Intuitive visualization of gestures in expressive timing: A case study on the final ritard

Maarten Grachten[1,i], Werner Goebl[2,ii], Sebastian Flossmann[1,i], and Gerhard Widmer[1,3,i]

[1] *Department of Computational Perception, Johannes Kepler University, Linz, Austria*
[i] {maarten.grachten,sebastian.flossmann,gerhard.widmer}@jku.at
[2] *Sequence Production Lab, McGill University, Montreal, Canada*
[ii] werner.goebl@mail.mcgill.ca
[3] *Austrian Research Institute for Artificial Intelligence, Vienna, Austria*

ABSTRACT

Expressive timing is vital for the aesthetic quality that makes us appreciate performed music. It is a largely tacit skill that musicians acquire by practice. A long-standing intuition is that expressive timing is closely related to the concept of motion. This view leads naturally to the adoption of a dynamical systems approach to the study of expressive timing. A well-known visualization technique from dynamical systems theory is the phase-plane representation. The application of this technique, that highlights the dynamic aspects of the data, is demonstrated in a case study on the final ritard in performances of Schumann's Träumerei. We argue that expressive gestures are visible in a clear and intuitive manner in the phase-plane representations. Another striking aspect of the phase-plane trajectories is their suggestion of human gestural motion.

I. INTRODUCTION AND RELATED WORK

Expressive timing is vital for the aesthetic quality that makes us appreciate performed music. It is a largely tacit skill that musicians acquire by practice. Empirical research concerning expressive timing dates back to the beginning of the 20th century and has boomed over the past decennia. Several distinct approaches have been taken to model and more in general to gain understanding of the phenomenon. Two main strategies for modeling expressive timing can be characterized as knowledge-driven and data-driven respectively. A good example of the former is the 'analysis-by-synthesis' method used by Sundberg et al. (1991). The latter approach is exemplified by the application of machine learning algorithms, such as by Widmer (2002).

A long-standing intuition is that music is closely related to the concept of motion. This idea has been elaborated by Truslit (1938) (see also Repp (1993)), who claimed that music is essentially the transmission of motion information from the musician to the listener. This analogy between music and motion has inspired several kinetic models of expressive timing (and dynamics) (Todd, 1995; Friberg and Sundberg, 1999). These models are derived from the laws of physical motion from classical mechanics. In particular musical circumstances these models provide quite accurate descriptions of expressive timing observed in actual performances. However, as pointed out by Repp (1992); Honing (2006), these models are limited in the sense that they model expressive timing as a function of score time, and neglect any structural aspects of the music.

An important implication of the analogy between motion and music, in particular expressive timing, is the view that the timing of notes in music performances is the result of some underlying process, and that this process, however complex it may be, is governed or constrained by certain principles. This perspective renders the problem of understanding expressive timing in music as a special case of a general branch of problems that occur throughout empirical science, in which we want to learn something about a dynamical system through the limited observation of its behaviour. A dynamical system is formally defined by a collection of states (the state-space, or phase-space) together with an evolution function that maps any state to the next state in time. Usually, the evolution function is assumed to be smooth. The observations of such a system typically come in the form of a series of measurements of some quantity through time. From the resulting time-series we wish to infer qualitative information about the underlying system. In particular, we want to reconstruct the behaviour of the system as a trajectory through its state-space (or phase-space). Even if the dimensions of the state space are unknown, as is often the case in practice, principally the state-space can still be reconstructed from the data.[i] This is achieved by representing the time-series as a trajectory through an m-dimensional space that is defined by taking m values of the time-series with a fixed time delay τ (the *lag*). This technique is called *delay reconstruction*, and is a common tool in nonlinear time-series analysis (Kantz and Schreiber, 1997). Such a dynamical approach has been applied to various aspects of music, including acoustic modeling (Schoner, 1997), and gesture-based virtual instrument control (Métois, 1996).

Of course the dynamical system we assume to underlie the expressive timing of musical performances is highly complex, especially relative to the number of observations we have in the form of time-series of tempo or IOI values describing the expressive timing of performances. As a result, state-space reconstruction by techniques such as delay reconstruction is likely to be unfeasible. Nevertheless, visualization techniques related to state-space representations are useful for exploring observations from dynamical systems. In particular, *phase-plane* visualization can facilitate observations of the 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 behaviour 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 the context of 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. 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. We intend to demonstrate the suitability of these phase-plane representations as a visualization technique for expressive timing. We argue that expressive gestures are visible in a clear and intuitive manner in the phase-plane representations. Moreover, a striking aspect of the phase-plane trajectories is their suggestion of human gestural motion. In particular cyclic phase-plane trajectories corresponding to periodic patterns in timing data illustrate this.

In section II, we discuss the phase-plane representation of expressive timing as an alternative to the more conventional times-series plots. We show the phase-planes of some archetypal curve segments to demonstrate the relation between the two. Then, we introduce the method of computing the phase-planes representations from a series of tempo or IOI values. This involves a functional approximation of the data using splines. In section III, we discuss the final ritards of three performances of Schumann's Träumerei (Opus 15, Nr. 7) using phase-planes. We end the paper with some conclusions and future work in section IV.

II. 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.

A. 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 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.[ii] 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

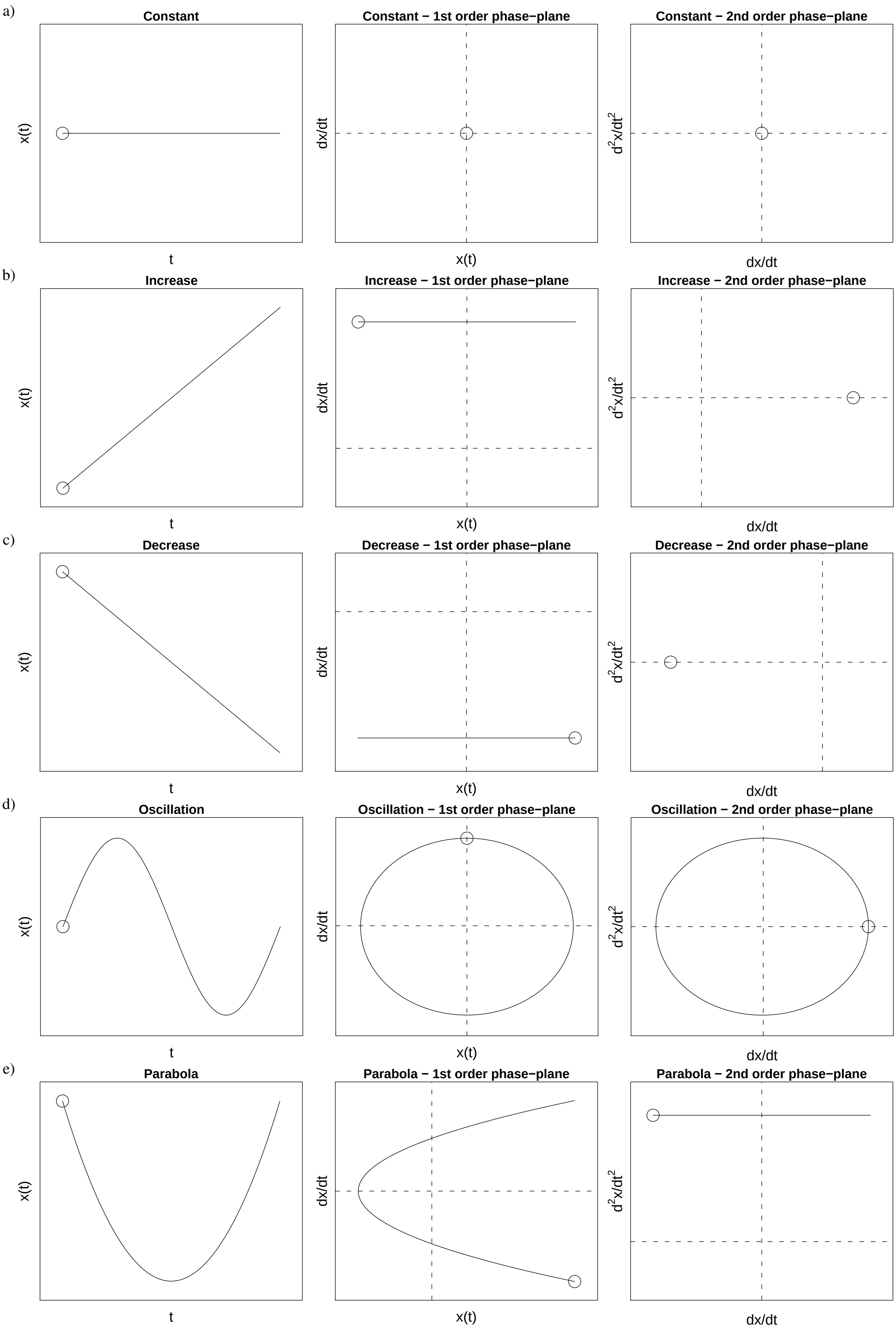

Figure 1: Examples of five basic curve types (first column), and their first and second order phase-plane trajectories (second and third columns respectively); Horizontal and vertical dashed lines represent x and y axes respectively; Circles indicate the beginning of the curves/trajectories; Units are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

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).

B. 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 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 of the spline S, $D[S(t)]$ and $D^2[S(t)]$.

III. CASE STUDY

In this case study we illustrate the phase-plane visualization for the final ritards of three performances of Schumann's Träumerei. The relevant score fragment is shown in figure 2.

The score fragment exists of four motifs, or melodic gestures. Some motif boundaries are accentuated in the score by commas, as a hint for performance. Even if the ritardando indication is located at the start of the first motif, this motif in most measured performances exhibits a short accelerando. This is arguably a preparation for the ritardando itself, which in most cases starts effectively at the second motif (starting at IOI number 246). For this reason, we will be looking at the last three motifs (the IOI range 246–255). IOI's correspond to half beats, implying 8 IOI values per measure.[iii]

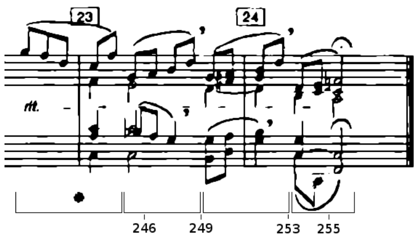

Figure 2: Last two measures of Schumann's Träumerei (Op 15. nr. 7); The brackets below the score (annotated by authors) indicate the grouping of notes into melodic gestures; The numbers at the bottom (annotated by authors) indicate the IOI number; (figure adapted from Repp (1992))

Figure 3 shows the IOI curves and corresponding phase-plane trajectories of the performances of Horowitz (1982) (figure 3a), Brendel (1980) (figure 3b), and Argerich (1984) (figure 3c), for the IOI range 245–255. The first column contains the measured IOI values (normalized across the complete performance) as circles connected by dashed lines, together with the fitted splines, as solid curves. The positions of the breakpoints in the spline are indicated by diamonds.

The fitted splines were computed as described in section II. A limitation of the smoothing using the $PENSSE$ criterion is that when the roughness penalty λ is low enough to allow for a good fit of relatively sudden fluctuations in the data, the splines tend to also exhibit large fluctuations *in between* data points. To solve this issue, an extra data point was inserted in between every consecutive pair of data points, by linear interpolation (these points are not shown in the plot). The splines were subsequently fitted to the linearly interpolated time-series. The number of breakpoints and their positions were determined using a simple heuristic based on the angle of the linearly interpolated curve.

The second column in figure 3 shows the first-order phase-plane trajectories corresponding to the fitted splines. The horizontal axis shows normalized IOI values, and the vertical axis shows the first derivative of IOI values. Analogously, the third column shows the second-order phase-planes, with the first derivative of IOI values horizontally, and the second derivative vertically. Thus, for each row of plots it holds that the vertical dimension in one plot is identical to the horizontal dimension in the next plot to the right. The horizontal and vertical dashed lines indicate the origin in the phase-planes.

The presence of the ritard is easily observed in the time-series plots, from the increasing trend in the IOI values. In the first-order phase-plane, this trend corresponds to a leftward movement of the trajectory (independent of the vertical movement). In addition to the ritard effect, it is clear that the grouping of the score into three motifs is somehow reflected in the IOI curves: In accordance with common knowledge of how the temporal grouping structure of the score is communicated (Palmer, 1997), peaks in the time-series plots can be observed at, or close to the motif boundaries (at IOI numbers 249 and 253). Argerich prolongs IOI 252, the IOI *before* the IOI that spans the last motif boundary, and shortens IOI 253. This may indicate a focus on the lower staff, for which the penultimate motif effectively ends at IOI 252.

Note that whereas the expressive gestures that accentuate the motivic structure are apparent from the time-series plots, they correspond more clearly to individual visual entities in the first-order phase-plane trajectories. The gestures, that consist of a lengthening of IOI's, possibly preceded by a shortening of IOI's, appear as partial circular segments in the first-order phase-plane trajectory. The parts of the circular trajectory above and below the horizontal axis correspond to the lengthening and shortening of IOI's, respectively. In each of the three plots in the second column of figure 3, the trajectories consist of three such forms of increasing size. The increase in size indicates that the expressive gestures become increasingly pronounced for the subsequent motifs. This is effect is observed consistently in the three performances. The major difference between the performances is in the shortening of IOI's as part of the gesture, corresponding to the lower half of the circular forms. In Horowitz's performance, this effect is virtually absent, whereas in Argerich's performance the gestures appear as full circles, indicating a strong IOI shortening in addition to the IOI lengthening. Brendel's performance is intermediate in this respect, showing small downward loops as part of the gestures. Note also that the performers are very consistent the shaping of the gestures for the subsequent motifs.

In the second-order phase-planes, shown in third column of figure 3, the trajectories do not move rightwards, but remain more centered at the origin. Only the last two IOI values significantly depart from the origin.[iv] This is because the second-order phase-plane abstracts from IOI durations, which implies that similar gestures, even at different tempos, map to the same region in the phase-plane. In this phase-plane, the 'intensity' of the gesture is expressed in the scale of the corresponding trajectory segment. For example, in the second-order phase-plane of Argerich's performance, the outward spiraling trajectory indicates a repetition of the same gesture but with increasing 'intensity', or amplitude.

It is worth mentioning the similarities that the phase-plane trajectories bear to the form of the physical gestures that humans intuitively use to make to mimic the dynamics of processes such as music or speech. The circular movements that are characteristic for the phase-plane trajectories of expressive timing in performances, are also typical for arm movements in such gestures. Moreover, it seems natural to express more dramatic changes with eccentric movements, as is the case in the phase-planes. Bearing in mind Truslit's hypothesis (Truslit,

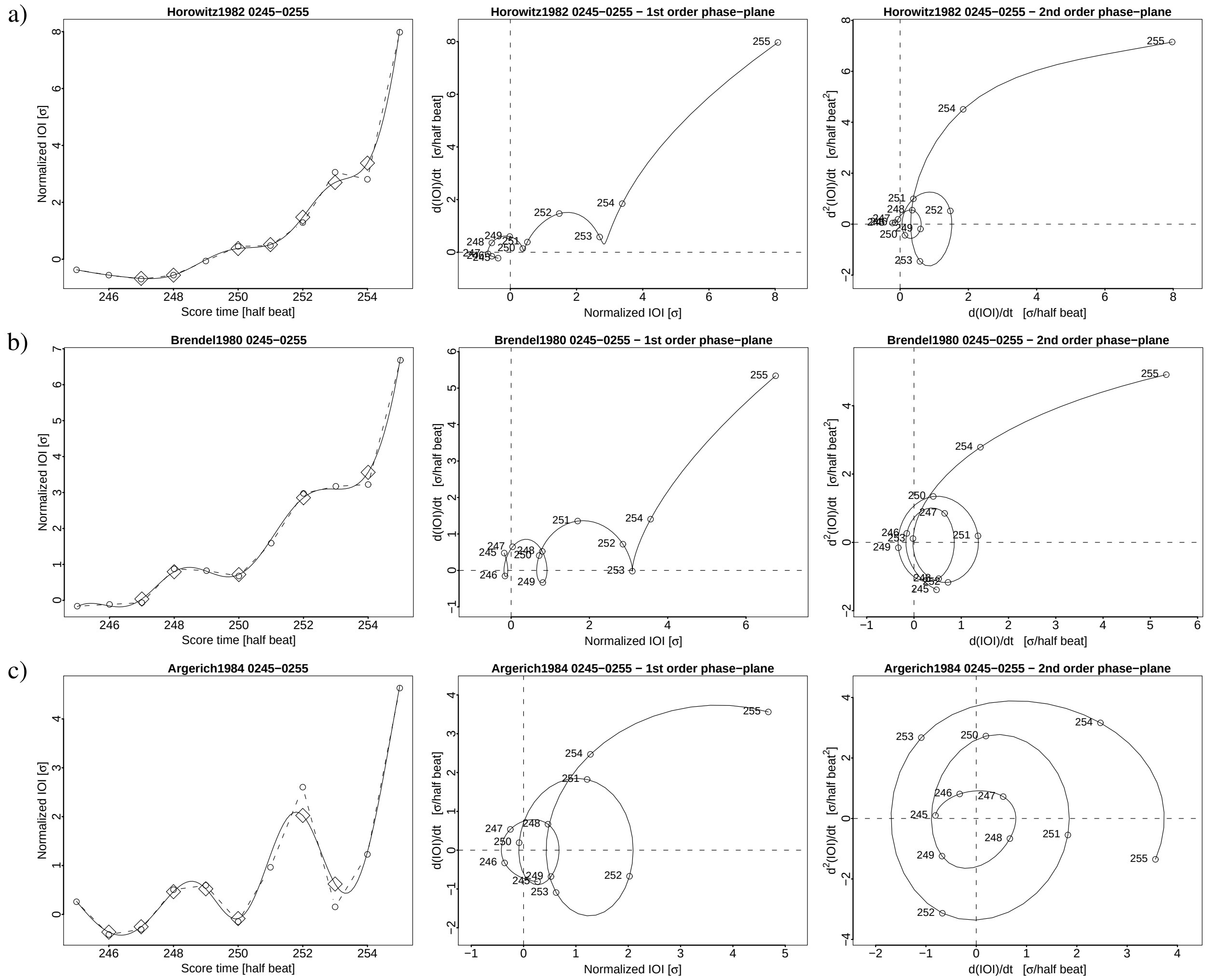

Figure 3: Fitted IOI curves and corresponding phase-plane trajectories for three performances at IOI range 245-255 from Schumann's Träumerei: a) Horowitz (1982); b) Brendel (1980); c) Argerich (1984); In the first column, the circles connected by dashed lines are the measured IOI values (normalized); The solid line is the fitted spline; Diamonds indicate the breakpoints of the spline; The phase-plane trajectories are annotated with IOI numbers; The axis units are shown in square brackets in the axis labels; σ denotes the standard deviation of the normalized IOI values

1938) that the essence of music is to transmit forms of natural motion, it can be argued that the phase-plane method applied here appears to transform expressive timing data into a representation that is closer to such forms of motion.

IV. CONCLUSIONS AND FUTURE WORK

We have demonstrated the use of phase-plane representations for the visualization of expressive timing in musical performances. Such visualizations show performances as trajectories through a two-dimensional plane that combines tempo curves and its derivatives. The essential difference to conventional time-series plots is that the *change* in the curves is represented explicitly as a dimension. An effect of this is that expressive timing gestures (patterns of changes in timing) tend to be more pronounced in phase-plane plots.

A striking aspect of phase-plane visualization of expressive timing is its suggestion of human gestural motion. Although this observation is based only on a similarity in appearance, it is an interesting question to what extent the phase-plane trajectories parallel the gestures humans make to mimic expressive aspects of music.

Apart from the allusion to physical motion, the phase-plane, by its focus on the derivatives of tempo, sometimes yields radically different phase-plane trajectories for curves that appear to be only slight variations in a time-series plot. This makes it potentially a valuable tool for modeling tempo curves by mathematical functions (as in Todd (1985); Friberg and Sundberg (1999)). This possibility is discussed in more detail in Grachten et al. (2008).

ACKNOWLEDGMENTS

This work is funded by the Austrian National Science Fund (FWF) under project number P19349-N15.

NOTES

[i]There is a relation between the dimensionality of attractors in the state space and the number of observations needed for reconstruction, as specified by the *delay embedding theorem* (Takens, 1981)

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

[iii]The use of the term interonset interval (IOI) might wrongly suggest that precisely one measurement value is given for each interval between consecutive onsets. Rather, measurement values are present for each eighth note interval, independently of the presence or absence of note onsets. A more correct term would be interbeat-interval (IBI) at half beat level

[iv]Unfortunately the adjustment of the plotting range needed to accommodate the large curves, somewhat obscures the view of the main part of the trajectory, mostly so in the case of Horowitz

References

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.

Grachten, M., Goebl, W., Flossmann, S., and Widmer, G. (2008). Phase-plane visualizations of gestural structure in expressive timing. In *Proceedings of the fourth Conference on Interdisciplinary Musicology*, Thessaloniki, Greece.

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. (2006). Computational modeling of music cognition: a case study on model selection. *Music Perception*, 23(5):365–376.

Kantz, H. and Schreiber, T. (1997). *Nonlinear time series analysis*. Cambridge University Press, Cambridge, UK.

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. (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.

Repp, B. H. (1993). Music as motion: A synopsis of Alexander Truslit's (1938) Gestaltung und Bewegung in der Musik. *Psychology of Music*, 21(1):48–72.

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.

Takens, F. (1981). Detecting strange attractors in turbulence. In Rand, D. A. and Young, L.-S., editors, *Dynamical Systems and Turbulence*, volume 898 of *Lecture Notes in Mathematics*, pages 366–381. Springer-Verlag.

Todd, N. (1995). The kinematics of musical expression. *Journal of the Acoustical Society of America*, 97 (3):1940–1949.

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. (2002). Machine discoveries: A few simple, robust local expression principles. *Journal of New Music Research*, 31(1):37–50.