University of Groningen

Musicologist-driven writer identification in early music manuscripts

Niitsuma, Masahiro; Schomaker, Lambert; van Oosten, Jean-Paul; Tomita, Yo; Bell, David

Published in:
Multimedia Tools and Applications

DOI:
[10.1007/s11042-015-2583-8](https://doi.org/10.1007/s11042-015-2583-8)

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Publisher's PDF, also known as Version of record

Publication date:
2016

[Link to publication in University of Groningen/UMCG research database](#)

Citation for published version (APA):

Niitsuma, M., Schomaker, L., van Oosten, J-P., Tomita, Y., & Bell, D. (2016). Musicologist-driven writer identification in early music manuscripts. *Multimedia Tools and Applications*, 75(11), 6463-6479. <https://doi.org/10.1007/s11042-015-2583-8>

Copyright

Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license. More information can be found on the University of Groningen website: <https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment>.

Take-down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): <http://www.rug.nl/research/portal>. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.

Musicologist-driven writer identification in early music manuscripts

Masahiro Niitsuma⁴ · Lambert Schomaker² ·
Jean-Paul van Oosten² · Yo Tomita¹ · David Bell³

Received: 1 July 2014 / Revised: 3 February 2015 / Accepted: 23 March 2015 /

Published online: 2 May 2015

© Springer Science+Business Media New York 2015

Abstract Recent renewed interest in computational writer identification has resulted in an increased number of publications. In relation to historical musicology its application has so far been limited. One of the obstacles seems to be that the clarity of the images from the scans available for computational analysis is often not sufficient. In this paper, the use of the Hinge feature is proposed to avoid segmentation and staff-line removal for effective feature extraction from low quality scans. The use of an auto encoder in Hinge feature space is suggested as an alternative to staff-line removal by image processing, and their performance is compared. The result of the experiment shows an accuracy of 87 % for the dataset containing 84 writers' samples, and superiority of our segmentation and staff-line removal free approach. Practical analysis on Bach's autograph manuscript of the Well-Tempered Clavier II (Additional MS. 35021 in the British Library, London) is also presented and the extensive applicability of our approach is demonstrated.

Keywords Image processing · Writer identification · Domain driven data mining · Optical music recognition

✉ M. Niitsuma
mniitsuma@media.ritsumeikai.ac.jp

L. Schomaker
l.schomaker@ai.rug.nl

¹ School of Creative Arts, Queen's University of Belfast, Belfast, UK

² Department of Artificial Intelligence, University of Groningen, Groningen, Netherlands

³ School of Electronics, Electrical Engineering and Computer Science, Queen's University Belfast, Belfast, UK

⁴ Department of Media Technology, College of Information Science and Engineering, Ritsumeikan University, Shiga, Japan

1 Introduction

Handwriting analysis has been widely used for writer identification as a common alternative to physiological biometrics. Although physiological biometrics can be strong modalities for person identification because of their reduced variability and high complexity, cooperating subjects are required by the invasive character of the process. Applicability to many domains is therefore limited. In contrast, behavioural biometrics such as handwriting are less invasive and they have been actively investigated from computer science disciplines with a high identification rate [16].

However, only a few attempts have been made at writer identification in musical scores, despite the huge potential and the possible impact it would have on the domain. Writer identification has been a critical issue in musicology, playing a prominent role in authenticating unknown pieces and identifying their chronology [10]. However, it lacks a method for subject evaluation, thereby holding a high risk that it may provide false assumptions for further hypotheses. In Bach studies, a major revision was made to the chronology of Bach's works in the light of the ground-breaking research by [4] and [2] in the 1950s. Prior to this, scholars largely accepted the chronology [18] proposed in the late 19th century. This demonstrates the importance of a method for quantitative evaluation of hypotheses in musicological studies.

How to deal with staff-lines seems to be one of the most important issues when computational writer identification methods are applied to musicological problems. In terms of writer identification, staff-lines should be regarded as background and should be removed. Most of the previous studies of writer identification in musical notation have made this assumption. However, this assumption significantly limits application, as it requires sufficient quality of images to achieve the removal of staff-lines satisfactorily. Especially when original manuscripts have been subject to age related degradation, the result is far from satisfactory; some lines are visible and symbols can appear broken (Fig. 1). It seems to have hindered the evaluation of writer identification on data sets containing handwritings closely resembling each other (Fig. 2). Such homogeneous handwritings are often found in copies produced in the baroque era, and they tend to be available only

Fig. 1 Staff-line removal using Dalitz method. The original score is shown in Fig. 2 (b)

Fig. 2 Demonstration of copying activities common in the Baroque era: **(a)** Original autograph manuscript by Johann Sebastian Bach (JSB); **(b)** Anna Madalena Bach (AMB)'s copy; **(c)(d)(e)(f)** Other copyists' copies

as personal scans of poor quality. Moreover, comparatively little empirical information has emerged on the actual effect of prior staff-line removal on the writer identification rate.

It is necessary to consider how staff-line removal affects the accuracy of writer identification, and to evaluate writer identification methods using more realistic data sets, for example, those which contain images of poor quality and musical scores resembling each other, such as those used in historical musicology. This paper has two aims: (1) To propose a method that does not require the removal of staff-lines to identify the writers of musical scores; (2) To compare the accuracy rate by the presence or absence of staff-lines in personal scans and in a black and white microfiche owned by a musicologist.¹

¹It is often the case that libraries do not allow us to return to source.

2 Methodology

2.1 The contour-hinge probability distribution function

Although writer identification is one of the most important issues in musicology, methodologies seem to lack both objectivity and transparency. Musicologists' investigations of Bach's handwriting, for example, have concentrated on geometric analysis of musical symbols such as C-clefs. However, the chronological variance of Bach's handwriting is not necessarily limited to the shape recognisable by the untrained eye. Machine learning has a potential to capture the characteristics of scripts in different periods other than such shapes. In previous work [13], we have investigated this possibility by extracting geometric features from Bach's C-clefs and arranging them into the groups specified by the labelled training data. However, the segmentation of each C-clef was mainly done manually, therefore, its use is limited to experiments on small data sets.

On the other hand, more robust feature extraction methods have been investigated in computational writer identification research. Directional features such as slants seem to be effective for handwriting analysis, and yet they have rarely been investigated statistically by musicologists.²

Slant is determined by the biomechanics of the wrist/finger system in interaction with the trained pen grip. Its application was first investigated by Dooijes, and followed by Maarse's confirmation of its stability with rotation and within a writer's work [11]. Using single line text samples, an identification performance of 79 in 80 (99 %) for twenty writers was achieved by linear discriminant analysis with five online features [12].

Recently, Schomaker and Bulacu [17] proposed novel Contour-Hinge features which capture not only slant but also curvature information which is caused by the differential phase modulation of the wrist and the finger system, as shown in Fig. 3. They reported that an accuracy of 84 % has been reached using the Contour-Hinge feature for 250 writers. Moreover, the feature extraction does not involve segmentation. This makes the Contour-Hinge feature the most suitable method for writer identification from poor quality scans.

Imagine a hinge laid on the surface of the image, and place its junction on top of every contour pixel, then open the hinge and align its legs along the contour. Consider the angles ϕ_1 and ϕ_2 that the legs make with the horizontal and count the instances in a two-dimensional array of bins indexed by ϕ_1 and ϕ_2 . The final normalised histogram gives the joint probability distribution function $p(\phi_1, \phi_2)$ quantifying the chance of finding in the image two hinged contour fragments oriented at the angles ϕ_1 and ϕ_2 , respectively (Fig. 3).

2.2 Staff-line removal by an auto encoder

The presence of staff-lines generate dominant angle probabilities in the Hinge feature space, which are not writer invariant. Dimensionality reduction can discard those as irrelevant information, thereby achieving staff-line removal not at pixel level but in feature space. We propose the use of an auto encoder to attempt staff-line removal in feature space.

²Javis [9] claims to have conducted his research this way. However, his application was not sufficiently transparent.

Fig. 3 (a) Schematic description of the extraction method of the Contour-Hinge. (b) Surface plots of the Contour-Hinge for two different writers. Each lot shows a unique landscape [1]

An auto encoder is a neural network, where the number of input neurons is the same as the number of the output neurons, and the number of hidden neurons is significantly smaller than the number of input and output neurons (Fig. 4). An auto encoder minimises the

Fig. 4 Schematic description for an auto encoder

difference between the input and the output of the network, whereby low-dimensional data is produced in the narrow hidden layer. This hidden layer can be interpreted as representing the principal components of the input pattern space. In fact, it is well known that an auto encoder with linear neurons leads to results strongly related to PCA (Principal Component Analysis). However, PCA requires inverse operations, which can be difficult on certain types of data. Moreover, on a large dataset containing high-dimensional data such as those used in this study, it is preferable to use an auto encoder. Neural dimensionality-reduction techniques are currently well established [8].

In the context of Hinge features, the residual—obtained by subtracting the reconstructed signal from hidden neurons, from input neurons—is assumed to have no angle information stemming from general attributes such as bar lines and staff-lines. In the residual, some dimensions have lower vector-element values (suppressed) whereas other dimensions have higher values. In the context of a feature vector, which consists of a probability distribution of Hinges, this implies that the virtual presence of a Hinge with $p(\phi_1, \phi_2)$ in the image is reduced, or enhanced.

No smoothing or optimisation has been performed to present the polar plot shown in Fig. 5, yet the predominance of (almost) horizontal angles clearly indicates that the auto encoder was able to detect the redundant information in the horizontal staff-lines. It should be noted that staff lines are often slightly curved and scans may be slightly rotated. Such factors introduce a natural spread around the horizontal.

Figure 6 shows the pixels with the most prominent Hinge shape $p(\phi_1, \phi_2)$ estimated by an auto encoder from the image shown in Fig. 2b. The result confirms the expectation that the horizontal staff-lines originally dominated the variance in the input vector. It is understood that most of the staff-lines are suppressed by the proposed method. It should be noted

Fig. 5 Demonstration of the most suppressed Hinges; the radius shows how many times the angle (ϕ) appears in the top 3 list in the hidden layers of an auto encoder trained by Bach's autograph manuscript P 967 of 6 Violin Sonatas and Partitas. It is understood that they are concentrated around 0 and 180 degrees—which are suppressed the most

Fig. 6 Staff-line removal by an auto encoder; The 200 most suppressed dimensions are illustrated. The pixels were extracted only if they have the dominant angle probabilities detected by an auto encoder. They mostly coincide with staff-lines. The original score is shown in Fig. 2b

that this pseudo staff-line removal is achieved not by image processing, but by dimensionality reduction in the Contour-Hinge feature space—which encodes the horizontality. The effect is widespread and sometimes reaches beyond the staff-lines, while some pixels are left even within staff-lines. Experimentation is necessary to ensure that the Hinges (curvatures) essential for writer/scribe identification, survive the suppression.

3 Experiment

There is a well-organised data set by Fornés et. al, accompanied by the staff-line removed version and writer information [5]. However, it seems to consist mainly of modern handwritten scores, which have fewer calligraphic features than those found in early music manuscripts. Our focus is on the evaluation of applicability to musicological problems—as the greatest demand for writer identification in music scores comes from musicologists whose study is concerned with determining if a certain composer is the writer of given original manuscripts.

To emulate the practical situation in musicology, it is important to use a data set containing more degraded and homogeneous handwriting. In this study, two data sets are considered; the first one (denoted by *Scan*) was gathered from an online archive and represents common colour scans in relatively fine quality. It includes music manuscripts by 44 different composers from the 16th century to the 20th century. The second one (denoted by *Fiche*) was gathered from personal scans and from a black and white microfiche owned by a musicologist, representing more common data in the domain. The second data set contains more degraded images of manuscripts, with less contrast between the background and foreground (Fig. 7).

The second data set (*Fiche*) is constituted of only the sources of the Well-Tempered Clavier II. While JSB's original manuscripts are among them (e.g. The British Library, London, Add. MS 35021, which consists of pairs of movements—21 preludes and fugues—partially in JSB's own hand and partially copies by Anna Magdalena produced under the composer's supervision), the majority are copies with many different kinds of errors and variants, some of which were supposedly made from the now lost autograph manuscripts.

Fig. 7 Comparison of common scans and microfiche data; **(a)** an example from data set *Scan*; **(c)** an example from data set *Fiche*. **(b)** shows the intensity profile along the third stave line from the top in **(a)**. **(d)** shows that in **(c)**

One of the most significant merits of these images in the present context is their homogeneity on the same keyboard setting, owing to the fact that they all originated from the composers manuscripts (as shown in Fig. 2).

For both data sets, effort was made to collect handwriting as similar as possible. For instance, both data sets contained the handwriting of JSB and his wife AMB (as shown in Fig. 2 **(a)(b)**). This kind of calibration is important not only to enhance the classifiers but also to evaluate the real applicability in the domain, and yet it has not been attempted before. In order to have comparable experimental conditions across all data sets, these data sets were made to contain four pages per writer (352 pages in total).

Feature extraction starts with convolution with sobel operator, followed by binarization [14]. 718-dimensional Hinge features were calculated and an auto encoder was applied using the top 200 hidden units. Averaged processing time takes one minute per one scan on dual quad cores, 1.86 GHz, while a big scan can take 20 minutes. For comparative purposes, four different versions were distinguished according to how they are processed before classification:

- *Scan*—the original data set;
- *Scan-R*—one processed by an auto encoder;
- *Scan-S*—one processed by staff-line removal by image processing;

- *Scan-RS*—one processed by both an auto encoder and staff-line removal by image processing.

For *Scan-S* and *Scan-RS*, staff-lines were removed by Dalitz's method [3] from each binarized image, to be compared with the proposed method.

To compare the performance of the classification method, a battery of classifiers were adopted as follows. Classification was evaluated by a 10-fold cross validation with repetition 100 times. We used the implementation included in the Weka package for the 5 classifiers listed below.³ The number of repetitions for Bagging was set to 100. Euclidian distance was used as a metric for k-NN and k was set to 1. These parameters were decided upon the basis of the results of preliminary experiment using different data from those used for the experiment.

- SVM
- RF
- Bagging
- k-NN
- Naive Bayes (NBayes)

4 Discussion

Table 1 shows the results of the experiment and Table 2 shows detailed accuracy of the proposed method. For all methods, with the exception of SVM and NBayes, the residual calculation without staff-line removal (-R) performed best. It is understood that essential information for writer identification was retained by the residual calculation. k-NN performed well despite its simplicity. Its performance is not far behind the result for SVM despite the significantly shorter execution time required compared to that of SVM. This implies that the Hinge feature effectively encodes the character of each writer, and its distance directly reflects the actual distance of handwriting. On the other hand, the performance of Bagging and NBayes seems poor. This was significantly improved by an auto encoder both with and without staff-line removal. In contrast, for SVM, and k-NN, the accuracy is degraded by the residual calculation when staff-line removal is attempted beforehand (-RS). This implies that SVM, and k-NN managed to capture subtle features discarded by further suppression after staff-line removal. This seems to be also the reason why SVM achieved the highest accuracy rate with its high ability for generalisation, using no staff-line removal and no residual calculation. As dimensionality reduction such as that obtained using neural networks, does not necessarily sustain the information that can be used for classification, it seems desirable if SVM can handle the raw data without any dimensionality reduction. However, as the execution time for SVM is much longer than that of other classification methods (except for Bagging), it seems preferable to use other classifiers such as k-NN with the residual calculation from the perspective of applications. Another advantage of the residual calculation is that we can ensure that classifiers work on writer invariant features, and not on others (e.g. staff-lines); otherwise, even if an accuracy rate is high with the current data set, its generalisation ability may be restricted. Moreover, residual calculation takes

³See <http://www.cs.waikato.ac.nz/ml/weka> for more information about Weka. Normalised poly kernel was used as the kernel function of SVM and sequential minimal optimisation was conducted [15]. The number of trees was set to 100 for RF.

Table 1 Recognition rate for each classifier; -R denotes residual; -S denotes staff-line removal

Data set	SVM	RF		k-NN		Bagging		NBayes	
Scan	86.90	77.79	•	82.57	•	57.17	•	35.98	•
Scan-R	86.20	82.95		83.86		60.66	•	45.57	•
Scan-S	86.06	79.40	•	82.11	•	46.34	•	31.08	•
Scan-RS	85.04	75.83	•	80.82		51.34	•	48.07	•
Average	86.05	78.99		82.34		53.88		40.17	
Fiche	89.38	82.94	•	87.75		57.38	•	43.06	•
Fiche-R	90.88	85.25	•	88.00		59.38	•	53.00	•
Fiche-S	85.81	79.19	•	82.19		42.38	•	24.94	•
Fiche-RS	84.56	76.88	•	76.50	•	50.00	•	35.44	•
Average	87.66	81.06		83.61		52.28		39.11	
Combined	87.45	79.18	•	84.02	•	54.06	•	32.67	•
Combined-R	87.33	82.71	•	84.68		56.89	•	40.29	•
Combined-S	85.04	79.14	•	81.74	•	44.05	•	21.94	•
Combined-RS	84.92	76.07	•	78.55	•	47.19	•	34.50	•
Average	86.18	79.28		82.25		50.55		32.35	

• statistically significant improvement or degradation compared to SVM

significantly shorter time (few milliseconds) than staff-line removal by image processing (few minutes). Although more investigation is needed on the effect of staff-line removal on writer identification rate, prior staff-line removal seems to be effectively substituted by dimensionality reduction in the Contour-Hinge feature space.

For the data set *Fiche*, the overall accuracy was slightly better than that of *Scan*, contrary to expectation. Unlike the case of *Scan*, residual version outperformed the others for all methods. The performance was still good (87.3 % by SVM) for the combined data set (84 writers). In conclusion, better results can be obtained without staff-line removal using residual calculation.

Figures 8, 9, 10 and 11 present some queries and hit lists used in practical situation where the final judgement is generally determined by experts based on exemplars at a small ‘distance’ from the subject manuscript. In Fig. 8, similar handwriting is found throughout all the images. It is notable that the distance between (a) and (b) is smaller than that between

Table 2 Detailed accuracy of each classifier on Combined-R

Classifier	TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area	Processing time (sec)
SVM	0.869	0.002	0.872	0.869	0.861	0.986	51.40
RF	0.818	0.002	0.828	0.818	0.809	0.980	5.380
k-NN	0.842	0.002	0.843	0.842	0.830	0.912	0.010
Bagging	0.589	0.005	0.586	0.589	0.570	0.948	136.3
NBayes	0.411	0.007	0.770	0.411	0.485	0.732	0.140

Fig. 8 Successful writer identification search and top 5 hit list: (a) query image: (AMB); (b) rank 1 (AMB-03-dist=116.00); (c) rank 2 (AMB-02-dist=116.00); (d) rank 3 (JSB, dist=176.91); (e) rank 4 (JSB, dist=243.35); (f) rank 5 (JSB, dist=260.96.)

(a) and (d). Because (d) is the copy of (a) by AMB and the musical content of (d) is identical to that of (a). This shows the stability of the Hinge feature within the writer's work. In Fig. 9, the first hit was not AMB, but the majority of the hits concentrated on the early 17th century. This implies—considering the fact that the data set consists of a wide variety of manuscripts dating from the 16th to 20th centuries—that the Hinge feature captured the character of the handwriting that was unique to a certain period. In Fig. 10, all the handwriting is in orchestral format and similar textures were found. This implies that the Hinge feature captures not only writer invariant features but also features unique to musical

Fig. 9 Unsuccessful writer identification search and top 5 hit list: (a) query image: (AMB); (b) rank 1 (Tartini, dist=270.97); (c) rank 2 (Handel, dist=275.65); (d) rank 3 (Biber, dist=291.57); (e) rank 4 (Biber), dist=295.62); (f) rank 5 (C. G. Neefe), dist=296.62)

Fig. 10 Unsuccessful writer identification search and top 5 hit list: (a) query image (Gluck); (b) rank 1 (Gluck, dist=186.33); (c) rank 2 (Wagner, dist=189.54); (d) rank 3 (Wagner, dist=198.55); (e) rank 4 (Gluck, dist=204.59) (f) rank 5 (Wagner, dist=207.16)

style. Even in microfiche data, the top 3 hits were always by the writer of the query image as shown in Fig. 11.

5 Case Study

In this section, a case study of the London autograph manuscript is considered. The London autograph consists of 22 folios. Several different layers have been identified by

Fig. 11 Successful writer identification search in microfiche data and top 5 hit list: (a) query image (Add14330); (b) rank 1 (Add14330, dist=89.11); (c) rank 2 (Add14330, dist=96.65); (d) rank 3 (Add14330, dist=153.10); (e) rank 4 (Brussels, dist=196.45); (f) rank 5 (D-Rp SM92, dist=209.57)

Table 3 Layers in the London manuscript classified by the proposed method; ‘f’ denotes folio, ‘r’ denotes ‘recto’ meaning the front side of a leaf, and ‘v’ denotes ‘verso’ meaning the back side. The number of layers does not necessarily correspond to the chronological order

Layer One (AMB)	f8r (1st half), f11r,f11v,f2r,f2v,f22r,f22v,f4v
Layer One (JSB)	f10r,f10v,f12r,f12v,f21r,f3r,f3v,f7v,f8r (2nd half),f8v
Layer Two (JSB)	f16r,f16v,f17r,f17v,f5v
Layer Three (JSB)	f18r,f6r,f6v
Layer Four (JSB)	f1r,f1v,f13r,f13v,f14r,f14v,f18v,f20r,f20v,f9r
Layer Five (JSB)	f15r,f15v,f19r,f19v,f21v,f4r,f5r,f7r,f9v

musicologists, however, those layers still lack objective justification, and therefore, warrant further investigation.

Firstly, an authentication was determined by using a binary classifier (JSB or AMB) trained by the data set used in the previous experiment. From this, layer 1 and the other layers were distinguished (Table 3).

Secondly, more a detailed grouping is considered. Though we can identify some groupings by k-means clustering, it is difficult to decide the proper number of clusters to be identified. To solve this, a model-based clustering analysis is considered. Model-based clustering analysis assumes the observation as a mixture distribution and estimates the class of each sample as hidden variables. Let $X = \{x_1, x_2, \dots, x_n\}$ be n samples generated from parameters $\Theta = \{\theta_1, \theta_2, \dots, \theta_c\}$, then model-based clustering tries to maximise $p(X|\Theta)$ by the EM algorithm [6].

Figure 12 shows the number of clusters versus the value of Bayesian information criterion (BIC), implying that 6 clusters is optimal. Table 3 shows the result of classification

Fig. 12 Number of clusters versus the value of BIC. ‘II’: spherical, equal volume; ‘VII’: spherical, unequal volume; ‘EEI’: diagonal, equal volume and shape; ‘VEI’: diagonal, varying volume, equal shape; ‘EVI’: diagonal, equal volume, varying shape; ‘VVI’: diagonal, varying volume and shape; ‘EEI’: diagonal, equal volume and shape [6]

Table 4 Layers in the London manuscript proposed by musicologists. The number of layers corresponds to the chronological order

Layer One (AMB)	f2r, f2v, f4r, f4v, f8r (1st half), f22r, f22v, f11r, f11v
Layer One (JSB)	f8r (2nd half), f8v, f7r, f7v, f10r, f10v, f12r, f12v, f16r, f16v, f17r, f17v, f21r, f21v, f5r, f5v
Layer Two (JSB)	f20r, f20v, f3r, f3v, f9r, f9v, f19r, f19v, f6r, f6v, f18r, f18v, f15r, f15v
Layer Three (JSB)	f13r, f13v, f1r, f1v, f14r, f14v

using the diagonal, varying volume, equal shape model. The interpretation still involves domain knowledge. Cluster 1 mostly corresponds to AMB's handwriting. This concurs with the result of the binary classification. Cluster 5 consists of the folios coming from the later stages. On the other hand, some of the results contradict musicologists' conclusions as shown in Table 4.⁴ For instance, f4r was misclassified to JSB. Although more analysis involving multiple evidence is required to interpret these clustering results, it surely gives a lot of insight in the situation where only limited evidence is available.

The analysis of the distance and the hit list could be also used as the verification of hypotheses proposed by users. Assuming the hypothesis proposed by the musicologists is correct, the average and variance of the distance seems small enough within the same layer by the same writer (JSB to JSB was 11.70 ± 2.95 and AMB to AMB was 3.33 ± 1.7 . On the contrary, JSB to AMB was 19 ± 1.668 , which is significantly larger than the others.). Moreover, 85 % of the top 1 hit for each folio was from the same layers. Every case where AMB was classified into JSB, it classified as layer 1. In other words, the distance between different layers by the same writer is, as required, less than the distance between different writers.

6 Conclusion

The results of this study indicate that the Hinge-feature is effective for writer identification without staff-line removal. This corresponds with the fact that humans can identify writers by investigating musical scores with staff-lines. The proposed method is powerful and extensive, especially in the case of degraded musical scores, requiring no staff-line removal and no segmentation. Moreover, it has the advantage of avoiding subjective influence caused by manual measurement or any such human intervention, thereby improving reproducibility.

However, some application-specific limitations have been observed. In the present study, one page of a music score is assumed to be written by the same person; however the fact is that manuscripts often contain later revisions as well. They often appear to have been transcribed in different periods, in different ink, or even by different hands. It is important that the different layers of revisions are correctly identified to evaluate the authenticity of musical contents. To handle this, features should be extracted from segmented smaller pieces of scores (e.g. from each bar). In the case where the number of available handwriting examples is small, this could be a solution, whereby multiple samples from one page are generated.

⁴Franklin's classification is considered [7]. This conclusion was not based on handwriting analysis but the studies on paper, rastral and the diplomatic features of the score.

References

1. Bulacu M, Schomaker L (2007) Text-independent writer identification and verification using textural and allographic features. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 29(4):701–17
2. Dadelsen Gv (1958) Beiträge zur Chronologie der Werke Johann Sebastian Bachs. *Tübinger Bach-Studien*, vol. 4–5, Trossingen, Hohner
3. Dalitz C, Droettboom M, Pranzas B, Fujinaga I (2008) A comparative study of staff removal algorithms. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 30(5):753–766
4. Dürr A (1957) Zur Chronologie der Leipziger Vokalwerke J. S. Bachs. *Bach-Jahrbuch* 44:5–162
5. Fornés A, Dutta A, Gordo A, Lladós J (2011) The ICDAR 2011 music scores competition: staff removal and writer identification. In: *Proceedings of International Conference on Document Analysis and Recognition*, pp 1511–1515. IEEE
6. Fraley C, Raftery AE (2000) Model-based clustering, discriminant analysis, and density estimation. *J Am Stat Assoc* 97:611–631
7. Franklin DO (1989) Reconstructing the Urpartitur for WTC II: a study of the ‘London Autograph’. In: *Bach Studies*, pp. 240–77. Cambridge University Press, Cambridge, UK
8. Hinton GE, Salakhutdinov RR (2006) Reducing the dimensionality of data with neural networks. *Science* 313(5786):504–507
9. Jarvis M (2007) Did Johann Sebastian Bach write the six cello suites? Ph.D. thesis, Charles Darwin University
10. Kobayashi Y, Beißwenger K (2007) Die Kopisten Johann Sebastian Bachs : Katalog und Dokumentation, Kassel, Bärenreiter
11. Maarse F (1987) The study of handwriting movement: Peripheral models and signal processing techniques, Ph.D. thesis, University of Nijmegen
12. Maarse F, Schomaker L, Teulings HL (1988) Automatic Identification of Writers. In: van der GC, Mulder G (eds) *HumanComputer Interaction: Psychonomic Aspects*. Springer, New York, pp 353–360
13. Niitsuma M, Tomita Y, Yan W, Bell D (2011) Classifying Bach’s Handwritten C-clefs. In: *Proceedings of 12th International Conference on Music Information Retrieval*, pp. 417–421
14. Otsu N (1975) A threshold selection method from gray-level histograms. *IEEE Transactions on Systems Man, and Cybernetics* 11(1):285–296
15. Platt JC (1999) Fast training of support vector machines using sequential minimal optimization. In: Schölkopf B, Burges CJC, Smola AJ (eds) *Advances in kernel methods*. MIT Press, pp 185–208
16. Schomaker L (2007) Advances in Writer Identification and Verification. In: *Proceedings of Ninth International Conference on Document Analysis and Recognition (ICDAR 2007)*, vol. 2, pp. 1268–1273. IEEE
17. Schomaker L, Bulacu M, Franke K (2004) Automatic writer identification using fragmented connected-component contours
18. Spitta P Johann Sebastian Bach. Leipzig: Breitkopf und Härtel, 2 vols. (1873–1880)

Masahiro Niitsuma is currently an assistant professor at Ritsumeikan University Japan, working on sound language processing and music informatics research. Before he came to Ritsumeikan, he received a BS in Information Technology and a MS in Computer Science from Keio University Japan, and Ph.D in Creative Arts from Queen’s University Belfast. He has published numerous articles on computational analysis and Baroque music and given papers at conferences. His articles are also published in such as *Journal of the Information Processing Society of Japan*; *Journal of the Information Processing & Management*. He is interested in more expert-oriented data mining from musical performance and music manuscripts, from both academic and musical perspectives to achieve ‘deeper’ knowledge discovery.

Lambert Schomaker (1957) is a full professor in Artificial Intelligence at the University of Groningen since 2001. His main interest is in pattern recognition and machine learning problems. He has contributed to over 140 peer-reviewed publications. His work is cited in 23 patents. In recent years his focus is on continuous learning systems and bootstrapping problems, where learning starts with very few examples. He has constructed a search engine, Monk, for historical handwritten documents in Western, Chinese and other script styles, containing hundreds of books and documents and rapidly growing. He is senior member of IEEE, member of the IAPR and is a member of a number of Dutch research programme committees in e-Science (NWO), Computational Humanities (KNAW), Computational science and energy (Shell/NWO/FOM). He is the chairman of the international Unipen foundation for benchmarking of pattern-recognition systems.

Jean-Paul van Oosten received the MSc degree cum laude in artificial intelligence from the University of Groningen, The Netherlands, in 2010. He received the IAPR Best (Oral) Paper award at the ICFHR 2012 and 2014 conferences and is currently a PhD student at the Institute of Artificial Intelligence and Cognitive Engineering in Groningen.

Yo Tomita is Professor of Musicology at School of Creative Arts, Queen's University Belfast, Northern Ireland, and Senior Fellow of the Bach-Archiv Leipzig. He received his doctorate from the University of Leeds in 1991 with a dissertation on the sources of Bach's *The Well-Tempered Clavier*, part II. He has published widely in Bach studies, from those pursuing to identify Bach's compositional choices and decisions as manifested in Bach's own scores to the reception history of Bach's music in the late 18th and early 19th centuries. His recent publication includes *Bach: The Baroque Composers* (Fahnhan: Ashgate, 2011) and *Exploring Bach's B-minor Mass* (Cambridge University Press, 2013), and is currently working on a two-volume monograph *The Genesis and Early History of Bach's Well-tempered Clavier, Book II: a composer and his editions, c.1738-1850* (Ashgate), and *The Cambridge Bach Encyclopedia* (Cambridge University Press).

David Bell graduated from Queen's University, Belfast in 1969 in Pure Mathematics, and has three research degrees in Computing topics. He became a full professor 1986. He has worked in Queen's University, Belfast since 2002, where he was the Director of Research in Knowledge and Data Engineering, a group focussing on research in Soft Computing until 2011, and he is now Professor Emeritus and Visiting Research Professor.

He has produced several hundred publications, and supervised nearly 40 PhDs to completion. He was prime investigator on many national projects and on about 18 EU-funded projects (eg MAP, ESPRIT, DELTA, COST, and AIM) in IT in the eighties and nineties. His activities have included: PC chairmanships (eg joint Programme Committee Chair of VLDB'93 and ICDE'97), conference PC memberships and journal editing/editorial board memberships eg *Computer Journal* and North-Holland's 'Information Systems'. He has been guest editor/co-editor of well-known journals such as IEEE 'Transactions on Knowledge and Data Engineering', on (eg) *Knowledge Discovery*, *Data Driven Data Mining* and *Semantic Web*. He is Author/Co-author of several books, including 'Distributed Database Systems' published by Addison-Wesley in 1992. He served on the UK Technology Foresight Panel for several years, and has been a member of a number of national and international advisory and funding groups, including, for example, a sub-group of the Wellcome Trust in the UK. His research interests are centred on data and knowledge management — the linking of reasoning under uncertainty, machine learning, and other artificial intelligence techniques with database work. This involves the exploration of other topics in computing, including, aspects of agent awareness and innateness.