Quantitative analysis of the relationship between linguistic tones and melody in jingju using music scores

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

Shuo Zhang
Music Technology Group,
Universitat Pompeu Fabra
Barcelona
ssz6@georgetown.edu

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

ABSTRACT

When lyrics of tonal languages are set to music, the pitch contour of the tones has to agree to a certain extent with the melodic contour to assure intelligibility. The relationship between the linguistic tones of the complex dialectal construct used in jingju (commonly known as Beijing or Peking opera) and its melody has been largely studied, but not definite consensus has been achieved among scholars. After reviewing the related literature, we present a first approach for the quantitative analysis of the relationship between linguistic tones and melody in jingju using a collection of machine readable music scores with tone category annotations for 7,283 syllables. We describe two statistical analyses performed in this collection regarding the melodic contour for each syllable and the pitch height relationship in 5,494 pairs of consecutive syllables. We argue that the obtained results contribute to supporting claims from the literature and complementing others, although some limitations of the approach might nuance the confidence of their validity.

CCS CONCEPTS

•**Information systems → Digital libraries and archives;** •**Applied computing → Sound and music computing;**

KEYWORDS

jingju music, Chinese linguistic tones, music scores analysis, corpus-driven research
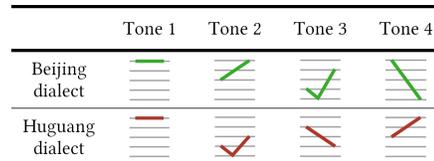
Figure 1: Pitch contour for each tone category in BJ and HG. Horizontal lines represent the division of speech pitch range into five levels.

pitch contours to assure understandability and avoid misconceptions. In jingju (also known as Beijing or Peking opera) circles, the importance of the agreement between these two elements is stated through the common saying *zi zheng qiang yuan* [2, 9, 12], literally meaning "correct characters, round tune". According to Yu Huiyong's explanation of this expression, in jingju, as in other sung genres, both language and music have their own set of rules, and a good musical arrangement has to respect both of them [2]. Failing in doing so can cause *dao zi*, literally "an inverted character", one that is difficult to understand, or even confused with a different one. This topic in jingju has been studied for over a century, but no consensus has been reached among scholars [15]. One of the reasons is the complexity of the linguistic construct used in jingju, which integrates elements from two dialects, Beijing (BJ) and Huguang[1] (HG),[2] which share the same four tonal categories,[3] but with different pitch contour realizations (Figure 1). However, there are no specific rules that state when a character should be pronounced as BJ or HG. Tonal information is related to melody in two ways, by the melodic contour of the sung syllable (Figure 2, a), or by the pitch height relationship between two consecutive syllables. In the last case, some authors consider the first note in the melodic contour of each syllable [3, 12] (Figure 2, b), while others, the last of the first syllable and the first of the following one [9, 10] (Figure 2, c). In this paper, we will refer to these two relationship types as syllabic contour and pairwise relationship respectively.

The consequence of the complexities described above, and the lack of consensus among scholars, is that there is no commonly

1 INTRODUCTION

When lyrics in tonal languages, such as Chinese are set to music, there needs to be an agreement between the linguistic and melodic

---

[1]Former term for designing the region corresponding to today's Hubei and Hunan provinces.
[2]Many authors [2, 3, 9, 11] also speak of the influence of the *Zhongzhou yun*, a literary dialect from central China that works as a linguistic framework for BJ and HG.
[3]In the literature they are usually called with their classical names, namely *yinping*, *yangping*, *shangsheng*, and *qusheng*. In this paper we will name them by numbers, respectively 1, 2, 3 and 4, according with contemporary linguistics. Some authors [3, 11, 12] analyze the influence in jingju of a tone category currently lost in BJ and HG, the *ru* tone, which mostly has rhythmic implications
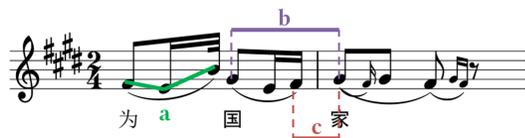
Figure 2: Example of syllabic contour (a), first-first pairwise relationship (b) and last-first pairwise relationship (c).

agreed general model to explain the relationship between linguistic tones and melody in jingju. However, from the literature reviewed for this study [2, 3, 8–12], we have observed some partial agreements: tone 1 is the most consistently described, as sung higher than neighboring syllables and with a lightly descending syllabic contour. Tone 2 is usually described as "starting low" (*di chu*) to then ascend. Tone 3 is the least defined. Tone 4 is commonly described as ascending, but with other possible syllabic contours. In this paper, we take this observations as our reference model.

When explaining the lack of consensus among Chinese scholars, Xu Zheng [15] argues that "the majority of authors obtained their understanding of linguistic tones in jingju through their personal experience, and this experience in its majority was obtained by listening or imitating."[4] Chinese sources [2, 3, 8–12][5] generally present a list of prescriptive rules, illustrated with several selected examples, without further analyses of full arias. Studies from outside China [4, 7] adopt an analytical approach, but based in one or few arias. We argue that a computational approach to this topic can contribute to the state of the art qualitative descriptions with empirical quantitative analyses performed on extensive data. In previous research [13, 14], we have analyzed 20 aria recordings, covering 1993 syllabic contours and approximately 1300 pairwise relationships. This approach presented some limitations, like a time consuming process of manual annotation, unsatisfactory results of predominant pitch extraction algorithms, and difficulty for pitch class definition due to the rich ornamentation, long glides and wide vibratos of jingju singing style.

In this paper we present a first approach to the automatic analysis of the relationship between linguistic tones and melody in jingju using machine readable music scores. In the remainder of the paper, first we briefly describe the score collection, detailing its benefits for this task, but also its limitations. Then, we introduce the automatic analyses performed in the collection, whose results are presented and discussed in the following section. The paper finishes with some concluding remarks.

2 METHODOLOGY

2.1 Dataset

Our study uses the Jingju Music Scores Collection (JMSC) recently built [6] as part of the CompMusic Jingju Music Corpus [5]. It contains 92 scores manually created using MuseScore and exported to MusicXML format, covering 897 lyrics lines. It is accompanied with related metadata and annotations regarding lyrics, elements

of the jingju musical system, line types, and the start and end offsets in the score for both the line and each of its three sections.[6] The tone categories of every syllable in each lyrics line have been computed automatically using Native-to-Pinyin,[7] a java utility developed by the second author based on the pinyin4j library. The results, manually curated by the first author and a Chinese native speaker knowledgeable about jingju, are added to the JMSC annotations. The neutral tone (without specific pitch contour) has also been annotated as tone 5. Thus, JMSC contains linguistic tones annotations for 7,283 syllables. The score collection, together with the metadata and annotations, is openly available for research purposes [6]. In this paper, we present the first use of the JMSC and these tone categories annotations, for the automatic analysis of the relationship between linguistic tones and melody in jingju.

2.2 Analyses

We have implemented two different analyses on the score collection, corresponding to the two relationship types described previously, namely syllabic contour and pairwise relationship. We have used the music21 toolkit [1], and the code is openly available[8].

Yu Huiyong [2] classifies syllabic contours as Level (L), Ascending (A), Descending (D), or a combination of the last two (AD, DA). Our analysis consists in assigning one of these labels to each syllable, adding a default-Level (dL) class, for those syllables that are sung with a single note. We take a rule base approach, that looks at the first and last notes of each syllable's melodic contour, and the intermediate peak that is the furthest from them (Figure 2-a is an example of DA). For those syllables which are sung with a melismatic prolongation called *tuoqiang*, we analyze only the first three notes, since some authors argue that tonal information in these cases is conveyed at the beginning of the syllabic contour [2, 10]. For detecting *tuoqiang*, we assign a duration threshold to each combination of *shengqiang* (melodic framework) and *banshi* (metrical pattern) in the scores collection, after running a melodic density analysis [6]. Finally, mordents, the most frequent ornament, defined in this paper as a semiquaver or shorter note between two of the same pitch, have been omitted.

For the pairwise relationship we have considered consecutive syllables in each of the three sections into which a lyrics line is usually divided, and whose start and end offsets are also annotated in the collection. Since these sections are usually separated by an instrumental interlude, the perceptual pitch relationship between the previous and following melodic contours might be lost. We consider the two relationship types described in the literature, namely last-first (Figure 2, c) and first-first (Figure 2, b), and assign a level (L), ascending (A) or descending (D) label according to the intervallic direction between the two notes in the pair.

As a final remark, an important issue arose regarding grace notes in both analyses. In jingju is common the use of grace notes either before or after a main note. However, the grace notes in the score loaded to music21 do not contain information about to which main note they relate. To avoid false relations, grace notes have been disregarded in our analyses.

3 RESULTS AND DISCUSSION

Table 1 shows the results for the analysis of 7,283 syllabic contours. We don't report results for tone 5 (just 2% of the the syllables) since it is not analyzed in the literature and linguistically has no specific pitch contour, but it depends on the preceding syllable. dL is the most frequent syllabic contour for each tone category, arguably due to the fact that 36.9% of the lines in the scores collection are sung in *liushui* and *kuaiban*, two *banshi* (metrical patterns) with a very syllabic style of singing. Since in these cases the tonal information might be expressed through pairwise relationships, in this discussion we focus on the rest of pitch contours. Table 1 shows that AD and DA pitch contours are not predominant in any tone category, and that L is the least frequent one in each of them except for tone 1, since its linguistic pitch contour is level both in BJ and HG (Figure 1). This tone's most frequent syllabic contour is D, as stated in the reference model described in section 1. In tone 1, dL presents a remarkably higher percentage than in other tones, suggesting that it might be sung with a single note also in *banshi* other than *liushui* and *kuaiban*. The most common syllabic contour for tones 2 and 3 is respectively A and D, although their inverted realizations, D and A, also present a high percentage. The most frequent syllabic contour for tone 4 is A.

Results for pairwise relationship analysis are shown in table 2, covering 5,494 pairs. For the reasons stated above, tone 5 is not considered here. Comparing last-first and first-first relationships, the most important changes occur in pairs involving tone 5, like 2-5, 3-5 and 4-5, as well as pair 4-2, a fact that might indicate that the first note of the second syllable lies within the pitch range of the previous one. Otherwise, first-first generally presents a percentage increase in the predominant direction of each pair, with the exception of 2-1, 2-2, 3-1 and 3-3, but highly significant in 1-2, 1-3, 1-4, 4-3, and 4-4. Direction L presents a low percentage in all cases except for 1-1 in the first-first relationship. The predominance of A in last-first relationship and L in first-first for 1-1 agrees with the predominant D pitch contour of tone 1 (Table 1), suggesting that both tones 1 start in the same pitch, but the first descends. Except for 1-1, tone 1 is predominantly sung higher than the preceding or following syllable, as observed in the reference model. Tone 2 is preferably sung lower than preceding or following syllables, with the exception of 4-2 in the first-first relationship. The percentage difference between the two relationship types for this pair can be explained by tone 4's predominantly ascending syllabic contour (Table 1), so that the following tone 2's starting pitch lies within tone 4's pitch range. When two tones 2 are consecutive, the first one is generally sung higher. Tone 3 tends to be sung higher than the preceding or following syllable, except if it is a tone 1. However, the percentage is much lower than for tone 1, suggesting that tone 3 is usually also sung lower than neighboring syllables. When two tone 3 are consecutive, the percentage difference between A and D is not significant enough, especially in the first-first relationship, indicating that both directions are almost equally present. Finally, tone 4 tends to be sung higher than a preceding tone 2 or other tone 4, but lower than a preceding or following tone 3, although the inverted direction is also quite frequent in each case, except for 4-3 and 4-4 in the first-first relationship, arguably due to the predominant ascending syllabic contour of tone 4.

All these results, however, should be taken with caution, due to a technical and a conceptual reasons. The omission of grace notes might affect the results since, as some authors point out, the tonal information might be conveyed via small melodic inflections [2, 10]. But it has to be considered that music scores in jingju, a tradition orally transmitted, are not authoritative works of a composer, but *a posteriori* transcriptions of a performance. Therefore, the use of grace notes or other symbols for the notation of these inflections, or the very decision of transcribing them at all, lies with the transcriber's interpretation and purpose. Consequently, even if the grace notes were included, there will always exist the possibility that the tonal information is distorted or even lost in the score.

4 CONCLUSIONS

Although the results here presented should be taken with caution, we argue that they help to observe some general trends that can contribute to the understanding of the relationship between linguistic tones and melody in jingju. Firstly, this relationship is uneven depending on tones categories. Tone 1 seems to be the more consistent one, with a predominantly descending syllabic contour and sung higher in any pairwise relationship. Tone 2 tends to be ascending and sung lower, but less consistently. In case of tones 3 and 4, predominant pitch contours are respectively descending and ascending, but results do not allow to determine a clearly predominant pairwise relationship. These results supports Stock's claim that the relationship between linguistic tones and melody is "neither automatic nor complete" [7]. Zhang Zaifeng [12] makes this claim even bolder, arguing that "in jingju singing, generally speaking, the only consideration is that when tone 1 is sung together with other tones, it shouldn't be sung lower that tones 2, 3, and 4, and thus there won't be any inverted character [*dao zi*] situation." As for which relationship is more informative for pairwise relationship analysis, results from first-first seem to provide a slightly better support for the observations in reference model. We might argue though that both of them complement each other, as it might be the case in 4-2, what suggests that listeners can obtain cues from both relationships to perceive tonal information. Finally, regarding the use of BJ and HG, our results seem to support a preference for HG, according to predominant syllabic contours of tones 3 and 4, and tone 2 preferred to be sung lower and tone 3 higher than neighboring syllables. However, percentages are not significant enough as to refute the influence of BJ.

Future research on this topic could considerably benefit from perceptual studies of jingju performers and listeners. But in any case, it must be considered that the melodies performed nowadays have been transmitted orally through a system that encourages imitation, and therefore the perceptual sensibility to tone categories, specially related to HG, no longer spoken by many performers and listeners today, might differ from that of previous generations.

Table 1: Pitch contour analysis results (Overall percentage / Percentage excluding dL)

	Tone 1 (1645)		Tone 2 (1724)		Tone 3 (1516)		Tone 4 (2252)		Tone 5 (146)
dL	699	(42.49%)	511	(29.64%)	522	(34.43%)	664	(29.48%)	
L	156	(9.48%/16.49%)	98	(5.68%/8.08%)	81	(5.34%/8.15%)	139	(6.17%/8.75%)	
A	110	(6.69%/11.63%)	**458**	**(26.57%/37.76%)**	281	(18.54%/28.27%)	**665**	**(29.53%/41.88%)**	
D	**471**	**(28.63%/49.79%)**	338	(19.61%/27.86%)	**368**	**(24.27%/37.02%)**	326	(14.48%/20.53%)	
AD	112	(6.81%/11.84%)	192	11.14%/15.83%)	146	(9.63%/14.69%)	269	(11.94%/16.94%)	
DA	97	(5.9%/10.25%)	127	(7.37%/10.47%)	118	(7.78%/11.87%)	189	(8.39%/11.9%)	

Table 2: Pairwise relationships analysis results (Percentage of each direction (column) per pair (row))

	last - first						first - first					
	A		L		D		A		L		D	
1-1	**126**	**(38.89%)**	114	(35.19%)	84	(25.93%)	51	(15.74%)	**160**	**(49.38%)**	113	(34.88%)
1-2	36	(8.85%)	96	(23.59%)	**275**	**(67.57%)**	10	(2.46%)	65	(15.97%)	**332**	**(81.57%)**
1-3	32	(18.29%)	33	(18.86%)	**110**	**(62.86%)**	17	(9.71%)	23	(13.14%)	**135**	**(77.14%)**
1-4	66	(18.23%)	70	(19.34%)	**226**	**(62.43%)**	33	(9.12%)	45	(12.43%)	**284**	**(78.45%)**
1-5	4	(16.0%)	3	(12.0%)	**18**	**(72.0%)**	3	(12.0%)	1	(4.0%)	**21**	**(84.0%)**
2-1	**248**	**(83.5%)**	16	(5.39%)	33	(11.11%)	**240**	**(80.81%)**	23	(7.74%)	34	(11.45%)
2-2	70	(21.88%)	83	(25.94%)	**167**	**(52.19%)**	97	(30.31%)	79	(24.69%)	**144**	**(45.0%)**
2-3	**116**	**(56.04%)**	28	(13.53%)	63	(30.43%)	**118**	**(57.0%)**	42	(20.29%)	47	(22.71%)
2-4	**166**	**(43.01%)**	71	(18.39%)	149	(38.6%)	**179**	**(46.37%)**	59	(15.28%)	148	(38.34%)
2-5	16	(37.21%)	8	(18.6%)	**19**	**(44.19%)**	**22**	**(51.16%)**	5	(11.63%)	16	(37.21%)
3-1	**192**	**(83.84%)**	16	(6.99%)	21	(9.17%)	**183**	**(79.91%)**	21	(9.17%)	25	(10.92%)
3-2	71	(27.52%)	50	(19.38%)	**137**	**(53.1%)**	74	(28.68%)	34	(13.18%)	**150**	**(58.14%)**
3-3	89	(38.03%)	43	(18.38%)	**102**	**(43.59%)**	96	(41.03%)	39	(16.67%)	**99**	**(42.31%)**
3-4	143	(35.93%)	90	(22.61%)	**165**	**(41.46%)**	131	(32.91%)	84	(21.11%)	**183**	**(45.98%)**
3-5	**20**	**(48.78%)**	3	(7.32%)	18	(43.9%)	15	(36.59%)	4	(9.76%)	**22**	**(53.66%)**
4-1	**342**	**(84.24%)**	29	(7.14%)	35	(8.62%)	**360**	**(88.67%)**	25	(6.16%)	21	(5.17%)
4-2	146	(32.44%)	88	(19.56%)	**216**	**(48.0%)**	**211**	**(46.89%)**	104	(23.11%)	135	(30.0%)
4-3	**167**	**(44.53%)**	45	(12.0%)	163	(43.47%)	**231**	**(61.6%)**	59	(15.73%)	85	(22.67%)
4-4	**206**	**(44.21%)**	73	(15.67%)	187	(40.13%)	**254**	**(54.51%)**	82	(17.6%)	130	(27.9%)
4-5	16	(43.24%)	3	(8.11%)	**18**	**(48.65%)**	24	**(64.86%)**	4	(10.81%)	9	(24.32%)
5-1	**13**	**(100.0%)**					**13**	**(100.0%)**				
5-2	3	(21.43%)	3	(21.43%)	**8**	**(57.14%)**	5	(35.71%)	3	(21.43%)	**6**	**(42.86%)**
5-3	1	(14.29%)			**6**	**(85.71%)**	2	(28.57%)			**5**	**(71.43%)**
5-4	**11**	**(55.0%)**	3	(15.0%)	6	(30.0%)	**12**	**(60.0%)**	3	(15.0%)	5	(25.0%)

REFERENCES

[1] Michael Scott Cuthbert and Christopher Ariza. 2010. music21: A Toolkit for Computer-Aided Musicology and Symbolic Music Data. In *11th International Society for Music Information Retrieval Conference (ISMIR 2010)*. Utrecht, Netherlands, 637–642.

[2] Yu Huiyong. 2008. *Qiang ci guanxi yanjiu*. Zhongyang yinyue xueyuan chubanshe, Beijing.

[3] Liu Jidian. 1992. *Jingju yinyue gailun*. Renmin yinyue chubanshe, Beijing.

[4] Rulan Chao Pian. 1972. Text Setting with the *Shipyi* Animated Aria. In *Words and Music: The Scholar's View*, Laurence Berman (Ed.). Harvard University Press, Cambridge, MA, 237–270.

[5] 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 (ISMIR 2014)*. Taipei, Taiwan, 313–318.

[6] 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 (ISMIR 2017)*. Suzhou, China. In press.

[7] Jonathan P. J. Stock. 1999. A reassessment of the relationship between text, speech tone, melody, and aria structure in Beijing Opera. *Journal of Musicological Research* 18, 3 (1999), 183–206. https://doi.org/10.1080/01411899908574757

arXiv:http://dx.doi.org/10.1080/01411899908574757

[8] Shu Tonglin. 2011. *Yu Shuyan "shiba zhang ban" changqiang qianxi*. Tianjin guji chubanshe, Tianjin.

[9] Zhu Weiying. 2004. *Xiqu zuoqu jifa*. Vol. 1. Renmin yinyue chubanshe, Beijing.

[10] Elizabeth Wichmann. 1991. *Listening to Theatre*. University of Hawaii Press, Honolulu.

[11] Fei Yuping. 2005. *Jingju zuoqu jifa*. Zhongguo xiqu chubanshe, Beijing.

[12] Zhang Zaifeng. 2011. *Zenyang chang hao jingju*. Hunan wenyi chubanshe, Changsha.

[13] Shuo Zhang, Rafael Caro Repetto, and Xavier Serra. 2014. Study of the Similarity Between Linguistic Tones and Melodic Pitch Contours in Beijing Opera Singing. In *15th International Society for Music Information Retrieval Conference (ISMIR 2014)*. Taipei, Taiwan, 343–348.

[14] Shuo Zhang, Rafael Caro Repetto, and Xavier Serra. 2015. Predicting Pairwise Pitch Contour Relations Based on Linguistic Tone Information in Beijing Opera Singing. In *16th International Society for Music Information Retrieval (ISMIR) Conference (ISMIR 2015)*. Málaga, Spain, 107–113.

[15] Xu Zheng. 2007. Jiushi nian lai jingju de shengdiao yanjiu zhi huigu. *Nankai yuyan xuekan* 10 (2007), 39–50.