# A CROSS-VERSION APPROACH FOR STABILIZING TEMPO-BASED NOVELTY DETECTION

**Meinard Müller**[1,3], **Thomas Prätzlich**[2,3], **and Jonathan Driedger**[1,2]
[1] Bonn University, [2] Saarland University, [3] MPI Informatik
{meinard,tpratzli}@mpi-inf.mpg.de, driedger@iai.uni-bonn.de

### ABSTRACT

The task of novelty detection with the objective of detecting changes regarding musical properties such as harmony, dynamics, timbre, or tempo is of fundamental importance when analyzing structural properties of music recordings. But for a specific audio version of a given piece of music, the novelty detection result may also crucially depend on the individual performance style of the musician. This particularly holds true for tempo-related properties, which may vary significantly across different performances of the same piece of music. In this paper, we show that tempo-based novelty detection can be stabilized and improved by simultaneously analyzing a set of different performances. We first warp the version-dependent novelty curves onto a common musical time axis, and then combine the individual curves to produce a single fusion curve. Our hypothesis is that musically relevant points of novelty tend to be consistent across different performances. This hypothesis is supported by our experiments in the context of music structure analysis, where the cross-version fusion curves yield, on average, better results than the novelty curves obtained from individual recordings.

## 1. INTRODUCTION

Music is highly structured data. Structure in music arises from repetitions, contrasts and homogeneity in musical aspects such as melody, dynamics, harmony, timbre or tempo [12]. The extraction of the musical structure from audio recordings is an important task in the field of music information retrieval. It consists of a segmentation problem, where the goal is to find the boundaries that mark the transitions between two structural parts, and a musically meaningful labeling (e.g. chorus, verse, first theme, second theme) of the segments, see [3, 12] for an overview. In many cases, segment boundaries are accompanied by a change in instrumentation, dynamics, harmony, tempo, or some other characteristics. The task of *novelty detection* is to specify points within a given audio recording where a human listener would recognize such a change [6,9,14,15].



Such points of novelty are not only of musical relevance, but also allow for speeding up further music analysis tasks [11].

In this paper, we present a general approach for stabilizing novelty-based segmentation techniques. Following [6], we first convert the audio signal into a suitable feature representation, compute a self distance matrix, and derive a novelty curve by detecting 2D corner points in this matrix. The choice of features (e. g. MFCCs, chroma features, tempogram features) depends on the musical aspects (e. g. timbre, harmony, tempo) of interest [9]. In the following, we consider the aspect of tempo using the cyclic tempogram features as proposed in [7] as an illustrative example. Particularly in classical music, there often exist many different recordings for a given piece of music. Even though all recordings follow the same musical score, two distinct versions may differ significantly in performance aspects regarding tempo, dynamics, or timbre. This is the reason why novelty detection results often vary across different audio versions.

The main contribution of this paper is to apply the novelty detection simultaneously to a set of different performances of a given piece. To this end, using a score-based MIDI reference, we convert the physical time axis (in seconds) of all version-dependent novelty curves into a common musical time-axis (in measures) . Then we combine the individual curves into a cross-version fusion curve, see Figure 1 for an overview. Assuming that the musically interesting points of novelty are consistent across the different versions, we expect the fusion curve to be more stable and musically meaningful than the individual curves. Applying our cross-version novelty detection approach for locating segment boundaries in music structure analysis, we show that the fusion curves yield, on average, better results than the version-dependent novelty curves of individual recordings. This effect becomes more prominent, when there is a high performance variance across the recordings, which is typically the case for the aspect of tempo.

Cross-version strategies have previously been applied for other music analysis tasks. For example, multiple performances are used in [1] to support tempo tracking, in [10] to stablize chord labeling, and in [8] to detect critical passages in a piece of music that are prone to beat tracking errors.

The remainder of this paper is organized as follows. In Section 2, we describe the various steps of our cross-version novelty detection procedure. Then, in Section 3,

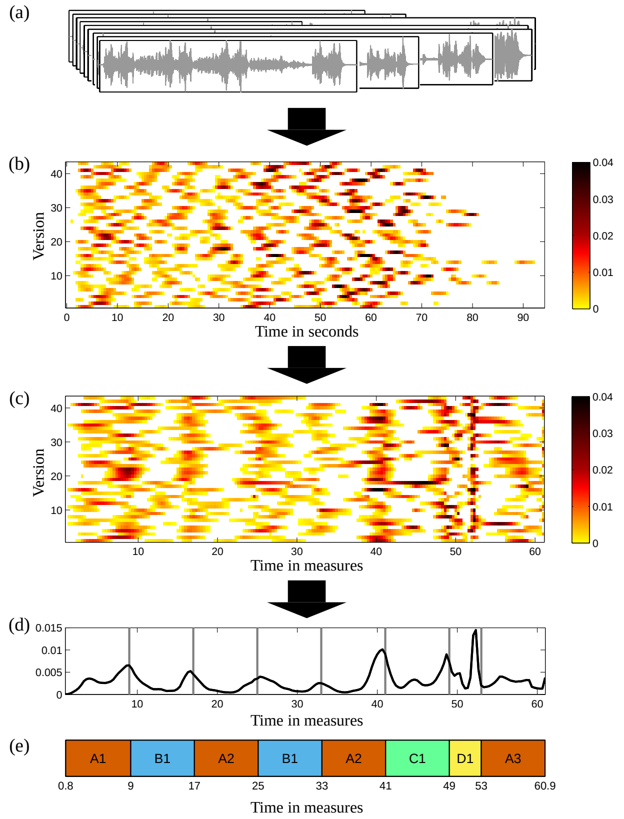


**Figure 1:** Overview of the cross-version novelty detection pipeline for Chopin's Mazurka Op. 7 No. 4. **(a)** Waveforms of several performances. **(b)** Individual novelty curves (color-coded) for 43 performances. Each row of the matrix corresponds to one novelty curve. **(c)** Individual novelty curves warped to a common musical time axis (in measures). **(d)** Fusion novelty curve. **(e)** Annotated structure and segment boundaries.

we give a detailed quantitative evaluation of our procedure within a structure analysis scenario for Chopin's Piano Mazurkas. Furthermore, we critically assess the results by a musically informed discussion of concrete examples. Finally, we conclude with Section 4 indicating future work.

## 2. CROSS-VERSION NOVELTY DETECTION

In this section, we describe the pipeline for our cross-version approach to novelty detection. For the purpose of illustration, we concentrate on the musical aspect of tempo using a cyclic tempogram feature representation (Section 2.1). As for the novelty detection, we follow a standard procedure based on 2D corner detection in self distances matrices (Section 2.2). Applying music synchronization techniques, we show how to warp the novelty curves onto a version-independent musical time axis (Section 2.3). Finally, we describe how to merge the novelty curves based on a late-fusion strategy (Section 2.4). This pipeline is also illustrated by Figure 1.

### 2.1 Cyclic Tempogram Features

In a first step, the given audio recording is transformed into a suitable feature representation that captures the musical aspects of interest. As an example, we consider the case of tempo-based novelty detection, even though our cross-version approach is applicable to any kind of feature representation. In the following, we revert to cyclic tempogram features as introduced in [7]. These features constitute a robust mid-level representation encoding local tempo information. In a first step, we capture changes in the signal's energy and spectrum [2] and then apply windowed autocorrelation methods [4]. Afterwards, the lag-axis is converted into a tempo axis specified in beats per minute (BPM), yielding a tempogram as shown in Figure 2c. Forming tempo equivalence classes by binning tempi that differ by a power of two and quantizing the values of the resulting cyclic tempogram yields an even more robust feature representation, see Figure 2d. In our experiments we use a feature resolution of 5 Hz (five feature vectors per second) and a feature dimension of 10 (ten feature values per vector). A free MATLAB implementation of these features is part of the tempogram toolbox.[1] For further details we refer to [7].

### 2.2 Novelty Curve

Let $X = (x_1, \ldots, x_N)$ denote the resulting feature sequence. To compute a novelty curve from this sequence, we employ a standard approach introduced by Foote [6]. To this end, an $N \times N$ self distance matrix $\mathbf{D}(n,m) := \mathbf{d}(x_n, x_m)$ is computed using the local distance function

$$\mathbf{d}(x_n, x_m) = 1 - \exp\left(\frac{\langle x_n, x_m \rangle}{\| x_n \| \| x_m \|} - 1\right),$$

for $1 \leq n, m \leq N$. Then, $\mathbf{D}$ is analyzed by correlating a kernel along its main diagonal. The kernel consists of an $M \times M$ matrix (with $M < N$) which has a $2 \times 2$ checkerboard-like structure weighted by a Gaussian radial function. This yields a *novelty curve*, the peaks of which indicate changes in the musical aspect represented by the feature type (in our case, tempo changes), see Figure 2e. We further process the novelty curve by subtracting a local average, see Figure 2f. In our experiments, a value $M$ corresponding to 7 seconds has turned out to be suitable, see Section 3.2 and Figure 3 for a further discussion of the parameter $M$.

### 2.3 Time Axis Conversion

The computed novelty curve depends on the performance characteristics of the underlying music recording. To make novelty curves comparable across different recordings of the same piece of music, we convert the version-dependent physical time axis (in seconds) to a version-independent musical time axis (in measures). To this end, we assume that we are given a score-like MIDI version of the piece with explicit beat and measure positions. Then, for a given music recording, we apply music synchronization techniques to automatically align the MIDI version with the audio version.[2] The alignment result allows for transferring the beat and measure positions specified by the MIDI

[1] `www.mpi-inf.mpg.de/resources/MIR/tempogramtoolbox`
[2] In our implementation, we revert to the high-resolution music synchronization approach described in [5].

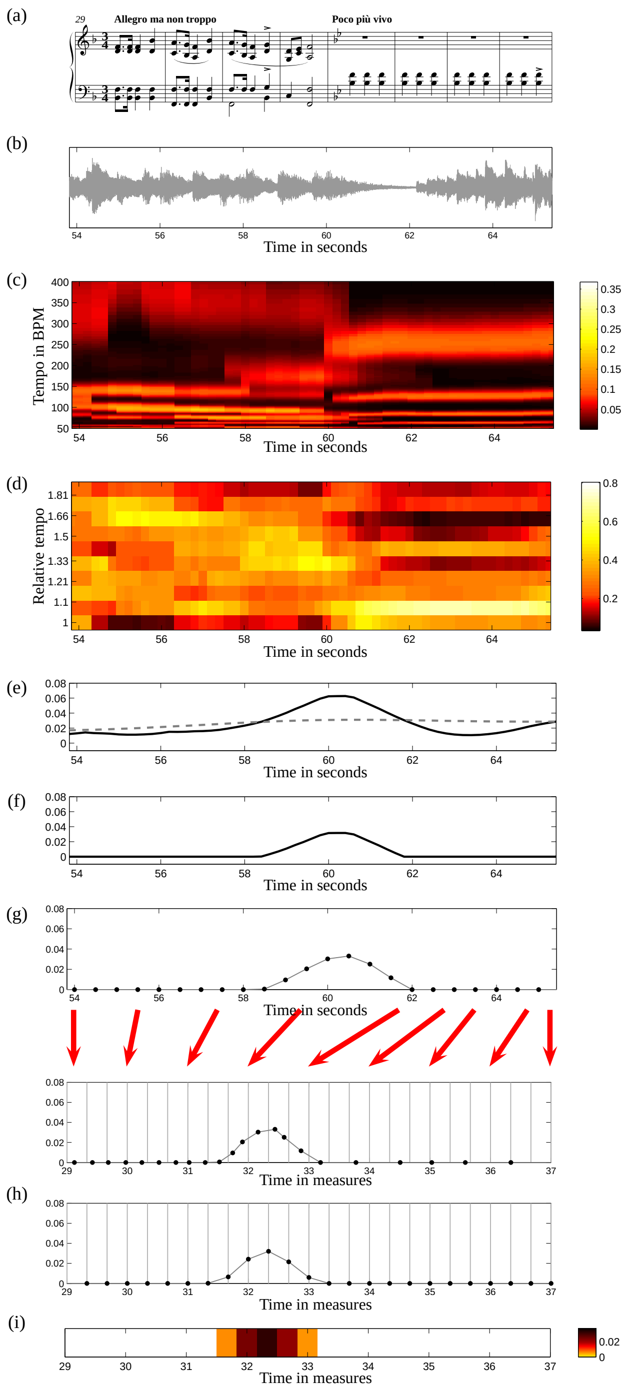


**Figure 2:** Novelty detection for a recording of Chopin's Mazurka Op. 68 No. 3. **(a)** Measures 29-36. **(b)** Waveform. **(c)** Tempogram. **(d)** Quantized cyclic tempogram. **(e)** Novelty curve (solid line) and its local average curve (dashed line). **(f)** Post-processed novelty curve. **(g)** Time axis conversion. **(h)** Resampled novelty curve. **(i)** Color-coded representation of (h).

version to the corresponding time positions in the audio version. Based on this information, we locally stretch and contract the time axis of the novelty curve computed from the recording to obtain a musical time axis, see Figure 2g. Finally, we interpolate and resample the novelty curve to obtain one value for each beat position of the piece of music, see Figure 2h and Figure 2i.

### 2.4 Fusion Novelty Curve

Being based on the same musical time axis, one can now directly compare novelty curves from different performances of the same piece of music. As an example, Figure 1b shows the original novelty curves (in some color-coded form) for 43 different performances of Chopin's Mazurka Op. 7 No. 4. No correlations across the different performances are visible. After the time axis conversion, as shown in Figure 1c, strong correlations between the different novelty curves become evident. For example, there is a tempo change at measure 52 for basically all performances.

To fuse the information across all novelty curves, we basically compute the average of the novelty curves. To become more robust to outliers, we first remove the $20\,\%$ smallest and largest novelty values for each beat position among all performances, and then compute the *fusion novelty curve* by taking the beat-wise arithmetic mean of the remaining values. The crucial observation is that a fusion novelty curve reveals a local maximum (peak) at those positions where a large number of individual novelty curves also possess a local maximum. In other words, the fusion novelty curve expresses the consistencies in the peak structures across the various recordings, see also Figure 1d.

## 3. EXPERIMENTS

Even though there are often significant differences in the way musicians interpret a piece of music, tempo changes are not arbitrary and there are musical reasons for a speed up or slow down. Our hypothesis is that the tempo changes that can be observed across a large number of different performances are of particular musical importance. Therefore, we conjecture that peaks of the fusion novelty curve are more relevant than the peaks of the individual novelty curves. To investigate our hypothesis, we have conducted various experiments on a dataset consisting of Chopin's Mazurkas (Section 3.1). Our quantitative evaluation in the context of music structure analysis (Section 3.2) as well as a discussion of various representative examples (Section 3.3) demonstrate that cross-version fusion curves yield, on average, better results than the novelty curves obtained from individual recordings.

### 3.1 Dataset and Annotations

We conduct our experiments on a Mazurka dataset, which consists of 2792 recorded performances for the 49 Mazurkas by Frédéric Chopin. These recordings were collected in the Mazurka Project[3] and have been previously used, e. g., for the purpose of performance analysis [13]. For each of the 49 Mazurkas, there are on average 57 different recordings (ranging from the early stages of music recording until today), as well as a MIDI file that represents the piece in an uninterpreted symbolic form. In particular, measure and beat positions are known in the MIDI file.

The Chopin Mazurkas are short piano compositions with a 3/4 time signature. These pieces have a relatively

[3] mazurka.org.uk

clear musical structure, where certain parts are repeated more or less in the same way. We have manually annotated each score-like MIDI file according to its musical structure. On average this leads to $9.4$ segment boundaries per Mazurka (disregarding segment boundaries at the beginning and end of the piece) and an average duration of $11.9$ measures per musical part, see also Table 1 for more details.

### 3.2 Quantitative Evaluation

As is the case for romantic piano music, most Mazurka performances reveal numerous local tempo changes which often indicate transitions of musical importance. Many of these transitions occur near segment boundaries between musical parts, where one can often observe tempo changes. Even though not all segment boundaries are characterized this way, we use them for a first quantitative evaluation to indicate the behavior of our cross-version fusion novelty curves. Let $\mathcal{B}$ denote the set of segment boundaries (specified in musical beats) for a given Mazurka.

For a novelty curve (with time axis given in musical beats), we perform some peak picking to determine a set $\mathcal{P}$ of relevant peak positions. Here a position is considered relevant if the novelty curve assumes at this position a global maximum over a window of length $\lambda$ centered at the corresponding position. In our experiments, the value $\lambda = 19$ beats has turned out to be meaningful, see also Figure 3. A peak position in $\mathcal{P}$ is considered to be *true* if there is a segment boundary in $\mathcal{B}$ in a $\delta$-neighborhood, otherwise it is considered to be *false*. This allows to define a precision (P), recall (R), and F-measure (F) for the set $\mathcal{P}$ relative to $\mathcal{B}$ . In our experiments, we choose $\delta = 3$ beats corresponding to a musical measure. In our evaluation, we further ignored all boundaries and all peaks in the first four and last four measures of a piece of music. The main reason for excluding these measures is that many of the recordings start and end with non-musical content such as silence or applause, which leads to spurious peaks at the positions where the music starts or ends. Also, synchronization errors typically occur in these regions.

Before we investigate the role of the various parameters, we first look at the results for a fixed parameter setting as indicated by Table 1. To better understand the effect of the cross-version approach, we computed P/R/F-measures in two different ways. First, for a given Mazurka, we computed individual P/R/F-measures for each performance using the version-dependent novelty curves and then averaged over all performances to obtain averaged individual P/R/F-measures. Secondly, we computed these measures from the fusion novelty curve to obtain cross-version P/R/F-measures. Table 1 shows the resulting averaged individual as well as cross-version P/R/F-measures for all of the $49$ Mazurkas. Furthermore, the last row of the table indicates the overall values averaged over all Mazurkas. As the main result, one can see that the overall $F$-measure obtained from individual novelty curves is $F = 0.39$, whereas the overall $F$-measure obtained from the fusion novelty curves is $F = 0.52$. In other words, the tempo-based novelty detection can indeed be improved when simultaneously analyzing a set of different performances.

| | | | | Ind.-Version | | | Cross-Version | | |
|---|---|---|---|---|---|---|---|---|---|
| Piece | #P | #M | #B | P | R | F | P | R | F |
| M06-1 | 49 | 112 | 7 | 0.29 | 0.60 | 0.39 | 0.50 | 1.00 | 0.67 |
| M06-2 | 51 | 96 | 9 | 0.42 | 0.63 | 0.50 | 0.47 | 0.89 | 0.62 |
| M06-3 | 47 | 98 | 12 | 0.23 | 0.31 | 0.26 | 0.13 | 0.18 | 0.15 |
| M06-4 | 46 | 40 | 9 | 0.65 | 0.40 | 0.49 | 0.83 | 0.63 | 0.71 |
| M07-1 | 55 | 104 | 9 | 0.30 | 0.52 | 0.38 | 0.44 | 0.89 | 0.59 |
| M07-2 | 51 | 120 | 14 | 0.40 | 0.45 | 0.42 | 0.36 | 0.36 | 0.36 |
| M07-3 | 65 | 105 | 11 | 0.31 | 0.44 | 0.36 | 0.47 | 0.64 | 0.54 |
| M07-4 | 43 | 60 | 7 | 0.51 | 0.60 | 0.55 | 0.75 | 0.86 | 0.80 |
| M07-5 | 46 | 20 | 12 | 0.61 | 0.35 | 0.44 | 1.00 | 0.60 | 0.75 |
| M17-1 | 52 | 100 | 10 | 0.25 | 0.35 | 0.29 | 0.07 | 0.10 | 0.08 |
| M17-2 | 55 | 68 | 3 | 0.21 | 0.65 | 0.32 | 0.25 | 1.00 | 0.40 |
| M17-3 | 51 | 168 | 10 | 0.26 | 0.64 | 0.37 | 0.42 | 1.00 | 0.59 |
| M17-4 | 93 | 132 | 10 | 0.23 | 0.49 | 0.31 | 0.35 | 0.67 | 0.46 |
| M24-1 | 61 | 96 | 10 | 0.30 | 0.40 | 0.34 | 0.46 | 0.60 | 0.52 |
| M24-2 | 66 | 120 | 16 | 0.38 | 0.48 | 0.42 | 0.50 | 0.64 | 0.56 |
| M24-3 | 55 | 79 | 6 | 0.26 | 0.46 | 0.33 | 0.45 | 0.83 | 0.59 |
| M24-4 | 76 | 186 | 20 | 0.42 | 0.59 | 0.49 | 0.65 | 0.85 | 0.74 |
| M30-1 | 50 | 53 | 4 | 0.34 | 0.57 | 0.42 | 0.50 | 0.75 | 0.60 |
| M30-2 | 60 | 64 | 7 | 0.48 | 0.60 | 0.53 | 0.78 | 1.00 | 0.88 |
| M30-3 | 63 | 111 | 10 | 0.30 | 0.45 | 0.36 | 0.50 | 0.60 | 0.55 |
| M30-4 | 65 | 139 | 14 | 0.34 | 0.51 | 0.40 | 0.47 | 0.62 | 0.53 |
| M33-1 | 55 | 48 | 4 | 0.43 | 0.62 | 0.51 | 0.50 | 0.75 | 0.60 |
| M33-2 | 70 | 143 | 16 | 0.34 | 0.47 | 0.39 | 0.33 | 0.44 | 0.38 |
| M33-3 | 50 | 48 | 3 | 0.28 | 0.61 | 0.38 | 0.33 | 1.00 | 0.50 |
| M33-4 | 74 | 224 | 19 | 0.24 | 0.40 | 0.30 | 0.27 | 0.37 | 0.31 |
| M41-1 | 56 | 139 | 14 | 0.26 | 0.39 | 0.32 | 0.19 | 0.29 | 0.23 |
| M41-2 | 63 | 68 | 7 | 0.43 | 0.57 | 0.49 | 0.75 | 0.86 | 0.80 |
| M41-3 | 40 | 78 | 13 | 0.44 | 0.45 | 0.44 | 0.57 | 0.73 | 0.64 |
| M41-4 | 45 | 74 | 9 | 0.42 | 0.58 | 0.48 | 0.62 | 0.89 | 0.73 |
| M50-1 | 49 | 104 | 6 | 0.18 | 0.48 | 0.26 | 0.20 | 0.50 | 0.29 |
| M50-2 | 58 | 127 | 10 | 0.31 | 0.56 | 0.40 | 0.56 | 0.90 | 0.69 |
| M50-3 | 74 | 208 | 10 | 0.18 | 0.57 | 0.27 | 0.21 | 0.70 | 0.33 |
| M56-1 | 42 | 204 | 13 | 0.18 | 0.41 | 0.25 | 0.11 | 0.23 | 0.15 |
| M56-2 | 53 | 92 | 8 | 0.31 | 0.60 | 0.41 | 0.54 | 1.00 | 0.70 |
| M56-3 | 57 | 220 | 15 | 0.23 | 0.51 | 0.32 | 0.30 | 0.67 | 0.42 |
| M59-1 | 63 | 142 | 9 | 0.19 | 0.42 | 0.26 | 0.17 | 0.44 | 0.25 |
| M59-2 | 63 | 111 | 4 | 0.10 | 0.41 | 0.16 | 0.20 | 0.75 | 0.32 |
| M59-3 | 66 | 154 | 11 | 0.20 | 0.44 | 0.28 | 0.30 | 0.55 | 0.39 |
| M63-1 | 46 | 102 | 9 | 0.26 | 0.46 | 0.33 | 0.27 | 0.44 | 0.33 |
| M63-2 | 65 | 56 | 4 | 0.33 | 0.61 | 0.43 | 0.38 | 0.75 | 0.50 |
| M63-3 | 88 | 76 | 8 | 0.41 | 0.55 | 0.47 | 0.78 | 0.88 | 0.82 |
| M67-1 | 44 | 60 | 7 | 0.23 | 0.31 | 0.26 | 0.14 | 0.17 | 0.15 |
| M67-2 | 41 | 72 | 6 | 0.33 | 0.54 | 0.41 | 0.45 | 0.83 | 0.59 |
| M67-3 | 46 | 56 | 3 | 0.30 | 0.75 | 0.43 | 0.43 | 1.00 | 0.60 |
| M67-4 | 59 | 112 | 6 | 0.26 | 0.71 | 0.38 | 0.40 | 1.00 | 0.57 |
| M68-1 | 46 | 84 | 11 | 0.53 | 0.63 | 0.57 | 0.50 | 0.50 | 0.50 |
| M68-2 | 65 | 84 | 11 | 0.49 | 0.54 | 0.51 | 0.77 | 0.91 | 0.83 |
| M68-3 | 51 | 60 | 9 | 0.61 | 0.55 | 0.57 | 0.78 | 0.78 | 0.78 |
| M68-4 | 63 | 63 | 7 | 0.35 | 0.37 | 0.36 | 0.47 | 0.58 | 0.52 |
| ∅ | 57.0 | 103.7 | 9.4 | 0.33 | 0.51 | 0.39 | 0.45 | 0.69 | 0.52 |

**Table 1:** Overview of the Mazurka dataset and precision (P), recall (R), and F-measures (F) for two different settings. The first four columns specify the Mazurka (e.g. `M06-1` refers to Mazurka Op. 6 No. 1), the number of performances (#P), the number of measures (#M), and the number of annotated segment boundaries (#B). The next three columns show the average individual P/R/F-measures obtained from individual performances and the last three columns show cross-version P/R/F-measures obtained from the fusion novelty curves. The used parameters are: $M \sim 7$ seconds, $\lambda = 19$ beats, $\delta = 3$ beats.

In the next experiments, we investigate the role of the kernel size parameter $M$ (see Section 2.2) and the neighborhood parameter $\lambda$ used in the peak picking. Figure 3 shows the cross-version P/R/F-measures averaged over all $49$ Mazurkas for various combinations of $M$ and $\lambda$. Generally, when increasing $\lambda$, the precision increases (Figure 3a) and the recall decreases (Figure 3b). This is not surprising, since an increase in $\lambda$ imposes stricter conditions on the peak picking (and the set of relevant peaks becomes smaller). The remaining peaks tend to be true (increase in precision), while fewer segment boundaries in $\mathcal{B}$ are detected (decrease in recall). The kernel size parameter $M$ has a minor influence on the final results. Only for large values of $\lambda$, smaller kernel sizes tend to be favorable. As

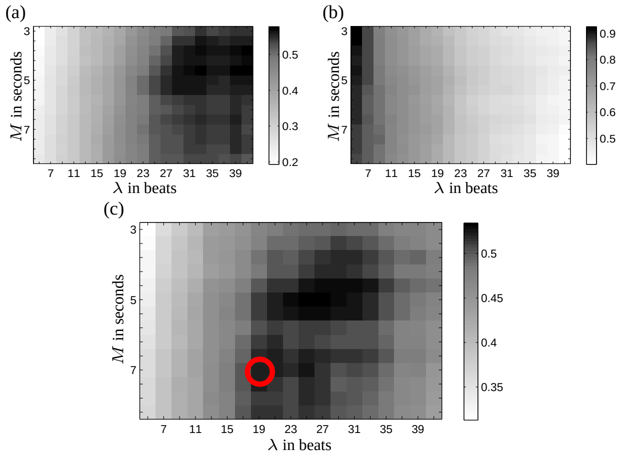


**Figure 3:** Average cross-version P/R/F-measures for different parameter settings. **(a)** Average precision values. **(b)** Average recall values. **(c)** Average F-measure values. The red circle indicates the parameter setting used in Table 1.

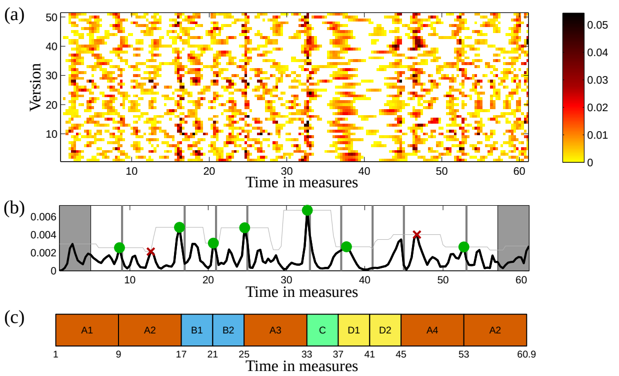


**Figure 4:** Novelty curves for `M68-3`. **(a)** Individual novelty curves (color-coded, musical time axis) for 51 performances. **(b)** Fusion novelty curve. True peaks are indicated by green discs and false peaks by red crosses. The gray areas at the beginning and end are left out in the evaluation. The thin gray curve indicates the peak picking condition introduced by the neighborhood parameter $\lambda$. **(c)** Annotated structure and segment boundaries.

for our main experiments, we favored comparatively larger kernel sizes (resulting in smoother novelty curves) and a smaller $\lambda$ (being less restrictive in the peak picking) choosing $M \sim 7$ seconds and $\lambda = 19$ beats. However, as also indicated by Figure 3c, the specific paramter setting is not of crucial importance and slightly changing the settings yields similar experimental results.

### 3.3 Qualitative Evaluation

For some Mazurkas this improvement is significant. For example, for the Mazurka Op. 7 No. 4 shown in Figure 1, the F-measure increases from $F = 0.55$ (individual) to $F = 0.80$ (cross-version). Also for the Mazurka Op. 68 No. 3 (Figure 4) the cross-version fusion approach stabilizes the tempo-based novelty detection improving the F-measure from $F = 0.57$ (individual) to $F = 0.78$ (cross-version).

However, there is also a number of Mazurkas where one has rather low P/R/F-measures—for the individual curves as well as for the fusion novelty curves. For example, for the Mazurka Op. 56 No. 1 shown in Figure 5, the F-

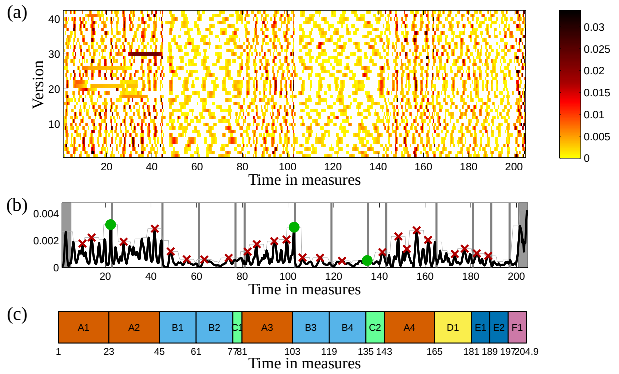


**Figure 5:** Novelty curves for `M56-1` as in Figure 4. **(a)** Individual novelty curves for 42 performances. **(b)** Fusion novelty curve. **(c)** Annotated structure and segment boundaries.

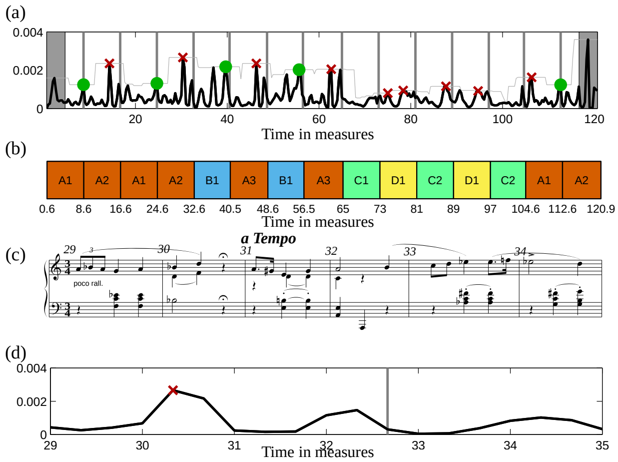


**Figure 6:** Detailed example on `M07-2`. **(a)** Fusion novelty curve. **(b)** Annotated structure and segment boundaries. **(c)** Score excerpt of measures 29-34. **(d)** Fusion novelty curve excerpt of measures 29-34.

measure even decreases from $F = 0.25$ (individual) to $F = 0.15$ (cross-version). In this piece, the annotated segments are rather long in comparison to the other Mazurkas. Listening to the performances reveals that each $A$-part consists of several phrases, which are shaped by most pianists using a characteristic tempo progression with a slow down and speed up at phrase boundaries. These tempo changes lead to a large number of consistent peaks, which are not reflected by our structure annotations (even though the peaks are musically meaningful) and sometimes also not captured by our peak picking ($\lambda$ being too restrictive). Also, in the other parts there are a number of false positive peaks of less musical significance. As this Mazurka shows, annotated segment boundaries do not need to go along with tempo changes and, vice versa, musically meaningful tempo changes may also occur within musical parts. Therefore, our quantitative evaluation within the structure analysis context, even though indicating meaningful general tendencies, is an oversimplification.

We now discuss some further typical examples where the fusion novelty curve reveals musically relevant tempo changes that do not concur with segment boundaries. Let us look at the fusion novelty curve for Mazurka Op. 7 No. 2 as shown in Figure 6a. Here one can notice strong peaks in

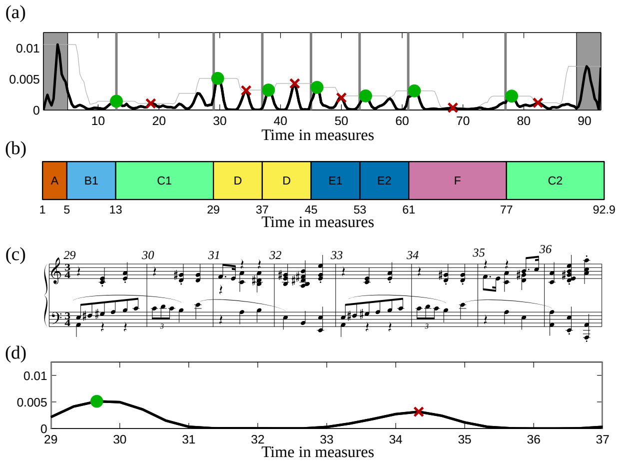


**Figure 7:** Detailed example on `M56-2`. **(a)** Fusion novelty curve. **(b)** Annotated structure and segment boundaries. **(c)** Score excerpt of measures 29-36. **(d)** Fusion novelty curve excerpt of measures 29-36.

the $A_2$-parts and $A_3$-parts located roughly two measures before segment boundaries, so that these peaks are considered false positives in our evaluation. Looking at the score of the piece reveals that there is actually a tempo instruction *a Tempo* just two measures before the respective segments boundaries, see Figure 6c. Most pianists realize this instruction by speeding up their performances, which leads to the musically relevant peaks captured by our cross-version novelty curve. As another example, let us look at the fusion novelty curve of Mazurka Op. 56 No. 2, see Figure 7. Here, two of the false peak positions in the fusion novelty curve are located in the middle of the two $D$-parts, see Figure 7a. A manual investigation showed that each of the eight-measure $D$-parts consists of two repeating four-measure phrases. This substructure is not reflected by our structure annotations. The pianists, however, shape the phrases by a pronounced tempo change. Furthermore, in the middle of the $C$-parts and the $F$-part, Figure 7 also shows some false positive peaks of no musical relevance. Here, an improved peak picking may remedy this problem.

## 4. CONCLUSIONS

In this paper, we introduced a cross-version approach for novelty detection capturing consistencies across different performances of a piece. Applying this concept to tempo-related audio features, we showed that the resulting fusion novelty curves perform better in revealing musically meaningful points of novelty than the individual curves. In the future, we plan to conduct similar experiments using different audio features that reflect not only tempo, but also harmony, timbre, and dynamics. Also, the described cross-version approach is generic in the sense that it can also be applied to other music analysis tasks beyond novelty detection. A stabilization effect has also been reported for chord labeling and beat tracking, and we plan to apply this concept to general structure analysis. Finally, we discussed that our evaluation of the novelty detection results based on segment boundaries indicates interesting general tendencies, but also constitutes an oversimplification. Here, future work must address the evaluation problem by including more musicological knowledge, e. g. by looking at expected tempo changes in the score, annotated by musically trained experts. On the other hand, our cross-version approach might not only be used for the task of audio segmentation, but may also aid as a performance analysis tool for musicologists.

**Acknowledgments:** This work has been supported by the Cluster of Excellence on Multimodal Computing and Interaction at Saarland University and the German Research Foundation (DFG MU 2686/5-1).

## 5. REFERENCES

[1] Andreas Arzt, Johannes Kepler, and Gerhard Widmer. Simple tempo models for real-time music tracking. In *Proceedings of the Sound and Music Computing Conference (SMC)*, 2010.

[2] Juan Pablo Bello, Laurent Daudet, Samer Abdallah, Chris Duxbury, Mike Davies, and Mark B. Sandler. A tutorial on onset detection in music signals. *IEEE Transactions on Speech and Audio Processing*, 13(5):1035–1047, 2005.

[3] Roger B. Dannenberg and Masataka Goto. Music structure analysis from acoustic signals. In David Havelock, Sonoko Kuwano, and Michael Vorländer, editors, *Handbook of Signal Processing in Acoustics*, volume 1, pages 305–331. Springer, New York, NY, USA, 2008.

[4] Daniel P.W. Ellis. Beat tracking by dynamic programming. *Journal of New Music Research*, 36(1):51–60, 2007.

[5] Sebastian Ewert, Meinard Müller, and Peter Grosche. High resolution audio synchronization using chroma onset features. In *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, pages 1869–1872, Taipei, Taiwan, 2009.

[6] Jonathan Foote. Automatic audio segmentation using a measure of audio novelty. In *Proceedings of the IEEE International Conference on Multimedia and Expo (ICME)*, pages 452–455, New York, NY, USA, 2000.

[7] Peter Grosche, Meinard Müller, and Frank Kurth. Cyclic tempogram – a mid-level tempo representation for music signals. In *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, pages 5522 – 5525, Dallas, Texas, USA, 2010.

[8] Peter Grosche, Meinard Müller, and Craig Stuart Sapp. What makes beat tracking difficult? A case study on Chopin Mazurkas. In *Proceedings of the 11th International Conference on Music Information Retrieval (ISMIR)*, pages 649–654, Utrecht, The Netherlands, 2010.

[9] Kristoffer Jensen. Multiple scale music segmentation using rhythm, timbre, and harmony. *EURASIP Journal on Advances in Signal Processing*, 2007(1):11 pages, 2007.

[10] Verena Konz, Meinard Müller, and Sebastian Ewert. A multi-perspective evaluation framework for chord recognition. In *Proceedings of the International Society for Music Information Retrieval Conference (ISMIR)*, pages 9–14, Utrecht, The Netherlands, 2010.

[11] Jouni Paulus and Anssi P. Klapuri. Music structure analysis using a probabilistic fitness measure and a greedy search algorithm. *IEEE Transactions on Audio, Speech, and Language Processing*, 17(6):1159–1170, 2009.

[12] Jouni Paulus, Meinard Müller, and Anssi P. Klapuri. Audio-based music structure analysis. In *Proceedings of the 11th International Conference on Music Information Retrieval (ISMIR)*, pages 625–636, Utrecht, The Netherlands, 2010.

[13] Craig Stuart Sapp. Comparative analysis of multiple musical performances. In *Proceedings of the International Conference on Music Information Retrieval (ISMIR)*, pages 497–500, Vienna, Austria, 2007.

[14] Douglas Turnbull, Gert Lanckriet, Elias Pampalk, and Masataka Goto. A supervised approach for detecting boundaries in music using difference features and boosting. In *Proceedings of the International Conference on Music Information Retrieval (ISMIR)*, pages 51–54, Vienna, Austria, 2007.

[15] George Tzanetakis and Perry Cook. Multifeature audio segmentation for browsing and annotation. In *Proceedings of the IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, pages 103–106, New Platz, NY, USA, 1999.