# Modeling and learning structural breaks in sonata forms

Laurent Feisthauer, Louis Bigo, Mathieu Giraud

# MODELING AND LEARNING STRUCTURAL BREAKS IN SONATA FORMS

**Laurent Feisthauer** **Louis Bigo** **Mathieu Giraud**
CRIStAL, UMR 9189, CNRS, Université de Lille, France
{laurent,louis,mathieu}@algomus.fr

## ABSTRACT

Expositions of Sonata Forms are structured towards two cadential goals, one being the Medial Caesura (MC). The MC is a gap in the musical texture between the Transition zone (TR) and the Secondary thematic zone (S). It appears as a climax of energy accumulation initiated by the TR, dividing the Exposition in two parts. We introduce high-level features relevant to formalize this energy gain and to identify MCs. These features concern rhythmic, harmonic and textural aspects of the music and characterize either the MC, its preparation or the texture contrast between TR and S. They are used to train a LSTM neural network on a corpus of 27 movements of string quartets written by Mozart. The model correctly locates the MCs on 14 movements within a leave-one-piece-out validation strategy. We discuss these results and how the network manages to model such structural breaks.

## 1. INTRODUCTION

### 1.1 Sonata Form

The classical sonata form shaped many musical works in the classical and the romantic period. It began to appear in the second half of the 18th century but was not formalized until the early 19th century. Recent theories on sonata forms emerged in the last decades, with various points of view [5, 7, 13], but nevertheless agree on its high-level structure, involving *Exposition*, *Development* and *Recapitulation* sections, and optional *Introduction* or *Coda* sections.

According to Hepokoski and Darcy, an Exposition may be either a *two-part exposition*, featuring two contrasting thematic zones, or a *continuous exposition*, with only one thematic zone [13]. The two-part exposition is characterized by two strong punctuation breaks (Figure 1). The first one is the *Medial Caesura (MC)* which closes the first part of the exposition. The second one is the *Essential Expositional Closure (EEC)*, which is a *Perfect Authentic Cadence (PAC)* that concludes the Secondary thematic

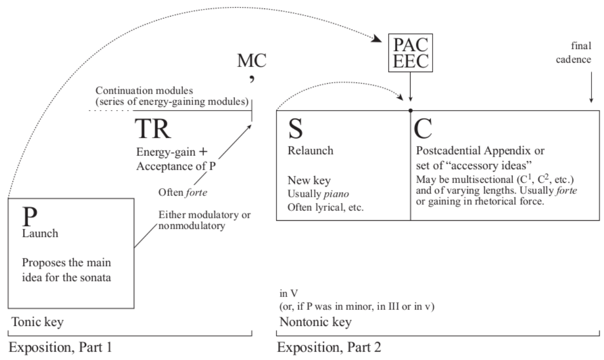


**Figure 1**. Structure of the two-part exposition in a sonata form, from Hepokoski and Darcy [13].

zone (S). The lack of a clearly articulated MC is the main difference between the two-part exposition and the continuous exposition [12]. Figure 1 shows the two-part exposition of a sonata form with its two punctuation points.

### 1.2 Formalizing Medial Caesuras

The MC is a "break in texture" [21] or a "textural change" [12] built around a cadence (most of the time, a half cadence (HC)), that acts as a boundary between TR and S. According to Hepokoski and Darcy, the MC has two fonctions [12, 13]: It closes the first part of the Exposition, concluding a process of energy gain initiated during the TR. It makes the second part available, opening a space for S thanks to that energy accumulated in TR. [1] Whether or not TR is modulatory, the MC is the point when the music reaches a structural dominant. [2] This dominant may be prolonged by neighbor motion, a repeated dominant pedal and a strong *forte*, and may be emphasized by *hammer strokes* (or *hammer blows*), that are repetitions of the final dominant chord. The whole process is concluded by the actual articulation of the MC. There can be a *general pause* on all voices, but there can also be a *caesura-fill*, that is a small melodic pattern or a sustained note or chord that "bridges the gap" to the S zone.Two examples of MCs are presented on Figures 2 and 3.



[1] Hepokoski and Darcy suggest that the MC "may be thought of as metaphorically analogous to the moment of the opening of elevator doors onto a higher floor." [13]

[2] This structural dominant is the *arrival point* of the associated HC [6], what follows being considered as *post-cadential*.

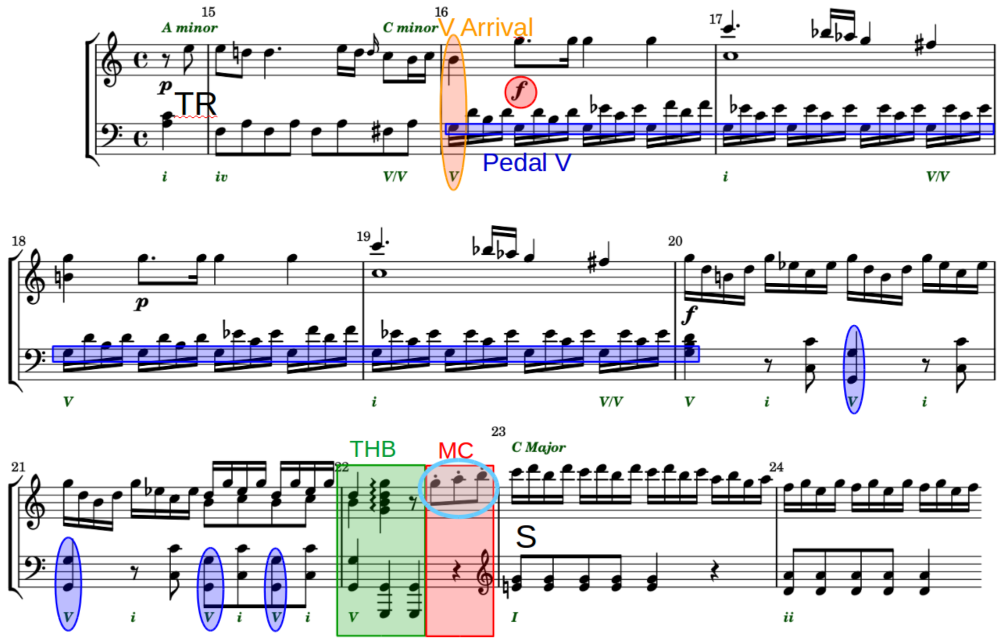


**Figure 2**. Mozart, *Piano sonata in A minor*, K310, 1st movement, mm15–24. The arrival point of the half cadence is on the downbeat of m16 (circled in orange). The measures 16 to 21 have a prolongational function. They are built on a *dominant pedal* (in blue) and a speeding-up harmonic oscillation between dominant and minor tonic of C minor. The dominant tension is reinforced by the *forte/piano/forte* contrast at m16, m18, and m20. This leads to a *triple hammer blow* (THB) at m22 (green) and the actual articulation of the MC on the fourth beat of m22 (red). Then *caesura-fill* at the right hand (lighter blue) leads to S and its new thematic unit.

Hepokoski and Darcy classify MCs according to the cadence occuring before the break and their position inside the Exposition [12]. A *first-level default* MC is associated with a HC in the secondary key [3] and can be denoted by V:HC MC (III:HC MC or v:HC MC in minor mode). This MC generally occurs between 25 and 50% of the length of the exposition, and sometimes at 60%. A *second-level default* MC considers a HC in the primary key (I:HC MC, between 15 and 45% of the length). A *third-level default* MC is caracterized by a PAC in the secondary key (V:PAC MC, III:PAC MC, or v:PAC MC), and occurs between 50 and 70%, sometimes at 75% of the length. The least encountered option, *fourth-level default*, is a PAC or an imperfect authentic cadence (IAC) in the primary key (I:PAC MC or IAC MC), generally at the end of P. In this case, S follows P without TR.

Another point of view on MCs is given by Richards [20], who identifies in 2013 seven *signals*, underlying the beginning of the secondary theme (S): tonic harmony of the new key, beginning function, preparation by a phrase-ending chord, textural gap of a medial caesura, change in texture, change in dynamics and characteristic melodic material. Each signal can be encountered in a *strong* or *weak* form. For example, the "Tonic harmony of new key" signal is strong when the chord encountered on the first downbeat is a tonic chord in the secondary key. It is weak when that chord is not the tonic of the secondary key, or when the chord is the tonic of another key due to a temporary modulation.

Somehow, a "textbook" MC – like the one displayed on Figure 2 – is heard when all these signals, either on MC or on start of S, are strong. Of course, MCs in the classical repertoire do not strictly follow these rules and are rather heard with many deformations, like the one presented on Figure 3.

### 1.3 Medial Caesura, Sonata Form and MIR

Working on high-level music structures is a challenge for Music Information Retrival (MIR) research [19]. To our knowledge, no previous study in MIR specifically targeted MC. However, several authors worked on sonata forms and designed algorithms to model or retrieve parts of its structure, on audio signals [15, 26] and on symbolic data [2, 22]. Sears and colleagues worked on cadences, by demonstrating that terminal events from cadential context are the most predictable thanks to a finite-context model (IDyOM) [23]. We previously worked on sonata form structure identification, modeling the MC as one state in Hidden Markov Models with 14 or 18 states and using a Viterbi algorithm to find back the sections from symbolic features [1, 4]. We also worked on PAC and HC detection thanks to features extraction and a SVM model. While PAC detection seems satisfying, HC detection was more challenging due to the lack of characteristic feature for this cadence type [3].

[3] The secondary key is often the dominant major (V) for major mode primary key and the relative major (III) or the dominant minor (v) for minor mode primary key.


**Figure 3**. Mozart, *String Quartet in D minor*, K421, 1st movement, mm22-24. This MC is weak: No dominant arrival, no lock on degree V, nor hammer strokes, and energy depletion rather than a gain. This MC weakness can be due to a I:HC MC denial on the third beat of m14 not shown on this figure (dominant arrival at m12). The composer delays the arrival of S and continues into TR, maybe to create surprise.

Modeling MCs is a challenging subject even amongst music theorists. As the MC is a striking event, playing a role in the high-level structure, one may wonder whether it is possible to model and predict MCs with computational musicology methods. This study tries thus to model features relevant to identify structural breaks such as the MC. We propose such features that may be specific to MCs, based on music theorist works [13, 20] (Section 2) and present a neural network model that we train on a corpus of expositions in Mozart string quartets (Section 3). We finally discuss the occurences of the features, detail how the network manages to model the MC, and propose perspectives (Sections 4 and 5).

## 2. FEATURES INDICATING THE MEDIAL CAESURA

We mostly here introduce features to model high-level signals leading to the Medial Caesura, taking inspiration from Hepokoski and Darcy as well as from Richards [13,20]. To find the MC, we want to model the MC, but also its preparation and the textural contrast with the beginning of S. We also use low-level features inspired by dedicated extraction software as jSymbolic [18]. This section details 13 *features* that are estimated on each beat of the music piece. The features are then used in the next section to train a neural network to model the MC. In contrast to frequent uses of neural networks that consist in automatically identify most relevant features, this research aims to validate the efficiency of a set of pre-determined theory driven features to model breaks such as the medial caesura.

In the following, given a onset $b$, the $[b, b+1[$ interval means that we consider each note actually sounding in that interval, including notes whose onset is $b$, or before, but still sounding on $b$, but excluding notes whose onset is $b+1$.

### 2.1 Rhythm, energy and textural features

Preparations of MCs are expected to be passages of high rhythmic intensity as a consequence of the repetition of the pedal and the *forte*. It might even be the most obvious way of gaining musical energy. We also expect a change of rythmic density between TR and S and a beat filled with silence on the articulation of the MC.

Modeling textural changes with precision is a challenging topic in computational musicology [11]. To try to capture these "breaks in texture" between the end of TR and the beginning of S, we implement the following low-level features, for each beat $b$:

- *f-rhythm-density* counts the number of notes in $[b, b+1[$,
- *f-rest* counts the number of voices not sounding on $b$.

We add these features, as the textural/energy change can be seen on the range of voices:

- *f-mean-pitch* is the mean of the MIDI values of pitches in $[b, b+1[$,
- *f-range-pitch* is the difference between the maximum and minimum MIDI values of pitches in $[b, b+1[$.

We also keep track of the position in the piece:

- *f-time* is the current beat number divided by the total number of beat in the score.

We propose one high-level feature specific to MCs:

- *f-hammer-blow* tracks double, triple or more hammer strokes (see THB on Figure 2). To estimate this feature, the set of pitches on $b$ are compared to sets of pitches on previous beats. This feature reaches its maximum when the set of pitches is the same for $b$ and for at least 2 previous beats.

### 2.2 Harmonic Features

We expect a (functional) dominant lock during the few beats before the articulation of the MC (whether it is a dominant on the primary key or on the secondary key).

Moreover, whether a TR is modulatory or not, we expect to encounter accidentals during TR or at least during the beginning of S. This is due to neighbor motion during the phase of prolongation of dominant and use of the dominant of the secondary key.

Algorithms for detecting local tonalities work well [17, 24] but tend to average variations over a window (often a few beats or measures). We propose rather two sets of features adapted to the detection of the MC.

**Functional harmony compatibility.** Functional harmonic analysis is a challenging problem in itself [9,25]. The idea here is to assert how compatible is a current harmony to

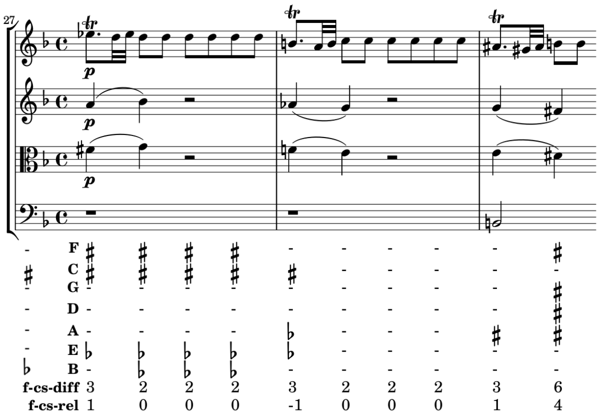


**Figure 4**. Estimation of the *current diatonic scale* on mm27-29 of Mozart's 1st movement of String Quartet No. 13 in D minor (K173). The D minor harmonic scale is given on the left. **f-cs-rel** stands for *f-current-scale-relative*.

a harmonic function, but without actually classifying the harmonies.

For a given functional harmony (as for example *f-predominant*, that may be either a ii or a IV/iv), we define a *compatibility profile* $h : P \to [-1, 1]$ that asserts how compatible a pitch $p \in P$ is to the given harmony. The pitch is given relative to the tonic of the primary key. Given a set of notes $c = \{p_1, p_2, ...\}$ from a given offset, the feature computes $\sum_{p \in c} h(p)$. The actual computation weights notes by their length in the given beat.

The compatibility profiles $h$ could be learned as profiles used in tonality detection [17]. We took here a simpler approach and encoded two pitches lists coming for musical knowledge, one with relevant pitches, the other with irrelevant pitches. We define five such features, each one with a particular list of relevant and irrelevant pitches:

| | relevant $h(p) = 1$ | irrelevant $h(p) = -1$ |
|---|---|---|
| *f-maj-tonic* | 1, 3, 5 | ♯4, 7 |
| *f-min-tonic* | 1, ♭3, 5 | ♮4, 7 |
| *f-predominant* | 2, 4, ♯4, ♭6, 6, 1, ♭2 | 3, 5, 7 |
| *f-dominant-of-dominant* | 2, ♯4, ♭6, 6, 1, ♭3 | 3, 4, 5, ♯5, 7 |
| *f-dominant* | 5, 7, 2, 4, 6, ♭6 | 1, ♯5 |

These pitch lists suppose that we have a pitch space with the pitch spelling information. Pitches $p$ that are not listed count for $h(p) = 0$.

**Harmony landscape and current scale.** To better capture the occurrence of new accidentals, we model the *current scale*. It is a diatonic scale containing the seven pitches with, for each of them, the last accidental encountered (Figure 4). We expect it to be different of its initial state at the beginning of the piece and to vary a lot just before the MC. We estimate two features on this scale :

- *f-current-scale-diff* counts the number of pitches differences in the current scale in $]b-1, b]$ related to the initial current scale.
- *f-current-scale-relative* further weights this count by $+1$ when the scale "gains" a sharp (or "loses" a flat) and by $-1$ in the other case.

| | Tonality | Cadence | MC | Tempo |
|---|---|---|---|---|
| K80.1 | G Major | V:HC | 1 | Adagio |
| K80.2 | G Major | V:HC | 1 | Allegro |
| K156.1 | G Major | I:HC | 2 | Presto |
| K156.2 | E minor | III:HC | 1 | Adagio |
| K157.1 | C Major | I:HC | 2 | N.A. |
| K157.2 | C minor | III:IAC | 4 | Andante |
| K158.1 | F Major | V:HC | 1 | Allegro |
| K159.1 | B♭ Major | V:HC | 1 | Andante |
| K159.2 | G minor | III:PAC | 3 | Allegro |
| K168.1 | F Major | I:HC | 2 | Allegro |
| K168.2 | F minor | V:HC | 1 | Andante |
| K169.1 | A Major | I:PAC | 4 | Molto Allegro |
| K171.3 | C minor | v:HC | 1 | Andante |
| K171.4 | E♭ Major | V:HC | 1 | Allegro assai |
| K172.1 | B♭ Major | V:HC | 1 | Allegro spiritoso |
| K172.2 | E♭ Major | V:HC | 1 | Adagio |
| K172.4 | B♭ Major | V:PAC | 3 | Allegro assai |
| K173.1 | D minor | v:HC | 1 | Allegro moderato |
| K387.1 | G Major | V:HC | 1 | Allegro vivace assai |
| K421.1 | D minor | III:PAC | 3 | Allegro |
| K428.1 | E♭ Major | V:PAC | 3 | Allegro non troppo |
| K428.2 | A♭ Major | I:HC | 2 | Andante con moto |
| K465.1 | C Major | V:HC | 1 | Adagio + Allegro |
| K465.4 | C Major | V:HC | 1 | Allegro |
| K499.3 | G Major | V:HC | 1 | Adagio |
| K589.1 | B♭ Major | V:PAC | 3 | Allegro |
| K590.1 | F Major | V:HC | 1 | Allegro moderato |

**Table 1**. The corpus contains 27 expositions (21 in major, 6 in minor) in Mozart String Quartets. "MC" denotes the MC type as designed by [12].

For example, on the Figure 4, the current scales are compared with the D minor harmonic scale (primary key, with B♭ and C♯). On the downbeat of measure 28, three pitches are changed (A♭, E♭, B♮), so *f-current-scale-diff* is 3 and *f-current-scale-relative* is $1 - 2 = -1$.

## 3. LEARNING STRATEGY

### 3.1 Network Layout

A Long Short-Term Memory neural network (LSTM) is built to predict the position of the Medial Caesura in the pieces of the corpus (Figure 5).

The network takes vectors of values describing the beats of the pieces as input. The identification of a Medial Caesura at a particular beat requires also to look at several past beats and possibly future beats. This is partly taken into account by the LSTM, but we further directly provide to the network the feature values over a time window. A

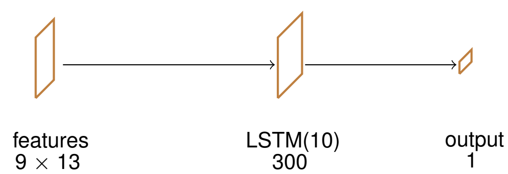


**Figure 5**. Network layout

vector describing a beat includes thus the 13 feature values corresponding to this beat but also those corresponding to the $p$ previous beats and the $n$ next beats. In this experiment, we set $p = 4$ and $n = 4$ which results in input vectors of size $9 \times 13 = 117$.

The input layer is fully connected to a recurrent hidden layer with 300 units. The time step of the LSTM is set to 10, meaning that 10 consecutive vectors are used to compute the probability of a MC occuring at one beat. Finally, the hidden layer is fully connected to the output layer that is a single unit. A sigmoid function scales the output value as an estimated probability in the interval $[0, 1]$.

### 3.2 Model Training

To avoid overfitting, the position of the Medial Caesura in a piece of the corpus is predicted with a model that has been trained on the whole corpus minus the piece itself. This is referred as *leave-one-piece-out validation* process.

The 13 features are computed at every beat of every piece in the training set. Each piece is represented as a sequence of *feature vectors* of size 117, each vector being associated with a specific beat.

Every feature vector of the training set is associated with a *label* having a value 1 (presence of an annoted MC in the next 5 beats) or 0 (otherwise). During the training, pairs *(feature vector, label)* are presented to the network by batches of size 200. An Adam optimization algorithm updates the unit weights to minimize a binary cross-entropy loss function over 60 epochs.

Given the small number of MCs in the corpus, there was no preliminary selection of a separated test set. This research primarily focuses on the validation of the musical features rather than on the classifier itself. For these reasons, the number and sizes of hidden layers, the batch size and the number of epochs as the optimization algorithm were selected among the most common values given the dimension and the quantity of input data. We did not try the optimize the hyper-parameters to avoid over-fitting.

## 4. EVALUATION

### 4.1 Two-part expositions in Mozart's String Quartets

Mozart wrote 23 string quartets totaling 86 movements, including 42 in sonata form [16]. Many of these quartets are encoded as `.krn` Humdrum files [14] that we downloaded from the `humdrum-mozart-quartets` repository at `github.com/musedata/`. Some of these movements were left out because of unavailable clean encodings or of other technical inconsistencies including the absence of Medial Caesura. The corpus finally contains 27 two-part expositions totaling 4179 beats. Medial Caesura annotations were taken from the sonata form annotation dataset we proposed in [1] and available at `www.algomus.fr/data`. These annotations include P, TR, MC, S, and C sections. Table 1 lists these 27 movements with their MC type. We denote by K171.4 the 4th movement of K171.

| | P | TR | MC | S | C |
|---|---|---|---|---|---|
| *f-rhythm-density* | 0.442 | 0.517 | 0.524 | 0.557 | **0.577** |
| *f-hammer-blow* | 0.080 | 0.087 | **0.143** | 0.070 | 0.059 |
| *f-rest* | **0.223** | 0.194 | 0.099 | **0.229** | 0.181 |
| *f-mean-pitch* | 0.770 | 0.768 | **0.786** | **0.782** | 0.766 |
| *f-range-pitch* | 0.373 | 0.434 | **0.508** | 0.420 | 0.441 |
| *f-time* | 0.053 | 0.150 | 0.189 | 0.270 | **0.350** |
| *f-maj-tonic* | **0.679** | 0.621 | 0.593 | 0.575 | 0.595 |
| *f-min-tonic* | **0.650** | 0.598 | 0.581 | 0.561 | 0.588 |
| *f-predominant* | **0.568** | 0.553 | **0.576** | 0.545 | 0.521 |
| *f-dominant-of-dominant* | 0.493 | 0.537 | **0.594** | 0.561 | 0.525 |
| *f-dominant* | 0.629 | 0.670 | **0.737** | 0.699 | 0.719 |
| *f-current-scale-diff* | 0.065 | 0.182 | **0.262** | 0.228 | 0.237 |
| *f-current-scale-relative* | 0.008 | 0.069 | 0.094 | **0.146** | 0.134 |

**Table 2**. Average value of the features according to the section on the whole corpus.

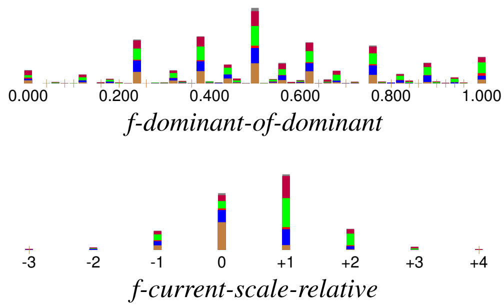


**Figure 6**. Distribution of two features relevant for MC identification. From bottom to top, P (brown), TR (blue), MC (red), S (green), and C (purple).

### 4.2 Implementation

We encoded feature extraction within the Python music21 framework [10]. The pitch space is Base40, modeling full pitch spelling information. Features described in Section 2 are computed at each beat of each piece and have their values scaled between 0 and 1 through min-max normalization. The values of the features are available as open data at `www.algomus.fr/data`. The neural network has been implemented with the Python framework Keras [8].

### 4.3 Features distribution

Table 2 shows the average values of each feature depending on the sections, and Figure 6 details the distribution of two relevant features. Several features have larger values on the MC, notably *f-dominant-of-dominant* and *f-dominant*. As expected, the features on tonic harmonies have higher values on P and TR, while features on dominant harmonies have higher values on the MC, S, and C sections. The *current-scale* features are very low on P (initial tonal stability), but are then mostly activated on the other sections as the music moves to another key, and preferably one with more sharps. The *f-current-scale-diff* is maximal around the MC, reflecting the harmonic oscillations and tonal instability at this place. This behaviour is less visible on *f-current-scale-relative*. Indeed, we observe sometimes here a modal instability (as in Figure 2) that means more flats before the MC in major mode. Other relevant features

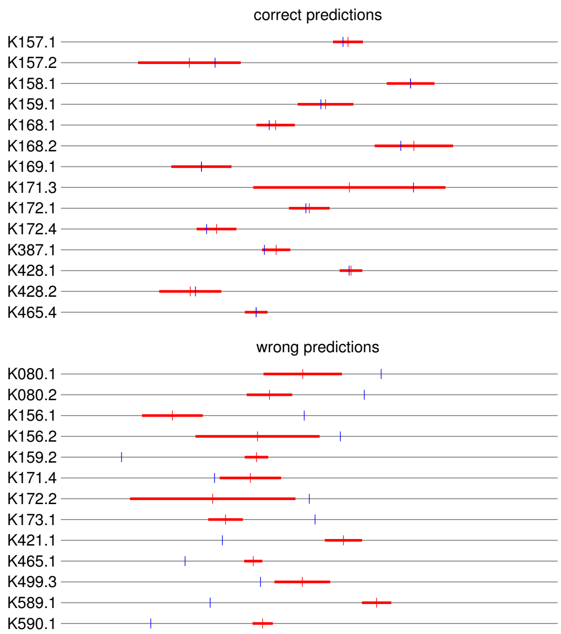


**Figure 7**. MC in the reference annotation (red marks on bold segments, representing $\pm 4$ beats) and predicted by the model (blue marks).

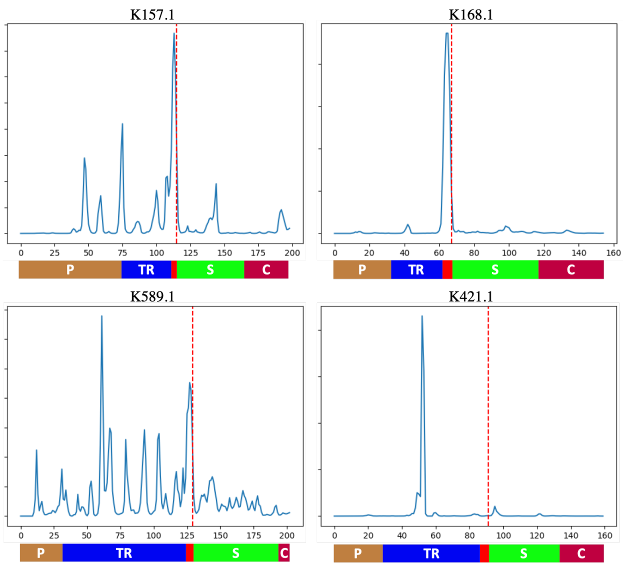


**Figure 8**. Probability curves for each beat to be an MC along four pieces of the corpus, two with correct MC predictions (top) and two with wrong MC predictions (bottom). Below each curve is the structure of the exposition in the reference annotation (P/TR/MC/S/C), and the red dashed lines emphasize the position of the MCs.

for the MC are, as expected, *f-hammer-blow*, and *f-range-pitch*, emphasizing the octave leap down often found on hammer blows.

### 4.4 MC prediction

In order to predict the position of the MC in an unseen piece, the sequence of vectors representing all beats are presented to the network. The position of the MC is identified at the offset where the network estimates the maximum probability. We consider that a prediction is correct when the predicted MC is less than 4 beats before or after the annotated MC which seems reasonable given the progressive aspect of the MC phenomena.

Figure 7 displays the location at which the MC is identified as the MC original annotations for each piece of the corpus. The network correctly locates the MC of 14 of the 27 pieces of the corpus. This is a improvement from our previous work [1] where the model found only 8 MCs out of the same 27 pieces. Due to the small size of the corpus, we did not find any significant correlation between the accuracy of the prediction and the piece mode, tempo, or MC type. For example, MCs are correctly estimated in 11 out of 21 pieces in major mode and in 3 out of 6 pieces in minor mode.

Figure 8 displays the estimation of the probability of having a MC at each beat of four pieces of the corpus. These probabilities are computed by different models that have been trained on the whole corpus, except on the piece on which the prediction is performed. The model works well on some pieces. In K157.1, the highest peak predicts well the MC. The second highest peak, on beat 76, is also noteworthy as it is an HC ending the P section. Other peaks are not well explained. In K168.1, there is a unique peak at the correct position of the MC. The model fails to predict the MC in other pieces. In K421.1 (see also Figure 3), a MC is wrongly predicted with high confidence about 40 beats too early, at measure 14. This false prediction is actually a *denied MC*. In K589.1, there are more candidates for the MC location, but with low estimated probabilities, under 0.2. Another peak triggers the detection around beat 60, where there are two beats with only rests.

## 5. CONCLUSION AND PERSPECTIVES

We proposed theory driven features modeling structural breaks such as the Medial Caesura. Trained with only these features and without any other note information, the network succeeds in identifying about half of the MCs of the corpus. This is notable given the diversity of the realisations of MCs in such a small corpus. The success of the predictions does not seem to be correlated with the tempo, mode or the MC type.

The model might probably be improved both by enlarging the corpus and by taking into account additional elements that can not be retrieved from the files used in these experiments, such as dynamics. Features were selected based on music theory. It could be worth learning also more lower-level elements used in their computations, such as note pitches and durations. Furthermore, the performance of the model might be improved by considering additional musical features that have been proposed for other MIR tasks such as cadence detection or sonata form retrieval.

## 6. REFERENCES


[1] Pierre Allegraud, Louis Bigo, Laurent Feisthauer, Mathieu Giraud, Richard Groult, Emmanuel Leguy, and Florence Levé. Learning sonata form structure on Mozart's string quartets. *Transactions of the International Society for Music Information Retrieval*, in revision.

[2] Adriano Baratè, Goffredo Haus, and Luca A Ludovico. Music analysis and modeling through Petri nets. In *International Symposium on Computer Music Modeling and Retrieval (CMMR 2005)*, pages 201–218, 2005.

[3] Louis Bigo, Laurent Feisthauer, Mathieu Giraud, and Florence Levé. Relevance of musical features for cadence detection. In *International Society for Music Information Retrieval Conference (ISMIR 2018)*, Paris, France, 2018.

[4] Louis Bigo, Mathieu Giraud, Richard Groult, Nicolas Guiomard-Kagan, and Florence Levé. Sketching sonata form structure in selected classical string quartets. In *International Society for Music Information Retrieval Conference (ISMIR 2017)*, pages 752–759, 2017.

[5] William E. Caplin. *Classical Form: A Theory of Formal Functions for the Instrumental Music of Haydn, Mozart, and Beethoven*. Oxford University Press, 1998.

[6] William E. Caplin. The classical cadence: Conceptions and misconceptions. *Journal of the American Musicological Society*, 57:51–117, 2004.

[7] William E. Caplin, James Hepokoski, and James Webster. *Musical Form, Forms & Formenlehre – Three Methodological Reflections*. Leuven University Press, 2009.

[8] François Chollet et al. Keras. `https://github.com/fchollet/keras`, 2015.

[9] Nathaniel Condit-Schultz, Yaolong Ju, and Ichiro Fujinaga. A flexible approach to automated harmonic analysis: Multiple annotations of chorales by bach and prætorius. In *International Society for Music Information Retrieval Conference (ISMIR 2018)*, pages 66–73, 2018.

[10] Michael Scott Cuthbert and Christopher Ariza. music21: A toolkit for computer-aided musicology and symbolic music data. In *International Society for Music Information Retrieval Conference (ISMIR 2010)*, pages 637–642, 2010.

[11] Mathieu Giraud, Florence Levé, Florent Mercier, Marc Rigaudière, and Donatien Thorez. Towards modeling texture in symbolic data. In *International Society for Music Information Retrieval Conference (ISMIR 2014)*, pages 59–64, 2014.

[12] James Hepokoski and Warren Darcy. The medial caesura and its role in the eighteenth-century sonata exposition. *Music Theory Spectrum*, 19(2):115–154, 1997.

[13] James Hepokoski and Warren Darcy. *Elements of Sonata Theory: Norms, Types, and Deformations in the Late-Eighteenth-Century Sonata*. Oxford University Press, 2006.

[14] David Huron. Music information processing using the Humdrum toolkit: Concepts, examples, and lessons. *Computer Music Journal*, 26(2):11–26, 2002.

[15] Nanzhu Jiang and Meinard Müller. Automated methods for analyzing music recordings in sonata form. In *International Society for Music Information Retrieval Conference (ISMIR 2013)*, pages 595–600, 2013.

[16] Alec Hyatt King. *La Musique de chambre de Mozart*. Arles : Actes Sud, 1968.

[17] Carol L. Krumhansl and Edward J. Kessler. Tracing the dynamic changes in perceived tonal organisation in a spatial representation of musical keys. *Psychological Review*, 89(2):334–368, 1982.

[18] Cory McKay, Julie Cumming, and Ichiro Fujinaga. JSymbolic 2.2: Extracting features from symbolic music for use in musicological and MIR research. In *International Society for Music Information Retrieval Conference (ISMIR 2018)*, pages 348–354, 2018.

[19] David Meredith. *Computational Music Analysis*. Springer, 2015.

[20] Mark Richards. Sonata form and the problem of second-theme beginnings. *Music Analysis*, 32(1):3–45, 2013.

[21] Charles Rosen. *Sonata Forms*. W. W. Norton, 1980.

[22] David Sears, William E. Caplin, and Stephen McAdams. Perceiving the classical cadence. *Music Perception*, 31(5):397–417, 2014.

[23] David R. W. Sears, Marcus T. Pearce, William E. Caplin, and Stephen McAdams. Simulating melodic and harmonic expectations for tonal cadences using probabilistic models. *Journal of New Music Research*, 47(1):29–52, 2016.

[24] David Temperley. What's key for key ? the Krumhansl-Schmuckler key-finding algorithm reconsidered. *Music Perception*, 17(1):65–100, 1999.

[25] David Temperley and Daniel Sleator. Modeling meter and harmony: A preference-rule approach. *Computer Music Journal*, 23(1):10–27, 1999.

[26] Christof Weiß and Meinard Müller. Quantifying and visualizing tonal complexity. In *Conference on Interdisciplinary Musicology (CIM 2014)*, pages 184–188, 2014.