Cocco, C. , Bavaud, F. : *Correspondence Analysis, Cross-Autocorrelation and Clustering in Polyphonic Music*. In: Lausen, B. et al. (Eds.) Data Science, Learning by Latent Structures, and Knowledge Discovery, pp. 401–410. (Series: Studies in Classification, Data Analysis, and Knowledge Organization). Springer, Heidelberg (2015)

Correspondence Analysis, Cross-Autocorrelation and Clustering in Polyphonic Music

Christelle Cocco[1] and François Bavaud[2]

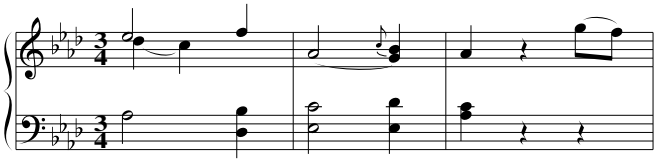
[1] University of Lausanne christelle.cocco@unil.ch
[2] University of Lausanne francois.bavaud@unil.ch

Abstract. This paper proposes to represent symbolic polyphonic musical data as contingency tables based upon the duration of each pitch for each time interval. Exploratory data analytic methods involve weighted multidimensional scaling, correspondence analysis, hierarchical clustering, and general autocorrelation indices constructed from weighted temporal neighborhoods. Beyond the analysis of single polyphonic musical scores, the methods sustain inter-voices as well as inter-scores comparisons, through the introduction of ad hoc measures of configuration similarity and cross-autocorrelation. Rich musical patterns emerge in the related applications, and preliminary results are encouraging for clustering tasks.

1 Introduction

This paper aims to produce an exploratory data analysis of symbolic polyphonic musical data represented as contingency tables, which count the duration of each pitch for each time interval, given a predefined partition of the musical score into equal durations. This representation, not so far from the piano-roll representation or from the Chroma representation for audio files (see e.g. Müller and Ewert (2011) or Ellis and Poliner (2007)), has the advantage of representing digital polyphonic music, being usable with common data analytic methods, such as correspondence analysis and being aggregation-invariant (Section 2).

In Section 3.1, analyses of whole music pieces are proposed, by means of correspondence analysis and a flexible autocorrelation index able to deal with general neighborhoods. Both methods grasp intrinsic structures of musical scores and provide pattern visualizations. Multiple voices within a single musical score are analyzed through soft multiple correspondence analysis and a cross-autocorrelation index (Section 3.2). Finally, based on the choice of the contingency table, a similarity measure, aimed to cluster music pieces according to composers, is proposed and illustrated (Section 3.3).



	Time intervals t with $\tau = \quarternote$								
pitches j	29	30	31	32	33	34	35	36	37
0	0	4	0	4	4	0	4	0	0
1	4	0	4	0	0	4	0	0	0
3	4	4	0	4	4	4	0	0	0
5	0	0	4	0	0	0	0	0	2
7	0	0	0	0	0	4	0	0	2
8	4	4	0	4	4	0	8	0	0
10	0	0	4	0	0	4	0	0	0
z	0	0	0	0	0	0	0	4	0

	Time intervals t with $\tau = \dotted{\halfnote}$		
pitches j	11	12	13
0	4	8	4
1	8	4	0
3	8	12	0
5	4	0	2
7	0	4	2
8	8	8	8
10	4	4	0
z	0	0	4

Fig. 1. Transposed display of the contingency table $X = (x_{tj})$, giving the duration of each pitch (in units of sixteenth note), for τ equal to a quarter note (top) and to a dotted half note (bottom). Extract of the *3rd movement of the Beethoven's Piano Sonata No. 1 in F minor, Op. 2, No. 1.*

2 Data Representation

In this contribution, symbolic music files are used, and especially files in *Humdrum* **kern format, as they are well structured with all voices, independent of the performer and freely available on the web (`http://kern.ccarh.org/`). Moreover, *Humdrum extras* (`http://extra.humdrum.org/`) are used when modifications, such as transposition, are needed, as well as to transform **kern files in *Melisma* format (`http://www.link.cs.cmu.edu/music-analysis/`), easily handleable for the representation proposed in this paper. Note that the representation proposed in the followings could have also been obtained with other digital files, such as ABC or MIDI files, especially if the latter is performed with a constant tempo.

Each musical score is represented, with all repeated passages, as a contingency table $X = (x_{tj})$ crossing *pitches* ($j = 0, ..., m$) and *time intervals* ($t = 1, ..., n$). The table gives the duration of each pitch in each time interval. Notice that the repetition of notes of the same pitch within a time interval is not coded. In more detail, MIDI note numbers (1 to 128) are transformed in a 12-note octave-equivalent pitch set using a modulo 12, where 0 stands for C; 1, for C♯ or D♭; 2, for D; etc. Moreover, a *true rest z* is added whenever no

note is played. Thus, j can take on 13 different values: 0 to 11 and z. Regarding time intervals, each one has a constant *duration* of τ which can take any value, such as a sixteenth note, a measure or a number of milliseconds. Consequently, the total duration of the musical score is $\tau_{\text{tot}} = n\tau$. An example of the transposed contingency table is given in Figure 1 for two different values of τ.

Besides the advantage to deal with polyphonic music, this representation is aggregation-invariant in the sense that doubling τ amounts to summing counts within two consecutive parts. So, considering an interval T made out of smaller intervals t, the new counts are $\tilde{x}_{Tj} = \sum_{t \in T} x_{tj}$. Lavrenko and Pickens (2003) and Morando (1981) use a quite similar representation, except that the former do not take into account the duration of and between notes and the latter bases his representation upon the succession of chords. However, in contrast to the present representation, theirs are not aggregation-invariant.

Then, as a second step, the contingency table $X = (x_{tj})$ is normalized to $\Xi = (\xi_{tj})$ in order that the sum of each row $\sum_j \xi_{tj} = \xi_{t\bullet}$ equals to 1, that is $\xi_{tj} = \frac{x_{tj}}{x_{t\bullet}}$. Thus, the same importance is given to each time interval, regardless of the duration and the number of pitches.

3 Methods and Applications

3.1 Single Score Analysis

Correspondence Analysis

To perform the correspondence analysis (CA) on the Ξ matrix, an equivalent method is used which consists in applying a weighted multidimensional scaling on the chi-squared dissimilarities between time intervals $\hat{D} = (\hat{D}_{st})$ and between pitches $\check{D} = (\check{D}_{ij})$:

$$\hat{D}_{st} = \sum_j \rho_j (q_{sj} - q_{tj})^2 \qquad \check{D}_{ij} = \sum_t f_t (q_{ti} - q_{tj})^2 \tag{1}$$

where $f_t = 1/n$ is the relative weight of time intervals, $\rho_j = \xi_{\bullet j}/n$ is the relative weight of pitches and $q_{tj} = \xi_{tj} n / \xi_{\bullet j}$ is the independence ratio.

In a nutshell, scalar products between time intervals $\hat{B} = (\hat{b}_{st})$ and between pitches $\check{B} = (\check{b}_{ij})$ are computed from the dissimilarity matrices as:

$$\hat{B} = -\frac{1}{2} H^f \hat{D} (H^f)' \qquad \check{B} = -\frac{1}{2} H^\rho \check{D} (H^\rho)'$$

where $H^f = I - \mathbf{1} f'$, $H^\rho = I - \mathbf{1}\rho'$ are the corresponding centering matrices. Then, weighted scalar products $\hat{K} = (\hat{k}_{st})$ and $\check{K} = (\check{k}_{ij})$ are defined as:

$$\hat{k}_{st} = \sqrt{f_s f_t}\, \hat{b}_{st} \qquad \check{k}_{ij} = \sqrt{\rho_i \rho_j}\, \check{b}_{ij} \tag{2}$$

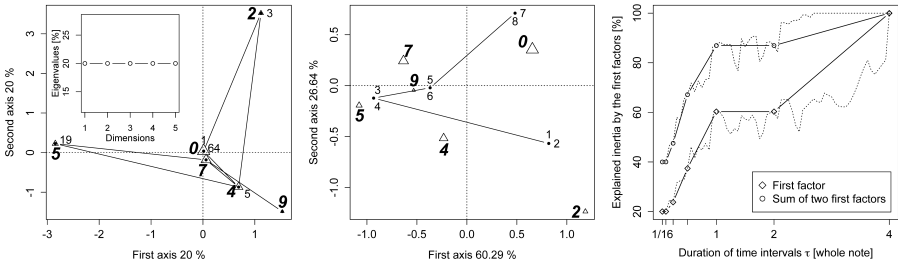


Fig. 2. CA on *Are you sleeping? in C major*. Left: scree graph and biplot with τ equal to an eighth note. Triangles with large-sized figures in italic stand for pitches (triangle size is proportional to the quantity of the pitch in the music piece) and full circles, sometimes with small-sized figures, represent time intervals and are linked in consecutive order according to the time progression. Middle: biplot with τ equal to a measure. Right: explained inertia by the (two) first factors according to τ. Dotted lines represent results for all durations and solid lines stand for results for integer divisors of τ_{tot}.

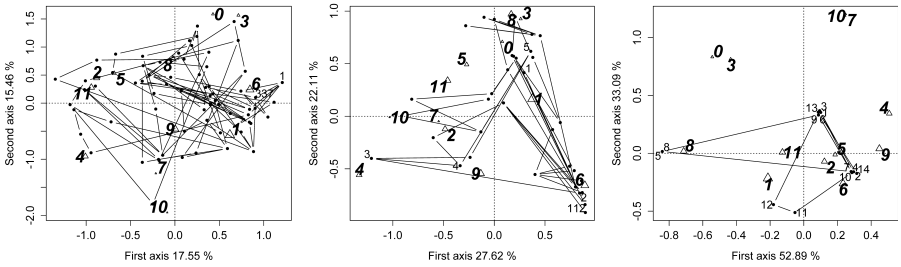


Fig. 3. CA on *Mazurka Op. 6, No. 1 in F♯ minor by Chopin*. Left: biplot with with τ equal to a quarter note. Middle: biplot with with τ equal to a measure. Right: biplot with with τ equal to eight measures.

The spectral decomposition of the matrix $\hat{K}$ (respectively $\check{K}$) provides the eigenvectors $u_{t\alpha}$ (resp. $v_{j\alpha}$) and the corresponding eigenvalues λ_α (identical for both matrices) from which stem the factor coordinates for time intervals ($x_{t\alpha}$) and for pitches ($y_{j\alpha}$):

$$x_{t\alpha} = \frac{\sqrt{\lambda_\alpha}}{\sqrt{f_t}} u_{t\alpha} = \frac{1}{\sqrt{\lambda_\alpha}} \sum_{j=1}^{m} \rho_j q_{tj} y_{j\alpha} \qquad y_{j\alpha} = \frac{\sqrt{\lambda_\alpha}}{\sqrt{\rho_j}} v_{j\alpha} = \frac{1}{\sqrt{\lambda_\alpha}} \sum_{t=1}^{n} f_t q_{tj} x_{t\alpha}$$

An example of this formalism for the well-known French monophonic nursery melody *Frère Jacques* (*Are you sleeping?* in English) is given in Figure 2. The graph on the left shows the result obtained with τ equal to a eighth note, which means that no more than one pitch is played during each time interval, i.e. the representation is totally monophonic. In that case, chi-squared dissimilarities between time intervals are "star-like", i.e. of the form $\hat{D}_{st} = a_s + a_t$

(see e.g. Critchley and Fichet (1994)). Consequently all λ_α are equal and data are difficult to compress by factor analysis. When τ is equal to a measure (graph in the middle), the graph reveals the structure of the music piece, with each measure played two times. Note the "horseshoe effect" resulting from the temporal ordering of time intervals. The right graph highlights that when increasing the duration τ, the percentage of explained inertia climbs, except when τ is smaller than or equal to a eighth note, the smallest duration of a note, and between τ equal to a whole note (corresponding to a measure) and equal to two whole notes, due to the repeated structure of the piece.

Another example is given in Figure 3 for a Mazurka by Chopin, with three different interval durations. The structure emerges more clearly for large values of τ. In particular, the right graph, with τ equal to eight measures, reveal the similar (e.g. 1, 3, 6, 9 and 13) and different passages (e.g. 2 against 3).

While these two examples clearly highlight the structure of the piece, results are less comprehensible when a motif is transposed in the same piece or when a true rest appears. In fact, in the latter case, the first factor often exclusively expresses the contrast between true rests and pitches.

Autocorrelation Index

Consider now the neighborhood analysis between ordered time intervals, represented by the rows of Ξ. Temporal neighborhoods can be defined by a non-negative symmetric *exchange matrix* $E = (e_{st})$ obeying $e_{t\bullet} = e_{\bullet t} = f_t = 1/n$. The associated *autocorrelation index* (Bavaud et al. (2012)) is calculated as:

$$\delta := \frac{\Delta - \Delta_{\text{loc}}}{\Delta} \in [-1, 1] \tag{3}$$

where Δ is the (global) inertia and Δ_{loc} is the local inertia:

$$\Delta := \frac{1}{2}\sum_{st} f_s f_t \hat{D}_{st} = \frac{1}{2n^2}\sum_{st}\hat{D}_{st} \qquad \Delta_{\text{loc}} := \frac{1}{2}\sum_{st} e_{st}\hat{D}_{st} \tag{4}$$

Thus, the autocorrelation index measures the difference between the overall variability of chi-squared interval dissimilarities and the local variability within some neighborhood defined by E, generalizing the usual "immediate left-right neighborhood" (see e.g. Morando (1981) for a musical data-analytic approach). A large positive (resp. negative) autocorrelation means that the pitches distributions are more (resp. less) similar in the neighborhood than in randomly chosen intervals.

Among all possible exchange matrices, it turns out to be convenient to define a *periodic* exchange matrix, with a neighborhood at temporal *distance* (or *lag*) r (right and left) of the current interval, $E^{(r)}$:

$$e^{(r)}_{st} = \frac{1}{2n}[1(t = (s \pm r) \bmod n) + 1((s \pm r) \bmod n = 0) \cdot 1(t = n)]$$

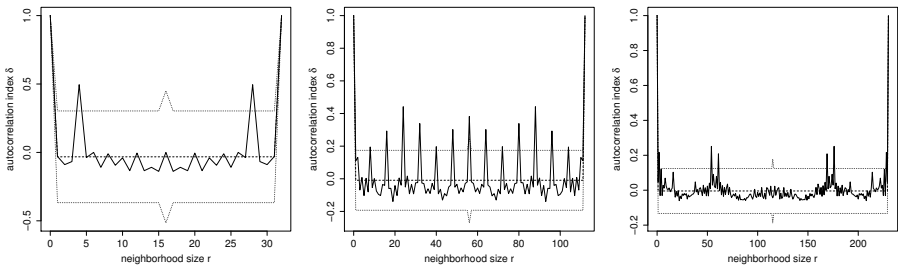


Fig. 4. Autocorrelation index according to the lag r varying from 0 to n (solid line), the expected value (dashed line) and $u_{0.975} = 1.96$ times the standard deviation (dotted line). Left: *Are you sleeping?* with τ equal to a quarter note. Middle: *Mazurka Op. 6, No. 1 by Chopin* with τ equal to a measure. Right: *sonata L. 12 (K. 478) by Scarlatti* with τ equal to a measure.

For statistical testing of the autocorrelation index, see e.g. Cliff and Ord (1981) and Bavaud (2013). Note that, in contrast to the usual autocorrelation function in time series analysis (see e.g. Box and Jenkins (1976)) which considers a single numerical variable, the autocorrelation index can deal with multiple simultaneous categorical variables.

The autocorrelation index is computed on three musical scores (Figure 4). As expected, $\delta = 1$ for $r = 0$ and the figures are symmetric, since the neighborhood is periodic ($E^{(r)} = E^{(n-r)}$). Moreover, noticeable peaks appear in all graphs. For the monophonic music piece *Are you sleeping?*, the highest value ($\delta = 0.495$) appears for $r = 4$ which corresponds to the duration of a measure. In fact, due to the systematic repetition of each measure, at each point the same pitches are played at a distance equal to four, sometimes on the left, sometimes on the right. For the Chopin's piece, peaks occur each eight measures as expected by the results obtained in Figure 3. Finally, for the Scarlatti's sonata, there are two remarkable peaks ($\delta = 0.25$ and $\delta = 0.21$), for $r = 54$ and $r = 61$ measures, corresponding to the length of the two repeated parts of the piece, which compose the whole piece.

3.2 Between Voices Analysis

Soft Multiple Correspondence Analysis

Let Ξ^v denote the row-normalized contingency table for *voice* $v = 1, ..., V$ occurring in a music piece. The complete contingency table of the musical score obtains as $\Xi^{\text{COMP}} = (\Xi^1|\Xi^2|...|\Xi^V)$, on which a CA is carried out. Whereas an usual multiple correspondence analysis (MCA) is computed on a disjunctive table, the present procedure is applied to row cells containing, due to row-normalization, the pitch *proportions* of the voice during a given t, and hence constitutes a soft variant of MCA.

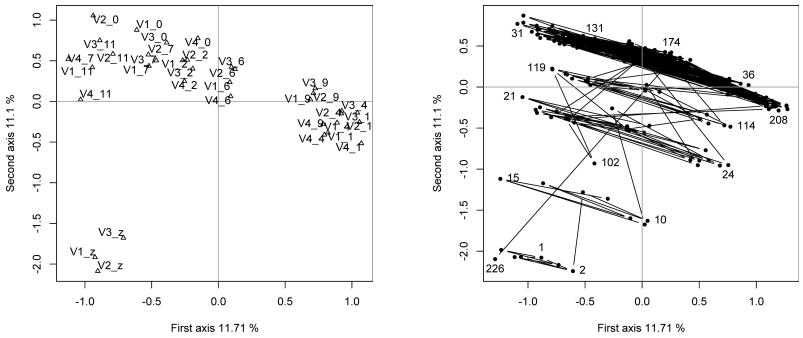


Fig. 5. Soft MCA on the *Canon in D Major by Pachelbel* with τ equal to a quarter note. Left: factor coordinates for the pitches, whose names are preceded by V1 for violin I, V2 for violin II, V3 for violin III and V4 for Harpischord. Right: factor coordinates for the time intervals.

Figure 5 shows the results obtained for the Pachelbel's canon. On the right graph, different zones appear depending on the number of instruments which are playing. For instance, in the bottom zone, only the harpsichord is playing, and so there are true rests for the three violins.

Cross-autocorrelation Index

Define the "raw" coordinates of the voice Ξ^v as ${}^*\xi^v_{tj} = \sqrt{\rho^v_j}(q^v_{tj} - 1)$, with the property that the associated squared Euclidean distances $D_{st} = \sum_j ({}^*\xi^v_{sj} - {}^*\xi^v_{tj})^2$ are equal to the chi-squared distances $\hat{D}_{st}$ of equation (1).

To extend the autocorrelation index to two voices (α and β), one proposes a *cross-autocorrelation* index for multidimensional variables Ξ^α and Ξ^β , which measures the similarity between the pitch distribution of α and the pitch distribution of β within a fixed lag or, more generally, a defined neighborhood, namely:

$$\delta(\Xi^\alpha, \Xi^\beta) := \frac{\Delta(\Xi^\alpha, \Xi^\beta) - \Delta_{\mathrm{loc}}(\Xi^\alpha, \Xi^\beta)}{\sqrt{\Delta(\Xi^\alpha)\Delta(\Xi^\beta)}} \in [-1, 1]$$

In the latter, $\Delta(\Xi^v)$ is the inertia of the voice v (see the first part of (4)), $\Delta(\Xi^\alpha, \Xi^\beta) = \frac{1}{2}\sum_{st} f_s f_t D^{\alpha\beta}_{st} = \sum_s f_s \sum_j {}^*\xi^\alpha_{sj}\,{}^*\xi^\beta_{sj} - \sum_j {}^*\bar{\xi}^\alpha_j\,{}^*\bar{\xi}^\beta_j$ is the cross-inertia between the voice α and the voice β, where $D^{\alpha\beta}_{st} = \sum_j ({}^*\xi^\alpha_{sj} - {}^*\xi^\alpha_{tj})({}^*\xi^\beta_{sj} - {}^*\xi^\beta_{tj})$ is the cross-dissimilarity between two time intervals of two voices, and finally $\Delta_{\mathrm{loc}}(\Xi^\alpha, \Xi^\beta) = \frac{1}{2}\sum_{st} e_{st} D^{\alpha\beta}_{st} = \sum_s f_s \sum_j {}^*\xi^\alpha_{sj}\,{}^*\xi^\beta_{sj} - \sum_{st} e_{st} \sum_j {}^*\xi^\alpha_{sj}\,{}^*\xi^\beta_{tj}$ is the local cross-inertia between voices α and β.

In particular, $\Delta(\Xi, \Xi) = \Delta(\Xi)$ and $\Delta_{\mathrm{loc}}(\Xi, \Xi) = \Delta_{\mathrm{loc}}(\Xi)$, so $\delta(\Xi, \Xi) = \delta(\Xi) = \delta$ given in (3). It must be noticed that this formalism works in this

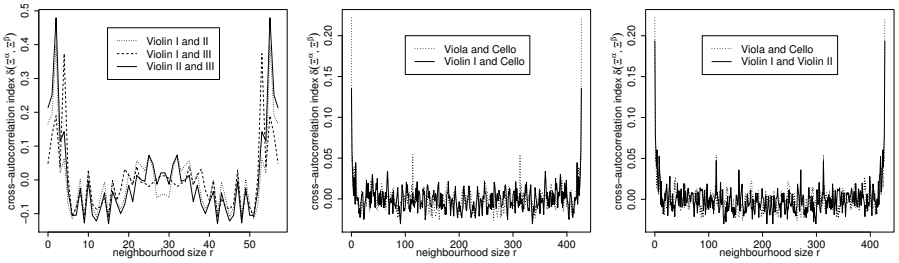


Fig. 6. Cross-autocorrelation index according to the lag r varying from 0 to n. Left: *Canon in D Major by Pachelbel* with τ equal to a measure. Middle and right: *first movement of the String Quartet No. 1 in F major, Op. 18 by Beethoven* with τ equal to a measure.

specific context because $f_t^\alpha = f_t^\beta = f_t = \frac{1}{n}$ due to the normalization of Ξ or Ξ^v and since all voices have the same number of time intervals.

This cross-correlation index is computed on two multiple-voice music pieces with the same exchange matrix as the one proposed for the autocorrelation index (Figure 6). For the Pachelbel's canon, highest peaks on the left graph appear at $r = 2$ for the cross-autocorrelation between violins I and II and between violins II and III, and at $r = 4$ between violins I and III, corresponding to the lag of two or four measures between the starts of each violin. For the Beethoven's string quartet (center and right graphs), peaks at $r = 0$ reveal largest melodic similarities between violin I and violin II on the one hand, and between viola and the cello on the other hand. Moreover, both graphs exhibit large peaks at $r = 114$ measures, corresponding to a repetition in the music piece.

Thus, the cross-autocorrelation index allows the comparison of different voices of a music piece. It can also be implemented to compare two music piece variants. See e.g. Ellis and Poliner (2007), who apply cross-correlation on audio files.

3.3 Between Scores Analysis

To measure the configuration similarity between two musical scores a and b, a weighted dual version of the RV-Coefficient proposed by Robert and Escoufier (1976) is computed:

$$\mathrm{CS}_{ab} = \frac{\mathrm{Tr}(\check{K}^a \check{K}^b)}{\sqrt{\mathrm{Tr}((\check{K}^a)^2)\mathrm{Tr}((\check{K}^b)^2)}}$$

where $\check{K}^a$ (resp. $\check{K}^b$) is the weighted scalar product between pitches of the musical score a (resp. b) as defined in the second part of the equation (2). By construction, the components of $\check{K}^a$ (or $\check{K}^b$) are zero for a pitch absent in

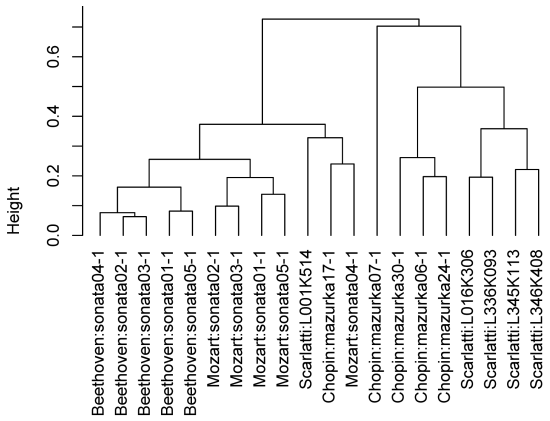


Fig. 7. Hierarchical clustering of 20 music pieces with the Ward aggregation method.

the corresponding musical score. Both $\check{K}^a$ and $\check{K}^b$ depend upon the reference duration τ, chosen as identical for both music pieces.

Define the dissimilarity between two musical scores as $D_{ab} = 1 - \mathrm{CS}_{ab}$. This dissimilarity can be seen as a generalization of the well-known cosine distance (see e.g. Weihs et al.(2007)), and turns out to be squared Euclidean. Usual clustering methods between musical scores, based upon D_{ab}, can in turn be applied.

Figure 7 presents the results obtained with an agglomerative hierarchical clustering on a dataset made up of 20 music pieces written by four composers:

- **Scarlatti:** Sonatas L. 1 (K. 514), L. 16 (K. 306), L. 336 (K. 93), L. 345 (K. 113) and L. 346 (K. 408). They all have a 2/2 time signature.
- **Mozart:** First movement of sonatas n°1, 2, 3, 4 and 5.
- **Beethoven:** First movement of sonatas n°1, 2, 3, 4 and 5.
- **Chopin:** Mazurkas Op. 6 (No. 1), Op. 7 (No. 1), Op. 17 (No. 1), Op. 24 (No. 1) and Op. 30 (No. 1).

For comparison sake, the 20 music pieces are all transposed in C, with a common τ value of one measure. Although the dataset is small, this first result is encouraging, producing well-grouped music pieces with respect to each composer, especially for Beethoven.

4 Conclusion

The present data-analytic treatment of musical scores is based upon two primitives, namely a dissimilarity matrix and a neighborhood matrix between time intervals, defined with respect to a reference duration. It covers and generalizes well-known multi-categorical, factorial and time-series techniques, and is able

to treat polyphonic pieces, as well as performing between-voices and between-scores analyses, with encouraging clustering results. Its modest computational cost makes it amenable to the automatic treatment of large symbolic musical data sets. Furthermore, it allows the consideration of flexible alternatives, both for the dissimilarity matrix (other than the chi-square) and for the exchange matrix (other than periodic neighborhood), deserving further investigation.

So far, exploratory analyses are interpretable in a fairly satisfactory way, although the complex factorial structures exhibited by rich music pieces certainly deserve further attention. In the near-future agenda,within the present formalism, we hope to progress in the automatic detection of τ, motif recognition and large dataset clustering or classification.

References

BAVAUD, F. (2013): Testing Spatial Autocorrelation in Weighted Networks: the Modes Permutation Test. *Journal of Geographical Systems, 15, 233–247.*

BAVAUD, F., COCCO, C. and XANTHOS, A. (2012): Textual autocorrelation: formalism and illustrations. In: *11èmes Journées internationales d'analyse statistique des données textuelles*, 109-120.

BOX, G.E.P. and JENKINS, G.M. (1976): *Time series analysis: forecasting and control.* Holden-Day.

CLIFF, A. D. and ORD, J. K. (1981): *Spatial Processes: Models and Applications.* Pion, London.

CRITCHLEY, F. and FICHET, B. (1994): The partial order by inclusion of the principal classes of dissimilarity on a finite set, and some of their basic properties. In: B. Van Cutsem (Eds.): *Classification and Dissimilarity Analysis.* Springer, New York, 5-65.

ELLIS, D. P. W. and POLINER, G. E. (2007): Identifying 'Cover Songs' with Chroma Features and Dynamic Programming Beat Tracking. In: *IEEE International Conference on Acoustics, Speech and Signal Processing, 2007*, IV-1429–IV-1432.

LAVRENKO, V. and PICKENS, J. (2003): Polyphonic Music Modeling with Random Fields. In: *Proceedings of the eleventh ACM international conference on Multimedia.*, Berkeley, CA, 120–129.

MORANDO, M. (1981): L'analyse statistique des partitions de musique. In: Benzécri, J.-P. et al. (Eds.): *Pratique de l'analyse des données, tome 3: Linguistique et lexicologie*, Dunod, Paris, 507-522.

MÜLLER, M. and EWERT, S. (2011): Chroma Toolbox: Matlab Implementations for Extracting Variants of Chroma-based Audio Features. In: *Proceedings of the 12th International Conference on Music Information Retrieval*, 215-220.

ROBERT, P. and ESCOUFIER, Y. (1976): A Unifying Tool for Linear Multivariate Statistical Methods: The RV-Coefficient. *Journal of the Royal Statistical Society. Series C (Applied Statistics), 25, 257–265.*

WEIHS, C., LIGGES, U., MÖRCHEN, F. and MÜLLENSIEFEN, D. (2007): Classification in Music Research. *Advances in Data Analysis and Classification, 1, 255–291.*