Phase-plane visualizations of gestural structure in expressive timing

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

[1]Department of Computational Perception, Johannes Kepler University, Linz, Austria
[2]Sequence Production Lab, McGill University, Montreal, Canada

Background in music performance research 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 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.

Background in computing, mathematics, and statistics A common situation in the natural sciences is the limited observability of complex systems, e.g. the weather. The observed variables typically exhibit little regularity due to the influence of unknown and interacting underlying factors. A helpful method in the study of such systems is the graphical presentation of typical trajectories through its phase-space, or state-space, the space of all the possible states the system can be in. The system state is represented by state variables. The complexity of the system is intrinsically related to the number of state variables that are necessary to completely describe its behaviour. The phase-plane is a two-dimensional plot of two state variables against each other. Such plots can give insights in the temporal behaviour of the system, and sometimes allow for a qualitative description of the behaviour.

Aims Tempo measured as the reciprocal of the interbeat-interval defines tempo at discrete points in time. In between the musical events that imply the beat there is no way of inferring the tempo and therefore a tempo-curve as a continuous function of score time probably does not correspond to any percept in the listener (Desain and Honing, 1993). However, in order for the rhythmical structure of a piece to be perceptible, tempo must satisfy certain smoothness constraints. This coherence between consecutive tempo perceptions is expressed by the representation of tempo as a continuous function of time.

One of the goals of the study of music performance is to develop models of expressive timing. A particularly interesting set of models are those that are grounded in the metaphor of physical motion (e.g. Todd (1989)). Although such models have limitations (Honing, 2004), we do believe that the interpretation of expressive timing as the result of an underlying continuous process can prove fruitful. Consequently, we intend to show that visualization and analysis methods common in dynamical systems theory can be helpful in developing models of expressive timing.

Main Contribution In this paper we illustrate the use of phase-plane representations for investigating expressive timing in music. We survey of the relation between the phase-plane plots and the more common time series plots, and describe the construction of the phase-plane representation from the time series using a functional approximation of the data. Then we illustrate its use by showing discussing the phase-plane representations of performance fragments from Schumann's Träumerei by two different pianists.

Implications Compared to time series plots, phase-plane representations emphasize the dynamical aspects of expressive timing data. This makes it arguably a suitable tool for studying and modeling expressive gestures in timing. As shown by the examples discussed in the paper, it may be easier in phase-plane representations to determine which model fits best to a particular tempo-curve.

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 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 (Clarke, 1988), like metrical, phrase, and voice structure. Furthermore, timing and other temporal aspects such as global tempo and articulation play a role in the communication of semantic content, including emotional (Gabrielsson and Juslin, 1996), and sensorial (e.g. dark, light, heavy, soft) (Canazza et al., 1997) information. In addition to the establishment of 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 behaviour of some complex system based on limited observation. Examples of this can be found in population biology, and meteorology. In cognitive science and developmental psychology, such a dynamical systems perspective has been proposed as a viable alternative to symbolic and connectionist approaches (Beer, 2000). This approach aims at developing models that describe how the state of the system changes over time. Rather than predicting individual state trajectories, the aim is to gain insight in the qualitative structure of the state-space that results from the influence and interaction of possibly unknown factors.

In the case of expressive music performance we are not completely ignorant about these factors. As described above, past research has revealed valuable insights about how factors like musical structure and intended mood influence expressive timing. Also, there has been a long-standing metaphor of (the performance of) music as a form of motion (Truslit, 1938; Friberg and Sundberg, 1999). Several approaches exist that take kinetic models as a basis for modeling expressive tempo, most notably Todd (1985). Still, we are far from giving a full account of expressive timing.

A dynamical view of expressive timing

A helpful method in the study of a dynamical system is the graphical presentation of typical trajectories through its phase-space, or state-space, the space constituted by all of the possible states the system can be in. The system state is represented by state variables. The complexity of the system is intrinsically related to the number of state variables that are necessary to completely describe its behaviour. An example of a relatively simple system in mechanics is the simple pendulum. Its behaviour is completely determined by its position and velocity (the derivative of position with respect to time). In that case the phase-plane showing velocity versus position provides a complete overview of the behaviour of the system. By determining for example the attractors in the state-space, the phase-plane representation allows for a qualitative description of the behaviour of the system.

Of course expressive timing in music performance is vastly more complex than the behaviour of pendula or any other didactic example from mechanics, and it is highly unlikely that a complete and adequate description of expressive timing by a system of differential equations is feasible. However, this does not imply that a dynamical perspective cannot help to give a better understanding of the phenomenon. An example of the application of phase-plane visualization in music performance research is the *performance worm* (Langner and Goebl, 2003), which displays the performance as a trajectory through the loudness-tempo space. In the work presented here, we will consider only timing information. The phase-planes we investigate are constructed by plotting tempo against its derivative (the first-order phase-plane), and the first and second derivatives of tempo against each other (the second-order phase-plane).

In the rest of the paper we will present and discuss the phase-plane representation for studying expressive timing in music. We discuss the main differences to conventional time series plots of tempo, and illustrate the representation using some archetypal curve forms. Then we explain how phase-plane trajectories are computed from measured tempo curves. Finally, we demonstrate the phase-plane visualization using performances of Schumann's Träumerei.

Phase-plane plots versus time series plots

An obvious question that comes to mind when considering such phase-plane plots 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 in 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.

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 $x'(t)$ vs $x(t)$ plane, that is, the first derivative of $x(t)$ against $x(t)$ itself. The last column shows the second-order phase-planes, formed by $x''(t)$ vs $x'(t)$. 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.[i] 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 $x'(t)$ 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 $x'(t) = 2t$ 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 $x''(t) = 2$. 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).

From time series to phase-plane trajectories

The concept of a tempo curve, even when ubiquitous in expressive music performance research, is not straight-forward. Given that tempo can be loosely defined as the rate at which

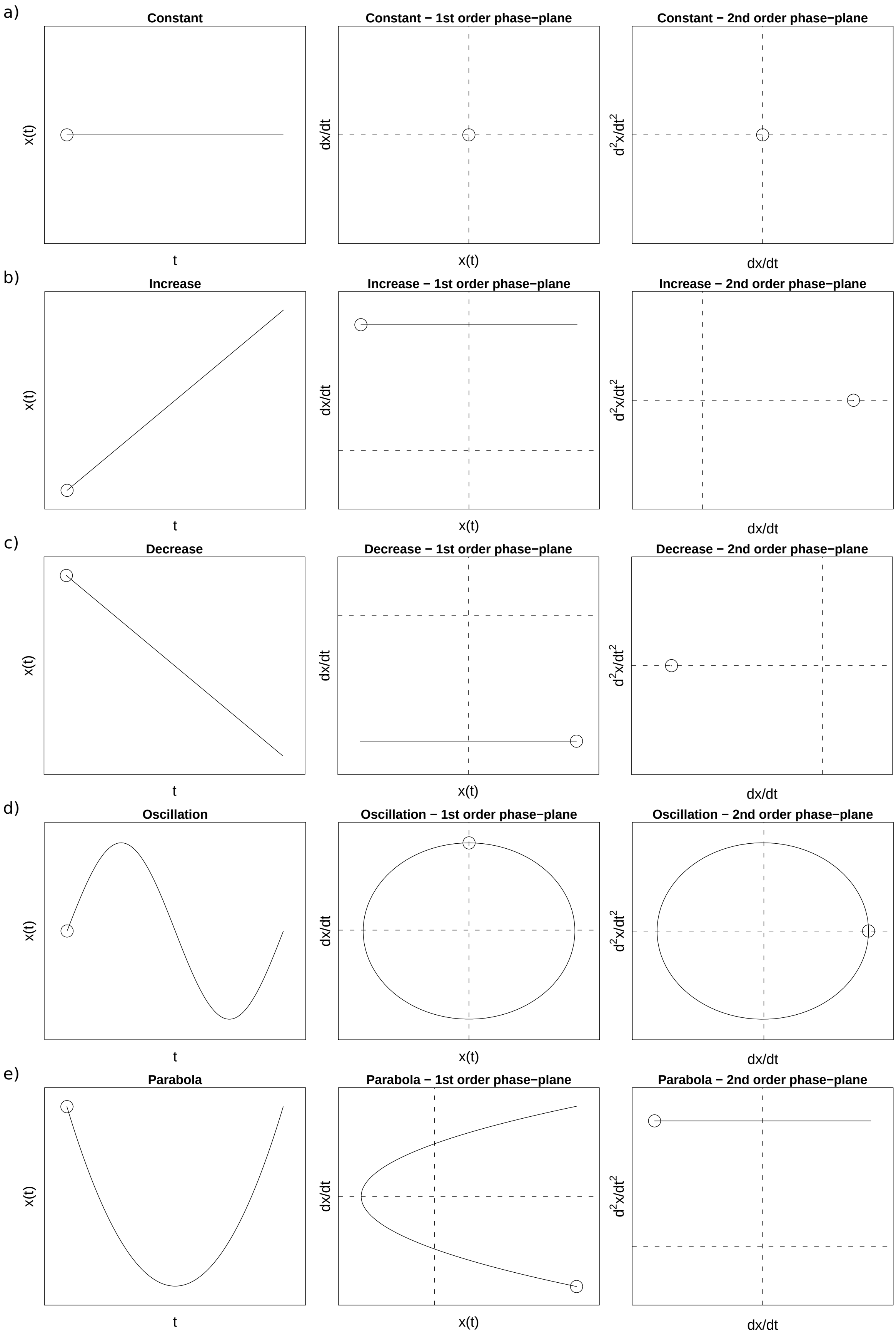

Figure 1: Examples of five basic curve types (first column), and their first and second order phase-plane trajectories (second and third columns respectively); Horizontal and vertical dashed lines represent x and y axes respectively; Circles indicate the beginning of the curves/trajectories; Units are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

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 justifiable to interpolate the time series of tempo values to obtain a tempo curve. However, the curve is not meaningless in the way a curve drawn between the outcomes of rolling a dice would be meaningless. In order for the rhythmical structure of a piece to be perceptible, tempo must satisfy certain smoothness constraints Honing (2004). 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 statistics, 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 behaviour 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 data, which is formally represented as:

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

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:

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

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:

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

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:

$$PENSSE = SSE^2 + \lambda PEN$$

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:

$$PEN = \int [D^2 x(s)]^2 ds$$

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 (1997). 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:

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

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 $D[S(t)]$ and $D^2[S(t)]$.

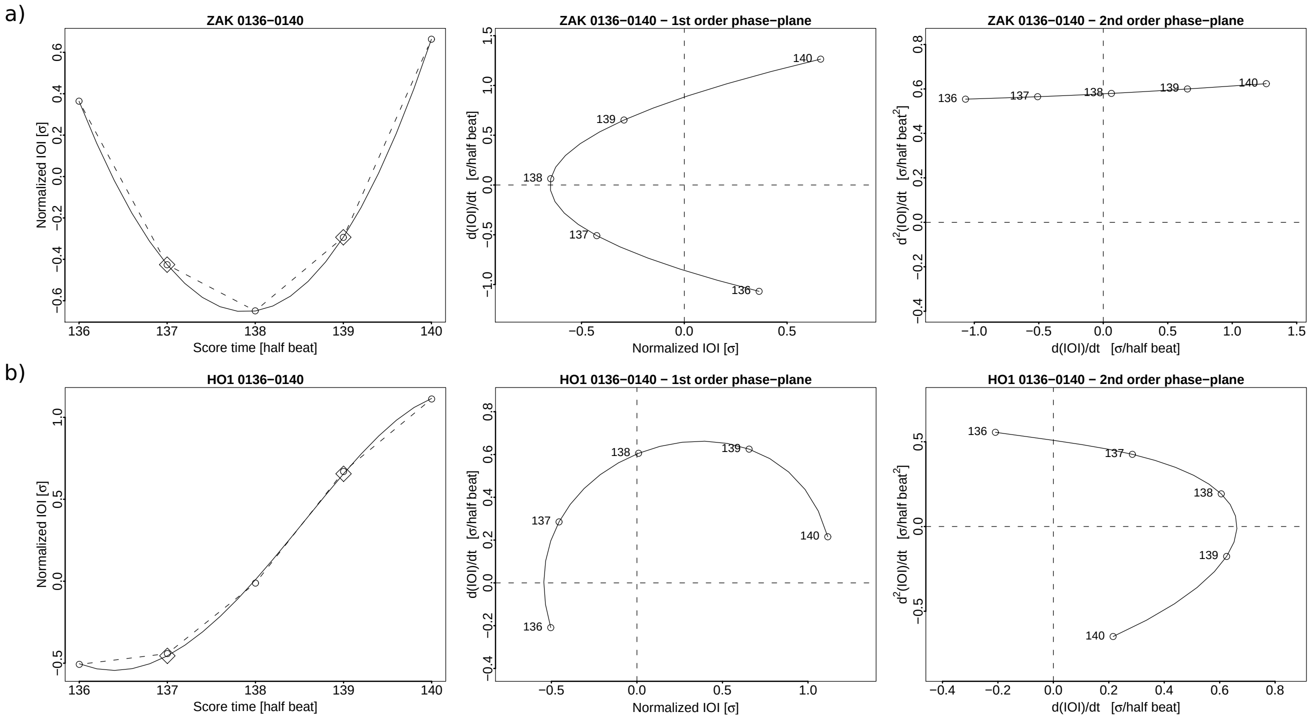

Figure 2: Fitted IOI curves and corresponding phase-plane trajectories for two exemplary performances of a melodic gesture (MG2, fourth instance, half beats 136–140) from Schumann's Träumerei, by Zak (1960) (a), and Horowitz (1947) (b) respectively; In the first column, the circles connected by dashed 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

Phase-planes of expressive gestures in Schumann's Träumerei

We will illustrate the phase-plane visualization for two performances of a motif (or melodic gesture) from Schumann's Träumerei. We will do this in relation to an earlier study of expressive timing in that piece Repp (1992).[ii] Repp describes the results of an extensive study of expressive timing of 28 performances of this piece by renown pianists. A part of this study is a detailed investigation of the timing of notes in particular a motif, or *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 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.[iii] 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 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 show relatively much ‘space in between the 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 what the trajectories looks like. 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.

Conclusions and future work

We have demonstrated the representation of tempo-curves as trajectories in the phase-plane, the 2-dimensional plane that shows a function against its derivative. This representation highlights 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 can give a first indication of the class of models that could fit a particular tempo curve, which is a benefit in the modeling of expressive timing, as in Repp (1992); Todd (1985); Friberg and Sundberg (1999).

Our investigation into the phase-plane representation of tempo-curves has been preliminary and informal until now. We intend to use this representation in more elaborate and systematic experimentation of expressive timing. In particular, clustering a set of performances of the same musical fragment based on phase-plane representations looks a promising approach to decide whether the performances can be adequately represented by a single model. In combination with pattern-finding, the phase-plane trajectories of complete pieces could be analyzed to find an ‘idiom’ of expressive gestures, or alternatively, to reconstruct the phrase structure of the performed piece, as in Grachten and Widmer (2007). Furthermore, an interesting topic from a phase-plane perspective is the *final ritard*, a well-studied expressive gesture for which several models have been proposed Friberg and Sundberg (1999); Honing (2004).

Acknowledgments This work is funded by the Austrian National Science 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

Beer, R. D. (2000). Dynamical approaches to cognitive science. *Trends in Cognitive Sciences*, 4(3).

Canazza, S., De Poli, G., and Vidolin, A. (1997). Perceptual analysis of musical expressive intention in a clarinet performance. In Leman, M., editor, *Music, Gestalt, and Computing: studies in cognitive and systematic musicology*, number 1317 in Lecture Notes in Artificial Intelligence, pages 441–450. Springer.

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).

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.

Gabrielsson, A. and Juslin, P. N. (1996). Emotional expression in music performance: Between the performer's intention and the listener's experience. *Psychology of Music*, 24(1).

Grachten, M. and Widmer, G. (2007). Towards phrase structure reconstruction from expressive performance data. In *Proceedings of the International Conference on Music Communication Science (ICOMCS)*, Sydney, Australia.

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.

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

Ramsay, J. and Silverman, B. (1997). *Functional Data Analysis*. Springer-Verlag.

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.

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. (1989). A computational model of rubato. *Contemporary Music Review*, 3 (1).

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.

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

[ii] The data used here originates from that study

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