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 MFCC 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

- point out usage of pitch histograms/key profiles/pitch class profiles in MIR tasks (genre classification, key detection, anything else)
- plot krumanshl vs. various temperley profiles
- differences between midi-derived key profiles and audio-derived [4, 9]
- what do we want to do here: analyze the differences between audio-based key profiles between genres
- how we do it: use a data set that is annotated both with genres and keys, compute/visualize differences, use MDS and classification to analyze differentiability
- how might this data be of use

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 [10], 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. [2] 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; Table **??** gives an overview of the number of annotated files and their mode.

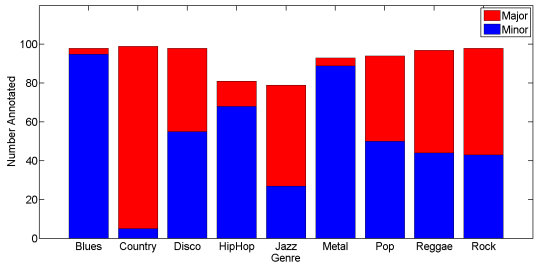

Figure 1. Per genre Major/Minor distributions

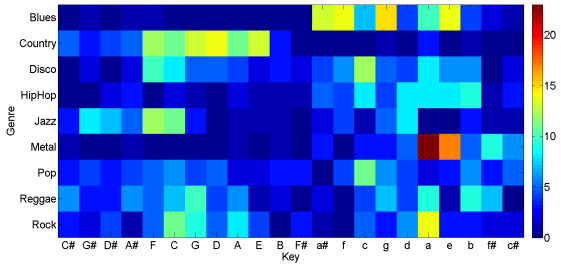

Figure 2. Key distribution per genre

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 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

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

[2] github.com/alexanderlerch/data_set

It is worth noting that Li and Chan presented another set of key annotations for the GTZAN data set [7]. An examination of the two independent annotations reveals some disagreements between the two annotations: overall, about 85 percent of the data is labeled identically. ICH VERSTEHE DIE ZAHLEN HIER NICHT: 67 + 119 IST DOCH MEHR ALS 15%?? Of the differences where there was disagreement on whether a key modulation occurred 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 [8]. 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).

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" [6] 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., [5]). Temperley's key profiles are extracted from symbolic data rather than from audio [11, 13].

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

4.2 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 [12] and finally the overall median major/minor pitch profiles are denoted by Mdn.

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.

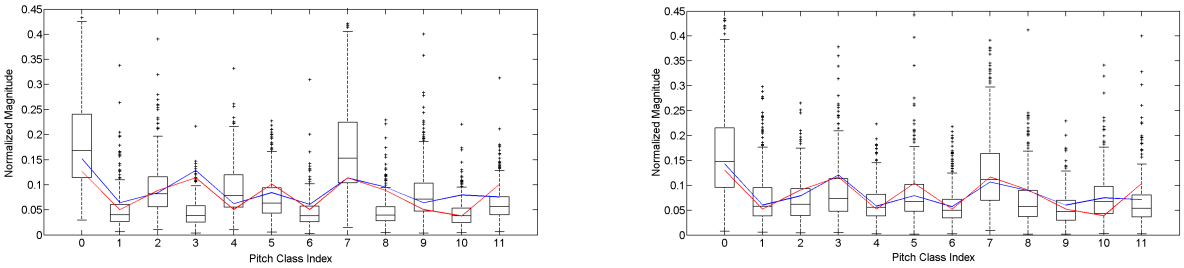

Figure 3. Major (left) and minor (right) key profiles for the complete data set, in comparison with two widely-used key profiles (Krumhansl in red and Temperley in blue).

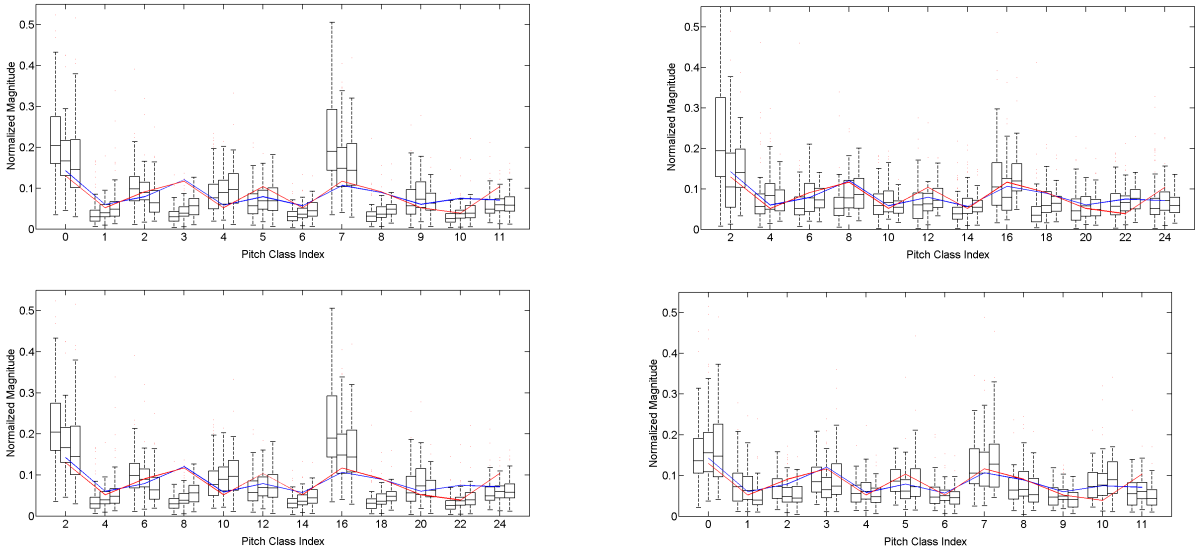

Figure 4. Major (left) and minor (right) key profiles for the various genres, in comparison with the Krumhansl and Temperley profiles.

To further explore the genre relationships we use Multidimensional 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 [7]. 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.

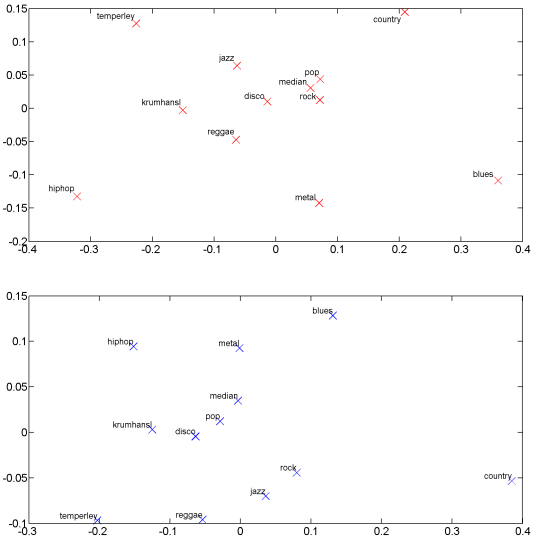

Figure 5. MDS plots for Major and Minor profiles using L1 distance

The MFCC were calculated using a freely available MATLAB implementation. [3] 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) 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. 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 MFCC [14].

Table 3 summarizes the results of the SVM classification for the different key profile computations and how they perform when combined with the MFCC.

Figures 6, 7 and 8 display the confusion matrices of the classification results. PLEASE ADD ANOTHER ROW TO THE MATRICES WITH THE SUM OVER COLS - WE HAVE TO REDO ALL OF THESE ANYWAY, WE DONT WANT THEM AS PNGs. DESCRIBE AND DISCUSS ME.

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

Feature	Major	Minor	All
KP0	35.35 ± 2.53	37.90 ± 1.39	35.04 ± 1.97
KP1	37.24 ± 2.35	34.72 ± 2.21	35.91 ± 1.65
KP2	37.74 ± 2.29	36.36 ± 2.58	32.36 ± 2.08
KP3	40.33 ± 2.04	39.66 ± 3.33	33.83 ± 0.92
MFCC	57.26 ± 1.50	64.33 ± 1.69	58.25 ± 2.55
KP0+MFCC	59.17 ± 1.98	66.84 ± 2.57	61.80 ± 1.76
KP1+MFCC	61.88 ± 1.34	64.27 ± 2.22	62.86 ± 1.73
KP2+MFCC	61.53 ± 1.65	62.08 ± 2.49	61.48 ± 1.38
KP3+MFCC	61.96 ± 1.42	67.37 ± 1.46	61.48 ± 1.98

Table 3. Classification accuracy for different feature combinations.

	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

5.1 Results and discussion

- table: profile-based classification worse than timbre classification
- table: ground truth shifting better than other features, except when combining modes
- table: FFT shift in the same range as Max shift
- table: unshifted not much worse than shifted
- major confusion matrix: many classifications into country, minor confusion matrix: many classifications into metal — why?

	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

COULD YOU ALSO GENERATE THE NORMALIZED CONFUSION MATRICES, SO THAT THE ACCURACY PER CLASS IS ON THE DIAGONAL? I DON'T WANT TO ADD THEM TO THE PAPER¡ I JUST WANT TO LOOK - MAYBE THERE IS SOMETHING MORE EVIDENT THAN WITH THE ABSOLUTE NUMBERS.

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. 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 3-4% indicating that the pitch chro-ma distributions contain at least some genre-relevant information.

FROM THE OVERALL PROFILE SECTION ABOUT JAZZ: This agrees with the classification results of Section 5, where Jazz was one of the least confused major genres.

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

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] 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] Emilia Gómez. Tonal description of polyphonic audio for music content processing. *INFORMS Journal on Computing*, 18(3):294–304, August 2006. 00103.

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

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

[7] 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.

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

[9] Hendrik Purwins, Benjamin Blankertz, and Klaus Obermayer. A new method for tracking modulations in tonal music in audio data format. In *IEEE-INNS-ENNS International Joint Conference on Neural Networks (IJCNN'00)*, volume 6, page 6270–6275, Como, 2000. 00039.

[10] 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.

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

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

[13] 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.

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