STUDY OF THE SIMILARITY BETWEEN LINGUISTIC TONES AND MELODIC PITCH CONTOURS IN BEIJING OPERA SINGING

Shuo Zhang, Rafael Caro Repetto, Xavier Serra
Music Technology Group, Universitat Pompeu Fabra
ssz6@georgetown.edu, {rafael.caro, xavier.serra}@upf.edu

ABSTRACT

Features of linguistic tone contours are important factors that shape the distinct melodic characteristics of different genres of Chinese opera. In Beijing opera, the presence of a two-dialectal tone system makes the tone-melody relationship more complex. In this paper, we propose a novel data-driven approach to analyze syllable-sized tone-pitch contour similarity in a corpus of Beijing Opera (381 arias) with statistical modeling and machine learning methods. A total number of 1,993 pitch contour units and attributes were extracted from a selection of 20 arias. We then build Smoothing Spline ANOVA models to compute matrixes of average melodic contour curves by tone category and other attributes. A set of machine learning and statistical analysis methods are applied to 30-point pitch contour vectors as well as dimensionality-reduced representations using Symbolic Aggregate approXimation(SAX). The results indicate an even mixture of shapes within all tone categories, with the absence of evidence for a predominant dialectal tone system in Beijing opera. We discuss the key methodological issues in melody-tone analysis and future work on pair-wise contour unit analysis.

1. INTRODUCTION

Recent development in signal processing and cognitive neuroscience, among other fields, has revived the research on the relationship between speech and musical melody [10]. Singing in tone languages offers a particularly convenient entry point to compare musical and speech melodies, allowing us to gain insight into the ways the prosody of a particular language shapes its music. In a tone language, as opposed to an intonation language, the pitch contour of a speech sound (often a syllable) can be used to distinguish lexical meaning. In singing, however, such pitch contour can be overridden by the melody of the music, making the lyrics difficult to decode by listeners.

In such consideration, musicologists have observed that features of the prosody of the local dialect often play an important role in shaping the melodic characteristics of the regional operas in China [9, 15]. On the other hand, it is generally assumed that Beijing opera had incorporated linguistic tone systems from both the Hu-Guang (HG) dialect and Beijing (BJ) dialect [22]. [1] Xu [19] reviewed 90 years of research on the dialect tone system in Beijing opera, and concluded that there is no agreement as to which system is predominant in shaping the melodic characteristics of the genre.

In sum, previous work indicates that the overall degree and manner of the melody-tone relationship is not entirely clear, partly due to the limitation that music scholars typically were not able to go beyond analyzing a few arias by hand [19]. In this paper, we propose a novel approach to melody-tone similarity by applying statistical modeling and machine learning methods to a set of 20 arias selected from a corpus of 381 arias of Beijing opera audio recording. The research questions are defined as follows: (1) How similar are syllable-sized melodic contours within a given tone category? (2) How similar is the ”average” melodic contour to its corresponding prototype contour in speech in the same tone category? (3)Which tone system (BJ or HG) better predicts the shape of melodic contours?

Following preprocessing, we apply clustering algorithms and statistical analysis to 30-point feature vectors of pitch contours, as well as dimensionality-reduced feature vectors represented symbolically using the Symbolic Aggregate approXimation (SAX) algorithm [8]. Special considerations are given to existing hypotheses regarding the distribution of the tone systems in Beijing opera. Lastly, we build Smoothing Spline ANOVA Models to compute matrixes of average melodic contour curves by tone category and other attributes.

2. KEY ISSUES IN STUDYING MELODY-TONE SIMILARITY

2.1 Beijing Opera: Performance Practice

Several features of Beijing opera may explain why the melody-tone relationship remains challenging. First, the composition process of Beijing opera assumes no designated composer for any given opera. Rather, each opera is composed by re-arranging prototype arias from a inventory of arias according to the rhythmic type, role type, tempo, and

Attribution: Shuo Zhang, Rafael Caro Repetto, Xavier Serra . “STUDY OF THE SIMILARITY BETWEEN LINGUISTIC TONES AND MELODIC PITCH CONTOURS IN BEIJING OPERA SINGING”, 15th International Society for Music Information Retrieval Conference, 2014.

[1] A schematic representation of the four tones in these two systems is shown in Figure 1.

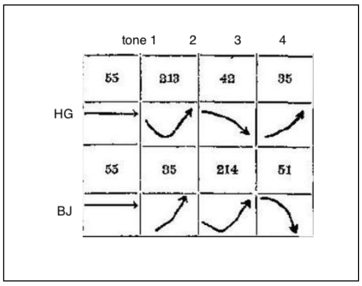

Figure 1. Schematic representation of the BJ and HG tone system

other factors. Lyrics, taken from libretto, are relatively fixed whereas the specific melody may change according to each performer / performance. While there has been evidence [15] that the performers do take special consideration of tones in their singing, it is not clear in what manner and to what degree they are doing this. Next,we discuss several key issues and assumptions as a bridge from previous work to the novel approach proposed in this paper.

2.2 Key Issues in Studying Tone-Melodic Similarity

First, the melody-tone relationship as a problem of tone perception (and production). A key assumption underlying previous works is that speech comprehension in tone language crucially depends on the availability of tone contour information. However, recent development in tone perception and modeling has challenged this view and revealed the great capacity of human listeners in identifying words in tone languages, with the minimum amount of tone information and despite a great amount of variances in tone signals. We consider the following aspects of evidence: (1) tone contour shapes in connected speech deviates significantly from its standard (canonical) shapes due to co-articulation, limit of maximum speed of pitch change in vocal folds, and other factors [18], introducing a great amount of variances; (2) Patel et al [11] demonstrated that Mandarin speech in monotone is over 90% intelligible to native speakers in a non-noise background, pointing to a low entropy (i.e., high predicability) of the information that is carried by segmental contrast in context; (3) Gating experiments [21] have demonstrated that Mandarin speakers are able to correctly detect tone category based on only the initial fractions of second of a tone signal. From these evidence, we should use caution in making the aforementioned assumption about tone information in music. Similarly, we may also expect to find a even larger amount of variation in the syllable-sized melodic contours in a given tone category. [2]

Second, we define the hypotheses and specific goals in this work. We observe that in the review of tone systems in Beijing opera [19] , one key assumption is that one of the two underlying dialectal systems must dominate. However, we also find evidence in the literature [22] that one may expect to find an even mixture of contours from both dialects. [3] In this work, we consider both hypotheses and find our data to be more consistent with the second hypothesis.

3. DATA COLLECTION

3.1 Beijing Opera Audio Data Collection

The music in Beijing opera is mainly structured according to two basic principles, *shengqiang* and *banshi*, which in a broad sense define respectively its melodic and rhythmic components [17]. On top of these two structural principles, the system of role-types impose particular constrains to the execution of *shengqiang* and *banshi*. The interaction of these three components, hence, offers a substantial account of Beijing opera music. Our current collection includes 48 albums, which contain 510 recordings (tracks) featuring 381 arias and over 46 hours of audio [14].

The current study focuses on a small selection of 20 arias from the corpus to serve as a manageable starting point of the melody-tone relationship analysis. This set is selected according to a number of criteria: (1) we selected only *yuanban*, a rhythmic type in which the duration of a syllable sized unit bears the most similarity to that of speech; (2) we selected both types of *shengqiang*, namely *xipi* and *erhuang*; (3) we selected five role types: D(*dan*), J(*jing*), LD(*laodan*), LS(*laosheng*), and XS(*xiaosheng*). For each combination of *shengqiang* and role types, we selected two arias, yielding a total of 20 arias for analysis.

3.2 Data Preprocessing

The vocal frames of the audio recordings of the 20 arias are partially-automatically segmented into syllable sized unit with boundary alignment correction by hand. [4] The segmentation is implemented as timestamps of a TextGrid file in the speech processing software Praat [2]. The textgrid is later integrated with the metadata labels from the annotation process.

Following segmentation, we annotate the audio with lyrics extracted from the online Beijing opera libretto database `jingju.net`. The Chinese-character lyrics files are converted into romanized pinyin form with tone marks in the end (1,2,3, or 4) using an implementation of Java library

[2] We must bear in mind also that speech tones are generated under a different mechanism than pitch contours in singing. For one thing, the latter has a more planned mechanism of design - the composition of the music. In speech, as the qTA model has demonstrated [12], speakers may have a pitch target (defined by a linear equation) in mind during articulation, but the actual F0 realization is subject to a set of much complex physiological and contextual linguistic factors, which may be modeled by a third-order critically damped system [12]. This complication does not exist in music: in singing, a singer can realize the exact F0 target as planned. Therefore, we propose that approaches that directly compute similarity between melodic and linguistic tone F0 contours should be ruled out.

[3] Some cite three dialects [22], HuGuang, Beijing, and ZhongZhou YinYun.

[4] Automatic segmentation using forced-alignment with machine-readable form of the score is currently being developed. For the current study, we used the result of a trained spectral-based classifier [3] that is able to separate the pure instrumental frames of the audio signal from those frames that contain both vocal and instrumental parts. The result of this segmentation is in many cases the voiced segment (vowel) of a syllable, which is precisely the unit of our analysis.

pinyin4j. A Praat Script is implemented to automatically parse the romanized lyrics files and to annotate the segmented audio files. The metadata attributes (*shengqiang*, role type, artist, duration, tone category, and word) are also automatically annotated for each segmented unit.

3.3 Pitch Contour Extraction

We then proceed to the extraction of F0 values for each annotated pitch contours of interest. The F0 is computed using the MELODIA salience function [13] within the Essentia audio signal processing library in Python [1], in order to minimize the interference of background instrumental ensemble to the computation of F0 of the primary vocal signal. All rows of F0 values associated with a specific pitch contour is automatically assigned a unique pitch contour id for the convenience of analysis in later stages. For the sake of analysis, we produce down-sampled 30-point F0 vectors by using equidistant sampling across each pitch contour [5]. All F0 values are normalized so that each contour has a mean F0 of 0 and sd of 1. A 5-point averaging sliding window is applied to smooth the signal (the ends are smoothed with progressively smaller smooths).

4. PROPOSED APPROACH

In this section we overview the methodology employed in the analysis of the extracted pitch contour dataset. As discussed above in 2.2, all methodology are boiled down to addressing the research question (1), which attempts to analyze and describe the variances and clusters found in melodic contours of each tone category and across categories. Research question (2) and (3), both of which involve comparing music with speech melody, can only be addressed indirectly by the average curves computed by the SSANOVA model for each tone category.

4.1 Time Series Representation

Finding the best measure of time series representation in the current tasks requires careful thought and experimentation. The challenge of this task is that, it is somewhat different from a standard pitch contour similarity task addressed in previous MIR work. In a standard task, such as query-by-humming (QBH), the goal of the task is to match the melody as precisely as possible. However, in this task, our goal is in a way to model the human perception of tone. An important capacity of human cognition is its capacity to abstract away the commonalities from groups of pitch contours with much different fine detail variations. In this study, we experiment with the Symbolic Aggregate approXimation (SAX) [8] representation of pitch contour vectors. [6]

Even though SAX representation is mostly used outside of MIR, it has been applied to the QBH task [16]. It transforms the pitch contour into a symbolic representation with a user-designated length (nseg=desired length of the feature vector) and alphabet size (m), the latter being used to divide the pitch space of the contour into m equiprobable segments assuming a gaussian distribution of F0 values. It is in principle very suitable for the current task as discussed above, as it is able to transform the fast-changing time varying signal of pitch contour into a coarse representation of abstract ”shapes”, which models the human cognition [7].

In this work, we rely on the SAX representation (1) as a effective and economic way to represent the shapes of time series in statistical analysis; and (2)as a coarse symbolic representation for clustering. To ensure the validity of SAX to reflect the true shape of the original 30-point vector, we experiment with different parameters and use four different ways to evaluate the effectiveness of the SAX representation (dicussed below).

4.2 Methodology

As discussed in 2.2, we consider two different analytical approaches in this work based on the two hypotheses regarding the distribution of tone systems in Beijing opera.

In the first hypothesis (H1), we assume that there is one predominant tone system (BJ or HG) in Beijing opera. We define a time-series clustering task with the goal of clustering all tone contours into four big clusters, corresponding to four tone categories. Using dynamic time warping (DTW) as the distance measure, we perform K-means Clustering and Agglomerative Clustering (hierarchical) on the 30-point pitch vectors. Using the lower bounding mindist distance measure defined for SAX-based symbolic representation, we also perform K-means Clustering on the SAX string vectors of length 5 (alphabet size is 3).

In the second hypothesis (H2), we expect an even mixture of tone systems and tone shapes in all tone categories. In this scenario, our goal is to perform exploratory cluster analysis on the distribution of contours shapes within each tone categories. More specifically, we perform statistical and clustering analysis on the SAX-based shapes within and across tone categories. In addition, we investigate distribution of attributes associated with each sub-cluster of shape.

We infer from literature [22] that regardless of the distribution of tone systems, the first tone is expected to have the most consistent flat shape if a reaonably strong correlation is assumed between linguistic tone and melodic contour (Notice that tone 1 has the same flat shape across dialects in Figure 1). More specifically, a musicological analysis by hand reveals that the most predominant shape in tone 1 is flat or flat with a final fall (henthforce referred to as Hypothesis 3, or H3, also inferred from rules described in [22]).

[5] The unvoiced part in the beginning of the syllable is skipped in the down-sampling. In addition, the downsampling strategy is also fine tuned in order to filter out the spurious pitch values computed by MELODIA in the beginning portion of the voiced segments.

[6] SAX is the first symbolic representation for time series that allows for dimensionality reduction and indexing with a lower-bounding distance measure. In classic data mining tasks such as clustering, classification, index, etc., SAX is as good as well-known representations such as DWT and DFT, while requiring less storage space. [8].

[7] Strictly speaking, the Gaussian assumption is not met in the pitch space musical notes. However, due to the nature of the task that does not require precise mapping, we use the original SAX implementation without revising the Gaussian assumption.

Lastly, we build a Smoothing Spline ANOVA model with the goal of (1) computing average pitch contours for each tone category, and (2) quantifying the variances accounted for by each predictor variable in different tone categories. Smoothing splines are essentially a piecewise polynomial function that connects discrete data points called knots. It includes a smoothing parameter to find the best fit when the data tend to be noisy, estimated by minimizing the following function:

$$G(x) = \frac{1}{n}\sum_{all\ i}(y_i - f(x_i))^2 + \lambda \int_a^b (f''(u))^2\,du \qquad (1)$$

where n is the number of data points, λ is the smoothing parameter, and a and b are the x coordinates of the endpoint of the spline.

The Smoothing Spline ANOVA (SSANOVA) is of the following form, each component of f is estimated with a smoothing spline:

$$\begin{aligned} f = \mu + \beta x + main\,group\,effect + smooth(x) \\ + smooth(x;group) \end{aligned} \qquad (2)$$

where the main group effects correspond to the smoothing splines for each dataset, smooth(x) is the single smoothing spline that would be the best fit for all of the data put together, and the interaction term smooth(x;group) is the smoothing spline representing the difference between a main effect spline and the smooth(x) spline [4]. [8]

5. RESULTS AND DISCUSSION

Evaluation of SAX representation. Experimentation with different values of nseg and alphabet size shows that, in order to capture the abstract nature of tone perception and to minimize the effect of large amount of noise in pitch movements, a limit of nseg <=3 must be placed. This is a reasonable limit considering that linguists use only two or three segments to represent tone contours in any tone language [9]. In this work, we use nseg=2 and alphabet size of 3. This choice of parameterization is evaluated as a sufficient representation for the perception of pitch contour shapes in four different ways.

First, a perceptual evaluation is carried out by having a human listener judge the shape of the contours as flat, rising, or falling (n=50). The result shows that the SAX representation achieves a 88% accuracy. Second, hierarchical clustering is performed on all contours in a given tone category. The result is then compared with the SAX labels. Figure 2 shows that in addition to meaningful groupings of SAX-labeled shapes, the clustering results also indicate that there are subgroups of shapes within the SAX-shape groups (especially SAX-flat group) that is more similar to falling or rising shapes. Third, we used SAX representation to select a subset of contours from all four tones [10], and performed a hierarchical clustering task that showed success in separating the four tones. Finally, we performed a classfication task, in which the 30-point pitch vectors from a tone category are classified into SAX class labels with a mean accuracy of 85.2%.

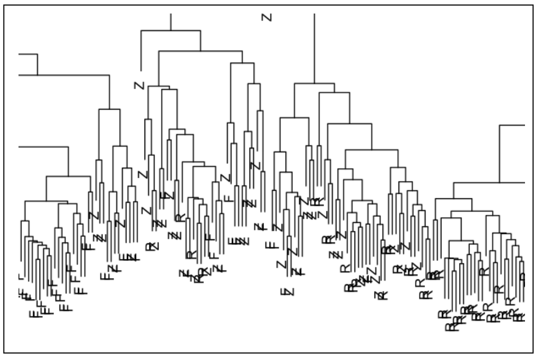

Figure 2. Hierarchical clustering on tone 4 with SAX labels (zoomed in), F=Falling, R=Rising, Z=Flat

Clustering of 4 tones (H1). Unsupervised K-means Clustering with 30-point vectors cannot learn any meaningful grouping of tone categories regardless of the number of desired clusters (performed in data mining tool Weka [7] with Euclidean distance, and in *R* with DTW distance, numOfClust varied within [4,10], otherwise default setting). Likewise, hierarchical clustering with DTW distance cannot find any meaningful groupings of tone labels at any level. This shows that we cannot find a distinct, predominant shape for a given tone category, and failure to cluster melodic contours into meaningful groups that correpond to four tones.

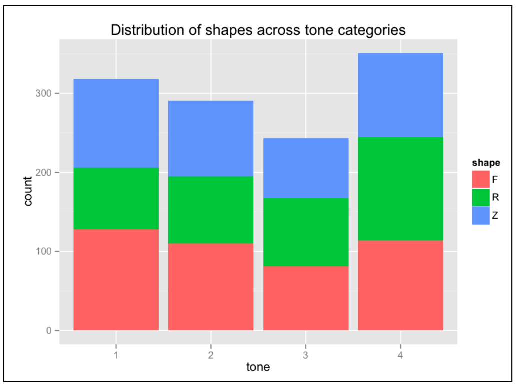

Figure 3. Distribution of shapes across five tones,F=Falling, R=Rising, Z=Flat

Exploratory within-category shape analysis (H2 and H3). First, we use the validated SAX representations to compute the distribution of three shapes rising(R), falling(F), flat(Z) within each tone category. Figure 3 shows that consistent with H2, each tone category consists of a even mixture of all shapes, with the absence of a dominant shape

[8] The SSANOVA does not return an F value. Instead, the smoothing parameters of the components smooth (x)and smooth (x); group are compared to determine their relative contributions to the equation [4]. In this paper, we use the implementation of gss package in statistical computing language *R*.

[9] In linguistics convention, high tone=H, low tone= L, rising=LH, falling=HL, falling rising=HLH, etc.

[10] Tone1-"bb", tone2-"ac", tone3-"bac", tone4-"ca", a<b<c in pitch space.

[11] . To get a more fine-grained analysis of the distributions of shapes, a two-sample test on hypothesis of population proportion is performed across tones and shapes. Results show that the proportion of rising is significantly different across four tones from the proportion of flat (χ^2 = 17.4065, df = 4, p = 0.002) or falling (χ^2 = 18.238, df = 4, p = 0.001). The proportion of flat and falling are not significantly different (p = 0.96). Furthermore, a one-sample test show that, only the proportion of rising shape is significantly different across four tones (χ^2 = 21.9853, df = 4, p = 0.0002), whereas the proportion of flats and fallings are not significanlty different across tones (p = 0.23 and p = 0.19). Finally, a two-sample pairwise test of hypothesis of population proportion shows that the proportion of rising is significantly different between tone 1 and tone 3 (χ^2 = 7.3403, df = 1, p = 0.007), tone 1 and tone 4 (χ^2 = 12.1243, df = 1, p = 0.0005), but not between tone 2, tone 3, tone 4 (except with the difference between tone 2 and tone 4 that reached significance at p = 0.04). Therefore, with the exception of tone 1 and tone 2 (p=0.22, tone 2 seem to behave more similarly to tone 1), the proportion of rising is highly significantly different between in tone 1 and other tones, whereas no strong significant differences are found among other tones. This result supports the H3 discussed above in asserting that tone 1 is mostly consisted of a mixture of flat and falling shapes (to be more specific, flat and flat-falling in H3).

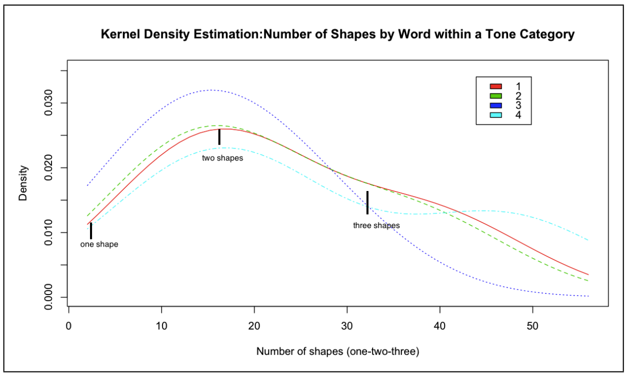

Figure 4. Kernel density estimates of number of shapes by word across four tones

Analysis of shapes by attributes. We report the analysis of attributes (artist, word, duration, position, *shengqiang, banshi* in the current context) and its correlation with a sub-cluster of shapes within a tone category. First, we performed a classification task using the shape as class label and the listed attributes. Results show that the mean accuracy is around 42%. Second, we analyze the consistency in which a word is sung as a contour shape (a word is defined as a syllable bearing a particular tone) to probe the contour similarity at the word level. Results show that among the words that appear more than once (a mean of 58% of words in our dataset), it is most likely to take on 2 different shapes at different instances, with a lower probability of taking on the same shape or even more different shapes. Figure 4 shows a kernel density estimates of the number of shapes by word in different tones. This result indicates a strong likelihood of inconsistency in singing the same word with the same tone at different contexts.

model parameter	levels (nominal)	R-squared (T1)	R-squared (T2)	R-squared (T3)	R-squared (T4)
word	468	0.178	0.0772	0.0566	0.0667
artist	15	0.0885	0.0606	0.0465	0.042
shengqiang	2	0.027	0.0235	0.0154	0.0123
position	4	0.028	0.0211	0.0189	0.0103
role type	5	0.029	0.0273	0.0242	0.018
all	na	0.032	0.028	0.0249	0.201

Table 1. SSANOVA Model comparison

SSANOVA. Results of the SSANOVA models comparison and R-suqared values (Table 1) indicate that word and artist are the best predictors of all the predictor variables (as well as all combinations of predictor variables not shown here). However, it is noticeable that the even the best model only explains less than 20% of the variance among all pitch curves in a given tone category [12] . This indicates a large amount of variation in the shape of the contours. On the other hand, the consistently larger value of R-squared for tone 1 indicates positive evidence for a more consistent shape in tone 1, as stated in the H3 discussed above.

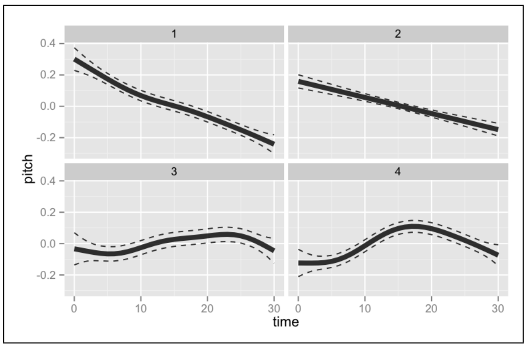

Figure 5. Average curves computed by the time+word SSANOVA model.

Average curves of four tones are computed based on this model (Figure 5), with confidence intervals shown in dashed lines. The interpretation of these average curves should be done with caution, because of the low R squared value and large standard error in the model. In particular, tone 1 and tone 2 has average contours that differ from both HG and BJ system; tone 3 and tone 4 show resemblance to BJ and HG system, respectively.

[11] Tone 5 is a neutral tone whose contour shape depends on the tone the precedes it. It exists in our dataset but is not under consideration in the current analysis

[12] And also notice that the R-squared value is highly correlated with the number of levels in the nominal attributes.

6. CONCLUSION AND FUTURE WORK

This work constitutes a preliminary step in the computational approaches to the linguistic tone-melodic contour similarity in Beijing opera singing. In this work, we focused on the single-syllable sized contours by adopting different methodologies based on competing hypothesis of tone systems. We have demonstrated the effectiveness of SAX-based representations in tasks of shape analysis and time-series mining. The results indicate a even mixture of shapes within each tone category, with the absence of a dominant tone system in Beijing opera. In addition, we found evidence supporting the hypothesis that tone 1 is sung with more consistent shape than other tones. Overall, our results point to low degree of similarity in single-syllable pitch contours. Given the discussion and methodology proposed here, we expect future research on pairwise syllable contour similarity analysis to yield more promising results.

7. ACKNOWLEDGEMENT

This research was funded by the European Research Council under the European Union Seventh Framework Program, as part of the CompMusic project (ERC grant agreement 267583).

8. REFERENCES

[1] Bogdanov, D., Wack N., Gmez E., Gulati S., Herrera P., Mayor O., et al.: ESSENTIA: an Audio Analysis Library for Music Information Retrieval. International Society for Music Information Retrieval Conference (ISMIR'13). 493-498.(2013).

[2] Boersma, Paul.: Praat, a system for doing phonetics by computer. Glot International 5:9/10, 341-345.2001.

[3] Chen,K.: Characterization of Pitch Intonation of Beijing Opera. Master Thesis, Music Technology Group, Universitat Pompeu Fabra, Barcelona, 2013.

[4] Davidson, L.: Comparing tongue shapes from ultrasound imaging using smoothing spline analysis of variance. Journal of the Acoustic Society of America, 120(1).2006.

[5] Gautheir,B., Shi,R, Xu,Y.: Learning phonetic categories by tracking movements. Cognition, 103, (2007),80-106.

[6] Gu,C.: Smoothing Spline ANOVA Models. New York: Springer- Verlag.2002.

[7] Mark Hall, Eibe Frank, Geoffrey Holmes, Bernhard Pfahringer, Peter Reutemann, Ian H. Witten: The WEKA Data Mining Software: An Update; SIGKDD Explorations, Volume 11, Issue 1.2009.

[8] Lin,J., Keogh,E., Wei,L.,and Lonardi,S.: Experiencing SAX: a novel symbolic representation of time series. Data Mining and Knowledge Discovery. Oct.2007, Vol.15, Issue.2, pp107-144.2007.

[9] Pian,R.C.: Text Setting with the Shipyi Animated Aria. In Words and Music: The Scholar's View,edited by Laurence Berman, 237-270. Cambridge: Harvard University Press,1972.

[10] Patel,A.: *Music, Language, and the Brain*, Oxford Press, 2008.

[11] Patel, A. D., Xu, Y. and Wang, B.: The role of F0 variation in the intelligibility of Mandarin sentences. In Proceedings of Speech Prosody 2010, Chicago.(2010).

[12] Prom-on, S., Xu, Y. and Thipakorn, B.: Modeling tone and intonation in Mandarin and English as a process of target approximation. Journal of the Acoustical Society of America 125: 405-424.(2009).

[13] Salamon, J and Gomez E: "Melody Extraction from Polyphonic Music Signals using Pitch Contour Characteristics", IEEE Transactions on Audio, Speech and Language Processing, 20(6):1759-1770.2012.

[14] Serra,X.: "Creating Research Corpora for the Computational Study of Music: the case of the CompMusic Project", AES 53rd International Conference, January 27-29th 2014, London (UK).

[15] Stock, J: A Reassessment of the Relationship Between Text, Speech Tone, Melody, and Aria Structure in Beijing Opera. Journal of Musicological Research (18:3): 183-206. 1999.

[16] Valero,J: Measuring similarity of automatically extracted melodic pitch contours for audio-based query by humming of polyphonic music collections. Master's Thesis, MTG, DTIC, UPF, 2013.

[17] Wichmann, E.: Listening to the theatre: the aural dimension of Beijing opera. Honolulu: University of Hawaii Press. 1991.

[18] Xu, Y. and Sun X.: Maximum speed of pitch change and how it may relate to speech. Journal of the Acoustical Society of America 111: 1399-1413.2002.

[19] Xu,Z. 2007: Jiu shi nian lai jingju de sheng diao yan jiu zhi hui gu. (Overview of Ninety Years of Research of Tone Relations in Peking Opera). Nankai Journal of Linguistics, 10(2):39-50.

[20] Yu,H.: Qiang ci guan xi yan jiu. (Study on the relationship between tone and melody). Beijing: Central Conservatory Publishing House.2009.

[21] Lai, Yuwen and Jie Zhang. : Mandarin lexical tone recognition: the gating paradigm. In Emily Tummons and Stephanie Lux (eds.), Proceedings of the 2007 Mid-America Linguistics Conference, Kansas Working Papers in Linguistics 30. 183-194.2008.

[22] Zhu, Weiying: Xiqu Zuoqu Jifa.(The Composition Techniques for Chinese Operas.). Beijing: Renmin Yinyue Chubanshe.2004.