2016-6

NeoMI: a New Environment for the Organization of Musical Instruments

Carolien Hulshof
Musical Instruments Museum, Brussels, c.hulshof@mim.be

Xavier Siebert
University of Mons, xavier.siebert@umons.ac.be

Hadrien Melot
University of Mons, hadrien.melot@umons.ac.be

NeoMI : a new environment for the organization of musical instruments

Carolien Hulshof
Musical Instruments Museum, Brussels
c.hulshof@mim.be

Xavier Siebert
Mathematics and Operational Research, University of Mons
xavier.siebert@umons.ac.be

Hadrien Mélot
Computer Science Department, University of Mons
hadrien.melot@umons.ac.be

1. INTRODUCTION

The current system to classify musical instruments, (Hornbostel-Sachs), is conceptually and practically outdated, because it has a reducing effect by only considering morphological features (Weisser et al., 2011). Our research project NeoMI aims at developing a new environment for the organization of musical instruments that takes into account their many aspects. The aim is to develop an environment consisting of an integrated, unhierarchical and flexible tool to organize the musical instruments. Without reducing the complexity and the richness of these multifaceted objects, it includes the manifold aspects of musical instruments into a unique environment. To that end, the system is based on temporary grouping of instruments among their "peers", according to user-based criteria. This allows an important variability in the precision level: it can be used to group instruments according to a single-criterion (such as the presence on the instrument of an anthropomorphic decoration), or to constitute a corpus of very specific instruments (for example, instruments equipped with devices contributing to provide buzzing sounds), or, on the contrary, to constitute a group of similar instruments made by the same maker, at the same place, over time. NeoMI aims at providing a flexible and pertinent tool for managing museum collections, as well as a fruitful and innovative conceptual framework for research. It explores three different axes: (1) the instrument as an artefact (production time and place, maker, morphological features, etc.); (2) the instrument in its social/cultural context; (3) the instrument as a tool for music. In this paper we focus on the latter, and study the sound-based classification (Fourer et al., 2014; Dupont et al., 2010) of one family of instruments: the fiddles, or bowed chordophones.

2. METHODS

To form a sound-based classification of fiddles, many sound recordings of different fiddle types were gathered from libraries, personal archives and online sources. Effort has been made to ensure that fiddles are included with diverse geographic provenances. The recordings were edited in the Musical Instruments Museum using SoundStudio[1] to get smaller samples of 2 to 4 seconds with minimal environmental noise. Representative samples -referred to as the MIM database from now on- have thus been created for the following fiddle types (number of sound samples between parentheses):

Endingidi (10), a one-string spike tube fiddle from the Baganda people in Uganda;
Erhu (14), a two-string spike tube fiddle from China;
Haegum (9), a two-string spike tube fiddle from Korea;
Hardingfele (20), a folk violin with 4 playing strings and 4 sympathetic strings from Norway;
Imzad (15), a one-string spike bowl fiddle from the Touareg people in Northern Africa
Izeze (17), a spike fiddle from the Wagogo people in Tanzania with one to four strings;
Kamanche (9), a spike bowl fiddle from Iran with four strings;
Kiiki (31), a half-spike bowl fiddle with one string from Chad;
Mamokhorong (10), a one-string fiddle with a tin can resonator from Lesotho;
Masenqo (11), a one string spike fiddle with a rhombus-shaped resonator from the Amhara in Ethiopia;
Morin khuur (18), a two-string fiddle with a horsehead scroll from Mongolia;
Njarka (15), a one-string spike bowl fiddle from the Songhay people in Mali;
Orutu (10), a spike tube fiddle with one string from the Luo people in Kenya;
Ruudga (10), a one-string spike bowl fiddle from the Mossi people in Burkina Faso;
Sarangi (9), an classical Indian fiddle with three playing strings and up to 35-37 sympathetic strings.

The timbre of the MIM instruments was studied using a set of 22 sound features from MirToolbox (Lartillot et al., 2008). Two other databases were also used to test the relevance of the proposed methods as well as to select a subset of discriminating features:

1. MIS: recorded in standardized conditions by the Electronic Music Studios of the University of Iowa, USA[2].

2. PHIL: recorded by musicians from the Philharmonic Orchestra of London, UK[3].

[1] http://felttip.com/ss/

[2] http://theremin.music.uiowa.edu/MIS.html

[3] http://www.philharmonia.co.uk/explore/make_music

Several classification algorithms (K-nearest neighbors (kNN), naïve Bayes, Support Vector Machines (SVM)) were applied to each database.

We started with the MIS and PHIL databases, for which 30% of the sounds were used as a test set to estimate the percentage of correct classifications, while the other 70% were used as a training set.

Those results were compared with a complete exploration of all the combinations of 22 features from MirToolbox: a set of 13 MFCC coefficients, centroïd, spread, skewness, kurtosis, brightness, flatness, entropy, roll frequencies, and the mean of the signal's envelope. This feature selection allowed us to select a subset of 14 features that gives a better classification performance.

Afterwards, the MIM database (15 fiddle types) was grouped into classes using either all features or the subset of features identified by feature selection. Because the MIM database is too small to allow 30% of the sounds to be kept aside, we performed an n-fold cross-validation, with a stratified scenario to preserve the percentage of samples for each class and n=9, which corresponds to number of samples in the smallest class.

A multidimensional scaling approach was then used to represent the results in two dimensions.

3. RESULTS

3.1 MIS and PHIL databases

The confusion matrices for the MIS and PHIL databases are shown in Figures 1 and 2, respectively, using one representative classifier (kNN with k=3). Confusion matrices with other classifiers (kNN with k=1,5; Naïve Bayes; SVM) are similar.

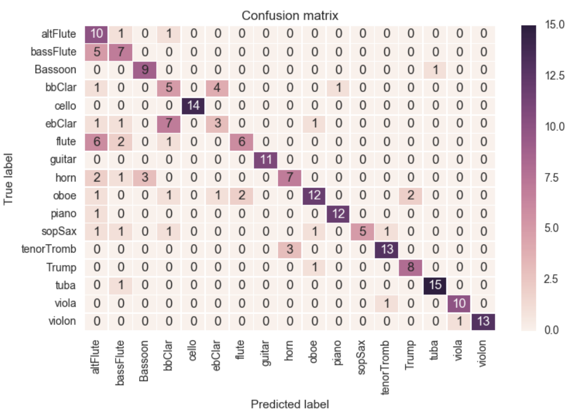

Figure 1. Confusion matrix for the MIS database

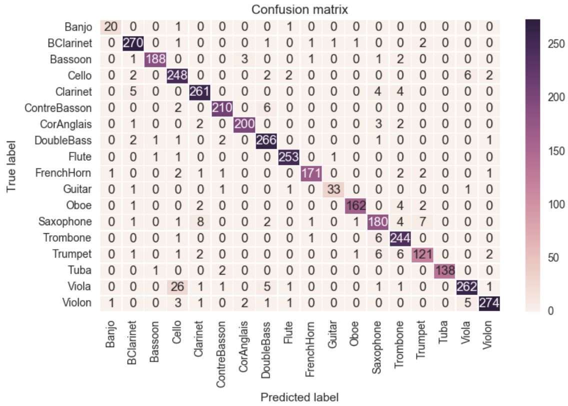

Figure 2. Confusion matrix for the PHIL database

In Figures 1 and 2, the numbers in the diagonal indicate a correct classification, while the off-diagonal ones reflect a confusion between the true and predicted labels.

For the MIS database, the precision is 77%, while the recall is 73%. Some confusion occurs for example among the different types of flutes (altFlute, bassFlute and flute) or among clarinets. This indicates some difficulty to distinguish between instruments of the same family or whose timbre is similar.

For the PHIL database, precision and recall are both around 95%. This reflects the fact that the PHIL database is bigger, but mostly that it contains shorter recordings, each producing a specific note, which simplifies the task of the classifier. Some confusion occurs for example between Cello and Violin, which makes sense considering the proximity of these instruments.

As mentioned in Section 2, these performances have been measured on the test set composed of 30% of the samples.

To improve these results, we performed feature selection, starting from the observation that not all 22 features from MIRTOOLBOX were contributing efficiently to the classification. We thus performed a complete combinatorial analysis to find the best combinations among the 22 descriptors from the MIRTOOLBOX, by comparing the best results obtained with several classifiers: k nearest neighbours (kNN) with k values ranging from 1 to 5, naive Bayes and SVM. The results in Figure 3 show indeed that the classification rate reaches a maximum between 10 to 15 features, before decreasing progressively when increasing the number of features until 22.

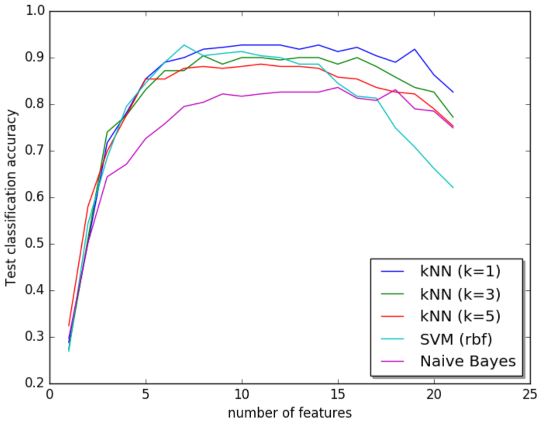

Figure 3. Number of features and accuracy

A study of the frequency of appearance of each feature in the most accurate combinations (i.e., more than 85% accuracy) of features is shown in Figure 4, which shows that features with indices 9 to 14 are less efficient.

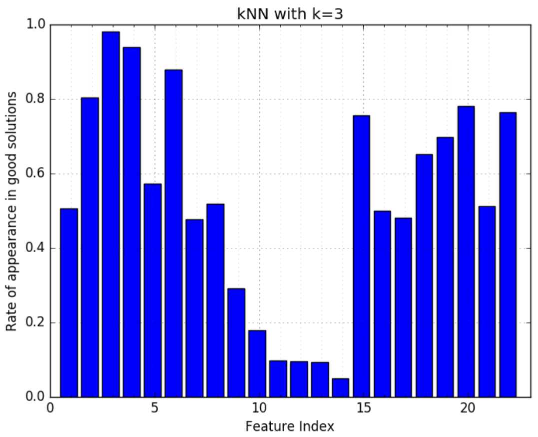

Figure 4. Efficiency of features, measured by the frequency of appearance of each feature in the solutions with more than 85% accuracy in the MIS database.

Removing the features 9 to 14 from the set of features used for the classification leads to the confusion matrices shown in Figures 5 and 6, for the MIS and PHIL databases, respectively.

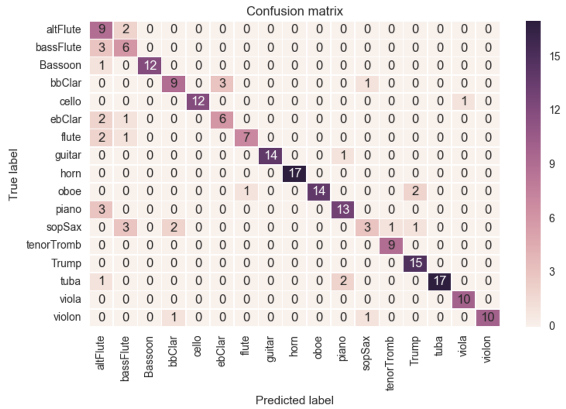

Figure 5. Confusion matrix for the MIS database, with a subset of features.

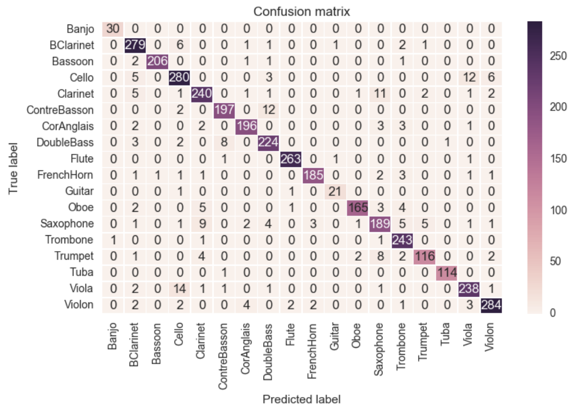

Figure 6. Confusion matrix for the PHIL database, with a subset of features.

For the MIS database, the precision has now increased to 86%, and the recall to 84%. However, for the PHIL database, the precision and recall remain stable around 94%.

The slight variations in the PHIL database upon feature selection (95% to 94%) are probably caused by the fact that a different subset of 30% of sounds is chosen each time.

3.2 MIM database

The confusion matrix for the MIM database is shown in Figure 7, with one representative classifier (kNN with k=3).

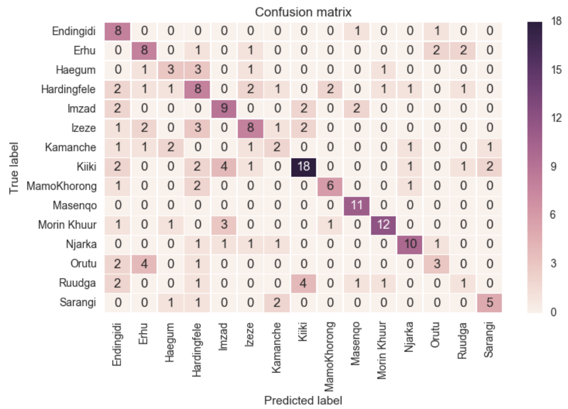

Figure 7. Confusion matrix for the MIM database

Considering the proximity of the instruments involved - the fiddle family- it is not surprising that the confusion matrix is less accurate than for the MIS and PHIL databases.

Some tendencies can be extracted but have to be interpreted with caution. For example, the Kiiki family seems to be fairly homogeneous. However, it is also the most populated (31 instruments), which has a tendency to bias the classification by attracting other instruments (such as Imzad, Izeze or Ruudga) in this category. Another class that appears quite homogeneous is Masenqo. Endingidi, on the contrary, has a high recall (most Endingidi have indeed been classified as Endingidi) but a low precision (several instruments from the Hardingfele, Imzad, Izeze, Kamanche, Kiiki, Mamokhorong, Orutu and Ruudga types have been misidentified as Endingidi).

We also tried the feature selection to classify the MIM sounds with the subset of features, giving us a confusion matrix as shown in Figure 8.

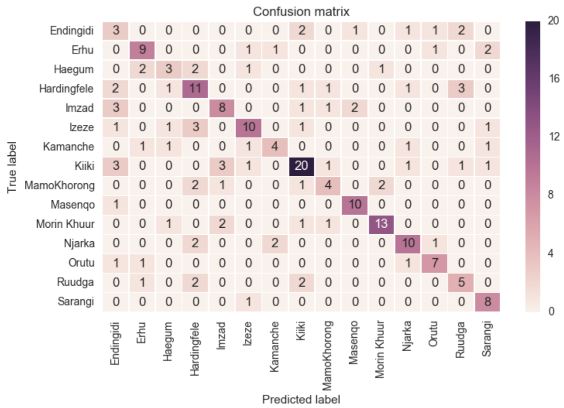

Figure 8. Confusion matrix after feature selection

The new confusion matrix shows a slight overall improvement; all fiddle types have a higher recall, except Endingidi and Mamokhorong.

To visualize and to be able to interpret the results, we computed the distance matrices between predicted classes of instruments (Figure 9), and represented them using a multidimensional scaling (MDS) approach (Cox et al., 2000), as shown in Figure 10.

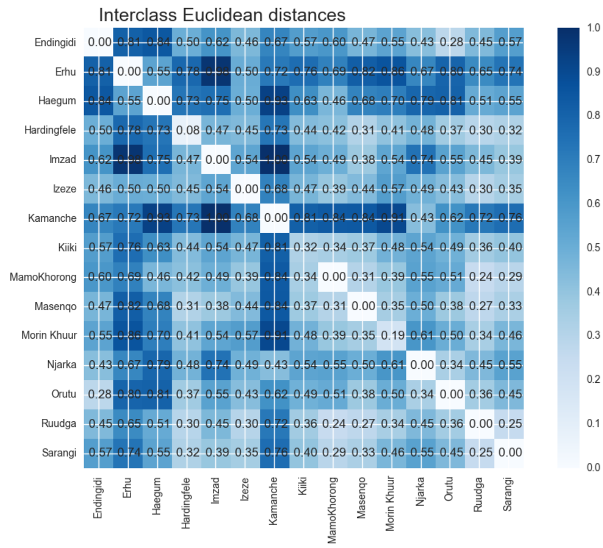

Figure 9. Interclass Euclidian distances

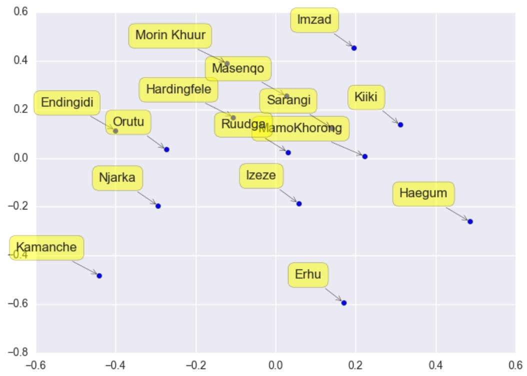

Figure 10. Distances between fiddle types

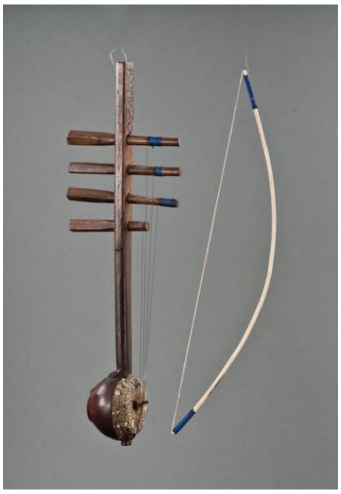

Figure 12. Tanzanian izeze, mim inv. 2014.273.001. © mim, photo Simon Egan

4. DISCUSSION

A visual representation using an MDS approach leads to some interesting questions. For example, based on their morphology and geographic distribution one would not expect a close proximity between the Indian Sarangi (Figure 11) and the Tanzanian Izeze (Figure 12):

Figure 11. Indian sarangi, mim inv. 1972.003. © mim, photo Simon Egan

However, as shown in Figure 10, these two fiddle types are quite close to each other. This leads to new questions: is it because they both possess sympathetic strings? Does the playing technique play a role in their similarity? Another question arises when looking at the Imzad, a fiddle from the Touareg people in Northwest Africa (Figure 13), and the Njarka, a fiddle from the Songhai people in Mali; both are single string spike fiddles with a calabash resonator, played with a horsehair bow (Figure 14):

Figure 13. Touareg imzad, mim inv. 2009.002. © mim, photo Simon Egan

Figure 14. Njarka from Mali, RMCA inv. MO.1967.63.777. © RMCA Tervuren

However, apparently there are certain qualities that make them appear far from each other in Figure 10. How can we explain this distance? Not all distances between the different fiddle types are surprising, though - to the human ear, the Endingidi and Orutu sound very much alike, and they are indeed quite close to each other in the graph in Figure 10.

5. CONCLUSION

Confusion matrices show that a classification based on sound features is efficient for two databases (MIS and PHIL) containing various kinds of instruments.

Our results indicate that it is also feasible with the MIM database, containing only various fiddle families.

The interest of the sound-based classification is that it allows us to discover possible new links between certain instruments, for example between different fiddle types, as shown on the visualization using an MDS approach. Furthermore, at the dawn of the 21st century, the persistent use of a conceptual framework designed in the 19th century is a problem. Indeed, classificatory systems are not a mere way to sort objects: they are also (and often implicitly) a conceptual ground and a basis for research. The NeoMI project aims therefore to induce an important change of scientific paradigm: from a linear thought to a truly multidimensional one, in which the relative importance of features is adjusted according to the needs of the research.

6. REFERENCES

Cox T.F. and Cox M.A.A. (2000). *Multidimensional scaling.* CRC Press.

Dupont S., Frisson C., Siebert X., and Tardieu D. (2010). Browsing sound and music libraries by similarity. *128th Audio Engineering Society (AES) Convention*, London UK.

Fourer D., Rouas J.-L., Hanna P. and Robine, M. (2014). *Automatic timbre classification of ethnomusicological audio recordings*, International Society for Music Information Retrieval Conference (ISMIR)

Lartillot O., Toiviainen P., and Eerola T. (2008). A matlab toolbox for music information retrieval. *Data Analysis, Machine Learning and Applications*. Berlin: Springer, pp. 261-268.

Weisser, S. and Quanten, M. (2011). *Rethinking Musical Instrument Classification: Towards a Modular Approach to the Hornbostel-Sachs System*, Yearbook for Traditional Music 43, pp. 122—146.