A Computational Model for Discriminating Music Performers

Efstathios Stamatatos

Austrian Research Institute for Artificial Intelligence
Schottengasse 3, A-1010 Vienna
stathis@ai.univie.ac.at

Abstract

In this study, a computational model that aims at the automatic discrimination of different human music performers playing the same piece is presented. The proposed model is based on the note level and does not require any deep (e.g., structural or harmonic, etc.) analysis. A set of measures that attempts to capture both the style of the author and the style of the piece is introduced. The presented approach has been applied to a database of piano sonatas by W.A. Mozart performed by both a French and a Viennese pianist with very encouraging preliminary results.

1 Introduction

Studying music performance is one of the most active research areas in computational musicology. Various empirical approaches attempt to model the performance of musical pieces by human experts based mainly on elementary structure analysis of music [1], [2]. Little attention has been paid so far to the development of computational tools able to discriminate between music performers without any external assistance. To the best of our knowledge there is no published study dealing with this subject.

However, the music performer identification problem offers a good testing ground for the development of computational musicology theories since it is a well defined task where the results of a given approach can be evaluated objectively. Moreover, different approaches can be compared by applying them to the same data and reliable conclusions regarding the accuracy of each approach can be extracted. On the other hand, the conclusions drawn by a performer identification study can be taken into account in the designing of other, more practical and useful, tools that try to solve traditional problems.

In this study we try to answer the following questions:

- Are the differences and similarities between different music performers computationally traceable?
- What level of analysis is required for extracting reliable classification results?
- What are the measures that best distinguish between different music performers?
- Can the existing theories of music performance be useful in the development of a performer identification system?

In this paper, a set of parameters that try to capture the stylistic properties of a given performance of a musical piece is introduced. The main idea is that information for both the performance and the musical piece itself should be taken into account. Thus, in addition to parameters dealing with the deviation of the human performer from the score in terms of timing, articulation, and dynamics, the proposed set contains piece-dependent parameters that attempt to represent the stylistic properties of the musical piece. The existing KTH set of generative rules for music performance [3], [4] is used for providing the piece-dependent information that, in essence, includes the deviations of a machine-generated performance from the score.

The proposed approach is based on the note level and does not require any deep (e.g., structural, harmonic, etc.) analysis. Experiments on a database of piano sonatas by W.A. Mozart, performed by both a French and a Viennese pianist, show that the presented tool is able to distinguish accurately between them.

The rest of this paper is as follows: Section 2 describes the proposed model in detail. Section 3 includes the experimental results while in Section 4 the conclusions drawn by this study are given and future work directions are proposed.

2 The Proposed Model

In order to quantify the performance of a musical piece, the relative distance between the performance and the score, in terms of timing, articulation and dynamics, is used. Given two discrete vectors of values $x=\{x_1,..., x_n\}$ and $y=\{y_1,..., y_n\}$, the relative distance $D(x, y)$ between them as used in this paper is defined as follows:
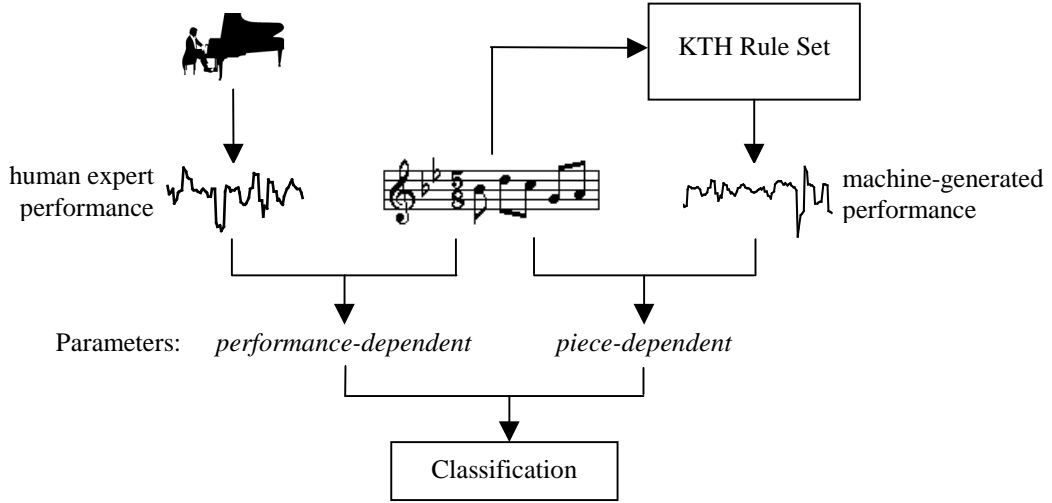
Figure 1. The proposed methodology.

$$D(x, y) = \frac{\sum_{i=1}^{n} \frac{(x_i - y_i)}{x_i}}{n}$$

The three *performance-dependent* parameters used in this study, which correspond to deviations in terms of timing, articulation, and dynamics, respectively, are following:

- $D(IOI_{nominal}, IOI_{measured})$ timing
- $D(IOI_{nominal}, OTD_{measured})$ articulation
- $D(SL_{nominal}, SL_{measured})$ dynamics

where $IOI_{nominal}$ is the nominal Inter-Onset Interval, extracted from the score, and $SL_{nominal}$ is the default Sound Level, while $IOI_{measured}$, $OTD_{measured}$, and $SL_{measured}$ is the inter-onset interval, the Off-Time Duration and the sound level, respectively, as measured in the actual performance. It has to be noted that only the soprano voice is taken into account. Note also that the off-time duration of a note n_i is defined as the difference between the offset of n_i and the onset of n_{i+1}. Recent studies show that the relative amount of staccato for one tone is independent from the IOI [5], [6]. However, the distance of OTD from IOI is quite effective for discriminating between performers (see Section 3).

The values of the above parameters usually depend on the characteristics of the musical piece. For providing the classifier with appropriate information about the stylistic properties of the piece, a set of similar measures that are obtained by a machine-generated performance is introduced. To this end, we use a subset of the well-known KTH set of generative rules for music performance [3], [4], [7]. In more detail, only the rules that can be applied on the note level and do not require any special analysis (e.g., phrase boundary detection, harmonic analysis, etc.) are used. The rules employed in this study are given in Table 1.

KTH-rule	Affected performance variables
Durational Contrast	IOI, SL
Double Duration	IOI
High Loud	SL
Leap Articulation	OTD
Leap Tone Duration	IOI
Faster Uphill	IOI
Repetition Articulation	OTD
Duration Contrast Articulation	OTD
Punctuation	IOI, OTD

Table 1. The KTH rules that have been employed in this study (k=1 for all the rules).

The machine-generated performance is compared with the score and the following *piece-dependent* parameters are obtained:

- $D(IOI_{nominal}, IOI_{rule})$ timing
- $D(IOI_{nominal}, OTD_{rule})$ articulation
- $D(SL_{nominal}, SL_{rule})$ dynamics

where the IOI_{rule}, OTD_{rule}, and SL_{rule} are the inter-onset interval, the off-time duration and the sound level, respectively, as measured in the rule-generated performance.

Thus, for each performance of a musical piece a vector of six parameters is extracted. This vector can then be processed by a standard classification method to obtain the most likely performer. The proposed methodology is illustrated in Figure 1.

3 Experiments

The ideal testing ground for the presented approach would be a database of enough musical pieces performed several times by many human experts with different musical styles. The available database

Parameters included	Guess Actual	Entremont	Batik	Total samples
Perfromance-dependent parameters only	Entremont	**33**	1	34
	Batik	5	**38**	43
Performance-dependent and piece-dependent parameters	Entremont	**32**	2	34
	Batik	1	**42**	43

Table 2. Confusion matrix for the cross validated data. Comparable results for using performance-dependent parameters only and the entire set of parameters. Correct guesses are in boldface.

that best matches these requirements is a collection of piano sonatas by W.A. Mozart performed by Philippe Entremont and Roland Batik in machine-readable form. Specifically, the database we used includes parts of the sonatas KV 279, 280, 281, 282, 283, 284, and 333 played by both pianists. Each sonata movement has been divided in sections and repetitions manually provided in total 34 samples for Entremont and 43 samples for Batik[1]. Moreover, each sample has been matched against the score [2].
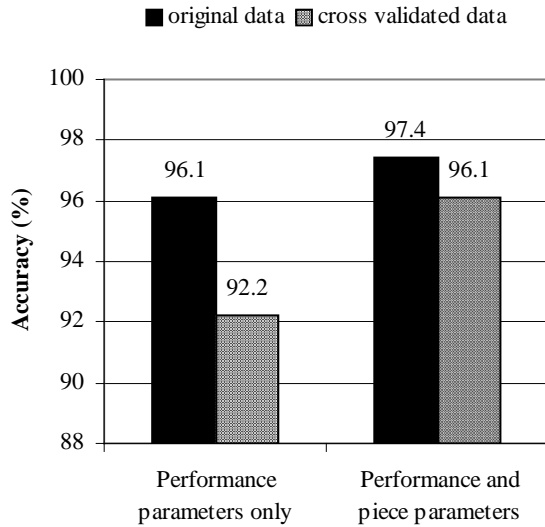

Figure 2. Accuracy of the proposed model. Comparative results for performance-dependent parameters only and the entire set of parameters.

The proposed methodology has been applied to this data set providing a six-parameter vector for each sample. Then, *discriminant analysis*, a standard technique of multivariate statistics [8], has been used to classify the produced vectors. The data then were cross validated, that is, each sample was considered as unseen case and classified based on the remaining samples (i.e., leave-one-out methodology). The results of the classification procedure are given in the confusion matrix of Table 2. The corresponding

classification results when only the performance-dependent parameters are taken into account are given as well. The total classification accuracy for both the original and the cross validated data is given in Figure 2. Note that the original data columns refer to the application of the classification model to the training data (i.e., no unseen cases). It is clear that the performance-dependent parameters alone can give quite reliable results. However, there is a significant improvement when the piece-dependent parameters are included in the parameter vector.

Parameter	$|t|$ value
$D(IOI_{nominal}, IOI_{measured})$	2.883
$D(IOI_{nominal}, OTD_{measured})$	8.823
$D(SL_{nominal}, SL_{measured})$	7.951
$D(IOI_{nominal}, IOI_{rule})$	1.321
$D(IOI_{nominal}, OTD_{rule})$	1.731
$D(Sl_{nominal}, SL_{rule})$	2.245

Table 3. Absolute *t* values for both performance-dependent and piece-dependent parameters.

In order to explore the contribution of each parameter to the classification model, we applied *linear regression analysis* and obtained the *t* values for each parameter. The absolute *t* value is an indication of the importance of the parameter. The higher the absolute *t* value, the more important the contribution of the parameter to the classification model. The results are given in Table 3 and confirm the results of the Table 2 since the performance-dependent parameters proved to be the most significant ones. In more detail, the articulation and the dynamics parameters seem to be the ones that contribute the most to the classification model. From the piece-dependent parameters, the dynamics parameter seems to be the most significant.

Moreover, for giving an indication to the reader as concerns the differences between the two pianists in terms of the used parameters, Table 4 shows an interpretation of the standardized coefficients of the regression function. Thus, Entremont's performances

---

[1] There are more samples for Batik than Entremont since more repetitions of some sections were available for the former.

Parameter	Entremont	Batik
Timing	+	−
Articulation	+	−
Dynamics	−	+

Table 4. An interpretation of the standardized regression coefficients illustrating the differences between the two pianists.

are usually characterized by a higher average deviation of timing and articulation, and a lower average deviation of dynamics than Batik's performances. In other words, the greater the average deviation of timing and articulation and the lower the average deviation of dynamics, the more likely for Entremont to be the performer.

Parameter	$|t|$ value
$D(IOI_{nominal}, IOI_{measured})$	2.721
$D(IOI_{nominal}, OTD_{measured})$	7.461
$D(SL_{nominal}, SL_{measured})$	4.407
$D(IOI_{nominal}, IOI_{rule\_dc})$	0.847
$D(SL_{nominal}, SL_{rule\_dc})$	0.962
$D(IOI_{nominal}, IOI_{rule\_dd})$	0.496
$D(SL_{nominal}, SL_{rule\_hl})$	0.597
$D(IOI_{nominal}, OTD_{rule\_la})$	0.037
$D(IOI_{nominal}, IOI_{rule\_ltd})$	0.043
$D(IOI_{nominal}, IOI_{rule\_fu})$	0.476
$D(IOI_{nominal}, OTD_{rule\_ra})$	1.802
$D(IOI_{nominal}, OTD_{rule\_dca})$	0.013
$D(IOI_{nominal}, IOI_{rule\_punc})$	1.539
$D(IOI_{nominal}, OTD_{rule\_punc})$	1.092

Table 5. Absolute t values for both performance-dependent and decomposed piece-dependent parameters.

In the last experiment, the contribution of each KTH rule to the classification model is examined. In this case, only one rule is taken into account for producing the machine-generated performance. The measured parameters correspond to the affected performance variables of the rule under examination. For instance, the durational contrast rule affects both IOI and SL (see Table 1), so two parameters are obtained. This procedure is followed for each rule providing in total eleven new piece-parameters that replace the three old piece-parameters. Linear regression has been applied to the model consisting of the performance-dependent parameters and the new decomposed piece-dependent parameters. The absolute t values for each parameter are given in Table 5. As can be seen, the repetition articulation rule, the punctuation rule, and the durational contrast rule provide the most important piece-dependent parameters. On the other hand, the leap articulation rule, the leap tone duration rule, and the durational contrast articulation rule seem to contribute the least to the classification model.

4 Conclusions

In this paper we presented a computational model for automatically discriminating music performers. The proposed vector that attempts to capture the stylistic properties of the performance consists of both performance-dependent and piece-dependent parameters. These parameters represent average deviations in terms of timing, articulation, and dynamics for the real performance and for a machine-generated performance. Alternative average parameters, e.g., the absolute relative distance, may also contribute significant information and they will be considered in future experiments.

Preliminary results that have been presented are very encouraging since the proposed model succeeded on discriminating between two human experts playing the same piano sonatas. However, the proposed approach has to be tested on various heterogeneous data sets comprising more candidate performers for extracting more reliable results.

The requirements of the presented method are quite limited since it can be applied on the note level and does not involve any computationally-hard analysis. On the other hand, the high importance of the punctuation rule, as suggested by Table 5, is a strong indication that at least structural analysis could improve considerably the classification results. Note that this rule automatically locates small tone groups and marks them with a lengthening of the last note and a following micropause.

Another aspect that has to be examined is the possibility of segmenting a sample into parts of equal length, in notes, and applying the presented methodology to each part rather than the whole sample. In that case, it would be possible to test the proposed model in data sets where only limited training samples are available for each performer.

Acknowledgments

References

[1] Repp, B. 1992. Diversity and Commonality in Music Performance: An Analysis of Timing Microstructure in Schumann's 'Traümerei'. *Journal of the Acoustical Society of America*, 92(5), pp. 2546-2568.

[2] Widmer, G. 2001. Using AI and Machine Learning to Study Expressive Music Performance: Project Survey and First Report. *AI Communications*, 14.

[3] Friberg, A. 1991. Generative Rules for Music Performance: A Formal Description of a Rule System. *Computer Music Journal*, 15(2), pp. 56-71.

[4] Friberg, A. 1995. *A Quantitative Rule System for Musical Performance*. Doctoral dissertation, Royal Institute of Technology, Sweden.

[5] Bresin, R., and Battel, G.U. 2000. Articulation strategies in expressive piano performance. Analysis of legato, staccato, and repeated notes in performances of the Andante movement of Mozart's sonata in G major (K 545). *Journal of New Music Research*, 29 (3), pp. 211-224.

[6] Bresin, R., and Widmer, G. 2000. Production of staccato articulation in Mozart sonatas played on a grand piano. Preliminary results. *Speech Music and Hearing Quarterly Progress and Status Report*, Stockholm: KTH, 4, pp. 1-6.

[7] Friberg, A., Bresin, R., Frydén, L., and Sundberg, J. 1998. Musical Punctuation on the Microlevel: Automatic Identification and Performance of Small Melodic Units. *Journal of New Music Research*, 27(3), pp. 271-292.

[8] Eisenbeis, R., and Avery R. 1972. *Discriminant Analysis and Classification Procedures: Theory and Applications.* Lexington, Mass.: D.C. Health and Co.