Available online at www.sciencedirect.com
SCIENCE DIRECT®
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 © 2004 Elsevier B.V. All rights reserved.
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.