tones.

1^{st} Octave	F(Hz)	2^{nd} Octave	F(Hz)	3^{rd} Octave	F(Hz)
F_4	349	F_5	698	F_6	1396
$F^\#_4/G^b_4$	370	$F^\#_5/G^b_5$	740	$F^\#_6/G^b_6$	1480
G_4	392	G_5	784	G_6	1568
$G^\#_4/A^b_4$	415	$G^\#_5/A^b_5$	830	$G^\#_6/A^b_6$	1660
A^d_4	***428***	***A^d_5***	***856***	***A^d_6***	***1712***
A_4	440	A_5	880	A_6	1760
$A^\#_4/B^b_4$	466	$A^\#_5/B^b_5$	932	$A^\#_6/B^b_6$	1864
B_4	494	B_5	988	B_6	1976
C_5	523	C_6	1046	C_7	2092
$C^\#_5/D^b_5$	554	$C^\#_6/D^b_6$	1108		
D_5	587	D_6	1174		
$D^\#_5/E^b_5$	622	$D^\#_6/E^b_6$	1244		
$*E^d_5$	***641***	***E^d_6***	***1282***		
E_5	659	E_6	1318		

Although the Shabbaba tones are theoretically form two octaves of an Arabian scale, but during performance the Shabbaba player fits his blow into the instrument in a way that lets him/her perform just one polyphonic octave each of its tones is a harmony of the tone and its 8^{th} (octave tone). The Nay player imitates the Shabbaba by following this, mostly when performing traditional songs or non-rhythmic melodies (Taqaseem). Frequencies of half flat tones are estimation. Even though the whole interval in music is divided to 9-Kuma, the ¾ interval has different standards according to Arabian regions. For example; It is 7-Kuma in Al-Sham region (Syria, Lebanon, Jordan and Palestine), 6-Kuma in Egypt, 8-Kuma in Maghreb (Morocco, Algeria, Tunisia and Libya), and it also differs in Iraq according to the geographical region inside the country. The half flat tone is therefore given the average value of its preceding- and following-tone frequencies.

III. MIR SYSTEM OVERVIEW

Music Information Retrieval (MIR) process is essentially based upon AMT process and database system. The AMT system accepts an acoustic input processed through several stages and ultimately generates a full score as an output [12]. Instead of score generation, the proposed MIR system that shares similar input and processing stages is capable of identifying melodies with musical features similar to these of the input. Upon prevision of a certain melody portion, the MIR system extracts its musical features and searches in a pre-prepared musical database for a melody that matches the extracted features-pattern of the input. The matched melody which represents the desired output will then be retrieved from the musical database [13]. Simplified block diagram for the proposed system is shown in Fig. 1.

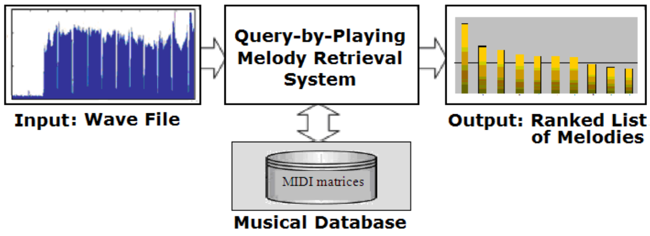
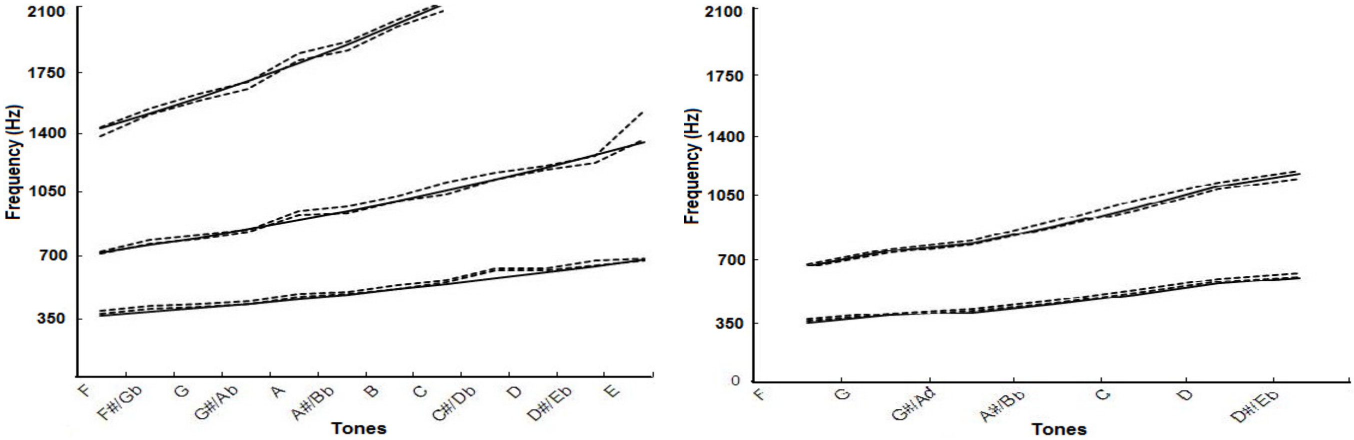


Fig. 1. Simplified block diagram of the MIR system.

A comprehensive musical database for all tones that could be played by Nay and Shabbaba are initially prepared and used in the following analysis. This database consists of 37 Nay tones and 14 Shabbaba tones. Each of these reference tones is chosen to be 5s long and is carefully played and recorded to ensure maximum quality of reference tones. Each tone is then segmented into 100 samples, each of 50 ms period. Cross correlation and Fourier analysis are then applied on the obtained samples. The obtained results indicated that Fourier analysis is more accurate and simpler than the cross correlation for pitch detection. Such results are also found in correlation with the findings that were previously reported in references [5], [11]. It was mentioned that cross correlation in AMT is usually used for separation process, which is to separate different instruments sounds but Fourier analysis is more accurate only for segmentation of one instrument's sound, especially for the flute.

IV. TIME-FREQUENCY ANALYSIS

Standard and experimentally measured frequency ranges for the Nay and Shabbaba chromatic scales are investigated in this section. Experimental values are obtained by recording five seconds of playing while changing the instrument positioning during performance with the aim of obtaining lowest and highest frequencies. The obtained results for each octave are shown in Fig. 2(a). In this figure, the standard frequency tone is presented in bold lines while the experimentally measured low and high frequency tones are presented in dotted lines, as illustrated. In the first octave, standard frequencies are mostly below measured frequencies. Besides, the tone D^b could not be distinguished from the tone D because the instrument is not capable of playing the former tone unless under certain circumstances; such as professionalism and slow performance. In the second and third octaves, the standard frequency lines fit in between the upper and lower measured lines, except for A and E. The latter tones are performed by closing half of the Nay hole, which bounds accuracy by the circumstances mentioned above. Shabbaba follows Nay "Nawa" in results but since there is no need to close half of a Shabbaba holes to perform the tones A and E, all tones in the second octave fit in between measured lines as shown in Fig. 2(b).



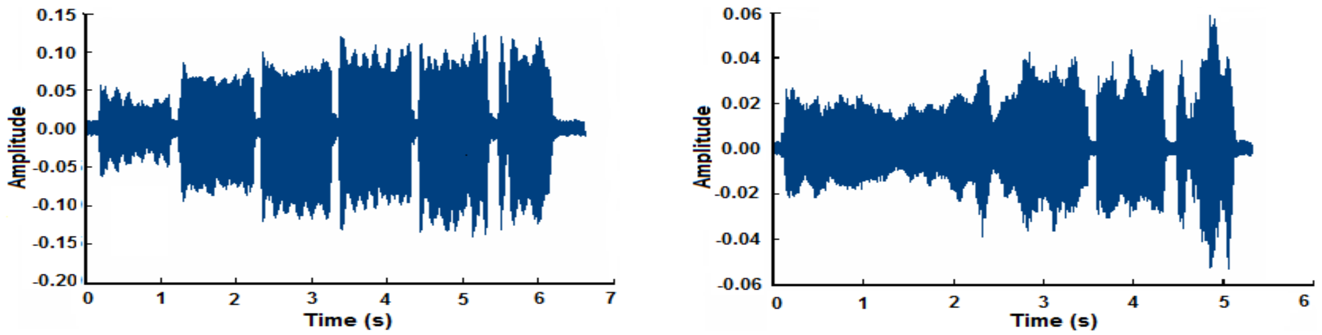
(a) Nay (b) Shabbaba

Fig. 2. Frequency ranges for the Nay and Shabbaba tones.

The time-frequency features of the Nay and Shabbaba tones are compared by playing the first six tones of the common Arabian scale "Bayat" using both instruments. The power spectrums and spectrograms of the Nay and Shabbaba are shown in Fig. 3. Looking at the power spectrums, it can be noticed that the sound of both instruments goes stronger as tones go up. The uniqueness of the harmonized Shabbaba temper is confirmed by its spectral analysis as shown in Fig. 4. For the Nay, it is obvious that the fundamental frequency of this tone is the most repeated as illustrated in Fig. 4(a). In contrast, the Shabbaba tone has two repeated fundamental frequencies, as illustrated in Fig. 4(b). These fundamental frequencies are belonging to A^d_4 and A^d_5.

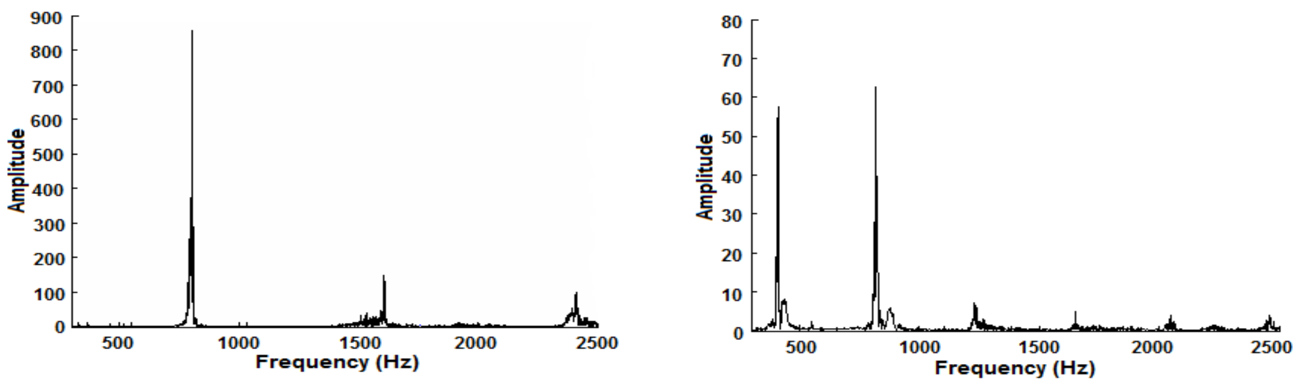
V. THE MIR PROCESS

Natural music signals, which are pseudo-periodic, can be modelled by a strictly periodic signal time-warped by an invertible function. They repeat, but each cycle is not exactly the same as the previous, and the cycles tend to change in a smooth way over time [14]. The pitch detection of a signal is therefore not as easy as detecting the period of oscillation. The proposed MIR process involves several stages shown in the simplified block diagram of Fig. 5. Of these, the pitch detection of input tones is the most challenging issue. The stages of MIR process can be summarized as follows.



(a) Nay (b) Shabbaba

Fig. 3. Power spectrums of the Nay Nawa and Shabbaba tones.



(a) Nay (b) Shabbaba

Fig. 4. Spectral analysis of the Nay Nawa and Shabbaba tones.

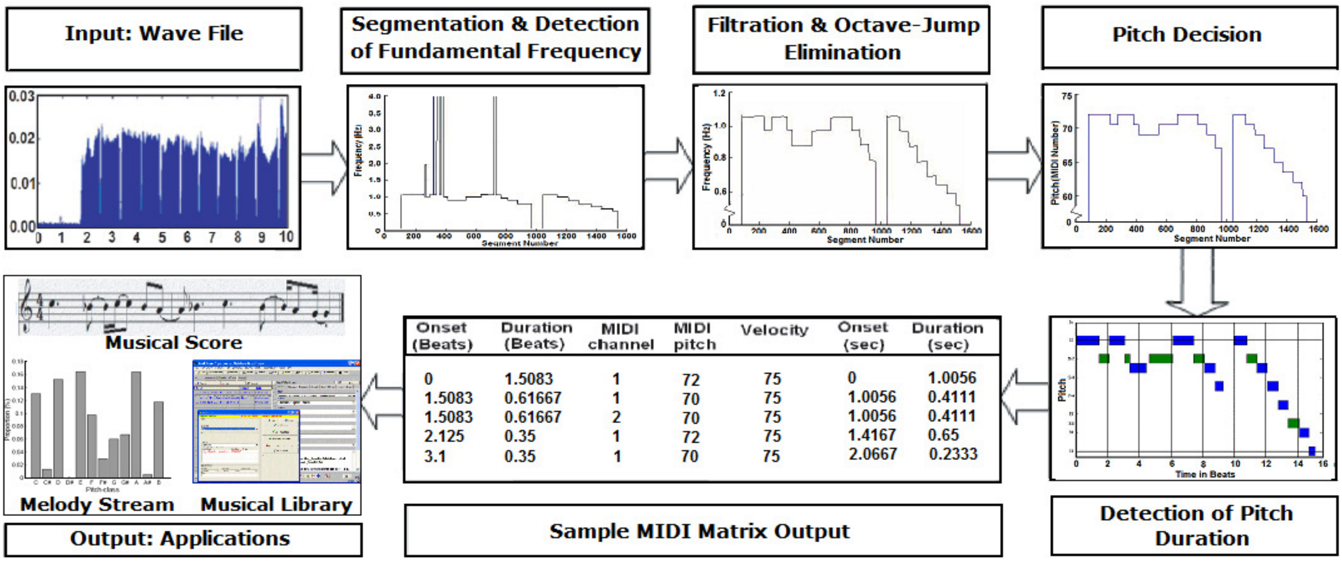


Fig. 5. Stages of the MIR process.

A. Segmentation and Fundamental Frequency Detection

The tone segmentation and fundamental frequency detection are performed using the following algorithm [5]:

{

(a) Extract one segment of length *x* (ms), designated (*f*).

(b) *F*← fast Fourier transform of (*f*).

(c) Fourier Spectrum (*FS*) ← abs (*F*).

(d) Find the index (frequency) at the maximum *FS*.

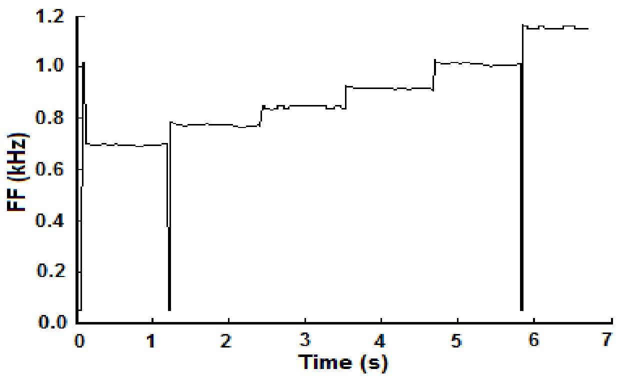
}

The segment length is initially set to 50 ms which was found optimal for a pure time-frequency analysis. It should be noted here that careful selection of the segment length is an essential requirement as it will affect performance of the following stages; especially pitch decision and identification of tone duration. Sample output of this stage is shown by the time-frequency plot of Fig. 6. This figure demonstrates the capability of the proposed system in detecting both the Nay and Shabbaba pitches. Yet, it insists on treating Shabbaba tones has one octave of polyphonic sound rather than two octaves of melodic sound. This means that the tone A^d_5 will be considered for the next processing phases irrespective of whether the detected fundamental frequency is belonging to the range of A^d_4 or A^d_5.

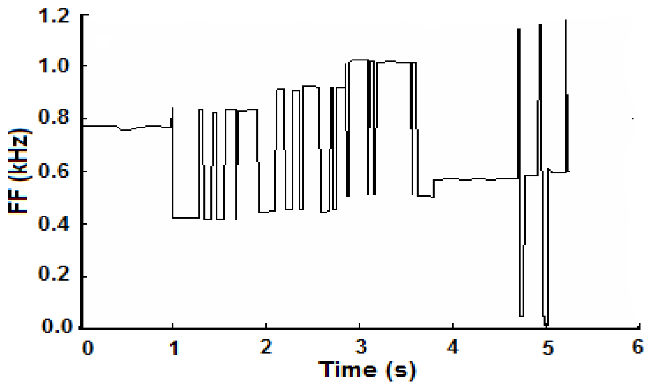
B. Filtration and Octave-Jump Elimination

The Frequency range of the Nay “Nawa” from lowest tone F_4 to highest tone C_7 is (349-2092) Hz. When the detected Fundamental Frequency (FF) of a segment is found out of this range, it is considered a miss-detected segment. Generally, there are two possible reasons for misdetection; octave jumps and non-musical noise. Experiments have shown that possibility of octave jumps increases with higher pitches. Two strategies for avoiding octave jumps are considered. The first that is related to the concept of octave jump is detecting the first harmonic rather than the FF [12]. Therefore, if the detected frequency is under or equal to the range of C_8 (4184 Hz), it will be considered as an octave jump and thus divided by two.

The second strategy is directed to those jumps appearing as pulses; an algorithm for minimum note duration is developed and discussed in the next section. Non-musical noise accompanies the musical recording ending up with a detected frequency above the highest range of an octave jump [6]. Consequently, if the detected frequency is above the range of C_8, it is considered to be a non-musical noise. After several experiments, it has been considered acceptable to assign the segment having FF above this range with the FF value of the previous segment.



(a) Nay



(b) Shabbaba

Fig. 6. Time-fundamental frequency Characteristics of Nay and Shabbaba.

C. Pitch Decision

A Look-Up-Table (LUT) is constructed for the Nay "Nawa". The frequency range of each tone is determined to be bounded by two points, the point in the middle of the interval between the desired tone and its precedent, and the point in the middle of the interval between the desired tone and its following tone. Then, training is done to adapt the table to the individual instrument used in this experiment and accordingly some modifications are made to range boundaries. Results illustrated in Fig. 2 are also taken into consideration during adaptation with the player's instrument. As shown in Table 2, the critical regions are those where Arabian quarter tones exist (the tone range is narrower than that of the occidental music tones). It can be noticed that higher pitches have wider frequency ranges and thus have improved accuracy. This demands musical performing accuracy to be within acceptable limits. Moreover, the importance of adapting the system to the user's personal instrument arises here.

Table 2. Sample of the LUT for Nay "Nawa".

Tone	Frequency Range (Hz)	Tone	Frequency Range (Hz)
F_5	680-720	B_5	960-1017
$F^{\#}_5/G^b_5$	720-762	C_6	1017-1077
G_5	762-807	$C^{\#}_6/D^b_6$	1077-1142
$G^{\#}_5/A^b_5$	807-843	D_6	1142-1210
$\underline{\boldsymbol{A^d_5}}$	843-868	$D^{\#}_6/E^b_6$	1210-1263
A_5	868-906	$\underline{\boldsymbol{E^d_6}}$	1263-1300
$A^{\#}_5/B^b_5$	906-960	E_6	1300-1355

The pitch decision process can be mathematically formulated as follows:

$$Tone_n(f_o) = \begin{cases} silence & f_o < 325 \\ LUT_{value} & 325 \le f_o \le 2155 \\ tone\left(\frac{f_o}{2}\right) & 2155 < f_o \le 4310 \\ tone_{n+1}(f_o) & 4310 < f_o \end{cases} \quad (1)$$

f_o = *fundamental frequency.*
n = *segment number*

D. Pitch Duration

An algorithm is designed and implemented to count all sequential similar tone segments and end up with segments count. The main objective of this algorithm is to discard tones of possible octave jumps appearing as pulses and pitches with sequential segments count less than minimum note duration [15]. The variable, minimum note duration, is configured in an inverse relationship with playing tempo. The duration of a discarded tone is added to the following note and therefore errors concerning spurious notes are eliminated. The proposed algorithm is given as follows:

{
(a) Count (n) = ∑ (n^{th} similar sequential tones)
(b) Segment time = 1/ ∑ (overlapped segments per second)
(c) Duration (n) = count (n) × segment time
(d) If duration (n) < minimum note duration:
- Duration (n) = 0
- Duration (n+1)= Duration (n) + Duration (n+1)
}

E. MIDI Matrix Creation

In this stage, onset and duration values of each pitch are used to create a MIDI matrix. MIDI files contain instructions detailing what notes to play, order, duration and volume of each note. As the MIDI matrix cannot encode Arabian pitches, the authors suggested the assignment of each pitch with its nearest standard MIDI pitch value in first MIDI channel (for example: $\boldsymbol{A^d}$ is assigned with $\boldsymbol{A^b}$ MIDI value and $\boldsymbol{E^d}$ is assigned with $\boldsymbol{E^b}$). Such assignment is repeated for the second MIDI channel to distinguish Arabian tones from standard MIDI tones. Velocity or loudness is assigned one value (75) for all pitches as dynamics are not of a major concern in the present work. The MIDI tool box library [16] is then used to form the Arabian MIDI matrix as a MIDI file.

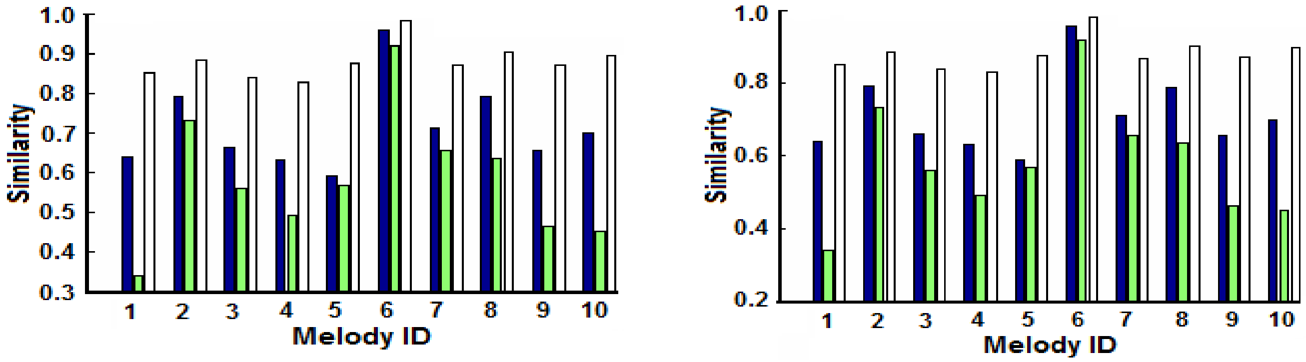
VI. DATABASE AND MELODY RETRIEVAL

The resulted MIDI matrix is ready now to be used for several applications such as obtaining the musical score of the performed melody, composing accompaniment lines for the retrieved score [17], analyzing pitch classes or intervals within the musical stream, and retrieving musical information from digital melody indexed libraries. The later application, "query-by-playing", is experimented in this paper for both the Nay and the Shabbaba. An experiment aims to retrieve Nay and Shabbaba melodies from MIDI databases are performed. Such a "query-by-playing" experiment for the Nay is presented by saving 10 MIDI matrices corresponded to 10 well-known Jordanian traditional songs. The Nay player is then played the exact melody of one of these songs in front of a simple commercial microphone. The system is to analyze the signal, compose its MIDI matrix and compare it with all matrices saved in the database. The most relevant matrices are retrieved in order. The same steps are repeated for Shabbaba. In this work, experiments focus on recognition of the melody pattern while searching within the MIDI database will be the subject of a future research.

A relatively small but sufficient musical database is used to test the performance of the proposed system. This is justified since all of the chosen melodies are belonging to the same genre (Jordanian Folklore). These melodies have a common limited tone range that does not exceed the dominant tone of the Arabian scale "Bayat". Moreover, all melodies are relatively short and follow a smooth melody stream with no wide intervals in general. This yield to a high level of melodies similarity and consequently the information retrieval process will be more challenging.

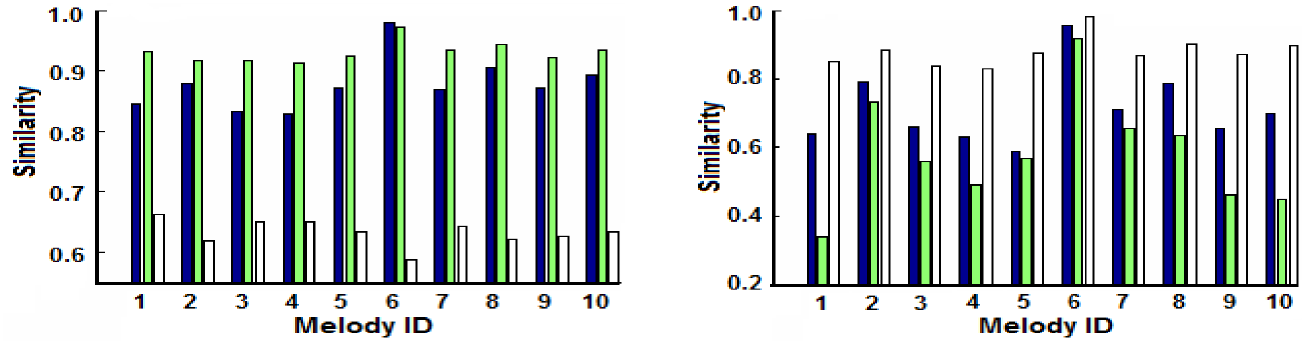
VII. RESULTS AND DISCUSSIONS

Benefiting from MIDI Tool Box [16], two tests are done for each instrument. In each test and for MIDI matrices comparison reasons, either the representation or the distance metric is left constant and the other is variable. In the first test, the metric and its taxicab norm are constant, but the representation is variable. It is tested for three representations: contour, pitch class, and interval. These representations are all succeeded to retrieve the played melody as illustrated in Fig. 7. Yet, depending on a musical perspective, the contour representation is the most convenient when the retrieval process is done by playing part



(a) Nay (b) Shabbaba

Fig. 7. Taxicab norm versus pitch classes, intervals and melodic contour.



(a) Nay (b) Shabbaba

Fig. 8. Melodic contour versus taxicab, Euclidean and cosine.

of the melody rather than playing it as a whole [13]. Therefore, in the second test the representation and its contour are constant while the metric is variable. It is also tested for metrics: taxicab norm, Euclidean distance and cosine of the angle between vectors. All metrics are also succeeded to retrieve the played melody as illustrated in Fig. 8. Yet, the taxicab norm provided the best results in terms of similarity percentage and distance between the played melody and most relevant melody to it. Success in this challenging experiment assures the possibility to retrieve other Arabian melodies belonging to other genres.

REFERENCES

[1] C. Xu, N. C. Maddage and X. Shao, "Automatic music classification and summarization," *IEEE Trans. on Speech and Audio Processing*, Vol. 13, No. 3, May 2005, pp. 441 – 450.

[2] F. M. Michael, "Music information retrieval," *IAML - IASA Congress*, Olso, August 12, 2004.

[3] J. Sounders, "Real-time discrimination of broadcast speech/music," *Proc. ICASSP96*, Vol. 2, Atlanta, GA, 1996, pp. 993–996.

[4] E. Scheirer and M. Slaney, "Construction and evaluation of a robust multi-feature music/speech discriminator," *Proc. ICASSP97*, vol. 2, 1997, pp. 1331–1334.

[5] M. A. Al-Taee, M. S. Al-Rawi and F. M. Al-Ghawanmeh, "Time-frequency analysis of the Arabian flute (Nay) tone applied to automatic music transcription," *Proc. ACS/IEEE Int. Conf. on Computer Systems and Applications (AICCSA'2008)*, Doha – Qatar March 31 – April 4, 2008.

[6] M. A. Al-Taee and F. M. Al-Ghawanmeh, "Analysis and features recognition of Al-Nay musical tones," *Proc. 6th Arabian Conf. on Modern Arts*, Irbid-Jordan, October 19-22, 2008.

[7] M. S. El-Farjani, *Scales of Arabian Music*. Al-Jamaheriah Publishing Company, 1st ed., Tripoli, 1986.

[8] M. T. Al-Ghawanmeh, *Traditional Jordanian Melodies*, Rozana publishing company, Irbid-Jordan, 1997.

[9] A. Abdulla, "Traditional woodwind instruments and their famous players in Iraq," *Conf. Proc. of the Arab Society for Music*, Tunis – Tunisia, 2008.

[10] M. T. Al-Ghawanmeh, "Traditional woodwind instruments in Jordan," *Conf. Proc. of the Arab Society for Music*, Tunis – Tunisia, 2008.

[11] W. Birmingham, R. Dannenberg and B. Pardo, "Query by hamming with the vocal search system", *Communication of the ACM*, Vol. 49, No. 8, 2006, pp. 49-52

[12] G. Gerhard, *Computer Music Analysis*, School of Computing Science, Simon Fraser University, Burnaby, Technical Report CMPT TR 97-13, 2002.

[13] R. J. McNab, L. A. Smith, D. Bainbridge D. and L. H. Witten, "The New Zealand Digital Library Melody index," *D-Lib Magazine*, May 1997

[14] C. Dan, S. Yoram and M. David, "Optimal multi-pitch estimation using the EM algorithm for co-channel speech separation," *IEEE-ICASSP*, 1993, Vol. 2, pp. 728-731.

[15] J. Bello, G. Monti and M. Sandler, "Techniques for automatic music transcription". *Proc. of the Int. Symposium on Music Information Retrieval*, Music IR 2000, Plymouth, Massachusetts, October 23-25, 2000.

[16] T. Eerola and P. Toiviainen, *MIDI Toolbox: MATLAB Tools for Music Research.* University of Jyväskylä: Kopijyvä, Jyväskylä, Finland, 2004.

[17] M. T. AL-Ghawanmeh, R. Haddad and F. M. AL-Ghawanmeh, "The appliance of music information retrieval system for Arabian woodwinds in e-learning and music education," *Proc. 3rd Int. Conf. of the Greek Association of Primary Music Education Teachers*, Athens, Greece, 8-10 May, 2009.