Phase-plane representation and visualization of gestural structure in expressive timing

Maarten Grachten[1], Werner Goebl[1], Sebastian Flossmann[1], and Gerhard Widmer[1,2]

[1] *Department of Computational Perception, Johannes Kepler University, Linz, Austria*
[2] *Austrian Research Institute for Artificial Intelligence, Vienna, Austria*
{maarten.grachten,werner.goebl,sebastian.flossmann,gerhard.widmer}@jku.at

Abstract

For the past few decades there has been considerable scientific interest in expression in music performance (Gabrielsson, 2003). A particularly relevant aspect of music performance is expressive timing, that is, the intentional fluctuations of tempo during a performance. Accordingly, expressive timing has been one of the major topics in music performance research. As an expressive parameter, timing is used to clarify the musical structure of the piece (Clarke, 1988), among other things. The problem of explaining expressive timing in music performances can be regarded as a special case of a very wide range of problems where we want to learn experimentally about the temporal behavior of some *dynamical system* based on limited observation. A common way of studying data in dynamical systems theory is by *phase-plane* representation. In this paper we argue that phase-plane representations of expressive timing provide a useful way of visualizing data, and furthermore, we show that such representations are promising in the context of performer characterization and identification.

1 Introduction and related work

For the past few decades there has been considerable scientific interest in expression in music performance (Gabrielsson, 2003). A particularly relevant aspect of music performance is expressive timing, that is, the intentional fluctuations of tempo during a performance. Accordingly, expressive timing has been one of the major topics in music performance research. A well-known function of timing as an expressive parameter is the clarification of structural aspects of the music, like metrical, phrase, and voice structure (Clarke, 1988; Palmer, 1997). Furthermore, timing and other temporal aspects such as global tempo and articulation play a role in the communication of semantic content, including emotional (Juslin and Sloboda, 2001), and sensorial (Canazza et al., 2003) information. In addition to establishing such global relationships, more detailed accounts of expressive timing have been given using several distinct methodologies, such as *analysis-by-synthesis* (Sundberg et al., 1991), and machine learning (Widmer, 2003). A particularly profound and relatively large-scale analysis of expressive timing can be found in Repp (1992).

The problem of explaining expressive timing in music performances can be regarded as a special case of a very wide range of problems where we want to learn experimentally about the temporal behavior of some *dynamical system* based on limited observation, as for example in population biology, and meteorology. This dynamical systems paradigm aims at models that describe how the state of the system changes over time. Rather than predicting individual *state-space* trajectories, the focus is on the qualitative structure of the state-space that results from the influence and interaction of possibly unknown factors and constraints. Such a dynamical approach has been applied to various aspects of music, including acoustic modeling (Schoner, 1997), gesture-based virtual instrument control (Métois, 1996), and analysis of temporal/pitch complexity of compositions (Boon and Decroly, 1995).

In the case of music performance, as mentioned above, past research has already revealed some valuable insights about the way factors like musical structure and intended mood influence expressive timing. In addition, there has been a long-standing metaphor of music as a form of motion (Truslit, 1938; Friberg and Sundberg, 1999; Todd, 1992). This metaphor has led to kinematic models of expressive dynamics and timing in music performance (Todd, 1992; Friberg and Sundberg, 1999). Even if the proposed models have limitations, and the motion metaphor maybe incomplete as an explanation for expressive phenomena (Desain and Honing, 1996; Honing, 2005), we believe that it is worthwile to explore the dynamical systems view of expressive performance in more depth.

As in any field concerned with data analysis, visualization techniques are often very helpful for studying observations from dynamical systems. In particular, *phase-plane* visualization can reveal characteristics of time-series data that are less evident from plots of the data against time (in the rest of the paper we will refer to this as *time-series plots*). The phase-plane is a two-dimensional plot of some aspects of a dynamical system. In the case of a simple pendulum for example, a useful phase-plane is the one that plots the velocity against the position of the pendulum, as it completely describes the behavior of the system. If multiple signals are observed from the system, phase-plane trajectories can be drawn by plotting the signals against each other. An example of this in expressive music performance is the *performance worm* (Langner and Goebl, 2003), which visualizes performances by plotting loudness versus tempo.

In this paper we choose a different phase-plane method, that exclusively represents tempo information.[1] We focus on first-order and second-order phase-planes. The former plots the derivative of tempo versus tempo, whereas the latter plots the second versus the first derivative of tempo. After introducing the visualization method using schematic examples and describing the procedure for computing phase-plane trajectories from expressive performances in section 2, we review two expressive gestures in performances of Schumann's Träumerei (section 3). Finally, in section 4, we describe an experiment in which we determine the effects of several parameters of phase-plane representation on tasks like performer identification.

2 Phase-plane plots versus time-series plots

An obvious question that comes to mind when considering phase-plane representations of a function (or a time-series) as an alternative to plotting the function against time, is what new insights it can possibly give. After all, the derivatives are fully determined by the function, they don't convey any information that is not contained in the function itself. Rather than providing new information, phase-plane representations show a new perspective on the data, just like for example a transformation of a function from the time to the frequency domain provides a new perspective. The essential difference from time-series plots is that the time dimension is implicit rather than explicit in the phase-plane. Whether this is an advantage depends on the intended kind of analysis. For example, if the aim is to get an impression of the trends in absolute tempo over the course of a performance, a time-series plot may be more useful than a phase-plane plot. On the other hand, if the focus is on the particular form that the change of tempo takes, then phase-plane plots may provide better insight. The reason for this is that the tempo trajectory in the phase-plane expresses exclusively the change in tempo — any episodes of constant tempo are projected to a single point in the phase-plane. As opposed to time-series plots, where tempo trajectories by definition advance steadily in one dimension (time), in phase-plane plots the change of tempo is expressed in two dimensions, leading to trajectories that are visually more distinct than their equivalent time-series plots. We will illustrate this shortly.

This emphasis on the dynamic aspects of tempo in phase-plane representations is in accordance with the observation that the expressive use of timing is mainly manifested through the momentary fluctuations of tempo. Absolute tempo, or large scale trends in tempo are not commonly regarded as the principal expressive parameters, even if they do belong to the expressive degrees of freedom of the performer.

2.1 Examples of basic curve types

To get a feel for how to interpret phase-plane representations, we briefly discuss the phase-plane trajectories of some archetypal curves. In the first column of figure 1, five basic curves are shown as a function x of time t. The second column shows the corresponding first-order phase-planes, representing the curve as a trajectory through the dx/dt vs $x(t)$ plane, that is, the first derivative of $x(t)$ against $x(t)$ itself. The

[1] The phase-plane visualization method was introduced in Grachten et al. (2008). With the current paper, we extend this work, by a quantitative and qualitative evaluation of the phase-plane representation.

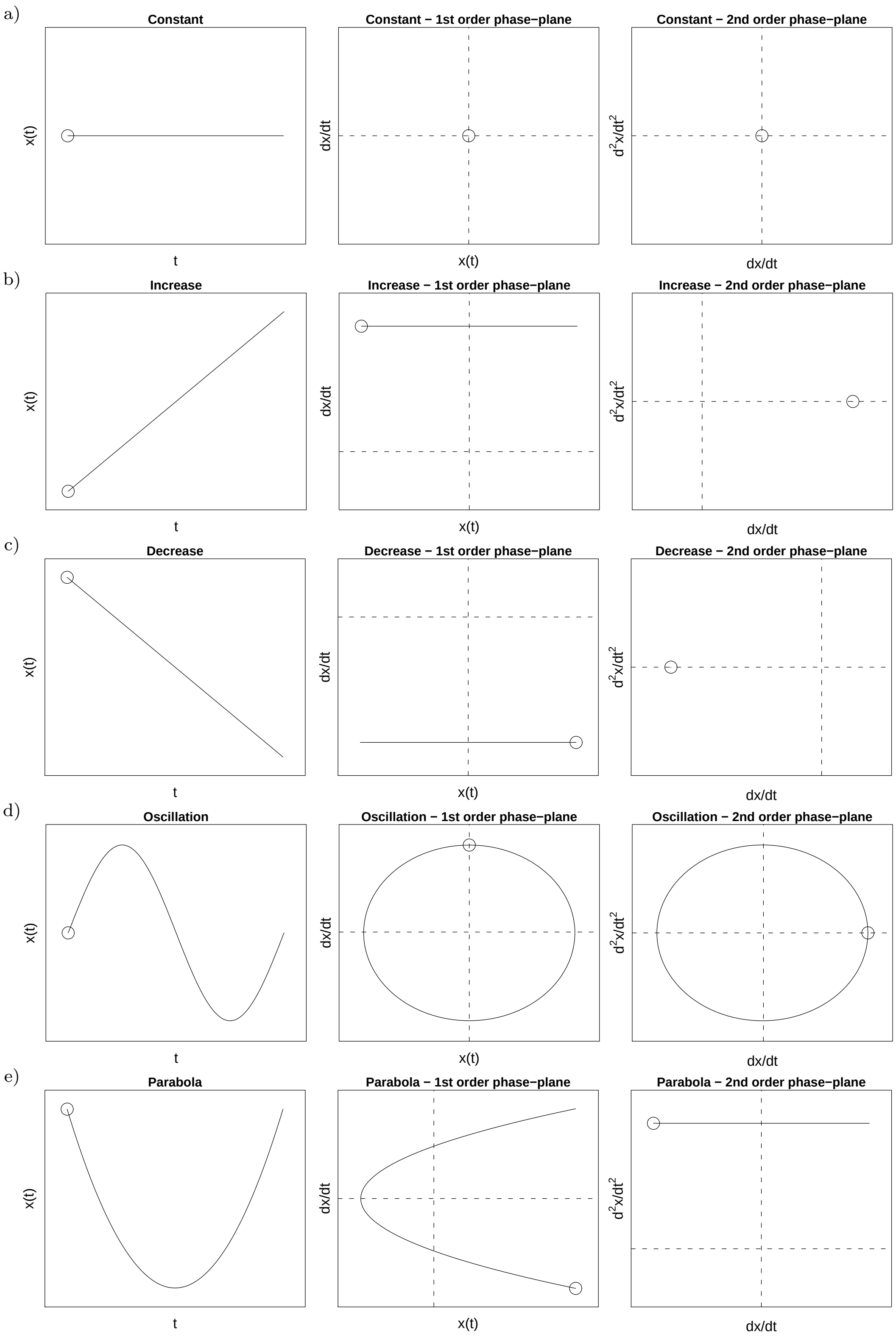


Figure 1: Examples of five basic curve types (first column), and their first and second order phase-plane trajectories (second and third columns respectively); Horizontal and vertical dashed lines represent x and y axes respectively; Circles indicate the beginning of the curves/trajectories; Units are arbitrary

last column shows the second-order phase-planes, formed by d^2x/dt vs dx/dt. The circles indicate the beginning of the curves, and their corresponding phase-plane trajectories. The horizontal and vertical dashed lines indicate the origin in the phase-planes.

Note that constant tempo (row (a) in figure 1) corresponds to a single point in the phase-planes, as all derivatives are zero.[2] Constant change of tempo (rows (b) and (c)) leads to a displacement along the $x(t)$ (horizontal in the first-order phase-plane) axis and a constant offset along the dx/dt axis (vertical in the first-order phase-plane, and horizontal in the second-order phase-plane).

Row (d) shows one period of a simple harmonic, or oscillatory motion. This type of motion is defined by a second order differential equation which has sinusoidal functions as its solutions. Such functions correspond to a circular motion in both phase-planes, where the end position of the trajectory is equal to its starting position. This example illustrates how, as the time dimension is implicit, repeated curve segments map to the same trajectory in the phase-plane. Note that due to the derivative relationship between the vertical dimension with respect to the horizontal dimension, the movement of any phase-plane trajectory is necessarily clockwise around the origin. More precisely, the trajectory always moves leftward below the horizontal axis, and rightward above it, and is exactly perpendicular to the horizontal axis at the time of crossing it.

Finally, row (e) shows a parabolic curve $x(t) = t^2$. Since its first derivative $dx/dt = 2t$ is linear in time, the first-order phase-plane is also a parabola, with the horizontal and vertical axis interchanged. The second-order phase-plane trajectory is a straight line segment, since $dx/dt = 2$. Note that although it is hard to visually distinguish the parabola from a semiperiod of a simple oscillation in the time-series plot (first column), the phase-plane trajectories of both types of curves are very distinct. This is a particularly interesting feature for mathematical modeling of tempo curves, such as in Todd (1985), and Repp (1992).

2.2 From time-series to phase-plane trajectories

The concept of a tempo curve, even if ubiquitous in expressive music performance research, is not straightforward. Given that tempo can be loosely defined as the rate at which events take place, it is inherently related to a temporal context of events, rather than a single point in time. For the sake of quantifying tempo over the course of a performance, it is commonly measured as the reciprocal of the interval between two consecutive metrical beats (IBI), and this value is associated either with the first or the second of the beats for which the IBI was measured. As the tempo quantity is undefined in the absence of events, it is questionable whether tempo is perceived as a constant entity by humans (Desain and Honing, 1993), and therefore whether it is justified to interpolate the time-series of tempo values to obtain a continuous tempo curve. On the other hand, there are reasons to consider the construction of continuous and smooth tempo curves from discrete timing information as appropriate. Firstly, in order for the rhythmical structure of a piece to be perceptible, tempo must satisfy certain smoothness constraints Honing (2004). Furthermore, Dixon et al. (2006) argue that tempo perception of human listeners discards a certain variability in performed note onsets, implying a beat-grid that is slightly smoothed with respect to the exact timing of onsets. This coherence between consecutive tempo perceptions is expressed by the representation of tempo as a continuous function of time.

The problem of finding a function that fits to a series of data values is well-known in statistical data analysis, since a very common situation in empirical studies is to have a series of measurement values that we hypothesize or assume to be result of some process of which the behavior can be adequately described by some smooth function. As is unavoidable in any measurement, the measured values will include measurement errors and possibly other distortions of the values that we actually intended to measure. This view is known as the *signal plus noise* model, which is formally represented as:

$$\mathbf{y} = x(\mathbf{t}) + \mathbf{e} \tag{1}$$

where $\mathbf{y}$ is a vector of length n containing the measured values, $\mathbf{t}$ is a vector of length n containing the time values associated with each measurement, x is the unknown function that we wish to estimate, and $\mathbf{e}$ is a vector of length n containing the error values associated with each measurement. The function x is often chosen to be of the form:

[2] We interpret the curves as tempo curves, although these remarks of course hold independently of the interpretation of the dimensions

$$x(\mathbf{t}) = \mathbf{c}'\phi \tag{2}$$

that is, a linear combination of a set of K basis-functions ϕ, where $\mathbf{c}$ is a vector of length K containing the weight for each basis-function. The fitting of the function x to the data $\mathbf{y}$ can be done by minimizing the summed squared error:

$$SSE = ||\mathbf{y} - \mathbf{\Phi c}||^2 \tag{3}$$

where $\mathbf{\Phi}$ is a n by K matrix such that $\mathbf{\Phi}_{i,k}$ contains $\phi_k(i)$, value of the k-th basis-function at sampling point i.

As the number of basis-functions K increases, the fit to the data becomes better, reducing the *bias* of the estimation. But large values for K also increase the *variance* of the estimation, resulting in a less smooth fitted curve. To take the smoothness constraint into account, a penalty term for roughness is included the quantity that is minimized:

$$PENSSE = SSE^2 + \lambda PEN \tag{4}$$

The relative importance of the penalty term is controlled by the smoothing parameter λ. The penalty term quantifies the roughness as the integrated square of the second derivative of x:

$$PEN = \int [D^2x(s)]^2 ds \tag{5}$$

This minimization criterion is independent of the choice of the system of basis-functions ϕ. There is a wide variety of bases that can be sensibly used. Typical bases are Fourier series and polynomials. Furthermore, with a slight change of the minimization criterion, kernel smoothing (e.g. using a Gaussian kernel) can be construed as a special case of basis expansion with one basis-function $\phi(t) = 1$.

In the work described here, we use a B-spline basis for smoothing, as described in Ramsay and Silverman (2005). B-splines are *piecewise polynomial*. This means that the spline consists of segments defined by a series of breakpoints, and on each of those segments S the B-spline is a polynomial. A B-spline S is defined by an order m, and a sequence of breakpoints τ, and is computed from a set of basis-functions B:

$$S(t) = \sum_{k=1}^{m+L-1} c_k B_k(t, \tau) \tag{6}$$

Here, $B_k(t, \tau)$ is the value at point t of the k-th basis-function. L is the number of intervals as defined by the breakpoint sequence τ. The basis-functions are themselves B-splines, with compact support. They are constructed recursively from B-splines of a lower degree. B-spline smoothing of data is achieved by choosing the coefficients $\mathbf{c} = (c_1, \cdots, c_{m+L-1})$ such that the criterion $PENSSE$ is minimized.

After computing $\mathbf{c}$, the phase-plane representations are obtained by computing the first and second derivatives of the spline S, $D^1S(t)$ and $D^2S(t)$.

3 Phase-planes of expressive gestures in Schumann's Träumerei

We will illustrate the phase-plane visualization for two performances of a melodic gesture (or motif) from Schumann's Träumerei. We will do this in relation to an earlier study of expressive timing in that piece (Repp, 1992).[3] Repp describes the results of an extensive study of expressive timing of 28 performances of this piece by renowned pianists. A part of this study is a detailed investigation of the timing of notes in a particular melodic gesture (labeled MG2 in Repp, 1992) present in the piece. A majority of the performances showed an IOI pattern that could be modeled quite well with a parabolic curve segment, where the curvature of the fitted parabola varied from performance to performance. However, the goodness of fit of the curves to the measured IOI's was only informally assessed.

Figure 2 shows the IOI curves (the reciprocals of the tempo curves) and corresponding phase-plane trajectories of two performances of MG2, by Zak (1960) and Horowitz (1947) respectively. The first

[3] The data used here originates from that study

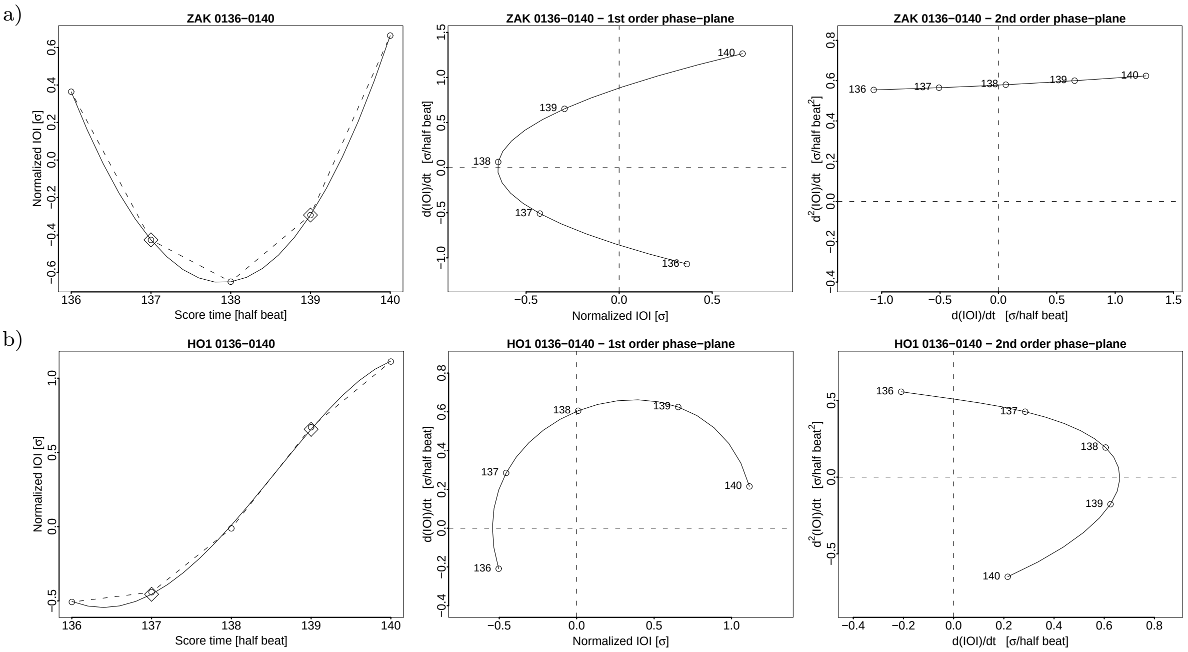


Figure 2: Fitted IOI curves and corresponding phase-plane trajectories for two exemplary performances of a melodic gesture (MG2, fourth instance, half beats 136–140) from Schumann's Träumerei, by Zak (1960) (a), and Horowitz (1947) (b) respectively; In the first column, the circles connected by dashed lines are the measured IOI values (normalized), and the solid line is the fitted spline; Diamonds indicate the breakpoints of the spline; the phase-plane trajectories are annotated with half beat numbers; The units are shown in square brackets in the axis labels; σ denotes the standard deviation of the normalized IOI values

column shows the normalized measured IOI values as circles connected by dashed lines, together with the fitted splines, as solid curves. The splines are constructed from cubic polynomials, and are thus of order 4. The breakpoints are identical for both examples, and their positions (at half beats 137 and 139) are chosen manually to provide a good fit to the data with a relatively low number of breakpoints.[4] The roughness penalty λ is set to .001.

The phase-plane trajectories of Zak's performance (figure 2a) indeed bear a striking resemblance to those of the parabola shown before, in figure 1e. The first-order phase-plane trajectory strongly resembles a rotated parabola, and the second-order phase-plane trajectory is approximately a straight horizontal line segment going from left to right. Consequently, this is a performance that can be very well modeled with a parabola. Horowitz's performance (figure 2b) on the other hand, is apparently not a prototypical instance of a parabola. A parabola fitted to this performance would show a rather poor fit, especially due to the non-constant curvature in the IOI data. This is confirmed by the phase-plane trajectories of the fitted spline, which are rather different from those of the parabola in figure 1e. In particular, the first-order phase-plane trajectory is circular rather than parabolic, and also the second-order phase-plane trajectory is curved rather than straight. Especially the first-order phase-plane trajectory suggests oscillatory motion.

Discussion

The purpose of this example is not so much to challenge the hypothesis that this particular performance gesture can be adequately modeled with a parabola (that would require a more thorough investigation), but to show that the phase-plane visualization can 'amplify' differences between time series plots. As the example illustrates, in some cases where one may be inclined to apply the same model for two tempo (or IOI) curves, the phase-planes show very distinct trajectories for the two curves.

We believe that phase-plane visualizations can be of help in modeling expressive tempo precisely for

[4] A common practice in spline fitting is to put a breakpoint at each data point

this reason. It is important to note that the phase-plane trajectories of common functions like a parabola or a sinusoidal, did not emerge because we used those functions as a model. This illustrates the flexibility of the spline basis expansion for fitting the data.

However, care must also be taken when interpreting the phase-plane trajectories. Firstly, the examples shown here are based on a small number of data points, and thus the trajectories to a certain degree describe 'space' in between the measured data points. Secondly, the fact that the phase-plane trajectories tend to diverge more than the time series plots has the downside that small artifacts of the fitting (for example some ripples in the curve between two data points) can have a large impact on the appearance of the trajectories. Therefore, when interpreting phase-plane trajectories, it is essential that the fit of the basis expansion to the data is also inspected, to verify that the major forms in the trajectories are not due to curve fitting artifacts.

4 An assessment of alternative phase-plane representations

In the previous section we have illustrated the use of the phase-plane representation to visually compare expressive timing gestures of different performers. We have shown the phase-plane trajectories in different *spaces*, in particular the first-order phase-plane, and the second-order phase-plane. The smoothness of the spline-approximation of the measured data was chosen manually. Affine transformations, which are typically used to compare geometric shapes in a scale, position, and rotation independent way, were not applied in the examples given.

The question which phase-plane order is most informative, at which degree of smoothing, and with or without applying affine transforms, is likely to depend on the purpose of the analysis. Generally speaking, we would like to know whether similarities between phase-plane trajectories tend to be indicative of musically relevant commonalities between the performances that those phase-plane trajectories represent. In the current experiment, we limit ourselves to two such aspects: the identity of the performer, and the place of occurrence within the musical piece.

In the literature there is evidence that performers have, sometimes strongly, idiosyncratic ways of performing music (Repp, 1998). On the other hand, it is known that timing profiles of performances tend to have a component that is common among performers, mainly due to generally adopted (but mostly implicit) conventions of how to convey musical structure through timing (Todd, 1985). The context of the occurrence of a phrase may therefore have an effect on how it is played. For example, a phrase or musical fragment that is repeated several times throughout a piece, may be played differently depending on whether it occurs at the beginning, middle, or ending.

Parallel to these two aspects of expressive performance, in the experiment described here, we focus on the comparison of phase-plane trajectories for two contrary purposes. The first is performer identification, in which the task is to partition a set of phase-plane trajectories such that for each performer there is a group that contains all trajectories of that performer. The second task is to group fragments of phase-plane trajectories according to their position in the piece. We regard the accuracy on each of these tasks as an indicator of how well the phase-plane representation reflects the performer's characteristics, and the score context, respectively. Furthermore, we investigate whether, and how, the choice of the parameters of representation (i.e. phase-plane order, the degree of smoothing, normalization) affects this balance.

4.1 Data and Procedure

To address these questions, we use data sets containing phase-plane trajectories describing the performances of musical fragments. Each data set contains performances of a single musical piece by several performers. Rather than using the entire performances, we select those parts that correspond to the occurrences of a single musical fragment which is repeated multiple times[5]. Thus, with K performers and N occurrences of the selected fragment, a data set is a collection of $K \times N$ phase-plane trajectories. Each phase-plane trajectory is labeled with a pair of identifiers (k, n), where k denotes the k-th performer, and n indicates that the instance is the n-th occurrence of the fragment in the piece. We can then compare a clustering of the data set that was computed from the phase-plane trajectories, to the performer-partitioning of the data set on the one hand, and to the *order-of-occurrence*-partitioning on

[5] The pieces are selected to contain at least three occurrences of a single musical fragment

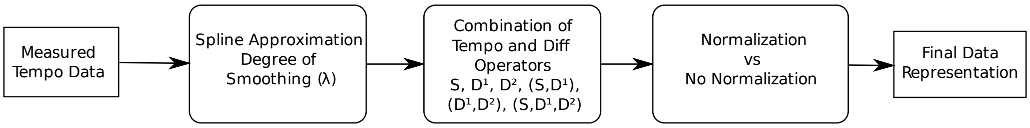

Measured Tempo Data
Spline Approximation Degree of Smoothing (λ)
Combination of Tempo and Diff Operators S, D¹, D², (S,D¹), (D¹,D²), (S,D¹,D²)
Normalization vs No Normalization
Final Data Representation

Figure 3: Processing steps from measured data to the data used for clustering

the other. This gives us a way to evaluate different representations of the performance data with respect to the partitioning tasks.

Because our prime interest is in the assessment of experimental parameters that concern the representation of the data, rather than parameters that concern the decision-theoretic aspect of clustering, we avoid direct partitional clustering techniques (such as k-means, or EM clustering). Instead, we opt for hierarchical clustering, and using full knowledge of the instance labels, select a subset of nodes from the dendrogram to form an optimal partitional clustering.

In the rest of this subsection we will respectively discuss the alternative data representations of tempo curves that we consider, the musical material used in the experiment, and the evaluation of results.

4.1.1 Alternative Data Representations

The tempo measurements from the performances are represented as a series of timestamp/tempo pairs, i.e. $(\mathbf{t}, \mathbf{y}) = (t_1, y_1), \cdots, (t_M, y_M)$, where t_i denotes the time at which the i-th beat occurred (the whole piece spanning M beats), and y_i the tempo value at that time, measured as the reciprocal of the IBI between beat i and beat $i-1$. The vector $\mathbf{y}$ is normalized by subtracting the mean and dividing by the standard deviation of the values. We use S to denote the spline approximation of the data as described in subsection 2.2, that is $S(\mathbf{t}) \approx \mathbf{y}$.

The parameter λ in equation (4) weights the impact of the roughness penalty of the functional approximation on the minimization criterion. Low values of λ will provide better approximations $S(\mathbf{t})$ of the measured tempo values $\mathbf{y}$, but at the cost of increasing fluctuations in between the measurement points $\mathbf{t}$. Higher values of λ will result in smoother curves, that may not fully approximate the measured values however. In the experiment we use spline-approximations with varying λ values, that jointly cover the effective range of the parameter.

To produce the phase-plane trajectories of S, we use the first and second derivatives of S, D^1 and D^2 respectively, which can be obtained in closed form from S. In principle, any subset of (S, D^1, D^2) can be used to represent the data. In the rest of the paper, we will refer to such subsets as *spaces*. By a trajectory of a musical fragment spanning beats k through l in the space (S, D^1, D^2), we mean a sequence $v_k, \cdots, v_l$, where $v_i = \langle S(t_i), D^1(t_i), D^2(t_i) \rangle$ (and analogous for other spaces). In the current experiment, we compare the six (redundant) spaces S, D^1, D^2, (S, D^1), (D^1, D^2), and (S, D^1, D^2).

Another issue is the relative size, position, and orientation of phase-plane trajectories. If we treat phase-plane trajectories as geometrical shapes annotated with *landmark points*, then a natural method of comparison would be by *Procrustes analysis* (Goodall, 1991). This method is common in biomedical shape comparison, and consists in applying scaling, translation, and rotation transformations to one shape to make it maximally similar to another, before calculating a measure of distance between the landmark points that define the shapes. The interpretation of phase-plane trajectories as geometrical shapes however does not fit completely with the musical aspects that the trajectories represent. In particular, using a distance measure that is rotation-invariant would imply that, for example, a ritardando (a lower hemi-circle) could be matched perfectly to an accelerando (an upper hemi-circle), which is obviously undesirable. On the other hand, translation and scaling invariance may be sensible characteristics of a distance measure, since they allow us to ignore differences in the absolute range and the 'intensity' of gestures. The translation and scaling of trajectories to a common position and size is equivalent to normalization of the data.

The above steps in the construction of the final data representation are shown schematically in figure 3.

Composer, Piece	Fragment name	Fragment start positions (in beats)	Frag. length (in beats)	Cumulative frag. length (% of piece)	No. of performances
Chopin, Op. 10, No. 3	A	3 (repeated at 67, 491)	16	±16%	8
Chopin, Op. 10, No. 3	B	20 (84, 508)	22	±21%	8
Chopin, Op. 28, No. 17	A	14 (62, 206, 434)	35	±26%	10
Schumann, Op. 15, No. 7	MG1/2	3 (35, 67, 99, 131, 163)	9	±28%	28

Table 1: Description of the data used

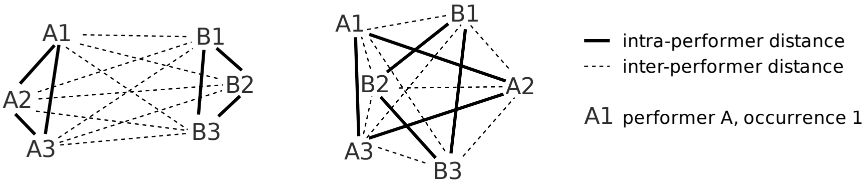


Figure 4: Left: small intra-performer distances and large inter-performer distances (that is, high intra/inter performer gap); Right: intra-performer and inter-performer distances are similar (that is, low intra/inter performer gap)

4.1.2 Musical Material

For this experiment we used four distinct fragments from three classical piano pieces from the romantic period, as shown in table 1. We manually selected score fragments that occurred identically at least three times throughout the piece. An additional criterion was that the fragments be of reasonable duration, roughly speaking two bars or longer, to avoid fragments that are too short to extract any meaningful expressive information. All timing data was extracted by manual annotation from commercial CD recordings of performances by professional pianists. The performances of Schumann's piece (Träumerei) are the recordings used by Repp (1992).

4.1.3 Evaluation of Clustering

For each data set, all pairwise distances are computed between the data instances. The distance between a pair of trajectories is calculated as the root summed square (RSS) of the pairwise differences of the pairs of trajectory points (all trajectories in a data set have the same number of points). We are primarily interested in how the organization of trajectories of different performers is, given a particular choice of representation, and the RSS distance measure. Figure 4 shows an example with two imaginary organizations for three occurrences of a musical fragment performed by two different performers (A and B). Note that in the organization on the left, it will be easy to separate the data instances by performer based on the distance measure, whereas in the organization on the right, it will be impossible. To quantify this, we introduce the notion of *inter-intra performer gap* (IIPG). For a given performer k, this quantity is the average distance between k's trajectories and the trajectories of other performers, minus the average distance among k's own trajectories.

More formally, let $\langle k,n\rangle$ be the data instance containing the trajectory the k-th performer playing the n-th occurrence of the musical fragment, and let $d_C(\langle k,n\rangle,\langle l,m\rangle)$ be the distance between data instances $\langle k,n\rangle$ and $\langle l,m\rangle$ in space C. Then, the IIPG of the k-th performer in C is defined as:

$$IIPG_C(k) = \frac{2}{N(N-1)}\sum_{n=2}^{N}\sum_{m=1}^{n-1} -d_C\left(\langle k,n\rangle,\langle k,m\rangle\right) + \frac{1}{K-1}\sum_{l\neq k}^{K} d_C\left(\langle k,n\rangle,\langle l,m\rangle\right) \qquad (7)$$

where K is the number of different performers in the data set, and N the number of occurrences of the musical fragment. Rather than looking at the IIPG of individual performers, we will focus on the suitability of different spaces for performer identification. Therefore, we additionally define the IIPG of a space C as the average IIPG per performer:

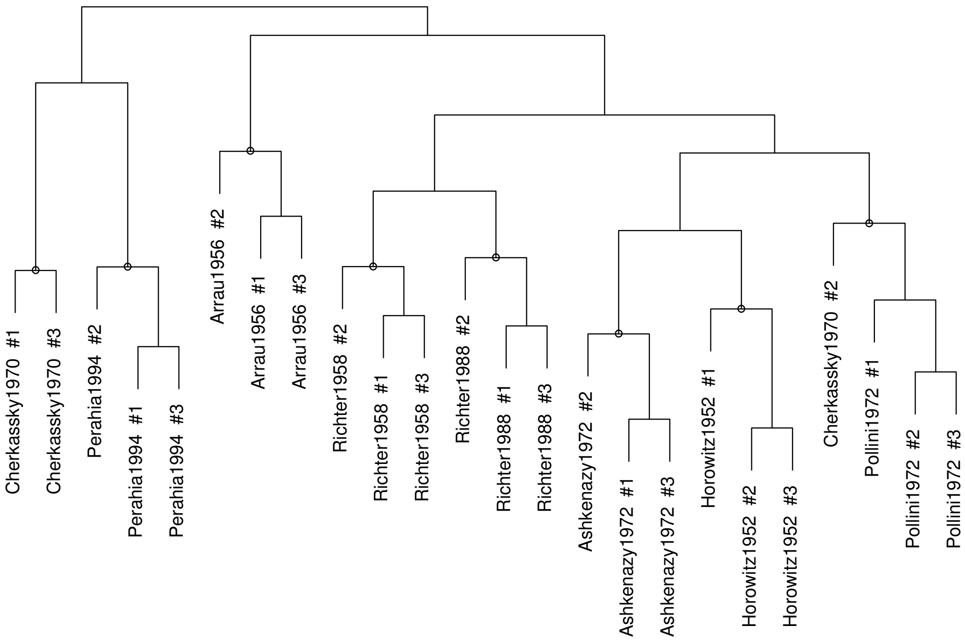


Figure 5: Dendrogram obtained from hierarchically clustering the Chopin, Op. 10, No. 3 (fragment B) performance data set. The clustering is obtained by considering trajectories in the space D^2, the second derivative of tempo. The parameter λ for controlling the roughness penalty is optimized for performer identification ($\lambda = 10^{-4}$); The labels show performer/year and order-of-occurrence; The circled nodes jointly form the clustering with maximal average cluster F-score, given the performer/year labels

$$IIPG_C = \frac{1}{K}\sum_{k=1}^{K} IIPG_C(k) \tag{8}$$

To be able to compare IIPG's across spaces and data sets, we normalize the distances within each space and data set.

Figure 5 shows a *dendrogram*, as obtained by a *complete link* hierarchical clustering of a data set, based on the distances between data instances. In the dendrogram, two sets of instances S_1 and S_2 are joined by a node at a height h precisely when $RSS(s_i, s_j) \leq h$ for any $s_i \in S_1$, and $s_j \in S_2$.

The dendrogram is evaluated by comparing it to the performer partitioning and the order-of-occurrence partitioning, respectively. This is done by computing an F-score for each node in the dendrogram, and selecting the subset of nodes that has the highest average F-score, and partitions the data (that is, each data instance belongs to exactly one node).

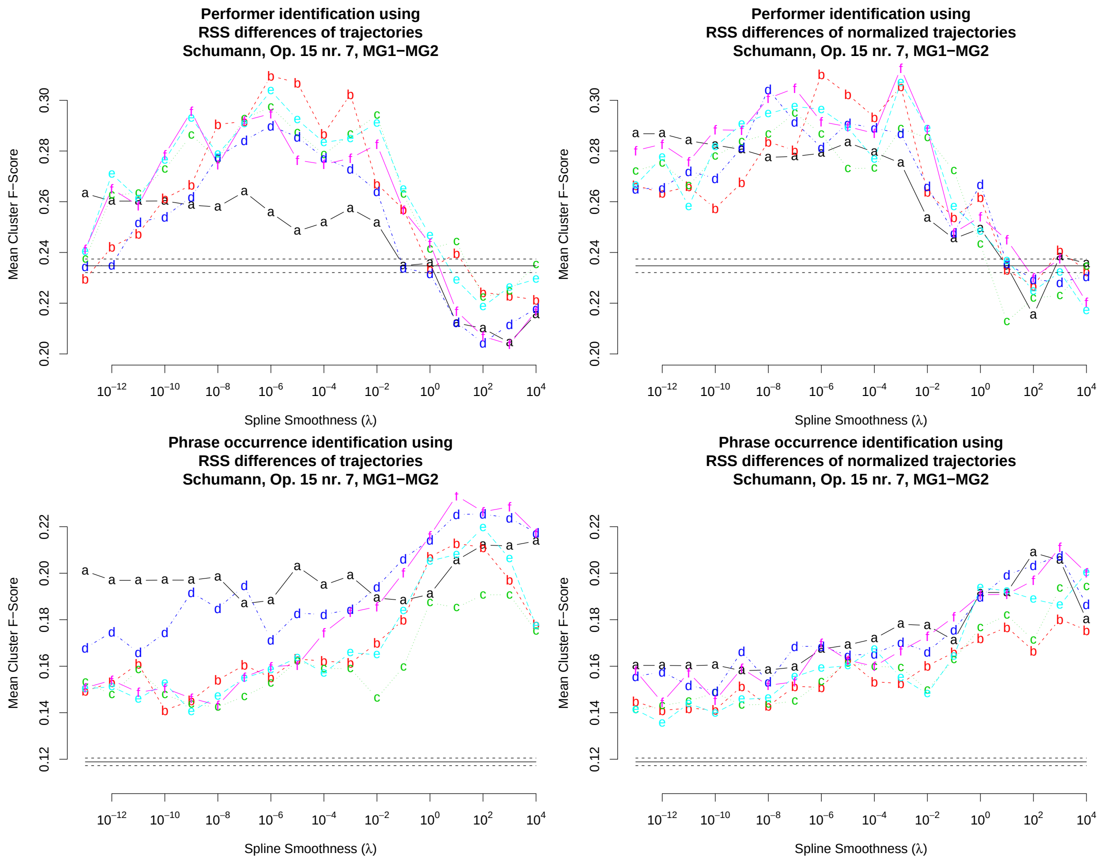
The F-score of a given node C is computed as follows: Let i be the most frequent label in C, and let $1 \leq k_i \leq |C|$ be the number of occurrences of i in C. Furthermore, let K_i be the total number of instances with label i in the data set. Then, we define the F-score of C in terms of its precision and recall:

$$prec(C) = \frac{k_i}{|C|} \tag{9}$$

$$rec(C) = \frac{k_i}{K_i} \tag{10}$$

$$F\text{-}score(C) = \frac{2 \cdot prec(C) \cdot rec(C)}{prec(C) + rec(C)} = \frac{2k_i}{|C| + K_i} \tag{11}$$

These definitions are in accordance with the standard interpretations of precision, recall, and F-score, used in information retrieval, except for the fact that the label i is not fixed in advance, but is chosen to



Performer identification using RSS differences of trajectories Schumann, Op. 15 nr. 7, MG1–MG2

Performer identification using RSS differences of normalized trajectories Schumann, Op. 15 nr. 7, MG1–MG2

Phrase occurrence identification using RSS differences of trajectories Schumann, Op. 15 nr. 7, MG1–MG2

Phrase occurrence identification using RSS differences of normalized trajectories Schumann, Op. 15 nr. 7, MG1–MG2

Figure 6: Mean cluster F-scores as a function of spline smoothness, for different combinations of tempo and derivatives. Legend: **a**: tempo (S); **b**: first derivative of tempo (D^1); **c**: second derivative of tempo (D^2); **d**: (S,D^1); **e**: (D^1,D^2); **f**: (S,D^1,D^2); Horizontal lines indicate estimated baseline F-score (solid) with standard error (dashed); Left column: non-normalized trajectories; Right column: normalized trajectories; Top row: performer identification; Bottom row: phrase-occurrence identification

be the label that for a given C maximizes $prec(C)$. Since we have $K_i = K_j$ for any pair of labels (i, j), the label that maximizes $prec(C)$ also maximizes $rec(C)$, and therefore $F\text{-}score(C)$.

4.2 Results and Discussion

With the definition of the F-score for dendrogram nodes (i.e. subsets of the data set), and the mean F-score per node for partitionings of the data set, we have a means of quantifying how successful a given phase-plane representation is for the performer and order-of-occurrence identification tasks, respectively. This enables us to systematically evaluate different representations.

Figure 6 shows the evaluations of different representations for the Schumann data set. Each plot shows the mean cluster F-score of the optimal partitioning as a function of spline smoothness. Each curve represents a different space (see the caption in figure 6 for a legend). The top row shows the results when the clustering is optimized for the performer identification task, the bottom row shows the results when clustering is optimized for the order-of-occurrence task. Finally, the plots in the left column were generated using non-normalized trajectories, and plots in the right column using normalized trajectories.

The figure presents a few interesting outcomes of the experiment. Most notably, the choice of the degree of smoothing substantially affects the accuracy for both identification tasks. Moreover, the optimal degree of smoothing is lower for performer identification than for order-of-occurrence identification. Whereas performer identification benefits from medium levels of smoothing (λ values in the range 10^{-6}–10^{-2}), order-of-occurrence identification works best with high degrees of smoothing (λ values close to

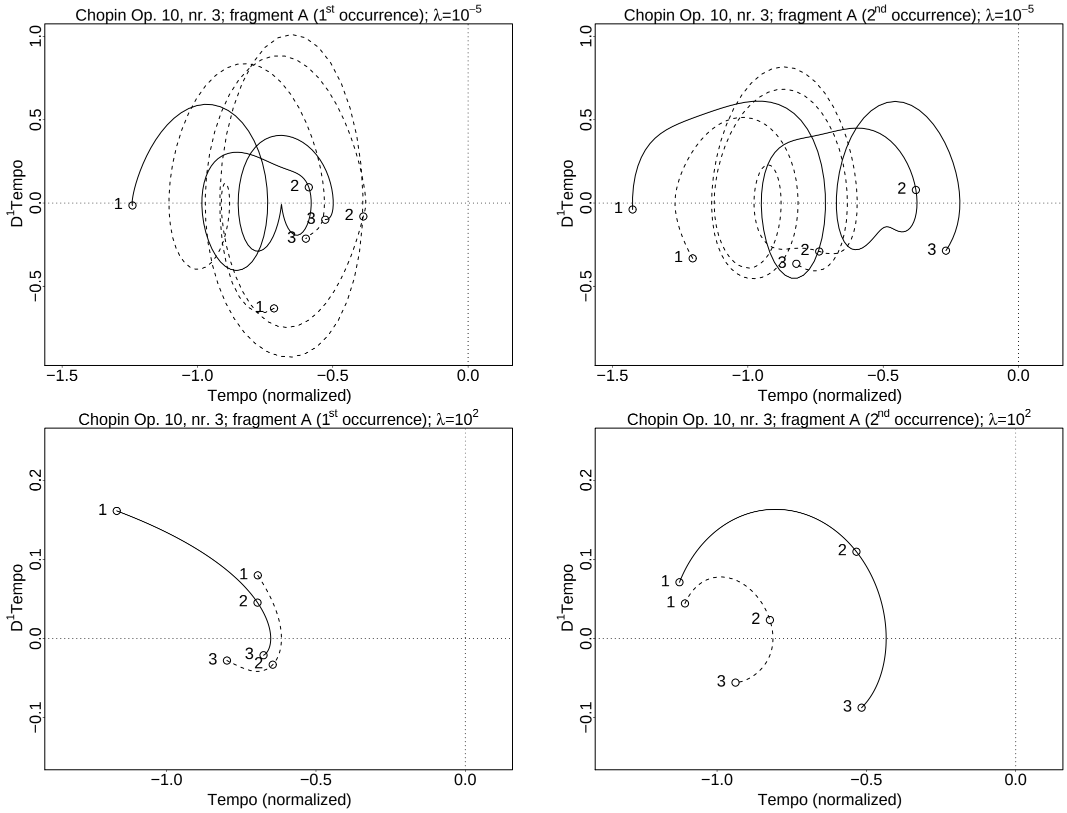


Figure 7: First-order phase-plane trajectories for Ashkenazy 1972 (solid curves), and Richter 1988 (dashed curves) recordings of Chopin, Op. 10, No. 3, fragment A. The numbers along the trajectories indicate the starting of consecutive bars; Left column: first occurrence of fragment; Right column second occurrence; Top: unsmoothed trajectories, bottom: smoothed trajectories

10^4). This difference is confirmed with a t-test on λ values weighted by mean cluster F-score ($t = 3.2176$; $p < .001$).

As an example of this effect at the level of individual phase-plane trajectories, consider the four plots in figure 7. The figure displays the trajectories of two pianists, Ashkenazy (solid curves), and Richter (dashed curves), both playing the first and second instances of the same fragment (left and right columns respectively). The trajectories are plotted in the first-order phase-plane, using low and high levels of smoothing (top and bottom row respectively). Notice how, with little smoothing (top row), each of the two pianists has its own characteristic phase-plane form, that is relatively constant throughout the two occurrences of the same fragment. Richter plays the fragment with more pronounced gestures, while keeping a fixed base tempo, whereas Ashkenazy slightly increases tempo throughout the fragment in three less pronounced gestures. Unsurprisingly, when λ is increased such that any note level detail of the timing profile is eliminated (bottom row), the differences between the two pianists fade. Instead, it seems that the trajectories now show mainly a distinction between the first occurrence (left plot) and the second (right plot).

This result suggests that the aspects of phase-plane trajectories that are performer specific are in the details, whereas the global form of the trajectory is rather determined by the score context. This claim is only tentative however, because it must be borne in mind that the issue of different but not performer-specific *performance strategies*, has not been taken into account in this study. Such strategies may influence either the detailed or global form of trajectories, or both.

Returning to figure 6, a slight effect of normalizing on mean cluster F-score can be observed. For performer identification, normalization of trajectories increases mean cluster F-scores over the whole range of λ. Interestingly, in the case of order-of-occurrence identification, results are better using non-normalized trajectories. Both effects are present when the results of all four data sets are taken together. A paired Wilcoxon signed rank test revealed higher accuracies for normalized trajectories in performer identification ($V = 41587$, $p < .05$), and higher accuracies for non-normalized trajectories in order-of-

					95% conf. interv.		
space A	space B	*IIPG*$_A$	*IIPG*$_B$	*IIPG*$_A$ − *IIPG*$_B$	lower	upper	p
D^1	S	0.843	0.679	0.164	0.026	0.302	0.009
D^2	S	0.849	0.679	0.171	0.033	0.308	0.006
(S, D^1)	S	0.815	0.679	0.136	-0.002	0.274	0.056
(D^1, D^2)	S	0.883	0.679	0.204	0.066	0.342	0.000
(S, D^1, D^2)	S	0.880	0.679	0.201	0.063	0.339	0.000

Table 2: Results of a Tukey HSD test of the intra/inter performer gap (IIPG) for tempo-only representation (S), versus representations that include tempo derivatives; Considered trajectories are normalized; Roughness penalties of considered trajectories range from 10^{-6} to 1

occurrence identification ($V = 58743$, $p < .0001$).

The fact that order-of-occurrence identification is more accurate with non-normalized trajectories indicates that absolute tempo is a relevant factor for this task. This is plausible, since the higher level musical structures of which the repeated fragments form part, may impose an evolution of the tempo curve that differentiates tempo between its constituent structures.

Lastly, we consider the effect of the choice of space (the combination of tempo and its derivatives to represent trajectories). The plots in figure 6 show that for the Schumann data set, the different spaces (curves **a** to **f**) follow roughly the same trends. A notable exception is the tempo-only space (curve **a**) in the non-normalized condition for performer identification, shown in the top left plot. For moderate degrees of smoothing, which tend to be optimal for performer identification, the tempo-only representation yields substantially lower F-scores. This effect is also observed in the other data sets. The top right plot in figure 6 suggests that normalizing trajectories partly alleviates this, since curve **a** is slightly raised.

Upon inspection of the joint data sets however, it turned out that even in the normalized condition, the spaces that include derivatives on average lead to higher F-scores. This is confirmed by an analysis of variance, which showed an effect of space on the size of IIPG[6]. we use Tukey's Honest Significant Difference test as a *multiple comparison procedure* to find the differences in IIPG as an effect of space. For all spaces involving derivatives of tempo except (S,D^1), this test shows a significant ($\alpha = .05$) increase in IIPG with respect to the tempo-only space. The results of the test are summarized in table 2. Between spaces involving tempo derivatives no significant differences were found (these comparisons have been omitted in table 2).

5 Conclusions and future work

Phase-plane representations of expressive timing of music performances include tempo derivatives in addition to tempo itself. Such representations highlight the dynamic aspects of expressive timing. A consequence of this is that functions that look similar in time series plots, such as a parabola and a semiperiod of a simple harmonic oscillation, have qualitatively different phase-plane trajectories, since their derivatives are different. As such, phase-plane trajectories may suggest a particular class of functions that could fit a particular tempo curve. This can be a benefit in the modeling of expressive timing, as in Repp (1992); Todd (1985); Friberg and Sundberg (1999).

In addition, different representations of phase-plane trajectories were compared experimentally. Results indicate that phase-plane trajectories are most performer specific when moderate degrees of smoothing are used, and normalization is applied. Highly smoothed and non-normalized trajectories on the other hand, are more indicative of the order of occurrence of the fragment in the musical piece. The experiment also showed that considering tempo derivatives in addition to tempo in general facilitates the distinction between performers based on their performances.

In conclusion, we argue that phase-plane representations of expressive timing are an interesting alternative to conventional time-series plots of expressive timing information. Not only do they allow for intuitive visualizations of expressive gestures, but they also seem relevant for the characterization of the timing of individual performers.

[6] The analysis was performed on the IIPG measure rather than F-score since the F-score data, by repeated summarizing of intermediate results, is too sparse to make a reliable claim

In this study, phase-plane trajectories were compared using simple root sum squared differences of trajectory coordinates. A more advanced method of comparison would do more justice to the *gestalt* aspect of trajectories, for example through a qualitative characterization of trajectories, ideally in terms of a small set of prototypical forms. Such an *alphabet* of phase-plane trajectory fragments may obtained in a data-driven way, as in Widmer et al. (2003).

Acknowledgments

This work is funded by the Austrian National Research Fund (FWF) under project number P19349-N15. We are grateful to Bruno Repp for providing the expressive timing data for the performances of Schumann's Träumerei.

References

Boon, J. and Decroly, O. (1995). Dynamical systems theory for music dynamics. *Chaos*, 5(3):501–508.

Canazza, S., De Poli, G., Rodá, A., and Vidolin, A. (2003). An abstract control space for communication of sensory expressive intentions in music performance. *Journal of New Music Research*, 32(3):281–294.

Clarke, E. F. (1988). Generative principles in music. In Sloboda, J., editor, *Generative Processes in Music: The Psychology of Performance, Improvisation, and Composition*. Oxford University Press.

Desain, P. and Honing, H. (1993). Tempo curves considered harmful. In "Time in contemporary musical thought" J. D. Kramer (ed.), Contemporary Music Review. 7(2).

Desain, P. and Honing, H. (1996). Physical motion as a metaphor for timing in music: the final ritard. In *Proceedings of the 1996 International Computer Music Conference*, pages 458–460, San Francisco. ICMA.

Dixon, S., Goebl, W., and Cambouropoulos, E. (2006). Perceptual smoothness of tempo in expressively performed music. *Music Perception*, 23(3):195–214.

Friberg, A. and Sundberg, J. (1999). Does music performance allude to locomotion? a model of final ritardandi derived from measurements of stopping runners. *Journal of the Acoustical Society of America*, 105(3):1469–1484.

Gabrielsson, A. (2003). Music performance research at the millennium. *The Psychology of Music*, 31(3):221–272.

Goodall, C. (1991). Procrustes methods in the statistical analysis of shape. *Journal of the Royal Statistical Society. Series B (Methodological)*, 53(2):285–339.

Grachten, M., Goebl, W., Flossmann, S., and Widmer, G. (2008). Intuitive visualization of gestures in expressive timing: A case study on the final ritard. In *Proceedings of the International Conference on Music Perception and Cognition*.

Honing, H. (2004). When a good fit is not good enough: a case study on the final ritard. In *Proceedings of the ICMPC*, pages 510–513.

Honing, H. (2005). Is there a perception-based alternative to kinematic models of tempo rubato? *Music Perception*, 23(1):79–85.

Juslin, P. and Sloboda, J., editors (2001). *Music and Emotion: Theory and Research*. Oxford University Press.

Langner, J. and Goebl, W. (2003). Visualizing expressive performance in tempo-loudness space. *Computer Music Journal*, 27(4):69–83.

Métois, E. (1996). *Musical Sound Information: Musical gestures and embedding synthesis*. PhD thesis, Massachusetts Institute of Technology.

Palmer, C. (1997). Music performance. *Annual Review of Psychology*, 48:115–138.

Ramsay, J. and Silverman, B. (2005). *Functional Data Analysis (2nd ed).* Springer.

Repp, B. H. (1992). Diversity and commonality in music performance - An analysis of timing microstructure in Schumann's "Träumerei". *Journal of the Acoustical Society of America*, 92(5):2546–2568.

Repp, B. H. (1998). A microcosm of musical expression: I. Quantitative analysis of pianists' timing in the initial measures of Chopin's etude in E major. *Journal of the Acoustical Society of America*, 104(2):1085–1100.

Schoner, B. (1997). State reconstruction for determining predictability in driven nonlinear acoustical systems. Master's thesis, Rheinisch-Westphalische Technische Hochschule Aachen.

Sundberg, J., Friberg, A., and Frydén, L. (1991). Common secrets of musicians and listeners: an analysis-by-synthesis study of musical performance. In Howell, P., West, R., and Cross, I., editors, *Representing Musical Structure*, Cognitive Science series, chapter 5. Academic Press Ltd.

Todd, N. (1992). The dynamics of dynamics: A model of musical expression. *Journal of the Acoustical Society of America*, 91:3540–3550.

Todd, N. P. (1985). A model of expressive timing in tonal music. *Music Perception*, 3:33–58.

Truslit, A. (1938). *Gestaltung und Bewegung in der Musik.* Chr. Friedrich Vieweg, Berlin-Lichterfelde.

Widmer, G. (2003). Discovering simple rules in complex data: A meta-learning algorithm and some surprising musical discoveries. *Artificial Intelligence*, 146(2):129–148.

Widmer, G., Dixon, S., Goebl, W., Pampalk, E., and Tobudic, A. (2003). In search of the Horowitz factor. *AI Magazine*, 24(3):111–130.