# PITCH CONTOUR SEGMENTATION FOR COMPUTER-AIDED JINGJU SINGING TRAINING

**Rong Gong, Yile Yang, Xavier Serra**
Music Technology Group
Universitat Pompeu Fabra
Barcelona, Spain
{rong.gong,yile.yang,xavier.serra}@upf.edu

## ABSTRACT

Imitation is the main approach of jingju (also known as Beijing opera) singing training through its inheritance of nearly 200 years. Students learn singing by receiving auditory and gestural feedback cues. The aim of computer-aided training is to visually reveal the student's intonation problem by representing the pitch contour on segment-level. In this paper, we propose a technique for this purpose. Pitch contour of each musical note is segmented automatically by a melodic transcription algorithm incorporated with a genre-specific musicological model of jingju singing: bigram note transition probabilities defining the probabilities of a transition from one note to another. A finer segmentation which takes into account the high variability of steady segments in jingju context enables us to analyze the subtle details of the intonation by subdividing the note's pitch contour into a chain of three basic vocal expression segments: steady, transitory and vibrato. The evaluation suggests that this technique outperforms the state of the art methods for jingju singing. The web prototype implementation of these techniques offers a great potential for both in-class learning and self-learning.

## 1. INTRODUCTION

Jingju is a traditional Chinese theater which combines music, vocal performance, mime, dance, and acrobatics. It arose in the late 18th century and became fully developed and recognized by the mid-19th century. The form was extremely popular in the Qing dynasty court and has come to be regarded as one of the cultural treasures of China [1].

Singing, one of the most basic means of performing in jingju, is fundamentally different from the music system in the West. There are currently four main role-types in jingju: *sheng*, *dan*, *jing*, *chou*. Different role-types have widely different singing styles. For example, the role of *laosheng* (elderly man) mainly uses the real voice, whereas the role of *dan* (woman) uses mainly falsetto [2].

Imitation is the main method of jingju professional singing training nowadays. During the training class, the teacher has absolute authority and his/her singing is seen as the standard to which the student's imitation should be as close as possible. Students do not have a big space for developing their own singing style until graduation.

The basic unit of jingju arias is couplets. Each couplet consists of two lines: opening line and closing line. The teacher sings firstly one line of a couplet which usually contains 7 or 10 syllables, then the student imitates. Finally, the teacher will give comments from the perspectives of intonation, rhythm, timbre, loudness and phonation. The assessing process is conducted by comparing the student's singing performance with the teacher's version on multilevel time scales which includes line-level, syllabic level, note-level and segment-level. The state of the art computer-aided singing training methods [3–6] devote most of their efforts on line-level and note-level intonation assessment because according to the research on Western music [7], the intonation is the predominant perceptual dimension for the musicians to judge the goodness of the singing performance. Similarly, intonation accuracy is also the very basic assessment dimension for jingju singing training, such that it is the research object of this work. Other assessment dimensions include phonation, rhythm, loudness and timbre.

In jingju singing assessment, the teacher would pay much attention on various vocal expressions in terms of intonation which are represented as different pitch contour segments, such as vibrato and transitory segments. Different vocal expressions will be dealt with on their characteristics in the singing assessment process. For the steady segment, we mainly evaluate its length and average pitch. For the transitory segment, we evaluate its slope, starting and ending pitches. For the vibrato, its rate, extent, length and average frequency will be evaluated.

The preliminary step of this task is to transcribe the singing performance to pitch contours of musical notes. This problem can be solved by melodic transcription algorithms [8]. Then we perform the segmentation of note's pitch contour in order to perform the comparison with a small granularity. We concentrate on three main pitch contour segment categories: steady, transitory and vibrato segments. We limit our discussion in this paper to these because they are three of the most important vocal expressions that determine the correctness of the intonation of jingju singing.

This paper is organized as follows: Section 2 presents an overview about the related techniques. Section 3 de-

scribes the jingju singing specific approach of pitch contour segmentation. Section 4 presents the dataset and the evaluation of these techniques. Section 5 introduces a web prototype for jingju singing training and Section 6 draws the conclusion of this work.

## 2. BACKGROUND

Mauch [9] defined melodic transcription as symbolizing the pitch (or fundamental frequency, $f_0$) track to the stable segments containing pitch and duration features, that is to say, the algorithms compute a time-pitch representation which needs to be further processed in order to detect note events with a discrete pitch value, an onset time and an offset time [8]. [9] introduced a melodic transcription algorithm whose first stage - pYIN is a modification of YIN algorithm [10] which outputs multiple pitch candidates together with their probabilities and then finds a smooth path through the candidates by using a HMM. The note tracking stage is performed as Viterbi-decoding of an HMM which simplified the approach of Viitaniemi [11] by replacing the observation probability distribution with a simple Gaussian Distribution and removing the duration model. In addition, it doesn't deal with notes quantised to the integer MIDI scale which allows a more fine-grained analysis and makes this simplified approach more prone to the singing performance evaluation.

Recently, the increasing interest of the MIR community in the application of music analysis techniques to non-Western music has drew attention to the fact that different musical genres necessitate different analysis techniques [12]. The incorporation of genre-specific knowledge such as rhythmic structure and tonality improves the transcription accuracy [13]. Musicological model of [11] which includes key signature probabilities and bigram note transition probabilities is also the implementation of this concept. However, because occurrence probabilities of different note values given the key have not been estimated for jingju singing, the genre-specific tonality knowledge cannot be directly used.

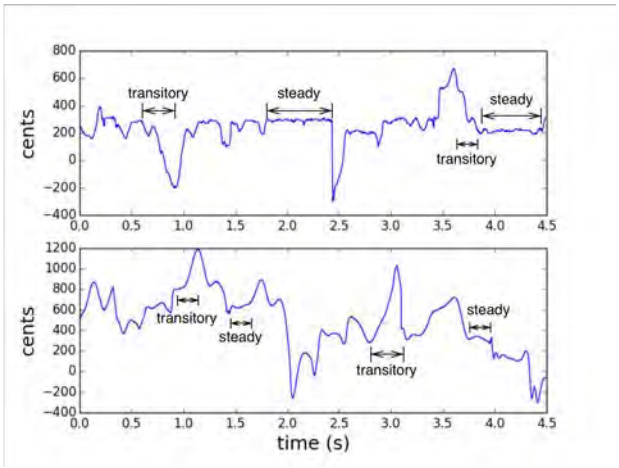


**Figure 1**. Pitch contour examples of Hindustani vocal (upper) and jingju vocal (lower). 0 cents is 261.6 Hz (C4).

The related problem of pitch contour segmentation has been studied for Hindustani vocal music, [14, 15] detect the steady segment in Hindustani vocal music by tracking the predominant peaks from the pitch histogram (pitch class distribution), then fit a 3 degrees polynomial to the remaining transitory segments. [16] obtains a canonical representation of the pitch contour in terms of straight lines and critical points of inflection, then recognizes two vocal expressions - *andolan* and *meend* by using their templates. In contrast to Hindustani music, the intonation of jingju singing shows more variability in steady segments (Figure 1). However, the transitory segment of jingju singing doesn't contain complicated ornaments such as undulating glide [17] and is thus less complicated. Consequently, a specific segmentation techniques will be developed for the adaptation of the intonation characteristics of jingju singing.

Both frequency domain and time domain algorithms are employed in vibrato detection. In the former, we exploit the fact that the spectrum of the vibrato pitch contour represents a predominant peak in the considered frequency range [18]. In the latter, vibrato period is measured with the temporal distances between two local maxima or minima. The presence of vibrato is estimated by several parameters, such as mean and variance of vibrato period, number of distances [19]. [20] examined the vibrato statistics of two jingju singing role-types - *laosheng* and *dan*, which shows a similar vibrato rate and extent to Western opera singing.

## 3. APPROACH

Considering that the steady segment in jingju singing is much less 'stable' than that of Hindustani vocal music (Section 1), we develop here a new segmentation method (Figure 2). Firstly, we perform the preliminary segmentation by using a modified melodic transcription algorithm to obtain the pitch contour of each note. Secondly, we conduct the finer segmentation within each note's pitch contour by employing the standard deviation of the cumulative differences of local extrema (StdCdLe) as the criterion. Thirdly, we classify the segmented pitch contours into three categories: linear trend, vibrato and 'others'. Then, we perform a new round of segmentation only for the pitch contours in 'others' category, and reclassify the results into linear trend category. Finally, a straight line is fitted to the pitch contours of linear trend category by linear regression, and a refinement process is performed by concatenating the over-segmented pitch contours according to three criteria: the fitted slope, pitch and timing distances.

### 3.1 Preliminary Segmentation

We use Mauch's algorithm to extract the fundamental frequency contour (pitch contour) and segment it into musical notes [9]. With the aim of improving the melodic transcription accuracy specifically for jingju singing, we incorporate a genre-specific musicological model into the note tracking step: bigram note transition probabilities estimated from a jingju singing score dataset. Bigram note

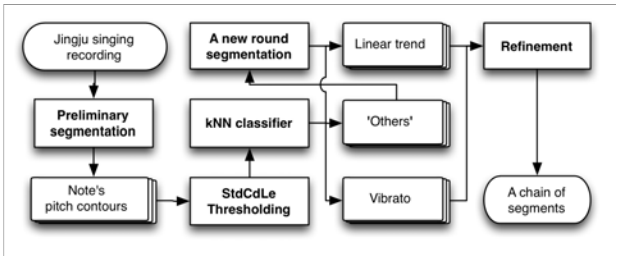


**Figure 2**. The block diagram of pitch contour segmentation.

transition probabilities define the probabilities of a transition from one note to another. Distinctive musical traditions represents distinctive bigram probabilities. For example, the major third note transition probability of jingju singing is much lower than that of Germany and Poland folk songs [11]. Unlike the probabilities estimated from a dataset, original algorithm provides an empirical, gaussian-shaped likelihoods which do not correspond to actual fact, for example, the excessively high probability of self-transition.

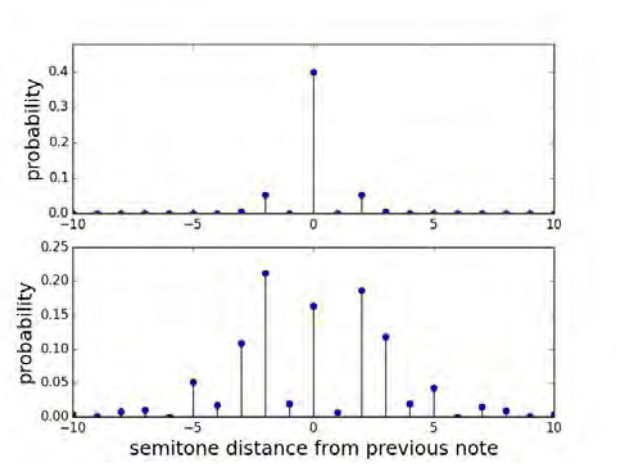


**Figure 3**. Central part of the note transition probabilities: original (upper) and those estimated from jingju singing score dataset (lower).

We estimate the bigram note transition probabilities (Figure 3) independently of the musical key from a jingju singing score dataset [1]. This dataset focuses on three jingju role-types: *dan* (female), *laosheng* (elderly man) and *jing* (painted-face). It contains 62 arias and represents 20827 notes. They are manually transcribed from printed sheet music into MusicXML format.

### 3.2 StdCdLe Thresholding

The strong variability of the jingju singing pitch contour (Section 1) urges us to reconsider the concept of each segment category. The "pure" steady segment as appeared in Hindustani vocal music is hard to find, which thus forces us to search for some uniformity among the variability during the segmentation process, such as the uniform rise and fall of a pitch contour. In this work, we use the criterion - StdCdLe to measure the uniformity of the variability of a segment, and assume that the StdCdLe of a segment should be lower than a given threshold (Figure 4).

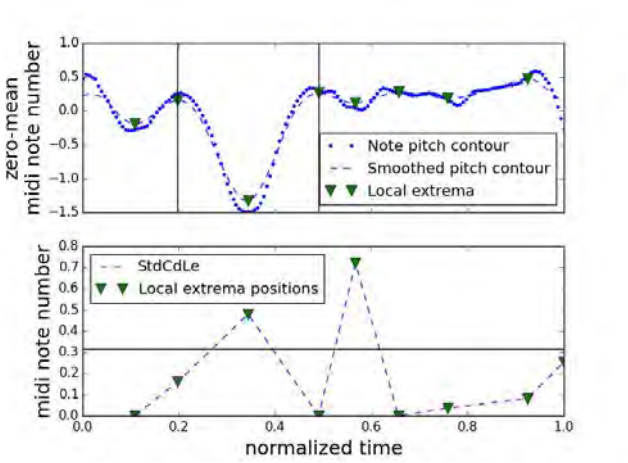


**Figure 4**. The note's pitch contour (upper): vertical black solid lines represent the segmentation points. The StdCdLe (bottom): horizontal black solid line represent the segmentation threshold.

The time interval of each contour obtained by the modified melodic transcription algorithm is normalized to [0,1]. Then, the pitch contour is subtracted by its mean value and passed through a moving average filter to remove the vocal jitter. After that, the local extrema in cents $A = \{\alpha_1, \alpha_2, ..., \alpha_M\}$ are detected on each pitch contour as the potential segmented points and its forward difference is denoted as $\triangle A = \{\alpha_2 - \alpha_1, ..., \alpha_M - \alpha_{M-1}\} = \{\delta_1, \delta_2, ..., \delta_{M-1}\}$ Finally, we segment the note's pitch contour on the basis of thresholding iteratively its StdCdLe. Whenever the StdCdLe exceeds a threshold $th_s$, the current local extremum is set as a segmentation point $B = \{\beta_1, \beta_2, ..., \beta_N\}$, and the cumulative vector $C$ is reset to empty. Otherwise, the current difference is appended to the cumulative vector (Algorithm 1).

**Algorithm 1** StdCdLe Thresholding Function

**function** StdCdLe($\triangle A$) : $B$
  $j \leftarrow 1, k \leftarrow 1$
  **for** $i \leftarrow 1, 2, ..., M-2$ **do**
    $C[j] \leftarrow \delta_i, j \leftarrow j+1$
    **if** std$(C) > th_s$ **then**
      $\beta_k \leftarrow i+1, k \leftarrow k+1$
      $C \leftarrow$ empty array, $j \leftarrow 1$
    **end if**
  **end for**
  **return** $B$
**end function**

### 3.3 Pitch Contour Classification

The aim of this step is to classify the preliminary segments into three categories: linear trend, vibrato and 'others'. The linear trend category is supposed to contain the segments with linear characteristic, such as steady and transi-

[1] https://github.com/jingju-SMC2016-PCS/SMC2016

tory ones. The 'others' category contains those which can be further segmented. The selected features are listed in Table 1.

| Category | Features | Dim. |
|---|---|---|
| Linear regression | Regression coefficients | 2 |
| | R-squared | 1 |
| | Mean squared error | 1 |
| | Fitting curve crossing | 1 |
| Others | Vibrato rate | 1 |
| | Extrema number | 1 |
| | Contour length | 1 |
| | Standard deviation | 1 |

**Table 1**. Pitch contour classification features and their dimensions (Dim.)

The linear regression feature category is used for describing the linearity of the segment, of which the 'fitting curve crossing' is calculated in a similar way as the 'zero crossing' but using the fitting curve as the crossing axis. This feature is designed for the classification of vibrato segment which tends to have a large 'fitting curve crossing' number. Due to the error of pYIN pitch estimation algorithm, the segment may contain outliers. We employ the measure of Cook's distance [21] to exclude the points in the segment which exhibit a large degree of influence on the estimated coefficients. The vibrato rate is calculated by using the frequency domain approach [18]. A kNN is chosen as the classifier considering its simplicity.

Due to the 'non linear' characteristic, the segments classified into 'others' category will be segmented a new round by simply using the local extrema as the segmentation points. The expected results will be reclassify into linear trend category which only consists of steady and transitory segments.

### 3.4 Refinement

The aim of the refinement step is to eliminate the phenomenon of over-segmentation which is mainly brought by the preliminary segmentation step (Section 3.1) in the edge region of the note's pitch contour. The approach is to concatenate the adjacent segments which meet certain given conditions (Figure 5).

We use the same linear regression technique mentioned in Section 3.3 to fit the segments, then extract their slopes. The slope, the starting and ending pitches of the fitted line of the segment $i$ are respectively denoted as $k_i$, $f_i^s$ and $f_i^e$. The starting and ending times of the segment $i$ are respectively denoted as $t_i^s$ and $t_i^e$. The concatenation conditions are listed as following:

$$|k_{i+1} - k_i| < th_{\text{slope}} \quad (1)$$

$$|t_{i+1}^s - t_i^e| < th_{\text{time}} \quad (2)$$

$$|f_{i+1}^s - f_i^e| < th_{\text{pitch}} \quad (3)$$

The condition 1 verifies that the adjacent segments both have the similar slope. The condition 2 and condition 3

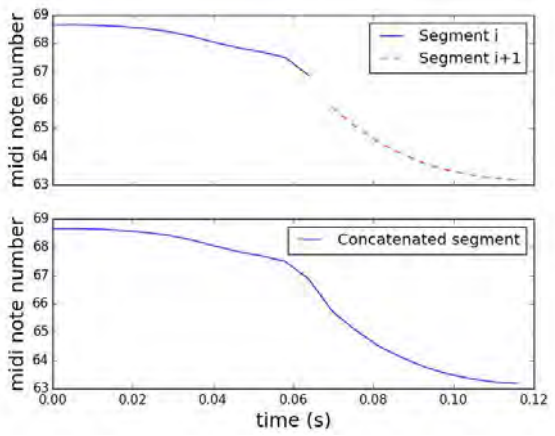


**Figure 5**. An example of concatenation of adjacent segments which meet the given conditions.

make sure that they are well connected in terms of boundary timing and pitch. If the adjacent segments are both flat pitch segments, that is to say:

$$|k_i|, |k_{i+1}| < th_{\text{flat pitch}} \quad (4)$$

only conditions 1 and 2 need to be fulfilled to achieve the concatenation. If the adjacent segments are both not flat pitch segments but have the same slope signs:

$$k_i k_{i+1} > 0 \quad (5)$$

all of the three conditions need to be fulfilled to perform the concatenation. If the adjacent segments are both not flat pitch segments and have the different slope sign, they can not be concatenated.

## 4. EVALUATION AND RESULTS

### 4.1 Dataset

The *a cappella* singing audio dataset used for melodic transcription and pitch contour segmentation tasks coming from MTG and C4DM [2] [22] focuses on two most important jingju role-types [23]: *dan* (female) and *laosheng* (elderly man). It contains 41 interpretations of 33 unique arias sung by 13 jingju singers. The melodic transcription, Std-CdLe thresholding, pitch contour classification and segmentation ground truth are manually annotated to all of the arias[1] . The melodic transcription ground truth represents 7686 notes. The average note duration is 0.38s and standard deviation of note duration is 0.37s. The pitch contour segmentation ground truth represents 14467 segments. The average segment duration is 0.21s and standard deviation of note duration is 0.25s. All of them are manually annotated in Sonic Visualizer [3] .

Since no parameter needs to be optimized in the preliminary segmentation step, all of the audio dataset and the annotated ground truth for the melodic transcription are used to evaluate its accuracy. For the rest of the steps, the dataset

[2] http://isophonics.net/SingingVoiceDataset
[3] http://www.sonicvisualiser.org/

| Algorithms | COnPOff | | | COnP | | |
|---|---|---|---|---|---|---|
| | F-measure | Precision | Recall | F-measure | Precision | Recall |
| Baseline | 0.718 | 0.716 | 0.720 | 0.757 | 0.755 | 0.759 |
| Modified | **0.730** | 0.736 | 0.724 | **0.769** | 0.774 | 0.762 |

**Table 2**. Results for melodic transcription.

| Algorithms | COnPOff | | | COnP | | |
|---|---|---|---|---|---|---|
| | F-measure | Precision | Recall | F-measure | Precision | Recall |
| Baseline | 0.284 | 0.307 | 0.264 | 0.534 | 0.592 | 0.487 |
| Proposed | **0.388** | 0.480 | 0.326 | **0.642** | 0.793 | 0.539 |

**Table 3**. Results for pitch contour segmentation.

is randomly split into 4 parts with the constraint that each part is selected without role-type bias and contains almost an equal number of segments. 3 of them are reserved as the training set for the purpose of parameter optimization. Another part is used as the test set to evaluate the pitch contour segmentation accuracy.

### 4.2 Evaluation Metrics

According to [9], note-based evaluation can expose more subtle details than frame-wise evaluation. We adopt the note-based evaluation metrics described in [24] for the melodic transcription evaluation: COnPOff and COnP. COnPOff takes into account correct note onset time (+/- 50 ms), pitch (+/- 0.5 semitones) and offset (+/- 20% of the ground truth note duration or +/- 50 ms, whichever is larger) is the most strict metric. COnP (correct note onset time and pitch) is the relax metric.

For the pitch contour segmentation evaluation, we adopt the similar metrics. However, since there is no demand for the pitch correctness, we simplify the metrics as: COnOff (correct segment onset and offset) and COn (correct segment onset).

### 4.3 Parameters Optimization

The parameters which need to be optimized are: StdCdLe threshold $th_s$ in Section 3.2, the number of kNN neighbors $K$ in Section 3.3 and $th_{\text{slope}}$, $th_{\text{time}}$, $th_{\text{pitch}}$, $th_{\text{flat pitch}}$ in Section 3.4. We use the grid search algorithm to perform the optimization. Since the kNN classification is a learning algorithm, a 3-fold cross-validation is adopted on the training set to measure its performance metric - misclassification rate. As $K$ increases, the misclassification rate firstly goes down, then stabilizes. The optimal $K$ is set at the beginning of the stable zone. The cross-validation technique is not employed during the optimization process of other parameters, because these steps do not contain learning algorithms and the performance metric - F-measure of COnOff can be report directly by sweeping these parameters on the training set without validation. Table 4 lists the search bounds or sets and the optimal results. The complete results of the grid search can be found on the web page[1] .

| Parameters | Search bounds or sets | OR |
|---|---|---|
| $th_s$ | $[0.01, 1.0]$ with step 0.01 | 0.22 |
| $th_{\text{slope}}$ (deg) | $[10, 90]$ with step 10 | 60 |
| $th_{\text{flat pitch}}$ (deg) | $[10, 90]$ with step 10 | 30 |
| $th_{\text{time}}$ (sec) | {0.01,0.02,0.05,0.1,0.2,0.5} | 0.02 |
| $th_{\text{pitch}}$ (semitone) | {0.1,0.2,0.5,1,2,5,10} | 5 |
| $K$ | $[3, 20]$ with step 1 | 13 |

**Table 4**. Search bounds or sets, optimal results (OR) of the optimization process for each parameter.

### 4.4 Results and Discussion

For the melodic transcription algorithms evaluation, we run a modified version of Mauch's algorithm which incorporated with the jingju singing bigram note transition probabilities. The results (Table 2) show that the modified algorithm slightly outperforms the original one (baseline), which proves of the necessity of the incorporation of genre-specific musicological model into transcription system [12, 13].

For the pitch contour segmentation algorithms evaluation, we choose Ganguli's method [15] as the baseline. The proposed method largely outperforms the baseline method (Table 3) because, firstly, as we mentioned above, the steady segment in jingju singing contains more variability than that in Hindustani vocal music, thus the 'pitch histogram discretisation' approach used by Ganguli which only searches 'pure' steady segment fails; secondly, the baseline method is more likely to lose effectiveness on segmenting the vibrato in jingju singing which mainly rides on a steady segment, whereas that in Hindustani vocal music mainly appears as undulating glide - nearly periodic oscillation rides on a glide-like transition [17]. In general, the better adaptation of our method to jingju singing characteristics leads to a better segmentation accuracy.

## 5. WEB PROTOTYPE

The web prototype[4] (Figure 6) is designed for desktop browser and realized using mainly JavaScript. The data visualization and interaction are handled in the frontend, uti-

[4] https://dunya.compmusic.upf.edu/smc-2016

lizing primarily two JavaScript libraries D3.js and wavesurfer.js. The implementation uses Web Audio API and HTML5 technologies. The recorded audio from the user is sent to a backend server set up by Python, where the computation of the algorithms is done, and the results are visualized dynamically in the frontend.

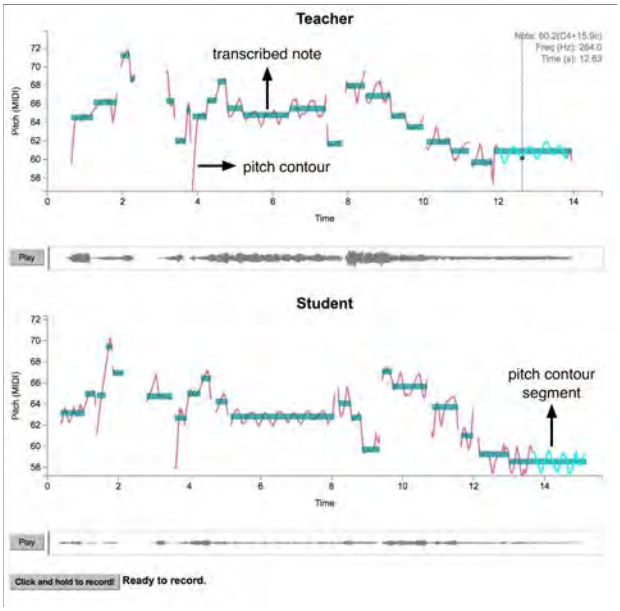


**Figure 6**. A web prototype for jingju singing training.

The web prototype consists of mainly two graphs, one of the teacher, which is the imitation target, and the other of the student, which reflects the students singing. By comparing the two graphs, the tool provides useful analysis to help the students find the difference and improve with practice. Each graph consists of a pitch contour of the singing melody, and notes derived from the pitch contour. Hovering over the mouse on the graph will highlight current pitch contour segment or note, and the aligned one on the other graph. The pitch (in Hz, cents and letter notation), duration and starting, end time information of the selected note and its aligned one will be shown on the up-right corner of each graph. Clicking and holding the record button will start recording. Releasing the record button the student graph will be updated with the analysis of the recording. A trainer and a student of jingju singing from NACTA (National Academy of Chinese Theatre Arts, China) have tried using this interface during the training class. They both indicated that it can provide a visual representation for the singing intonation and timing, and effectively help identify the problem in a rapid and intuitive way.

## 6. CONCLUSION AND FUTURE WORK

In this paper we have presented two techniques for computer-aided jingju singing training, their evaluations and a web prototype implementation.

The melodic transcription results suggest that the incorporation of the genre-specific musicological model - bigram note transition probabilities of jingju singing can successfully increase the transcription accuracy to this type of music in terms of note-based evaluation. The study of pitch contour segmentation shows that our proposed algorithm is able to search for the uniformity among the high variability included in the steady segments of jingju singing and achieve a better segmentation accuracy than the state of the art method. The web prototype provides intonation analysis with a fine granularity, and helps the students find the difference by comparison with the teacher.

With the deepening of our research on jingju singing, more musicological models or knowledge will be exploited to optimize the pitch contour segmentation algorithms. Since the syllabic level phonation accuracy occupies an important place in jingju singing training, the effort will be also focused on syllable-related techniques, such as syllable segmentation. Finally, the research of a better web interface which can effectively help in the training process and the systematic evaluation of its effectiveness will be included in our future works.

## 7. REFERENCES

[1] C. Mackerras, *The rise of the Peking Opera, 1770-1870: social aspects of the theatre in Manchu China*. Clarendon Press, 1972.

[2] E. Wichmann, *Listening to Theatre: The Aural Dimension of Beijing Opera*. University of Hawaii Press, 1991.

[3] E. Molina, "Automatic scoring of singing voice based on melodic similarity measures," Master thesis, 2012.

[4] W.-H. Tsai and H.-C. Lee, "An automated singing evaluation method for Karaoke systems." IEEE, May 2011, pp. 2428–2431.

[5] O. Mayor, J. Bonada, and A. Loscos, "The singing tutor: Expression categorization and segmentation of the singing voice," in *In Proceedings of the AES 121st Convention*, 2006.

[6] R. Schramm, H. d. S. Nunes, and C. R. Jung, "Automatic Solfège Assessment," in *International Society for Music Information Retrieval Conference*, Málaga, Oct. 2015.

[7] J. M. Geringer and C. K. Madsen, "Musicians' Ratings of Good versus Bad Vocal and String Performances," *Journal of Research in Music Education*, vol. 46, no. 4, pp. 522–534, Dec. 1998.

[8] E. Benetos, S. Dixon, D. Giannoulis, H. Kirchhoff, and A. Klapuri, "Automatic music transcription: challenges and future directions," *Journal of Intelligent Information Systems*, vol. 41, no. 3, pp. 407–434, 2013.

[9] M. Mauch, C. Cannam, R. Bittner, G. Fazekas, J. Salamon, J. Dai, J. Bello, and S. Dixon, "Computer-aided Melody Note Transcription Using the Tony Software: Accuracy and Efficiency," in *Proceedings of the First International Conference on Technologies for Music Notation and Representation*, May 2015.

[10] A. de Cheveigné and H. Kawahara, "YIN, a fundamental frequency estimator for speech and music," *The Journal of the Acoustical Society of America*, vol. 111, no. 4, pp. 1917–1930, 2002.

[11] T. Viitaniemi, A. Klapuri, and A. Eronen, "A probabilistic model for the transcription of single-voice melodies," in *2003 Finnish Signal Processing Symposium, FINSIG'03*, 2003, pp. 59–63.

[12] X. Serra, "A Multicultural Approach in Music Information Research," in *Proceedings of the 12th International Society for Music Information Retrieval Conference (ISMIR-2011)*, Miami (Florida), USA, Oct. 2011, pp. 151–156.

[13] A. Klapuri and M. Davy, Eds., *Signal Processing Methods for Music Transcription*. Boston, MA: Springer US, 2006.

[14] S. Gulati, J. Serrà, K. K. Ganguli, and X. Serra, "Landmark Detection in Hindustani Music Melodies," in *International Computer Music Conference/Sound and Music Computing Conference*, Athens, Greece, Sep. 2014, pp. 1062–1068.

[15] K. K. Ganguli and P. Rao, "Discrimination of melodic patterns in indian classical music," in *2015 Twenty First National Conference on Communications (NCC)*, Feb. 2015, pp. 1–6.

[16] S. S. Miryala, K. Bali, R. Bhagwan, and M. Choudhury, "Automatically Identifying Vocal Expressions for Music Transcription." in *International Society for Music Information Retrieval Conference*, 2013, pp. 239–244.

[17] C. Gupta and P. Rao, "Objective Assessment of Ornamentation in Indian Classical Singing," in *Speech, Sound and Music Processing: Embracing Research in India*. Springer Berlin Heidelberg, 2012, pp. 1–25.

[18] N. Kroher, "Automatic Characterization of Flamenco Singing by Analyzing Audio Recordings," Master thesis, Universitat Pompeu Fabra, 2013.

[19] S. Rossignol, P. Depalle, J. Soumagne, X. Rodet, and J.-L. Collette, "Vibrato: detection, estimation, extraction, modification," in *Digital Audio Effects Workshop (DAFx'99)*, 1999.

[20] L. Yang, M. Tian, and E. Chew, "Vibrato characteristics and frequency histogram envelopes in Beijing opera singing," *5th International Workshop on Folk Music Analysis*, pp. 139–140, 2015.

[21] R. D. Cook, "Detection of Influential Observation in Linear Regression," *Technometrics*, vol. 19, no. 1, pp. 15–18, 1977.

[22] D. A. A. Black, M. Li, and M. Tian, "Automatic Identification of Emotional Cues in Chinese Opera Singing," in *13th Int. Conf. on Music Perception and Cognition (ICMPC-2014)*, 2014, pp. 250–255.

[23] R. Caro Repetto and X. Serra, "Creating a Corpus of Jingju (Beijing Opera) Music and Possibilities for Melodic Analysis," in *15th International Society for Music Information Retrieval Conference (ISMIR-2014)*, Taipei, Taiwan, Oct. 2014, pp. 313–318.

[24] E. Molina, A. M. Barbancho, L. J. Tardón, and I. Barbancho, "Evaluation Framework for Automatic Singing Transcription," in *Proceedings of the 15th International Society for Music Information Retrieval Conference (ISMIR-2015)*, 2014, pp. 567–572.