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

Scale Degree Profiles from Audio Investigated with Machine Learning Techniques

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

[1] *Berlin University of Technology, Berlin, Franklinstraße 28/29, 10587, Germany*

[2] *Fraunhofer FIRST (IDA), Berlin,Kekuléstraße 7, 12489, 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 respective analyzed part of music. 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 set includes pieces from Bach's Well-Tempered Clavier (WTC), part I and II, the sets of preludes that encompass a piece in every key by Chopin (op. 28), Alkan (op. 31), Scriabin (op. 11), Shostakovich (op. 34), and the fugues of Hindemith's 'ludus tonalis' (one fugue for each pitch class, neither major nor minor). For the purpose of investigation we employ supervised and unsupervised machine learning techniques. 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 (2) a visualization technique called Isomap which reveals in its 2-dimensional representation some additional structure 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

A probe tone rating [8] is a psychological profile that portrays prominence of pitch classes in a certain context. A constant Q Profile is a pitch class profile derived from audio, and closely related to probe tone rating [11]. Transposed to the tonic we achieve the constant Q scale degree profile. It describes the strengths of different scale degrees depending on mode, key, style, instrument, and composer.

We may use the knowledge which constant Q profile stems from which composer and which piece. This information can serve as a 'teacher' in supervised learning, e.g. the Support Vector Machine [14].

In the 12-dimensional constant Q scale degree profiles, we can look for clusters or for projections on lower dimensions, by means of the Self-Organizing Feature Map [7], or Correspondence Analysis [5].

By using the constant Q profile technique as a simple auditory model in combination with the Self-Organizing Feature Map, an arrangement of keys emerges that resembles results from psychological experiments [8], and from music theory [3; 12]. But in this work we want to reveal the potential of profiles that are all transposed to the same key note.

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 unfolded.

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 [2]. 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

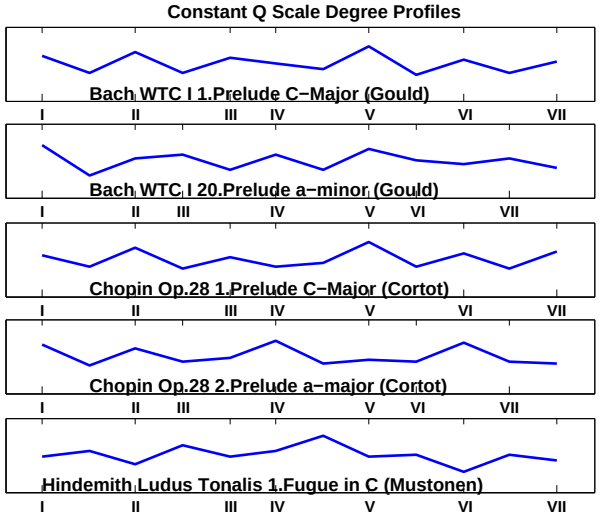

Fig. 1: Constant Q scale degree profiles for selected pieces from Bach, Chopin and Hindemith. Scale degrees are shown on the horizontal axis. In Chopin and Bach the peaks are related to the diatonic scale and to psychological probe tone ratings (Cf. Section 2.1).

to extract pitch and pitch classes, since pitch perception works logarithmically as well. (Cf. [11] for technical details.)

A constant Q scale degree profile is derived from a constant Q profile by transposing it to the tonic, 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

We will discuss the concept of cross-validation and the Receiver Operator Characteristics.

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 the 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 pieces (most pieces exist only in one interpretation and form a singleton equivalence class). In the 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. [10]. 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).

2.4. Supervised Learning

We will introduce two Classifiers, that will proof 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 $\mathcal{C}_i$ is Gaussian distributed $\mathcal{N}(\mu_i, \Sigma_i)$ 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 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 one is in real-world problems 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. By modifying the estimated covariance matrices according to $\hat{\Sigma_i} \mapsto (1-\lambda)\hat{\Sigma_i} + \lambda \sum \hat{\Sigma_i}$, 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 eigenval-

ues 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. [4], 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)[14] 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 [9].

2.6. K-Means Clustering

Cluster analysis [6] 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 repetitions led to the same result.

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

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

The function m_1 assigns data points that are close to the center of cluster 1 value near 1, while point that ambiguously lie between the clusters get values near 0.5 in both functions, m_1 and m_2.

2.7. Isomap for Visualization

For Isomap [13] pairwise dissimilarities of items are provided, e.g. Euclidean distances. The idea is that from one item another item can not reached directly but by taking a route passing by other items that lie in a chain, so that every item on the route lies in the $\epsilon-$ or k-nearest neighborhood of its preceding item. The dissimilarity then assigned to a pair of items is the closest route on a chain in the described manner. The dissimilarity matrix is transformed so if fulfills the triangular equation and then subject to multidimensional scaling. Finally the items can be projected on the most prominent Eigenvectors of the multidimensional scaling.

2.8. Musical Corpus

For musical material we choose Bach's Well-Tempered Clavier (WTC) and various compositions that were modeled after this cycle of pieces. The musical data under investigation consists of the following corpus of 226 constant Q scale degree profiles: (1) Bach's WTC I & II Fugues & Preludes -'wtc' (Gould-'GG', Feinberg-'SF', 96 profiles), (2) Chopin's Preludes -'cp' (Cortot-'AC' , Pogorelich- 'IP', 28 profiles), (3) Alkan's Preludes -'ap' (Mustonen-'OM' , 25 profiles), (4) Scriabin Preludes op.11- 'sp11' (various players-'nn', 21 profiles), (5) Shostakovich Preludes op.34 -'sp' (Mustonen, 24 profiles), and (6) Hindemith's Fugues from 'Ludus Tonalis'-'lt' (Mustonen, 12 profiles).

3. RESULTS

Classification, clustering, and visualization is now performed on the basis of audio represented as constant Q scale degree profiles.

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 Operator Characteristics. 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

	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 Operator 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.

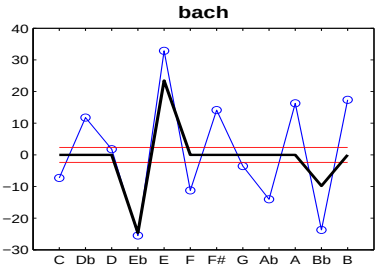

Fig. 2: Significance of scale degrees for mode discrimination. The horizontal axis denotes the scale degree for the profiles transposed to c. The vertical axis denotes significance. The horizontal red lines indicate the significance level for the t-test of error level 1%. If the blue line ('o') is above 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.

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 we clearly see the significance of the both thirds. But all other scale degrees are significant as well, except the major second and the fifth. In Chopin the significance of these scale degrees is much less. In the t-test only five scale degrees are significant, some of them close to insignificance. In Shostakovich 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 seventh have some significance.

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 mode 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 – especially the minor ones – concentrate in a smaller region. Bach clearly separates with a wide margin, leaving only the chromatic a–minor prelude of WTC II somewhat 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 (especially Chopin's $D\flat, B\flat, A\flat$, 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 modulation to $G\flat$–major includes the $d\flat$, which is the small third of $B\flat$. Alkan's $A\flat$–major prelude contains mostly mi-

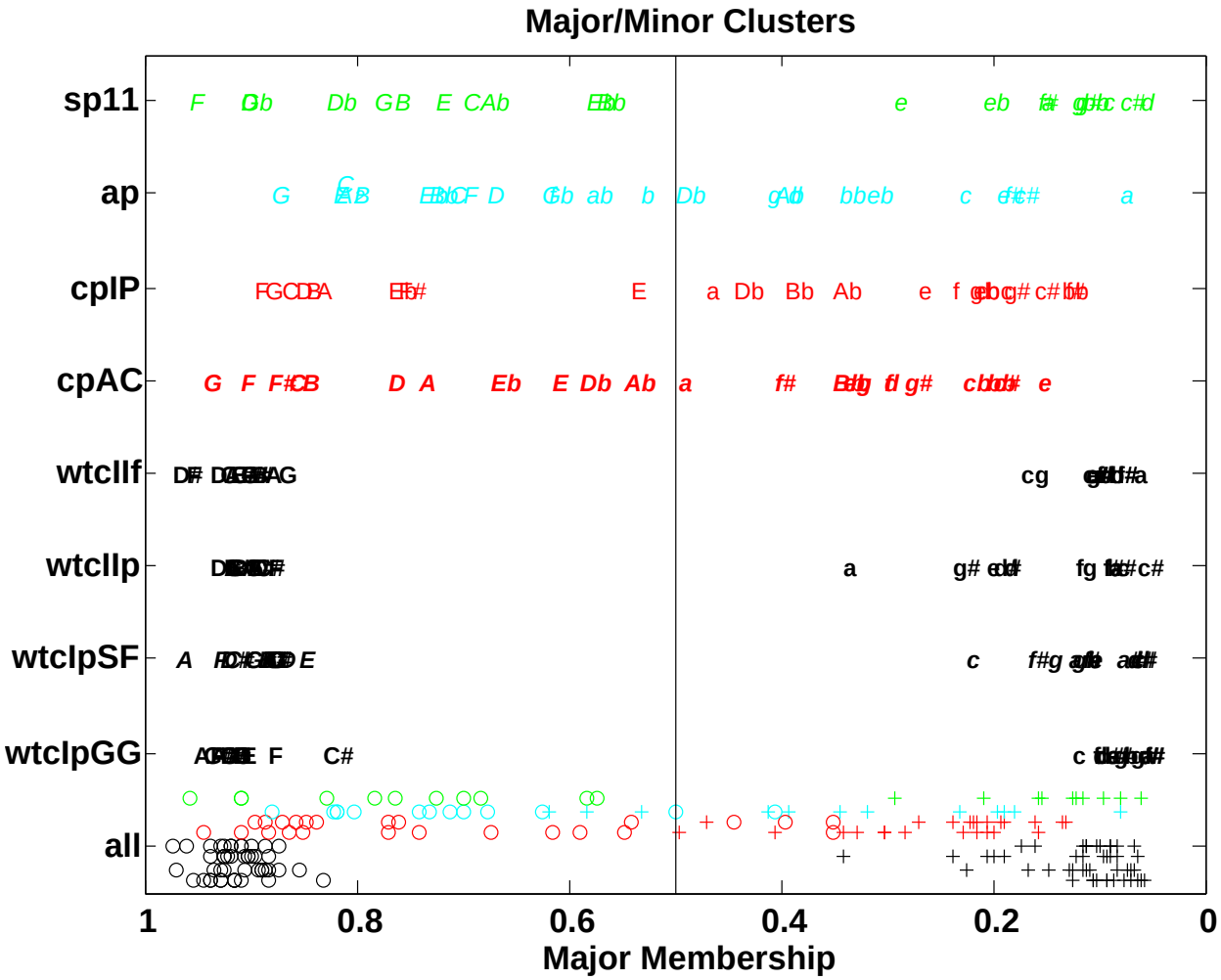

Fig. 5: 'Majorness' in k-means clustering ($k = 2$) of constant Q scale degree profiles. At the bottom of the graph, the pieces are shown again without labels, 'o' indicating major, '+' indicating minor, with a typical colors assigned to each composer. The vertical axis shows the different groups of 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). 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)

nor passages. Alkan's b-minor prelude has major middle part that modulates in major keys. Alkan's ab–minor prelude has a major middle part and an ongoing low frequency cluster-like figure. Alkan's f minor prelude uses unusual scales and has a major middle part.

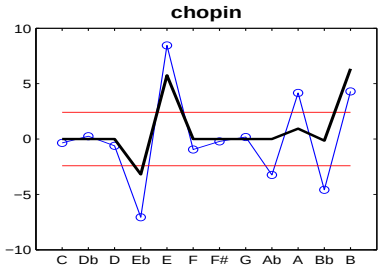

Fig. 3: Chopin: The t-test identifies only thirds, sixths, and sevenths to be significant for mode discrimination. LPM emphasizes thirds and the major seventh. (Cf. Figure 2 for an explanation of the curves)

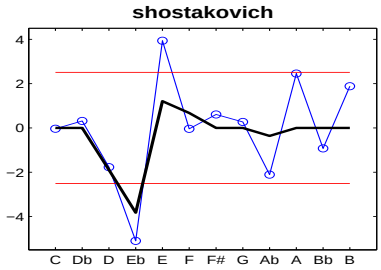

Fig. 4: According to the t-test, in Shostakovich only the thirds allow for mode discrimination. (Cf. Figure 2 for an explanation of the curves.)

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

We use the Isomap with k=2 nearest neighbor for visualization (Figures 7,6,10,11,9). The accumulated residual curves for the Isomap show that a two-dimensional projection accounts for explaining 35 % of the variance in the data.

Hindemith (Figure 6) occupies all outlier positions very far away from the center. Only A–major lies more closely to the center. Bach's major preludes cluster densely on a small spot while his minor preludes concentrate on a prolonged strip (Figure 7). As in Figure 5, the margin between the two clusters is wide. It is not arbitrarily 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 chromatism that dominates it. Shostakovich (Figure 11)

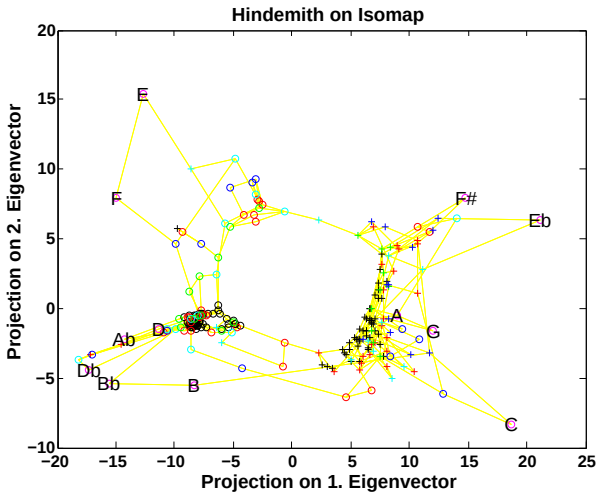

Fig. 6: Landscape of all pieces. A two-dimensional projection of the Isomap with k-nearest neighbor topology ($k = 2$), trained with the entire corpus: Bach (black), Chopin (red), Alkan (cyan), Scriabin (green), Shostakovich (blue), and Hindemith (magenta). As in Figure 5, 'o'/'+' denote major/minor respectively. The two-dimensional projection accounts for 35 % of 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.

does not separate well between major/minor. The A-major Prelude is off. Scriabin (Figure 10) makes up two major/minor clusters clearly separated by a wide margin. In Alkan (Figure 9) $G\sharp$-major is very far in in the minor region, and $f, g\sharp$-minor are in the major region. In Chopin (Figure 8) major and minor intermingle.

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

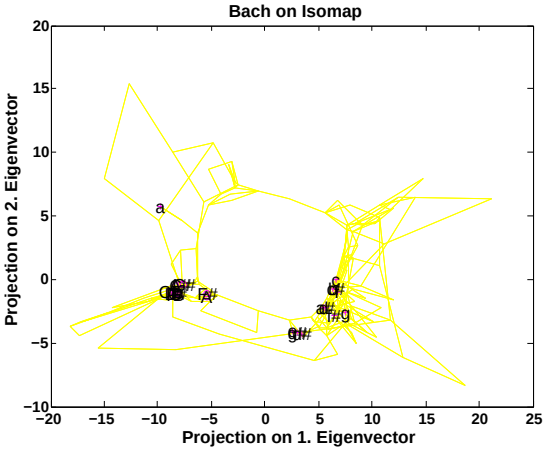

Fig. 7: Bach's preludes of WTC II are labeled. Two densely concentrated well separated clusters are formed: a highly compressed major cluster and a minor cluster in the form of a strip. The very chromatic a-minor prelude of WTC II is an outlier relative to these clusters. (Cf. Figure 6 for all Bach pieces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

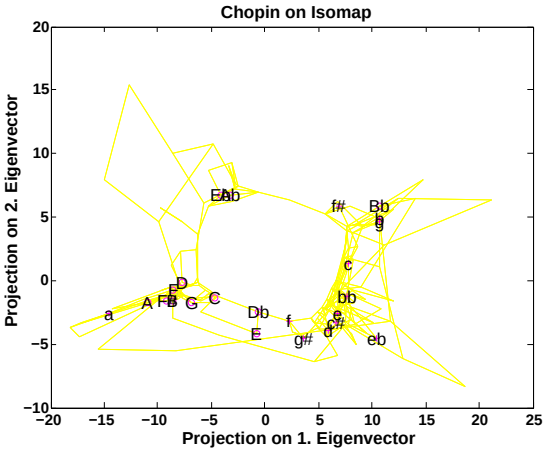

Fig. 8: Chopin's Preludes op.28 in Cortot's interpretation are highlighted. They distribute around the circle. Major/minor are not clearly separated. See all Chopin Preludes marked as red symbols in Figure 6.

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

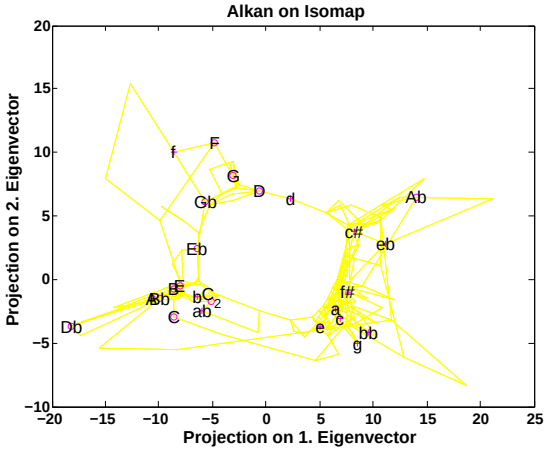

Fig. 9: Alkan concentrates somewhat on two wide clusters, but major and minor intermingle a little. $D\flat$-major is an outlier. (Cf. Figure 6 for details.)

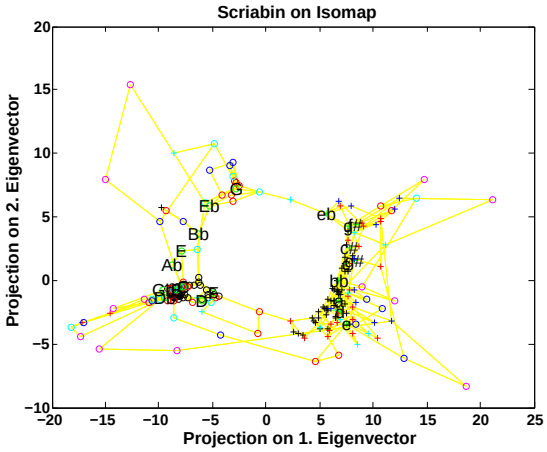

Fig. 10: Scriabin clearly separates into a major and a minor cluster. (Cf. Figure 6 for details.)

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. The compactness of clusters can be musically interpreted as the composer's consistency in the use of major/minor.

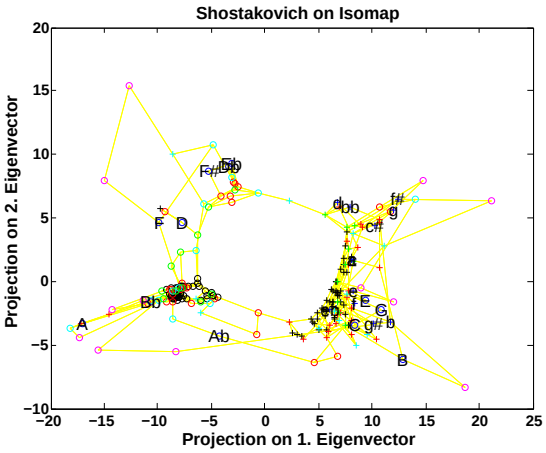

Fig. 11: Shostakovich is widely distributed with *A*-major as an outlier. (Cf. Figure 6 for details.)

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 extend certain features do manifest in constant Q scale degree profiles, like key character and performer. [1] 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.

Bibliography

[1] 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.

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

[3] Leonhard Euler. *Opera Omnia*, volume 1 of *3*, chapter Tentamen novae theoriae musicae. Stuttgart, 1926.

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

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

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

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

[8] Carol L. Krumhansl and E. J. Kessler. Tracing the dynamic changes in perceived tonal organization in a spatial representation of musical keys. *Psychological Review*, 89:334–68, 1982.

[9] 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.

[10] 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.

[11] 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.

[12] A. Schoenberg. *Structural functions of harmony.* Norton, New York, 2. edition, 1969.

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

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