SINGING PITCH EXTRACTION FROM MONAURAL POLYPHONIC SONGS BY CONTEXTUAL AUDIO MODELING AND SINGING HARMONIC ENHANCEMENT

Chao-Ling Hsu[1] **Liang-Yu Chen**[1] **Jyh-Shing Roger Jang**[1] **Hsing-Ji Li**[2]

[1]MediaTek-NTHU Joint Lab
Dept. of Computer Science,
National Tsing Hua Univ., Taiwan
{leon, davidson833, jang} @mirlab.org

[2]Innovative Digitech-Enabled Applications &
Services Institute (IDEAS)
Institute for Information Industry, Taiwan
lihsingji@iii.org.tw

ABSTRACT

This paper proposes a novel approach to extract the pitches of singing voices from monaural polyphonic songs. The hidden Markov model (HMM) is adopted to model the transition between adjacent singing pitches in time, and the relationships between melody and its chord, which is implicitly represented by features extracted from the spectrum. Moreover, another set of features which represents the energy distribution of the enhanced singing harmonic structure is proposed by applying a normalized sub-harmonic summation technique. By using these two feature sets with complementary characteristics, a 2-stream HMM is constructed for singing pitch extraction. Quantitative evaluation shows that the proposed system outperforms the compared approaches for singing pitch extraction from polyphonic songs.

1. INTRODUCTION

Melody, usually represented by the pitch contour of a lead vocal in a song, is considered as one of the most important elements of a song. It is broadly used in various applications, including singing voice separation, music retrieval, and musical genre classification.

Since Goto [1] proposed the first melody extraction system by employing a parametric model trained by statistical methods in 1999, more and more work has been proposed in the literature [2-8]. Because harmonic structures of a singing voice are very noticeable in spectrogram even in a polyphonic song, they are commonly used as cues for extracting the singing melody [1][4-6]. However, they neglect the contextual information of music.

Ryynänen et al. [8] used both acoustic and musicological models to generate hidden Markov models (HMMs) for a singing melody transcription system. The musicological models determine the transition probabilities between the adjacent notes. Li et al. [2] also utilized an HMM where the transition probability was estimated from the labeled training data. However, they only considered the transition between adjacent notes; the concurrent pitches generated by other musical instruments, such as chords, were not considered.

While the concurrent pitches are usually the obstacles in singing pitch extraction, we try to utilize them as the cues to extract the melody. Generally speaking, melody is composed of a series of notes and is decorated by chords. The chords here represent the concurrent pitches accompanying the melody. These notes and chords progress according to some underlying music rules to make the song euphonious. Therefore, we use an HMM to learn these rules from actual song data by observing their spectrograms. Note that we do not identify the chords explicitly. Instead, we use the energy distribution of each semitone to train the contextual audio model. In addition, in order to utilize the harmonics information as cues to extract the singing pitches, we also model the energy distribution of harmonics by using the proposed normalized sub-harmonic summation (NSHS) to enhance the harmonic structures of the sound sources especially for those of the singing voices. By synergizing these two techniques, the accuracy of singing pitch extraction is improved significantly.

The rest of this paper is organized as follows. Section 2 describes the proposed system in detail. The experimental results are presented in section 3, and section 4 concludes this work with possible future directions.

2. SYSTEM DESCRIPTION

Fig. 1 shows the overview of the proposed system. Two streams of features are extracted from the spectrogram and the NSHS map, respectively, of the input polyphonic song. A 2-stream HMM is then employed to decode the input songs into the most likely unbroken pitch vectors. On the other hand, the MFCCs (Mel-frequency cepstral coefficients) are extracted to perform the voiced/non-voiced detection. Lastly, the singing pitch vectors are produced by integrating the results of these two processes. The following subsections explain these blocks in detail.

2.1 Features Extraction from a Spectrum

This block extracts two types of features, including MFCCs and ESI (Energy at Semitones of Interests).

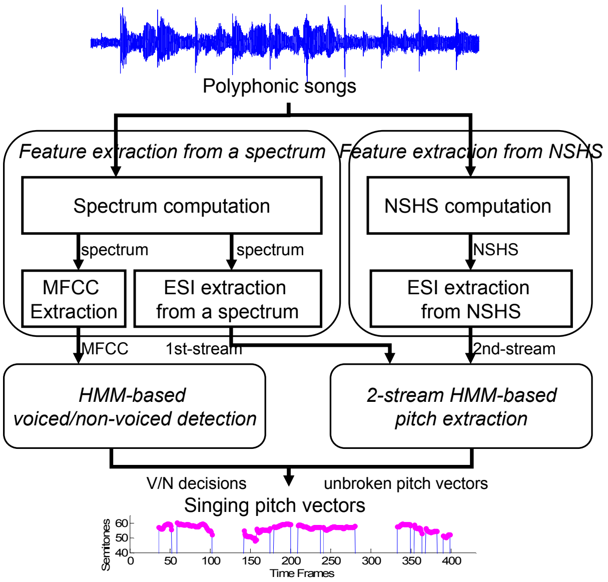

Figure 1. System overview

MFCCs are the features for a 2-state HMM for voiced/non-voiced detection. ESI is the 1st-stream feature for a 2-stream HMM for pitch extraction. Since most of the songs nowadays follow the twelve-tone equal temperament, it is intuitive to employ semitone scale to model the relations between melody and chords. For each integer semitone of interests within the range $[40,72]$, we identify its maximum energy as an element of the feature vector. Take semitone 69 for example, the search range in semitone is $[68.5,69.5]$, corresponding to a frequency bin of $[427.47,452.89]$ in terms of Hertz. Then we find the maximum power spectrum within this range as the feature associated with semitone 69. Since there are 33 elements within semitone of interests, the length of the feature vector of ESI is also 33.

More specifically, the ESI computed from a spectrum in the time frame t can be obtained as follows:

$$v_t(m) = \max_{f_m - \frac{f_m - f_{m-1}}{2} \le f < f_m + \frac{f_{m+1} - f_m}{2}} \left(P_t(f)\right), \tag{1}$$

where $P_t(*)$ is the power spectrum calculated from short time Fourier transform (STFT), $m = 0,1,..,M-1$, M is the total number of semitones that are taken into account, and f_m is the frequency of mth semitone in the selected pitch range.

Note that we also need to record the maximizing frequency within each frequency bin in order to reconstruct the most likely pitch contours.

2.2 HMM-based Voiced/Non-voiced Detection

This block employs a continuous 2-state HMM to decode the mixture input into voiced and non-voiced segments, similar to the one proposed by Fujihara et al. [9]. Note that the "voiced" here indicates the voiced singing voice, and "non-voiced" indicates the unvoiced singing voice and music accompaniments. Given the MFCC feature vectors $X = \{x_0, \cdots, x_t, \cdots\}$ of the input mixtures, the problem is to find the most probable sequence of voiced/non-voiced states, $\hat{S} = \{s_0, \cdots, s_t, \cdots\}$:

$$\hat{S} = \arg\max_{S} \left\{ p(s_0)p(x_0 \mid s_0) \prod_t \{p(s_t)p(x_t \mid s_t)p(s_t \mid s_{t-1})\} \right\}, \tag{2}$$

where $p(x \mid s)$ is the output likelihood of a state s, $p(s_t \mid s_{t-1})$ is the state transition probability from state s_{t-1} to s_t, and $p(s_t)$ is the prior of the state s_t. Note that $p(s_t \mid s_{t-1})$ and $p(s_t)$ can be obtained from the actual song data with manual annotations.

2.3 Features Extraction from NSHS

This block extracts the 2nd-stream feature vector which represents the energy distributions of the enhanced harmonic structures of singing voices. The harmonic structures can be enhanced by sub-harmonic summation (SHS) proposed by Hermes [10]:

$$H_t(f) = \sum_{n=1}^{N} h_n P_t(nf), \tag{3}$$

where $H_t(f)$ is the sub-harmonic summation value of the frequency f at time frame t, $P_t(*)$ is the power spectrum calculated from STFT, n is the index of harmonic components, N is the number of the harmonic components in consideration, and h_n is the weight indicating the contribution of the nth harmonic component. Usually we set $h_n = h^{n-1}$, where $h \le 1$. In order to further enhance the harmonics of singing voices, we propose the use of normalized SHS (NSHS) defined as follows:

$$\hat{H}_t(f) = \frac{\sum_{n=1}^{N_f} h_n P_t(nf)}{\sum_{n=1}^{N_f} h_n}, \tag{4}$$

where the number of harmonic components N_f depend on the frequency under consideration:

$$N_f = floor\left(\frac{0.5 f_s}{f}\right), \tag{5}$$

with f_s being the sampling rate. The reason of the modification is based on the observation that most of the energy in a song in located at the low frequency bins, and the energy of the harmonic structures of the singing voice seems to decay slower than that of instruments [2]. Therefore, when more harmonic components are considered, energy of the vocal sounds is further strengthened. Although some percussive instruments (e.g. cymbals)

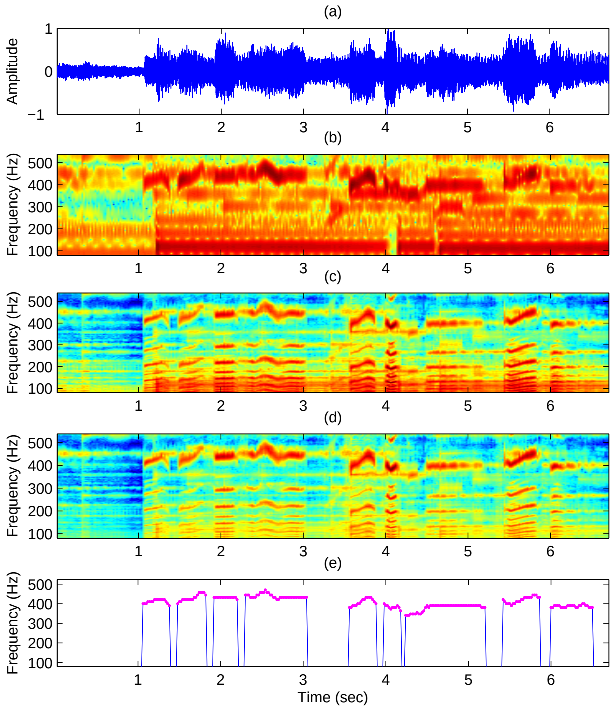

Figure 2. The energy distributions of a sample clip Amy_4_05 in MIR-1K dataset at 0 dB SNR. The distributions are computed within the frequency range [80.06, 538.58], or [39.5, 72.5] in terms of semitones. (a) The waveform of the mixture. (b) The spectrogram. (c) The SHS map. (d) The proposed NSHS map. (e) The manually labeled pitch vector of the singing voice.

present high energy at higher frequency bins, their non-harmonic nature does not affect the NSHS much.

Figure 2 illustrates the energy distributions of a traditional spectrogram, original SHS map, and the proposed NSHS map. By comparing the spectrogram in 2(b) and the SHS map in 2(c), it is obvious that most of the energy of accompaniments in the spectrogram is attenuated in the SHS map. However, the energy in the lower frequency bins remains high. The proposed NSHS map shown in 2(d) further attenuates the low-frequency energy and enhances the sub-harmonic structure of the singing voice. As a result, after the enhancement by the NSHS map, the pitch of the singing voice can be extracted much easier.

Based on the proposed NSHS, we can extract a 33-element feature vector of ESI for each given frame, as explained in Section 2.1. The feature vector is sent to the 2-stream HMM for pitch extraction.

2.4 2-Stream HMM-based Pitch Extraction

We employ a 2-stream HMM to model the relationship between the adjacent melody pitches and their corresponding audio context. Given the 1st-stream ESI feature vectors $V = \{v_0, \cdots, v_t, \cdots\}$ from spectrogram and the 2nd-stream ESI feature vectors $C = \{c_0, \cdots, c_t, \cdots\}$ from NSHS map, our goal is to find the most likely sequence of pitch states, $\hat{R} = \{r_0, \cdots, r_t, \cdots\}$:

$$\hat{R} = \arg\max_{R} \left\{ p(r_0) p(v_0, c_0 \mid r_0) \prod_t \{ p(r_t) p(v_t, c_t \mid r_t) p(r_t \mid r_{t-1}) \} \right\}, \tag{6}$$

where $p(r_t \mid r_{t-1})$ is the state transition probability from pitch state r_{t-1} to r_t, $p(r_t)$ is the prior of the pitch state r_t, and $p(v, c \mid r)$ is the joint output likelihood of the pitch state r defined as:

$$p(v, c \mid r) = p_v(v \mid r) p_c(c \mid r), \tag{7}$$

where $p_v(v \mid r)$ and $p_c(c \mid r)$ are the state likelihoods of feature vectors v and c, respectively, given the state r. This is a typical multi-stream HMM which is broadly used in speech processing [11]. The state likelihoods (or conditional observation likelihoods), transition probabilities, and priors of eq. (6) and (7) can all be obtained from the actual song data with manually annotated pitch contours.

Figure 3 shows the benefits of applying a 2-stream HMM instead of using a single-stream feature from either the spectrum or the NSHS. Each of the plots is a state-frame likelihood table where the vertical axis indicates the pitch state of each semitone and horizontal axis indicates time frames. The likelihood is computed for each state and time frame. All likelihood in the same time frame is normalized to zero mean and unity variance for better visualization. The ideal singing pitches are overlaid as solid lines. Figure 3(a) shows $p_v(v \mid r)$ of each state which utilizes audio context as cues to extract the singing pitches. Figure 3(b) shows $p_c(c \mid r)$ of each state which indicates the likelihood that an enhanced singing harmonic structure is presented, and Figure 3(c) shows the joint likelihood $p_v(v \mid r) p_c(c \mid r)$. Figure 3(d) and (e) illustrate the overall maximum likelihoods (up to a given frame time and pitch state) of single-stream HMMs using feature vectors V and C, respectively. More specifically, the value of each point in the figure represents the maximized accumulated likelihood of the previous pitch states sequence including the transition probabilities. Again, for better visualization, each column in these two tables is normalized to have zero mean and unity variance. The joint likelihood using the 2-stream HMM is shown in Figure 3(f). It can be observed that the likelihood of the singing pitch states at around 1.2 and 5.6 seconds are low in Figure 3(d), but they are recovered in Figure 3(f) by combining with the likelihood in Figure 3(e). In addition, the likelihood of the states that are not corresponding to the singing pitches between 1.5 and 5.3 seconds in Figure 3(e) are diminished in Figure 3(f) as well. Furthermore, both single-stream HMMs exhibit high likelihood for the singing pitch states. Therefore, after combining the likelihood using the 2-stream HMM, the likelihood of the singing pitch states are higher than that of the other states, and the pitches of singing voices can thus be extracted more accurately.

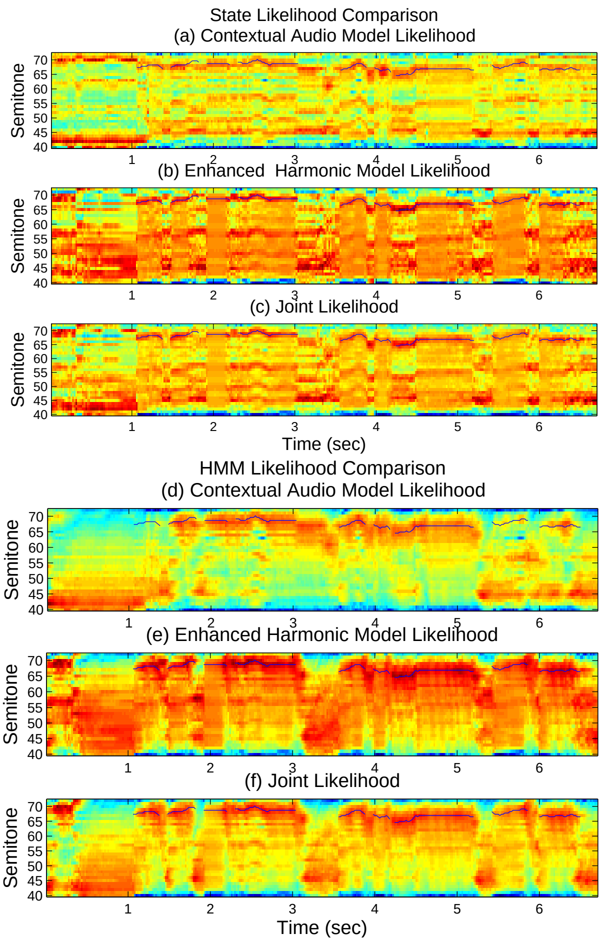

Figure 3. The state likelihood and HMM likelihood comparison for the clip Amy_4_05 in MIR-1K dataset at 0 dB SNR. (a) to (c) and (d) to (f) show the likelihood of contextual audio model, the likelihood of enhanced harmonic model, and the join likelihood of state likelihood and HMM likelihood, respectively. The solid line indicates the manually labeled pitch vector of the singing voice.

3. EVALUATION

Two datasets were used to evaluate the proposed approach. The first one, MIR-1K[1], is a publicly available dataset proposed in our previous work [12]. It contains 1000 song clips recorded at 16 kHz sample rate with 16-bit resolution. The duration of each clip ranges from 4 to 13 seconds, and the total length of the dataset is 133 minutes. These clips were extracted from 110 karaoke songs which contain a mixed track and a music accompaniment track. These songs were selected (from 5000 Chinese pop songs) and sung by our colleagues in the lab, consisting of 8 females and 11 males. Most of the singers are amateurs with no professional training. The music accompaniment and the singing voice were recorded at the left and right channels, respectively. The second dataset, called commercial set for short, contains 178 song clips from commercial CDs, and the total length of the dataset is about 25 minutes. The ground truth of the voiced/non-voiced segments and pitch values of the singing voices were first estimated from the pure singing voice and then manually adjusted for these two datasets.

	MIR-1K	Commercial set
Precision	87.48 %	91.14 %
Recall	86.03 %	91.78 %
Overall accuracy	81.52 %	87.12 %

Table 1. Performance of voiced/non-voiced detection

All songs are mixed at 0 dB SNR, indicates that the energy of the music accompaniment is equal to the singing voice. Note that the SNRs for commercial pop songs are usually larger than zero, indicating that our experiments were set to deal with more adversary scenarios than the general cases.

3.1 Evaluation for Voiced/Non-voiced detection

The evaluation was performed via two-fold cross validation with the MIR-1K dataset. The dataset was divided into two subsets of similar sizes (487 vs. 513, recorded by disjoint subjects). In addition, the commercial set was also evaluated by using all MIR-1K for training. The reason for not using the commercial set for training the voiced/non-voiced model is because its size is too small.

39-dimensional MFCCs (12 cepstral coefficients plus a log energy, together with their first and second derivatives) were extracted from each frame. The MFCCs were computed from STFT with a half-overlapped 40-ms Hamming window. Cepstral mean subtraction (CMS) was used to reduce channel effects.

Two 32-component GMMs were trained for voiced frames and non-voiced frames, respectively. All GMMs had diagonal covariance matrices. Parameters of the GMMs were initialized via k-means clustering algorithm and were iteratively adjusted via expectation-maximization (EM) algorithm with 30 iterations. Each of the GMMs was considered as a state in a fully connected 2-state HMM, where the transition probabilities and the weight of each GMMs were obtained through frame counts of the labeled dataset. For a given input song mixture, Viterbi algorithm was used to decode the mixture into voiced and non-voiced segments.

Table 1 shows the performance of voiced/non-voiced detection. The precision is the percentage of the frames that are correctly classified as voiced over the frames that are classified as voiced. The recall is the percentage of the frames that are correctly classified as voiced over all the voiced frames. The effects of the results will be discussed in the following subsections.

3.2 Evaluation for Singing Pitch Extraction

The MIR-1K dataset was divided into two subsets in the same way as subsection 3.1 for two-fold cross validation, and the commercial set was evaluated by using all MIR-1K for training. The spectrum of each frame was com-

[1] The MIR-1K dataset is available at http://unvoicedsoundseparation.googlepages.com/mir-1k

puted from STFT with a half-overlapped 40-ms window and zero padding to 2^{14}. In addition, the pitch range for computing ESI (for both spectra and NSHS) was [40-0.5, 72+0.5] in semitones or [80.06, 538.58] in Hertz, which is similar to the common singing frequency range used in [2]. The compression factor h for computing NSHS was set to 0.99. At last, a 33-dimentional feature vector v_t from spectra and a 33-dimentional feature vector c_t from NSHS were extracted for each frame.

Two diagonal 8-component GMMs, Γ_V and Γ_C, were trained for each of the 33 semitone models by using feature vectors v_t and c_t, respectively. Parameters of the GMMs were initialized via k-means clustering algorithm and were iteratively adjusted via EM algorithm with 30 iterations. Each of the $(\Gamma_V, \Gamma_C)_m$ pairs (with $m=[0,32]$) was considered as a state in an HMM, where the transition probabilities and the prior of each GMM were obtained through frame counts of the labeled dataset. For a given input mixture, Viterbi algorithm was used to decode the mixture into a sequence of pitch states $\hat{R}$. By tracking the maximizing frequency (which generates ESI at each semitone) for each pitch state, we can then reconstruct the optimum pitch contour.

In order to evaluate the proposed method, eight other approaches were used for comparison. For simplicity, we use SPEC and NSHS to indicate ESI that were extracted from a spectrum or a NSHS, respectively. In addition, HMM, DP, and MAX are used to indicate different schemes for extracting the singing pitches. More specifically, HMM represents the proposed HMM approach; DP represents the approach of dynamic programming over spectrum/NSHS directly (to be detailed next); MAX is simply maximum-picking over spectrum/NSHS.

The goal of the DP method is to find a path $f=\left[f_0,\cdots,f_i,\cdots,f_{n-1}\right]$ that maximizes the score function:

$$score(f,\theta)=\sum_{t=0}^{n-1} Y_t(f_t)-\theta\times\sum_{t=1}^{n-1}\left|f_t-f_{t-1}\right|, \qquad (8)$$

where $Y_t(f_t)$ is a feature vector extracted from spectrum/NSHS at the frame t and frequency f_t. The first term in the score function is the sum of energy of the pitches along the path, while the second term controls the smoothness of the path with the use of a penalty term θ (which is set to 2 in this study). If θ is larger, then the computed path are smoother. In particular, the MAX approach sets θ to be zero so that maximizing the above objective function is equivalent to maximum-picking of the features from spectrum/NSHS of each frame.

The DP method employs dynamic programming to find the maximum of the score function, where the optimum-valued function $D(t,m)$ is defined as the maximum score starting from frame 1 to t, with $f_t=m$:

$$D(t,m)=Y_t(m)+\max_{k\in[0,32]}\left\{D(t-1,k)-\theta\times\left|k-m\right|\right\}, \qquad (9)$$

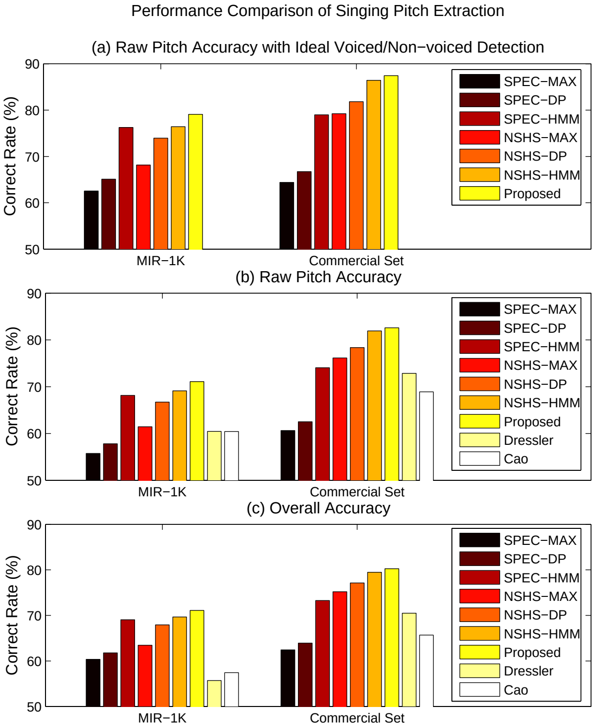

Figure 4. Performance comparison for singing pitch extraction. (a) Raw pitch accuracy with ideal voiced/non-voiced detection. (b) Raw pitch accuracy. (c) Overall accuracy.

where n is the number of frames, $t=[1,n-1]$, and $m=[0,32]$. The initial conditions are $D(0,m)=Y_0(m)$, and the optimum score is equal to $\max_{m\in[0,32]} D(n-1,m)$.

Moreover, the “Dressler” approach indicates a melody extraction method proposed by Dressler [4] which ranked first from 2005 to 2006 in the MIREX task of audio melody extraction. We obtained the software from her for comparison purpose. The “Cao” approach indicates the method proposed by Cao et al. [5], which was re-implemented by us for comparison.

Figure 4 shows the performance comparison for the singing pitch extraction. Figure 4(a) shows the raw pitch accuracy with ideal voiced/non-voiced detection, where the correct rate is computed over the frames that were labeled as voiced in the reference files. Figure 4(b) shows the raw pitch accuracy with automatically detected voiced/non-voiced segments. Figure 4(c) shows the overall accuracy where all frames are taken into account for computing the correct rate. In other words, Figure 4(a) shows the performance of the singing pitch extraction alone, assuming ideal voiced/non-voiced detection. On the other hand, the Figure 4(b) and (c) shows the performance in a practical situation where the results are affected by the errors of voiced/non-voiced detection. Since Dressler’s and Cao’s method perform singing voice detection implicitly, their performance is only shown for the cases of raw pitch and overall accuracy in Figure 4(b) and (c). Note that Dressler’s method was designed not only to extract the melody from vocal songs, but also from non-vocal music which contains no singing voice, so the performance may not be as good as the other approaches that solely designed for vocal songs.

The proposed system achieved 71.10% and 80.24% overall accuracy in MIR-1K and commercial set, respectively. Experiments show that performance is significantly improved by applying the proposed HMM and by the NSHS in both datasets. Two points are worth noting. Firstly, while NSHS enhances the harmonic structures of both the singing voices and chords, the energy enhancement of chords is relatively weaker. Therefore, the improvement of using HMM over the MAX and DP approaches is much larger by using spectrum-based ESI than NSHS-based. This shows that the chord information embedded in spectrum-based ESI does help for extracting the singing pitches. Secondly, when spectrum-based ESI are replaced by NSHS-based ESI, the performance of MAX and DP is improved significantly. It shows that the NSHS does help for reducing the interference of non-singing pitches. By taking the advantages of both approaches, the proposed method therefore performs significantly better than the compared approaches.

4. CONCLUSIONS

In this paper, we propose a new singing pitch extraction system by employing a 2-stream HMM to model the relation between adjacent notes and between melody and chords. By modeling the energy distribution in spectrogram and in the proposed NSHS map, the performance is significantly improved. Besides, the improvement of the performance is quite similar in different datasets which confirms the robustness of the proposed approach.

The proposed NSHS only applies a simple weight function for harmonic components; the performance can be further improved by optimizing it with the training scheme proposed by Klapuri [13]. In addition, the raw pitch accuracy with ideal voiced/non-voiced detection of the proposed system is much higher than that of the overall accuracy (6~8%). Therefore it is also one of our future directions to improve the voiced/non-voiced detector by not only using MFCCs but also considering the voice vibrato information as proposed by Regnier et al. [14].

It is worth noting that the evaluation is performed by using our dataset, MIR-1K, which contains more song clips than that used in MIREX (less than 20 minutes, and only 7 minutes of them are publicly available). It allows researchers to evaluate and compare their systems with others easily by using the more comprehensive dataset.

5. ACKNOWLEDGEMENT

This study is conducted under the "III Innovative and Prospective Technologies Project" of the Institute for Information Industry which is subsidized by the Ministry of Economy Affairs, the Republic of China. The authors would also like to thank K. Dressler for providing her pitch tracking software.

6. REFERENCES

[1] M. Goto, “A Real-Time Music Scene Description System: Predominant-F0 Estimation for Detecting Melody and Bass Lines in Real-World Audio Signals,” *Speech Communication*, vol. 43, no. 4, pp. 311–329, 2004.

[2] Y. Li and D. L. Wang, “Detecting Pitch of Singing Voice in Polyphonic Audio,” *IEEE ICASSP*, pp. 17–20, 2005.

[3] G. E. Poliner and D. P. W. Ellis, “A Classification Approach to Melody Transcription,” *6th ISMIR*, pp. 161-166, 2005.

[4] K. Dressler, “An Auditory Streaming Approach on Melody Extraction,” *Extended abstract for 7th ISMIR*, 2006.

[5] C. Cao, M. Li, J. Liu and Y. Yan, “Singing Melody Extraction in Polyphonic Music by Harmonic Tracking,” *8th ISMIR*, 2007.

[6] V. Rao and P. Rao, “Melody Extraction Using Harmonic Matching,” *Extended abstract for 9th ISMIR,* 2008.

[7] J.-L. Durrieu, G. Richard and B. David, “An Iterative Approach to Monaural Musical Mixture De-soloing,” *IEEE ICASSP*, pp. 105-108, 2009.

[8] M. Ryynänen and A. Klapuri, "Transcription of the Singing Melody in Polyphonic Music," *7th ISMIR*, pp. 222-227, 2006.

[9] H. Fujihara, M. Goto, J. Ogata, K. Komatani, T. Ogata, and H. G. Okuno, “Automatic Synchronization between Lyrics and Music CD Recordings Based on Viterbi Alignment of Segregated Vocal Signals,” *ISM*, pp. 257–264, 2006.

[10] D. J. Hermes, “Measurement of Pitch by Subharmonic Summation,” *Journal of Acoustic Society of America*, vol.83, pp. 257-264, 1988.

[11] S. J. Young, G. Evermann, M. J. F. Gales, T. Hain, D. Kershaw, X. Liu, G. Moore, J. J. Odell, D. Ollason, D. Povery, V. Valtchev, and P. C. Woodland: *The HTK Book (for HTK version 3.4)*, Cambridge University, 2006.

[12] C. L. Hsu and J. S. Jang, “On the Improvement of Singing Voice Separation for Monaural Recordings Using the MIR-1K Dataset,” *IEEE Trans. Audio, Speech, and Language Processing*, accepted.

[13] A. Klapuri, "Multiple Fundamental Frequency Estimation by Summing Harmonic Amplitudes," *7th ISMIR*, 2006.

[14] L. Regnier and G. Peeters, “Singing Voice Detection in Music Tracks using Direct Voice Vibrato Detection,” *IEEE ICASSP*, pp. 1685-1688, 2009.