Creating an A Cappella Singing Audio Dataset for Automatic Jingju Singing Evaluation Research

Rong Gong
Music Technology Group
Universitat Pompeu Fabra
Barcelona, Spain
rong.gong@upf.edu

Rafael Caro Repetto
Music Technology Group
Universitat Pompeu Fabra
Barcelona, Spain
rafael.caro@upf.edu

Xavier Serra
Music Technology Group
Universitat Pompeu Fabra
Barcelona, Spain
xavier.serra@upf.edu

ABSTRACT

The data-driven computational research on automatic jingju (also known as Beijing or Peking opera) singing evaluation lacks a suitable and comprehensive a cappella singing audio dataset. In this work, we present an a cappella singing audio dataset which consists of 120 arias, accounting for 1265 melodic lines. This dataset is also an extension our existing CompMusic jingju corpus. Both professional and amateur singers were invited to the dataset recording sessions, and the most common jingju musical elements have been covered. This dataset is also accompanied by metadata per aria and melodic line annotated for automatic singing evaluation research purpose. All the gathered data is openly available online[1].

KEYWORDS

a cappella singing, automatic jingju singing evaluation, audio recording dataset

1 INTRODUCTION

1.1 Short presentation of jingju music

The music of jingju has been receiving increasing attention from MIR researchers in the last years. A brief jingju MIR research bibliography can be referred to [8]. Music in jingju has been deeply conventionalized according to the following three elements that build its musical system:

- *shengqiang*: melodic framework associated with a particular emotional atmosphere. There are two main *shengqiang*, namely *xipi* and *erhuang*, which define the musical identity of jingju.
- *banshi*: rhythmic transformations of the *shengqiang*s melodic framework, which can be classified into two categories: metered and non-metered.
- role-type: acting profile which the performer belongs to. There are four broad categories of role-types: *sheng*, *dan*, *jing* and *chou*, where *chou* is focused on other disciplines than singing like reciting, acting or acrobatics. For the role-types focused on singing, their style is a variation of either the male or female style, represented respectively by *laosheng* and *dan*.

The structure of the lyrics determines the musical structure of the arias. The basic lyric unit for jingju arias is the couplet, and each *shengqiang* defines a melodic line for the opening line of the couplet, and another one for the closing line. One single aria is usually set to only one *shengqiang*, but it can contain different *banshi*.

1.2 Automatic jingju singing evaluation

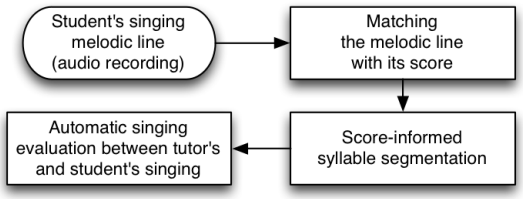

Figure 1: Flow chart of the research project.

The most prominent aspect of jingju music is singing. The ultimate goal of our research project is to automatically evaluate the jingju a cappella singing of a student in the scenario of jingju singing education, see figure 1. Jingju is extremely demanding in the clear pronunciation and accurate intonation for each syllabic or phonetic singing unit. During the initial learning stages, students are required to imitate completely tutor's singing. Therefore, the automatic jingju singing evaluation system we envision is based on this training principle and measures the intonation and pronunciation similarities between the student's and the tutor's melodic lines. Before measuring the similarities, the a cappella singing melodic lines should be matched with their scores [4]; then the score-informed method will be used to segment these lines into syllabic or phonetic units in order to capture the temporal details [6]. Considering that our research mainly uses data-driven methods, it is thus necessary to build a relatively large a cappella singing audio dataset in order to better train and validate the computational models for automatic jingju singing evaluation.

1.3 Existing jingju music audio datasets

The CompMusic corpus [7] is formed by a collection of commercial recordings, as well as their metadata. Another jingju music corpus gathered in [10] also consists of commercial recordings, annotated for structural segmentation analysis. These recordings are

[1] https://doi.org/10.5281/zenodo.842229

DLfM'17, Shanghai, China

DOI:

all mixed with instrumental accompaniment, which means clean singing voice should be separated during the preprocessing step if we want to take advantage of these recordings for the automatic singing evaluation research. However, singing voice separation itself is a very challenging research task. One of the state-of-the-art source separation algorithms [2] was tried by the authors, which was not able to produce a clean singing voice while preserving its characteristics.

The only jingju a cappella singing dataset we can find [1] is small and not complete enough since it only contains around 1 hour recordings of 31 unique arias, and its annotations were made for the task of mood recognition instead of automatic singing evaluation. This dataset has been used intensively in our previous studies for the subtasks of pitch contour segmentation [5], acoustic modeling of phonemes [4] and syllable segmentation [3, 6]. The results of these studies revealed that such small dataset is the main bottleneck which prevents building robust computational models with better generalization ability.

The main goal of the dataset presented in this paper is to offer a comprehensive and complete resource for the study of automatic jingju singing evaluation. The recordings of this dataset show the most common *shengqiang*, *banshi* and role-types in the jingju singing educational scenario. Both professional and amateur recordings were collected so that similarity or evaluation models can be built between them. The remaining of the paper is structured as follows. In the next section the jingju a cappella singing audio dataset is described in detail. In the last section, we present some concluding remarks and point out future work.

2 RECORDING THE DATASET

2.1 Artists

We invited 5 professional singers from NACTA (National Academy of Chinese Theatre Arts, all of them have rich experience in stage performance and teaching) and another 4 amateur singers from jingju associations in non-art schools to the recording sessions. All singers were asked to sing their familiar arias with the constraint that these arias should be representative ones for each school.

Jingju singing is accompanied by an instrumental ensemble including minimum 8 melodic and percussive instruments in heterophony. It is customary that when it is short of instrument players or in the scenario of singing practice, only the primary instrument of the ensemble - *jinghu*, is used as the accompaniment because the other melodic instruments follow the melody played by the *jinghu* [11]. 7 singers (3 professional and 4 amateur) were singing along with the accompaniment of commercial audio recordings; other 2 professional singers were accompanied by 2 professional *jinghu* players. We show the detail information of the singers and *jinghu* players in table 2.

2.2 Recording setup

Most of the recording sessions have been conducted in professional recording rooms by using professional equipment, see table 1 for the detail information, where we use two recording equipment sets and two recording rooms:

- Set 1: M-Audio Luna condenser microphone + RME Fireface UCX audio interface + Apple GarageBand for Mac DAW;
- Set 2: Mojave MA-200 condenser microphone + ART voice channel microphone preamp + RME Fireface 800 audio interface + Adobe Audition 2.0 DAW;
- Room 1: The conference room in NACTA's business incubator with reflective walls, carpet-covered floor, conference furniture and medium room reverberation;
- Room 2: The sound recording studio in Institute of Automation, Chinese Academy of Science, with acoustic absorption and isolation.

Table 1: Equipment and rooms for recording each artist.

Recording equipment	Recording rooms	Artists' names
set 1	room 1	LIAO Jiani
Partly with set 1 in room 1 and partly with set 2 in room 2		SHAO Yakun
set 2	room 2	SUN Yuzhu, SONG Ruoxuan, TIAN Hao, LONG Tianming, LIU Hailin, SONG Weihao, XU Jingwei, ZHANG Lantian, FU Yanchen

When commercial audio recordings were used as the accompaniment, singers were recorded while listening to the accompaniment sent through their monitoring headphone. Otherwise, when *jinghu* players were used as the accompaniment, to simultaneously record both singing and *jinghu* without crosstalk, we placed them separately in two different recording rooms and used two recording channels. However, they were still able to have visual communication through a window and monitor each other through headphones.

3 DESCRIPTION OF THE DATASET

In total, 21 recording sessions were conducted, which resulted in a dataset containing around 9 hours audio recordings of 120 arias - 74 of them are sung by professional singers and 46 are sung by amateur singers. Table 3 shows the distribution of aria recordings per role-type and *shengqiang*; *banshi* is not included here because some arias contain more than one. However, since the main melodic unit is the line, the information given in Table 4 is a better representation of the dataset's potential for the study of the automatic jingju singing evaluation.

3.1 Coverage, completeness, quality and reusability

As stated previously, the purpose of the jingju a cappella singing audio dataset is to offer a comprehensive and complete resource for the study of automatic jingju singing evaluation as described in section 1. Based on this purpose, we evaluate four criteria for corpus creation - coverage, completeness, quality and reusability, as defined by [9].

Coverage: The dataset includes the three main role-types - *laosheng*, *dan* and *jing*. For *laosheng* and *dan* role-types, both professional and amateur singings have been recorded. 96 *laosheng* and

Table 2: Detail information of the recording artists, according to role-type, singing school, affiliation, level and use of accompaniment. NACTA: National Academy of Chinese Theatre Arts, USTB: University of Science and Technology Beijing, Renmin: Renmin University of China

		Artist's name	Role-type	Singing school	Affiliation	Level	Use of accompaniment
Singers	Professional	SUN Yuzhu (female)	*dan*	CHENG Yanqiu	NACTA	4th year undergraduate	audio recordings
		LIAO Jiani (male)	*laosheng*	YU Shuyan & YANG Baoseng		4th year undergraduate	audio recordings
		SHAO Yakun (female)	*jing*	QIU Shengrong		4th year undergraduate	audio recordings
		SONG Ruoxuan (female)	*dan*	MEI Lanfang		graduated	FU Yanchen
		TIAN Hao (male)	*laosheng*	YU Shuyan		graduated	ZHANG Lantian
	Amateur	LIU Hailin (male)	*dan*	CHENG Yanqiu	USTB	–	audio recordings
		SONG Weihao (male)	*dan*	XUN Huisheng	USTB	–	audio recordings
		LONG Tianming (male)	*laosheng*	YU Shuyan & YANG Baoseng	USTB	–	audio recordings
		XU Jingwei (male)	*laosheng*	–	Renmin	–	audio recordings
Jinghu players	Professional	ZHANG Lantian	–	–	NACTA	3rd year undergraduate	–
		FU Yanchen	–	–		graduated	–

Table 3: Content of the jingju a cappella audio dataset, according to role-type and *shengqiang*, Format of each cell: professional|amateur aria number

	laosheng	*dan*	*jing*	Total
xipi	16|17	19|4	10|0	45|21
erhuang	13|10	4|8	4|0	21|18
sipingdiao	–	4|2	–	4|2
nanbangzi	–	2|1	–	2|1
fanerhuang	0|2	1|1	–	1|3
fansipingdiao	–	1|1	–	1|1
Total	29|29	31|17	14|0	74|46

105 *dan* melodic lines in the dataset were sung both by professional and amateur singers, which allows us to analyze vocal techniques between the professional and amateur versions of the same melodic lines, and build computational similarity models by using these lines. The dataset also includes the two main *shengqiang* - *xipi* and *erhuang*, and a few auxiliary ones, such as *sipingdiao*, *nanbangzi*, where the information of the auxiliary *shengqiang* is not presented in table 4. In terms of *banshi*, the whole range of metered ones is represented in the dataset - *yuanban*, *manban*, *kuaiban*, *erliu*, *liushui*, *sanyan* and its three variations. Besides these metered *banshi*, there are a few auxiliary ones, whose occurrence is very punctual in performance. The thorough coverage of the most common *shengqiang*, *banshi* allows to train a singing evaluation model with good generalization ability.

Completeness: The dataset contains the metadata of the recordings and annotations both at the recording and the line level, organized in separate spreadsheets. For the recordings, the metadata contains the title of the work in Chinese, role-type, *shengqiang*, *banshi*, whether it contains *jinghu* accompaniment. As for the lines, each of them is annotated with the role-type, *shengqiang*, *banshi*, line type, that is, opening or closing, the lyrics for the whole line and the related score in the score collection [8].

Quality: We conducted a small number of the recordings in room 1 and these recordings contain medium room reverberation and minor background noise. However, apart from those, the other recording sessions have been done in room 2 and those recordings are dry, clean and of good quality.

Reusability: The a cappella singing and *jinghu* accompaniment audio recordings, their metadata are available online for research. Due to copyright issues, the commercial accompaniment audio recordings are available on request. All the audio and metadata files in this dataset are licensed under Creative Commons Attribution-NonCommercial 4.0 International.

3.2 Integration in the corpus

The scores in our corpus have been gathered with the purpose of studying jingju singing regarding its musical system elements [8]. The a cappella audio recordings and the scores in our corpus are related through the melodic line which both of them represent. As described in [8], scores and recordings are not directly related, however, the melodic contour and the lyrics are common in the majority of the pieces. Taking into account these considerations, 257 of the 705 lines (36.31%) for *laosheng* and 180 of the 512 lines (35.16%) are related between the score collection and the a cappella singing audio dataset. The related scores are vital in the automatic singing evaluation algorithm since the score-informed method is used for the syllable segmentation step (figure 1).

Table 4: Content of the jingju a cappella audio dataset per melodic line for role-types of *dan*, *laosheng* and *jing*, according to *shengqiang* and *banshi*. On the upper heading, *da* stands for *dan*, *ls* for *laosheng*, *eh* for *erhuang* and *xp* for *xipi*. Format of each cell: professional|amateur melodic line number (related lines in the score collection [8]).

	daeh	*daxp*	*lseh*	*lsxp*	*jieh*	*jixp*	Total
yuanban	0|27	58|22 (20|22)	54|31 (39|20)	38|19 (6|0)	26|0	33|0	209|99 (65|42)
manban	12|44 (0|8ft	14|4 (0|4)	35|25 (32|12)	39|28 (18|28)			100|101 (50|52)
kuaiban	–	6|0	–	50|29	–	58|0	114|29
sanyan	7|6	–	17|0	–	–	5|0	29|6
kuaisanyan	16|16	–	22|13 (22|11)	–	10|0	–	48|29 (22|11)
zhongsanyan	–	–	0|4	–	–	–	0|4
mansanyan	–	–	6|6 (6|6)	–	–	–	6|6 (6|6)
erliu	–	39|12 (7|12)	–	22|84	–	8|0	69|96 (7|12)
liushui	–	85|45 (61|45)	–	32|34 (32|23)	–	17|0	134|79 (93|68)
daoban	1|0	2|0	2|0	4|2	1|0	5|0	15|2
sanban	–	15|0	1|0	1|9 (0|1)	2|0	7|0	26|9 (0|1)
yaoban	1|0	2|1 (0|1)	1|0 (1|0)	21|12	–	9|0	34|13 (1|1)
huilong	1|0	–	4|0	–	1|0	–	6|0
Total	38|93 (0|8)	221|84 (88|84)	142|79 (100|49)	207|217 (56|52)	40|0	142|0	792|473 (293|185)

3.3 Potential of the dataset

Apart from the great potential for automatic singing evaluation, the dataset allows many other sorts of musicological research. Since the dataset contains partial recordings of the *jinghu* accompaniment, they are a useful resource for the analysis of the performing interaction between the singing and *jinghu* lines. Some *laosheng* and *dan* arias were sung by both female and male singers, and they will be of benefit for analyzing differences in male and female timbre in these role-types. The scores and recordings which share lines allow a combined analysis, such as linguistic tone and melody relationship analysis. Finally, the audio recordings along with their annotations also will be beneficial for some basic MIR research tasks on jingju singing, such as melody extraction, structural segmentation, key detection and audio-to-lyrics alignment.

4 CONCLUSIONS

In this paper, we have presented a jingju a cappella singing audio dataset for the study of automatic jingju singing evaluation. This dataset presents both professional and amateur singings and the most common *shengqiang*, *banshi* and role-types. It has been integrated into our existing CompMusic jingju corpus. Some potential usages of this dataset apart from the automatic singing evaluation have been discussed. The audio dataset, together with its annotated metadata per aria and melodic line are openly avaiable online.

In the future work, we intend to extend the dataset and increase the shared melodic lines between professional and amateur singings by recording other amateur singers. At the same time, we will exploit the potential of this dataset by annotating it in terms of melodic line and syllable boundaries, then retrain the phoneme acoustic model and syllable segmentation model presented in section 1.3. Finally, we plan to conduct perceptual experiments to measure the similarities between professional and amateur singing syllables. These similarities along with the audio recordings will be used as the training dataset to build automatic jingju singing evaluation models.

ACKNOWLEDGMENTS

This research was funded by the European Research Council under the European Unionfis Seventh Framework Program, as part of the CompMusic project (ERC grant agreement 267583). We are thankful to WANG Xin for providing the recording equipment and to LIU Yiting for providing the recording room in NACTA.

REFERENCES

[1] Black Dawn A.A., Li Ma, and Tian Mi. August, 2014. Automatic identification of emotional cues in Chinese opera singing. In *Proc. of the 13th International Conference on Music Perception and Cognition and the 5th Conference for the Asian-Pacific Society for Cognitive Sciences of Music (ICMPC 13-APSC0M 5)*. Seoul, South Korea.

[2] Pritish Chandna, Marius Miron, Jordi Janer, and Emilia Gómez. 2017. Monoaural audio source separation using deep convolutional neural networks. In *International Conference on Latent Variable Analysis and Signal Separation*. Grenoble, France, 258–266.

[3] Rong Gong, Nicolas Obin, Georgi Dzhambazov, and Xavier Serra. 2017. Score-Informed Syllable Segmentation for Jingju a Cappella Singing Voice with Mel-Frequency Intensity Profiles. In *International Workshop on Folk Music Analysis*. Málaga, Spain.

[4] Rong Gong, Jordi Pons, and Xavier Serra. 2017. Audio to Score Matching by Combining Phonetic and Duration Information. In *18th International Society for Music Information Retrieval Conference*. Suzhou, China.

[5] Rong Gong, Yile Yang, and Xavier Serra. 2016. Pitch Contour Segmentation for Computer-aided Jingju Singing Training. In *Proceedings of the Sound and Music Computing Conference 2016*. Hamburg, Germany.

[6] Jordi Pons, Rong Gong, and Xavier Serra. 2017. Score-informed Syllable Segmentation for A Cappella Singing Voice with Convolutional Neural Networks. In *18th International Society for Music Information Retrieval Conference*. Suzhou, China.

[7] Rafael Caro Repetto and Xavier Serra. 2014. Creating a Corpus of Jingju (Beijing Opera) Music and Possibilities for Melodic Analysis. In *15th International Society for Music Information Retrieval Conference*. Taipei, Taiwan, 313–318.

[8] Rafael Caro Repetto and Xavier Serra. 2017. A Collection Of Music Scores for Corpus Based Jingju Singing Research. In *18th International Society for Music Information Retrieval Conference*. Suzhou, China.

[9] Xavier Serra. 2014. Creating Research Corpora for the Computational Study of Music: the case of the CompMusic Project. In *AES 53rd International Conference on Semantic Audio*. AES, AES, London, UK, 1–9.

[10] Mi Tian and Mark B. Sandler. 2016. Towards Music Structural Segmentation Across Genres: Features, Structural Hypotheses, and Annotation Principles. *ACM Trans. Intell. Syst. Technol.* 8, 2, Article 23 (Oct. 2016), 19 pages.

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