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

Pitch chroma are a popular feature for many MIR tasks. Using the GTZAN data set we investigate the distributions of pitch chroma for 9 different genres, the degree to which these genres can be identified using these distributions and different strategies for achieving key-independence; namely transposition of the chroma according to its maximum value and 12-point FFT. We find that combining pitch chroma with commonly used MFCCs can lead to small increase in classification accuracy using a Support Vector Machine. Furthermore these results show that the imposing key-independence has a surprisingly small affect on performance.

1. INTRODUCTION

Musical genre recognition is a well studied field in MIR [*ref]. As with any classification task, the feature used to summarize tracks is an extremely important concern. Previous work in this area has shown that timbral features, particularly Mel Frequency Spectrum Coefficients, are especially suited to the task of predicting genre. While MFCCs are suited to picking up on instrumental and timbral difference between genres, some work has shown that they are not totally independent of harmonic or tonal information [*ref li]. Here we investigate the extent to which tonal information can be used to discern genres y examining the distributions of pitch chroma within each of 9 different genres. First we examine the overall distribution of keys for each genre. Next different methods for transposing pitch chroma to a key-independent representation are introduced and we look at the chroma distributions using chroma box plots, inter-genre distance and mulch-dimensional scaling. In the final section we test the degree to which the chroma distribution separate gen-res by performing classification with an SVM using chroma and combined chroma/MFCC features.

2. DATA SET

The data set used was the GTZAN collection, which is popular 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. First we examine the general layout of the data set, specifically with respect to the key distributions within genres. Key annotations for each track were produced manually and are publicly available. [1] Instances that included key modulations or where hard to identify have been omitted. The Classical genre has not been annotated. The number of annotated files reflects the difficulty of unambiguously identifying the key — the number of annotated files is shown in the Table 1.

genre	# major	# min	# annotated
blues	3	95	98
classic	0	0	0
country	94	5	99
disco	43	55	98
hiphop	13	68	81
jazz	52	27	79
metal	4	89	93
pop	44	50	94
reggae	53	44	97
rock	55	43	98
overall	361	476	837

Table 1. Number of annotated data set entries

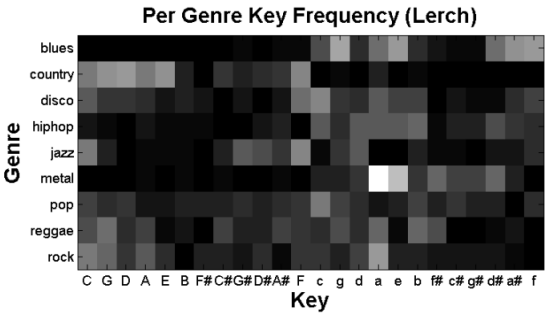

Figure 1. Key distribution per genre (brighter means more frequent): SHIFT BOTH CMAJOR AND AMINOR TO THE MIDDLE

Figure 1 visualizes the distribution of keys for each genre. The distribution of keys in each genre proved unsurprising. The relation of major v.s. minor modes is very skewed for blues and metal (predominantly minor) as well as country (predominantly major), while the genres disco, pop, reggae, and rock appear to be quite balanced. Jazz tracks tend

[1] github.com/alexanderlerch/data_set

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-min or e-min, keys that are well-suited to the electric guitar and bass and correspond to the two lowest open strings.

Li and Chan presented another set of key annotations for the GTZAN data set [5]. 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 a key modulation occured 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* occured, 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). DISCUSS DIFFERENCES MORE IN DETAIL??

3. FEATURE EXTRACTION

We extract the key profiles per file. The term key profile as we use it here is 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) [6]. 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 [2]. 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 XXX, the hop size is XXX.

3.2 Key profile

We assume that the pitch chroma 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 nine indices to the left so that the bin of pitch class A lands on the first index).

We investigated the following approaches to obtaining a key-independent representation.

3.2.1 Transposition by ground truth

The overall pitch chroma of each song is shifted by the root note index annotated in the ground truth.

	B	C	D	H	J	M	P	Rg	Rk	K	T
B	0	0.29	0.31	0.47	0.33	0.23	0.28	0.32	0.23	0.33	0.37
C	0.29	0	0.17	0.46	0.28	0.28	0.13	0.22	0.12	0.27	0.33
D	0.31	0.17	0	0.29	0.13	0.24	0.14	0.11	0.13	0.13	0.21
H	0.47	0.46	0.29	0	0.18	0.3	0.35	0.27	0.38	0.2	0.25
J	0.33	0.28	0.13	0.18	0	0.24	0.17	0.15	0.21	0.12	0.18
M	0.23	0.28	0.24	0.3	0.24	0	0.25	0.23	0.2	0.19	0.33
P	0.28	0.13	0.14	0.35	0.17	0.25	0	0.14	0.1	0.18	0.23
Rg	0.32	0.22	0.11	0.27	0.15	0.23	0.14	0	0.15	0.1	0.15
Rk	0.23	0.12	0.13	0.38	0.21	0.2	0.1	0.15	0	0.18	0.26
K	0.33	0.27	0.13	0.2	0.12	0.19	0.18	0.1	0.18	0	0.15
T	0.37	0.33	0.21	0.25	0.18	0.33	0.23	0.15	0.26	0.15	0

Figure 4. Manhattan distance between the major profiles of all genres and three overall profiles MAYBE INCLUDE THE OVERALL MEDIAN VALUES HERE AS WELL

3.2.2 Transposition by max

The overall pitch chroma of each song is shifted by the index of the maximum of this pitch chroma. This can be interpreted as the simplest possible root note estimation.

3.2.3 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.

4. RESULTS

4.1 Overall key profiles

Figure 2 shows the overall key profiles in a box plot in comparison with two widely-used measures. Krumhansl's "Probe Tone Ratings" [4] are not really a key profile, but has been frequently used in audio key detection since they seem to correlate well (e.g., [3]). Temperley's key profiles [7] are derived from symbolic data rather than from audio. MAKE SURE THE TEMPERLEY PROFILES ARE THE SAME AS IN THE IZMIRLI PAPER AND THIS IS THE CORRECT CITATION (DONT HAVE THE BOOK).

4.2 Genre-specific key profiles

The key profiles of the six most populated genres is plotted in Fig. **??**.

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 Figs. 4 and 5. 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 and T the Temperley profile [7].

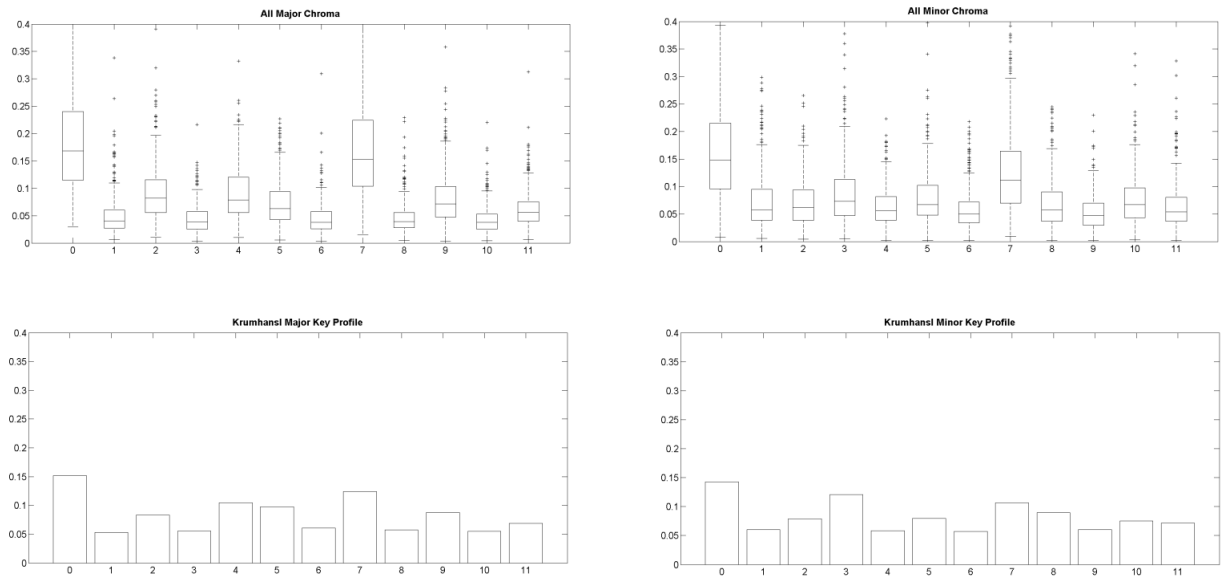

Figure 2. Major (left) and minor (right) key profiles for the complete data set, in comparison with two widely-used key profiles (Krumhansl and Temperley) CAN YOU OVERLAY THE BOX PLOTS WITH THE KRUMHANSL/TEMPERLEY GRAPHS? SHOULDNT THE WHISKERS BE THERE FOR ALL PITCHES?

	B	C	D	H	J	M	P	Rg	Rk	K	T
B	0	0.22	0.21	0.27	0.16	0.17	0.17	0.25	0.24	0.25	0.36
C	0.22	0	0.3	0.45	0.25	0.3	0.29	0.26	0.24	0.39	0.42
D	0.21	0.3	0	0.22	0.19	0.12	0.07	0.12	0.13	0.14	0.19
H	0.27	0.45	0.22	0	0.27	0.22	0.21	0.26	0.33	0.18	0.27
J	0.16	0.25	0.19	0.27	0	0.21	0.14	0.2	0.22	0.24	0.28
M	0.17	0.3	0.12	0.22	0.21	0	0.11	0.19	0.16	0.17	0.29
P	0.17	0.29	0.07	0.21	0.14	0.11	0	0.1	0.12	0.13	0.21
Rg	0.25	0.26	0.12	0.26	0.2	0.19	0.1	0	0.12	0.18	0.22
Rk	0.24	0.24	0.13	0.33	0.22	0.16	0.12	0.12	0	0.21	0.26
K	0.25	0.39	0.14	0.18	0.24	0.17	0.13	0.18	0.21	0	0.16
T	0.36	0.42	0.19	0.27	0.28	0.29	0.21	0.22	0.26	0.16	0

Figure 5. Manhattan distance between the minor profiles of all genres and three overall profiles MAYBE INCLUDE THE OVERALL MEDIAN VALUES HERE AS WELL

4.4 Discussion

Overall each genre roughly exhibits the expected major pattern with some exceptions, for example, Jazz. The Jazz distribution is noticeably more flat than other genres — it doesn't exhibit the same peakiness at the root and fifth for example and the relative ratios of each interval are noticeably more uniform. This agrees with the classification results of Section 5, where Jazz was one of the least confused major genres.

For the minor examples, the distributions look much more uniform with relatively more mass in non-diatonic bins — for example the flat-second (C#) and-sharp fifth (G#) genres.

Some facts revealed from the key profile distances, the most mutually distinct genres when considering major tracks are Country and Jazz while the the most similar are Rock and Pop. For minor tracks, Reggae and Pop are most similar while Reggae and Rock are the most distinct.

5. CLASSIFICATION

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. To investigate this effect, we train and evaluate a Support Vector Machine using libSVM [1]. The SVM hyper-parameters (C,) representing the margin and kernel widths were tuned using a grid-search and 5-fold cross validation on a separate stratified split of the data (I HAVE NO IDEA WHAT THAT MEANS). Best performance was achieved using a non-linear SVM with Radial Basis Function kernel. We investigated the following scenarios: (i) the whole key-labeled data set with no differentiation between major and minor and (ii) the major/minor subsets individually in order to control for the differences in major/minor distribution between keys. 10-fold cross validation accuracy was calculated using the raw pitch chroma, max-shifted key profile, the Fourier pitch chroma and the key profile shifted according to the ground truth. We compare the results to a classifier based on a set of MFCC features as well as a combination of MFCC and key profile features. MFCC features were calculated according to REFXXX: WHAT COMPUTATION DID YOU USE?. The mean and standard deviation of 12 MFCCs form a 24-dimensional timbre feature vector.

Table 2 summarizes the results of the SVM classification for the different key profile computations, and Tab. 3 shows the mode-independent results when combined with the MFCC features.

Figures 6, 7, and 8 display the confusion matrices of the

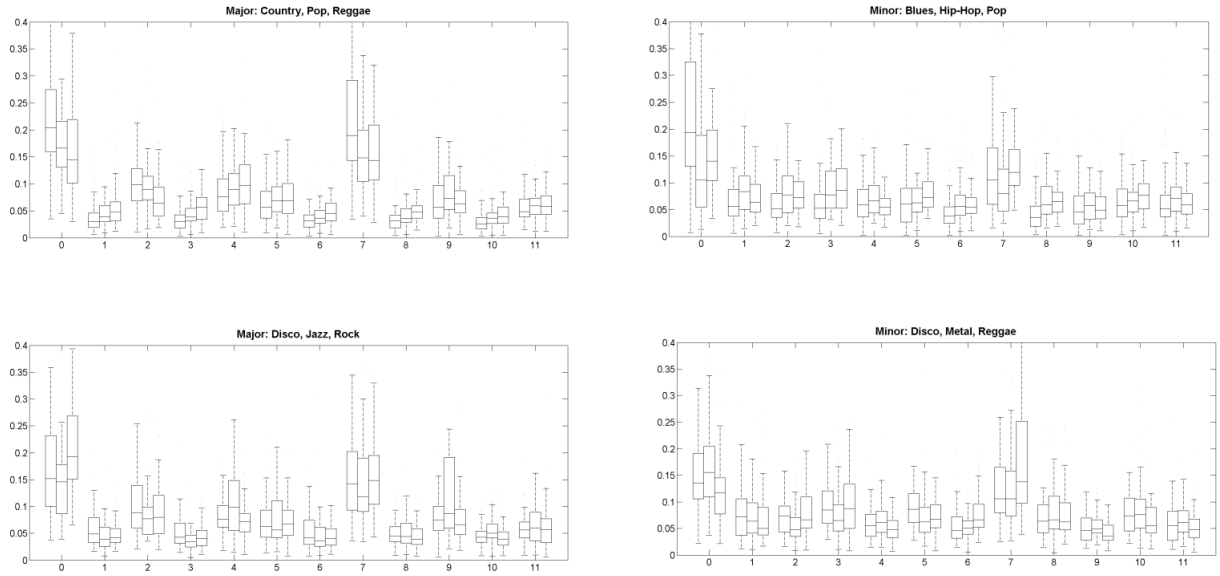

Figure 3. Major (left) and minor (right) key profiles for the various genres, in comparison with two widely-used key profiles (Krumhansl and Temperley) CAN YOU OVERLAY THE BOX PLOTS WITH THE KRUMHANSL/TEMPERLEY GRAPHS? SHOULDNT THE WHISKERS BE THERE FOR ALL PITCHES?

Feature	Major	Minor	All
Raw	35.35 ± 2.53	37.90 ± 1.39	35.04 ± 1.97
Max	37.24 ± 2.35	34.72 ± 2.21	**35.91 ± 1.65**
FFT	37.74 ± 2.29	36.36 ± 2.58	32.36 ± 2.08
GT	**40.33 ± 2.04**	**39.66 ± 3.33**	33.83 ± 0.92

Table 2. Classification accuracy for different key profile computations

Feature	All
MFCC	55.01 ± 2.96
MFCC+Raw	59.31 ± 2.08
MFCC+Max	59.14 ± 1.7
MFCC+FFT	57.93 ± 1.66
MFCC+GT	**59.94 ± 2.02**

Table 3. Classification accuracy for different key profile computations in combination with MFCC features

classification results. PLEASE ADD ANOTHER ROW TO THE MATRICES WITH THE SUM OVER ROWS - BUT WE HAVE TO REDO ALL OF THESE ANYWAY, WE DONT WANT THEM AS PNGs. DESCRIBE AND DISCUSS ME.

5.1 Results and discussion

One fact revealed in this analysis is the relatively small performance increase observed by sorting the chroma, even when using the ground truth directly. By sorting the chroma any information on the genre given by the key has been removed (for example as a prior estimate based on the key distributions in each genre) which would, it is hoped, reveal any differences in the distributions. However, whether or not this technique improves performance depends on the initial distribution of pitch chroma; suppose genre groups are perfectly separable given their key-independent chroma distributions and then for each song we transpose it into a new key with some related probability. Since cyclic-transpositions correspond to reflections this process is unlikely to make the data significantly more or less separable, which may explain the small in-crease in performance when using sorted chroma. I STILL CANNOT FOLLOW HERE! Splitting the data into major and minor results in slightly improved accuracy, however in this case we have added information by using the ground truth and realistically harmony estimation should be done automatically which would be an additional source of error. Unsurprisingly MFCCs vastly outperformed pitch chroma features alone. However, combining features results in a slight performance increase in overall accuracy of around 4-5

	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	0	0	0	2	7	6	34
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

Figure 6. CONFUSION MATRIX 1, major

	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	3	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

Figure 7. CONFUSION MATRIX 2, minor

	B	C	D	H	J	M	P	Rg	Rk	Total
B	54	3	3	2	1	3	5	4	5	80
C	5	55	6	1	0	1	3	3	6	80
D	4	5	40	8	2	2	9	3	7	80
H	0	0	9	44	13	3	2	9	0	80
J	1	1	1	12	58	1	3	0	3	80
M	3	0	4	3	2	63	0	1	4	80
P	2	9	10	7	6	0	41	5	0	80
Rg	4	4	3	12	1	0	3 4	9	4	80
Rk	4	11	18	0	3	7	4	4	29	80

Figure 8. CONFUSION MATRIX 3, pc + mfcc

MENTION RANDOM CLASSIFICATION RATE - SHOULD BE AROUND 11%, but we have nonuniform class distributions.

LETS NOT FORGET THAT WE DO NOT ACTUALLY WANT TO CLASSIFY, WE JUST WANT TO SHOW THAT THE GENRES ARE SEPARABLE WRT THEIR KEY PROFILE.

DISCUSS CONFUSION MATRICES

6. CONCLUSION

- some genres are separable, but overall the similarities between the key profiles outweigh the differences
- mention most similar and most separate genres
- not a proposal to use these profiles but to investigate whether there are genre dependent differences or not.

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

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

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

[5] 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.

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

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