A FORMALIZATION OF RELATIVE LOCAL TEMPO VARIATIONS IN COLLECTIONS OF PERFORMANCES

Jeroen Peperkamp **Klaus Hildebrandt** **Cynthia C. S. Liem**
Delft University of Technology, Delft, The Netherlands
jbpeperkamp@gmail.com {k.a.hildebrandt, c.c.s.liem}@tudelft.nl

ABSTRACT

Multiple performances of the same piece share similarities, but also show relevant dissimilarities. With regard to the latter, analyzing and quantifying variations in collections of performances is useful to understand how a musical piece is typically performed, how naturally sounding new interpretations could be rendered, or what is peculiar about a particular performance. However, as there is no formal ground truth as to what these variations should look like, it is a challenge to provide and validate analysis methods for this. In this paper, we focus on relative local tempo variations in collections of performances. We propose a way to formally represent relative local tempo variations, as encoded in warping paths of aligned performances, in a vector space. This enables using statistics for analyzing tempo variations in collections of performances. We elaborate the computation and interpretation of the mean variation and the principal modes of variation. To validate our analysis method despite the absence of a ground truth, we present results on artificially generated data, representing several categories of local tempo variations. Finally, we show how our method can be used for analyzing to real-world data and discuss potential applications.

1. INTRODUCTION

When performing music that is written down in a score, musicians produce sound that subtly differs from what is written. For example, to create emphasis, they can vary the time between notes, the dynamics, or other instrument-specific parameters, such as which strings to use on a violin or how to apply the pedals on a piano. In this paper, we focus on variations in timing, contributing a method to detect local tempo variations in a collection of performances.

Solving this problem is made difficult by the fact that it is not clear what we are trying to find: there is generally no ground truth that tells us what salient variations there are for a given piece. Furthermore, it is difficult to discern whether a given performance is 'common' or 'uncommon'.

To overcome this, we propose an approach for statistical analysis of relative local tempo variations among performances in a collection. To this end, we elaborate the computation of the mean variation and the principal modes of variation. The basis of the approach is the insight that after normalization, the set of possible tempo variations, represented by temporal warping paths, forms a convex subset of a vector space. We test our approach on artificially generated data (with controllable variations in a collection), and on recorded real performances. We discuss two applications: analysis of tempo variations and example-guided synthesis of performances.

2. RELATED WORK

2.1 Performance Analysis

Most closely related to the present work are the works in [9, 11] and [21, 22], focusing on statistical comparison of performances, targeting local tempo variations without ground truth. [9, 11] focus especially on temporal warping paths with respect to a reference performance. Furthermore, [10] analyzes main modes of variation in comparative analysis of orchestral recordings. We differ from these works in offering a more formalized perspective on variation, a more thorough and controlled validation procedure on artificially generated data, and ways to perform analyses with respect to a full collection of performances, beyond a single reference performance.

Further work in comparative performance analysis considered features such as dynamics [6]: here, it was shown that dynamic indications in a score do not lead to absolute realizations of loudness levels. [8] and [1] provide comparative analyses on many expressive features, although the latter work also finds that musicians find it difficult to think about the aspects of their performance in the quantitative fashion that is common in the MIR literature.

The absence of a clear-cut ground truth also poses challenges when automatically creating a natural-sounding rendition of a piece of music, as noted in [3] as well as [26]. Indeed, the system in the latter work explicitly relies "on a 'correct' or 'appropriate' phrase structure analysis", suggesting it is not trivial to get such an analysis.

Quite some work has also gone into the task of structure analysis, e.g. [12, 14–16, 18, 19, 23]. It turns out, however, that for some genres, the structure may be perceived ambiguously, as observed with professional annotators [23], performers [17] and listeners [24].

Attribution: Jeroen Peperkamp, Klaus Hildebrandt, Cynthia C. S. Liem. "A formalization of relative local tempo variations in collections of performances", 18th International Society for Music Information Retrieval Conference, Suzhou, China, 2017.

2.2 Dynamic Time Warping

For obtaining temporal warping paths between performances, we use Dynamic Time Warping (DTW). In a nutshell, DTW matches points from one time series to points from another time series such that the cumulative distance between the matched points is as small as possible, for some suitable distance function; the matching can then be interpreted as a warping path. A thorough overview of DTW is given in [13].

3. FORMAL ANALYSIS FRAMEWORK

We start with a formalization of tempo variations and then describe the proposed statistical analysis. The tempo variations we consider can be described by warping paths, which can be obtained from recordings of performances by using DTW.

3.1 Formal Properties

We wish to compare tempo variations between different performances of a piece. In this section, we consider an idealized setting in which only the local tempo is varied. In the next section, we will discuss how this can be used for analyzing variations in actual performances.

For our formal framework, we first need a representation of a performance. We will call the reference performance $g : [0, l_g] \to \mathbb{R}^d$, with l_g the length of the performance and d the dimensionality of some suitable feature space in which the performance can be represented. Other performances in a collection, displaying tempo variations with respect to the reference performance, can be defined as follows:

Definition 1. *A performance of g with varied tempo is a function $f = g \circ \psi : [0, l_f] \to \mathbb{R}^d$, with l_f and d defined as above, and $\psi : [0, l_f] \to [0, l_g]$ a function with nonnegative derivative, i.e., $\dot{\psi} \geq 0$. We call ψ a tempo variation.*

For the analysis of tempo variations between f and g, we distinguish between average and relative tempo variation. The average tempo variation can be observed by looking at the length of the interval over which the functions are parametrized; it is simply the difference in average overall tempo of each performance. Clearly, the longer the interval, the slower the performance is on average. There is more structure in the details, of course, which is what the relative variations attempt to capture. Specifically, this refers to an analysis of tempo variations given that the performances are parametrized over an interval of the same length, for instance, the unit interval.

Now, to implement the concept of relative tempo variations, we first reparametrize the performances over the unit interval. Given $f : [0, l_f] \to \mathbb{R}^d$, we consider the normalized performance $f^* = f \circ \rho : [0, 1] \to \mathbb{R}^d$, where $\rho : [0, 1] \to [0, l_f]$ is given by $\rho(t) = l_f t$. Now we can go into more detail about these relative tempo variations.

3.1.1 Structure of the Set of Relative Tempo Variations

Relative tempo variations can be described by reparametrizations that relate the performances in question. Due to the normalization of the performances, the reparametrizations map the unit interval to itself. The relative tempo variations φ and their derivatives $\dot{\varphi}$ are characterized by the following two properties:

Property 1. $\varphi(0) = 0$, $\varphi(1) = 1$.

Property 2. $\dot{\varphi}(n) \geq 0$ *for any* $n \in [0, 1]$.

Examples of such relative tempo variations are shown in Figure 1 (left), along with insets to see what happens when one zooms in. When working with the normalized performances, every performance with varied tempo f^* of a reference performance g^* has the form $f^* = g^* \circ \varphi$.

The benefit of splitting average and relative variation is that the set of relative variations has a geometric structure: the following lemma shows that it is a convex set in an vector space. This enables us to use classical methods from statistical analysis to analyze the relative tempo variations, as explained in Section 3.2.

Lemma 1. *Convex combinations of relative tempo variations are relative tempo variations.*

Proof. Let $\alpha = (\alpha_1, \ldots, \alpha_m)$ be a vector of nonnegative numbers, $\alpha_i \geq 0$, with unit ℓ_1 norm, $\sum_{i=1}^m \alpha_i = 1$, and let $\varphi_i : [0, 1] \mapsto [0, 1]$ be relative tempo variations ($1 \leq i \leq m$). We show that $\varphi = \sum_{i=1}^m \alpha_i \varphi_i$ is a relative tempo variation. As a sum of functions on the unit interval, φ is also a function on the unit interval. Since the α_i sum to 1, $\sum_{i=1}^m \alpha_i \varphi_i(0) = 0$ and $\sum_{i=1}^m \alpha_i \varphi_i(1) = 1$, which means that Property 1 holds. Finally, since all α_i are nonnegative, $\dot{\varphi} \geq 0$ is also maintained. □

3.2 Analysis of Prominent Variations

In the following, we consider a set of performances (with varied tempo) and show how our approach allows us to compute statistics on the set. Explicitly, we take the mean and perform principal component analysis (PCA). As a first step, we reparametrize the performances over the unit interval $[0, 1]$, as described above. We distinguish two settings for our analysis. First, we describe a setting in which we consider one reference performance. An example of such a reference performance in practice is a rendered MIDI, which has a linear timing to which we relate the actual performances in the set. In the second setting, we avoid the use of a reference performance by incorporating all pairwise comparisons between performances.

3.2.1 Comparing to the Reference Performance

Comparing a set of performances $\{f_1, f_2, \ldots, f_n\}$ to a reference g^* means obtaining for each normalized performance f_i^* the corresponding relative tempo variation φ_i, such that $f_i^* = g^* \circ \varphi_i$. Lemma 1 shows that we can build a continuous set of relative tempo variations by building convex combinations. Geometrically speaking, we consider the simplex spanned by the φ_i. Though not needed

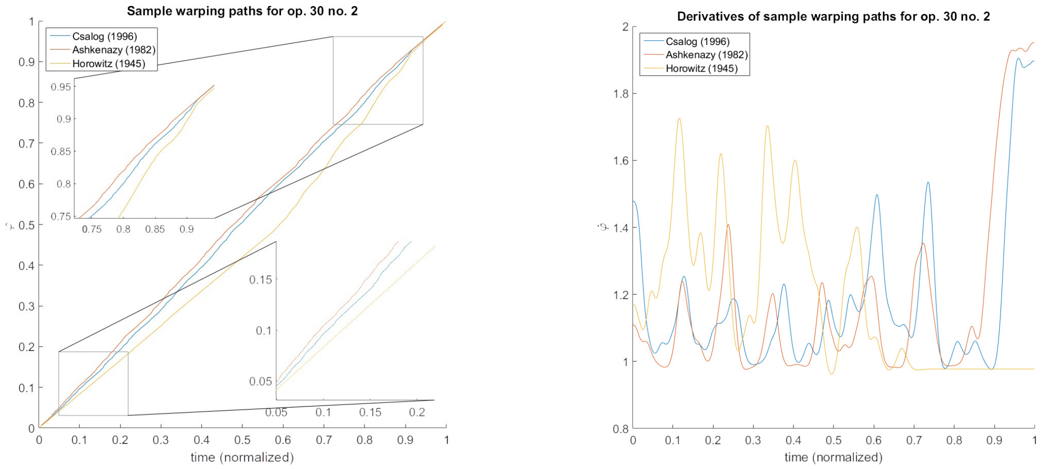

Figure 1. Several reparametrizations φ relating professional human performances of Chopin's Mazurka op. 30 no. 2 to a deadpan MIDI version. Original φ with zoomed insets (left) and their derivatives $\dot{\varphi}$ (right).

for our analysis, extrapolation out of the simplex is possible, as long as Property 2 is satisfied.

A particularly interesting convex combination for our purposes is the mean of the set of performances. The mean relative tempo variation $\bar{\varphi}$ can be computed by setting all the α_i to the same value in Lemma 1 above. The mean of the normalized performances $\{f_i^*\}$ is given as $g^* \circ \bar{\varphi}$. To obtain the mean of the performances, $g^* \circ \bar{\varphi}$ is linearly rescaled to the average length of the performances f_i. The mean $\bar{\varphi}$ gives information about which local tempo variations away from g^* are the most prevalent among the performances under analysis. Of course, the mean does not capture the variance in the set, for example, deviations in opposite directions, as when some performers speed up and others slow down, which would be evened out.

The variance in a set can be analyzed using PCA. To perform a PCA on the set φ_i, we need a scalar product on the space of relative tempo variations. Since these are functions on the unit interval, any scalar product on this function space can be used. For our experiments, we used the L^2-scalar product of the derivatives of the functions (in other words the Sobolev H_0^1-scalar product). The reason for using a scalar product of the derivatives, rather than the function values, is that the derivatives describe the variations in tempo, and the function values encode the alignment of the performance. See Figure 1 (right) for an example of how this brings out the variation. Once a scalar product is chosen, we construct the covariance matrix, whose entries are the mutual scalar products of the functions $\varphi_i - \bar{\varphi}$ (the distance of the tempo variations to the mean). The eigenvectors of the covariance matrix yield the principal modes of variation in the set φ_i. These express the main variations away from the mean in the set and the eigenvalues indicate how much variance there is in the set of performances by how much of the variance is explained by the corresponding modes. The modes express the tendency of performers to speed up or slow down observed in the set of performances.

3.2.2 Incorporating All Pairwise Comparisons

When using a reference performance, one has to choose which performance to use as g^*, or to produce an artificial performance for g^* (as we do in Section 4). This way, the comparison becomes dependent on the choice of g^*, which may not be desirable, as there may be 'outlier' performances that would not necessarily be the best choice for a reference performance (though other things can be learned from them [17]).

To avoid the need to choose g^*, we propose an alternative analysis using all pairwise comparisons. This means obtaining reparametrizations φ for every pair of performances f^* and g^* such that $f^* = g^* \circ \varphi$. This makes sense, as it is not guaranteed that for three normalized performances f^*, g^* and h^* and reparametrizations φ_i and φ_j such that $g^* = f^* \circ \varphi_i$ and $h^* = g^* \circ \varphi_j$, we would get $h^* = f^* \circ \varphi_i \circ \varphi_j$. In other words, reparametrizations may violate the triangle inequality, so we obtain more information by taking into account all possible reparametrizations.

The same techniques can be applied once we have the (extended) set of reparametrizations φ. That is, we can take the mean of all the φ or perform a PCA on them. Empirically, it turns out there tends to be repeated information in the reparametrizations, which results in a certain amount of natural smoothing when taking the mean; this effect can be seen in Figure 3.

4. EXPERIMENTAL VALIDATION

In Section 3, we considered a collection of performances with tempo variations as compared to a reference performance. To perform the analyses described, we take the following steps. First, we map the audio into some suitable feature space; we take the chroma features implemented in the MIRtoolbox [7] to obtain sequences of chroma vectors. We then normalize these sequences to functions over the unit interval. Finally, we use DTW to compute the relative tempo variations φ that best align the performances.

Explicitly, let $f^*, g^* : [0,1] \to \mathbb{R}^d$ be sequences of chroma vectors (in our case, $d = 12$, as analysis at the semitone resolution suffices). Then DTW finds the function φ that satisfies Properties 1 and 2 and minimizes $\|f^* - (g^* \circ \varphi)\|_2$, i.e., the L^2 norm of the difference between f^* and the reparametrized g^*. We generate φ in this way for all performances in the collection.

Our goal is to analyze variations between performances. Local tempo variation should be reflected in φ, provided there is not too much noise and the same event sequence is followed (e.g. no inconsistent repeats). The way we bring out the local tempo variation is by taking the derivative $\dot{\varphi}$ (cf. Section 3.2). A derivative larger/smaller than 1 indicates that the tempo decreases/increases relative to the reference performance. Since the tempo variations are given as a discrete functions, we need to approximate the derivatives. We do this by fitting a spline to the discrete data and analytically computing the spline's derivative.

To avoid the ground truth issue mentioned in Section 2, we devise several classes of artificial data, representing different types of performance variations for which we want to verify the behavior of our analysis. We verify whether the analysis is robust to noise and uniform variation in the overall tempo (the scalar value mentioned in Section 3). Furthermore, we consider different types of local tempo variations, which, without loss of generalization, are inspired by variations typically expected in classical music performances.

In the previous section, we mentioned two possible analysis strategies: considering alignments to a reference performance or between all possible pairs of performances. Since the artificial data are generated not to have outliers, it is difficult to apply the analysis that uses all possible pairs to the artificial data. We therefore focus on the case of using a single reference performance, although we will briefly return to the possibility of using all pairs in Section 5.

4.1 Generating Data

The data were generated as follows. We start with a sequence $g \in \mathbb{R}^{12\times m}$ of m 12-dimensional Gaussian noise vectors. Specifically, for each vector g_i, each element $g_{i,j}$ is drawn from the standard normal distribution $N(0, 1)$. We then generate a collection $\mathcal{C}$ of 'performances' based on g, for seven different variation classes. We normalize the vectors in $\mathcal{C}$ such that each element is between 0 and 1, as it would be in natural chroma vectors. The classes are defined as follows:

Class 1: Simulate minor noise corruption. A new sequence c is generated by adding a sequence $h \in \mathbb{R}^{12\times m}$ of 12-dimensional vectors, where each element $h_{i,j} \sim N(0, \frac{1}{4})$, so $c = g + h$. We expect this does not lead to any significant alignment difficulty, so the derivative of the resulting $\bar{\varphi}$ (which we will call $\dot{\bar{\varphi}}$) will be mostly flat.

Class 2: Simulate linear scaling of the overall tempo by stretching the time. Use spline interpolation to increase the number of samples in g, to simulate playing identically, but with varying overall tempo. If there are n sequences generated, vary the number of samples from $m - \frac{n}{2}$ to $m + \frac{n}{2}$. Since this only changes 'performances' on a global scale, this should give no local irregularities in the resulting $\dot{\bar{\varphi}}$.

Class 3: Simulate playing slower for a specific section of the performance, with sudden tempo decreases towards a fixed lower tempo at the boundaries, mimicking common tempo changes in an A-B-A song structure. Interpolate the sequence to have 1.2 times as many samples between indices $l = \frac{1}{3}m - \frac{1}{2}X$ and $h = \frac{2}{3}m + \frac{1}{2}X$, where $X \sim U(0, \frac{m}{10})$ (the same randomly drawn X is used in both indices). We expect $\dot{\bar{\varphi}}$ to be larger in the B part than in A parts. Since in different samples, the tempo change will occur at different times, transitions are expected to be observed at the tempo change intervals.

Class 4: A variation on class 3. Simulate a disagreement about whether to play part of the middle section slower. Let $k = h - l$. With a probability of 0.5, do not interpolate the section from $l + \frac{k}{3}$ to $h - \frac{k}{3}$. We expect similar results as for class 3 with the difference that in the middle of the B part, we expect an additional jump in $\dot{\bar{\varphi}}$. In the B part, $\dot{\bar{\varphi}}$ will jump to a lower value, which should still be larger than the value in the A part since only half of the performances decrease the tempo.

Class 5: Simulate a similar A-B-A tempo structure as in class 3, but change the tempo gradually instead of instantly over intervals of size roughly $\frac{1}{6}m$. From index $l_1 = \frac{1}{4}m - \frac{1}{2}X$ to $l_2 = \frac{5}{12}m + \frac{1}{2}X$, gradually slow down to 120% of the original tempo by interpolating over a quadratic query interval [1], then gradually speed up again the same way between indices $h_1 = \frac{7}{12}m - \frac{1}{2}X$ and $h_2 = \frac{3}{4}m + \frac{1}{2}X$. Here, $X \sim U(0, \frac{1}{18}m)$ and is drawn only once. Here again, we expect to see smaller values of $\dot{\bar{\varphi}}$ in the A parts and a higher value in the B part. Due to the gradual change in tempo, we expect a more gradual transition between A-B and B-A.

Class 6: A variation on class 5. Instead of varying the interval using X, vary the tempo. First speed up the tempo by a factor $1.3 + Y$ times the starting value (with $Y \sim U(-\frac{1}{10}, \frac{1}{10})$), then gradually slow down to a lower tempo and again speed up before the regular tempo of A is reached again. Here we expect to see a peak in $\dot{\bar{\varphi}}$ at the transition from A to B, before the lower value in the B part is reached and again a peak in the transition from B to A.

Class 7: Another variation on class 5: disagreement about speeding up or slowing down. Toss a fair coin ($p = 0.5$); on heads, gradually increase the tempo between l_1 and l_2 to $1.2 + Y$ times the starting value and decrease it again between h_1 and h_2 as in class 5. On tails, decrease the tempo to $0.8 + Y$ times the starting value between l_1 and l_2 and increase it again between h_1 and h_2, with $Y \sim U(-\frac{1}{10}, \frac{1}{10})$. We expect this to give much more noisy alignment, though there may be a more stable area in $\dot{\bar{\varphi}}$ where the tempos do not change, even though they are different.

[1] Normal linear interpolation corresponds to a constant tempo curve, but if the tempo curve changes linearly, the query interval for interpolation becomes quadratic.

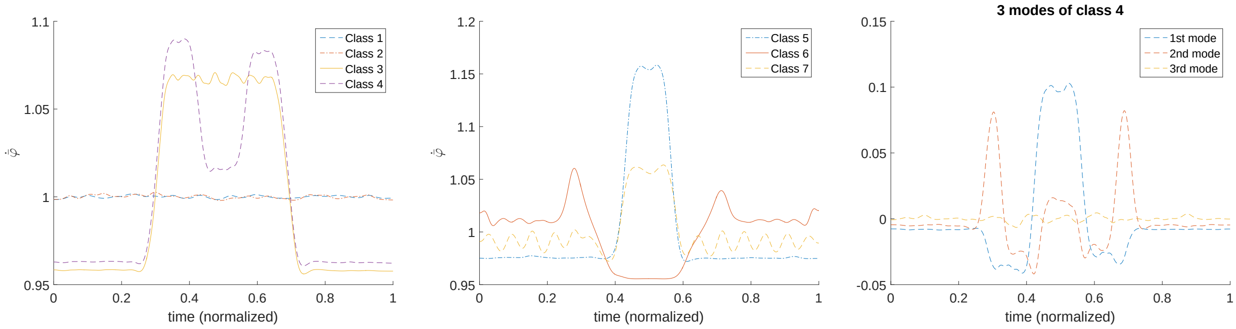

Figure 2. On the left: $\dot{\bar{\varphi}}$ for class 1–4. In the middle, $\dot{\bar{\varphi}}$ for class 5–7. On the right: the first three PCA modes for class 4.

When running our analysis on the classes of artificial data thus generated, we always took $m = 500$ and generated 100 sequences for each class. We used Matlab to generate the data, using 2017 as the seed for the (default) random number generator. A GitHub repository has been made containing the code for the analysis and for generating the test data [2]. The experiment was run 100 times, resulting in 100 $\bar{\varphi}$s and 100 sets of PCA modes; we took the mean for both and show the results in figures: Figure 2 (left and middle) show the derivatives when taking the mean (each time) as described in Section 3, while Figure 2 (right) shows what happens when taking the PCA, as also described in Section 3. We show the first three modes because these empirically turn out to cover most (around 90%) of the variance.

4.2 Discussion

We now briefly discuss what the analyses on artificial data tell us. First of all, the observed outcomes match our expectations outlined above. This demonstrates that our analysis can indeed detect the relative tempo variations that we know are present in performances of music.

We want to note that Figure 2 shows the derivatives of the relative tempo variation. For example, for class 3, all performances are shorter than the reference; therefore, they are stretched during the normalization. Consequently, the $\dot{\bar{\varphi}}$ in part A in the normalized performance is smaller than 1. This effect could be compensated by taking the length of the performances into account.

The PCA modes provide information about the variation in the set of performances. Figure 2 shows the first three modes found in Class 4. These three modes are the most dominant and explain more than 90% of the variation. The first mode has a large value in the middle part of the B section. This follows our expectation as only 50% of the performances slow down in this part, hence we expect much variation in this part. In addition, there are small values in the other parts of the B section. This is due to the fact that the performances do not speed up at the same time, so we expect some variation in these parts. Note that the principal modes are linear subspaces, hence sign and scale of the plotted function are arbitrary. An effect of this is that the modes cannot distinguish between speeding up the tempo or slowing it down. Since the first mode captures the main variation in the middle part of the B section, in the second mode the transitions between A and B are more emphasized. The third mode emphasizes the transitions too.

Finally, we note that it becomes possible to zoom in on a particular time window of a performance, in case one wants to do a detailed analysis. A hint of this is shown in Figure 1, left, where zoomed versions of φ are shown in insets. We have defaulted in our experiments to analyzing performances at the global level, and consider it future work to explore what information will be revealed when looking at the warping paths up close.

5. APPLICATIONS

Now that we have validated our approach, we describe several applications in which our method can be employed.

5.1 Analyzing Actual Performances

As mentioned in Section 3, we can analyze relative differences between a chosen reference performance and the other performances, or between all possible pairs of performances. We have access to the Mazurka dataset consisting of recordings of 49 of Chopin's mazurkas, partially annotated by Sapp [21]. Note that our analysis can handle any collection of performances and does not require annotations. Since we have no ground truth, it is difficult to make quantitative statements, but in this and the following subsection, we will discuss several illustrative qualitative examples.

In Figure 3, we show $\dot{\bar{\varphi}}$ for Mazurka op. 30 no. 2 for both approaches. Taking all pairs into consideration results in lower absolute values, as well as an apparent lag. For both approaches, it turns out the most important structural boundaries generally show up as the highest peaks. Another feature that stands out in both plots is the presence of peaks at the beginning and end. These can be interpreted as boundary effects, but we believe the final peak also is influenced by intentional slowing down by the musicians in a final retard [25].

Another example of applying the analysis on all pairs of performances is given in Figure 4. Here, we see two more

[2] `https://github.com/asharkinasuit/ismir2017paper`.

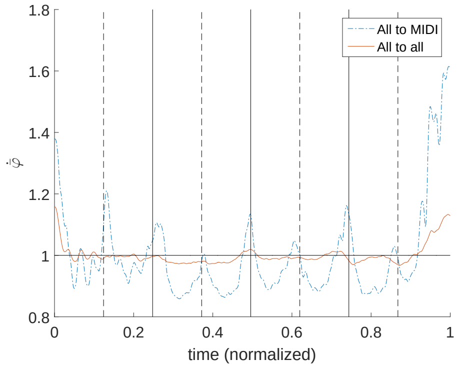

Figure 3. Sample showing $\dot{\bar{\varphi}}$ for Mazurka op. 30 no. 2, comparing warping to a deadpan MIDI and warping everything to everything. Note the smoothing effect in the latter case. Salient structural parts are indicated with vertical lines: repeats (dotted) and structural boundaries (solid).

interesting features of the analysis. Firstly, it tends to hint at the musicians' interpretation of the structure of the piece (as also in Figure 3); the start of the melody is indicated with the vertical dashed line. Most performers emphasize this structural transition by slowing down slightly before it. However, the time at which they slow down varies slightly (compare this to e.g. class 3 and 5 of our artificial data). This will show in φ, and consequently in $\dot{\bar{\varphi}}$. Secondly, we note that ornaments tend not to vary tempo as much: the thin section in the figure is closer to 1 than the peak near the start of the melody. This helps corroborate Honing's results, e.g. [2, 5].

5.2 Guiding Synthesis

For the performances in question, we know the piece that is performed and we have a score available. A direct acoustic rending of the score (via MIDI) would sound unnatural. Now, reparametrizations and their means are just functions, which we can apply to any other suitably defined function. Following the suggestion in [20] that a generated 'average' performance may be more aesthetically pleasing, we can now use these functions for this: by applying the $\bar{\varphi}$ derived from a set of performances to a MIDI rendition, a more natural-sounding result will indeed be obtained. As an example, we ran our analysis on Chopin's mazurka op. 24 no. 2 with the MIDI rendition as reference performance and applied the resulting reparametrization to the MIDI[3]. Note that, as in Figure 3, the tempo naturally decreases towards the end.

Applying $\bar{\varphi}$ directly to audio is not the only thing that we can do. One possibility is exaggeration of tempo variation. To amplify sections that show major tempo variation, we can modify the φ by squaring it. Alternatively, to better display the tempo variations in an individual performance, we can rescale the function $\varphi - \bar{\varphi}$, capturing the difference of the actual performance to the mean in a performance

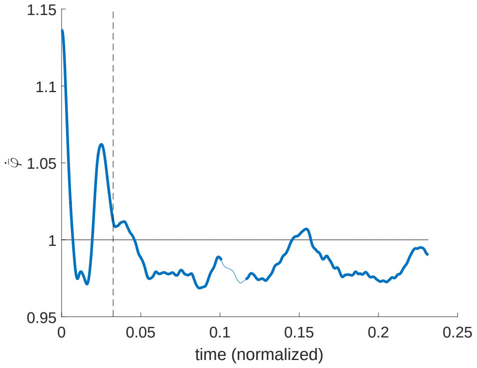

Figure 4. $\dot{\bar{\varphi}}$ of the start of mazurka op. 17 no. 4. The start of the melody is marked with a vertical dashed bar, while the *delicatissimo* section is drawn in a thinner line.

collection. Such modifications offer useful analysis tools for bringing out more clearly the sometimes subtle effects employed by professional musicians.

Another possibility is to take φ from various sources, e.g., by generating φ for several different reference performances, and applying them to a MIDI rendition with various coefficients to achieve a kind of mixing effect. Finally, the principal modes of variation in the set can be used to modify the tempo in which the MIDI is rendered. Example audio files are available on request for any of these different ways of rendering musical scores using information from actual performances.

6. CONCLUSIONS AND FUTURE WORK

We have presented a formal framework for analyzing relative local tempo variations in collections of musical performances, which enables taking the mean and computing a PCA of these variations. This can be used to analyze a performed piece, or synthesize new versions of it.

Some challenges may be addressed in the future. One would be to give a more rigorous interpretation to the case of taking all pairwise comparisons into account. Furthermore, quantification of variation still presently is used in a relative fashion; our analysis indicates some amount of variation, but further interpretation of this amount would be useful. One might also substitute other DTW variants that can e.g. deal more intuitively with repeat sections [4].

Furthermore, while the studied variation classes were inspired by local tempo variations in classical music performances, it should be noted that our framework allows for generalization, being applicable to any collection of alignable time series data. Therefore, in future work, it will be interesting to investigate applications of our proposed method on other types of data, such as motion tracking data.

7. REFERENCES

[1] A. Benetti Jr. Expressivity and musical performance: practice strategies for pianists. In *2nd Performance Studies Network Int. Conf.*, 2013.

[3] See `https://github.com/asharkinasuit/ismir2017paper`, which includes the original for comparison.

[2] P. Desain and H. Honing. Does expressive timing in music performance scale proportionally with tempo? *Psychological Research*, 56(4):285–292, 1994.

[3] S. Flossmann, M. Grachten, and G. Widmer. Expressive Performance Rendering with Probabilistic Models. In *Guide to Computing for Expressive Music Performance*, pages 75–98. Springer, 2013.

[4] M. Grachten, M. Gasser, A. Arzt, and G. Widmer. Automatic alignment of music performances with structural differences. In *ISMIR*, 2013.

[5] H. Honing. Timing is Tempo-Specific. In *ICMC*, 2005.

[6] K. Kosta, O. F. Bandtlow, and E. Chew. Practical Implications of Dynamic Markings in the Score: Is Piano Always Piano? In *53rd AES Conf. on Semantic Audio*, 2014.

[7] O. Lartillot and P. Toiviainen. A matlab toolbox for musical feature extraction from audio. In *Int. Conf. Digital Audio Effects*, pages 237–244, 2007.

[8] E. Liebman, E. Ornoy, and B. Chor. A phylogenetic approach to music performance analysis. *Journal of New Music Research*, 41(2):195–222, 2012.

[9] C. C. S. Liem and A. Hanjalic. Expressive Timing from Cross-Performance and Audio-based Alignment Patterns: An Extended Case Study. In *ISMIR*, pages 519–524, 2011.

[10] C. C. S. Liem and A. Hanjalic. Comparative analysis of orchestral performance recordings: an image-based approach. In *ISMIR*, 2015.

[11] C. C. S. Liem, A. Hanjalic, and C. S. Sapp. Expressivity in musical timing in relation to musical structure and interpretation: a cross-performance, audio-based approach. In *42nd AES Conf. Semantic Audio*, 2011.

[12] L. Lu, M. Wang, and H. Zhang. Repeating pattern discovery and structure analysis from acoustic music data. In *6th ACM SIGMM Int. Workshop on Multimedia Information Retrieval*, pages 275–282. ACM, 2004.

[13] M. Müller. *Fundamentals of Music Processing: Audio, Analysis, Algorithms, Applications*. Springer, 2015.

[14] M. Müller and S. Ewert. Joint Structure Analysis with Applications to Music Annotation and Synchronization. In *ISMIR*, pages 389–394, 2008.

[15] M. Müller and F. Kurth. Enhancing similarity matrices for music audio analysis. In *IEEE Int. Conf. Acoustics, Speech and Signal Processing*, volume 5. IEEE, 2006.

[16] O. Nieto and T. Jehan. Convex non-negative matrix factorization for automatic music structure identification. In *IEEE Int. Conf. Acoustics, Speech and Signal Processing*, pages 236–240. IEEE, 2013.

[17] M. Ohriner. What can we learn from idiosyncratic performances? Exploring outliers in corpuses of Chopin renditions. In *Proc. of the Int. Symp. on Performance Science*, pages 635–640, 2011.

[18] Y. Panagakis, C. Kotropoulos, and G. R. Arce. ℓ_1-graph based music structure analysis. In *ISMIR*, 2011.

[19] J. Paulus and A. Klapuri. Music structure analysis by finding repeated parts. In *Proc. of the 1st ACM Audio and Music Computing Multimedia Workshop*, pages 59–68. ACM, 2006.

[20] B. H. Repp. The aesthetic quality of a quantitatively average music performance: Two preliminary experiments. *Music Perception: An Interdisciplinary Journal*, 14(4):419–444, 1997.

[21] C. S. Sapp. Comparative Analysis of Multiple Musical Performances. In *ISMIR*, pages 497–500, 2007.

[22] C. S. Sapp. Hybrid numeric/rank similarity metrics for musical performance analysis. In *ISMIR*, pages 501–506, 2008.

[23] J. Serrà, M. Müller, P. Grosche, and J. L. Arcos. Unsupervised music structure annotation by time series structure features and segment similarity. *IEEE Trans. Multimedia*, 16(5):1229–1240, 2014.

[24] J. B. L. Smith, I. Schankler, and E. Chew. Listening as a Creative Act: Meaningful Differences in Structural Annotations of Improvised Performances. *Music Theory Online*, 20(3), 2014.

[25] J. Sundberg and V. Verrillo. On the anatomy of the retard: A study of timing in music. *Journal of the Acoustical Society of America*, 68:772–779, 1980.

[26] G. Widmer and A. Tobudic. Playing Mozart by Analogy: Learning Multi-level Timing and Dynamics Strategies. *Journal of New Music Research*, 32(3):259–268, 2003.