Computational Analysis of Gamelan Gong Kebyar Tuning for Geographic Classification

Lluís Solsona, Alexandre Vilanova, Stefano Scola
Universitat Pompeu Fabra

March 16, 2024

Abstract

Pitch detection algorithms have not been extensively explored for non-eurogenetic musical traditions, despite the rich theoretical foundation provided by ethnomusicology. In an effort to address this gap, our research takes a practical approach by examining contemporary pitch detection methods and applying them to a case study. We propose a pitch detection tool for gamelan gong kebyar, a traditional Balinese ensemble. By focusing on this particular case study, we hope to provide insights that can be applied to develop tools for other music traditions.

1 Introduction

In Western music, the pitch organization is mostly based on the 12 tone equal temperament system. Unlike Western musical traditions, there is no standardized temperament within the Balinese gamelan, a tradition consisting in an orchestra of bronze metallophones and gongs. Instead, tuning is left to the discretion of skilled craftsmen known as pande. This results in a rich variety of tunings in different ensembles. Moreover, while Western music typically uses A4 (440 Hz) as the reference tone to establish a standard tuning, in the gamelan musical tradition, there is no standard tuning, and no specific reference tone is utilized. Given these differences, it can be concluded that tools for estimating tuning, which are implemented and based on Western music, may not be directly applicable to gamelan music. In the following sections, an attempt to estimate tuning from gamelan music is presented.

Tuning within the gamelan tradition is a highly developed practice, with different styles emerging based on the preferences of individual pande. There are two established reference systems or scales. These include the seven-tone pelog and the five-tone slendro, which represent aproximate interval relationships rather than absolute frequencies, and their interpretation during the tuning process contributes to the geographical variation observed in tuning styles.

The pelog scale gives rise to modes, which are typically composed of five selected tones. Each type of gamelan uses a specific scale and mode. Gong kebyar is tuned to a 5-tone mode derived from the pelog, known as selisir. Due to the complex nature of gamelan tuning, this study focuses exclusively on examples of gamelan gong kebyar, a widespread ensemble on the island of Bali since its birth in the early 20th century.

2 Objectives of the research

- Review the state of the art in musical pitch detection algorithms in order to realize an implementation that determines the tuning given a single metallophone recording, from a gong kebyar ensemble.
- Review the state of the art in ethnomusicological research in order to obtain meaningful categories to characterize the tuning of a given recording of a gamelan gong kebyar.

3 Methodology

This section describes our methodological approach, which will be further discussed in the next section.

3.1 Preprocessing

In order to efficiently analyze the harmonic content of the input sound, our approach includes a preprocessing step that decomposes the original signal. For this purpose, we evaluate the effectiveness of harmonic percussive source separation (HPSS) [Driedger et al., 2014].

Initially, the spectrogram is computed using the STFT implmentation of librosa [McFee et al., 2015]. Subsequently, HPSS is applied on the magnitude spectrogram.

The concept behind HPSS involves processing the spectrogram both horizontally and vertically. This approach aims to capture harmonic and percussive aspects separately. The resulting spectrograms are then used to construct a masking, which is applied to the original spectrogram to separate harmonic and percussive components, thus yielding two distinct spectrogram representations.

The librosa implementation of HPSS uses as default a kernel size of 31, representing the size of the horizontal and vertical windows used to process the input spectrogram. The values within these windows are processed via median-filtering, wherein the mean of the selected segment is computed. By doing so, the algorithm distinguishes percussive events as they are not prolonged in time, therefore happening "vertically", from harmonic events, that are distributed "horizontally". Furthermore, the margin parameter (also referred to as the separation factor in the original publication) influences the generation of the two masking components. Essentially, when applying the masking, this parameter determines whether a certain frequency bin should belong to the percussive or harmonic component.

3.2 Pitch Extraction and Segmentation

Once we have separated the harmonic components from the percussive, we proceeded to extract the fundamental frequency present in the audio. Pitch is estimated in the frequency domain by applying parabolic interpolation. The stable regions of the estimated fundamental frequency are selected using a moving window, where the standard deviation is checked against a given threshold. A visualization of this step can be found in Figure 1a.

3.3 Tone Identification

Once a list of pitch classes is obtained, we use properties of intervalic structure to infer the ding. Since in a gamelan tuning process, the pande usually uses the pemade – one of the higher pitched metallophones – as a reference point only pemade recordings are considered in the first implementation fo this algorithm.

For a recording to be valid, it must also contain the top six notes of the register of the instrument. From there, an algorithm looks for three notes in a row whose pairwise distances do not exceed 220 cents, and we label the lowest of these as ding. The following ones are labeled with the successive notes.

The method is intended to be effective regardless of the notes featuring the recording, as long as the above condition is met. It also works if the middle note and at least the top two notes are present. If more notes of the higher register appear, there should be no missing notes in between.

3.4 Tuning Characterization and Regency Classification

Given the non-existence of an absolute frequency value for notes, we base our characterization of a pitch on its similarity to different reference intervallic patterns.

Once extracted and labeled the set of tones present in a audio recording, is to make sense of the resulting frequencies. On one hand we aim to give insights on the kind of tuning, on the other hand we aim to predict the geographical origin of the gamelan.

In order to accomplish the first task, we take into account specific domain knowledge on gamelan tuning generalization. Pande and musicians use specific terms to characterize gamelan tunings. These profiles are called begbeg, tirus and sedang – being the last a middle point between the former two, as illustrated in figure 1. Each of these correspond to an approximate interval profile [Vitale and Sethares, 2021]. Similarity is calculated integrating distances into a vector and comparing it with the given profiles computing their euclidean distance.

In regards to the prediction the geographical origin of the gamelan, Andrew Toth's extensive study documenting the frequencies of the keys and gongs of 49 gong kebyar ensembles from seven regencies serves as a valuable resource. We calculate the similarity again, but this time using the averages of the tunings of the 49 gamelans classified by regency.

4 Methodological Discussion

An examination of the frequency of the third octave ding, the reference pitch, shows a wide distribution. While there is an interesting concentration between 540 and 545 Hz, indicating a trend, this is limited to a total of 9 gamelan instruments spread over 4 regions. Looking at the full sample of gamelans, the ding takes values in a range of 400 cents (major third), with at least one value every 20 cents. This already indicates that using the 12 equal temperament model is not sufficient to study tuning, nor will chroma representation.

Let us observe and discuss the results of the HPSS. First of all, although the classification between percussive and harmonic is not trivial, given the hybrid nature of certain instruments, HPSS has proven to be an effective tool for the objective of this research. Working with monophonic recordings of idiophones concentrates the vertical component in the attacks. The key, deliberately forged to mute the harmonics as much as possible, results in a spectrum clean enough to be efficiently separated into harmonic and percussive components when the Margin parameter is set to 16. This behavior was observed consistently, even with the data constraints.

Additionally, for a comparative analysis, we have employed CREPE [Kim et al., 2018], a convolutional-based deep neural network which operates on the waveform. The latter has been reconsructed from the harmonic and percussive spectrograms using the Griffin and Lim algorithm [Perraudin et al., 2013]. Onsets were then computed from the reconstructed percussive audio to segment the pitch estimations. In this particular scenario, it has been observed that a traditional signal processing technique outperforms the deep learning model. The pitch contour estimated by CREPE exhibits more variation compared to piptrack. One possible reason for the difference in results is that these two algorithms operate in different domains: CREPE on the waveform domain, while piptrack operates on the frequency domain. The input sound to CREPE consisted of a reconstruction of the harmonic spectrogram. Although the reconstruction is perceptually acceptable, it might introduce some artifacts in the waveform domain, causing CREPE to perform poorly. In Figure 2 we can see a comparison of the two algorithms.

Finally, let us compare the similarity metrics used to compare tunings: Euclidean distance and cosine similarity. The Euclidean distance is computed as the square root of the sum of squared element-wise differences between two vectors. It measures the actual distance in space and tends to be less effective in high-dimensional spaces as it is sensitive to the magnitude of the vectors. On the other hand, cosine similarity is computed as a normalized dot product. It is more suitable in higher-dimensional spaces as it measures the difference in the direction of two vectors, making it less sensitive to the magnitude of the vectors. Overall, it shall be noted that the entries in the 3 tuning vectors do not largely differ, meaning that they are close to each other. The latter explains why cosine similarity yields similar results for all the 3 vectors. In fact, since the space is not highly dimensional, and given the need of accurately analyze the intervals, euclidean distance is thought to be more suitable and it provides an element-wise comparison.

5 Conclusion

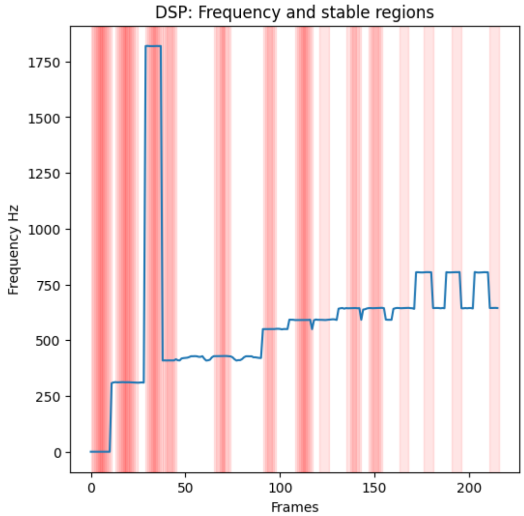
The proposed implementation presents a starting point for automatic tuning recognition and classification of gamelan ensembles that can be developed to adapt to any type of Javanese or Balinese ensemble.

Although the results are promising, and given the limited size of our sample, more data needs to be collected to properly asses the performance of the tool. The steps to be developed in the following include the retrieval of pemade recordings with the required metadata.

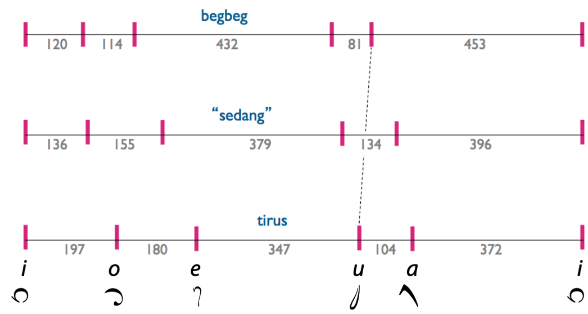
The gamelan usually consists of bamboo flutes and chordophones such as the rebab or kacapi. Tuning instruments of this type, non-idiophones, requires an approach to various points in the pipeline, such as the tool used to extract the melody.

The gamelan gong kebyar and other Balinese ensembles produce beating when the full ensemble performs together. This limits the applicability of our algorithm, as it must be applied to instruments alone, in addition to other limitations such as the restriction to the third octave.

A Appendix



(a) Stable Regions of Fundamental Frequency



(b) Interval profiles, taken from [Vitale and Sethares, 2021]

Figure 1

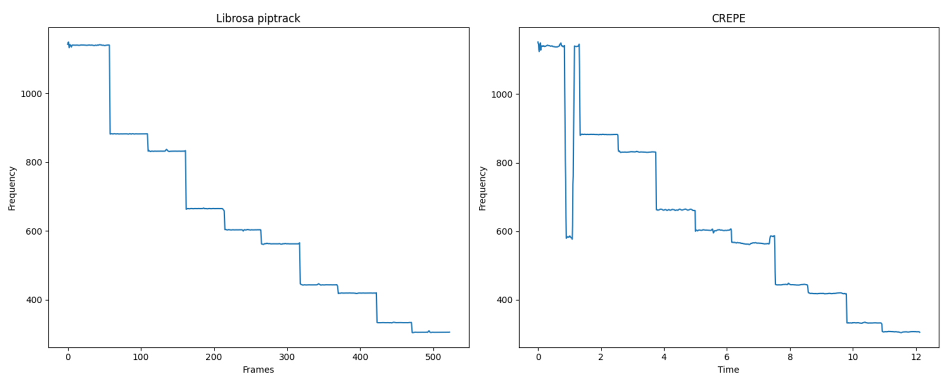


Figure 2: Pitch extraction, comparing piptrack with CREPE

References

[Driedger et al., 2014] Driedger, J., Müller, M., and Disch, S. (2014). Extending harmonic-percussive separation of audio signals. In *ISMIR*, pages 611–616.

[Kim et al., 2018] Kim, J. W., Salamon, J., Li, P., and Bello, J. P. (2018). Crepe: A convolutional representation for pitch estimation. In *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 161–165. IEEE.

[McFee et al., 2015] McFee, B., Raffel, C., Liang, D., Ellis, D. P., McVicar, M., Battenberg, E., and Nieto, O. (2015). librosa: Audio and music signal analysis in python. In *SciPy*, pages 18–24.

[Perraudin et al., 2013] Perraudin, N., Balazs, P., and Søndergaard, P. L. (2013). A fast griffin-lim algorithm. In *2013 IEEE workshop on applications of signal processing to audio and acoustics*, pages 1–4. IEEE.

[Vitale and Sethares, 2021] Vitale, W. and Sethares, W. (2021). Balinese gamelan tuning: The toth archives. *Analytical Approaches to World Music*, 9(2).