Learning Sonata Form Structure on Mozart's String Quartets

Pierre Allegraud, Louis Bigo, Laurent Feisthauer, Mathieu Giraud, Richard Groult, Emmanuel Leguy, Florence Levé

TISMIR

Allegraud, P., et al. (2019). Learning Sonata Form Structure on Mozart's String Quartets. *Transactions of the International Society for Music Information Retrieval,* 2(1), pp. 82–96. DOI: https://doi.org/10.5334/tismir.27

RESEARCH

Learning Sonata Form Structure on Mozart's String Quartets

Pierre Allegraud*, Louis Bigo*, Laurent Feisthauer*, Mathieu Giraud*, Richard Groult†, Emmanuel Leguy* and Florence Levé*,†

The musical analysis of large-scale structures, such as the classical sonata form, requires to integrate multiple analyses of local musical events into a global coherent analysis. Modelling large-scale structures is still a challenging task for the research community. It includes building large and accurate annotated corpora, as well as developing practical and efficient tools in order to visualize the analyses of these corpora. It finally requires the conception of effective and properly evaluated MIR algorithms.

We propose a machine learning approach for the sonata form structure on 32 movements from Mozart's string quartets. We release an open dataset, encoding two reference analyses of these 32 movements, totaling more than 1800 curated annotations, as well as flexible visualizations of these analyses. We discuss the occurrence in this corpus of melodic, harmonic, and rhythmic features induced by pitches, durations, and rests. We investigate whether the presence or the absence of these features can be characteristic of the different sections forming a sonata form. We then compute the emission and transition probabilities of several Hidden Markov Models intended to match the structure of sonata forms at several resolutions. Our results confirm that the sonata form is better identified when the parameters are learned rather than manually set up. These results open perspectives on the computational analysis of musical forms by mixing human knowledge and machine learning from annotated scores.

Keywords: Computational Music Analysis; Music Structure; Musical Form; Sonata Form

1 Introduction

1.1 Sonata form

The large-scale structure referred to as *sonata form* is a post-hoc formalization of a widely used composer practice since the middle of the 18th century. It is built on a *piece-level tonal path* concept involving both a *primary thematic zone (P)* and a contrasting *secondary thematic zone (S)* (**Figure 1**). This creates a polarization between two tonalities and induces a dramatic turn to the piece. The sonata form can be viewed as an evolution of both aria and concerto Baroque forms (Rosen, 1980; Hepokoski and Darcy, 2006). Greenberg (2017) investigated how sonata-form recapitulation may have come from both the double return of the tonic key and the parallel endings in a two-part movement.

A number of works composed by Haydn, Mozart and Beethoven are recognized as in sonata forms, especially first movements of string quartets, concerti, symphonies, and piano sonatas. However, the theories about the *classical sonata form* were introduced almost fifty years after its early golden era (Reicha, 1824; Marx, 1845; Czerny, 1848). One of its earliest formalizations seems to be the *grande coupe binaire* that Reicha (1824) described 30 years after Mozart died. The sonata form finally became a normative structure for several generations of romantic composers, being transmitted both through explicit teaching as well as implicit exposure.

Nowadays, sonata forms are still taught in music analysis, music history and composition lectures. They are also the focus of recent academic studies (Ratner, 1980; Rosen, 1980; Hepokoski and Darcy, 1997; Caplin, 1998, 2001; Hepokoski, 2002; Larson, 2003; Miyake, 2004; Hepokoski and Darcy, 2006; Gjerdingen, 2007; Greenberg, 2017). The past decades have seen a revival of the *Formenlehre* tradition in the classical era (Caplin et al., 2009). In Caplin (1998)'s theory of formal functions, small functional units at the idea level (e.g., basic idea, contrasting idea) are combined to form units at the phrase level (e.g., presentation, antecedent), which in turn are combined to form units at the theme level (e.g., sentence, period, etc.). This bottom-up approach builds up to the whole sonata form, paving the way to the three large-scale functions that are characteristic of sonata form: *Exposition, Development, and Recapitulation*, possibly including two other functions, *Introduction* and *Coda*. In this study, we rathe follow the *Sonata Theory* of Hepokoski and Darcy

* CRIStAL, UMR 9189, CNRS, Université de Lille, FR

† MIS, Université de Picardie Jules Verne, Amiens, FR

Corresponding author: Mathieu Giraud (mathieu.giraud@univ-lille.fr)

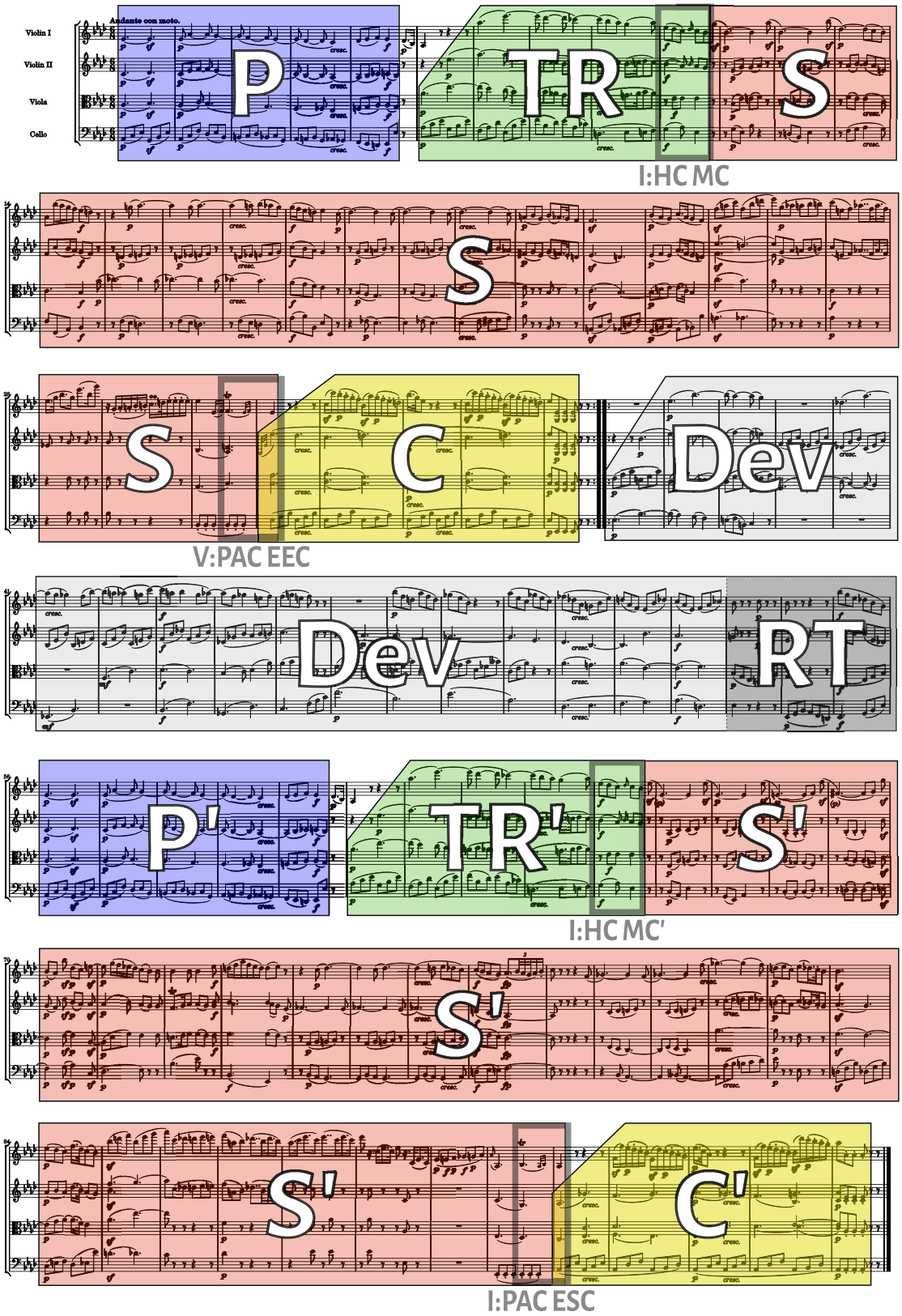


Figure 1: *Andante con moto* of the String Quartet #16 in E♭ Major, K 428, 2nd movement. Encoded in Lilypond by Maurizio Tomasi for the Mutopia Project. This slow movement has a sonata form, as detailed in Section 2.1. Following notations of Hepokoski and Darcy (2006), the primary themes (P/P') are followed by transitions (TR/TR'), ended with Medial Caesuras (MC/MC') – they are here Half Cadences (HC) in the main tonality (I). In the exposition, the secondary theme (S) and the conclusion (C) are here in the tonality of the dominant (V, E♭ major). In the recapitulation, both S' and C' come back to the main tonality. In the exposition, the S theme ends with a perfect authentic cadence (PAC) named essential expositional closure (EEC), whereas, in the recapitulation, the S' theme ends with an essential structural closure (ESC). Between the exposition and the recapitulation, the development (Dev) moves to other keys and is concluded by a retransition (RT) focusing on the dominant of the primary key.

(2006), where sonata form is viewed as an *"ordered system of generically available options permitting the spanning of ever larger expanses of time"* (ibid., p. 15). Their detailed formalization of the successive sections of the sonata form seems adequate to develop computational models.

1.2 MIR, high-level structure, and sonata form

On the one hand, *"analyzing a sonata form"*, which implies identifying the boundaries of its successive sections, often requires a number of musicological judgments that are piece-specific, which makes its automation difficult. Being strongly linked to music history, music analysis may indeed include ideas that involve the singularity of the piece, a comparison between composers as well as some aesthetic considerations. On the other hand, music analyses are often built upon specific *analytical elements*, like themes or patterns that structure the harmony and the texture of the piece. Analyses can therefore be modelled with Music Information Retrieval (MIR) algorithms that can be properly evaluated. Finally, the identification of a large-scale structure such as the sonata form requires the combination of these local features to reach a piece-level analysis, which is itself a challenge for MIR research. We previously reviewed research on *computational analysis of musical form* (Giraud et al., 2015). Chen et al. (2004) proposed to segment the musical piece into sections called "sentences", clustering phrases predicted by the LDBM algorithm by Cambouropoulos (2001). Rafael and Oertl (2010) built a global structure from patterns extracted by the algorithm from Hsu et al. (1998). Some studies, such as by Hamanaka et al. (2016), have attempted to compute large-scale structures as theorized by Schenker (1935) or later by the Generative Theory of Tonal Music (GTTM) of Lerdahl and Jackendoff (1983). Other works also modeled specific large-scale features, such as tonal tension (Lerdahl and Krumhansl, 2007; Farbood, 2010).

MIR modeling of high-level structures has also been employed in the field of *music generation*, wherein algorithms often have difficulties in producing long-term coherence. Herremans and Chew (2017) proposed to formulate this task as a combinatorial optimization problem. Nika et al. (2016) used harmonic scenarios to produce structured music improvisation. Medeot et al. (2018) elaborated a Recurrent Neural Network trained on a dataset of structural elements.

Finally, some research in the MIR community specifically targets sonata form structure: Jiang and Müller (2013) detected exposition/recapitulation pairs in Beethoven piano sonatas with self-similarity matrices. They also traced transpositions and harmonic changes through the different parts. Weiß and Müller (2014) proposed a model of "tonal complexity" and mapped it on sections of sonata forms. Baratè et al. (2005) introduced a model of sonata form structure based on Petri Nets. We previously proposed a model based on a Hidden Markov Model (HMM) emitting analytical features (Bigo et al., 2017). This model relied on human expertise, following the layout of sonata form as presented by Hepokoski and Darcy (2006). This previous approach was applied to a small set of pieces and the parameters of the model were hard-coded, based on music theory assumptions.

1.3 Contributions

Reproducible MIR research needs to be grounded on publicly available datasets. Here, we systematically study a corpus containing most of the sonata-form movements in Mozart's string quartets, and we release an open dataset providing two independent analyses of each movement, encoded manually, based on formal modeling of sonata form (Section 2). Extending the approach we introduced before (Bigo et al., 2017), we propose several models of sonata form using Hidden Markov Models for which parameters, emission probabilities, and transition probabilities are automatically learned on the corpus. The states of the HMMs represent the different sections of a sonata form and the observations consist of binary analytical features computed through the pieces (Section 3). We discuss the relationship between the occurrences of these features and the sonata form sections.

The results show that the sonata form is better identified when the parameters are learned rather than manually set up. We also study how the granularity of the model (i.e. the number of possible states) influences the success of the detection (Section 4).

2 The Mozart Sonata-Form String Quartet Corpus

2.1 Annotating sonata form

Annotating musical structure is challenging, subjective, and may involve different hypotheses from the analyst. Although different analysts might model sonata forms differently, there are points of consensus. In this work, we follow the notations of Hepokoski and Darcy (2006). Basically, a sonata form is built by following a *piece-level tonal path* involving a *primary thematic zone (P)* and a contrasting *secondary thematic zone (S)*. This is illustrated in **Figure 1** on a specific movement.

More precisely, the structure goes through the following parts:

- possibly an *introduction* (Intro);
- an *exposition* (Exp), including a thematic zone P in the main tonality (denoted by I), and a thematic zone S in an auxiliary tonality (usually the tonality of the dominant of I, denoted by V, for major-mode sonata movements). A transition (TR) bridges the two themes and triggers the modulation between the two tonalities. The transition ends with a perfect authentic or half cadence called the *Medial Caesura* (MC) (Hepokoski and Darcy, 1997), with "a decisive change of texture" (Rosen, 1980). The S zone generally concludes with a Perfect Authentic Cadence (PAC) called the Essential Expositional Closure (EEC). It is followed by a closing zone (C) rounding off the exposition by reinforcing the key of the EEC. The exposition is generally repeated once;
- a *development* (Dev) characterized by tonal instability, in which the existing themes are transformed and new themes can be introduced, possibly closed by a *retransition* (RT), that modulates back to the main tonality;
- a *recapitulation* (Rec) of P and S themes, now both in the tonality of the tonic, possibly including elements that were added throughout the development.

Recapitulation follows a layout analogous to the exposition (P', TR' ended with MC', S' ended with an Essential Structural Closure (ESC), C'). The transition TR' is generally the section that varies the most, in comparison with the exposition, as it does no longer need to include a modulation. One can often hear a move to the subdominant degree that remains in the home key, and thus resolves a "large-scale dissonance" (as called by Rosen (1980)) created by the exposition and intensified by the development;

- and possibly a *coda* (Coda).

Figure 2 displays layouts of sonata form at different granularity, including the sections described above along with short transitional sections. Some of these sections or transitional states may be skipped, leading to forward transitions between non-adjacent states. These models are seen as topologies of Hidden Markov Models, detailed in Section 3.

2.2 The corpus

The corpus used in this work includes 32 sonata-form movements of string quartets composed by Mozart. The pieces are encoded as `.krn` Humdrum files (Huron, 2002) downloaded from http://github.com/musedata/humdrum-mozart-quartets. These files were originally available from http://kern.humdrum.org and encoded by Edmund Correia, Jr. and Frances Bennion.

Between 1770 and 1790, Mozart composed 23 string quartets totaling 86 movements (King, 1968). We denote by K171.4 the 4th movement of K171. Out of these 86 movements, 42 are in sonata form, including 4 *rondo sonata* movements (K171.4, K173.1, K465.4, and K499.4), and 6 movements with special forms (K155.2, K168.2, K170.3, K171.1, K458.1, and K499.1). Special forms may include sections in unusual places, as for example the introduction and a "written" repeat of P' and TR' before the Coda in K171.1, or a strong bithematic unity (K168.2, *continuous exposition* in K458.1 "The Hunt" and K499.1). Ten out of these 42 sonata forms were left out because of unavailable clean encoding (K158.2, K160.1, K160.2, K160.3, K169.2, K170.3, K458.4, K464.1, K499.4, K575.1). Note that the dataset does not include pieces with an unusual sonata-form structure, such as K387.2, which is a minuet in sonata form without development, or K387.4, which is a fugue-sonata.

The corpus finally includes 19 first movements, 10 slow movements, and 3 final movements; 26 movements are in a major key and 6 are in a minor key.

2.3 Reference analyses

A reference annotation requires an agreement on a set of sections that need to be identified but also on the location of their boundaries. Some structural elements, such as the location of the cadences or the boundaries of the S theme, are especially subject to debate, and some of them may even be non-pertinent. For instance, there may be no precise border between P and TR. Reference datasets with divergent analyses may thus be particularly helpful. Following the above notations, we encoded two sets of analyses of the 32 sonata forms included in the corpus (**Figure 3**):

- The set F is an encoding of elements found in *Mozarts Streichquartette* by Marius Flothuis (1998). This book contains complete analyses of the quartets, including descriptions of P/TR/S/C sections in exposition and recapitulation that we formally encoded. Flothuis did not use the notations of Hepokoski and Darcy (2006) and took some liberties with the names of the sections. We freely interpreted his writings to match as much as possible the proposed model.
- The set A is our own analysis written following the notations described by Hepokoski and Darcy (2006). These analyses were checked by two curators. As Flothuis we encoded P/TR/S/C section boundaries, but also MC, EEC and ESC cadences, notable structures in the development and RT, as well as some patterns and some harmonic progressions. **Figure 4** shows how these analyses map onto some of the 18 possible sections. They have between 8 and 16 (average 11.9) of these 18 sections.

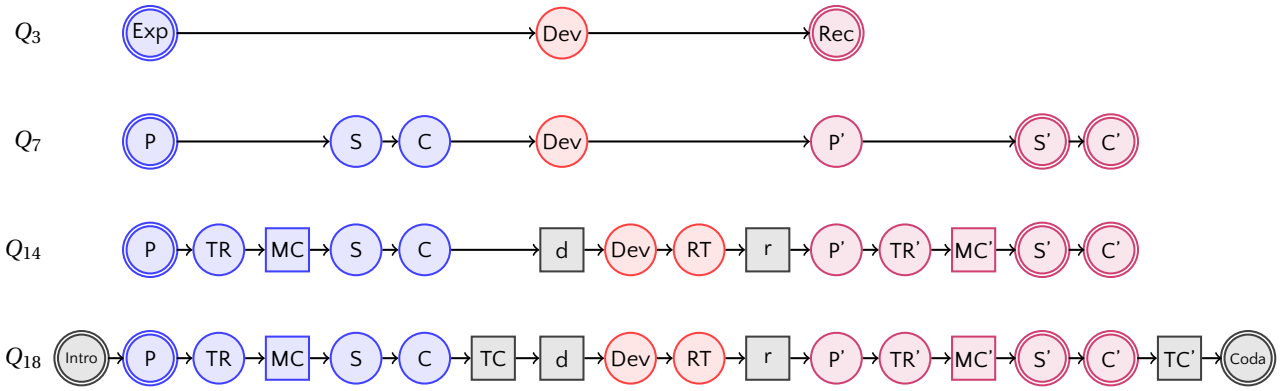
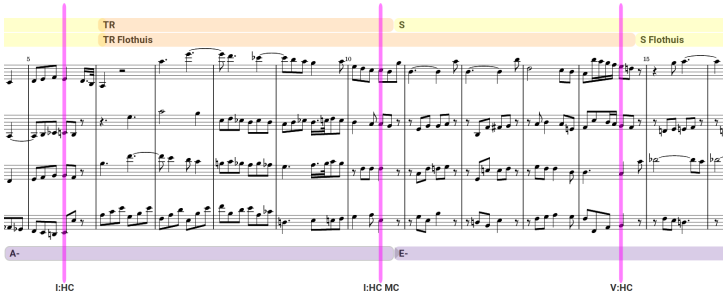


Figure 2: Model topologies describing the most common sonata form structure at several resolutions. The set of states Q_n has n states. Q_3 and Q_7 model the basic sections of the sonata form. Q_{14} (used by the model of Bigo et al. (2017)) and Q_{18} further model Intro, TR, RT and Coda sections as well as transitional states between these sections, represented with squares: the medial cesuras MC and MC', but also short transitions between the end of the closing zone and the complete end of the exposition (transition after the closing zone, TC), between the exposition and the development (d), between the development and the recapitulation (r), and between the recapitulation and the Coda (TC'). Initial and final states are circled twice.




```
{ "type": "Structure", "start": 0, "dur": 15, "tag": "P"},
{ "type": "Structure", "start": 0, "dur": 15, "tag": "P Flothuis"},

{ "type": "Structure", "start": 15, "dur": 14, "tag": "TR"},
{ "type": "Structure", "start": 15, "dur": 26, tag": "TR Flothuis"},

{ "type": "Structure", "start": 29.5, ..., "tag": "S"},
{ "type": "Structure", "start": 41.5, ..., "tag": "S Flothuis"},

{ "type": "Cadence", "start": 13.5, "tag": "I:HC"},
{ "type": "Cadence", "start": 28.5, "tag": "I:HC MC"},

{ "type": "Tonality", "start": 0,    "dur": 29, "tag": "A-"},
{ "type": "Tonality", "start": 29.5, "dur": 80, "tag": "E-"},
```


Figure 3: Extract of the reference analysis for the second movement of the String Quartet #16 in E♭ major (K428.2), as viewed on http://www.dezrann.net/ (left) and represented as a `json` file (right). The Primary theme (P) ends with a half cadence in the primary key (I:HC). Here a Transition zone (TR) begins, which stops on different beats according to the references. The A analysis starts the secondary (S) theme after the HC in the primary key on measure 10, whereas the F analysis rather starts it on measure 14 (HC on the dominant key). Onsets in the `json` file are expressed in quarter notes.

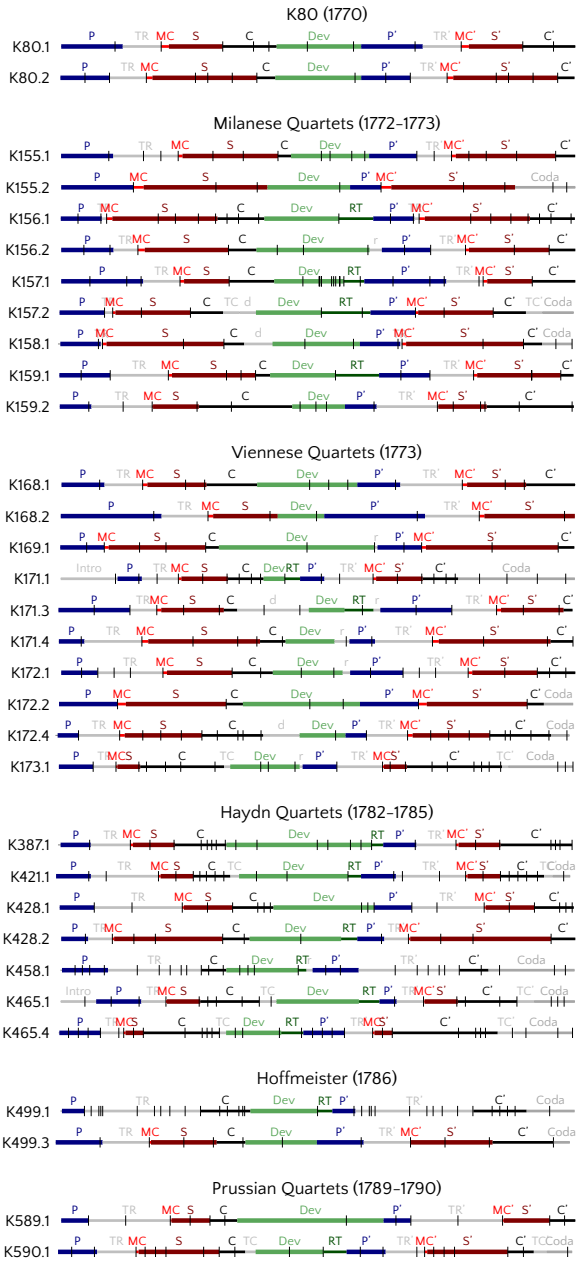


Figure 4: Reference analysis A of 32 sonata-form and sonata-form-like movements in Mozart string quartets. The analyses are projected on the 18 states of Q_{18}. Vertical lines show cadences.

The two encodings were done independently. They total 1939 labels, including more than 600 section labels and more than 500 cadences.

Despite some divergences (see **Figure 3**), 77% of the P/TR/S/C labels of A start at the same location in F. The majority of the differences between A and F occur when annotating the start of C. Indeed, Flothuis usually identifies the end of the S section on the first encountered PAC. On the contrary, Caplin (1998) usually extends S until a last strong PAC providing a conclusion to the theme or to a group of themes, and keeps in C only post-cadential material called *codettas*. We follow here the *first-PAC rule* as stated and nuanced by (Hepokoski and Darcy, 2006, p. 120 and 156):

> "(…) one could not consider S to be completed if either it or its cadential material is immediately restated. The PAC that ends the first statement of S proposes an EEC: by repeating the melody or a portion thereof, the composer reopens the PAC and shifts the EEC forward to the next PAC."

Indeed, Mozart frequently "reopens" PACs by repeating S material. He often restates the immediately preceding cadential progression and sometimes expands it. Thus, we identify an EEC when we encounter a PAC if what follows has not been heard shortly before.

Finally, 3 out of these 32 movements are differently annotated in the two sets of analyses: We see some movements as sonata forms, while Flothuis favors the loosened *two-part form* (K155.2, K168.2, K172.2). Moreover, he did not consider the form including a continuous exposition without a medial caesura (K458.1, K499.1).

2.4 Corpus availability

The annotation sets described above are distributed as Supplementary Files and at http://www.algomus.fr/data/ under the Open Database License (ODbL v1.0). These analyses are encoded as json files containing *labels*, each label being defined by a type (Structure/Cadence/Harmony), by an onset and possibly by a duration (**Figure 3**, right). Moreover, they are available through Dezrann, an interactive web platform for music annotation and analysis (Giraud et al., 2018, http://www.dezrann.net/).

3 Detection and Learning Strategy

As in (Bigo et al., 2017), we consider a finite alphabet of binary *analysis features* $\mathcal{A} = \{\alpha_1, \alpha_2 ...\}$ that may be present or absent at each quarter note and a Hidden Markov Model predicting the structure based on these features. Analysis features describe harmony, melody, or other local elements. In this section, we present the different models used in our experiments (section 3.1), the analysis features selected for this study (section 3.2), and the learning method used to set up the parameters of the model (section 3.3).

3.1 Hidden Markov Models to match sonata form structure

A Hidden Markov Model $\mathcal{M}_n = (Q_n, \pi, \tau, T, E)$ on $\mathcal{A}$ is defined by a set of n states $Q_n = \{q_1, ...q_n\}$ corresponding to the successive sections of sonata form. We experimented with different sets of states targeting several model topologies (**Figure 2**):

- The 3 states Q_3 = {Exp, Dev, Rec} and the 7 states Q_7 = {P, S, C, Dev, P′, S′, C′}, where the exposition and recapitulation parts of Q_3 are decomposed into thematic parts, match the most recognizable sections of sonata form;
- The 14 states Q_{14} and the 18 states Q_{18} are closer to sonata form structure as described by Hepokoski and Darcy (2006). They add the transitions TR, RT, TR', and (for Q_{18}) the Intro and Coda sections, and also model as short-lasting states the transitions between larger sections (MC, TC, d, r, MC', TC', see details in **Figure 2**).

The probabilities of the initial state and of the final state are respectively represented by $\pi = (\pi_1, ...\pi_n)$ and $\tau = (\tau_1, ...\tau_n)$. $T(i, j)$ is the *transition probability* – i.e. the probability that the state q_i goes to the state q_j, and $E(i, \alpha_k)$ is the *emission probability* – i.e. the probability that the state q_i emits the feature α_k.

Since several features can be predicted at the same step, any state may output simultaneously a set of symbols $A \subset \mathcal{A}$. If these emissions are independent events, the probability that the state q_i outputs the set A is

$$E(i, A) = \prod_{\alpha \in A} E(i, \alpha) \cdot \prod_{\alpha \in \mathcal{A} \setminus A} \left(1 - E(i, \alpha)\right)$$

Given an integer t, we define a *path* in $\mathcal{M}$ by a t-tuple of integers $P = (p_1, ..., p_t) \in [1, n]^t$, meaning that the path goes through the t states $q_{p_1} ...q_{p_t}$. We also consider a sequence of sets of symbols $A_1 ... A_t \in \mathcal{P}(\mathcal{A})^t$, where $\mathcal{P}(\mathcal{A})$ *is the set of subsets of* $\mathcal{A}$.

The probability that the model $\mathcal{M}$ follows a path $P = (p_1, ... p_t)$, entering by an input state p_1 and exiting from an output state p_t, while outputting the sequence $A_1 ... A_t$, one state outputting some symbols at each step, is given by:

$$\begin{aligned} prob(P, A_1 ... A_t) = \pi_{p_1} \quad & \cdot \quad E(p_1, A_1) \\ & \cdot \quad \prod_{i=2}^{t} T(p_{i-1}, p_i) \cdot E(p_i, A_i) \\ & \cdot \quad \tau_{p_t} \end{aligned}$$

Starting from a sequence of sets of symbols $A_1 ... A_t$, the Viterbi algorithm (Viterbi, 1967; Rabiner, 1989) finds the path P that maximizes $prob(P, A_1 ... A_t)$.

3.2 Analysis features

In (Bigo et al., 2017), we selected binary features "according to whether their presence or absence could be characteristic of (...) sections in a sonata form". We first included these features:

- Pattern features: repeated *candidate P pattern* (`pat:P`) and *candidate S pattern* (`pat:S`) that may be characteristic for P and S. These patterns are extracted from the highest voice (first violin), but successive occurrences may be found in other voices. The P candidate pattern is searched by a relatively strict variant of the Mongeau and Sankoff (1990) algorithm forbidding any transposition, whereas the S candidate pattern is searched with some transposition between the first occurrence and a next one – thus targeting a pattern that should appear in S' rather than again in S. Additional length and position constraints account for the balance of the sonata form, such as ending the candidate P pattern and starting the candidate S pattern before one-third of the length of the piece (Bigo et al., 2017).
- Harmonic features: *local tonalities* on 2-measure windows (2 × 7 `ton:x` features, minor and major) based on the algorithm of Krumhansl and Kessler (1982) using pitch class profiles adapted from Temperley (1999), heuristic detection of *Perfect Authentic Cadences* (`cad:PAC`), *Imperfect Authentic Cadences where both chords are in root position* (`cad:rIAC`), and pedals (`ped`), with the rule-based algorithms of Giraud et al. (2015), and finally features possibly involved in the preparation of half-cadences, such as *chromatic upward bass movements* (`harm:#`) and *diminished seventh or augmented second intervals* (`harm:7`).
- Features combining melody and/or harmony and/or rhythm: *full rests* (`rest`), *unisons* (`unison`), and finally *long harmonic sequences* (`seq`) where at least two voices repeat a pattern consecutively in different tonalities, the voices following the same (possibly diatonic) transpositions, for a duration of at least twenty quarter notes (Giraud et al., 2012).

We added the following two new features that may match more closely particular sections of the sonata form, like the Medial Caesura (**Figure 5**):

- *Rhythm break*. In both exposition and recapitulation, the end of the transition between the primary and the secondary theme is often enhanced by a dense and repetitive rhythm that is broken by the half-cadence of the Medial Caesura to enhance its closure effect (Hepokoski and Darcy, 1997). The feature `break` detects the interruption of repetitive rhythms, in any voice, that consist of at least 15 consecutive notes that have the same duration.
- *Triple hammer blow*. This striking event generally consists of three strongly repeated onsets preceding a

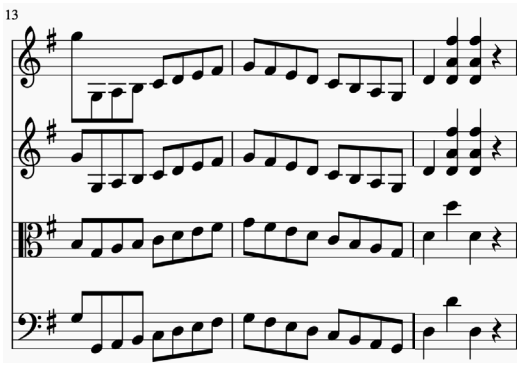


Figure 5: Medial Caesura in Allegro K80.2, measure 15. This half cadence (HC) has a very simple but very efficient tonic/dominant schema. It is reinforced by the sudden change of texture (`break`) between the unison in eighth notes and the *triple hammer blow* (`hammer`) that accentuates the dominant chord on *D*.

rest that separates the MC from the secondary theme (Hepokoski and Darcy, 1997). The feature `hammer` detects at any voice three repeated notes followed by a rest.

All the features consider only information on note pitches and durations as well as on rests. They do not look at any other information such as annotation marks, dynamics, or repeat bars. In particular, in almost all the pieces of the corpus, repeat bars are found at the end of the exposition and could ease the analysis. However, even without this repeat bar, this boundary is almost always unambiguous and can be predicted by automated methods.

The absence or presence of each feature is computed at every quarter note in every piece of the corpus. Features occurring at the limit between two sections are counted in both sections.

Note that all features are somewhat heuristic and may not be perfect. Nevertheless, the next section will show that some of them are significantly present or absent in some sections of the sonata form and that they may be used to learn the sonata-form structure.

3.3 Maximum likelihood parameter estimation

The parameters of the HMM can be learned by relating the section boundaries that are manually annotated in the whole corpus and the analysis features that are computed at each quarter note.

Let $\mathbb{T}(i,j)$ and $\mathbb{E}(i,\alpha)$ be the observed counts of transitions and emissions on the learning corpus, and $duration(i) = \sum_{k=1}^{n} \mathbb{T}(i,k)$ the total duration of the section i on the learning corpus. Any transition or emission probabilities can be computed by the following ratios:

$$\mathbb{T}(i,k) = \frac{\mathbb{T}(i,j)}{duration(i)} \qquad E(i,\alpha) = \frac{\mathbb{E}(i,\alpha)}{duration(i)}$$

To prevent zero probabilities, pseudo-counts with a very small ϵ are added to every value of $\mathbb{E}$ as well as to every value $\mathbb{T}(i,j)$ with $i \leq j$ (preventing backward transitions). Note that we considered that the features are independent both in the learning phase and when using the models. This is not true in the general case, especially for features that are mutually exclusive such as the tonality features, but this nevertheless allows for a practical approximation.

4 Evaluation and Results

Our experiments, including the computation of the analysis features and the HMM parameters, and the implementation of the Viterbi algorithm were done in python3 within the music21 framework (Cuthbert and Ariza, 2010), extended with analytic labels (Bagan et al., 2015). Every analytical feature was computed at each quarter note of every piece included in the corpus. Their occurrences in the corpus are discussed below.

To avoid overfitting, the learning strategy was evaluated with a *Leave-One-Piece-Out cross-validation* strategy. The sonata-form structure was predicted on each of the 32 pieces by the four HMMs described above, their parameters being learned on the 31 remaining pieces of the corpus. The cross-validation process was conducted on the whole corpus as the size and the heterogeneity of the corpus did not allow to have a separate test set dedicated to a final evaluation. Note that we did not identify any hyperparameter in the model that we tried to optimize, apart from the various topologies and feature subsets that are discussed below.

The results of the computation of the analysis features, as well as the learned probabilities, can be downloaded from http://www.algomus.fr/data/.

4.1 Discussion on feature statistics

Table 1 shows the number of occurrences of the computed features within the 18 sections of the sonata form as indicated by the annotation set A. Comparing occurrences of features or other elements against their expected number in "random" situations helps to evaluate their significance (Conklin and Anagnostopoulou, 2001). For example, the first primary zones (P) span 1130 quarter notes, that is 7.9% of the 14318 quarter notes of the corpus. In all the corpus, `ton:I` is activated on 4491 quarter notes. Should this feature be randomly distributed, `ton:I` would be activated on about 354 = 4491 × 7.9% quarter notes in P. However, there are actually 553 quarter notes out of these 1130 quarter notes in P where `ton:I` is activated.

For each feature and each section, *p*-values are estimated by an exact Fisher test computed by the Python `scipy` package. Fisher tests are computed independently. To account for the large number of tests, both on features and on sections, only features with *p*-values under 10^{-4} are considered as significant, either by their presence (**bold,** *) or their absence (***italic,*** *). For example, as expected, the feature `ton:I` is significantly present in P and significantly absent in S (both times $p < 10^{-30}$). The ≫ and ≪ symbols between two adjacent columns show the features which can be considered as significant to distinguish these two states, again with a 10^{-4} threshold on another Fisher test. For example, the feature `ton:II` is significantly more present in TR than in P ($p < 10^{-9}$), even if it is not significantly present in TR compared to all sections.

Table 1: Feature tallies for sections of the sonata form on the 32 movements of the corpus of Mozart string quartet movements in sonata form (see Section 4.1). The table shows, for each feature, the number of quarter notes where this feature occurs followed by its number of occurrences on quarter notes labeled as each of the sections in the reference annotation A, as well as, in gray, its expected number should the feature be random or uniformly distributed across the quarter notes. Bold, italic, and the ≫ and ≪ symbols indicate an estimation of the significance of their presence or absence compared to all the other sections as well as to adjacent sections. The total numbers of quarter notes can differ slightly from the sums of the different sections due to rounding of non-integer lengths on some sections (see Figure 3).

		Status																				
Features	**quarters**	**Intro**		**P**		**TR**		**MC**		**S**		**C**		**TC**		**d**		**Dev**		**RT**		...
`pat:P`	2448	35 20	≪	**858*** 193	≫	**341*** 235		15 13	≫	***23**** 250	≫	***0**** 200		***0**** 13		***0**** 12		***10**** 385		***8**** 50		
`pat:S`	3008	***0**** 25		***0**** 237	≪	**482*** 289		**36*** 16		**686*** 308	≫	**304*** 246		6 16		***0**** 14		***8**** 473		***0**** 61		
`ton:I`	4491	41 38		**553*** 354	≫	***311**** 432		22 23		***229**** 460		***222**** 367		18 24	≫	***0**** 22	≪	***360**** 707	≪	115 91		
`ton:II`	510	0 4		***18**** 40	≪	74 49		**12*** 2		72 52		**79*** 41		1 2		**14*** 2	≫	107 80		2 10		
`ton:III`	479	0 4		27 37		54 46		4 2		70 49		**64*** 39		8 2		5 2		**125*** 75		9 9		
`ton:IV`	1734	**31*** 14		**186*** 136	≫	***104**** 166		1 9		***34**** 177		***36**** 141	≪	21 9		4 8		***168**** 273		25 35		
`ton:V`	2514	***0**** 21		***41**** 198	≫	**467*** 242		**36*** 13		**683*** 257		**468*** 205	≫	***1**** 13		4 12		***327**** 396		70 51		
`ton:VI`	479	6 4	≫	***12**** 37	≫	54 46		2 2		**98*** 49		**66*** 39		0 2		2 2	≫	90 75	≫	***0**** 9		
`ton:VII`	386	9 3		20 30		30 37		0 2		***14**** 39		23 31		4 2		0 1		**94*** 60		1 7		
`ton:i`	892	3 7		84 70		***52**** 85		3 4		***16**** 91		***23**** 72	≪	12 4		**22*** 4		148 140	≪	**42*** 18		
`ton:ii`	534	4 4		33 42	≫	***10**** 51		0 2		***13**** 54		***17**** 43		1 2		0 2		**156*** 84		8 10		
`ton:iii`	356	6 3		***5**** 28	≪	**70*** 34		4 1		**68*** 36		46 29		0 1	≪	**12*** 1	≫	48 56		14 7		
`ton:iv`	349	**12*** 2		***6**** 27		***0**** 33		0 1		16 35	≪	**51*** 28		0 1		8 1		**114*** 54		18 7		
`ton:v`	460	0 3		22 36	≪	69 44		5 2		38 47		***8**** 37		3 2		1 2		**162*** 72	≫	2 9		
`ton:vi`	1052	0 8		***46**** 83		73 101		2 5		112 107		***51**** 86	≪	14 5	≫	0 5	≪	**368*** 165	≫	***0**** 21		
`ton:vii`	187	3 1		***0**** 14	≪	22 18		0 0		21 19		**36*** 15		0 1		4 0		14 29		0 3		
`cad:PAC`	416	4 3		20 32		22 40	≪	9 2		48 42		**72*** 34		1 2		0 2		***29**** 65		4 8		
`cad:rIAC`	142	2 1	≫	16 11		8 13		3 0		15 14		9 11		0 0		0 0		29 22		1 2		
`harm:#`	144	2 1		13 11		18 13		6 0	≫	7 14		5 11		1 0		1 0		27 22		1 2		
`harm:7`	1122	4 9		***49**** 88	≪	116 108		0 5		***68**** 115		86 91	≪	**18*** 6		3 5		**271*** 176		17 22		
`ped`	971	10 8		**116*** 76	≫	76 93		0 5		***42**** 99		66 79		1 5		0 4		186 152		20 19		
`rest`	331	6 2		35 26	≫	***11**** 31	≪	**12*** 1	≫	15 33	≪	38 27		4 1		2 1		39 52	≪	18 6		
`seq`	1254	24 10		***57**** 99		***61**** 120		2 6		95 128		***60**** 102		3 6	≪	**29*** 6	≫	**420*** 197	≫	***0**** 25		
`unison`	685	16 5		**91*** 54	≫	43 65		7 3		43 70		59 56	≪	**24*** 3		**27*** 3	≫	***68**** 107		12 13		
`break`	482	1 4		24 38		50 46	≪	**12*** 2	≫	36 49		47 39		5 2		1 2		74 75		9 9		
`hammer`	268	0 2		14 21		***8**** 25	≪	**14*** 1	≫	**49*** 27		20 21		0 1		0 1		52 42		7 5		
Total	14318	122		1130		1378		76		1468		1171		78		71		2255		292		...

Table 1 (continued): Feature tallies for sections of the sonata form on the 32 movements of the corpus of Mozart string quartet movements in sonata form.

Features	quarters	…	r		P'		TR'		MC'		S'		C'		TC'		Coda
		States															
pat:P	2448		5 5	≪	770* 184	≫	317* 243		15 12	≫	20* 264	≫	0* 218		0* 17		32* 125
pat:S	3008		0 6		1* 227	≪	444* 299		34* 15		692* 324	≫	307 269	≫	0* 20		7* 153
ton:I	4491		20 10		582* 339	≫	524* 447		44* 23		640* 484		497* 401	≪	17 31	≪	295* 229
ton:II	510		0 1		16* 38		28 50		5 2		25* 54	≪	56 45		0 3		0* 26
ton:III	479		1 1		18 36		24* 47		0 2		16* 51	≪	46 42		6 3	≫	1* 24
ton:IV	1734		0 3		156 130		230* 172		9 9		269* 187		256* 155		16 12		188* 88
ton:V	2514		4 5		36* 189	≪	132* 250		13 13		112* 271		77* 224		10 17		34* 128
ton:VI	479		3 1		18 36		35 47		0 2		58 51		27 42		0 3		7* 24
ton:VII	386		0 0		17 29	≪	80* 38		3 2		35 41		27 34	≪	12* 2	≫	18 19
ton:i	892		5 2		115* 67		148* 88		10 4		109 96		46* 79	≪	21* 6	≫	32 45
ton:ii	534		0 1		44 40		64 53		2 2		81 57		36 47		8 3		58* 27
ton:iii	356		0 0		9* 26		18 35		0 1		26 38		28 31		2 2		0* 18
ton:iv	349		0 0		9* 26		9* 34		0 1		23 37	≪	78* 31		0 2		4 17
ton:v	460		0 1		24 34		43 45		0 2		33 49		27 41		2 3		21 23
ton:vi	1052		0 2		57 79		77 104		2 5		104 113		68 94		2 7		77 53
ton:vii	187		0 0		0* 14	≪	34 18		1 0		23 20		17 16		7 1		5 9
cad:PAC	416		0 0		26 31		25 41		8 2		46 44		73* 37		1 2		28 21
cad:rIAC	142		0 0		18 10		5 14		1 0		17 15		9 12		0 0		9 7
harm:#	144		0 0		14 10		14 14		4 0		16 15		7 12		2 1		6 7
harm:7	1122		1 2		53* 84		93 111		0 5		100 121	≪	196* 100		12 7		35 57
ped	971		1 2		131* 73	≫	67 96		0 5		47* 104	≪	84 86		2 6	≪	120* 49
rest	331		5 0		35 24	≫	14 32	≪	14* 1	≫	16 35		33 29		3 2		31 16
seq	1254		0 2		58* 94	≪	172* 125		3 6		118 135		120 112		0 8		32* 64
unison	685		8* 1		96* 51	≫	44 68		7 3		43* 73		49 61	≪	27* 4	≫	22 35
break	482		3 1		24 36		52 48	≪	14* 2	≫	37 52		52 43		5 3		36 24
hammer	268		1 0		14 20		8* 26	≪	12* 1	≫	46 28		16 23		0 1		10 13
Total	14318	…	32		1081		1427		74		1545		1280		99		732

Although most features are not specific to a section, many of them differ significantly from one section to another and confirm their pertinence for the task of sonata form detection. A first observation is that the expected tonal path is confirmed by the `ton:x` features. Indeed, `ton:I` is met for most of the P quarter notes while `ton:V` and `ton:III` (dominant and relative major tonalities) are significantly present in S. This highlights the opposition between the two tonal zones of the exposition. As expected, this "large-scale dissonance" is resolved by the recapitulation. Indeed, both P' and S' are characterized by a high prevalence of `ton:I`.

Another result considering the tonality features is the symmetry between TR and TR'. Whereas TR usually induces an ascending fifth move from `ton:I` to `ton:V`, our results confirm that, in TR', Mozart often moves to `ton:IV` (called a *tonal adjustment* by Caplin (1998) or a *feint* by Rosen (1980) and Hepokoski and Darcy (2006)) in order to reach S' in `ton:I` with a move of the same interval.

The Perfect Authentic Cadences (PAC) are significantly present in C and C', and only there. Indeed, S and S' generally end with a strong structural EEC and ESC although the rest of S and S' do not significantly contain cadences.

The thematic pattern `pat:P` is significantly present for P and P', but also for TR and TR'. This is because the starts of TR and TR' are often the same. The thematic pattern `pat:S` is significantly present for S and S', but also for TR, C, TR' and C'. This is because the part of the exposition that is exactly transposed often starts (contrarily to **Figure 1**) inside TR and continues through S' and C'.

Features `break`, `harm:#`, and `rest` are especially significant on MC and MC'. Some of these features are triggered by the themes in P/P' or S/S' at relevant places. Long harmonic sequences and pedals significantly appear in the developments, but they are also present in other sections. In the small transitional sections before the development (TC, d), before the recapitulation (r), and before the Coda (TC'), many unisons are encountered, but again they are significantly found at other places as well.

4.2 Ability to retrieve the sonata-form structure

We evaluate the performance of the four HMMs with learned parameters ($\mathcal{M}_3$, $\mathcal{M}_7$, $\mathcal{M}_{14}$, and $\mathcal{M}_{18}$), as well as the HMM with hard-coded parameters proposed previously (Bigo et al., 2017) that we call $\mathcal{M}_{14}^*$.

4.2.1 Evaluation measures

Tables 2 (focus on quarter notes) and **3** (focus on boundaries) show the performance of the five HMMs using the cross-validation process described above on the 32 pieces of the corpus.

Table 2 shows F_1-measures for all the considered classifiers and for each predicted label. The top table further shows the confusion matrix for $\mathcal{M}_{18}$ that details for each predicted label (rows), the number of corresponding quarter notes in the reference annotation (columns). For example, the second row shows that 36 quarter notes are predicted as P but are labeled Intro in the reference annotation (*false positives*), whereas 751 quarter notes are labeled as P (*true positives*).

To evaluate the fact that the model is able to learn *transition probabilities*, we also compared the learned models to HMMs with "equal" transition probabilities (restricted to forward transitions) but with learned emission probabilities. We also show the best F_1-measure for "fixed" classifiers always predicting the same section. For example, the "fixed" classifier for Q_{18} on P always predicts P on the 14318 quarter notes of the corpus and has an F_1-measure of 0.15, far below the F_1-measure of 0.69 obtained by $\mathcal{M}_{18}$.

In **Table 3**, the first four columns (*main boundaries*) show the results of the evaluation on four boundaries (starts of sections S, Dev, P' and S') corresponding to milestones in the tonal path of sonata form. The last four columns (*all boundaries*) show results of the evaluation while considering the boundaries of all modeled sections. In what follows, the prediction of a section boundary is considered as "correct" (+ or =) if its distance from the corresponding boundary in the reference annotation is at most 3 measures.

4.2.2 Prediction evaluation

For the majority of the sections, the learned HMMs have much better F_1-measures than HMMs with equal transition probabilities, showing that the model can benefit from learned transitions.

Using the HMM $\mathcal{M}_{14}^*$ with hard-coded parameters successfully predicted 27 main boundaries (22%) and 89 out of all boundaries (25%). **Table 3** shows that learning parameters using the very simple $\mathcal{M}_3$ model gives a bad prediction, with 24 main boundaries correctly predicted. Indeed, as $\mathcal{M}_3$ merges P and S themes, even most tonality features are not very significant.

Better predictions are achieved by $\mathcal{M}_7$, $\mathcal{M}_{14}$, and $\mathcal{M}_{18}$. The model $\mathcal{M}_{14}$ correctly predicts 47 main boundaries (38%) and 125 (35%) out of all boundaries, improving the results obtained by the HMM with hard-coded parameters. F_1-measures are also improved for most of the sections. Even better results are obtained with $\mathcal{M}_{18}$ (41% and 38%). However, $\mathcal{M}_{18}$ models many sections. Some of the 18 corresponding states rarely appear over the pieces of the corpus to be consistently learned by the model, as shown by the very low F_1-measure on sections Intro, TC, d, RT, and TC'. For example, the Intro section is found in only two movements in the whole corpus, leading to incorrect predictions between Intro and P sections.

Note that many false positives reported in the confusion matrix for $\mathcal{M}_{18}$ come from only a few pieces. Indeed, 132 of the 134 = 49 + 85 quarter notes predicted as Dev instead of Intro or P come from the wrong prediction on K465.1 (see below and **Figure 7**), and 60 out of the 61 = 25 + 21 + 15 quarter notes predicted as C' instead of C, Dev, or RT come from the wrong prediction of K171.1 (data not shown).

Table 3 also shows the results on $\mathcal{M}_{18}$ while restricting the set of features. This confirms that `pat:P` and `pat:S` features are important to ground the prediction, but other features also contribute, even if the cadence features do not appear to improve the detection.

Finally, **Figure 6** details the success of the prediction for the start of each section. Apart from the trivial start of

Table 2: Classification results, with F_1-measures of the five studied HMMs as well as of baseline models on the 14318 quarter notes of the corpus against the reference A. The confusion matrix is detailed for $\mathcal{M}_{18}$: Each column denotes the quarter notes of a section in the reference analysis, and the rows show how these quarter notes are classified (after cross-validation (c-val.)) by $\mathcal{M}_{18}$. Underlined values are discussed in the text.

Q_{18}	Intro	P	TR	MC	S	C	TC	d	Dev	RT	r	P'	TR'	MC'	S'	C'	TC'	Coda
Intro	***0***	154	30	3	·	·	·	·	·	·	·	·	·	·	·	·	·	·
P	36	***751***	238	12	4	·	·	·	·	·	·	·	·	·	·	·	·	·
TR	1	86	***175***	10	121	47	·	28	35	·	·	16	32	4	29	·	·	·
MC	1	4	19	***3***	6	·	·	·	·	·	·	·	·	·	1	·	·	·
S	1	·	608	27	***588***	357	2	·	11	·	·	·	·	·	30	9	·	·
C	1	2	40	6	364	***355***	10	·	202	·	5	6	·	·	·	38	·	0
TC	·	23	68	·	5	1	***0***	21	114	·	·	12	·	·	·	·	·	·
d	3	·	29	·	5	2	6	***6***	101	·	·	9	·	·	·	·	·	·
Dev	49	85	134	11	268	353	60	16	***1320***	87	3	67	56	2	62	110	32	36
RT	30	24	20	·	30	12	·	·	393	***141***	14	57	51	3	12	·	·	5
r	·	·	·	·	·	·	·	·	20	25	***2***	49	2	·	·	·	·	·
P'	·	·	1	·	1	·	·	·	0	·	7	***713***	282	11	35	14		
TR'	·	·	·	·	·	·	·	·	·	·	·	46	***161***	4	174	3	·	1
MC'	·	·	1	·	1	·	·	·	·	·	·	2	18	***8***	6	·	·	3
S'	·	·	14	3	73	14	·	·	·	·	·	7	549	20	***471***	393	·	16
C'	·	·	·	·	·	25	·	·	21	15	·	58	197	10	353	***213***	32	58
TC'	·	·	·	·	·	4	·	·	34	·	·	9	45	3	42	49	***11***	12
Coda	·	·	·	·	·	·	·	·	2	24	·	28	32	8	328	463	24	***587***
quarter notes	122	1130	1378	76	1468	1171	78	71	2255	292	32	1081	1427	74	1545	1280	99	732
F_1 ($\mathcal{M}_{18}$, c-val.)	0.00	**0.69**	**0.18**	**0.05**	**0.38**	**0.32**	0.00	**0.05**	**0.53**	**0.26**	0.03	**0.66**	**0.18**	**0.15**	**0.30**	**0.19**	**0.07**	**0.53**
F_1 (equal)	0.00	0.56	0.14	0.04	0.29	0.24	0.00	0.00	0.30	0.12	**0.20**	0.42	0.02	0.00	0.19	0.15	0.00	0.26
F_1 (fixed)	**0.02**	0.15	**0.18**	0.01	0.19	0.15	**0.01**	0.01	0.27	0.04	0.00	0.14	**0.18**	0.01	0.19	0.16	0.01	0.10

Q_{14}	P	TR	MC	S	C	d	Dev	RT	r	P'	TR'	MC'	S'	C'
quarter notes	1130	1378	76	1468	1250	71	2255	292	32	1081	1427	74	1562	2095
F_1($\mathcal{M}_{14}$, c-val.)	**0.76**	0.17	**0.05**	**0.38**	**0.28**	**0.05**	**0.58**	**0.25**	0.03	**0.66**	0.18	**0.15**	0.28	0.56
F_1 ($\mathcal{M}_{14}^*$)	0.66	**0.35**	0.03	0.27	0.26	0.04	0.16	0.14	0.02	0.29	**0.33**	0.09	**0.29**	**0.61**
F_1 (equal)	0.40	0.05	0.00	0.20	0.04	0.00	0.16	0.08	**0.11**	0.23	0.00	0.00	0.12	0.31
F_1 (fixed)	0.15	0.18	0.01	0.19	0.16	0.01	0.27	0.04	0.00	0.14	0.18	0.01	0.20	0.26

Q_7	P	S	C	Dev	P'	S'	C'
quarter notes	2582	1471	1321	2580	2580	1565	2095
F_1 ($\mathcal{M}_7$, c-val.)	**0.65**	**0.37**	**0.25**	**0.68**	**0.54**	**0.33**	**0.54**
F_1 (equal)	0.50	0.36	0.23	0.44	0.39	0.18	0.37
F_1 (fixed)	0.31	0.19	0.17	0.31	0.31	0.20	0.26

Q_3	Exp	Dev	Rec
quarter notes	5374	2580	6240
F_1 ($\mathcal{M}_3$, c-val.)	**0.76**	**0.57**	**0.85**
F_1 (equal)	0.41	0.30	0.68
F_1 (fixed)	0.55	0.31	0.61

P, the boundary being the best predicted is the start of P', that is the start of the recapitulation.

Whereas the hard-coded $\mathcal{M}_{14}^*$ predicts 9 starts of P' exactly or within 1 measure compared to A, models $\mathcal{M}_3$, $\mathcal{M}_7$, $\mathcal{M}_{14}$, and $\mathcal{M}_{18}$ respectively predict 10, 15, 17, and 18 such boundaries. As P' always appears in the reference, no spurious P' is predicted. This success in detecting the start of P' is likely to come from the correlation between this section and features representing both the thematic patterns `pat:P` and the tonality `ton:I` which is strongly captured by the model as **Table 1** attests. TR and TR' sections are badly predicted, especially on their start,

Table 3: Number of boundaries predicted exactly or within one measure (+), within between 2 and 3 measures (=), beyond 3 measures (–) or not predicted (!), compared to the reference analysis A. The bottom part of the table shows results obtained with a subset of features.

	main boundaries (total: 124)				all boundaries			
	+	=	–	!	+	=	–	!
$\mathcal{M}_{14}^{*}$	23	4	54	43	68	21	154	115
$\mathcal{M}_{18}$	34	17	53	20	90	45	147	104
$\mathcal{M}_{14}$	31	16	56	21	87	38	146	87
$\mathcal{M}_{7}$	35	12	61	16	70	15	101	30
$\mathcal{M}_{3}$	16	8	40	0	46	8	42	0
$\mathcal{M}_{18}$, no `pat:P/pat:S`	13	7	97	7	32	29	229	96
$\mathcal{M}_{18}$, no `ton:*`	3	11	100	10	32	31	236	87
$\mathcal{M}_{18}$, no `cad:*`	35	16	57	16	90	40	159	97
$\mathcal{M}_{18}$, only `ton:*`	3	8	104	9	24	27	247	88
$\mathcal{M}_{18}$, no break features	33	12	61	18	85	36	168	97

which may be caused by the blend between P/P' and TR/TR' in our model.

As a global result, $\mathcal{M}_{18}$ correctly predicts the sections of 8 movements, only some sections of 20 movements, and incorrectly the sections of 4 movements.

4.3 Discussion on representative movements

Figure 7 illustrates 6 representative predictions performed by $\mathcal{M}_{18}$.

The structure of the *Adagio* K172.2 is almost perfectly predicted. Almost all sections in the reference analysis are found (7 out of 10, since the model does not predict C, C' nor Coda) and their starts are estimated on the correct beat or within 1 measure. The prediction for the *Andante con moto* K428.2 (see again **Figure 1**) is good in the exposition. The results in the recapitulation are degraded by the missing S' section in the prediction, the C' section starting far too early.

The model $\mathcal{M}_{18}$ predicts spurious Intro and/or Coda sections in different pieces such as in K428.1 or K428.2. This is due to the rarity of these sections in the corpus. These artifacts are not seen on $\mathcal{M}_{7}$ or $\mathcal{M}_{14}$. In K428.1 and K465.1, both $\mathcal{M}_{14}$ and $\mathcal{M}_{18}$ globally fail in predicting a pertinent structure, especially because they predict a too long development. Using a feature on the repeat bars would improve these predictions.

The *Allegro* K458.1 "The Hunt" is an example of a *continuous exposition* (Hepokoski and Darcy, 2006), with no MC/MC' or S/S' sections. The model nevertheless predicts these sections, and fails on many subsequent sections. Note that the reference F identifies an S section, but not at the same place as the one estimated by the model.

The *Allegro* K465.4 has a *rondo sonata* form: The movement follows the typical tonal path of sonata form, but the first theme P acts like a *chorus* that may be reused at other places – here also in Dev and Coda. It is another

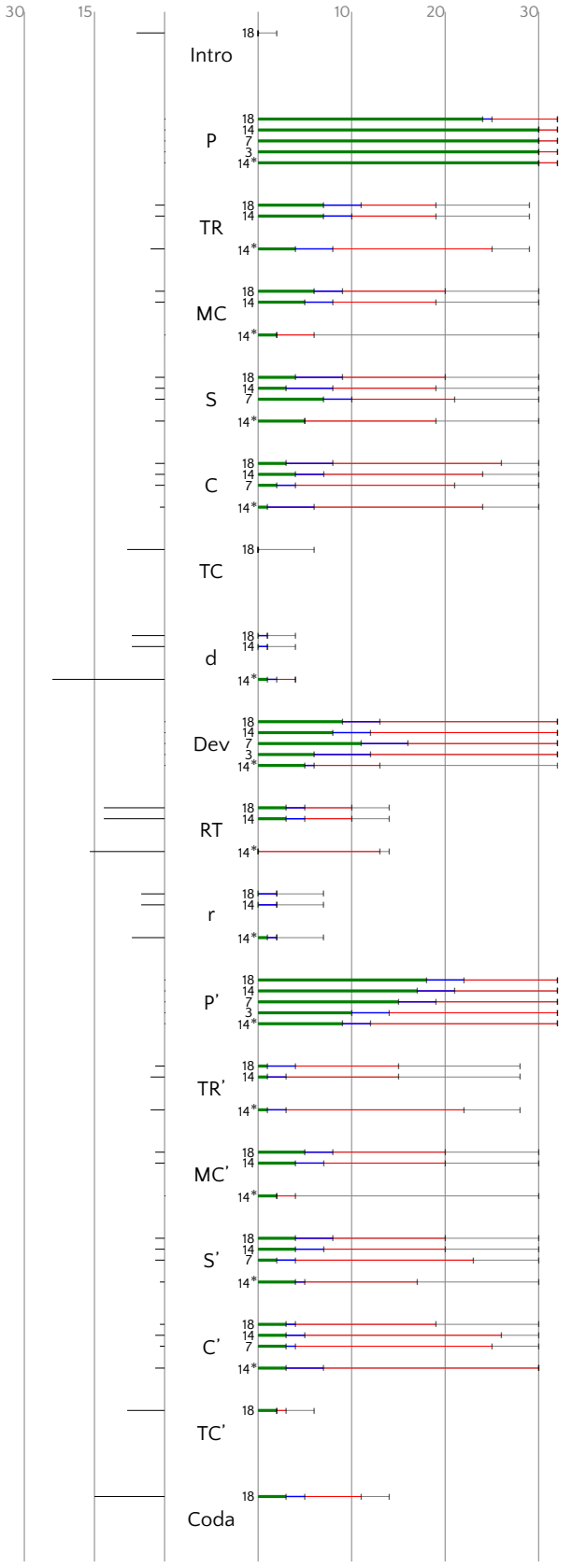


Figure 6: Detection precision (relative to the reference analysis A) of the five HMMs. Boundaries are predicted exactly or within 1 measure (green, + on Table 3), within between 2 and 3 measures (blue, =), more than 3 measures (red, –), or not predicted at all (gray, !). The lines at the left show the numbers of the spurious sections falsely predicted by the models.

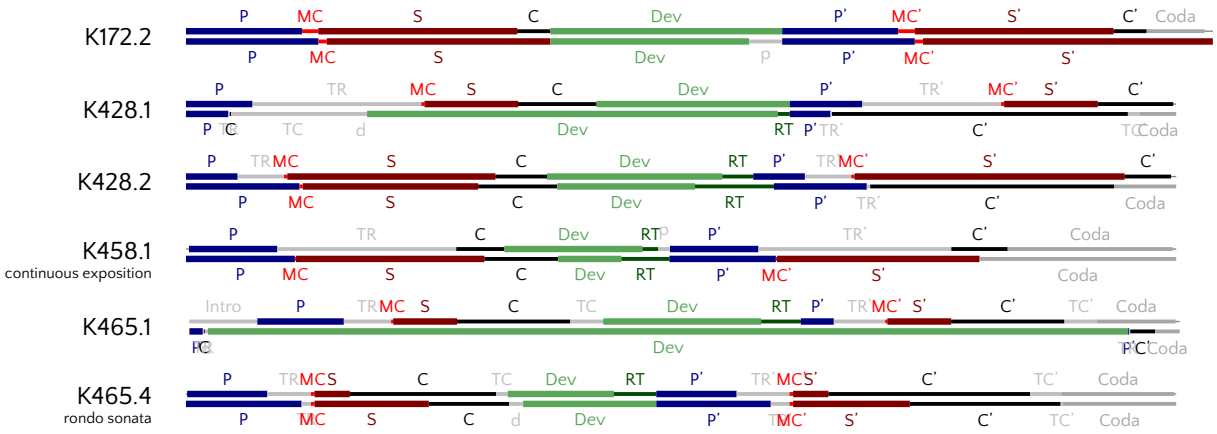


Figure 7: Comparison between the reference analysis A (top) and the predicted analysis by $\mathcal{M}_{18}$ (bottom) on six string quartet movements.

example of well-predicted form: the model correctly predicts the occurrence of 7 of the 15 sections annotated in the reference at the right beat or its neighborhood (P/MC/S/Dev/P'/MC'/S'). The end of S (and the start of C) is predicted at measure 104, whereas both the reference analyses A and F indicate it at measure 70, at the most satisfying and conclusive PAC. Since conclusions C and C' are very long and group several units, other analysts could reasonably agree with the model by including such thematic parts in S and S'. As in K465.4, the four rondo sonata forms in the corpus show satisfying results, even if the models have difficulty in correctly estimating the start of C.

5 Conclusions

We presented a new set of sonata-form annotations on 32 movements of Mozart string quartets and described how thematic, harmonic and rhythmic features are distributed across this corpus. Connecting both computed features and manual section annotations allows to learn parameters of Hidden Markov Models, enabling to retrieve some section boundaries of sonata form with better precision than manually set parameters.

Therefore, large music corpora can be analyzed by mixing human knowledge and learning from annotated scores. Somehow, this may be similar to the way composers learned and refined sonata form in a period of more than 150 years. On the one hand, the learning of emission and transition probabilities might reflect the human process of learning sonata form through instruction. On the other hand, modeling sonata form with unsupervised machine learning methods could be compared to the human process of learning sonata form by exposure without being aware of it.

Future directions of research include the modeling of sonata form with other learning models, either supervised, by following other theories of sonata form (e.g. Caplin (1998)) or unsupervised, as with HMMs by using the Baum-Welch algorithm. Recurrent neural networks may also provide better results, especially with layouts allowing to learn the positions where features tend to appear inside a section. However, the relatively small size of the corpus will be challenging for any such learning method.

Improvements might be obtained by enlarging the corpus and the set of selected features, including features using additional score elements, other than just notes. Pattern features could be extended. In particular, one may look for candidate patterns playing roles not only in the themes but also in the development. The impact of taking into account features at other resolutions than quarter notes could also be studied, especially when the tactus is not on quarter notes. Note also that most of our corpus is in the major mode. Further data could lead to the training of different models for major and minor keys.

Finally, other model topologies could analyze with more flexibility elaborated variations of sonata forms – especially continuous expositions as mentioned above – or focus on specific parts, such as the *rotations* in the development (Hepokoski and Darcy, 2006).

Acknowledgements

This project is partially funded by French CPER MAuVE (ERDF, Région Hauts-de-France) and by a grant from the French Research Agency (ANR-11-EQPX-0023 IRDIVE). We thank the editor and the anonymous reviewers for their insightful comments.

Competing Interests

The authors have no competing interests to declare.

References

Bagan, G., Giraud, M., Groult, R., & **Leguy, E.** (2015). Modélisation et visualisation de schemas d'analyse musicale avec music21. In *Journées d'Informatique Musicale (JIM 2015)*.

Baratè, A., Haus, G., & **Ludovico, L. A.** (2005). Music analysis and modeling through Petri nets. In *International Symposium on Computer Music Modeling and Retrieval (CMMR 2005)*, pages 201–218. DOI: https://doi.org/10.1007/11751069_19

Bigo, L., Giraud, M., Groult, R., Guiomard-Kagan, N., & **Levé, F.** (2017). Sketching sonata form structure in selected classical string quartets. In *18th International Society for Music Information Retrieval Conference (ISMIR 2017)*, pages 752–759.

Cambouropoulos, E. (2001). The local boundary detection model (LBDM) and its application in the study of expressive timing. In *International Computer Music Conference (ICMC 2001)*.

Caplin, W. E. (1998). *Classical Form: A Theory of Formal Functions for the Instrumental Music of Haydn, Mozart, and Beethoven*. Oxford University Press.

Caplin, W. E. (2001). The classical sonata exposition: Cadential goals and form-functional plans. *Tijdschrift voor Muziektheorie*, *6*(3), 195–209.

Caplin, W. E., Hepokoski, J., & **Webster, J.** (2009). *Musical Form, Forms & Formenlehre – Three Methodological Reflections*. Leuven University Press. DOI: https://doi.org/10.2307/j.ctt9qf01v

Chen, H.-C., Lin, C.-H., & **Chen, A. L. P.** (2004). Music segmentation by rhythmic features and melodic shapes. In *IEEE International Conference on Multimedia and Expo (ICME 2004)*, pages 1643–1646.

Conklin, D., & **Anagnostopoulou, C.** (2001). Representation and discovery of multiple viewpoint patterns. In *International Computer Music Conference (ICMC 2001)*, pages 479–485.

Cuthbert, M. S., & **Ariza, C.** (2010). music21: A toolkit for computer-aided musicology and symbolic music data. In *11th International Society for Music Information Retrieval Conference (ISMIR 2010)*, pages 637–642.

Czerny, C. (1848). *School of Practical Composition*. R. Cocks, London.

Farbood, M. (2010). A global model of musical tension. In *International Conference on Music Perception and Cognition (ICMPC 2010)*.

Flothuis, M. (1998). *Mozarts Streichquartette: Ein musikalischer Werkführer*. C. H. Beck.

Giraud, M., Groult, R., & **Leguy, E.** (2018). Dezrann, a web framework to share music analysis. In *International Conference on Technologies for Music Notation and Representation (TENOR 2018)*, pages 104–110.

Giraud, M., Groult, R., Leguy, E., & **Levé, F.** (2015). Computational fugue analysis. *Computer Music Journal*, *39*(2). DOI: https://doi.org/10.1162/COMJ_a_00300

Giraud, M., Groult, R., & **Levé, F.** (2012). Detecting episodes with harmonic sequences for fugue analysis. In *13th International Society for Music Information Retrieval Conference (ISMIR 2012)*, pages 457–462.

Gjerdingen, R. O. (2007). *Music in the Galant Style*. Oxford University Press.

Greenberg, Y. (2017). Of beginnings and ends: A corpus-based inquiry into the rise of the recapitulation. *Journal of Music Theory*, *61*(2), 171–200. DOI: https://doi.org/10.1215/00222909-4149546

Hamanaka, M., Hirata, K., & **Tojo, S.** (2016). Implementing Methods for Analysing Music Based on Lerdahl and Jackendoff's *Generative Theory of Tonal Music*. In Meredith, D., Editor, *Computational Music Analysis*, pages 221–249. Springer, Cham. DOI: https://doi.org/10.1007/978-3-319-25931-4_9

Hepokoski, J. (2002). Beyond the sonata principle. *Journal of the American Musicological Society*, *55*(2), 91. DOI: https://doi.org/10.1525/jams.2002.55.1.91

Hepokoski, J., & **Darcy, W.** (1997). The medial caesura and its role in the eighteenthcentury sonata exposition. *Music Theory Spectrum*, *19*(2), 115–154. DOI: https://doi.org/10.1525/mts.1997.19.2.02a00010

Hepokoski, J., & **Darcy, W.** (2006). *Elements of Sonata Theory: Norms, Types, and Deformations in the Late-Eighteenth-Century Sonata*. Oxford University Press. DOI: https://doi.org/10.1093/acprof:oso/9780195146400.001.0001

Herremans, D., & **Chew, E.** (2017). MorpheuS: Generating structured music with constrained patterns and tension. *IEEE Transactions on Affective Computing*. Early Access. DOI: https://doi.org/10.1109/TAFFC.2017.2737984

Hsu, J. L., Liu, C. C., & **Chen, A.** (1998). Efficient repeating pattern finding in music databases. In *International Conference on Information and Knowledge Management (CIKM 1998)*, pages 281–288. DOI: https://doi.org/10.1145/288627.288668

Huron, D. (2002). Music information processing using the Humdrum toolkit: Concepts, examples, and lessons. *Computer Music Journal*, *26*(2), 11–26. DOI: https://doi.org/10.1162/014892602760137158

Jiang, N., & **Müller, M.** (2013). Automated methods for analyzing music recordings in sonata form. In *14th International Society for Music Information Retrieval Conference (ISMIR 2013)*, pages 595–600.

King, A. H. (1968). *La Musique de chambre de Mozart*. Arles: Actes Sud.

Krumhansl, C. L., & **Kessler, E. J.** (1982). Tracing the dynamic changes in perceived tonal organisation in a spatial representation of musical keys. *Psychological Review*, *89*(2), 334–368. DOI: https://doi.org/10.1037//0033-295X.89.4.334

Larson, S. (2003). Recapitulation recomposition in the sonata-form first movements of Haydn's string quartets: Style change and compositional technique. *Music Analysis*, *22*(1–2), 139–177. DOI: https://doi.org/10.1111/j.0262-5245.2003.00178.x

Lerdahl, F., & **Jackendoff, R.** (1983). *A Generative Theory of Tonal Music*. MIT Press.

Lerdahl, F., & **Krumhansl, C. L.** (2007). Modeling tonal tension. *Music Perception*, *24*(4), 329–366. DOI: https://doi.org/10.1525/mp.2007.24.4.329

Marx, A. B. (1838, 1845). *Die Lehre von der musikalischen Komposition (volumes 2 and 3)*. Breitkopf & Härtel, Leipzig.

Medeot, G., Cherla, S., Kosta, K., McVicar, M., Abdalla, S., Selvi, M., Rex, E., & **Webster, K.** (2018). StructureNet: Inducing structure in generated melodies. In *19th International Society for Music Information Retrieval Conference (ISMIR 2018)*, pages 725–731.

Miyake, J. (2004). *The Role of Multiple New-key Themes in Selected Sonata-form Exposition*. PhD thesis, University of New York.

Mongeau, M., & **Sankoff, D.** (1990). Comparison of musical sequences. *Computers and the Humanities*, *24*(3), 161–175. DOI: https://doi.org/10.1007/BF00117340

Nika, J., Chemillier, M., & **Assayag, G.** (2016). Improtek: introducing scenarios into human-computer music improvisation. *Computers in Entertainment (CIE)*, *14*(2), 4. DOI: https://doi.org/10.1145/3022635

Rabiner, L. R. (1989). A tutorial on hidden Markov models and selected applications in speech recognition. *Proceedings of the IEEE*, *77*(2), 257–286. DOI: https://doi.org/10.1109/5.18626

Rafael, B., & **Oertl, S. M.** (2010). MTSSM – A Framework for Multi-Track Segmentation of Symbolic Music. *International Journal of Computer, Electrical, Automation, Control and Information Engineering*, *4*(1), 7–13. DOI: https://doi.org/10.2316/P.2010.674-008

Ratner, L. (1980). *Classical Music: Expression, Form, and Style*. Schirmer.

Reicha, A. (1824). *Traité de haute composition musicale*. A. Diabelli.

Rosen, C. (1980). *Sonata Forms*. W. W. Norton.

Schenker, H. (1935). *Der freie Satz*. Universal Edition.

Temperley, D. (1999). What's key for key? The Krumhansl-Schmuckler key-finding algorithm reconsidered. *Music Perception*, *17*(1), 65–100. DOI: https://doi.org/10.2307/40285812

Viterbi, A. (1967). Error bounds for convolutional codes and an asymptotically optimum decoding algorithm. *IEEE Transactions on Information Theory*, *13*(2), 260–269. DOI: https://doi.org/10.1109/TIT.1967.1054010

Weiß, C., & **Müller, M.** (2014). Quantifying and visualizing tonal complexity. In *Conference on Interdisciplinary Musicology (CIM 2014)*, pages 184–188.

How to cite this article: Allegraud, P., Bigo, L., Feisthauer, L., Giraud, M., Groult, R., Leguy, E., & Levé, F. (2019). Learning Sonata Form Structure on Mozart's String Quartets. *Transactions of the International Society for Music Information Retrieval,* 2(1), pp. 82–96. DOI: https://doi.org/10.5334/tismir.27

Submitted: 28 December 2018 **Accepted:** 31 October 2019 **Published:** 17 December 2019

]u[*Transactions of the International Society for Music Information Retrieval* is a peer-reviewed open access journal published by Ubiquity Press.

OPEN ACCESS