Genre-specific Key Profiles

Cian O’Brien  Alexander Lerch
Center for Music Technology
Georgia Institute of Technology
{cobrien30, alexander.lerch}@gatech.edu

ABSTRACT
The most common approaches to the automatic recognition of musical key are template-based, i.e., an extracted pitch chroma vector is compared to a template key profile in order to identify the most similar key. General as well as domain-specific templates have been used in the past, but to the authors' best knowledge there has been no study that evaluated genre-specific key profiles extracted from the audio signal. We investigate the pitch chroma distributions for 9 different genres, their distances, and the degree to which these genres can be identified using these distributions when utilizing different strategies for achieving key-invariance.

1. INTRODUCTION
The pitch chroma is a compact and robust representation of the tonal content of an audio signal. Automatic key detection systems commonly use the average pitch chroma of a music file in order to detect the musical key by comparing the extracted pitch chroma to a template key profile. In the literature, different strategies for deriving these templates have been proposed, such as based on human tonality perception [1], using diatonic models [2], extraction from MIDI data [3], and extraction from audio data [4]. Here we analyze the distributions of (pitch chroma based) key profiles extracted from different musical genres. The similarity of genre-specific key profiles is measured directly by computing inter-genre distances in Sect. 4 and indirectly by applying an SVM classifier for testing genre separability through key profiles (Sect. 5). The goal of this work is to investigate (i) how pitch chroma are distributed within each genre and (ii) the extent to which musical genres can be distinguished using only the tonal information contained in their pitch chroma profiles.

2. DATA SET
The data set used was the GTZAN collection.1 While this set is old and has obvious disadvantages [5], it is a well-known, widely-used, and easily available set for genre classification tasks. It consists of 1000 song excerpts divided into ten genres: Blues (B), Classical (Cl), Disco (D), Reggae (Rg), Pop (P), Metal (M), Rock (R), Jazz (J), Country (C) and Hip Hop (H). Key annotations for the tracks are publicly available.2 Tracks for which the key could not be unambiguously identified were excluded. The number of annotated files therefore reflects the number of unambiguously identifiable keys. For example, none of the excerpts from the Classical genre are annotated.

Figure 1 gives a detailed visualization of the modes (top) and key distribution (bottom) per genre. The tonics are sorted with respect to the circle of fifths (major modes are indicated in upper case letters and minor modes in lower case). The relation of major vs. minor modes is very skewed for blues and metal (predominantly minor) as well as country (predominantly major); the genres disco, pop, reggae, and rock have a more balanced distribution between modes. Jazz tracks tend to be clustered around flat keys which are favored by trumpet and saxophone players. The keys for country cluster around C-Maj with a tendency to sharp keys. The majority of metal tracks are in either a minor or e minor, keys well-suited to the electric guitar and bass (corresponding to the two lowest open strings).

3. FEATURE EXTRACTION
The pitch chroma is a commonly used feature in the field of MIR because it is a compact, robust, and mostly timbre-independent representation of the pitch content [6]. It is a 12-dimensional histogram-like octave-independent vector showing the “strength” of the 12 semitone classes (C, C#, D, ..., B) and is usually computed by converting the spectrum to semi-tone bands and summing the energy of all bands with the distance of an octave [7]. Here, the pitch chroma is extracted at a sample rate of 10kHz over a range of three octaves, starting from C at 130.8 Hz. The FFT block size is 8192, the hop size is 4096. The overall pitch chroma per file is a single 12-dimensional vector that is computed by taking the median of all pitch chromas per block.

The term key profile is used for the overall, tonic independent pitch chroma per file. Our hypothesis assumes that the key profiles of songs within one genre that have the same mode (major or minor) should be similar, but shifted circularly to the songs’ tonic. Under this assumption, each overall pitch chroma

---

1 http://marsyas.info/downloads/datasets.html

2 https://github.com/alexanderlerch/data_set
Temperley’s key profiles (derived from listening experiments on tonality), they correlate
with key profiles (compare [2]). Temperley’s key profiles.

The key profiles for minor have, compared to the major
distributions and number of annotated files (top) and key distributions
per genre (bottom).

can be “converted” to a key-profile by applying a circular shift.
In other words, the key profile is the tonic independent pitch
distribution (e.g., the pitch chroma of a song in A-Maj or a-
min is circularly shifted by 9 indices to the left so that the bin
of pitch class A lands on the first index).

4. KEY PROFILE ANALYSIS

Figure 2 shows the overall key profiles in a box plot in com-
parison with known profiles from the literature. While Krum-
hausl’s “Probe Tone Ratings” [1] are not exactly a key profile
(derived from listening experiments on tonality), they correlate
well with key profiles (compare [2]). Temperley’s key profiles
are extracted from symbolic data rather than from audio [3, 8].

The key profiles of the six most populated genres are plotted
in Fig. 3. The major distribution exhibits mostly a similar
pattern with prominent spikes at the tonic and the fifth. The
Jazz key profile is one example that is noticeably different: it is rather flat compared to the distributions of other genre’s.
It is to be expected that Jazz shows a wider range of pitches
and harmonies and has thus a more uniformly distributed key
profile.

The key profiles for minor have, compared to the major
profiles, less distinct minima for non-scale pitches; especially
the Blues profile is — with the exception of tonic and fifth —
basically uniformly distributed.

4.1 Inter-genre distances

In order to evaluate how distinct genres are with respect to
calculated using the Manhattan distance as shown in Tables 1 and 2.

Table 1: Genre distances for minor tracks using L1-norm

Table 2: Genre distances for major tracks using L1-norm

Genres for which the number of examples were less than 30
are grayed out. The labels are as introduced above, plus Kr
for the Krumhansl key profile, and Tp the Temperley profile [9].

The median major/minor profiles over all genres are denoted
by Mdn. With respect to major key profile distances, the most
similar genres are Rock and Pop while the most mutually
distinct genres are Country and Reggae. For minor tracks,
Disco and Pop are the most similar while Reggae and Blues
are the most distinct.

5. CLASSIFICATION

The distance results presented above indicate what genres are
most similar and dissimilar (with respect to their key profiles).
In order to directly investigate the separability in terms of the
key profiles, the extracted key profiles are used for the task of
musical genre classification — a well studied field in MIR [10].

The most widely used features in this area are timbre features
such as Mel Frequency Cepstral Coefficients (MFCC). MFCC
pick up on instrumental and timbral differences between gen-
res, although they are not totally independent of harmonic and
tonal properties [11]. A linear SVM classifier was trained
using extracted the key profiles. For comparison a linear SVM
was also trained using 24-dimensional timbre features vector
comprising the mean and standard deviation of the first 12
MFCCs. We used libSVM [12] and picked the SVM parame-
ters with a grid search and 5-fold cross validation on a separate
stratified split of the data. The classification is carried out for
the 9 classes described above.

For the distance measure presented above, the extracted pitch
chroma was shifted by the tonic from the ground truth (referred
to as KP3 below). While such key profiles can be used to show
similarity, they cannot be used in a general classification scenario
as no key label will be available. Therefore, we also evaluated
the following approaches to estimating a key-independent representation: (i) **KP0: unshifted** — the overall pitch chroma of each song is used as extracted; (ii) **KP1: transposition by max** — the overall pitch chroma of each song is shifted by the index of the maximum of this pitch chroma. Detecting the index of the maximum can be interpreted as the simplest possible tonic estimation; (iii) **KP2: Fourier transform** — the shift dependent on the tonic can be understood as the phase of the pitch chroma. The magnitude spectrum of the extracted pitch chroma is thus a phase-independent (and therefore tonic independent) representation; (iv) **KP3: transposition by ground truth** — the overall pitch chroma of each song is shifted by the tonic index annotated in the ground truth. Three classification scenarios have been evaluated: (i) only major keys, (ii) only minor keys, and (iii) the whole key-labeled data set without any differentiation between major and minor. All scenarios were carried out with the individual key profile features as well as with the combination of MFCCs and these features. The presented results are computed with 10-fold cross validation.

### 5.1 Results and discussion

Table 3 summarizes the results of the SVM classification for the different key profile computations and their performance when combined with the MFCCs.

A minimum result is the output of a hypothetical classifier that simply predicts the majority class (ZeroR). The classification accuracy for this minimal classifier for our data set would be 26% for major, 20% for minor, and 13% for the overall data set. The accuracy of a random pick is approximately 11%. Tzanetakis and Cook reported a 23% classification accuracy for the complete set with 10 classes (i.e., including samples with ambiguous tonality) using a GMM classifier with a set of simple pitch histogram features and a 47% accuracy for 10 MFCCs [13]. These numbers may serve as a base-line.
We presented an analysis of key profiles for different genres when combining major and minor keys in one data set, the Although not random, the overall classification performance similar range as inter-song differences between profiles. dent profiles as inter-genre differences are small and in a normalized pitch chroma. indicating the usefulness of the tonic-improvements by using the shifted key profiles instead of the genre differences. The classification results show modest some genres may indeed have distinct key profiles, but overall, classifying, adding an additional source of error to the analy-

Table 3: Average classification accuracy and standard deviation over folds for different feature combinations.

<table>
<thead>
<tr>
<th>Feature</th>
<th>Major</th>
<th>Minor</th>
<th>All</th>
</tr>
</thead>
<tbody>
<tr>
<td>KP0</td>
<td>35.35 ± 2.53</td>
<td>37.90 ± 1.39</td>
<td>35.04 ± 1.97</td>
</tr>
<tr>
<td>KP1</td>
<td>37.24 ± 2.35</td>
<td>34.72 ± 2.21</td>
<td>35.91 ± 1.65</td>
</tr>
<tr>
<td>KP2</td>
<td>37.74 ± 2.29</td>
<td>36.36 ± 2.58</td>
<td>32.36 ± 2.08</td>
</tr>
<tr>
<td>KP3</td>
<td>40.33 ± 2.04</td>
<td>39.66 ± 3.33</td>
<td>33.83 ± 0.92</td>
</tr>
<tr>
<td>MFCC</td>
<td>57.26 ± 1.50</td>
<td>63.33 ± 1.69</td>
<td>58.25 ± 2.55</td>
</tr>
<tr>
<td>KP0+MFCC</td>
<td>59.17 ± 1.98</td>
<td>66.84 ± 2.57</td>
<td>62.44 ± 1.76</td>
</tr>
<tr>
<td>KP1+MFCC</td>
<td>61.88 ± 1.34</td>
<td>64.27 ± 2.22</td>
<td>62.86 ± 1.73</td>
</tr>
<tr>
<td>KP2+MFCC</td>
<td>61.53 ± 1.65</td>
<td>62.08 ± 2.49</td>
<td>61.48 ± 1.38</td>
</tr>
<tr>
<td>KP3+MFCC</td>
<td>61.96 ± 1.42</td>
<td>67.37 ± 1.46</td>
<td>63.10 ± 2.39</td>
</tr>
</tbody>
</table>

6. CONCLUSION

We presented an analysis of key profiles for different genres and investigated inter-genre distances and separability using distance measures and classification. The results show that some genres may indeed have distinct key profiles, but overall, the similarities between key profiles seems to outweigh the genre differences. The classification results show modest improvements by using the shifted key profiles instead of the average pitch chroma, indicating the usefulness of the tonic-normalized pitch chroma.

Overall, the results support the notion of using genre-indepen-

dent profiles as inter-genre differences are small and in a similar range as inter-song differences between profiles.

7. REFERENCES

[11] T. L. Li and A. B. Chan, “Genre classification and the invariance of MFCC features to key and tempo,” in *Proceed-
[13] G. Tzanetakis and P. Cook, “Musical genre classification of audio signals,” *Transactions on Speech and Audio Pro-