COMPARING TEXTURE IN PIANO SCORES

Louis Couturier[1] **Louis Bigo**[2] **Florence Levé**[1,2]
[1] MIS, Université de Picardie Jules Verne, Amiens, France
[2] Univ. Lille, CNRS, Centrale Lille, UMR 9189 CRIStAL, F-59000 Lille, France
{louis.couturier, florence.leve}@u-picardie.fr, louis.bigo@univ-lille.fr

ABSTRACT

In this paper, we propose four different approaches to quantify similarities of compositional texture in symbolically encoded piano music. A melodic contour or harmonic progression can be shaped into a wide variety of different rhythms, densities, or combinations of layers. Instead of describing these textural organizations only locally, using existing formalisms, we question how these parameters may evolve throughout a musical piece, and more specifically how much they change. Hence, we define several distance functions to compare texture between two musical bars, based either on textural labels annotated with a dedicated syntax, or directly on symbolic scores. We propose an evaluation methodology based on textural heterogeneity and contrasts in classical Thema and Variations using the TAVERN dataset. Finally, we illustrate use cases of these tools to analyze long-term structure, and discuss the impact of these results on the understanding of musical texture.

1. INTRODUCTION

The term *texture* is used at various levels of description in the music domain. Initially related to the description of sound features, it is also used in symbolic representations of music to describe musical streams through a variety of concepts characterizing the volume and the organization of basic score elements such as notes and voices, which encompass high-level concepts such as monophony and polyphony [1] [1–4]. Between these two extremes, elements of musical texture include layers, voices, melodic or rhythmic patterns, articulation and instrumentation [2,5]. Huron interestingly summarizes it in three main ideas [3]: (1) the density of musical elements, (2) the diversity or inhomogeneity of elements, and (3) the overall sonic activity. The first two can be included in the notion of *compositional texture*, as opposed to orchestral (or timbral) texture [6]. Compositional texture, which is the object of study of the present work, is mostly embedded in the symbolic score. It is worth noting that some models of texture only focus on a particular musical dimension. Nordgren's categorization for instance deals with the vertical dimension only, with note doubling and spacing [7]. Conversely, other approaches focus on the time dimension, as a *complement* to harmony, as in [8], or [9] for style-transfer. Figure 1 shows multiple versions of the same musical theme, which is shaped into different compositional textures.

We aim at quantifying the differences of compositional texture in piano music. Previous studies provided local descriptions of texture [10–12]. Here, we question how textural dimensions may evolve through a whole piece of music. This objective requires the elaboration of dedicated tools to compare textures, more precisely to assess the distance, or dissimilarity, between two given textural configurations.

A number of Music Information Retrieval (MIR) tasks involve the search of similarities at various scales, from pattern detection [13, 14] to genre classification [15, 16]. In the audio domain, music similarity lies at the center of content-based recommender systems [17, 18]. At the level of the musical score, music similarity has also been extensively studied on specific musical notions such as melody [13–15,19–22], rhythmic pattern [23,24], or chord and harmonic progressions [25–27]. Classical approaches for computing similarities include edit-distance on string-based representation, or geometric distance on pianoroll-like representations [28]. New latent embeddings of music also emerged from development of deep neural networks, as well as metric learning methods, like [29]. Although measures of music similarity may ultimately reflect some similarities in the perception of music [30,31], we compare textural information based on symbolic music scores only. In particular, we focus here on studying different representations of texture in order to build interpretable dissimilarity measures.

In this work, we propose distance functions to quantify textural dissimilarity between musical bars from piano scores. We first detail four types of textural distance (Section 2). Then, we introduce estimators of textural heterogeneity and contrast, for longer musical extracts, and propose a dedicated methodology to evaluate our distances, using a dataset of Thema and Variations (Section 3). Finally, we provide use cases of such distances, especially in the context of form or structure analysis (Section 4).

[1] Polyphony, as a type of texture, has a stronger meaning than "the simultaneous presence of possibly more than one note". Here, it implies "two or more lines moving independently" [1]. Similarly, monophonic texture is not restricted to single melodies, but designates the presence of a unique musical line – possibly with note doubling or parallel motions.

Attribution: L. Couturier, L. Bigo, and F. Levé, "Comparing Texture in Piano Scores", in *Proc. of the 24th Int. Society for Music Information Retrieval Conf.,* Milan, Italy, 2023.

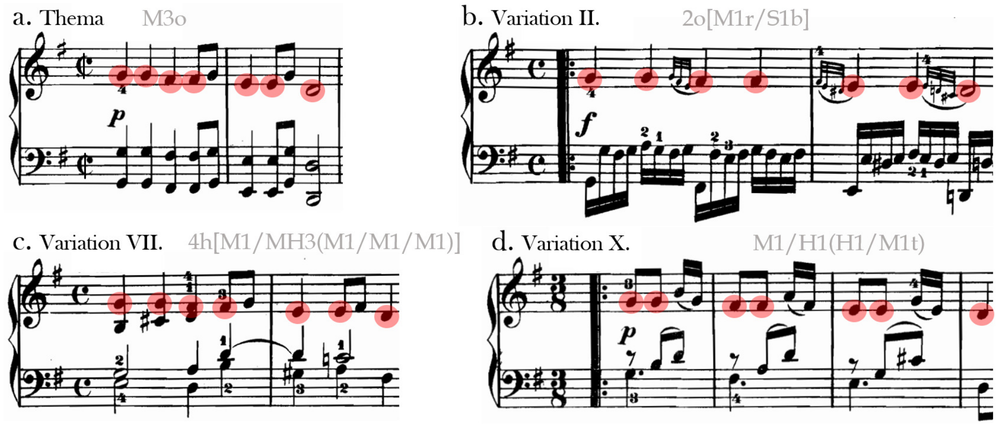

Figure 1: Examples of different textures from *Ten Variations in G on 'Unsere dummer Pöbel meint'* by W. A. Mozart (K. 455, 1784). A textural annotation, following the syntax defined in [10], is provided for each example. The melodic contour (circled in red) is shared among the variations, but the overall compositional texture changes. a. The theme is introduced in monophonic texture: three voices merge into a single musical idea, in parallel octave motions; b. There are now only two notes sounding at the same time: the vertical density decreases. But horizontal density is increased by sixteenth notes; c. A more homophonic texture: three or four threads, mostly synchronous; d. Here, we identify two layers of melody and accompaniment. In these last three bars, the harmony changes, but compositional texture is exactly the same.

2. DEFINING DISTANCES FOR COMPOSITIONAL TEXTURE

The distances that are designed in this paper aim at comparing compositional texture at the scale of individual musical bars. We focus on (polyphonic) piano scores of the Western Classical repertoire, with no voice separation.

2.1 Distances based on textural labels

Textural annotations have been produced in [6] for piano music, on Mozart's sonatas. For each annotated bar, a label enclose two levels of textural information: on the one hand, a set of keywords that indicate the presence of certain properties of the overall textural configuration, or in one of its layer (like parallelism, melodic or harmonic roles...); on the other hand, a vertical structuration of the musical content into main textural layers and sublayers [10]. We propose two distance functions based on this information.

2.1.1 Distance between textural elements

The first distance is based on a set of binary *textural elements* which have been defined in [6, section 3.2]. These indicators express the presence of atomic textural characteristics in a musical bar. They include specific functions of the musical layers: melodic (`M`), harmonic (`H`), or static (`S`), relationships between voices: homorhythmy (`h`), parallel (`p`) or octave (`o`) motions, as well as characteristic musical figures such as sustained (`t`) or repeated (`r`) notes, scale motives (`s`), oscillations (`b`), sparse horizontal density (`_`) and neat changes of texture in the bar (`,`).

A musical bar a is therefore abstractly represented by a vector *texel*(a) which comprised of the 12 textural elements from its label. The distance function d_{texel}returns the Hamming distance between the vectors. It is an integer between zero and 12 that corresponds to the number of textural elements that differ between two bar annotations:

$$d_{texel}(a,b) = \sum_{i=1}^{12} |\text{texel}_i(a) - \text{texel}_i(b)|$$

where a and b are two musical bars, and $\text{texel}_i(\cdot)$ the binary value of the i^{th} textural element of a given bar.

2.1.2 Textural diversity and density

At a higher level of description, textural annotation of piano scores mainly focus on grouping threads [2] of notes into distinct musical layers. Examples 1.a and 1.d both have three threads, but they are organized differently. In 1.a, they merge into a single layer. On the contrary, in 1.d, the threads are divided in 2 main textural layers: its texture is more *diverse* than 1.a without being thicker.

This grouping of threads is formalized in [10] under the terms of *density* (number of threads, of simultaneous sounding notes; the thickness or *density-number* in [2]) and *diversity* (number of distinct layers). These two dimensions allow to embed any textural label in the planar textural space represented in Figure 2 (left).

The *density-Diversity* distance d_{dD} separating two bars is defined as the Euclidean distance between their labels in this space. In previous examples, the first bar of 1.a and 1.d respectively have density-diversity coordinates of (3,1) and (3,2), resulting in a distance of 1.

Note that this distance only takes into account the vertical dimension of compositional texture (as in [7]). A drawback of restricting textural analysis to only two dimensions is that it is less sensitive to small textural fluctuation: as an example, the 1164 labels released by [6] only use 17 distinct combinations of density and diversity values. Nevertheless, this condensed description allows an interpretable

[2] The term 'thread' designates the most atomic elements that can be combined into musical 'layers' [5, p.65].

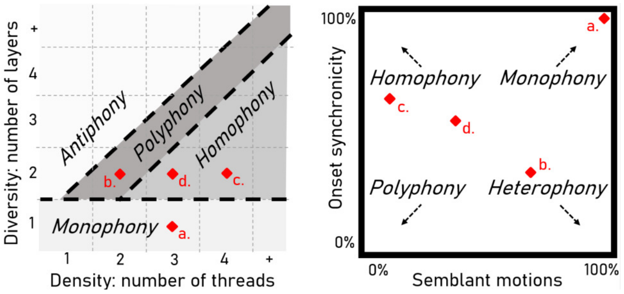

Figure 2: Schematic representations of textural spaces proposed respectively by Couturier et al. [6] (left) and Huron [3] (right). In both case, areas of the space are matched to *main types* of texture [1, 4]. The boundaries are not strict, though. The four examples of Figure 1 are also represented in these spaces.

approach for high-level analysis, as it reflects main textural strategies (see Figure 2).

2.2 Score-based distances

The distances defined in Section 2.1 are based on manually-annotated textural labels. Such textural annotations are however rarely available as their production requires substantial time and expert knowledge. In contrast, this section presents two distance functions that can systematically be computed on encoded musical scores.

2.2.1 *Adapting Huron's textural space*

Another two-dimensional textural space has been proposed by Huron in [3]. It is used in this article to categorize full musical pieces among *main types* of texture (see also [1]): polyphony, monophony, homophony and heterophony, represented on Figure 2. Instead of analyzing the quantity and grouping of musical threads (see Section 2.1.2), it relies on the relationships between them: the proportions of *onset synchronization* and *semblant motions*.

The original study used a pre-existing separation of voices in the pieces (such as Bach's Inventions and Sinfonias) to compute these features. For both of them, we provide an estimation of the value in the interleaved polyphonic case – i.e. without separation of voices. Note that it is not always possible to find a valid and unique voice separation in piano scores [32]. The details of the implementation, out of the scope of the article, can be found in the dedicated repository [1]. They are:

- The ratio of *onset synchrony* quantifies the degree of homorhythmy of the note onsets. A value of 1 indicates a perfect synchronization of note onsets, which is the case in monophonic (see Figure 1.a) or homophonic (chordal or hymnal) textures. This value decreases if note onsets happens while other notes are sustained. For example, 1.b has a value of 0.25: in this case, only one onset over four is fully synchronous.
- The ratio of *semblant motions* estimates to what extent the directions of pitch motions are similar. These feature has its maximal value in the case of monophony, once again, whereas the presence of multiple concurrent layers with opposite motions will reduce its value.

We use these dimensions to build a new distance d_{huron}(the *Huron* distance) between two bars, which is obtained by summing their differences of *onset synchrony* values and *semblant motions* values, using our implementation in the polyphonic interleaved case. This corresponds to the Manhattan distance between their respective coordinates in this textural space.

2.2.2 *Features of density*

We present a last set of three distances based on low-level textural features, focusing on vertical and horizontal density. On the one hand, vertical density refers to the thickness of the texture, the number of simultaneous notes – similarly to the density evoked in Section 2.1.2. On the other hand, horizontal density describes the volume of successive notes, and their position in time. For both dimensions, we use a value of volume and a value of dispersion:

- `vert_avg`: average thickness, in number of notes. After slicing the bar into successive pitch sets, we count the number of pitches in each slice, weighted by their duration.
- `vert_std`: standard deviation of the number of pitches in each onset of one or more notes.
- `horiz_avg`: average number of onsets per beat. The duration of one beat is inferred from the bar time signature.
- `horiz_std`: standard deviation of the regularity of onsets, i.e. around the average duration between successive onsets.

We use Manhattan distance to compare two vectors of features, computed on two target bars. We define and test three variants of this distance: based on the two horizontal features only (d_{h}), on the two vertical ones (d_{v}), or on all the four (d_{hv}).

2.3 Implemention details and release

The features are extracted from the musical scores using intermediate Tab-Separated Values (TSV) files, which contain a list of notes (see [33]). The code [1], in Python, includes a converter to this format from both Humdrum **kern [34] and musicXML formats, using music21 Python library [35].

3. EVALUATING TEXTURAL DISTANCES

3.1 Dataset

To evaluate the relevance of the distances proposed in the previous section, we use bars from classical *Thema*

[1] Available at `http://algomus.fr/code`.

and variations. [36] emphasizes the links between musical variations in general, and musical similarity. In the genre of Thema and variations, a theme is reproduced in short sections with various changes of (textural) parameters, but in a way that allow to recognize the original melodic contour and/or harmony; as in Figure 1. This structure has the advantage of providing both dissimilar examples (in distinct variations), and similar examples (in the same variation).

Although no explicit mention of textural homogeneity within variations has been found in musicological literature, authors more often insist on the higher contrast between distinct variations [37, p.570]. The genre of Thema and variations provides “the largest esthetic spectrum” [38], and this variety of content is valuable in our case. We rely on this fundamental assumption for the rest of the paper: *on average, a musical bar is more similar – in texture – to a bar from the same musical phrase, than to any other bar in another variation or piece*.

We use the TAVERN dataset [39], which consists in 27 sets of thema and variations by Mozart (10) and Beethoven (17). The variations are already segmented into structural phrases, totalling 1060 of them in the whole dataset. We take those phrases as structural units in which we use the score-based distances defined in section 2.2. Further annotations of texture would be required to apply label-based distances on this dataset.

Remark. The texture of phrases can vary within the same variation, to a lesser extent – this is generally the case in bipartite or tripartite variations, which is a common structure in this context [37, 38]. Changes of mode (major/minor) often occur, in general at least once per set of variations. This change is not considered as textural, but it is often accompanied by changes of other musical parameters that are in the scope of texture, so it would still add valuable information.

3.2 Heterogeneity and contrast

To evaluate textural dissimilarities on full musical extracts, we introduce two indicators:

- *heterogeneity*: the heterogenity (h_d) within a single set of bars corresponds to *the average distance between pairs of distinct bars from the set*, for a given distance function d.
- *contrast*: the contrast (c_d) between two sets of bars is defined as *the average distance value between pairs of bars from the two extracts*, for a given distance function d.

More formally, we have:

$$h_d(S) = \underset{\forall (m_i,m_j) \in S^2, i \neq j}{avg} d(m_i, m_j)$$

$$c_d(S_1, S_2) = \underset{\forall m_i \in S_1, \forall m_j \in S_2}{avg} d(m_i, m_j)$$

where avg is the arithmetic mean operator, S, S_1 and S_2 are sets of bars, and m_i, m_j denote bars/measures in those sets.

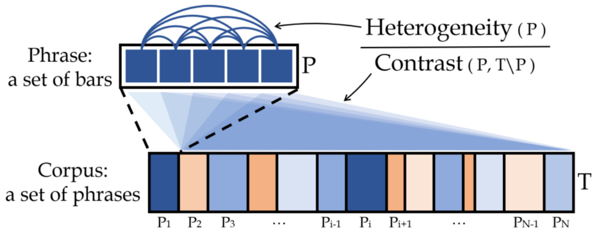

Figure 3: Schematic representation of the computation of heterogeneity of a specific phrase (arcs above) and contrast (links between the bars of phrase P and all the other bars outside P, in the corpus T). Our evaluation metric is the average value of this ratio for all the phrases of the corpus.

The heterogeneity is a measure of dispersion: a lower value means that samples in the extract are more similar between each other (given a distance d). We specifically ignore the comparisons of a bar with itself to reduce the influence of the size of S.

Let us illustrate those two indicators for the descriptor-based distance d_{hv} (Section 2.2.2) using examples Figure 1.a and Figure 1.d. We note S_a = {bars of Figure 1.a} = $\{a_1, a_2\}$ and S_d = {bars of Figure 1.d} = $\{d_1, d_2, d_3\}$. In S_a, the heterogeneity is simply equal to the distance between its two bars: small differences occur in horizontal density, but not in vertical density. We obtain a value of 0.6. To compute the *heterogeneity* in S_d, we have three possible unordered pairs of bars to compare ($\{d_1, d_2\}$, $\{d_1, d_3\}$ and $\{d_2, d_3\}$); however, the texture in these bars is precisely the same regarding d_{hv}, resulting a in value of zero of heterogeneity. The inequality $h_d(S_a) > h_d(S_d)$ can be interpreted as “S_a is more texturally heterogeneous than S_d, with regards to distance d”. The contrast c_d between S_a and S_d is 1.825. In general, the inequality $h_d(S_a) < c_d(S_a, S_d)$ means that a bar in S_a is, on average, more similar to other bars in S_a than bars in S_d.

3.3 Evaluation methodology

We evaluate how heterogeneous the texture is within each phrases of the TAVERN dataset, compared to the rest of the corpus. Under the assumption exposed in Section 3.1, we assess the quality of a textural distance d by looking for the lowest *average Relative Heterogeneity* on TAVERN phrases (T):

$$\text{aRH}_T(d) = \underset{\forall P_i \in T}{avg} \left(\frac{h_d(P_i)}{c_d(P_i, T \setminus P_i)} \right)$$

where T is the set of all the phrases in TAVERN dataset, P_i is the i^{th} phrase of the dataset, and d is a textural distance function between individual musical bars.

This process is schematized in Figure 3. For a given phrase, if the ratio between intra-phrase heterogeneity and inter-phrases contrast is very low, it means that the extract is rather homogeneous, and that this texture – or whatever the distance d represents – is rather specific to this extract compared to the rest of the corpus. If this ratio is above 1, it means that the bars in this phrase are more similar to other

Distance	d	$\mathrm{aRH}_T(d)$
Horiz. and Vert. density features	d_{hv}	0.51
Horizontal density features	d_{h}	0.39
Vertical density features	d_{v}	0.64
Huron's textural space	d_{huron}	0.72
Comparison: Pitch class content	d_{pc}	0.80

Table 1: Evaluation of textural distances using the *Average Relative Heterogenity* on phrases of the TAVERN dataset ($\mathrm{aRH}_T(d)$), to minimize.

bars outside the phrases than between themselves. Put differently: a value below 1 show that intra-phrase distances (heterogeneity) are smaller than inter-phrases comparison (contrast with the corpus). The value of $\mathrm{aRH}_T(d)$ is the average of this ratio on all the phrases of the corpus.

Remarks. We could directly compute values of *contrast* or *heterogeneity* on reference data, using different textural distance d_i and opt for the most convincing values. However, these values are not directly comparable if they are based on different distances: they are average values of specific distances, and thus follow their respective – and possibly very different – order of magnitude. Also note that the *contrast* is not a distance function (or metric) because the contrast between the same set of bars could be different from zero – if its bars that are not all the same. The functions presented in Section 2 *are* metrics, applied to different representations of texture in a musical bar.

3.4 Results

The results, for all score-based distances, are shown in Table 1. Using the distance based on all density features (d_{hv}), the aRH_T of 0.51 indicates that a musical bar is, on average, a twice more similar to bars in the same phrase than to the rest of the corpus. The use of horizontal density features alone (d_{h}) improves this value (0.39), highlighting the importance of the time dimension to discriminate between textures.

For comparison, we integrate an additional distance (d_{pc}) that describe not textural but harmonic content – computing Euclidean distance between pitch-classes profile. Its aRH_T of 0.80 is still below 1, which means that intra-phrase d_{pc}values (heterogeneity) are smaller than inter-phrase comparison (contrast with the corpus); this is not surprising in tonal music. But most importantly, this evaluation metric value is higher than for all other textural distances. This gap contributes to validate the use of Thema and Variations as a source of empiric ground truth examples of textural similarities.

3.5 Links between distances

In Table 2, we display correlations between all the distances defined in Section 2. They are computed on all distinct pairs of bars among 1160 from Mozart piano sonatas (K. 279, K. 280, K. 283), for which we have both textural annotations [6] and encoded scores [33]. We use Spearman correlation, that depicts similarities of rankings of these values.

	d_{hv}	d_{h}	d_{v}	d_{huron}	d_{dD}	d_{texel}
d_{hv}	1.00	"	"	"	"	"
d_{h}	**0.95**	1.00	"	"	"	"
d_{v}	**0.33**	0.05	1.00	"	"	"
d_{huron}	0.10	0.10	0.05	1.00	"	"
d_{dD}	0.19	0.03	**0.53**	0.03	1.00	
d_{texel}	0.10	0.06	0.14	0.01	0.20	1.00

Table 2: Spearman correlation between textural distances as defined in Section 2, evaluated on all pair of bars in three Mozart piano sonatas (K. 279, K. 280, K. 283).

	d_{hv}	d_{h}	d_{v}	d_{huron}	d_{dD}	d_{texel}
Horizontal density (time dimension)	×	×				×
Vertical density (thickness)	×		×		×	×
Semblant motions, parallelism				×		×
Roles of layers (melody, acc. ...)						×
Main types of texture (see Fig.2)				×	×	
Computed on symbolic scores	×	×	×	×		
Computed on annotated labels					×	×

Table 3: Summary of the distances defined in Section 2, and the different dimensions of compositional texture that they take into account.

Huron-space distance (d_{huron}) and texel distance (d_{texel}) seem independant from other distances. We explain it by the fact that our distances focus on different dimensions of texture, summarized in Table 3. In particular, d_{texel}covers a wide range of abstract concept and is the only function that deals with the roles of layers, which is difficult to approximate using low-level features. Otherwise, we find that using horizontal density features only (d_{h}) gives a very similar behavior than using all four density features (d_{hv}) – with a correlation of 0.95. Although Density-Diversity distance (d_{dD}) and vertical-feature distance (d_{v}) deal with very different level of abstraction, they correlates positively (0.53), as they both focus on the vertical dimension of texture.

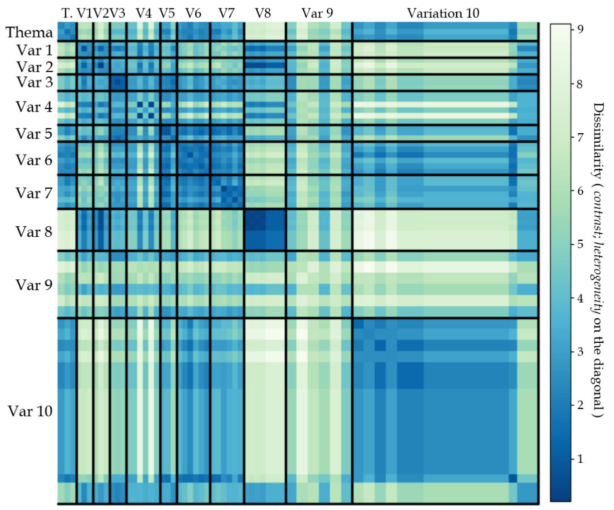

Figure 4: Textural dissimilarities between the phrases of *Ten Variations in G on 'Unsere dummer Pöbel meint'* by W. A. Mozart (K. 455, 1784). Intersections are colored according to *contrast* values using d_{hv}distance (Section 2.2.2), and *heterogeneity* of phrases on the diagonal (Section 3.2). The phrases are scaled according to their size in number of bars (totalling 338 in the whole piece). The most similar extracts are shown in dark blue, whereas light green indicates higher dissimilarity. – We identify blocks of consecutive similar variations, such as (1,2,3) or (5,6,7); inner structure of varations may reveal contrasting segments in the case of Variation 4; Variation 9 is very contrasted due to the alternation between chordal texture and fast melodic lines; the penultimate phrase comes back to the original texture of the Thema.

4. USE CASES FOR STRUCTURE ANALYSES

4.1 Long-term textural dissimilarities

The *contrast* defined in Section 3.2 can be used as a dissimilarity measure between any sets of bars, from individual pieces to entire corpora. It may also emphasize the relationships between sections, or phrases, of a given piece of music. Figure 4 shows an example of self-similarity matrix based on textural contrast between phrases of a piece in Thema and Variations form. Beyond the case of Thema and Variations, the contrast measure gives an overview of the piece macrostructure, and may even link thematic material up to transposition, such as recapitulation parts in sonata form. More generally, the proposed distances can lead to promising and original approaches for automatic structure segmentation.

4.2 Short-term textural changes

In this paper, we assume lower textural heterogeneity within phrases of thema and variations. But in the general case, changes of textures may occur in the middle of a phrase. Following the intuition that in-phrase texture changes mostly occur in openings and endings of phrases in the TAVERN dataset, we evaluate d_{hv}using the same methodology as in Section 3.3, but systematically ignore the last bar of each phrase of the corpus. We find that $\mathrm{aRH}_T(d_{\mathrm{hv}})$ decreases from 0.51 to 0.42. When removing each first bar instead, it drops to 0.34. In comparison, removing the second or third bar of the phrases increases the original value of $\mathrm{aRH}_T(d_{\mathrm{hv}})$ to respectively 0.55 and 0.54. This shows that the 'core' of phrases have slightly more textural homogeneity, and most importantly that openings and endings are less similar to the middle of phrases. Typical examples are transitional melodies and final chords in cadences – which often constrast with the rest of the phrase. We believe that our distance can be used to study more precisely these local changes of texture within short sections.

5. CONCLUSION AND FURTHER WORKS

The textural distances proposed in this paper give promising perspectives for the computation of multi-level similarities in symbolic music. On the one hand, comparing textural labels allows to rely on expert data, which is already known as texturally meaningful. This information already carries a lot of abstraction, but it is costly to produce in practice and can lead to a certain amount of subjectivity [40]. Moreover, the low amount of available annotations hinders our ability to evaluate the quality of these distances. On the other hand, using symbolic features that can be computed automatically is more practical, and also more objective. In further work, we plan to investigate the best features to use at a more global scale, as well as their relative contibution.

Although the proposed distances are drawn at the level of musical bars, we elaborated a more global dissimilarity measure to compare sets of several bars, and highlight textural contrast between and within structural sections of musical pieces. This measure made possible a quantitative evaluation of textural distances on a corpus of Thema and Variations, based on the assumption that texture is more dissimilar between two distinct variations, and more homogeneous within single variations.

Our distances capture different facets of compositional texture, at different levels of abstraction (see Table 3). Focusing on more atomic and independant textural aspects can enhance the precision and the interpretability of our analyses. However, ensuring a proper disentanglement of such dimensions remains a major challenge. Integrating Thema and variations in the evaluation methodology is a step further to link theoratical models of texture to concrete, and somewhat intuitive, examples. It contributes to a better understanding of some models of texture, but also of musical texture itself.

A potential continuation of this work is to broaden the scope of our experiments to other repertoires. We believe that the tools introduced in this paper are easily extendable to other styles of written polyphonic music, or to other instruments. In the meantime, the present experiments on Western classical piano music already offer promising opportunities of quantitative analyses of texture with regards to genre, style, form or harmony.

6. ACKNOWLEDGMENT

We want to thank Jean-Paul Chehab for fruitful discussion at the beginning of the project. We also thank Dinh-Viet-Toan Le, Alexandre D'Hooge and the rest of the Algomus team, as well as the anonymous reviewers for their constructive feedback. This research is partly funded by Région Hauts-de-France and by Agence Nationale de la Recherche (ANR), project TABASCO ANR-22-CE38-0001.

7. REFERENCES

[1] B. Benward and M. Saker, *Music: In Theory and Practice, Vol. I. Eigth Edition.* McGraw-Hill, 2008, ch. 7, p. 145–162.

[2] W. Berry, *Structural functions in music: A Probing Exploration of Conceptual Bases and Techniques for the Analysis of Tonality, Texture, and Rhythm.* Prentice-Hall, 1976.

[3] D. Huron, "Characterizing musical textures," in *International Computer Music Conference (ICMC 1989)*, 1989, pp. 131–134.

[4] J. Dunsby, "Considerations of texture," *Music & Letters*, vol. 70, no. 1, pp. 46–57, 1989.

[5] D. M. de Sousa, "Textural design: A compositional theory for the organization of musical texture," Ph.D. dissertation, Universidade Federal do Rio de Janeiro, 2019.

[6] L. Couturier, L. Bigo, and F. Levé, "A dataset of texture annotations in Mozart piano sonatas," in *International Society for Music Information Retrieval Conference (ISMIR 2022)*, 2022.

[7] Q. R. Nordgren, "A measure of textural patterns and strengths," *Journal of Music Theory*, vol. 4, no. 1, pp. 19–31, 1960.

[8] Z. Wang, D. Wang, Y. Zhang, and G. Xia, "Learning interpretable representation for controllable polyphonic music generation," in *International Society for Music Information Retrieval Conference (ISMIR 2020)*, 2020.

[9] O. Cífka, U. Şimşekli, and G. Richard, "Groove2Groove: one-shot music style transfer with supervision from synthetic data," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 28, pp. 2638–2650, 2020.

[10] L. Couturier, L. Bigo, and F. Levé, "Annotating symbolic texture in piano music: a formal syntax," in *Sound and Music Computing Conference (SMC 2022)*, 2022.

[11] M. Giraud, F. Levé, F. Mercier, M. Rigaudière, and D. Thorez, "Towards modeling texture in symbolic data," in *International Society for Music Information Retrieval Conference (ISMIR 2014)*, 2014, pp. 59–64.

[12] D.-V.-T. Le, M. Giraud, F. Levé, and F. Maccarini, "A corpus describing orchestral texture in first movements of classical and early-romantic symphonies," in *Digital Libraries for Musicology (DLfM 2022)*, 2022, pp. 22–35.

[13] D. Conklin, "Pattern in music," *Journal of Mathematics and Music*, vol. 15, no. 2, pp. 95–98, 2021. [Online]. Available: https://doi.org/10.1080/17459737.2021.1947404

[14] B. Janssen, W. B. de Haas, A. Volk, and P. van Kranenburg, "Finding repeated patterns in music: State of knowledge, challenges, perspectives," in *International Symposium on Computer Music and Multidisciplinary Research (CMMR 2013)*, 2013, pp. 277–297.

[15] A. Ferraro and K. Lemström, "On large-scale genre classification in symbolically encoded music by automatic identification of repeating patterns," in *Proceedings of the 5th International Conference on Digital Libraries for Musicology*, 2018, pp. 34–37.

[16] J. Ens and P. Pasquier, "Quantifying musical style: Ranking symbolic music based on similarity to a style," *arXiv preprint arXiv:2003.06226*, 2020.

[17] P. Knees and M. Schedl, *Music similarity and retrieval: an introduction to audio-and web-based strategies.* Springer, 2016, vol. 9.

[18] M. Schedl, P. Knees, B. McFee, and D. Bogdanov, "Music recommendation systems: Techniques, use cases, and challenges," in *Recommender Systems Handbook.* Springer, 2021, pp. 927–971.

[19] T. Collins, J. Thurlow, and R. Laney, "A comparative evaluation of algorithms for discovering translational patterns in baroque keyboard works," in *International Society for Music Information Retrieval Conference (ISMIR 2010)*, 2010, pp. 3–8.

[20] V. Velardo, M. Vallati, and S. Jan, "Symbolic melodic similarity: State of the art and future challenges," *Computer Music Journal*, vol. 40, no. 2, pp. 70–83, 2016.

[21] F. Karsdorp, P. van Kranenburg, and E. Manjavacas, "Learning similarity metrics for melody retrieval." in *ISMIR*, 2019, pp. 478–485.

[22] S. Park, T. Kwon, J. Lee, J. Kim, and J. Nam, "A cross-scape plot representation for visualizing symbolic melodic similarity." in *ISMIR*, 2019, pp. 423–430.

[23] F. Bruford, O. Lartillot, S. McDonald, and M. Sandler, "Multidimensional similarity modelling of complex drum loops using the groovetoolbox," in *International Society for Music Information Retrieval Conference (ISMIR 2020)*, 2020, pp. 207–215.

[24] D. Cocharro, G. Bernardes, G. Bernardo, and C. Lemos, "A review of musical rhythm representation and (dis) similarity in symbolic and audio domains," *Perspectives on Music, Sound and Musicology: Research, Education and Practice*, pp. 189–208, 2021.

[25] T. Rocher, M. Robine, P. Hanna, and M. Desainte-Catherine, "A survey of chord distances with comparison for chord analysis," in *ICMC*, 2010.

[26] W. B. De Haas, M. Robine, P. Hanna, R. C. Veltkamp, and F. Wiering, "Comparing approaches to the similarity of musical chord sequences," in *Exploring Music Contents: 7th International Symposium, CMMR 2010, Málaga, Spain, June 21-24, 2010. Revised Papers 7*. Springer, 2011, pp. 242–258.

[27] D. Bountouridis, H. V. Koops, F. Wiering, and R. C. Veltkamp, "A data-driven approach to chord similarity and chord mutability," in *2016 IEEE Second International Conference on Multimedia Big Data (BigMM)*. IEEE, 2016, pp. 275–278.

[28] K. Lemström and A. Pienimäki, "On comparing edit distance and geometric frameworks in content-based retrieval of symbolically encoded polyphonic music," *Musicae Scientiae*, vol. 11, no. 1_suppl, pp. 135–152, 2007.

[29] J. Lee, N. J. Bryan, J. Salamon, Z. Jin, and J. Nam, "Disentangled multidimensional metric learning for music similarity," in *ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2020, pp. 6–10.

[30] S. McAdams, S. Vieillard, O. Houix, and R. Reynolds, "Perception of musical similarity among contemporary thematic materials in two instrumentations," *Music Perception*, vol. 22, no. 2, pp. 207–237, 2004.

[31] E. Cambouropoulos, "How similar is similar?" *Musicae Scientiae*, vol. 13, no. 1_suppl, pp. 7–24, 2009.

[32] C. Finkensiep and M. A. Rohrmeier, "Modeling and inferring proto-voice structure in free polyphony," in *International Society for Music Information Retrieval Conference (ISMIR 2021)*, 2021, pp. 189–196.

[33] J. Hentschel, M. Neuwirth, and M. Rohrmeier, "The annotated Mozart sonatas: Score, harmony, and cadence," *Transactions of the International Society for Music Information Retrieval*, vol. 4, no. 1, pp. 67–80, 2021.

[34] D. Huron, "Music information processing using the humdrum toolkit: Concepts, examples, and lessons," *Computer Music Journal*, vol. 26, no. 2, pp. 11–26, 2002.

[35] M. S. Cuthbert and C. Ariza, "music21: A toolkit for computer-aided musicology and symbolic music data," in *International Society for Music Information Retrieval Conference (ISMIR 2010)*, 2010, pp. 637–642.

[36] A. Volk, W. B. de Haas, and P. Van Kranenburg, "Towards modelling variation in music as foundation for similarity," in *Proceedings of the 12th International Conference on Music Perception and Cognition and the 8th Triennial Conference of the European Society for the Cognitive Sciences of Music*. School of Music Studies, Aristotle University of Thessaloniki, 2012, pp. 1085–1094.

[37] W. E. Caplin, *Analyzing Classical Form: An approach for the classroom*. Oxford University Press, 2013.

[38] E. De Montalembert and C. Abromont, *Guide des genres de la musique occidentales*. Paris: Fayard Henry Lemoine, 2010.

[39] J. Devaney, C. Arthur, N. Condit-Schultz, and K. Nisula, "Theme and variation encodings with roman numerals (tavern): A new data set for symbolic music analysis," in *International Society for Music Information Retrieval Conference (ISMIR 2015)*, 2015.

[40] Y. Ni, M. McVicar, R. Santos-Rodriguez, and T. De Bie, "Understanding effects of subjectivity in measuring chord estimation accuracy," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 21, no. 12, pp. 2607–2615, 2013.