SEMITONE SHIFT DETECTION IN MUSIC AUDIO

Kartik Gupta
IIIT Hyderabad and
City University London
kartik.gupta
@research.iiit.ac.in

Daniel Wolff
Universitat Pompeu Fabra and
City University London
daniel.wolff.1
@city.ac.uk

Stephen Cottrell **Tillman Weyde**
City University London
{t.e.weyde,
stephen.cottrell.1}
@city.ac.uk

ABSTRACT

We present two datasets and an initial algorithm for the classification of semitone shifts. The semitone shift is a pattern in harmonic progressions, which has appeared relatively recently in Western tonal music. It is typically characterised by a change of key by one or two, sometimes more, semitones upwards and the repetition of a section of the music before and after the shift, which typically takes place towards the end of the piece. It is of musicological interest to recognise these key shifts automatically from audio recordings. The ultimate goal is to analyse large data collection to understand historically when this phenomenon started and how it spread and developed.

The two datasets comprise popular music with annotations on whether and where the audio recordings contain a semitone shift. The initial algorithm presented is based on self similarity and dynamic time warping (DTW) for the detection of semitone shifts, and it shows encouraging performance on the presented datasets.

1. INTRODUCTION

With increasing availability of music data in recent years there is scope for automating the analysis of musical data. We present here work towards a detector for a specific harmonic feature, the *semitone shift*.

The semitone shift is a transposition from one tonal area to another, but while the modulation from one key to another is an important characteristic of Western functional harmony, the upward transposition by one or two semitones of entire passages of music en bloc is a more recent phenomenon. From informal observation it seems to occur mostly in popular music, where the repetition of a complete section–for example the chorus–one or two steps higher than its preceding iteration is used as a means of generating tension and variety within the music. This is frequently encountered towards the end of piece. Well know examples are *Easy* by *The Commodores* (in the album version), *Livin' on a Prayer* by *Bon Jovi*, and *I Want It That Way* by the *Backstreet Boys*.

Because this technique is not rooted in classical music practice, it is reasonable to ask how and when it came about. It may be allied to the development of popular musical styles in the USA in the early 20th century, but we have yet no conclusive evidence of this. The technique is also widely encountered in pop music of the 1980s and 90s (for example in many of the hits produced by the songwriting/production team Stock, Aitken and Waterman). Quite why this phenomenon should be so common at this time is so far unclear.

Tracking this musical development manually, through the analysis of individual pieces, is a laborious task. But it is a task that is well suited to MIR particularly when applied to large corpora of music, i.e. using Big Music Data [6]. Once a robust detector is developed, we plan to track semitone shifts over a century or so of recorded music, and we hope to be able to show how this compositional strategy developed within Western music and to find clues how it relates to broader social and technological trends.

To our knowledge no specific detection algorithm for this task has been developed so far. There is previous work by Chai [1] aiming at detecting key changes in piano music. There is a chorus detection method by Goto [3] that detects the chorus, allowing for shifted repetitions. This paper proposes an approach based on the DTW and a modified version of the self similarity matrix proposed by Foote [2] which is a common tool for structure analysis [4]. This approach relates to our definition of the semitone shift as consisting of the shifted repetition of a segment.

2. TWO DATASETS WITH SEMITONE SHIFT ANNOTATIONS

Since there exists no openly available annotated datasets for this task yet, we created two datasets to train and evaluate algorithms, which we made accessible online [1]. In the first dataset, the audio recordings underlie copyright restrictions and hence only the annotations can be made public. The second dataset is based on music that is available through Creative-Commons licensing, allowing the audio recordings to be distributed with the annotations.

[1] http://mirg.city.ac.uk/datasets/ismir2016st

2.1 Dataset 1: Copyright-Restricted Music

This dataset consists of copyright protected music by the *Backstreet Boys* and *Michael Jackson*. Semitone shifts for this music were annotated manually and independently validated. The datasets are not balanced by the number of songs containing a semitone shift, but rather reflect a typical sample from the group's output.

The Backstreet Boys collection consists of 107 songs out of which 34 were annotated with a semitone shift. In the Michael Jackson collection, there were 92 songs in total out of which 21 were marked to contain a semitone shift. For each song, a boolean of whether a shift occurred as well as the time of the first and, if applicable, the second shift were recorded. The file furthermore contains the estimated tonic(s) of the source before the transposition(s). We provide links to YouTube [2] where the songs can be accessed. Moreover, identifiers from AcousticBrainz [3] [5] will be made available soon.

2.2 Dataset 2: Creative Commons Music

The creative commons dataset consists of songs downloaded from the Free Music Archive [4] which had Creative Commons – Attribution – Share Alike license. To this end, our dataset contains the details of each artist through links to their profile. In most of the songs downloaded there were no semitone shifts: In a total of 87 songs we only found 2 songs containing a semitone shift. We artificially increased the number of semitone shifts in this dataset by manually adding shifts to 48 songs we found suitable: First, a position for the region to be shifted was manually selected. We then inserted semitone-shifted versions of the segment using the Rubberband [5] library for transposition.

3. A SUBSEQUENCE-BASED CLASSIFIER

Our rule-based classifier uses chroma-similarity for detection of structural transposed recurrences. Beat-aligned chroma features were calculated for each song using the chromagram plugin [6] and the librosa [7] library respectively, resulting in the $allChroma$ features. The analysed song is then divided into overlapping segments of a predefined length $segLen$.

For each segment we calculate two cost vectors, *cShift* and *cOrig*. At index t, these contain the minimum subsequence-DTW costs of the original and semitone-transposed segment being matched to the $allChroma$ sequence of the entire song at position t. The cost vectors are subtracted to detect where the semitone transposed version of the segment is more similar (lower costs) than the original segment, and the result is concatenated to a differential cost matrix costMat. This is converted into a 2-D binary array, $betterTransposed$, in which then the largest rectangle is detected. The position of the rectangle is our estimate for the longest semitone-shift segment in the song.

```
Input: allChroma, segLen, hopSiz , finPart,
       minMatch
for segments(allChroma, segLen, hopSiz) do
    cOrig = DTW(segChroma, allChroma)
    cShift = DTW(transp(segChroma,semitone),
             allChroma)
    costMat.append(cShift - cOrig)
end
betterTransposed = binaryOpening(costMat < 0)
matchRegion = largestRectangle(betterTransposed)
if matchRegion > minMatch
  ∧ matchRegion after finPart then
    return True
else
    return False
end
```

Algorithm 1: A rule-based semitone shift classifier

3.1 Parameter Learning using Swarm Optimisation

We determined optimal parameter values using swarm optimization, in particular the implementation of the *pyswarm* Python toolbox [8]. We optimised the following parameters:

- $segLen$: The length of the segment to be considered for finding repetitions.
- $hopSiz$: The hop size between two segments.
- $finPart$: The minimum percentage a song must have elapsed before the shifted repetition.
- $minMatch$: The minimum length of the repeated part in the shifted semitone. This is important as sometimes very small parts in a different chord might seem like they are in the different tonic but the song has not shifted.

The range for these parameters were 2-32 beats, 1-32 beats, 25-90 and 8-64 beats. The optimal values found for these parameters are 20 beats, 8 beats, 30 and 32 beats, respectively, resulting in the following performances:
Accuracy : 0.88, Recall : 0.91, Precision : 0.77.

4. CONCLUSIONS

We have introduced the MIR task of detecting semitone shifts, presented two datasets for training and testing semitone shift detectors, as well as a rule-based classifier for semitone shift detection. The question of the history of the semitone shift is a good example of a musicological question that can be answered by MIR technology. The initial results are encouraging and we believe a method reliable enough for large scale application is within reach. In the future, we will improve the method and begin musicological analysis on large-scale datasets.

[2] http://www.youtube.com
[3] http://acousticbrainz.org
[4] http://www.freemusicarchive.org
[5] http://breakfastquay.com/rubberband/
[6] http://vamp-plugins.org/plugin-doc/qm-vamp-plugins.html#qm-chromagram
[7] https://github.com/librosa/librosa
[8] http://pythonhosted.org/pyswarm/

5. REFERENCES

[1] Wei Chai and Barry Vercoe. Detection of key change in classical piano music. In *ISMIR 2005, 6th International Conference on Music Information Retrieval, London, UK, 11-15 September 2005, Proceedings*, pages 468–473, 2005.

[2] Jonathan Foote and Matthew L. Cooper. Visualizing musical structure and rhythm via self-similarity. In *Proceedings of the 2001 International Computer Music Conference, ICMC 2001, Havana, Cuba, September 17-22, 2001*, 2001.

[3] Masataka Goto. A chorus section detection method for musical audio signals and its application to a music listening station. *IEEE TASPL*, 14:1783–1794, 2006.

[4] Jouni Paulus, Meinard Müller, and Anssi Klapuri. State of the art report: Audio-based music structure analysis. In *Proceedings of the 11th International Society for Music Information Retrieval Conference*, pages 625–636, Utrecht, The Netherlands, August 9-13 2010.

[5] Alastair Porter, Dmitry Bogdanov, Robert Kaye, Roman Tsukanov, and Xavier Serra. Acousticbrainz: a community platform for gathering music information obtained from audio. In *16th International Society for Music Information Retrieval Conference (ISMIR 2015)*, pages 786–792, Malaga, Spain, 26/10/2015 2015.

[6] Tillman Weyde, Stephen Cottrell, Jason Dykes, Emmanouil Benetos, Daniel Wolff, et al. Big data for musicology. In *Proceedings of the 1st International Workshop on Digital Libraries for Musicology*, pages 1–3. ACM, 2014.