Available online at www.sciencedirect.com
SCIENCE DIRECT®
ELSEVIER
Pattern Recognition Letters 26 (2005) 299–309
Pattern Recognition Letters
www.elsevier.com/locate/patrec

On musical stylometry—a pattern recognition approach

Eric Backer [a,*], Peter van Kranenburg [b]

[a] Faculty of Electrical Engineering, Mathematics and Computer Science, Delft University of Technology, Delft, Netherlands
[b] Faculty of Arts, Department of Musicology, University of Utrecht, Utrecht, The Netherlands

Received 21 July 2004; received in revised form 12 October 2004

Abstract

In this short communication we describe some experiments in which methods of statistical pattern recognition are applied for musical style recognition and disputed musical authorship attribution.

Values of a set of 20 features (also called "style markers") are measured in the scores of a set of compositions, mainly describing the different sonorities in the compositions. For a first study over 300 different compositions of Bach, Handel, Telemann, Mozart and Haydn were used and from this data set it was shown that even with a few features, the styles of the various composers could be separated with leave-one-out-error rates varying from 4% to 9% with the exception of the confusion between Mozart and Haydn which yielded a leave-one-out-error rate of 24%. A second experiment included 30 fugues from J.S. Bach, W.F. Bach and J.L. Krebs, all of different style and character. With this data set of compositions of undisputed authorship, the F minor fugue for organ, BWV 534 (of which Bach's authorship is disputed) then was confronted. It could be concluded that there is experimental evidence that J.L. Krebs should be considered in all probability as the composer of the fugue in question.

Keywords: Musical style recognition; Authorship attribution; Style markers; Machine learning

In memoriam

It was within the development of the international conferences on pattern recognition, a field of continuing growth in the early seventies, and the establishment of the International Association for Pattern Recognition (IAPR), starting from the first ICPR held in Washington, DC in 1973, that, I first met Azriel and after that, almost yearly in Board meetings and alike, aiming at serving the pattern recognition community in the context of a strong international association, and world wide organization of the series of biannual conferences. He was strongly driven and motivated to strengthen the organization and the impact of the IAPR.

* Corresponding author.
E-mail addresses: e.backer@ewi.tudelft.nl (E. Backer), p.vankranenburg@lodebar.nl (P. van Kranenburg).

0167-8655/$ - see front matter
doi:10.1016/j.patrec.2004.10.016

Azriel was most supportive in the process of founding *Pattern Recognition Letters* (PRL), in October 1982, and, in his capacity as president of IAPR, he established the fact that IAPR became the official sponsor of the journal PRL. My colleague in founding and managing the journal was Edzard Gelsema who regretfully passed away much too early on March 2, 2000. I had the privilege to serve as co-chairman, again together with Edzard Gelsema, in 1992, in organizing the 11th IAPR International Conference on Pattern Recognition, at The Hague, The Netherlands. Prof. Rosenfeld's compliments on the scientific contents, outspoken while being there, were meant in the context of his scientific ideas on image modeling and picture processing which were the major subjects at that time. We were proud of his judgments.

With this short contribution we want to honor a unique personality as Azriel Rosenfeld was and his life-long dedication to the pattern recognition community.

(Eric Backer)

1. Introduction

In the past decades, the ever-increasing power of computers made it possible to execute pattern recognition algorithms on a large scale. Those algorithms can also be of great value in authorship attribution, resulting in a research area called non-traditional authorship attribution (Love, 2002; Mason, 1985). This kind of research, tries to quantize the representation of the style of a certain author (text) or composer (music). Studies of this kind are called stylometric studies. It is not obvious what exactly has to be quantized but something in the structure of text or musical composition should bear the "fingerprint" of its maker. Many so-called style markers are developed in order to classify text or composition to certain styles and to discriminate between alternatives of authors and composers.

Interesting work has been done by Dannenberg and Watson (1997). They used machine learning tools to recognize the "mood" of music, such as lyrical, frantic, etc. They showed very low error rates, however, they do not mention all the features that were used. Also, the work of Pedro Ponce de León and José Iñesta is worth mentioning, (Ponce de León and Iñesta, 2003). They used self-organizing neural maps to classify musical styles. Extracted features included basic melody properties like number of notes, pitch range, etc.

The main problem of stylometry is the lack of an underlying theory, (Love, 2002). Many style markers turn out to be distinctive, but often it is not clear why. Until the study is done, it is not known which of the style markers (or which combination) will be the discriminator. As a method for automatically obtaining style markers would be very desirable but has not been developed up to now, we have to generate a large number of potentially interesting features (style markers) which it is hoped will be suitable for stylometric studies. This will be the subject of Section 3.

As it is the aim of this study to contribute to the problem of a disputed authorship of a specific composition, a fugue known as BWV 534, two experiments were defined to show that a pre-defined set of 20 style markers (low-level properties of counterpoint) could be successful.

Experiment 1. To indicate the difference between the style of J.S. Bach and other composers like Telemann and Handel, as well as to distinguish between composers, like Haydn and Mozart, whose styles are very alike.

Experiment 2. To test the hypothesis that the piece BWV 534 is not composed by J.S. Bach, and most likely is composed by J.L. Krebs and most likely is not composed by W.F. Bach (J.S. Bach's son).

It should be noted that for more than two decades, there are indeed a number scattered musicological contributions about the disputed authorship of J.S. Bach with respect to BWV 534 (Humphreys, 1985), though not conclusive. The conjecture that the piece could have been written by J.L. Krebs is just one of the outcomes of a more fundamental study of Peter van Kranenburg in his thesis (Kranenburg, 2004), about the disputed authorship of BWV 534. The application of pattern recognition methods on a large scale is thereby just an attempt to verify some of the presently formulated hypotheses.

2. Data and data preparation

A large corpus of encoded music is available from the Center for Computer Assisted Research in the Humanities at Stanford University (CCARH). [1] From this collection, a number of compositions are drawn to construct the dataset that is used in the present study.

The collection of encoded music at the CCARH consists almost entirely of music from the eighteenth and early nineteenth centuries. Not all of this is suited for our purpose. Many movements from cantatas, oratorios and operas have a basso continuo, which is not completely written out. So, some harmonic characteristics cannot be determined. These movements are only used when more than two other voices are active most of the time. In order to reduce the variance in the computed feature values, it is also important not to include too short compositions. After examining the behavior of the feature values a minimum of 30 bars is taken. Another issue is the presence of transposing instruments. Sometimes several parts had to be transposed. Apart from this, many files needed some adaptations before CPNView [2] could parse them. With these limitations in mind, a number of compositions is chosen from the CCARH library.

For experiment 1, the resulting dataset consists of the following groups of pieces:

- J.S. Bach: 40 cantata movements;
- J.S. Bach: 33 fugues from "Das Wohltemperierte Clavier";
- J.S. Bach: 11 movements from the "Kunst der Fuge";
- J.S. Bach: 8 movements from the violin concertos;
- G.F. Handel: 39 movements from the Concerti Grossi, op. 6;
- G.F. Handel: 14 movements from trio sonatas, op. 2 and op. 5;
- G.Ph. Telemann: 30 movements from the "Fortsetzung des Harmonischen Gottestdienstes";
- G.Ph. Telemann: 24 movements from the "Musique de table";
- F.J. Haydn: 54 movements from the string quartets;
- W.A. Mozart: 53 movements from the string quartets.

Of the three baroque composers works in different genres are added. Orchestral works as well as compositions for small instrumentation, and, in the case of J.S. Bach, works for keyboard. Of Mozart and Haydn only string quartets were added.

As mentioned above, the main point of interest is the difference between the style of J.S. Bach and the other composers. But it is also interesting to try to distinguish between composers whose style is very much the same. Especially the set with Haydn and Mozart will be challenging, since only compositions of the same genre are included.

For experiment 2, we have been collecting relevant material for comparison of each of the three—in this study considered—candidates of authorship of BWV 534.

- J.S. Bach: 11 fugues (different keys, different time signatures and different date of origin; it is assumed that all pieces have been composed by J.S. Bach);
- J.L. Krebs (pupil of J.S. Bach): 8 fugues (as above; all composed by J.L. Krebs);
- W.F. Bach (J.S. Bach's son): 5 fugues (as above; all composed by W.F. Bach).

In order to escape from the curse of dimensionality (and thus aiming at producing a sufficient amount of data), and at the same time making use of the length of a composition, we explore (overlapping) windowing over the entire composition as shown in Fig. 1.

Clearly, we are facing a trade-off between the number of fragments (as high as possible) and the variance of the feature values (as small as possible) computed on the basis of the number of bars in the fragment. From Fig. 2, we observe that a

---

[1] <http://www.ccarh.org>.

[2] CPNView: Donncha Ó Maidín (University of Limerick, Ireland).

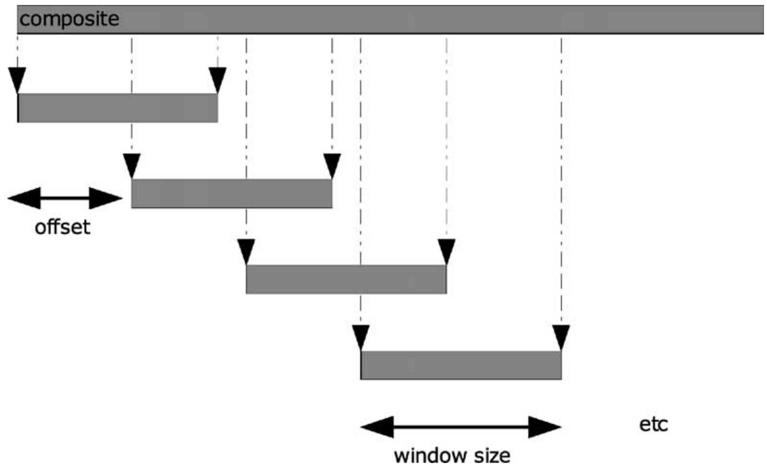

Fig. 1. Windowing over an entire composition.

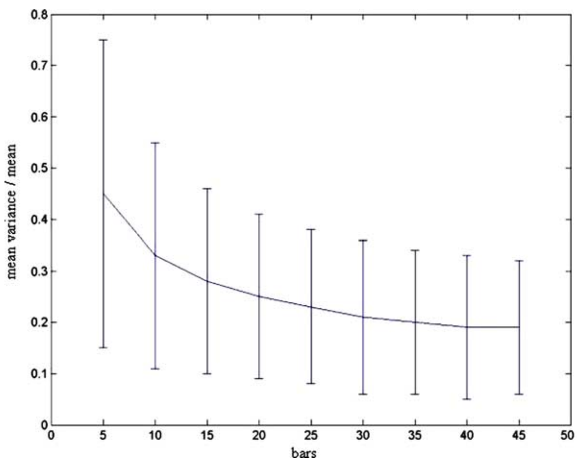

Fig. 2. Mean variance of feature values as a function of the number of bars in a window (fragments of the composition).

choice of 30 bars per fragment seems to be adequate to generate reliable feature values (in terms of variance). As a consequence, data points obtained from overlapping fragments will be close to each other in the feature space. Decomposing a composition (windowing) results in a number of related data points, enabling us to represent a composition as a cloud of data points on the basis of which global densities can be estimated. We note that the amount of overlap is a free parameter, which can be used in the analysis at later stage.

Also, as the data points generated from one composition are ordered in time, a composition is represented in the feature space as a path.

We denote a data set resulting from windowing as 30_10 data set if we use 30 bars as the window size and 10 bars as the offset. Likewise a 30_01 data set means 30 bars as the window size and only 1 bar offset (this is the largest overlap possible).

3. Features (style markers)

For each composition in the dataset, the values of 20 features are computed (Kranenburg and Backer, 2004). Most of these features are low-level properties of counterpoint. When composing polyphonic music, the composer must control the distances between the voices. The way he is doing this can be expected to be consistent for compositions in different genres and of different dates. Apart from the distances between the voices, some other features are computed which can be expected to be discriminative. Higher-level features (e.g. the key, modulations, the development of a theme, the use of certain motifs, etc.) are expected to be less suitable for our purpose, since they reflect the characteristics of the individual compositions. The following features are computed. Some of them come with an explanation, although in general, from a musicological point of view, much is just speculation.

3.1. StabTimeslice

The "stability" of the length of the successive time slices. With a time slice the time interval between two changes in the music is meant. This is shown in Fig. 3. The stability is computed by dividing the standard deviation of the lengths of the time slices by the mean length of the time slices. This normalization is necessary to compare pieces with different time signatures. So, when having a low value, the music is more like a steady stream, while a larger value indicates more diversity in rhythm.

3.2. DissPart

The fraction of the score that consists of dissonant sonorities. Consonants are: perfect primes, minor and major thirds, perfect fourths and fifths and minor and major sixths. But a fourth is considered dissonant if it is between the lowest voice and one of the upper voices. All other intervals are considered dissonant. The total duration of dissonant sonorities is divided by the total duration of the composition.

3.3. BeginBarDiss

The fraction of bars that begins with a dissonant sonority.

3.4. SonorityEntropy

For this feature, the concept "sonority" is used according to the definition of Mason (1985) In this definition sonority is a certain type of chord. So e.g. all the major triads are the same sonority, regardless of inversion or pitch. A unique number represents each sonority. For each sonority the total duration of all occurrences is computed. Then the probabilities of occurrence are estimated using these weighted frequencies. With this probabilities the entropy is computed according to:

$$-\sum_{i=1}^{N} p_i \log(p_i)$$

where N is the total number of sonorities and p_i the probability of occurrence of sonority i.

3.5. HarmonyEntropy

Mason also defines the concept "Harmony". It is much like sonority, but now difference is made in pitch. So e.g. a F-major triad and a G-major triad are the same sonority but different harmonies. Again the inversion is not taken into account. The value of this feature is computed the same way as the Sonority Entropy.

3.6. PitchEntropy

A list of occurrences of all pitches is made. Again the occurrences are weighted by the durations. Of the resulting list, the entropy is computed.

3.7. VoiceDensity

In a polyphonic composition not all voices are active during the whole composition. The voice density indicates the average number of active voices. This is normalized with the total number of voices. For this feature only bars that are strictly polyphonic are taken into account i.e. bars in which no voice has more than one note and in which more than one voice is active.

3.8. PartSeconds, PartThirds, PartFourths, PartAugFourths, PartDimFifths, PartSixths, PartSevenths, PartOctave

When combining the different voices of a polyphonic composition, the composer has to obey certain constraints. In many of these constraints the

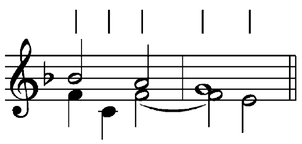

Fig. 3. Boundaries of the time slice.

vertical distances between the voices are important. This set of features measures the amount of a number of intervals between the different voice-pairs. Systematically all voice-pairs are examined. The total duration of all occurrences of each specific interval is computed and at the end divided by the total duration of all intervals in all voice-pairs. The intervals are taken modulo one octave. So e.g. a tenth is a third. When the same pitch occurs in more than one voice, it is taken into account once.

3.9. *ParThirds, parFourths, parSixths*

It can happen that in a voice pair two intervals of the same size succeed each other. This is called a parallel. For these three features the amount of parallel thirds, fourths and sixths is computed in the same way as the previous group of features. The total duration of all intervals involved in these parallels is added and divided by the total duration of all intervals in all voice pairs.

3.10. *StepSuspension*

When a dissonant is sounding between two voices, it often is suspended into a consonant by lowering the lower voice one step. This feature indicates how many dissonances are suspended this way. It is computed in the same way as the previous features.In the remaining these features are referred to by their index numbers. These can be found in Table 1.

4. Analysis

All experiments are carried out with the Matlab-toolbox *PRTools*. [3]

For both experiments, we perform feature selection using the *Floating Forward Selection* algorithm, proposed by Pudil et al. (1994). The FFS algorithm is applied to all possible class arrangements like {Bach}{all other composers}, {Bach}{Telemann}{Handel} etc. for experiment 1.

Table 1
The feature set (style markers)

Index	Feature
1	*StabTimeslice*
2	*DissPart*
3	*BeginBarDiss*
4	*SonorityEntropy*
5	*HarmonyEntropy*
6	*PichEntropy*
7	*VoiceDensity*
8	*PartSeconds*
9	*PartThirds*
10	*PartFourhs*
11	*PartAugFourths*
12	*PartDimFifths*
13	*PartFifths*
14	*PartSixths*
15	*PartSevenths*
16	*PartOctaves*
17	*ParThirds*
18	*ParFourths*
19	*ParSixths*
20	*StepSuspension*

Likewise, we have class arrangements like {J.S. Bach}{W.F. Bach}{J.L. Krebs}, {J.S. Bach}{W.F. Bach}, {J.S. Bach}{J.L. Krebs} and {W.F. Bach}{J.L. Krebs} for experiment 2. Fig. 4 shows the distributions of some of the "best" features to discriminate between the classes

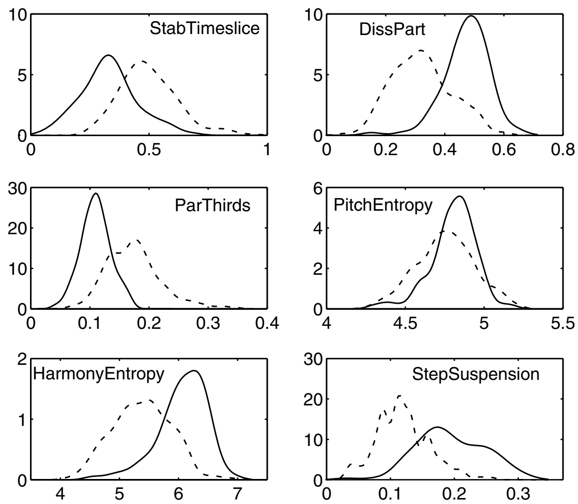

Fig. 4. Some of the "best" features for class arrangement {Bach}{notBach}; Bach = solid, not-Bach = dashed.

[3] <http://www.ph.tn.tudelft.nl/~bob/PRTOOLS.html>.

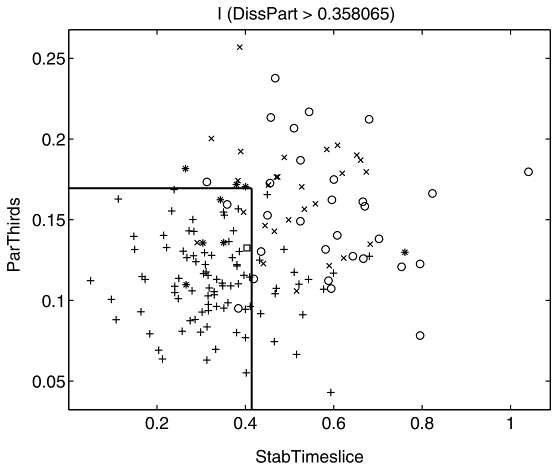

Fig. 5. Scatter plot with features that characterize Bach's style (+). Compositions with DissPart ⩽ 0.358 are not shown.

{Bach} and {not-Bach} and Fig. 5 a simple decision boundary obtained from these three features.

For comparison of the two extremes in class arrangements {Bach}{Telemann, Handel, Haydn, Mozart} (a two-class problem; a) and {Bach}{Telemann}{Handel}{Haydn}{Mozart} (as a five-class problem; b), we observe the following classification statistics when using a *k-nearest neighbor classifier* (see Table 2).

The five-class problem is obscured by the presence of compositions of Haydn and Mozart. The arrangement: {Haydn}{Mozart} yields a leave-one-out error (l-o-o-error) of 24.30%. Here we observe a significant limitation of the used music library (CCARH) as Haydn and Mozart were only represented by a collection of string quartets, obscuring the recognition results for the five-class arrangement. All other arrangements not including discrimination between Haydn and Mozart yield errors between 5% and 9%. A full account of the above results is given in (Kranenburg and Backer, 2004).

From these results we conclude that Bach's style can be isolated from the style of other composers with such a performance that it might be regarded as a valuable addition to the traditional methods of musical style classification. It offers a quantitative evaluation of the styles rather than the traditional qualitative descriptions. It is important not to see this as a replacement, but as an addition. Combining results from different viewpoints, will give more robust knowledge. The results of the above studies are a promise for the future, in which we can expect further increase in the computational power as well as further increase in the understanding and application of pattern recognition techniques.

This also means that this kind of research can be helpful in authorship disputes. This is the origin of experiment 2. [4]

Some of the features (from Section 2) are displayed in Fig. 6 for the class arrangement {J.S. Bach}{J.L. Krebs}{W.F. Bach}. The densities are estimated using a 30_01 data set (maximum overlap). None of the features are perfect discriminants, however a combination of six features used for training of a quadratic Bayesian classifier with 10-fold-cross-validation yields a (still optimistic) error of 1.2% for the discrimination of {J.S. Bach}{J.L. Krebs} arrangement, 1.6% for the discrimination of {J.S. Bach}{W.F. Bach} arrangement, and 1.6% for the discrimination of {W.F. Bach}{J.L. Krebs} arrangement.

We are using the Fisher Linear Discriminant transformation over the entire feature space to visually interpret the "best" two-dimensional scatter plot (discriminants 1 and 2). Fig. 7 shows the resulting scatter plot of the transformed data set with classes {J.S. Bach}, {J.L. Krebs} and {W.F. Bach} with in overlay—as an example—BWV 535, a fugue of J.S. Bach of which authorship is certainly not-disputed.

In order to interpret the features used in decision making of the different class arrangements, we generate the corresponding decision trees (C4.5).

We observe:

1. For the class arrangement {J.S. Bach}{J.L. Krebs} the decision tree (Fig. 8a) uses four features StabTimeslice, PartSeconds, PartThirds, and PartFourths; only one fragment of the 30_10 data set is misclassified.

[4] For all details about data: http://www.musical-style-recognition.net.

Table 2
Minimal errors with nearest neighbor classsifiers

Class arrangement	*k*	Feature subset	l-o-o-error (%)
Two-class problem (a)	5	1, 17, 2	6.62
Five-class problem (b)	11	2, 13, 8, 9, 1, 17, 5, 10, 14, 19, 11, 6, 7, 20	26.47

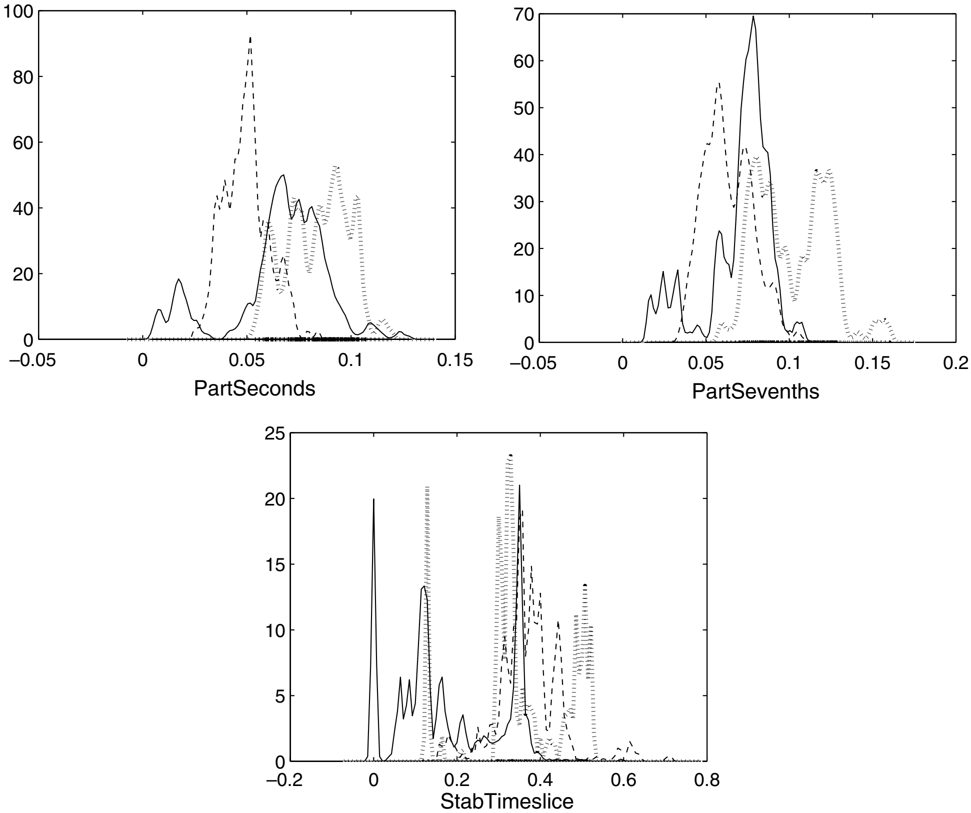

Fig. 6. Some densities of features for the class arrangement {J.S. Bach}{J.L. Krebs}{W.F. Bach}; J.S. Bach = solid, J.L. Krebs = stripes, and W.F. Bach = dashed.

2. For the class arrangement {J.S. Bach}{W.F. Bach} the decision tree (Fig. 8b) uses four features: ParthSevenths, StabTimeslice, PartOctaves, and StepSuspension; only one fragment is misclassified.
3. For the class arrangement {J.S. Bach}{W.F. Bach} the decision tree uses two features: BeginBarDiss and ParThirds; no errors occur.

We are now ready to classify the disputed f fugue for organ, BWV 534.

1. If we assume that J.S. Bach and W.F. Bach were the only candidates, we observe the following.
   The quadratic Bayesian classifier, trained with all features from the transformed Fisher space assigns all fragments of BWV 534 to {J.S. Bach}.
   With the six best features selected by FFS, 10 (out of 11) fragments are assigned to {J.S. Bach} for 30_10 data.
   The decision tree from Fig. 8b also assigns all fragments of BWV 534 to {J.S. Bach}. There-

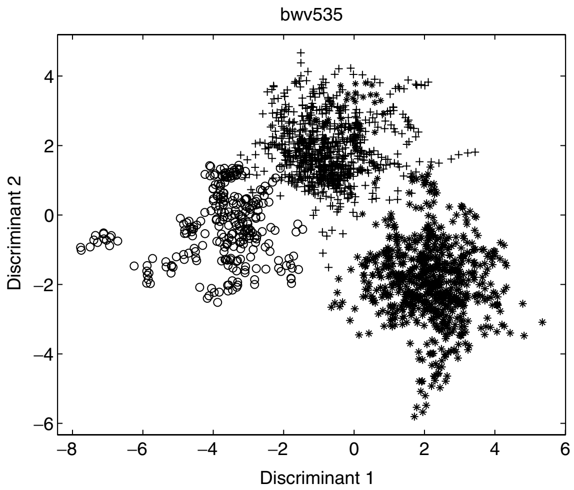

Fig. 7. Scatter plot of the fugue data set in the Fisher-transformed feature space; J.S. Bach (+), J.L. Krebs (∗) and W.F. Bach (○); BWV 535—as an example—in overlay (bold stars).

fore, it is save to conclude from these observations that the hypothesis of W.F. Bach being considered as the composer is false.

2. If we assume that J.S. Bach and J.L. Krebs were the only candidates, we observe the following. The quadratic Bayesian classifier, trained with all features of the transformed Fisher space assigns all fragments of BWV 534 to {J.L. Krebs}. Also, with the best six features selected by FFS, all fragments are assigned to {J.L. Krebs} for 30_10 data.
   The decision tree from Fig. 8a assigns five fragments to {J.S. Bach} and six fragments to {J.L. Krebs}. Therefore, it is still save to conclude that the style of BWV 534 resembles the style of {J.L. Krebs} more than the style of {J.S. Bach} and that the hypothesis of J.L.
   Krebs being considered as the composer, is true.
3. If we compare the styles of J.L. Krebs and W.F. Bach, we observe the following.
   The quadratic Bayesian classifier, trained with all features from the transformed Fisher space

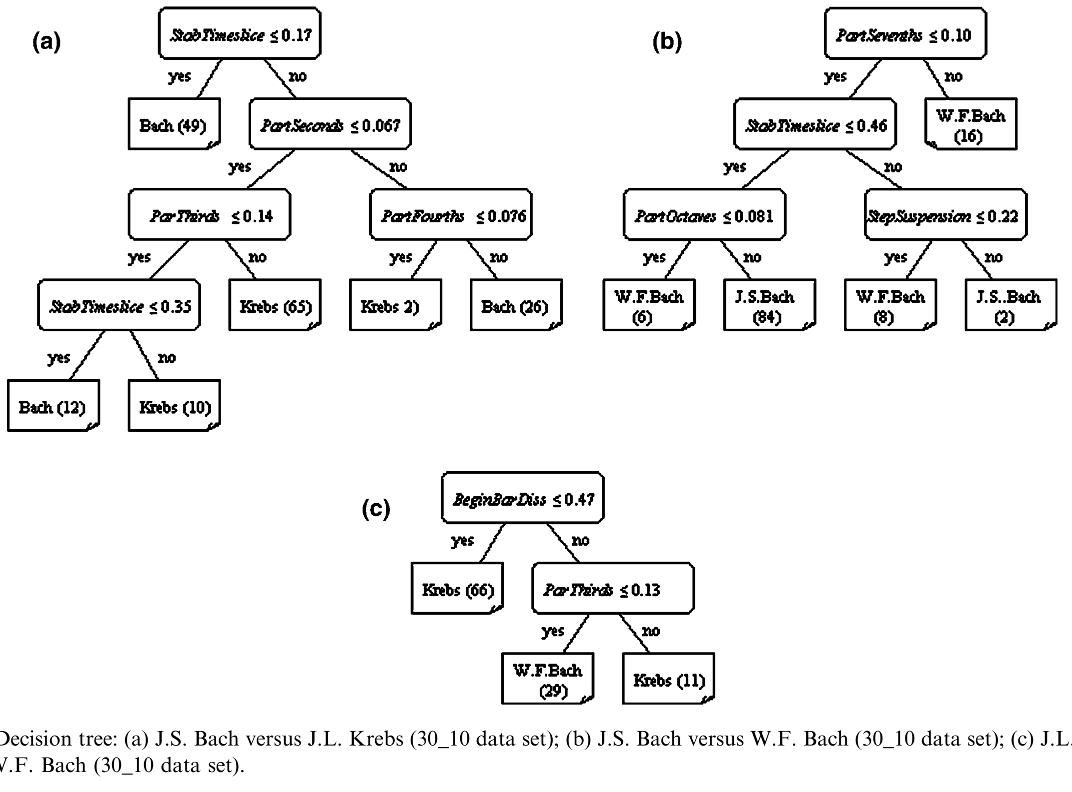

Fig. 8. Decision tree: (a) J.S. Bach versus J.L. Krebs (30_10 data set); (b) J.S. Bach versus W.F. Bach (30_10 data set); (c) J.L. Kerbs versus W.F. Bach (30_10 data set).

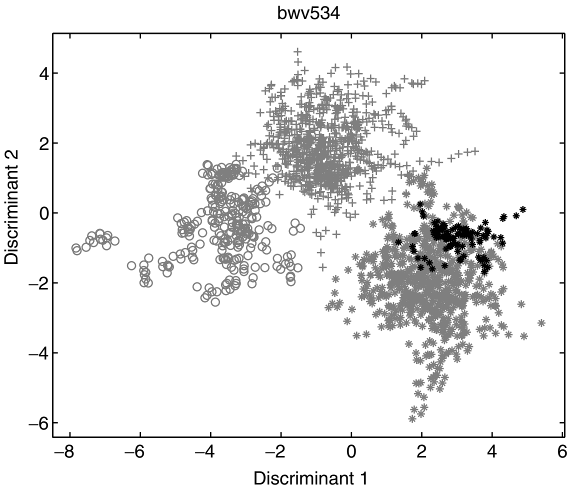

Fig. 9. Scatter plot of the disputed fugue BWV 534 (bold stars) in the Fisher-transformed feature space; J.S. Bach (+), J.L. Krebs (∗) and W.F. Bach (○).

assigns all fragments of BWV 534 to {J.L. Krebs}.
With the best six features selected by FFS, all fragments are assigned to {J.L. Krebs} for 30_10 data.
The decision tree from Fig. 8c also assigns all fragments of BMV 534 to {J.L. Krebs}. Therefore, it is save to conclude that if the choice had to be made between J.L. Krebs and W.F. Bach, there is no doubt in considering J.L. Krebs as the composer.

In Fig. 9, the fragments of the disputed fugue BWV 534 are displayed in overlay with the two-dimensional Fisher transformed feature space, nicely indicating how well the fragments fit into the available data of J.L. Krebs.

5. Conclusions

In this short communication, we have presented an attempt to apply pattern recognition techniques in the area of musical style characterization and disputed musical authorship.

First, we conducted an experiment to investigate how well the style of different composers could be identified. For that purpose, we designed 20 low-level properties of counterpoint to be measured in the represented score of a composition. It was concluded that it is very possible to isolate the style of J.S. Bach from other composers like Telemann, Handel, Haydn or Mozart. Given the positive outcome, it has been a challenge to enter the field of non-traditional author attribution.

So, we conducted a second experiment to investigate how well disputed musical authorship of a given composition could be solved if a limited number of alternatives are given a priori. In our case, the fugue BWV 534 which has been attributed to J.S. Bach but of which real authorship has been disputed on musicological grounds. His son, W.F. Bach and his pupil, J.L. Krebs have been put forward as serious candidates for true authorship.

From the experiment, it safely could be concluded that there is experimental evidence that J.L. Krebs has to be considered in all probability as the real composer.

Acknowledgement

The continuous support of Dr. Albert Clement, musicologist and Bach-expert of the University of Utrecht, The Netherlands, is very much appreciated.

References

Dannenberg, T., Watson, D., 1997. A machine learning approach to musical style recognition. In: Proceedings of the 1997 International Computer Music Conference, pp. 344–347.

Humphreys, D., 1985. Did J.S. Bach Compose the F minor Prelude and Fugue BWV 534? In: Williams, P. (Ed.), Bach, Handel, Scarlatti—Tercentenary Essays. Cambridge, pp. 173–184.

Kranenburg, P. van, Backer, E., 2004. Musical style recognition, a quantitative approach. In: Proceedings of the Conference on Interdisciplinary Musicology, Graz, Austria.

Kranenburg, P. van, 2004. De fuga in f (BWV 534) van Johann Sebastian Bach? Thesis. University of Utrecht. Available from: <www.Musical-Style-Recognition.net> (in Dutch).

Love, H., 2002. Attributing Authorship: an Introduction. Cambridge.

Mason, R.M., 1985. Modern Methods of Musical Analysis using Computers. Peterborough.

Ponce de León, P.J., Iñesta, J.M., 2003. Feature-driven recognition of musical styles. Lecture Notes in Computer Science 2652, 773–781.

Pudil, P., Novovicova, J., Kittler, J., 1994. Feature Search Methods in Feature Selection. Pattern Recognition Letters 15, 1119–1125.