COMPARISON OF THE SINGING STYLE OF
TWO JINGJU SCHOOLS

Rafael Caro Repetto, Rong Gong, Nadine Kroher, Xavier Serra
Music Technology Group, Universitat Pompeu Fabra, Barcelona
{rafael.caro, rong.gong, nadine.kroher, xavier.serra}@upf.edu

ABSTRACT

Performing schools (*liupai*) in jingju (also known as Peking or Beijing opera) are one of the most important elements for the appreciation of this genre among connoisseurs. In the current paper, we study the potential of MIR techniques for supporting and enhancing musicological descriptions of the singing style of two of the most renowned jingju schools for the *dan* role-type, namely Mei and Cheng schools. To this aim, from the characteristics commonly used for describing singing style in musicological literature, we have selected those that can be studied using standard audio features. We have selected eight recordings from our jingju music research corpus and have applied current algorithms for the measurement of the selected features. Obtained results support the descriptions from musicological sources in all cases but one, and also add precision to them by providing specific measurements. Besides, our methodology suggests some characteristics not accounted for in our musicological sources. Finally, we discuss the need for engaging jingju experts in our future research and applying this approach for musicological and educational purposes as a way of better validating our methodology.

1. MOTIVATION

This paper is a joint work between an ethnomusicologist (the first author) and a team of MIR researchers in the framework of the CompMusic project. In this project we exploit jingju music characteristics (and other music traditions) with the aim of pushing forward the state of the art in MIR. In last ISMIR Conference (Taipei, 2014) jingju music received significant attention, with a specific tutorial and several papers published by members of our team [1-3], as well as the work by Tian et al. [4]. In the present paper though, the motivation has been to test the potential of current MIR methodologies to support and enhance qualitative and descriptive musicological analyses of jingju music. To this aim, we have selected one of the more relevant aspects of jingju music appreciation, which is the singing style of different performing schools;
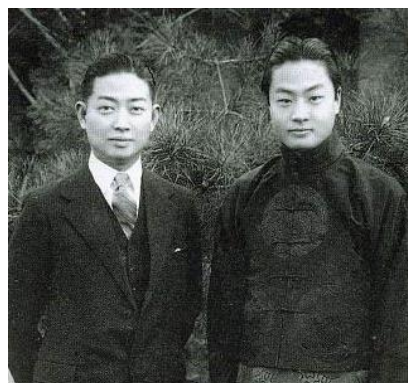
Figure 1. Mei Lanfang (left) and Cheng Yanqiu (right)

specifically, we have focused on two of the more popular ones, Mei school and Cheng school.

The paper is structured hence in the following sections. In the introduction we present the concept of jingju schools and the importance of singing style, as well as explain the purpose of the research undertaken for this paper. In the following two sections, we introduce the collection of recordings selected and the methodology proposed, and analyse the obtained results. In the discussion section we reflect on the challenges for expanding our research and present the direction of our future work. We conclude by summarising the musicological outcomes of the current research.

2. INTRODUCTION

Jingju is one of the genres of Chinese traditional theatre arts, arguably the most widespread and acclaimed one. Originally a folk art form, the actor traditionally was in charge of the whole creative process, from costumes and make-up, to acting, dancing, reciting, arranging the music and sometimes even writing (or improvising) the lyrics. In order to structure their performance, actors drew on a vast repertoire of predefined conventions handed down by tradition and which concerns every single aspect of this art. Characters of jingju plays are classified in acting categories or role-types, which define which set of conventions the actor who plays that role-type should master. The high complexity of such conventions requires the actor to specialize in the performance of just one role-type during his life-time career. Along jingju history, there were some outstanding actors that excelled in the mastery of these conventions and pushed forward the artistic standards of their respective role-types or the genre as a

whole. Some of these masters would bring their own personalities to their performances and created personal styles. In a tradition that is orally transmitted, this would result in the appearance of *liupai*, or performing schools.[1]

The first half of the 20th century was the period of major development of jingju and when the most renowned schools appeared. It saw the extraordinary development of the *dan* role-type, that portraying young or mid-aged female characters, but performed by male actors, due to social and political constrictions. Four of them gained the title of "four great *dan* actors" (*si da ming dan*),[2] and founded their own schools. Their strong personality, the context of market competition, and even their own physical condition caused schools of *dan* to be the ones with a greater degree of difference within one specific role-type. Among these four schools, those founded by Mei Lanfang (1894-1961) and Cheng Yanqiu (1904-1958) (Figure 1),[3] respectively named Mei and Cheng schools, are the most widespread and followed ones today, currently performed in its vast majority by female actresses. These are precisely the ones we chose for our study.

Each jingju school is highly associated to a particular repertoire of plays and generally to a predominant performance skill. In fact, this repertoire is formed of plays arranged by the school founder to precisely showcase its mastery in that specific skill. In the case of Mei and Cheng schools, singing is the most representative and acclaimed aspect of their art. This aspect concerns mainly two elements, the arrangement of new tunes[4] and the singing style. Among these two, the singing style is the feature that makes a performance instantly recognizable as belonging to any of these schools, and also one of the skills that performers, both professional and amateurs, put more effort to master. At the same time, it is one of the most important criteria for appreciating a performance among connoisseurs. The features that define singing style not only consist in the way voice is used in singing, but also in the very quality of the voice. Both of them should be considered not as natural personal qualities of particular actors, but as conventions that have to be trained and mastered. The resulting voice is to be understood hence as "an artificial voice, in the sense of displaying artifice, or art" [5], and followers of each school aim at mastering this voice quality as well.

Descriptions made in jingju musicology about singing style generally focus on its perceptual characteristics and the psychological profile of the characters conveyed through those characteristics. Wichmann [5] quotes a typical description of Mei school's singing style from *Zhongguo da baike quanshu* (China Great Encyclopedia) by Hu Qiaomu, in which timbre is described as "'sweet, fragile, clear and crisp, round, embellished and liquid.' This timbre is considered ideal for portraying 'natural, graceful and poised, dignified, gentle and lovely traditional women.'" In all the musicological sources consulted for this paper [5-10], description of singing style always includes this kind of terminology. However, since the aim of our research is add precision to musicological description, we have selected those characteristics for which audio features can be objectively computed. Table 1 shows the musicological characteristics selected and their corresponding audio features.

Characteristics	Audio features
Pitch register	Pitch histogram (1st degree)
Vibrato rate variability	Vibrato rate (SD)
Volume variability	Loudness (SD)
Brightness	Spectral centroid (mean) LTAS Tristimulus
Timbre variability	Spectral flux (mean)

Table 1. Musicological characteristics and their corresponding audio features considered in our study.

In the last few years there have been several studies about singing characteristics in different Chinese traditional theatre genres, like jingju [11-12] and kunqu [13-14]. In these studies different role-types have been compared in terms of several singing characteristics by analysing monophonic recordings produced by the authors. In our work, we look in depth to one particular role-type, *dan*, and analyse singing characteristics with explicit reference to its musicological descriptions and using commercial recordings. In the following section we describe the collection of recordings and explain the methodology proposed.

3. METHODOLOGY

For this study, we have selected a collection of recordings from our jingju music research corpus [1], according to two criteria: representativeness and comparability. In order to assure that these recordings are representative of their school, we have considered both the recording artist and the recorded aria. We have looked for artists whose school filiation is explicitly stated in the release's booklet, and arias belonging to plays for which we have literary evidence (mostly from [8]) that are representative of their school. In order to maximize comparability, we have searched for plays for which musicological literature specifically acknowledges a particular rendition from each of these two schools, as it is the case of *Su San qijie* according to [5, 8]. Since these are rare cases, due to the fact

---

[1] The translation of *liupai* as school can be subject to misinterpretation. Differently to other traditions, jingju schools do not imply training in specific institutions or affiliation to specific lineages. They consist in the transmission of the performance style of individual great masters, so that the reference is always the founder of the school, and not the teacher from whom the new actors or actresses learn.

[2] Quotes from Chinese sources are given in our translation.

[3] Picture from http://zh.wikipedia.org/wiki/程砚秋#/media/File:梅兰芳与程砚秋.JPG (detail).

[4] Since jingju music is created by applying pre-existing melodic conventions, it is customary to use the term arrangement (*bianqu*) instead of composition to refer to this process.

School	Work: *Play*. "Aria" (Character)	Recording: MusicBrainz ID	Length	Artist
Mei	`fhc`: *Feng huan chao*. "Ben ying dang sui muqin Haojing bi nan" (Cheng Xue'e)	`fhc-LYf`: a1e4b77b-88b0-4003-b688-66e39f579dc6	7:33	Li Yufu
		`fhc-SYh`: 4e3b46b2-9db7-4f52-af95-e43239a6c0e1	6:56	Shi Yihong
		`fhc-LSs`: 83d2fc7f-e1c1-4359-b417-ed9e519ecbb7	7:34	Li Shengsu
	`ssqj`: *Su San qijie*. "Yu Tangchun han bei lei mang wang qian jin" (Su San)	`ssqj-LSs`: 067b8f25-888a-4a08-a495-cbc402846b10	7:15	
Cheng		`ssqj-CXqd`: 87dbdf41-37ff-4f4a-83d4-7169d674579a	6:20	Chi Xiaoqiu
	`sln`: *Suo lin nang*. "Chunqiu ting wai feng yu bao" (Xue Xiangling)	`sln-CXq`: 11a44af7-e29a-4c50-aa38-6139d37ca306	3:21	
		`sln-LPh`: 3dcae41a-795c-4b7d-979b-1b52aa42dd3a	3:06	Li Peihong
		`sln-LGj`: 1e705224-0b44-48aa-a0de-6386cda9d517	3:15	Liu Guijuan

Table 2. Description of the recordings used in this paper. When applicable, short forms are provided.

that each school has developed its own specific repertoires, we have also searched for arias with similar music structure. This is the case of `fhc` and `sln`, arranged in the same *shengqiang* and similar *banshi*.[1] The resulting collection of recordings is described in Table 2, together with the abbreviations used throughout the paper. We argue that the size of this collection is appropriate for the current research since we are not performing a quantitative study. Instead we are using MIR methodologies for supporting qualitative descriptions.

Figure 2 describes the methodology proposed for this paper. Each of the recordings is ripped in a lossless compressed format with a sampling rate of 44.1 kHz. We manually identify the sections containing singing voice, for which we compute the predominant melody using the vamp plug-in version of Salamon and Gómez's algorithm [15], setting a pitch range threshold of 100 Hz to 1000 Hz, and the voicing parameter at its maximum level, 3.0. Given that a percentage of error results, pitch tracks are manually corrected (an average of 7.16% from the computed frames). In order to measure pitch register, we use the methodology proposed in [16] to compute pitch histograms and obtain the pitch of the first degree[2] from their peaks values. The algorithm presented in [17] is used to measure vibrato rate and extent, of which we calculate mean and standard deviation (SD). To measure the remaining features, we separate the singing voice from the accompaniment by computing a harmonic model analysis and synthesis using the methodology presented

in [18], and apply to it standard algorithms for the computation of those features as implemented in the Essentia library [19]. For loudness we use the *Loudness* algorithm, normalizing the resulting mean to a factor of 0.5, so that SD is better comparable. To measure brightness, we compute mean and SD of the spectral centroid using the *Centroid* algorithm. In order to better understand timbre qualities, we also compute tristimulus using the *Tristimulus* algorithm, and long-term-average spectrum (LTAS) using the implementation presented in [20]. Finally, to measure timbre variability, we compute spectral flux mean and SD using the *Flux* algorithm from Essentia.

In the following section, we analyse the results obtained for each school and relate them with their corresponding musicological characteristics.
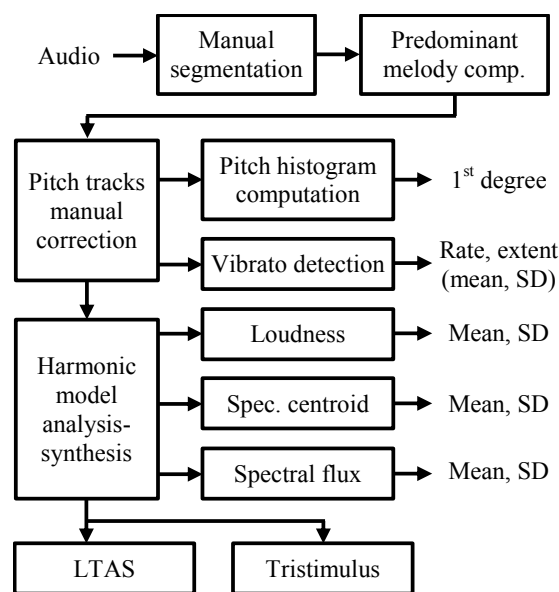

Figure 2. Block diagram of the methodology.

---

[1] *Shengqiang* is the musical convention in jingju that determines the melodic skeleton; *banshi* refers to the metrical pattern. For more detailed information about these concepts, please refer to [1, 5].

[2] We use "first degree" to translate *gongyin*. This term refers to the first degree of the sung scale, and in *jianpu* notation is notated with the number 1 (http://en.wikipedia.org/wiki/Numbered_musical_notation). Although common functions are shared, we consciously avoid the term tonic, for its implications with tonality, absent in jingju music.

School	Features								
	1st deg. (Hz)	Vibrato				Loudness	Spectral centroid (Hz)		Spectral flux
		Rate (Hz)		Extent (cents)					
		Mean	SD	Mean	SD	SD	Mean	SD	Mean	SD
Mei	335.41	4.757	0.728	137.325	37.466	0.279	2536.739	366.968	0.121	0.063
Cheng	323.35	6.090	0.963	111.101	41.157	0.387	2136.555	451.642	0.087	0.058

Table 3. Average measurement values from each of the features computed for each school.

4. ANALYSIS OF THE RESULTS

Table 3 shows a summary with the average measurement values from each of the features computed for each school.[1] In this section we analyse how these results relate to the musicological descriptions of their corresponding musical characteristics.

According to our musicological references, pitch register in Mei is higher than in Cheng. Since pitch range of arias for the *dan* role-type is consistent across plays, approximately an octave and a major third, we take the pitch of the first degree as an indicator of pitch register. However, this degree is rarely sung in arias of this role-type. This is due to one singing convention, according to which female role-types shift their pitch register a fifth higher than male role-types, so that the modal center becomes the fifth degree. To measure the pitch of the first degree hence, we compute a pitch histogram and assign a modal degree to each peak by listening to the recordings with the aid of scores. Since we observed that the peak corresponding to the sixth degree is usually the cleanest one, we take it as reference by assigning the value of 900 cents. Figure 3 shows the resulting pitch histograms for `ssqj-LSs` and `ssqj-CXq`, compared with the equal tempered scale. It has to be noted that in jingju there is no absolute standard pitch for tuning, but it depends on the actor's or actress' needs. Notwithstanding this, a pitch is commonly assumed as reference for each role-type; for the *dan* role-type first degree is expected to be around E4 (329.63 Hz) [21]. Results in Table 3 show that this is the case for both schools, although first degree in Cheng is in average 6.28 Hz (33.30 cents) lower than E4, and Mei 5.78 Hz (30.09 cents) higher. Consequently, first degree in Cheng is in average 63.39 cents lower than Mei, more than a semitone. Results for all the recordings show that in every case first degrees from Mei are higher than those for Cheng, although the smallest difference between recordings from each school is 1.31 Hz. These results hence invite us to support the musicological description for pitch register.

Besides the aforementioned results, Figure 3 suggests that pitch histograms can shed light upon other aspects of singing style. Chen [22] has used histograms to study

that, as Figure 3 shows, compared with the equal tempered scale the fourth degree, although seldom used, is sung at a higher pitch, what is common knowledge in musicological literature. Unexpectedly, the higher octave of the first degree appears in the histograms slightly shifted higher, especially in Mei school, for which we have not found literary evidence. Besides, peak shape differences, cleaner and with lower valleys for Cheng, also suggest differences in singing style, probably regarding vibrato and ornamentation. These observations invite us to argue that pitch histograms could be further exploited for the characterisation of singing style and explore features unnoticed or not explicitly accounted for in our references.[2]

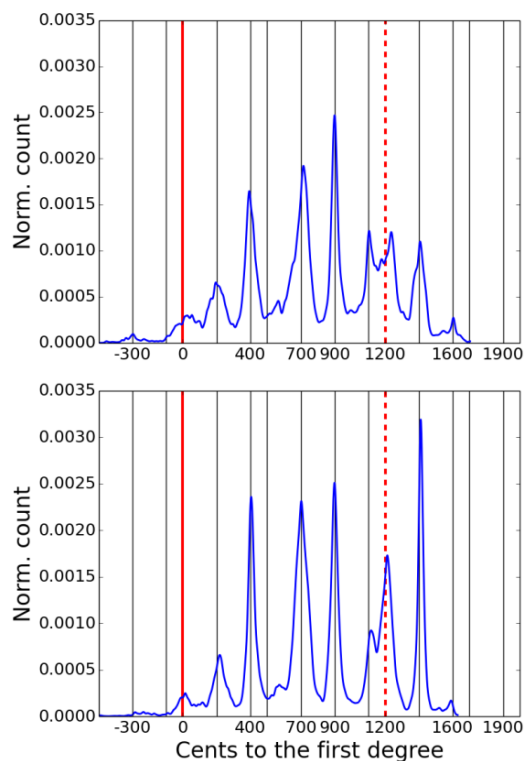According to the sources consulted, Cheng "excels in

Figure 3. Pitch histograms for `ssqj-LSs` (top) and `ssqj-CXq` (bottom). Vertical lines show the equal tempered scale, solid red line marks the first degree, and dotted red line marks its higher octave.

---

[1] Detailed results and more plots can be found in
https://github.com/jingjuschools/jingjuschoolsISMIR2015

[2] Differences in peaks height indicate different melodic preferences in each school, what concerns tune arrangement, an issue not considered in this paper.
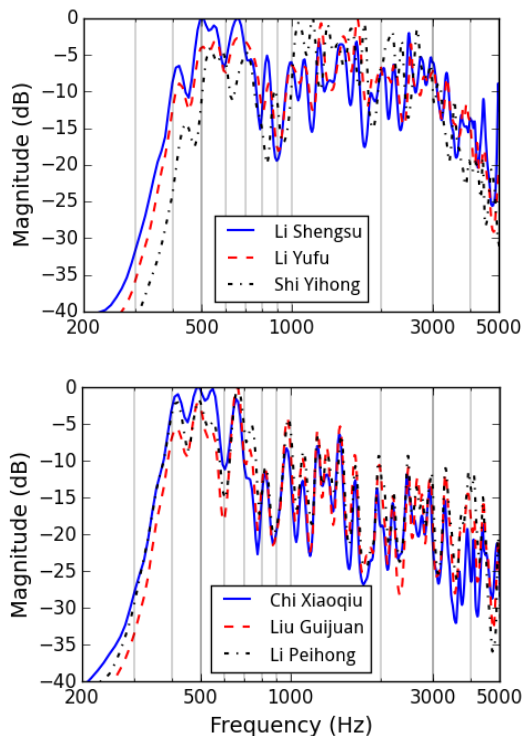
Figure 4. LTAS for the three recordings of `fhc` (top) and the three recordings of `sln` (bottom).

using slow and fast vibratos" [9]. Consequently, the feature that can better reflect this characteristic is the standard deviation of vibrato rate. As can be seen in Table 3, this value is higher in Cheng than in Mei, and this is also the case for every recording. However, the difference is less than 0.3 Hz, what is barely appreciable by human ear. Variability in vibrato extent, as reflected in our results, is slightly higher in Cheng than in Mei, but the difference in this case is even less significant. Besides, specific results from each recording are less consistent, since some instances from Mei school show higher SD in vibrato extent than others from Cheng, and vice versa. What our results clearly show though, is that vibrato in Mei is considerably slower and wider than in Cheng, a feature that is consistent across all the recordings. Interestingly enough, we have not found such a remark in our musicological sources.

Results obtained for loudness variance also support musicological description, which takes volume variability as a characteristic of Cheng school compared to Mei. Results for each recording are less consistent than for other features, finding one case in Mei with higher SD than the lowest value for a recording in Cheng. These results, however, ought to be taken carefully. Firstly, they might have been affected by possibly different mixing levels in the production process. Secondly, being loudness a perceptual feature, the algorithm used is an approximation to it by a simple modification of amplitude values, what prevent us to take them as a faithful representation of the characteristic measured.

Our musicological references agree that timbre in Mei school is brighter than in Cheng. To measure brightness, we have computed spectral centroid mean and SD. Values for the mean effectively show that Mei has brighter timbre than Cheng, as an average and for every recording in each school. Given the complexity of timbre, we have also looked at LTAS, a feature that has been used in [23] to study and compare vocal tract and formant structure. Figure 4 shows the LTAS for the three performances of `fhc` and `sln`. To focus on the region with greater loudness, plots show the region between 200 Hz and 5000 Hz, with a logarithmic scale in the x-axis. In these plots it can be observed how frequencies over 1000 Hz are considerably higher in Mei than in Cheng, as well as the peaks with the highest loudness, contributing thus to timbre brightness. Besides this information, LTAS allows us to compare vocal tract between performers in each school, so that we can observe that timbre similarity is higher in Cheng than in Mei. This method hence seems promising in order to characterise individual actors' or actresses' particularities within the overall requirements of the school.

Tristimulus has also been used to study and compare timbre qualities [24]. Figure 5 shows that in average both the second and the third of the three output components measured by tristumulus have higher values for Mei than for Cheng, what once more support the higher weight of higher partials in Mei. This figure also suggests a promising tool for classification according to timbre quality.

Finally, results for spectral flux, computed with the aim of measuring timbre variability, show a greater value in Mei than in Cheng, a difference which is consistent across all the recordings, with special homogeneity in Mei. Literature however remarks Cheng's timbre variability as a defining trait of this school. It is also interesting to notice that the SD value for spectral centroid also shows a higher value in Cheng, although is not as consistent as the spectral flux values across all the recordings. Since this descriptor was computed only for the sections of the recordings that contained singing, we re-
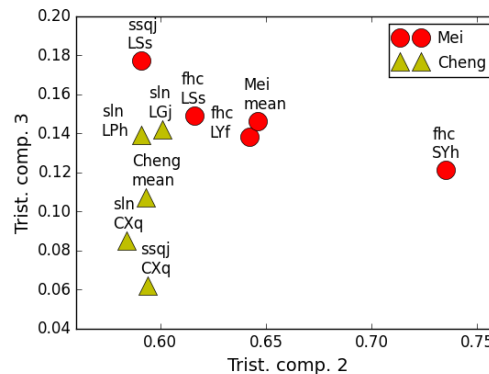

Figure 5. Scatter plot displaying values for the 2nd and 3rd tristimulus components for each recording and the mean for each school.

computed it setting different loudness thresholds for the transition frames in order to discard an influence of the segmentation. Results didn't change any tendency towards a higher value in Cheng than in Mei. How to approach timbre variability in jingju singing using audio analysis features remains hence an issue for future research, as discussed in the following section.

5. DISCUSSON

The analysis of the results presented in the previous section suggests that the methodology proposed in this paper is promising for the intended task, namely, supporting musicological description using audio analysis features. However, there are also some challenges that have to be addressed when extending this research in future work.

We aim to improve our methodology in the following two senses: automatizing most of the steps for audio analysis and improving existing algorithms to better fit jingju music characteristics. Some researchers in our team are currently working on improving the automatic segmentation method by Chen [22], adapting Ishwars' methodology [25] to jingju music for better extraction of predominant melody, and developing new algorithms for automatic computation of the first degree. Improvement of the harmonic model analysis and synthesis as presented in [18] is also to be undertaken in the near future.

Arguably, the bigger challenge for the continuation of this research is gaining the engagement of jingju musicologists. In this paper we have aimed to show how the present approach can benefit musicological work. Yet to that aim we have consciously avoided most of the terminology that is more commonly used by experts when describing singing style. The authors have not yet agreed on a methodology from an MIR point of view for approaching descriptions such as "sweet" (*tian*), "mellow" (*run*), "fragile" (*cui*), "round" (*yuan*), or "wide" (*kuan*). Even when considering a characteristic like timbre variability, whose study by means of spectral flux disagrees with musicological descriptions, we wonder how much of this disagreement is due to difficulties in establishing a common terminology between these two disciplines. From the field of MIR there have been recent calls for a better understanding of the musical content of commonly computed descriptors [26]. The authors argue that collaborative research as the one undertaken here would also encourage jingju experts to reflect on their terminology in terms of audio analysis features, and hopefully would gain complementary precision for those concepts.

Besides supporting musicological research, the use of audio features for qualitative analysis can be exploited for educational purposes. Jingju is a tradition that relies on oral transmission for training young actors and actresses. The key method in this training tradition consists on "teaching by mouth and heart" (*kou chuan xin shou*), that is, the teacher sings and the student repeats as many times as needed for achieving an acceptable standard. Recently,

new technologies are being used as part of this process. Students use their cell phones to record their teachers, and audio and video recordings of performances are easily accessible in the web. Technologies that could automatically evaluate the degree of similarity between the teacher's and the student's performance, and moreover offer a precise description of dissimilarities, would guide the trainee in better understanding his or her own learning process. The aim of such technologies would be performing qualitative analysis of audio recordings, similar to the ones implemented in this paper. Building upon the results obtained and the methodology tested in the current work, we have started to develop such educational tools. To this goal we will require closer collaboration with jingju experts. The involvement of these experts and the acceptance of the educational tools by jingju trainees will also provide better evaluation methods for our research.

6. CONCLUSIONS

The current paper has studied the potential of using audio analysis features for supporting and enhancing the musicological description of singing style in two jingju schools for the *dan* role-type, namely Mei and Cheng. Our results support the description given in musicological literature for most of the characteristics analysed, and add precision to them. Pitch register in Cheng school is in average 63.39 cents lower than Mei. Variability in vibrato rate is slightly higher in Cheng, what agrees with musicological description, but less than 0.3 Hz. Volume variability as a characteristic of Cheng school has been supported by the measure of loudness, whose SD is 38.71% higher in this school than in Mei. That this school is brighter in timbre than Cheng is supported by the mean value of its spectral centroid, 18.73% higher, but also by measurements in LTAS and tristimulus. Besides supporting these descriptions, our method also suggests some characteristics not explicitly accounted for in the sources consulted: vibrato in Mei is in average 1.33 Hz slower and 26.22 cents wider than in Cheng, pitch histograms suggest new characterisations for tuning and intonation, and LTAS looks promising for comparing vocal tracts of singers within a school. Only in the case of spectral flux, computed for studying timbre variability, the results do not support the musicological description, an issue that will be addressed in future research. In the light of these results, we have started to extend this methodology for the development of educational tools, a project in which we hope to gain the engagement of jingju experts, who could benefit from this approach for their own research.

7. ACKNOWLEDGEMENTS

8. REFERENCES

[1] R. Caro Repetto, and X. Serra: "Creating a Corpus of Jingju (Beijing Opera) Music and Possibilities for Melodic Analysis," *ISMIR 2014*, pp. 313-318, 2014.

[2] A. Srinivasamurthy, R. Caro Repetto, H. Sundar, and X. Serra: "Transcription and Recognition of Syllable based Percussion Patterns: The Case of Beijing Opera," *ISMIR 2014*, pp. 431-436, 2014.

[3] S. Zhang, R. Caro Repetto, and X. Serra: "Study of the Similarity between Linguistic Tones and Melodic Pitch Contours in Beijing Opera Singing," *ISMIR 2014*, pp. 343-348, 2014.

[4] M. Tian, G. Fazekas, D. Black, and M. Sandler: "Design and Evaluation of Onset Detectors Using Different Fusion Policies," *ISMIR 2014*, pp. 631-636.

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

[6] H. Li 李海涓: "Jingju qingyi 'Mei' 'Cheng' er pai zai changfa shang de tong yu yi" 京剧青衣"梅""程"二派在唱法上的同与异 (Similarities and differences in the singing techniques between the two schools of jingju *qingyi* Mei and Cheng), *Zhejiang yishu zhiye xueyuan xuebao*, Vol. 11, No. 1, pp. 50-55, 2013.

[7] R. Wang 汪人立: "Mei pai changqiang yinyue de meixue pinge" 梅派唱腔音乐的美学品格 (Character of music aesthetics in Mei school's singing), *Yishu bai jia*, 1996, No. 1, pp. 40-47.

[8] T. Wu 吴同宾, and Y. Zhou 周亚勋: *Jingju zhishi cidian* 京剧知识词典 (Dictionary of jingju knowledge), Tianjin renmin chubanshe, Tianjin, 2006.

[9] S. Yu 俞淑华: "Chuyi Chengpai de changqiang yu banzou" 刍议程派的唱腔与伴奏 (My humble opinion about Cheng school's singing and instrumental accompaniment), *Zuojia zazhi*, 2012, No. 1, pp. 209-210.

[10] S. Yu 俞淑华: "Lun jingju 'si da ming dan' de changqiang yinse" 论京剧"四大名旦"的唱腔音色 (Discussing singing timbre in jingju's 'four great *dan* actors'), *Ming zuo xinshang*, 2011, No. 33, pp. 114-115.

[11] J. Sundberg, L. Gu, Q. Huang, and P. Huang: "Acoustical study of classical Peking Opera singing," *Journal of Voice*, Vol. 26, No. 2, pp. 137-143, 2012.

[12] 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.

[13] L. Dong, J. Sundberg, and J. Kong: "Loudness and Pitch of Kunqu Opera," *Journal of Voice*, Vol. 28, No. 1, pp. 14-19, 2014.

[14] L. Dong, J. Kong, and J. Sundberg: "Long-term-average spectrum characteristics of Kunqu Opera singers' speaking, singing and stage speech," *Logopedics Phoniatrics Vocology*, Vol. 39, No. 2, pp. 72-80, 2014.

[15] J. Salamon, and E. Gómez: "Melody Extraction From Polyphonic Music Signals Using Pitch Contour Characteristics," *IEEE Transactions on Audio, Speech, and Language Processing*, Vol. 20, No. 6, pp. 1759-1770, 2012.

[16] G. K. Koduri, V. Ishwar, J. Serrà, X. Serra, and H. Murthy: "Intonation analysis of ragas in Carnatic music," *JNMR*, Vol. 43, No. 1, pp. 72-93.

[17] P. Herrera, and J. Bonada: "Vibrato Extraction and Parametrization in the Spectral Modeling Synthesis Framework," *DAFx*, 1998.

[18] X. Serra, and J. Smith: "Spectral Modeling Synthesis: A Sound Analysis/Synthesis Based on a Deterministic plus Stochastic Decomposition," *Computer Music Journal*, Vol. 14, No. 4, pp. 12-24, 1990.

[19] D. Bogdanov, N. Wack, E. Gómez, S. Gulati, P. Herrera, O. Mayor, G. Roma, J. Salamon, J. Zapata, and X. Serra: "ESSENTIA: an Audio Analysis Library for Music Information Retrieval," *ISMIR 2013*, pp. 423-498, 2013.

[20] T. Kinnunen, V. Hautamäki, and P. Fränti: "On the Use of Long-Term Average Spectrum in Automatic Speaker Recognition," *Proceedings of the 5th International Symposium on Chinese Spoken Language Precessing*, pp. 559-567, 2006.

[21] B. Cao 曹宝荣: *Jingju changqiang banshi jiedu: Xia ce* 京剧唱腔板式解读·下册 (Deciphering *banshi* in jingju singing: Second volume), Renmin yinyue chubanshe, Beijing, 2010.

[22] K. Chen: *Characterization of Pitch Intonation of Beijing Opera*, Master thesis, Universitat Pompeu Fabra, Barcelona, 2013.

[23] J. Sundberg: *The Science of the Singing Voice*, Northern Illinois University Press, Dekalb, 1987.

[24] M. Campbell, and C. Greated: *The Musician's Guide to Acoustics*, Oxford University Press, Oxford, 1987.

[25] V. Ishwar: *Pitch Estimation of the Predominant Vocal Melody from Heterophonic Music Audio Recordings*, Master Thesis, Universitat Pompeu Fabra, Barcelona, 2014.

[26] B. Sturm: "A Simple Method to Determine if a Music Information Retrieval System is a 'Horse'," *IEEE Transactions on Multimedia*, Vol. 16, No. 6, pp. 1636-1644, 2014.