# GENRE-SPECIFIC KEY PROFILES

**First author**
Affiliation1
author1@ismir.edu

**Second author**
**Retain these fake authors in**
**submission to preserve the formatting**

**Third author**
Affiliation3
author3@ismir.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 12-dimensional pitch chroma is a popular feature in music information retrieval, as it is a compact and robust representation of the tonal content of the audio signal. For example, 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 [5], using diatonic models [4], extraction from MIDI data [11], and extraction from audio data [**?**]. In this paper we analyze the distributions of key profiles extracted from audio with respect to musical genre. More specifically, we investigate the similarity of genre-specific key profiles and their separability in order to understand the tonal similarity of genres. Given the variety of different templates proposed in the literature, we search for evidence that musical genre impacts the key profile.

After we described the data set (Section 2) and the feature extraction (Section 3), we present the genre-specific key profiles and their inter-genre distances in Section 4. Section 5 uses an SVM classifier in order to estimate how separable the key profiles actually are, given different strategies for achieving key-invariance of the extracted pitch chromas.



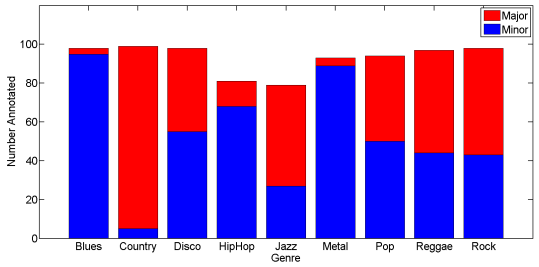


**Figure 1**. Per genre Major/Minor distributions

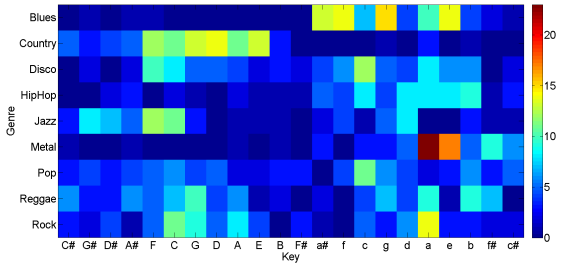


**Figure 2**. Key distribution per genre

## 2. DATA SET

The data set used was the GTZAN collection. [1] While this set is old and it is clear that it has its disadvantages [8], it is a well-known, widely-used, and, last but not least, easily available set for genre classification tasks. It consists of a collection of 1000 song excerpts divided into ten genres: Blues, Classical, Rock, Reggae, Pop, Metal, Rock, Jazz, Country and Hip Hop. Key annotations for the track are publicly available. Instances that included key modulations or with a not easily identifiable key have been omitted from the labels. For example, none of the excerpts from the Classical genre is not annotated. The number of annotated files therefore reflects the difficulty of unambiguously identifying the key; Figure 2 gives an overview of the number of annotated files and their mode.

Figure 2 gives a more detailed visualization of the key distribution per genre with the root notes sorted with respect to the circle of fifths and major modes (left) separated from minor modes (right). 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

[1] http://marsyas.info/download/data_sets

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 that are well-suited to the electric guitar and bass and correspond to the two lowest open strings. Rock is clustered around keys with few accidentals

It is worth noting that Li and Chan presented another set of key annotations for the GTZAN data set [6]. An examination of the two independent annotations reveals some disagreements between the two annotations: overall, about 85 percent of the data is labeled identically. Of the differences where there was disagreement on whether one key was clearly identifiable or not, the majority were found in the Blues genre with 67 out 127 total disagreements of this type. If we consider only examples where both analyses agreed a modulation had *not* occurred, there was a total of 119 disagreements. The most common differences were: major/relative-minor confusion (38), root/fifth confusion (35) and major/minor confusion (13).

## 3. FEATURE EXTRACTION

We extract the key profiles per file. We use the term key profile for the overall, root-note independent pitch chroma per file. The detailed extraction is explained below.

### 3.1 Pitch chroma

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

### 3.2 Key profiles

We assume that the key profile of a song in the same mode (major or minor) should be similar between songs within one genre, but is shifted circularly to the song's root note. Under this assumption, we can "convert" each pitch to a key-profile by applying a circular shift to make it root-note independent. In other words, the key profile is the root note independent pitch distribution (e.g., the pitch profile 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).

| Genre | B | C | D | H | J | M | P | Rg | R | Kr | Tp | Mdn |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| B | 0 | 0.35 | 0.40 | 0.68 | 0.44 | 0.34 | 0.35 | 0.43 | 0.31 | 0.51 | 0.59 | 0.33 |
| C | 0.35 | 0 | 0.27 | 0.60 | 0.32 | 0.32 | 0.17 | 0.34 | 0.20 | 0.41 | 0.47 | 0.19 |
| D | 0.40 | 0.27 | 0 | 0.33 | 0.12 | 0.21 | 0.11 | 0.12 | 0.11 | 0.15 | 0.25 | 0.09 |
| H | 0.68 | 0.60 | 0.33 | 0 | 0.27 | 0.41 | 0.42 | 0.28 | 0.43 | 0.24 | 0.29 | 0.40 |
| J | 0.44 | 0.32 | 0.12 | 0.27 | 0 | 0.25 | 0.15 | 0.17 | 0.17 | 0.12 | 0.24 | 0.13 |
| M | 0.34 | 0.32 | 0.21 | 0.41 | 0.25 | 0 | 0.20 | 0.21 | 0.16 | 0.27 | 0.41 | 0.19 |
| P | 0.35 | 0.17 | 0.11 | 0.42 | 0.15 | 0.20 | 0 | 0.19 | 0.08 | 0.23 | 0.29 | 0.05 |
| Rg | 0.43 | 0.34 | 0.12 | 0.28 | 0.17 | 0.21 | 0.19 | 0 | 0.17 | 0.13 | 0.23 | 0.17 |
| R | 0.31 | 0.20 | 0.10 | 0.43 | 0.17 | 0.16 | 0.08 | 0.17 | 0 | 0.23 | 0.30 | 0.06 |
| Kr | 0.51 | 0.41 | 0.15 | 0.24 | 0.14 | 0.27 | 0.23 | 0.13 | 0.23 | 0 | 0.15 | 0.21 |
| Tp | 0.59 | 0.47 | 0.25 | 0.29 | 0.24 | 0.41 | 0.29 | 0.23 | 0.30 | 0.15 | 0 | 0.28 |
| Mdn | 0.33 | 0.19 | 0.09 | 0.40 | 0.13 | 0.19 | 0.05 | 0.17 | 0.06 | 0.21 | 0.28 | 0 |

**Table 1**. Genre Distances for Major tracks using L1-norm

| Genre | B | C | D | H | J | M | P | Rg | R | Kr | Tp | Mdn |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| B | 0 | 0.29 | 0.27 | 0.30 | 0.27 | 0.18 | 0.22 | 0.33 | 0.28 | 0.28 | 0.39 | 0.18 |
| C | 0.29 | 0 | 0.46 | 0.59 | 0.36 | 0.45 | 0.43 | 0.43 | 0.36 | 0.53 | 0.53 | 0.41 |
| D | 0.27 | 0.46 | 0 | 0.17 | 0.20 | 0.16 | 0.10 | 0.15 | 0.19 | 0.14 | 0.20 | 0.12 |
| H | 0.30 | 0.59 | 0.17 | 0 | 0.28 | 0.21 | 0.17 | 0.23 | 0.33 | 0.17 | 0.27 | 0.18 |
| J | 0.27 | 0.36 | 0.20 | 0.28 | 0 | 0.23 | 0.16 | 0.19 | 0.16 | 0.24 | 0.27 | 0.15 |
| M | 0.18 | 0.45 | 0.16 | 0.21 | 0.23 | 0 | 0.13 | 0.26 | 0.20 | 0.14 | 0.31 | 0.09 |
| P | 0.22 | 0.43 | 0.10 | 0.17 | 0.16 | 0.13 | 0 | 0.15 | 0.16 | 0.13 | 0.25 | 0.05 |
| Rg | 0.33 | 0.43 | 0.15 | 0.23 | 0.19 | 0.26 | 0.15 | 0 | 0.19 | 0.18 | 0.23 | 0.18 |
| R | 0.28 | 0.36 | 0.19 | 0.33 | 0.16 | 0.20 | 0.16 | 0.19 | 0 | 0.25 | 0.33 | 0.15 |
| Kr | 0.28 | 0.53 | 0.14 | 0.17 | 0.24 | 0.18 | 0.13 | 0.18 | 0.25 | 0 | 0.16 | 0.16 |
| Tp | 0.39 | 0.53 | 0.20 | 0.27 | 0.27 | 0.31 | 0.25 | 0.23 | 0.33 | 0.16 | 0 | 0.28 |
| Mdn | 0.18 | 0.41 | 0.12 | 0.18 | 0.15 | 0.09 | 0.05 | 0.18 | 0.15 | 0.16 | 0.28 | 0 |

**Table 2**. Genre Distances for Minor tracks using L1-norm

## 4. KEY PROFILE ANALYSIS

### 4.1 Overall key profiles

Figure 3 shows the overall key profiles in a box plot in comparison with other reported profiles. While Krumhansl's "Probe Tone Ratings" [5] are not really a key profile (derived from listening experiments on tonality) they have been frequently used in audio key detection. They seem to correlate well with key profiles (e.g., [4]). Temperley's key profiles are extracted from symbolic data rather than from audio [9, 11].

### 4.2 Genre-specific key profiles

The key profiles of the six most populated genres are plotted in Fig. 4. The major distribution exhibits mostly a similar pattern with prominent spikes at the root note and the fifth. The Jazz distribution is one example that is noticeably different: it is rather flat and more uniform than the distributions of other genre's. It is to be expected that Jazz shows a wider range of pitches and harmonies and thus a more uniform pitch class distribution.

The distributions for minor have, compared to the major distributions, less distinct minima for non-scale pitches; especially the Blues distribution is — with the exception of root note and fifth — basically flat.

### 4.3 Inter-genre distances

In order to evaluate how distinct genres are with respect to their key profile, distance between all profiles were calculated using the Manhattan distance as shown in Tables 1 and 2. Genres for which the number of examples were less than 30 are grayed out. The labels are as follows: B is blues, C country, D disco, H hip-hop, J jazz, M metal, P pop, Rg Reggae, Rk is rock, K is the Krumhansl key profile, T the Temperley profile [10] and finally the overall median major/minor pitch profiles are denoted by Mdn.

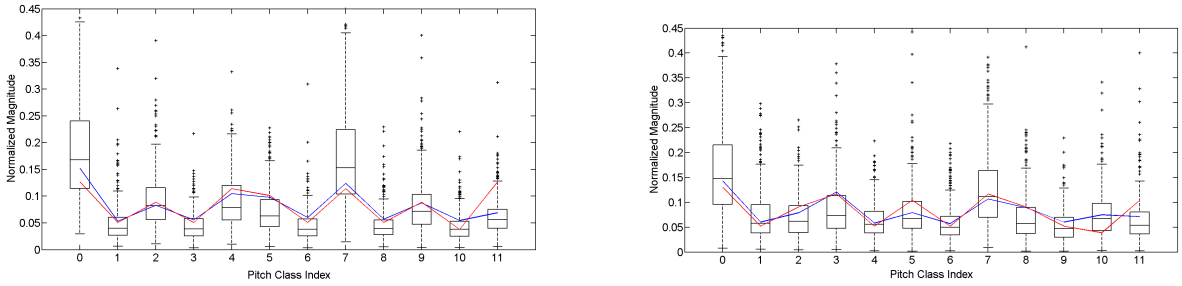


**Figure 3**. Major (left) and minor (right) key profiles for the complete data set, in comparison with two widely-used key profiles (Krumhansl in red and Temperley in blue).

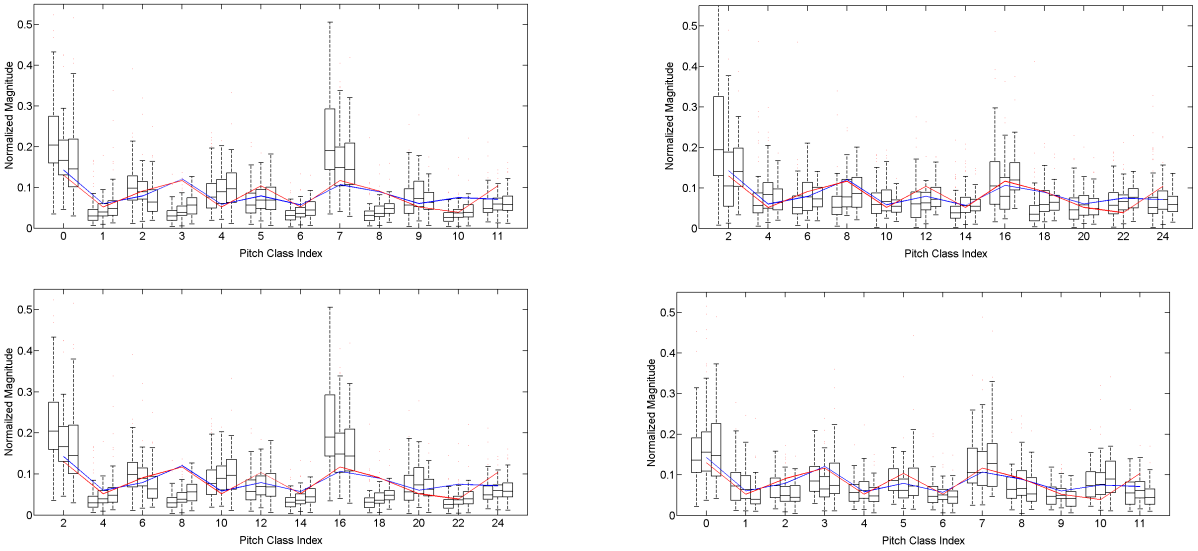


**Figure 4**. Major (left) and minor (right) key profiles for the various genres, in comparison with the Krumhansl and Temperley profiles.

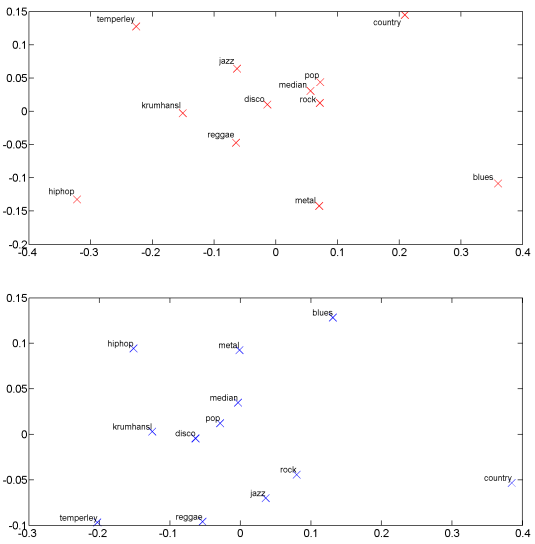


**Figure 5**. MDS plots for Major and Minor profiles using L1 distance

Looking at the distance matrices, we can see that for major keys, the most similar genres are Rock and Pop while the most mutually distinct genres when considering major tracks are Country and Jazz. For minor tracks, Reggae and Pop are the most similar while Reggae and Rock are the most distinct.

To further explore the genre relationships we use Multi-dimensional Scaling, a method for visualizing information contained in an arbitrary distance matrix by projecting the data into a space of smaller dimension while preserving the between-item distances as well as possible i.e. the Euclidean distance in the new space will approximate the dissimilarity in the distance matrix. These plots are presented in Figure 5.

## 5. CLASSIFICATION

Musical genre recognition is a well studied field in MIR [2]. The most widely and successfully used features in this area are timbre features such as Mel Frequency Cepstral Coefficients (MFCC). MFCC seem well suited to picking up on instrumental and timbral differences between genres, although they are not totally independent of harmonic and

tonal properties [6]. While the distance results indicate what genres are most similar and dissimilar, it does not allow an interpretation as to whether the genres distances do actually help to separate genres. In order to investigate how separable our genre-specific key profiles are, we train an SVM classifier with our features (see Sect. 3.2) and compare the results with an MFCC-based classifier. The MFCC were calculated using a freely available MATLAB implementation. [2] The mean and standard deviation of 12 MFCC form a 24-dimensional timbre feature vector. We used libSVM [1] and picked the SVM parameters 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.

While the KP3 feature (shifted by ground truth root note) are sufficient to prove or disprove separability, they can not used in a general classification scenario, as there will be no key label available. Therefore we also tested the following approaches to estimating a key-independent representation:

- **KP0: Unshifted**
  The overall pitch chroma of each song is used as extracted.

- **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 root note estimation.

- **KP2: Fourier transform**
  The shift dependent on the root note 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 root-note independent) representation.

- **KP3: Transposition by ground truth**
  The overall pitch chroma of each song is shifted by the root note index annotated in the ground truth.

We investigated 3 classification scenarios: (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 base line 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%.

[2] http://labrosa.ee.columbia.edu/matlab/rastamat/

| Feature | Major | Minor | All |
|---|---|---|---|
| KP0 | $35.35 \pm 2.53$ | $37.90 \pm 1.39$ | $35.04 \pm 1.97$ |
| KP1 | $37.24 \pm 2.35$ | $34.72 \pm 2.21$ | $35.91 \pm 1.65$ |
| KP2 | $37.74 \pm 2.29$ | $36.36 \pm 2.58$ | $32.36 \pm 2.08$ |
| KP3 | $40.33 \pm 2.04$ | $39.66 \pm 3.33$ | $33.83 \pm 0.92$ |
| MFCC | $57.26 \pm 1.50$ | $64.33 \pm 1.69$ | $58.25 \pm 2.55$ |
| KP0+MFCC | $59.17 \pm 1.98$ | $66.84 \pm 2.57$ | $62.44 \pm 1.76$ |
| KP1+MFCC | $61.88 \pm 1.34$ | $64.27 \pm 2.22$ | $62.86 \pm 1.73$ |
| KP2+MFCC | $61.53 \pm 1.65$ | $62.08 \pm 2.49$ | $61.48 \pm 1.38$ |
| KP3+MFCC | $61.96 \pm 1.42$ | $67.37 \pm 1.46$ | $63.10 \pm 2.39$ |

**Table 3**. Classification accuracy for different feature combinations.

The overall range of results is consistent with the results of Tzanetakis and Cook, who — for the complete set with 10 classes (i.e., including the ambiguous samples) using a GMM classifier — reported a 23% classification accuracy for a set of simple pitch histogram features and a 47% accuracy for 10 MFCCs [12].

Unsurprisingly, the MFCCs vastly outperformed the key profile features alone. Although not random, the overall classification performance given the key profiles is only mediocre. This indicates that while the key profiles provide genre-specific information, there are still a lot of similarities between genres. The combined feature set results in a slight performance increase in overall accuracy of around 3-4% indicating that the pitch chroma distributions contain some genre-relevant information.

For the Major and Minor subsets, we can observe a slight performance increase for the shifted profiles (most notably the KP3 profile, shifted by the ground truth), while this shifting actually results in a small decrease for the overall data set. That indicates that when combining Major and Minor keys in one data set, the root note information is actually more important for classification than the mode, since the root note information has been removed by the shifting. It also indicates that overall, the distances between Major and Minor profiles are larger than the distances between genre profiles.

It should be pointed out that neither the shifting by ground truth nor the split of the data set into Major and Minor are a realistic classification scenario, since this data is not available (or could be only be estimating the key before classifying, with the repercussion of adding an additional source of error to the analysis). Still, the objective of this analysis was to investigate the separability of root-note independent key profiles; we can at least observe some inter-genre separability.

Tables 4 and 5 display the confusion matrices of the classification results with the KP3 features. The confusion matrices give more detailed insight which genres are more separable than others, and also allow some insights on the classifier performance in the case of non-uniform class distributions.

The confusion matrices let us observe a strong tendency

| | B | C | D | H | J | M | P | Rg | Rk | Total |
|---|---|---|---|---|---|---|---|---|---|---|
| B | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 0 | 2 |
| C | 0 | 51 | 1 | 0 | 5 | 0 | 3 | 6 | 9 | 75 |
| D | 0 | 12 | 7 | 6 | 0 | 2 | 2 | 7 | 8 | 44 |
| H | 0 | 2 | 2 | 1 | 1 | 0 | 2 | 1 | 1 | 10 |
| J | 0 | 8 | 1 | 0 | 26 | 0 | 0 | 1 | 5 | 41 |
| M | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 2 | 0 | 3 |
| P | 0 | 17 | 2 | 0 | 0 | 0 | 4 | 7 | 5 | 35 |
| Rg | 0 | 11 | 4 | 0 | 1 | 0 | 2 | 7 | 18 | 43 |
| Rk | 0 | 20 | 4 | 0 | 2 | 0 | 2 | 4 | 13 | 45 |
| Total | 0 | 121 | 21 | 2 | 35 | 0 | 16 | 36 | 57 | |

**Table 4**. Confusion matrix, Major keys

| | B | C | D | H | J | M | P | Rg | Rk | Total |
|---|---|---|---|---|---|---|---|---|---|---|
| B | 51 | 0 | 1 | 3 | 4 | 14 | 2 | 0 | 1 | 76 |
| C | 2 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 0 | 4 |
| D | 1 | 0 | 14 | 8 | 0 | 12 | 4 | 3 | 2 | 44 |
| H | 30 | 0 | 2 | 32 | 0 | 13 | 1 | 4 | 0 | 55 |
| J | 12 | 0 | 1 | 0 | 8 | 0 | 0 | 0 | 0 | 21 |
| M | 14 | 0 | 4 | 14 | 0 | 34 | 3 | 1 | 1 | 71 |
| P | 7 | 0 | 8 | 7 | 0 | 15 | 1 | 1 | 1 | 40 |
| Rg | 2 | 0 | 2 | 8 | 0 | 9 | 3 | 10 | 1 | 35 |
| Rk | 13 | 0 | 4 | 1 | 0 | 7 | 1 | 5 | 3 | 34 |
| Total | 132 | 0 | 36 | 73 | 12 | 105 | 15 | 25 | 9 | |

**Table 5**. Confusion matrix, Minor keys

for the classifier to choose the majority class. For Major, this is the class country and for Minor the classes metal and blues. The genres with the highest classification accuracies are country and jazz for Major and blues, hip-hop and metal for minor. The results agree with the MDS plot in Section 4 in that the most distant genres in that plot seem to receive the highest classification accuracies. Given the overall range of the classification results, however, a more detailed interpretation would be most likely more speculation than based on evidence.

## 6. CONCLUSION

We presented an analysis of key profiles for different genres and investigated inter-genre distances and separability with a number of methods, including, distance measures, Multi-Dimensional Scaling, 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 how modest improvements by using the shifted key profiles instead of the average pitch chroma, indicating the usefulness of the root-note-normalized pitch chroma. These improvements, however, disappear when Major and Minor samples are not treated separately (which, in the case of classification, would be the normal case as the key is unknown). Possible other reasons for the very slight accuracy improvements might be found in the data set (definition of genres, sample selection, size, other deficiencies), but in general we believe that the results make sense and are conceivable.

Overall, the results support the notion of using genre-independent profiles as inter-genre differences are small and inter-song differences between profiles seem to be in a similar range.

## 7. REFERENCES


[1] Chih-chung Chang and Chih-jen Lin. LIBSVM: a library for support vector machines. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 2(3), 2011.

[2] Zhouyu Fu, Guojun Lu, Kai Ming Ting, and Dengsheng Zhang. A survey of audio-based music classification and annotation. *IEEE Transactions on Multimedia*, 13(2):303—319, April 2011.

[3] Takuya Fujishima. Realtime chord recognition of musical sound: a system using common lisp music. In *Proceedings of the International Computer Music Conference (ICMC)*, 1999.

[4] Özgür Izmirli. Template based key finding from audio. In *Proceedings of the International Computer Music Conference (ICMC)*, Barcelona, September 2005.

[5] Carol L Krumhansl. *Cognitive Foundations of Musical Pitch*. Oxford University Press, New York, 1990.

[6] Tom LH Li and Antoni B Chan. Genre classification and the invariance of MFCC features to key and tempo. In *Proceedings of the International Multimedia Modeling Conference (MMM)*, page 317–327, Taipei, 2011. Springer.

[7] Meinard Müller. *Information Retrieval for Music and Motion*. Springer, Berlin, 2007.

[8] Bob L Sturm. An analysis of the GTZAN music genre dataset. In *Proceedings of the 2nd International ACM Workshop on Music Information Retrieval with User-Centered and Multimodal Strategies (MIRUM),*, Nara, 2012. 00017.

[9] David Temperley. Bayesian models of musical structure and cognition. *Musicae Scientiae*, 8(2):175—205, 2004.

[10] David Temperley. The tonal properties of pitch-class sets : Tonal implication, tonal ambiguity, and tonalness. *Computing in Musicology*, 15:24–38, 2007.

[11] David Temperley and Elizabeth West Marvin. Pitch-class distribution and the identification of key. *Music Perception: An Interdisciplinary Journal*, 25(3):193—212, February 2008.

[12] George Tzanetakis and Perry Cook. Musical genre classification of audio signals. *Transactions on Speech and Audio Processing*, 10(5):293–302, July 2002. 01816.