Audio Engineering Society
Convention Paper
Presented at the 116th Convention
2004 May 8–11 Berlin, Germany

This convention paper has been reproduced from the author's advance manuscript, without editing, corrections, or consideration by the Review Board. The AES takes no responsibility for the contents. Additional papers may be obtained by sending request and remittance to Audio Engineering Society, 60 East 42nd Street, New York, New York 10165-2520, USA; also see www.aes.org. All rights reserved. Reproduction of this paper, or any portion thereof, is not permitted without direct permission from the Journal of the Audio Engineering Society.

Scale Degree Profiles from Audio Investigated with Machine Learning

Hendrik Purwins[1], Benjamin Blankertz[2], Guido Dornhege[2], and Klaus Obermayer[1]

[1]*Berlin University of Technology, Berlin, Franklinstraße 28/29, 10 587, Germany*

[2]*Fraunhofer FIRST (IDA), Berlin,Kekuléstraße 7, 12 489, Germany*

Correspondence should be addressed to Hendrik Purwins (`hendrik@cs.tu-berlin.de`)

ABSTRACT

In this paper we introduce and explore a method for extracting low dimensional features from digitized recordings of music performance: The so called constant Q scale degree profiles are 12-dimensional vectors that reflect the prominence of the 12 scale degrees in a section of a piece of music they are extracted from. Here we study the type and amount of information that is captured in those profiles when calculated from whole short pieces of piano music. The analyzed data encompass sets of preludes and fugues by Bach (WTC), Chopin (op. 28), Alkan (op. 31), Scriabin (op. 11), Shostakovich (op. 34), and Hindemith (Ludus Tonalis). In a supervised approach we investigated the ability of classifiers to recognize composers from profiles. As unsupervised methods we performed (1) a cluster analysis which resulted in one major and one minor cluster and indicated major/minor ambiguity and how clearly composers separate between major and minor, and (2) a visualization technique called Isomap which reveals in its 2-dimensional representation the degree of chromaticism of pieces apart from the major–minor duality. In summary it is astonishing how much information on a music piece is contained in the 12-dimensional profiles that can be calculated in a straight-forward manner from any digitized music recording.

1. INTRODUCTION

Most music is based on defined musical scales. Scales form a limited reservoir of pitches that serves like an alphabet of the musical language of that piece. The notion of pitch class abstracts from the octave position of a tone. Pitches that are related to each other by intervals of integer multiples of an octave are treated as one equivalence class. But it is not always appropriate to consider pitch classes instead of pitches: E. g. the pitches used by the pygmies of Aka in a single piece do not repeat in other octaves. Also in harmony of Western European music and Jazz the octave position of tones can only be exchanged according to specified rules. The selected pitch class material has a great impact on the style and character of a musical piece, e. g. the pitch class material of the major, minor, whole tone, or blues scale. Instead of giving binary information on whether a pitch class is contained in a piece of music or not, we can count the occurrences of pitch classes yielding a profile of pitch class frequency. Despite the complex interrelation between major/minor tonality and various musical features like voice leading, form, beat strength, musical rhetorical, gestalt, the mere frequencies of occurrences of pitch classes are a major cue for the percept of a tonality [1]. Entirely ignoring the temporal structure of a musical piece, in this reductionist view a major/minor tonality is a prominence profile of pitch classes. E. g. for C-major the tone c is most important. The other tones g, e of the tonic triad follow. The non-diatonic notes are least important. Provided one tone is emphasized as a reference, like the tonic in a major/minor tonality, we can view all other pitch classes relative to the tonic. The nuances of key character [2] disappeared slowly during the decades after the introduction of equal temperament (∼1800). Neglecting key character we will achieve scale degree profiles by transposing all pitch class frequency profiles to the same tonic. From a digitized recording it is possible to calculate pitch class intensities accumulated across the piece. Due to the constant Q transform used to generate them, they are called constant Q profiles. Although noisy they are closely related both to the profiles of frequencies of occurrence and of prominence [3]. The advantage of constant Q profiles compared to profiles of frequencies of pitch class occurrences is twofold: Constant Q profiles can be calculated when the played notes are not known and the piece cannot be transcribed. The constant Q profile is more similar to the sound representation in the human ear. From constant Q profiles, constant Q scale degree profiles can by generated by transposition. Constant Q profiles can be used for classification of keys and tone centers by correlation, Euclidean or other distances [3]. Inter key relations have been visualized by using a Self Organizing Feature Map [4, 5] or Correspondence Analysis [6, 7] in combination with averaged constant Q profiles. A toroidal arrangement of the keys emerged displaying the circle of fifths and resembling the configuration achieved with Multi Dimensional Scaling from pitch class prominence profiles [1].

The aim of this paper is to investigate, what the constant Q scale degree profiles can tell us about the stylistic relations between pieces. How does a composer discriminate between major and minor? Can the composer be determined based on the noisy and very compressed information contained in a constant Q scale degree profile? To what extent can a constant Q scale degree profile reveal the musical content of a piece?

Statistical analysis of inter style relations based on profiles of pitch class frequencies of occurrences had been done by [8] and [9], the latter using Principal Component Analysis, Linear Discriminant Analysis, and Hierarchical Clustering.

In Section 2.1 constant Q scale degree profiles are derived. In Section 2.3 the theoretical background for classification is outlined. Then k-means clustering (Section 2.6) and Isomap as a visualization tool (Section 2.7) are presented. The musical corpus is described in Section 2.8. By means of t-test and a sparse Linear Programming Machine, Section 3.2 derives how composers use pitch class intensities to discriminate between major and minor. Section 3.3 reveals how the mode of a piece is the most salient property emerging in the data using k-means clustering. Their position relative to the cluster centers reveals their 'majorness'. In some cases major/minor mode ambiguity and the modulation range is indicated. By the latter we mean the range of tone centers on the circle of fifths that are visited during the piece. In Section 3.4, from visualization by Isomap it can be seen how certain pieces distinguish from others due to their chromaticism, ambiguous or extended tonality.

2. METHODS

In this section we will introduce scale degree profiles. The we will discuss how performance of supervised learning can be measured (Section 2.3). Two classifiers will be presented: Regularized Discriminant Analysis and Support Vector Machines. K-Means Clustering and a visualization tool will be shown. Then the corpus of musical data will be covered.

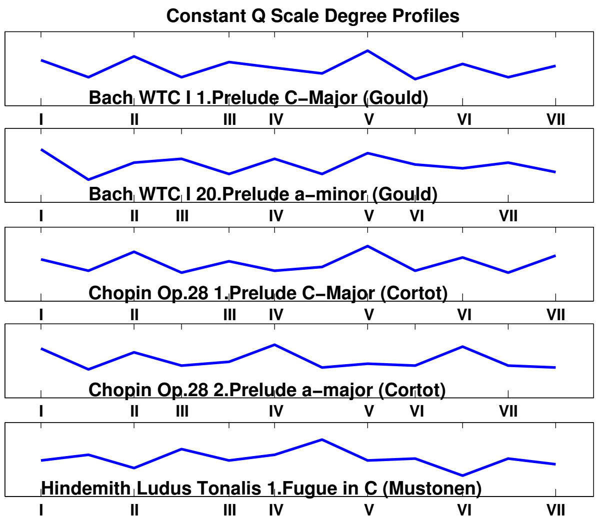

Fig. 1: Constant Q scale degree profiles for selected pieces from Bach, Chopin and Hindemith. Scale degrees are shown on the horizontal axis. In Chopin and Bach the peaks are related to the diatonic scale and to psychological probe tone ratings (Cf. Section 2.1).

2.1. Scale Degree Profiles

We focus our interest on the use of the 12 pitch classes in music. Instead of dealing with automatic transcription, we would like to generate a pitch class representation from audio directly. A short-time constant Q profile is calculated from some audio sequence. Successive profiles are summarized in a long-term constant Q profile. As underlying DSP method, serves the constant Q transform [10]. The later is based on a filter bank like FFT. But the center frequencies of the constant Q transform are equally spaced in the logarithmic frequency domain. Therefore the constant Q transform is well suited to extract pitch and pitch classes, since pitch perception works logarithmically as well. (Cf. [3] for technical details.)

Constant Q scale degree profiles are derived from constant Q profiles by transposition to the same reference key, that means the first entry in the profile corresponds to the keynote and the last entry corresponds to the major seventh.

2.2. Supervised versus Unsupervised Learning

Supervised learning can be used to classify data according to certain labels (e.g. key, mode, composer, performer). Unsupervised learning lets salient structural features emerge without requiring any assumption or having a specific question like a classification task according to predefined categories. While the performance of supervised methods can be quantified by – more or less – objective measures (cf. next section) the evaluation of unsupervised analyses is more delicate due to the exploratory nature and the lack of a specific goal that was fixed beforehand.

2.3. Evaluating Classification Algorithms

The concept of cross-validation and the Receiver Operating Characteristics are discussed.

2.3.1. Cross-Validation

Machine learning algorithms for classification work in two steps. First the learning algorithm is fed with some labeled data ('training set') from which regularities are extracted. After that the algorithm can classify new, previously unseen data. In order to validate how well a classification algorithm learns to generalize from given data, e.g., a technique called k-fold cross-validation is applied: the data set is randomly split into a partition of k equally sized subsets. Then the classifier is trained on the data of $k-1$ sets and evaluated on the hold-out set ('test set'). This procedure is done until every of the k sets was used as test set, and all k error ratios on those test sets are averaged to get an estimation of the 'generalization error', i.e., the error which the classification algorithm is expected to make when generalizing from given training data on which the classifier was trained to new data. Of course this quantity greatly depends on the structure and complexity of the data and the size of the training set. To get a more reliable estimate one can also do n-many partitions of the data set and do k-fold cross-validation of each partitioning ('n times k-fold cross-validation').

Technical note: With our music corpus some care has to be taken when doing cross-validation in order to avoid underestimating the generalization error. Since some pieces exist in versions from various performers all pieces are grouped in equivalence classes, each holding all versions of one specific piece. But most pieces exist only in one interpretation and form a singleton equivalence class. In cross-validation all pieces of one equivalence class are either assigned to the training or to the test set.

2.3.2. Receiver Operating Characteristic

Receiver Operating Characteristic (ROC) analysis is a framework to evaluate the performance of classification algorithms independent of class priors (relative frequency of samples in each class), cf. [11]. In a ROC curve the false positive rate of a classifier is plotted on the x-axis against the true positive rate on the y-axis by adjusting the classifier's threshold. In our plot all ROC curves from n times k-fold cross-validation procedure were averaged. To subsume the classification performance in one value the area under the curve (AUC) is calculated. A perfect classifier would attain an AUC of 1, while guessing has an expected AUC of 0.5, since false positive rate equals true positive rate in the latter case.

2.4. Supervised Learning

Two classifiers have proved successful in Section 3.1 for determining the composer based on constant Q scale degree profiles.

2.4.1. Regularized Discriminant Analysis

The Quadratic Discriminant Analysis (QDA) is a classification method which assumes that each class $\mathcal{C}_i$ has a Gaussian distribution $\mathcal{N}(\mu_i, \Sigma_i)$ with mean μ_i and covariance matrix Σ_i. Under this assumption with known parameters μ_i, Σ_i it is possible to derive a classification rule which has the minimum misclassification risk. The regions of the classes in input space are separated by a quadratic function. A related but simpler classifier is the Linear Discriminant Analysis (LDA) which makes the further assumption that the covariance matrices of all classes are equal ($\Sigma = \Sigma_i$ for all i). In this case the rule that minimizes the misclassification risk leads to a linear separation. Whether it is better to take QDA or LDA depends on the structure of the data. The covariance matrix of a Gaussian distribution describes in what way individual samples deviate from the mean. In classification problems where this deviation is class independent LDA should be prefered, otherwise QDA is based on the more appropriate model. So far the theory. Apart from that in real-world problems, one is faced with additional issues, even when we suppose that the Gaussian assumption is valid. The true distribution parameters μ_i and Σ_i are not known and thus have to be estimated ($\hat{\mu_i}$, $\hat{\Sigma}_i$) from given training data. If the number of training samples is small compared to the dimensionality d of the data p this estimation is prone to error and degrades the classification performance. This has two consequences. Even when the true covariance matrices are not equal, LDA might give better results than QDA because for LDA less parameters have to be estimated and it is less sensitive to violations of the basic assumptions. We modify the estimated covariance matrices according to $\hat{\Sigma}_i \mapsto (1-\lambda)\hat{\Sigma}_i + \frac{\lambda}{\sum|\mathcal{C}_j|}\sum|\mathcal{C}_j|\hat{\Sigma}_j$, with $|\mathcal{C}_i|$ being the class sample size. Then one can mediate between QDA (for $\lambda = 0$) and LDA (for $\lambda = 1$). This strategy is called regularization. On the other hand the estimation of covariance matrices from too little samples holds an inherent bias making the ellipsoid that is described by the matrix deviating too much from a sphere: Large eigenvalues are estimated too large and small eigenvalues are estimated too small. To counterbalance this bias a so called shrinkage of the covariance matrices towards the identity matrix I is introduced, $\hat{\Sigma}_i \mapsto (1-\gamma)\hat{\Sigma}_i + \gamma I \cdot \mathrm{trace}(\hat{\Sigma}_i)/d$. Of course regularization and shrinkage can also be combined which gives Regularized Discriminant Analysis (RDA), cf. [12], while LDA with shrinkage is called RLDA. The choice of parameters λ and γ is made in a model selection, e.g., by choosing that pair of parameters that results in the minimum cross-validation error on the training set.

2.5. Support Vector Machines

Support Vector Machines (SVMs)[13] are a popular classification tool. The method is based on the idea to use large margins of hyperplanes to separate the data space into several classes. Also non-linear functions, e.g. radial basis functions, can be used to obtain more complex separations by applying the kernel trick, cf. the review [14].

2.6. K-Means Clustering

Cluster analysis [15] is a technique of exploratory data analysis that organizes data as groups (clusters) of individual samples. The k-means clustering method is done by the expectation maximation (EM) technique in the following way: The data points are randomly separated into k clusters. First the mean of each cluster is calculated (E-step) and then each point is re-assigned to the mean to which it has the least Euclidean distance (M-step). The E- and the M-step are iterated until convergence, i.e., the M-step does not change the points-to-cluster assignment. It can be proven that this algorithm always converges, but runs with different initial random assignments could lead to different final configurations. In our experiment in Section 3.3 repetitions led to the same result.

To evaluate the outcome of the clustering procedure we introduce the following class membership function $m_i(p)$, with data point p. We scale down k-means clustering with the $k = 2$ class centers c_1, c_2 of the data points:

$$m_1(p) = \frac{\|p - c_2\|}{\|p - c_1\| + \|p - c_2\|} \quad (1)$$

A value near 1 is assigned to data points p close to the center c_1 of cluster 1 by function m_1. Points that ambiguously lie between the clusters get values near 0.5 in both functions, m_1 and m_2.

2.7. Visualization by Isomap

Often high dimensional data can be approximately described by a curved manifold of a lower dimension. For Isomap [16] data points in high dimensional space are given. Each of them is connected to its k nearest neighbors [17], with respect to a given dissimilarity, e. g. Euclidean distance. The idea is that from one point another point cannot be reached directly, only by taking a route via data points that are connected to each other. The graph distance assigned to a pair of data points is the length of the shortest path via connected points. Multi Dimensional Scaling [18] is performed on the graph distance matrix, so that projecting the points onto the eigenvectors with highest eigenvalues shows the configuration of points in a low dimensional Euclidean space that optimally preserves the graph distances. E. g. a two dimensional manifold hidden in the high dimensional data is visualized by Isomap as a planar display that appropriately reflects the inter point distances on the surface of the manifold.

2.8. Musical Corpus

As musical material we choose Bach's Well-Tempered Clavier and various sets of preludes and fugues that encompass a piece in every key. The 226 constant Q scale degree profiles under investigation are: (1) The preludes of WTC, part I and II, and the fugues of part II by Bach (1685-1750) are interpreted by Glenn Gould. Another interpretation of the preludes of part II is given by Samuel Feinberg. There are altogether 96 Bach profiles. (2) The 'Préludes' op. 28 by Chopin (1810-1849) are played by Alfred Cortot and Ivo Pogorelich, altogether 48 profiles. (3) The 25 'Préludes' op.31 by Alkan (1813-1888) are recorded from Olli Mustonen. (4) We use 21 profiles from the 'Preludes' op.11 by Scriabin (1872-1915), played by various pianists. (5) Mustonen played the 24 'Preludes' op.34 by Shostakovich (1906-1975). (6) 'Ludus Tonalis' by Hindemith (1895-1963) contains one fugue and one interludium for each pitch class as a tonic, but neither major nor minor. The 12 fugues are part of the corpus, presented by Mustonen.

It is an interesting question, what other references to Bach the latter pieces contain, apart from the set-up as a set of preludes and fugues in all keys.

3. RESULTS

Classification, clustering, and visualization can be performed on the basis of audio, in the form of constant Q scale degree profiles.

	LDA	RLDA	RDA	SVMrbf
Bach	0.79	0.79	**0.95**	**0.95**
Chopin	0.52	0.52	0.64	**0.73**
Alkan	0.43	0.43	0.72	**0.76**
Scriabin	0.65	0.65	**0.72**	0.69
Shostakovich	0.81	0.85	**0.86**	**0.86**
Hindemith	0.93	0.93	**0.97**	0.95

Table 1: For the classification of one composer vs. the rest, the performance of various classifiers is evaluated by a measure called the area under the curve describing the Receiver Operating Characteristics (cf. Section 2.3.2). The best classifiers are emphasized in boldface. Regularized Discriminant Analysis (Section 2.4.1) and Support Vector Machines (Section 2.5) with radial basis functions as kernels perform equally well.

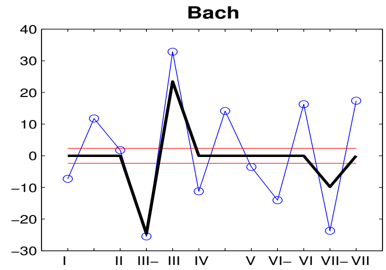

Fig. 2: Significance of scale degrees for mode discrimination. The horizontal axis denotes the scale degrees, '-' indicating the minor intervals. The vertical axis denotes significance. The horizontal red lines indicate the significance level for the t-test of error level 1%. If the blue line ('o') is above the upper or below the lower red line, the scale degree contributes significantly to major/minor discrimination. The black line indicates to what extent which scale degree is emphasized for discrimination in a sparse Linear Programming Machine (LPM). In Bach both t-test and LPM emphasize the thirds and the minor seventh. In addition t-test identifies every scale degree to be very significant, except the major second and the fifth.

3.1. Composer Classification

Composer classification works astonishingly well based on constant Q scale degree profiles with an appropriate classifier, especially considering that only very little and noisy data are provided. One Composer ist classified against all the rest. Table 1 shows the area under the curve of the Receiver Operating Characteristics. Instead of applying LDA, classification performance improves when using RDA, a compromise between LDA and QDA. An explanation could be (1) that the variance varies a lot between classes and (2) that some class sample sizes are quite small. RDA and SVM with radial basis functions are even for Bach and Shostakovich. RDA beats SVM slightly for Scriabin and Hindemith. SVM beats RDA for Chopin and Alkan.

3.2. Significance of Single Scale Degrees for Mode Separation

What does distinguish major and minor? To identify the significance of different scale degrees for mode separation we apply two methods: (1) two sided t-test with error level 1%, (2) Linear Programming Machine (LPM). The latter method is a classification algorithm that is similar to the SVM, but has a linear goal function. A special feature of the LPM is that it seeks to produce sparse projection vectors, i.e., it tries to find a good classification that uses as little feature components (in our case scale degrees) as possible. The components that are chosen by the LPM are thus most important for the discrimination task.

According to the t-test, in Bach (Figure 2) see the outstanding significance of both thirds and the minor seventh. All other scale degrees are equally significant, except the (almost) unsignificant major second and the fifth. In Chopin (Figure 3) the significance of these scale degrees is much less. In the t-test only thirds, sixths, and seventh are significant. The minor sixth is close to insignificance. LPM emphasizes the major seventh. In Shostakovich (Figure 4) only the thirds are significant according to the t-test. In Alkan thirds are significant. The minor seventh is slightly significant. For Scriabin thirds have high significance. But also sixths and sevenths have some significance.

As expected the thirds and to some degree the sixths and sevenths are significant for major/minor discrimination. During music history

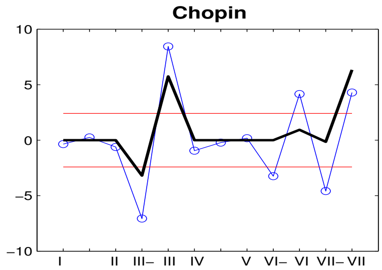

Fig. 3: Chopin: The t-test identifies only thirds, sixths, and sevenths to be significant for mode discrimination. LPM emphasizes thirds and the major seventh. (Cf. Figure 2 for an explanation of the curves.)

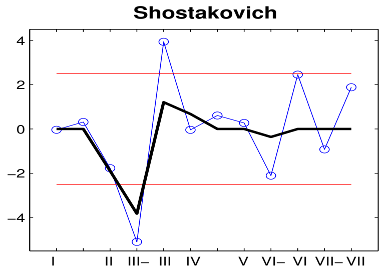

Fig. 4: According to the t-test, in Shostakovich only the thirds allow for mode discrimination. (Cf. Figure 2 for an explanation of the curves.)

from Bach to Shostakovich the significance of all pitch classes gets eroded due to increasing chromaticism and the continuing extension of tonality.

3.3. Clustering According to Mode

K-means clustering is performed on the corpus excluding Hindemith and Shostakovich, since they are the least tonal. Using Equation 1 as a major membership function yields clustering into a major (left) and minor (right) cluster (Figure 5). The results indicate a degree of 'majorness' of the keys. Pieces which lie on the borderline between major and minor may not be very typical representatives of that key. Bach's pieces concentrate in a smaller region. Bach clearly separates with a wide margin. This is due to the fact that in general Bach modulates only to tone centers that lie closely around the tonic on the circle of fifths. Usually he modulates in the range of a fifth lower and three fifths upper. Only the chromatic a–minor prelude of WTC II is a little off the Bach cluster. This will be discussed in Section 3.4. Chopin's a-minor prelude is (for Pogorelich and Cortot) almost on the borderline. This is related to the wired harmonical content of the piece, that makes it ambiguous in mode.

It helps to musically analyze pieces that are found in the 'wrong' cluster. There are three minor constant Q scale degree profiles ($f, b, a\flat$ by Alkan) on the side of the major cluster and six major constant Q scale degree profiles (Chopin's $B\flat$ played by Cortot and Pogorelich, $D\flat$, $A\flat$, and Alkan's $D\flat, A\flat$) within the minor cluster. For some of these pieces there is an intuitive musical explanation for their position in the other cluster. In Chopin's $B\flat$–major prelude the middle part in $G\flat$–major emphasizes the minor third and deemphasizes the leading note in $B\flat$–major. In addition the left hand figuration introduces chromaticism that later is taken over by the right hand also. Chopin's $A\flat$–major prelude is once found in the 'wrong' cluster (Pogorelich). In Cortot's interpretation it is the major piece second closest to the cluster border. The closeness to the minor cluster is especially due to modulations to E–major and therefore the emphasis of the minor third of $A\flat$–major. Also in the $A\flat$–major parts the minor seventh chord on $A\flat$ is frequently used. Alkan's $A\flat$–major prelude is chromatic. Within the length of one measure, all four flats alternate with notes a half tone above them, including enharmonic $c\flat/b\flat$. Therefore the enharmonic equivalents of the major and minor third are heard all the time. Alkan's $D\flat$–major prelude has a middle part in enharmonic $d\flat$–minor. Alkan's f-minor prelude has an F–major middle part. Alkan's $a\flat$–minor has a middle part in enharmonic $A\flat$–major.

Major/minor mode is the most salient property in the constant Q scale degree profiles. K-means clustering with the major membership function reveals the degree of 'majorness'. For most pieces around the borderline between clusters chromaticism or extended tonality holds. Mode separability of different composers is reflected in the clustering.

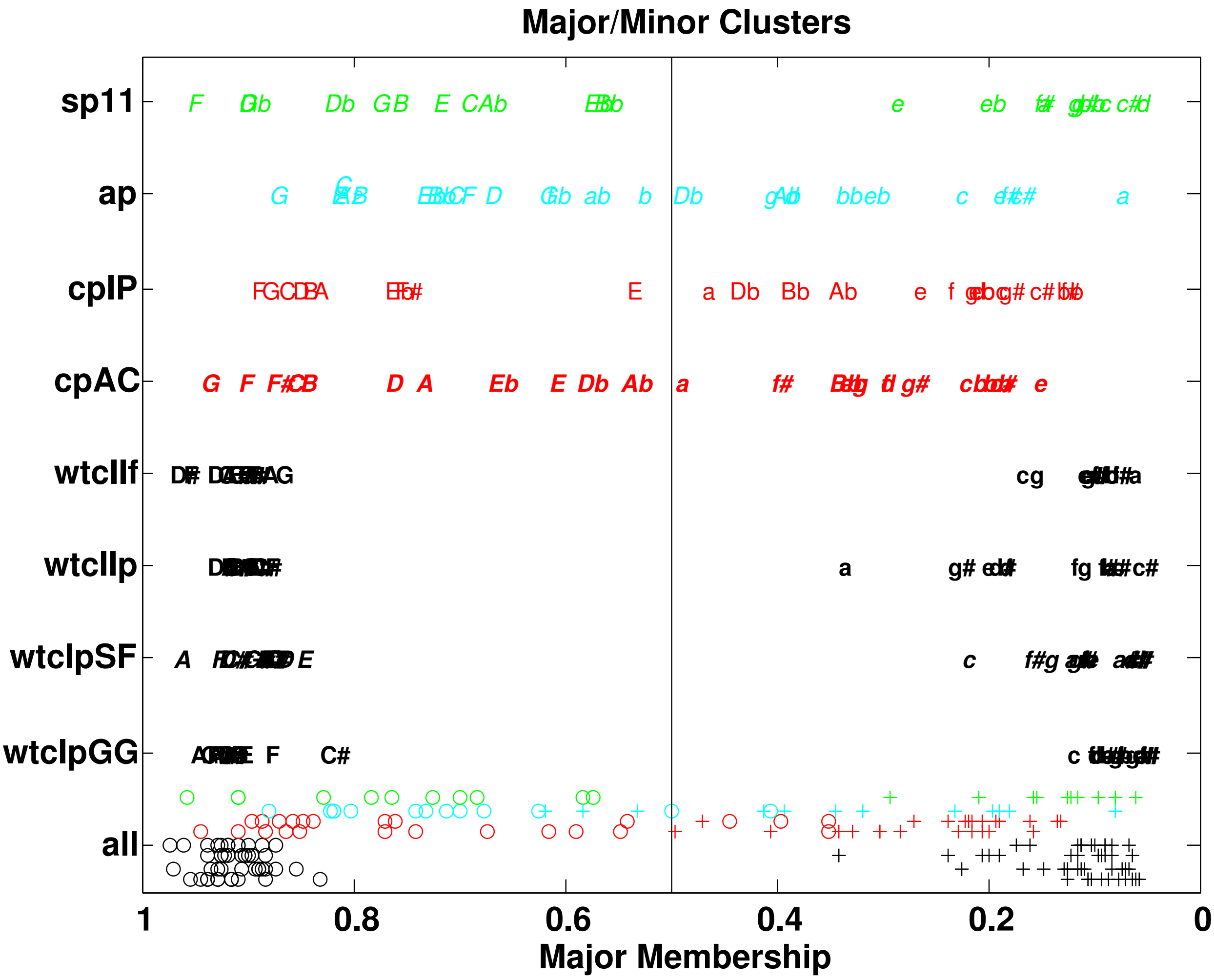

Fig. 5: 'Majorness' in k-means clustering ($k = 2$) of constant Q scale degree profiles. The vertical axis shows the different groups of pieces (from bottom to top): a summary of all pieces, then Bach's WTC (preludes-'p' and fugues-'f', black) performed by G. Gould ('GG'), and S. Feinberg ('SF', only WTC I preludes), Chopin's preludes ('cp', red), performed by A. Cortot ('AC') and I. Pogorelich ('IP'), Alkan's preludes ('ap', cyan), and Scriabin's preludes ('sp11', green). At the bottom of the graph, the pieces are shown again without labels, 'o' indicating major, '+' indicating minor, with the typical colors assigned to each composer. The horizontal axis indicates the mode membership for major (1 meaning typical major, 0 minor, 0.5 ambiguous). The vertical line in the middle at membership 0.5 splits the profiles into a left major and a right minor cluster. Bach and Scriabin clearly separate. In Chopin and Alkan more major/minor ambiguous pieces can be found, e.g. Chopin's a–minor prelude. (cf. 3.3)

3.4. Visualizing Inter Relations of Modes, Composers, and Pieces

How can we measure similarity among pieces? For each pair of pieces let us calculate the Euclidean distances between their constant Q scale degree profiles. A dissimilarity measure is then given by the shortest path length via pieces that have smallest Euclidean distance to their prede-

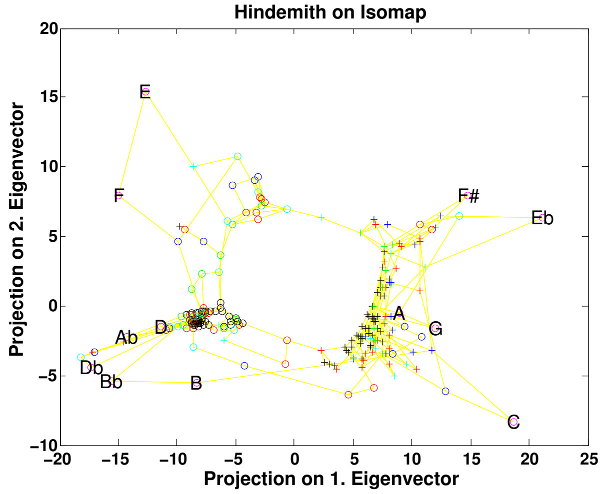

Fig. 6: 'Landscape' of all pieces. A two-dimensional projection of the Isomap with k-nearest neighbor topology ($k = 2$), trained with the entire corpus: Bach (black), Chopin (red), Alkan (cyan), Scriabin (green), Shostakovich (blue), and Hindemith (magenta). As in Figure 5, 'o'/'+' denote major/minor respectively. The two-dimensional projection accounts for explaining 86.2 % of the variance in the data. In this graph the projected pieces by Hindemith are labeled with their keynote. Hindemith occupies the outlier positions in the Isomap projection. Figures 7,8 9,10, and 11 are all the same curve, but with labels that belong to another composer.

cessor. We would like to analyze the net structure that is build up by connecting the constant Q scale degree profile of each piece to its two closest neighbors. We use the Isomap (Section 2.7) with k=2 nearest neighbor for visualization (Figures 6, 7,10,11,9). Under the assumption of the above given dissimilarity, Isomap reveals that the pieces form a circular structure with two 'poles' (major and minor) and some extensions to outliers. This two dimensional visualization accounts for 86.2 % of the variance in the data.

Compare the positions of the pieces in the visualization of Figure 6 to the clustering in Figure 5. On the Isomap Hindemith (Figure 6) occupies all outlier positions very far away from the center. Only A lies more closely to the center.

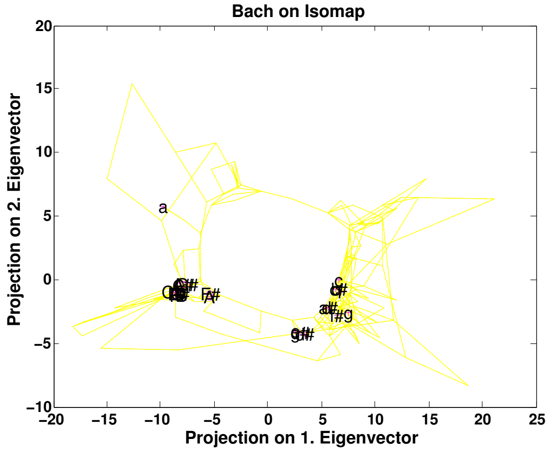

Fig. 7: Bach's preludes of WTC II are labeled. Two densely concentrated well separated clusters are formed: a highly compressed major cluster and a minor cluster in the form of a strip. The very chromatic a-minor prelude of WTC II is an outlier relative to these clusters. (Cf. Figure 6 for all Bach pieces labeled as black symbols.)

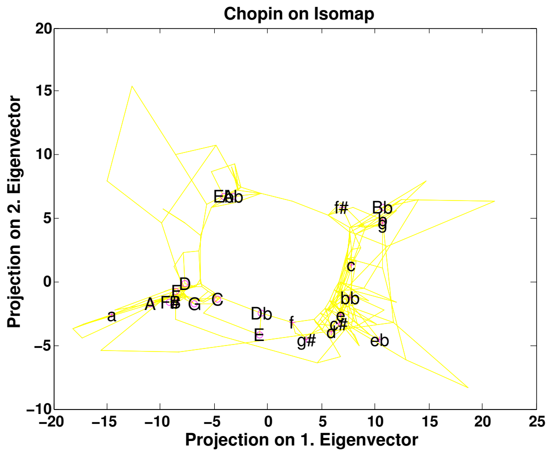

Fig. 8: Chopin's preludes op.28 in Cortot's interpretation are highlighted. They distribute around the circle. Major/minor are not clearly separated. See all Chopin preludes marked as red symbols in Figure 6.

Bach's major preludes cluster densely on a small spot while his minor preludes concentrate on a prolonged strip (Figure 7). As in k-means clustering of Figure 5, the margin between the two

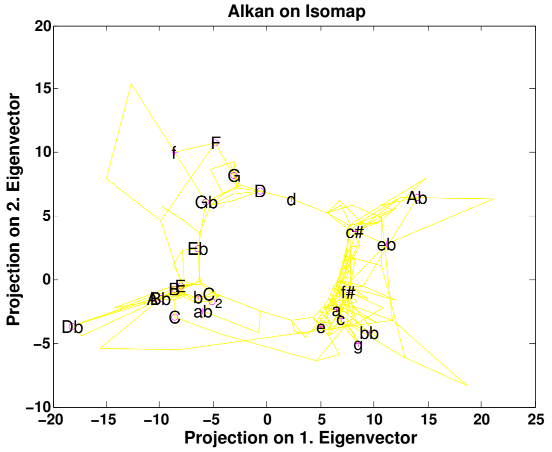

Fig. 9: Alkan concentrates somewhat on two wide clusters, but major and minor intermingle a little. $D\flat$–major and f–minor are outliers due to their modulation to the parallel of the tonic in the middle part, $A\flat$–major is a chromatic prelude. (Cf. Figure 6 for details.)

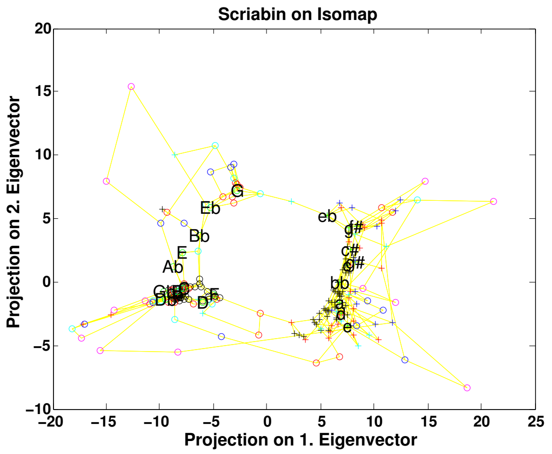

Fig. 10: Scriabin clearly separates into a major and a minor cluster. (Cf. Figure 6 for details.)

clusters is wide. It is not arbitrary that the a-minor prelude of WTC II is the only prelude outside the two main Bach clusters. This prelude also stands aside from the Bach minor cluster in Figure 5. The special position of this prelude is due to the chromaticism that dominates it. In Chopin (Figure 8) major and minor intermingle. In Alkan (Figure 9) a major and a minor cluster is found, but they are not clearly separated. Due to its chromaticism $A\flat$-major is very far in the minor region, and $f, a\flat$-minor are in the major region, because they modulate to the parallel key in their middle parts. $D\flat$–major is far off for the same reason. Arbitrarily it is next to Hindemith's fugue in $D\flat$. Scriabin (Figure 10) makes up two major/minor clusters clearly separated by a wide margin. Shostakovich (Figure 11) does not separate well between major/minor. The A-major prelude is off.

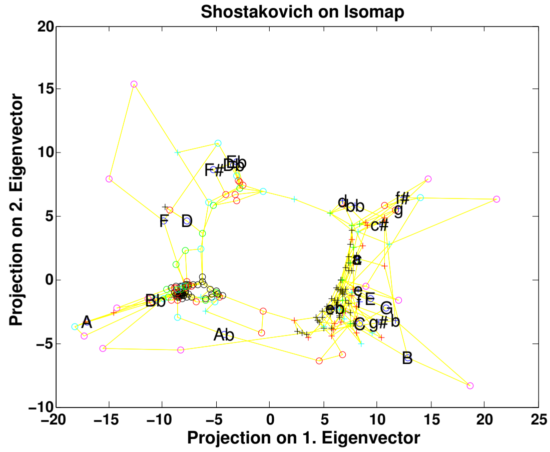

Fig. 11: Shostakovich is widely distributed with A-major as an outlier. (Cf. Figure 6 for details.)

The Isomap reveals stylistic characteristics of composers and pieces. Again distinguished positions on the 'landscape' often coinced with high chromaticism and a character unusual for that composer. Also stylistic coherence of a composer is visualized.

4. CONCLUSION AND FUTURE WORK

In this paper we introduced the constant Q scale degree profile. Audio in this compressed format had been subject to a wide range of machine learning methods. We successfully applied classification, clustering, and visualization to constant Q scale degree profiles for composer, mode, and style analysis.

The classification results show that constant Q scale degree profiles can substantially contribute to composer identification based on audio, provided the use of appropriate classifiers: RDA and

SVMrbf performed best among a group of classifiers.

K-means clustering of the constant Q scale degree profiles reveals that the 'majorness' is the most salient property in the data. Applying an appropriate mode membership function yields two clearly separated clusters: major and minor. The mode membership can be interpreted musically. For some of the pieces with 'ambiguous' membership to both the major and minor cluster center, musical analysis reveals an ambiguous major/minor content, chromaticism, or modulations to the parallel major/minor key or to tone centers far away from the tonic on the circle of fifths. The compactness of clusters can be musically interpreted in the way that the composer stays in a small compact neighborhood of tone centers around the tonic on the circle of fifths.

The Isomap visualization allows a stylistic comparison of pieces and the group of pieces by one composer. Some composers reside in the outlier positions (Hindemith) whereas other densely concentrate (Bach).

Considering that the constant Q scale degree profiles contain only a very special aspect in music—reducing high musical complexity to a small cue—it is remarkably good. It is promising to use this feature in combination with rhythmic and other features.

Based on an extended corpus of musical data, other questions of interest comprise the investigation to what extent certain features do manifest in constant Q scale degree profiles, like key character and performer. [19] describes key character as the property of the work of a some composers. A test design can be used to test the hypothesis, whether constant Q scale degree profiles signify the key character for that composer. It is also interesting whether the performer can be determined with a recording at hand.

The suggested machine learning methods are by no means restricted to the analysis of constant Q scale degree profiles for composer, mode, and style analysis. They could be promisingly used for instrument identification, beat weight analysis, harmonic analysis, and style investigation as well, just to name a few other possible applications.

5. REFERENCES

[1] C. Krumhansl. *Cognitive Foundations of Musical Pitch*. Oxford University Press, Oxford, 1990.

[2] Mark Lindley. A rudimentary approach to the history of the major and minor keys. Unpublished Manuscript, 2003.

[3] Hendrik Purwins, Benjamin Blankertz, and Klaus Obermayer. A new method for tracking modulations in tonal music in audio data format. In S.-I. Amari, C.L. Giles, M. Gori, and V. Piuri, editors, *International Joint Conference on Neural Networks*, volume 6, pages 270–275. IJCNN 2000, IEEE Computer Society, 2000.

[4] Teuvo Kohonen. Self-organized formation of topologically correct feature maps. *Biol. Cybern.*, 43:59–69, 1982.

[5] Hendrik Purwins, Benjamin Blankertz, and Klaus Obermayer. Computing auditory perception. *Organised Sound*, 5(3):159–171, 2000.

[6] Michael J. Greenacre. *Theory and Applications of Correspondence Analysis*. Academic Press, London, 1984.

[7] Hendrik Purwins, Thore Graepel, Benjamin Blankertz, and Klaus Obermayer. Correspondence analysis for visualizing interplay of pitch class, key, and composer. In Emilio Luis-Puebla, Guerino Mazzola, and Thomas Noll, editors, *Perspectives in Mathematical Music Theory*. Epos Verlag, Osnabrück, 2003.

[8] W. Fucks and J. Lauter. *Exaktwissenschaftliche Musikanalyse*. Westdeutscher Verlag, Köln, 1965.

[9] Jan Beran. *Statistics in Musicology*. Chapman & Hall/CRC, 2003.

[10] Judith Brown. Calculation of a constant Q spectral transform. *J. Acoust. Soc. Am.*, 89(1):425–434, 1991.

[11] Foster Provost, Tom Fawcett, and Ron Kohavi. The case against accuracy estimation for comparing induction algorithms. In *Proc. 15th International Conf. on Machine Learning*, pages 445–453. Morgan Kaufmann, San Francisco, CA, 1998.

[12] Jerome H. Friedman. Regularized discriminant analysis. *Journal of the American Statistical Association*, 84(405), 1989.

[13] Vladimir Vapnik. *Statistical Learning Theory*. Jon Wiley and Sons, New York, 1998.

[14] K.-R. Müller, S. Mika, G. Rätsch, K. Tsuda, and B. Schölkopf. An introduction to kernel-based learning algorithms. *IEEE Neural Networks*, 12(2):181–201, May 2001.

[15] Anil K. Jain and Richard C. Dubes. *Algorithms for Clustering Data*. Prentice Hall, 1988.

[16] J.B. Tenenbaum, V. de Silva, and J.C. Langford. A global geometric framework for nonlinear dimensionality reduction. *Science*, 290(5500):2319–2323, 2000.

[17] Brian D. Ripley. *Pattern Recognition and Neural Networks*. Cambridge University Press, 1996.

[18] R. N. Shepard. The analysis of proximities: Multidimensional scaling with an unknown distance function. I. *Psychometrika*, 1962.

[19] Wolfgang Auhagen. *Studien zur Tonartencharakteristik in theoretischen Schriften und Kompositionen vom späten 17. bis zum Beginn des 20. Jahrhunderts*. Peter Lang, Frankfurt a. M., 1983.