Late-Breaking / Demos Session Abstract for the
2018 International Society for Music Information Retrieval Society Conference

TOWARDS HMM-BASED GLISSANDO DETECTION FOR RECORDINGS OF CHINESE BAMBOO FLUTE

Changhong Wang[1], Emmanouil Benetos[1], Xiaojie Meng[2], Elaine Chew[1]
[1] Centre for Digital Music, Queen Mary University of London, UK
[2] Department of Chinese Music, China Conservatory of Music, China
{changhong.wang,emmanouil.benetos,elaine.chew}@qmul.ac.uk; mxj.yd@foxmail.com

ABSTRACT

Playing techniques such as ornamentations and articulation effects constitute important aspects of music performance. However, their computational analysis is still under-explored due to a lack of data and established methods. Focusing on the Chinese bamboo flute, we introduce a two-stage glissando detection system based on hidden Markov models (HMMs) with Gaussian mixtures. A rule-based segmentation process extracts glissando candidates that are consecutive note changes in the same direction. Glissandi are then identified by two HMMs (glissando and non-glissando). The study uses a newly created dataset of Chinese bamboo flute recordings. The results, based on both frame- and segment-based evaluation, achieve F-measures of 78% and 73% for ascending glissandi, and 65% and 72% for descending glissandi, respectively. The dataset and method can be used for performance analysis.

1. INTRODUCTION

Computational analysis of expressive music performance has attracted increased attention in the music information retrieval community over the past decade. For example, [1] introduced a method for automatic transcription of note embellishments in performance, [6] proposed mathematical models for automatic extraction and characterisation of vibrato effects and portamentos (note transitions) in string and voice performances, [4] introduced ways to detect legato and glissando in guitar performance.

The Chinese bamboo flute (thereafter referred to as CBF), also known as the *Dizi* or *Zhudi*, is one of the world's most ancient instruments, with a large repertoire of playing techniques. The acts of breathing, fingering, and tonguing, in interaction with a thin membrane glued to the CBF's resonance hole, create unique sounds and subtle sound nuances in performance, providing a rich platform for analysing playing techniques. The glissando figures prominently amongst the CBF playing techniques, and has a readily recognisable pattern; thus serves as our starting point for building a systematic methodology for the detection of CBF playing techniques.

EB is supported by a UK RAEng Research Fellowship (RF/128).

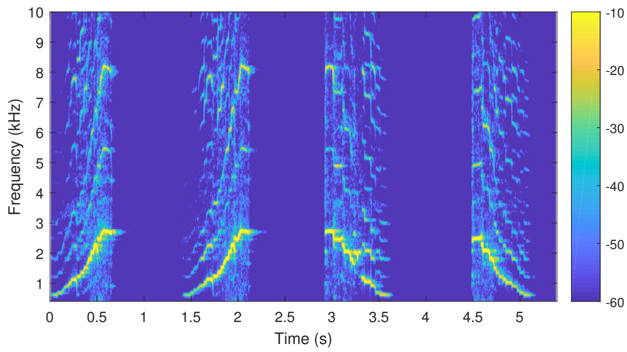

Figure 1. Spectrogram of ascending and descending glissando examples in Chinese bamboo flute playing.

Fig. 1 shows a spectrogram of a series of two ascending and two descending CBF glissandi. As can be seen, they resemble rapid scale segments. Glissando detection in CBF playing is not straightforward. CBF glissandi are less regular than the stair-like glissando patterns in piano and guitar playing [2]. For the same glissando type, variations exist in the ways they are executed between different players, different pieces, and even different parts in the same piece.

Hidden Markov models (HMMs) [5] are particularly well suited to capturing sequential information within time series. They enable decoding of consecutive note evolution while smoothing outlier variations in performed glissandi. In this paper, we propose and test a system combining rule-based segmentation and HMMs to detect glissandi in real-world monophonic CBF recordings. The system is trained, validated, and tested on a specially created CBF dataset.

2. DATASET

The glissando analysis dataset, CBF-GlissDB, part of the CBF dataset, comprises of recordings by ten expert CBF players from the China Conservatory of Music. All data was recorded in a professional recording studio using a Zoom H6 recorder at 44.1kHz/24bits. Each player performed both isolated glissandi covering all notes on the CBF and one full-length piece—*Busy Delivering Harvest* ≪扬鞭催马运粮忙≫ or *Morning* ≪早晨≫.

The fundamental frequency of each recording was estimated using the pYIN algorithm [3], all errors manually corrected by the first author using Sonic Visualiser [1], and glissandi annotated and verified by the CBF players.

3. METHOD

Fig. 2 shows the two-stage detection system, including rule-based segmentation and HMM-based identification.

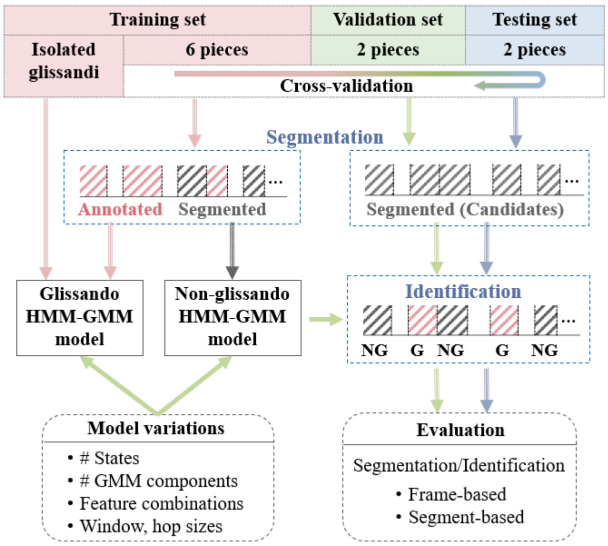

Figure 2. System diagram for glissando detection method.

3.1 Rule-based segmentation

To obtain potential glissando segments from the recordings, the pitch ground truth is first quantized to the nearest notes in twelve-tone equal temperament scale—16 notes in the CBF tonal range : G4-A6 for the C flute, and D5-E7 for the G flute (the only two flute types featuring in our dataset). Consecutive note changes in the same direction are then extracted as glissando candidates.

The CBF-GlissDB is subdivided into three subsets, namely, training (all isolated glissandi and 6 whole-piece recordings), validation (2 whole-piece recordings), and testing (2 whole-piece recordings). The segmentation stage is applied to all three subsets, but to different ends. For the training set, segmentation serves the purpose of obtaining false positives that are then used to train a non-glissando HMM. In validation and testing stages, the extracted segments serve as candidates to be assigned glissando (G) or non-glissando (NG) labels.

3.2 HMM-based identification

Since all glissando candidates (extracted in the previous stage) share similar pitch evolution characteristics, the input to the HMMs must possess sufficient discriminative power to distinguish glissandi from non-glissandi. We use a feature set consisting of both short-term (pitch change, intensity, intensity change) and long-term (note number, note duration, note range) features.

1. https://www.sonicvisualiser.org

Two HMMs (G and NG) with Gaussian mixture emissions are trained using the training dataset and calibrated using the validation dataset. The HMMs parameters are initialised using k-means and updated via an Expectation-Maximization algorithm [5]. During training, model parameters—the number of states, number of Gaussian mixtures, feature combinations, and window-hop sizes—are varied and the best performing model selected and applied to the testing data.

4. RESULTS & DISCUSSION

Because glissandi range approximately from 200 to 1100ms, both frame-based and segment-based evaluations are provided. Segments must overlap by more than 100ms with the ground truth to be considered correct.

The ten whole-piece recordings are randomly allocated to the training, validation, and testing subsets in a 6 : 2 : 2 ratio ; 5-fold cross-validation is then conducted. Table 1 gives the precision, recall, and F-measure figures for both ascending and descending glissandi.

Stage	Glissando direction	Frame-based(%)			Segment-based(%)		
		P	R	F	P	R	F
Rule-based segmentation	Ascending	7.0	93.2	12.6	8.2	98.4	15.2
	Descending	3.2	82.4	5.6	6.2	100.0	11.2
HMM-based identification	Ascending	81.2	77.8	78.0	73.4	79.4	73.2
	Descending	70.0	62.4	64.6	70.2	77.2	72.2

Table 1. Two-stage glissando detection system evaluation results (P=precision, R=recall, F=F-measure).

Note that apart from high recall in the segmentation stage, low precision which indicates a large number of false positives can be obtained for the NG-HMMs training, benefits the data balance in our system. Better identification performance of ascending than descending glissandi due to their more regular patterns. Future work will seek to compare these results with a fully automatic glissando detection system and consider other CBF playing techniques.

5. REFERENCES

[1] E. Benetos, S. Dixon, D. Giannoulis, H. Kirchhoff, and A. Klapuri. Automatic music transcription : challenges and future directions. *Journal of Intelligent Information Systems*, 41(3) :407–434, 2013.

[2] Y. P. Chen, L. Su, Y. H. Yang, et al. Electric guitar playing technique detection in real-world recordings based on F0 sequence pattern recognition. In *ISMIR*, pages 708–714, 2015.

[3] M. Mauch and S. Dixon. pYIN : A fundamental frequency estimator using probabilistic threshold distributions. In *ICASSP*, pages 659–663, 2014.

[4] T. H. Ozaslan and J. L. Arcos. Legato and glissando identification in classical guitar. In *7th Sound and Music Computing Conference*, pages 457–463, 2010.

[5] L. R. Rabiner. A tutorial on hidden Markov models and selected applications in speech recognition. *Proceedings of the IEEE*, 77(2) :257–286, 1989.

[6] L. Yang. *Computational Modelling and Analysis of Vibrato and Portamento in Expressive Music Performance*. PhD thesis, Queen Mary U of London, 2016.