A Dataset of Symbolic Texture Annotations in Mozart Piano Sonatas
Louis Couturier, Louis Bigo, Florence Levé

To cite this version:

HAL Id: hal-03860195
https://hal.science/hal-03860195
Submitted on 18 Nov 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L’archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Distributed under a Creative Commons Attribution 4.0 International License
A DATASET OF SYMBOLIC TEXTURE ANNOTATIONS IN MOZART PIANO SONATAS

Louis Couturier\textsuperscript{1}  
Louis Bigo\textsuperscript{2}  
Florence Levé\textsuperscript{1,2}

\textsuperscript{1} MIS, Université de Picardie Jules Verne, Amiens, France  
\textsuperscript{2} CRISTAL, UMR 9189 CNRS, Université de Lille, France

\{louis.couturier, florence.leve\}@u-picardie.fr, louis.bigo@univ-lille.fr

ABSTRACT

Musical scores are generally analyzed under different aspects, notably melody, harmony, rhythm, but also through their texture, although this last concept is arguably more delicate to formalize. Symbolic texture depicts how sounding components are organized in the score. It outlines the density of elements, their heterogeneity, role and interactions. In this paper, we release a set of manual annotations for each bar of 9 movements among early piano sonatas by W. A. Mozart, totaling 1164 labels that follow a syntax dedicated to piano score texture. A quantitative analysis of the annotations highlights some characteristic textural features in the corpus. In addition, we present and release the implementation of low-level descriptors of symbolic texture, that are preliminary experimented for textural elements prediction. The annotations and the descriptors offer promising applications in computer-assisted music analysis and composition.

1. INTRODUCTION

1.1 Texture and symbolic texture

Musical texture generally refers to two distinct levels of abstraction used to describe musical content [1]. On the one hand, \textit{sound} related texture, that can be referred to as \textit{orchestral texture}, results from orchestration, instrumentation and timbral characteristics of instruments and performances. On the other hand, \textit{symbolic texture}, or \textit{compositional texture}, results from the organization of notes, chords and voices in the musical score. Naturally, these notions are closely related, and Hérold studies both textural and instrumental factors of timbre [2], highlighting the impact of compositional texture on the final sound field. Symbolic texture, which is the focus of the dataset presented in this paper, can be described through high level musical concepts such as layer separation, diversity of sonic activities, layer roles, note density and interactions [3–7]. For instance, in piano music, an accompanying chord sequence can be performed in various ways, each one being identified by a specific texture contributing to a stylistic identity. Symbolic texture stands at the center of the compositional process and can sometimes be understood as a notion of style [8]. Musical style is however commonly associated with a whole piece [9] or section whereas the notion of symbolic texture considered in this work tends to describe much shorter time spans in the musical score.

1.2 Related work

Computational methods to analyze symbolic texture have been elaborated in various musical styles including classical string quartet music [10] and modern popular guitar music [11]. Given the wide range of playing modes offered by the instrument, piano music brings unprecedented challenges for the task of in-score texture identification. The present work builds on a recent formal syntax elaborated to classical piano music modeling [7]. As a crucial parameter of musical style, symbolic texture has also recently raised an important attention in the tasks of music generation [12] and style transfer [8,13], where musical texture is efficiently learned by deep neural networks but with limited perspectives of explicit categorization and musicological interpretation. Alternatively, the dataset proposed in the present work promotes a pedagogical and transparent expression of symbolic piano texture, aiming at facilitating its use to design computational tools intended to assist music pedagogy, analysis and composition.

The MIR – Music Information Retrieval – community dedicates an important part of its work effort in building musical expert annotations accompanying music datasets to facilitate training and evaluation of computational models. The classical repertoire, in particular its piano music, gives raise to a number and variety of annotation needs given the richness of its musical language. Key regions, cadences, phrases and harmonies were annotated in the Mozart piano sonatas from the New Mozart Edition (Neue Mozart Ausgabe) [14]. As other representative initiatives, the TAVERN dataset includes harmony and phrase annotations on Mozart and Beethoven’s piano variations [15] and the Fugue dataset includes form annotations specific to this genre [16]. Beyond piano music, classical string quartets have also raised a number of key, harmony and structure annotation efforts including repertoires of Haydn [17], Mozart [18] and Beethoven [19, 20]. Although a corpus has been proposed in [21], symbolic texture has still rarely...
The overall texture is split into independent textual layers that are described individually, separated by a /. They are ordered by descending register.

Example: The label M1/H2 includes two layers, as in examples a, b and c in Figure 1.

Function. The function of each layer is expressed with a combination of three specific labels being M for melodic function, H for harmonic function, and S for static function like pedals and ostinati.

Example: The label M1/H2 includes one layer with a melodic function and one layer with an harmonic function. As an other example, the label HS1 includes one single layer having both a harmonic and a static function like a persistent arpeggio.

Density. The density of a layer, also called thickness [1], corresponds to the number of voices it includes, expressed by an integer right after the function.

Example: The label M1/H2 includes one melodic layer with one voice and one harmonic layer including two voices.

Global density. The global density of a region corresponds to its global number of voices and is indicated before brackets surrounding the whole label. It is an approximation of the average number of notes perceived simultaneously. In most cases, the global density is equal to the sum of the layer's density, in which cases its notation is optional because redundant.

Example: The label 2[M1/H2] indicates a global density of 3 and can be simplified into M1/H2. However, the label 2[M1/H2], which can occur in certain types of sparse regions, cannot be simplified.

Internal organization of a layer. Additional elements can indicate the presence of relationships between voices: homorhythmy (h), parallel motion (p), octave (o) or characteristic musical figures: sustained notes (s), repeated notes (r), oscillations (b) or scales (a).

Example: The label M2p/H3hr can describe a melody doubled at the third accompanied by repeated three-note chords.

Sublayer decomposition. Each layer can optionally be decomposed into sublayers notated between parentheses.

Example: In the label M1/HS1(S1/M1) (see Figure 1.c), the second layer HS1 itself includes one static sublayer and one melodic sublayer enabling for example the expression of a single voice accompaniment consisting in an alternation between a single repeated pitch and a moving melodic line.

Sparsity. When a layer does not last during the full measure (as in Figure 1.b), or has too low horizontal density, it is considered as sparse and notated with ‘.’. This symbol is also used to annotate empty bars.
The excerpt of the first movement of Mozart’s Piano Sonata n.2 (K. 279, m.7-10) as seen on the web application used to annotate and review the dataset. Each annotated label can be put in doubt by the reviewer: once a feedback is opened over one given measure, comments can be added to exchange with the annotator. Feedbacks have three states: raised (in white – just submitted by a reviewer), in conflict (red – waiting for a consensus) and resolved (green, as in the figure).

<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Sampling</td>
<td>Western classical style piano music. Available data, all movements in Sonata Form.</td>
</tr>
<tr>
<td>Type of media diffusion</td>
<td>original files in .mscx and .tsv (Tab Separated Values), annotations in .tsv .txt and .dez (Dezrann format).</td>
</tr>
</tbody>
</table>

Table 1. Description of the annotated dataset.

2.2 Annotated corpus

Following the syntax presented in Section 2.1, we release manual annotations of texture at the bar level in the 3 movements of the 3 sonatas K. 279, K. 280 and K. 283 by W. A. Mozart, totaling a set of 1164 bar annotations in the 9 movements. Those early sonatas were all composed in the end of the year 1774, which presumes a style consistency and limits the shift of compositional practice that could be induced by the evolution of piano manufacture. Moreover, these movements present an interesting diversity in tonality, rhythmic signature and tempo, while covering a large variety of textures. All those movements are in Sonata Form, which opens perspectives to pursue research on the links between texture and form [22].

Table 1 synthesizes the properties of the dataset following the conventions proposed in [23].

2.3 Annotation procedure

2.3.1 Granularity of the annotations

Textural segments can have highly variable duration, which tends to complicate their annotation. As a compromise to facilitate statistics and computational processing of the annotations, we propose in this work to annotate a single texture label for each bar, as indicated in the proposition of syntax [7]. In some cases where the texture strongly shifts in the middle of a bar, a comma (,) is used to segment the label in two parts. In particular, this situation may occur on boundaries between musical phrases. By contrast, a texture can also remain unchanged over several consecutive bars, which will lead to a repetition of a same texture label over these bars.

2.3.2 Annotation and reviewing methodology

This section describes our annotation protocol which was inspired by [25] and illustrated in Figure 3. One expert $E$ annotated all the measures of the 9 movements with texture labels on Dezrann $^2$, a web platform dedicated to the annotation of musical analyses on scores [24]. The annotations were committed to a Git repository to trace the different stages of annotation and reviews. A Python script $^3$ was used to systematically check that the labels were consistently formed before to proceed to the review phase, relieving the reviewers from syntax checking to concentrate on the textural decomposition. Two reviewers $R_1$ and $R_2$ (distinct from the expert), with strong musical background and piano music knowledge, as well as knowledge of the texture syntax used, reviewed those annotations on Dezrann. They added a feedback label each time they disagreed or were unsure of the annotator label choice, detailing the reasons of the doubt and possibly providing a new proposition for the label. The expert $E$ then studied all the doubts, resolved the obvious ones and discussed with the two reviewers to reach a consensus on the remaining labels. Over the 1164 initial annotations of the annotator, 31% raised a reviewer feedback and 22% (256) were updated in

$^1$http://algomus.fr/data
$^2$http://dezrann.net
$^3$http://algomus.fr/code
the end. This substantial number emphasizes the importance of the reviewing phase as well as the complexity of texture annotation due to its variety and the high expressivity of the syntax. Besides correcting possible omissions or typos in the labels, the majority of conflicts ensured a more homogeneous and precise use of the syntax in the whole dataset – in each movements and between them. Hence, most of the discussions between the annotator and the reviewers involved several measures: consecutive ones sharing similar texture, cyclic resurgence of thematic materials (for example in both exposition and recapitulation of sonata form) or comparable textures across pieces. Note that in some cases (like in music analysis in general), it is possible to have different interpretations of the textural layers in a musical passage. The goal of the review here was not to blur those distinct possible analyses but to provide consistent labels that made sense for all.

2.3.3 Syntax knowledge and annotation reproducibility

The annotator followed the syntax recalled in Section 2.1 with the help of a catalog of common textural configurations provided in [7]. As shown in Section 2.3.2, the texture of a bar can sometimes be interpreted in various ways, leading to different labels. The consideration of the neighboring bars can also impact the texture estimation. To limit inconsistency in the labels, we provide annotation guidelines in addition to the dataset. These guidelines typically indicate the order of consideration of the different textural elements when estimating a label. They aim at encouraging a sense of normalization in the annotations, although such a normalization has not been strictly formalized. This aims at helping future annotators to understand and reproduce the annotation labels. Since the web platform used allows to share several analyses of the same musical piece, divergent labels could be further used ultimately to propose alternative analyses and diversify the dataset for machine learning applications.

3. CORPUS ANALYSIS

3.1 Common textures and layers

Among the 1164 annotated bars, 16.6% include two labels instead of one due to a substantial shift in texture in the middle of the bar (see Section 2.3.1), resulting in 1357 textural configurations. From this set, we can extract 2317 non-empty textural layers. Among the most common layers, we find simple single-voice melodies (M1) which totaled 24.9% of all written layers, followed by melodies with scale motives (M1s, 7.0%, see Figure 1.b) and with parallel motions (M2p, 3.7%), generally doubled at the third (see Figure 1.a) or at the sixth.

Another important family of textural layers encompasses repetitive accompaniment figures like arpeggios, in which notes of the current harmony are played one by one (HS1, here). A famous example is the Alberti bass, an idiomatic pattern which alternates notes in low-high-middle-high order (see Figure 1.c and the first half of Figure 2) and is typical of Mozart’s piano music. The combination

4 The proportions of all annotated textures and textural elements are provided in the dataset repository: http://algomus.fr/data.
element \( \text{II} \) varies between annotated movements, notably according to the tempo: this percentage is 26% higher in slower movements (the second of each full sonata) than in the average of the 6 others (84.4% versus 58.2%).

### 3.3 Density and diversity

We compute the diversity and the global density of each annotated textural configuration. As detailed in Section 2.1, the diversity corresponds to the number of stacked layers while the global density corresponds to the approximate number of monophonic textures that can be heard simultaneously. Figure 4 puts in relation these two values for the set of annotated labels. The diagonal is assimilated to polyphony where each voice sounds like a new distinct musical idea, an individual layer. On the contrary, the bottom row, with diversity of value 1, corresponds to monophonic texture; thus, any note is contributing to the same unique musical entity. This case is common at the end of structural parts and cadences, typically in homorhythm (see Figure 1.d). Homophonic textures, for example combinations of a main melody and accompaniment, are found between these two areas. Among the annotated non-empty textural configurations, 29.2% lay in the 2:2 combination of diversity/density, including the common case of melody and singe-voice accompaniment, presented earlier. Textures in 2:3 and 2:4 (three or four voices merged into two layers) illustrate denser homophonic variants, in which a melody may be doubled at the third or at the sixth, or an accompaniment made of simultaneous notes (in homorhythm). The combinations above the diagonal correspond to situations where the number of layers is higher than the number of voices (2:1 or 3:2). This can happen when two distinct lines alternate in antiphony (call & response): several textural layers are perceived whereas their notes do not overlap.

The diversity, which corresponds to the number of layers, rarely exceeds 2. However, Figure 4 shows that a systematic separation into two layers, which follows an intuitive organization of the score content between the two hands of the pianist, would not be sufficiently representative of the corpus.

In other repertoires, the number of voices – the vertical density – could be globally higher. Bach three-voice inventions would be mainly categorized as a continuous 3:3 polyphony, and some works of the Romantic Era including very large chords would be extremely dense vertically. Hence, it is easy to imagine that this textural space, similarly to the one described in [4], is prone to convey strong stylistic content. Moreover, the evolution of texture throughout a single piece of music could be modeled as a trajectory in this space, which offers promising future analytical perspectives.

### 4. APPLICATION: PREDICTION OF TEXTURAL ELEMENTS

We present in this section preliminary experiments of texture prediction.
The predictions of notes simultaneities and semblant relationships can be justified by the limited proportion of positive examples in the dataset (around 5.1%). Finally, the difficulty of annotating scale motives (s) in practice is reflected in their prediction result (F1-score of 0.36). From slow descending sequences of neighboring pitches to cells of short-duration notes that share the same repeated contour: many variants of this element can be found in the corpus, some of which were making consensus difficult to reach between annotator and reviewers. This issue can be adressed in the definition of the syntax as the need to refine the criteria and guidelines for the annotation of targeted concepts. The scale patterns sometimes span over wider regions than the duration of a bar, making the descriptors inadequate in this case. Providing a larger context to our models therefore appears as a promising perspective to improve the detection of this textural element.

4.2 Description of the models

Different machine learning models were compared to predict the presence of textural elements presented in Section 3.2 from the set of 62 descriptors detailed in Section 4.1. Descriptor values are given to the model as vectors of 62 floats extracted from each measure. The prediction of the presence of each textural element was formulated as a binary classification task leading to a dedicated model for each of the 14 textural elements. Whereas modern data-driven machine learning approaches tend to favor neural networks for their power of abstraction, we stick to simple Logistic Regression models applied on pre-processed high level features as they seemed more adapted to the limited size of our training set and are more easily interpretable. We used Scikit-learn [30] implementation with L-BFGS solver, a maximum of 100 iterations and L2-regularization. Decision Tree classifiers and Support Vector Machine with linear kernel were also tested without significant improvements. In all our models, output classes weights are balanced with respect to their proportion in the dataset. The evaluation criterion is the F1-score, using cross-validation with leave-one-piece-out strategy to avoid overfitting due to similarities and repetitions inside movements.

4.3 Results and interpretations

The results are presented in Table 2. We observe that the presence of melody (M) is very well detected (F1-score of 0.912), despite being highly unbalanced in the dataset (94.6% of labels contains at least one layer with a melodic function). Harmonic or static functions are quite more difficult to predict. They were also more difficult to annotate: the determination of these functions also involves a part of subjectivity, despite the efforts made in the reviewing process to ensure the consistency of annotations.

The predictions of notes simultaneities and semblant motions (h, p, o) obtain fair results. We could have expected better from them, as well as for t (sustained notes) and r (repeated notes), since they only involved rare divergences during reviews, their determination being more straightforward. Furthermore, they are closer to the use of defined descriptors – notably those which are related, for instance, to the number of notes played at the same time or the presence of certain harmonic intervals in onsets or slices. The success of predicting h+ compared to specific relationships h, p or o is meaningful: considering the fact that parallel motions are specific cases of homorhythmy, it seems more practical and sound to focus on this more general case.

The poor results of models to predict oscillation b can be justified by the limited proportion of positive examples in the dataset (around 5.1%). Finally, the difficulty of annotating scale motives (s) in practice is reflected in their prediction result (F1-score of 0.36). From slow descending sequences of neighboring pitches to cells of short-duration notes that share the same repeated contour: many variants of this element can be found in the corpus, some of which were making consensus difficult to reach between annotator and reviewers. This issue can be adressed in the definition of the syntax as the need to refine the criteria and guidelines for the annotation of targeted concepts. The scale patterns sometimes span over wider regions than the duration of a bar, making the descriptors inadequate in this case. Providing a larger context to our models therefore appears as a promising perspective to improve the detection of this textural element.

5. CONCLUSION AND FURTHER WORKS

We provide an open dataset of texture annotations that specifically handles piano classical music. This dataset contains annotations of 9 movements of Mozart’s piano sonatas, all in Sonata Form, and annotation guidelines are provided in order to allow for an easy extension of the dataset by the community. It could be completed with other movements of Mozart’s piano sonatas, but also with more diverse piano music from the Classical Era. An extension of the texture syntax to consider specific textures from other periods would also certainly bring new insights for the study of composition styles.

The dataset can be used for musicological purposes, for instance to study the links between texture and a variety of annotations – harmony, modulations, cadences, phrases and section boundaries – that are common subjects of interest in the MIR community. It also opens perspectives in MIR for computer-aided analysis and composition. As a first application, we implemented a set of textural descriptors and used them for the prediction of textural elements, obtaining encouraging results. An in-depth analysis of the relations between the textural descriptors and textural elements could help improving the prediction, especially on harmonic, static layer detection, or scale motives. This work will also allow to study the evolution of texture in the musical pieces and to better integrate this dimension for automatic generation of music following textural scenarios.
6. ACKNOWLEDGEMENT

We would like to thank Nathalie Hérold for fruitful discussions during the initial phase of the project. We also thank the Algomus team and the anonymous reviewers for their helpful feedback. This research is partly funded by Région Hauts-de-France.

7. REFERENCES


