Perception and Representation of Multiple Tempo Hierarchies in Musical Performance and Composition: Perspectives from a New Theoretical Approach

Jörg Langner
Hochschule für Musik Hannover

Reinhard Kopiez
Hochschule für Musik Würzburg

Bernhard Feiten
Deutsche Telekom/Technologiezentrum Darmstadt

ABSTRACT

Expressive changes of tempo in a musical performance are normally described using a zig-zag shaped tempo curve. In opposition to this one-dimensional perspective on tempo, it has recently been accepted that tempo is a multi-dimensional phenomenon. This multi-layer character of musical tempo implies the possibility of shaping time-layers independently from each other, for example, retardation and acceleration can happen simultaneously at different layers. The main goal of the presented study was to furnish proof of the hypothesized character of musical time by using a recently developed method which was based on the algorithms of the Theory of Oscillating Systems (TOS). Using a graphic representation of a so-called oscillogram, all

Correspondence address:
Reinhard Kopiez, Hochschule für Musik
Hofstallerei, 6-8
D-97070 Würzburg, Germany
E-mail: reinhard.kopiez@mail.uni-wuerzburg.de

1 The term 'tempogram' used in former publications (see Langner & Kopiez 1995, 1996) has been changed to 'oscillogram' starting from this publication.
tempo-layers contained in a composition can be visualized. The method's efficiency will be demonstrated in the following ways: (1) Through the use of a synthesised musical rhythm, the fundamental assumption of level-independent time-shaping, is demonstrated with the help of an oscillogram. (2) The analysis of a percussion performance will show that musicians really make use of this multi-level time-shaping. (3) An analysis of a small section of A. Honegger's Pacific 231 will prove that time-shaping is also used by composers.

**INTRODUCTION**

In past decades two methods have been established for the analysis of a musical performance: the analytical and the synthetical approach. The analytical approach, used for example in the studies by Seashore (1937), Repp (1992) or Epstein (1995), tries to extract the expressive parameter from a live performance by measuring the performance's acoustical or MIDI-data. The synthetical approach, represented mainly by the work of Sundberg (1991) and Friberg (1991) (for an overview see Kopiez 1996), describes a performance as determined by an expressive rule system (e.g. ritardando at the end of phrase). These rules are extracted intuitively from a live performance. If a convincing simulation (synthesis) can be generated by software using a set of rules, then we can assume that a musical performance can be described as the result of a limited set of performance-rules; both methods produced remarkable progress in the better understanding of the core of an expressive performance.

Despite this progress two questions remain unanswered: (a) how is a performance controlled by the performer and, (b) how is the performance stored in the player's memory? Existing models have tried to explain the storage of a performance in the player's memory with analogies to physical processes: Sundberg & Verillo (1980) used the ordinary experience of body-movement retardation as a controlling factor for the execution of a final retard; Todd (1992) used mechanical equations to describe the shaping of musical dynamics in analogy to the experience of body-acceleration.

The theoretical approach, which is also presented in this paper, tries to unite both aspects: the execution and storage of an expressive performance. The approach is demonstrated by using examples of multiple tempo-shaping. Before the theory can be described, one must first question what the relationship between creative processes and the shaping of time in a musical performance is. From the traditional, psychomusicological point of view – as represented by the work of Seashore (1938/1967) – expressive timing, as a deviation of metronomic regularity, seems to be a necessary, but not very complex „ingredient“ to an acceptable artistic performance. From this point of view, it is sufficient to describe timing as a one-dimensional phenomenon by the mean of so-called zig-zag shaped tempo curve. To this day, tempo curves are used in numerous publications.

Contrary to this simplified, graphical description of timing, historical sources of performance practice hint that the shaping of time by performers and composers must be more complex. E.g.: In the last 300 years there is no performance practice manual which omits a chapter on the shaping of time. In recent years, the fundamental source-study of Hudson (1994) – concerning the tempo rubato – shows the complex use of timing from the gregorian chant up until music heard today. Additionally, Chopin's pupils reported that one of the main goals of his teaching was to mediate the correct use of tempo rubato.

In keeping time Chopin was inexorable, and some readers will be surprised to learn that the metronom never left his piano. Even in his much maligned tempo rubato, the hand responsible for the accompaniment would keep strict time, while the other hand, singing the melody, would free the essence of the musical thought from all rhythmic fetters, either by lingering hesitantly or by eagerly anticipating the movement with a certain impatient vehemence akin to passionate speech\(^\dagger\) (cited after Eigeldinger 1986, 49).

What can be concluded from this short description? It can concluded that Chopin's shaping of time is at least two-dimensional, with each hand representing an independent dimension. The independent shaping of different time layers, however,
is not only found in the two hands of a pianist, but can also be found in music as well, which is not manually pre-divided in two layers of events, e.g.: this phenomenon can also be observed in percussion music. It can be assumed that this simple example of Chopin’s teaching gives us an insight into the complex reality of musical timing, which is not only two-dimensional, but multi-dimensional. However, despite this assumption, the study of experimental literature gives the impression that timing seems to be systematically underestimated in its complexity. In our investigation, we hypothesized a multi-layer character of timing both in performance and composition. Furthermore, we believe that musicians and composers make use of this complex character of musical timing in performances and compositions, and possess elaborated strategies to create and control a complex shaping of time.

In recent years, some researchers have given up the one-dimensional timing curve paradigm and tried to develop alternative methods for an adequate description of time-shaping processes: David Epstein’s (1995) extensive study *Shaping Time*, based on numerous timing-analyses of performances, is a milestone in timing-research. Epstein’s theory of timing ratios as a determining factor in the creation of a convincing timing architecture, both in composition and performance, was proven by his detailed, analytical results which showed that timing is shaped on numerous levels ranging from the bar unit to a whole movement. In a tempo-analytic study using Weite-Mignon player-piano paper rolls, Gottschewski (1996) (see also Danuser 1992, 318f.) showed, for example, that the pianist Carl Reinecke could shape his phrasing in the 2nd movement of Mozart’s Piano-Concerto KV 537 up to 6 metrical layers simultaneously; the whole movement was the largest timing unit used. The multi-dimensional timing structure can also be tested using the synthetic approach and musical examples from performance manuals. Carl Czerny (1839/1991, 25) gives a short (4 bars) practice example on agogic in his treatise on musical expression. However, we cannot create a satisfying musical realization of this example, even if Czerny’s intended timing was synthesized according to his instructions. It is only possible to finish our timing „handicraft work” successfully through the use of software which enables us to create multiple timing hierarchies, e.g. the software PRESTO developed by Mazzola (1993). The most far-reaching perspective on how timing is created in the players mind was discussed in the recently developed „inverse performance theory” by Mazzola & Zahorka (1995). This theory assumes that if music interpretation is looked upon as the result of a long decision-chain, starting with the primavista play and finishing with the concert performance, the decision-path could be traced backwards in order to understand how changes in musical thinking create a more differentiated timing-architecture.

In our investigation, we analysed the multi-dimensional character of musical time used by performers and composers in order to find out, if and how these time-layers are shaped independently from each other. In the next section we will illustrate our findings with some musical examples.

**METHOD**

The questions which have been raised can be viewed more closely with the help of the „Theory of Oscillating Systems” (TOS) (Langner, in preparation). This theory was based on the assumption that every musical progression triggers a series of oscillations or oscillation-like processes in the human perception. (Such oscillations can be regarded as the periodic firing of nerve cells.) This approach is not new in the psychology of music research – Gjerdingen (1992), Large & Kolen (1994), Parncutt (1994) and Todd (1994), for example, have all dealt with this subject – but, rather specific to our method is the assumption of a very broad spectrum of oscillating frequencies. Our model contains a set of 120 oscillators, each having a fixed frequency and arranged in logarithmic steps from 8 Hz down to 0.008 Hz. This lower limit may seem surprisingly low, (a frequency of 0.008 Hz has a period of vibration lasting over two minutes!) but our initial studies revealed that such low-frequency oscillations can indeed be significant in a musical sense and, furthermore, appear to correspond with the large artistic spans of tension and relaxation in a composition or interpretation. There are two publications by Langner & Kopiez (1995, 1996) which deal with this subject. However, these investigations were limited to the upper frequency range. We also concentrated on how oscillators are stimulated by the music. We began by looking at the loudness progression in decibels, which could be calculated from a sound recording; this was then used to detect the note onsets, i.e.
the respective point in time was determined within a few milliseconds. Together, loudness progression and onsets form the input for a computer program which calculates the strengths of the oscillations at every point in time. The algorithm required for this has a certain affinity with the Fourier transformation, refer to the publications of Langner & Kopiez (1995, 1996); the full mathematical description can be found in Langner (in preparation). This algorithm works in such a way as to produce plausible results even in simple cases, for example, if a musician plays quarter-notes at a tempo of M.M. = 120, a strong oscillation at 2 Hz is expected, and this is what the calculation yields. The results of the calculations are presented in pictorial form by mean of so-called oscillograms.

The calculation of an oscillogram needs the exact identification of the note onsets. It can be very difficult, however, to detect onset in acoustical complex signals, like orchestral music; therefore, we used software specifically designed to suite our needs. The software will be described in the following paragraph.

Calculation of Loudness Curves

The calculation of the loudness curve in music can be done in several ways. Ideally, the calculated curve corresponds to the subjectively perceived loudness; therefore, the best result can be obtained if the calculation method is based on a psychoacoustic model (see Zwicker & Fastl 1990). In the following paragraphs, different aspects and modifications of the calculation of loudness curves are discussed.

The easiest way to determine a loudness curve is through the measurement of the sound pressure level in decibel over time. The correlation between measured and perceived loudness is weak because psychoacoustic effects are considered only by using the logarithmic scale. Frequency-dependent effects, like the integration of sound energy in the auditory critical bands, is not considered. Furthermore, the integration and averaging parameters which determine the short-time-power have to be adapted to cover the whole frequency range. A favorable compromise must be found which not only offers a sufficient time-resolution, but avoids ripple caused by the low frequencies as well. Better results can be obtained by dividing the frequency range in sub-bands. The calculation of loudness, as proposed by Zwicker's model, divides the frequency range in critical bands so that they model the human auditory system. In our software program this feature was simulated with a sliding short time Fourier transformation. The partial frequency bands of the Fourier transformation were added to the critical bands, for example, the critical band width is about 100 Hz for low frequency and 3500 Hz for high frequencies. Furthermore, the effect of the outer ear and masking effects were considered, as well as the connection between the sound pressure level and the perceived loudness. In order to obtain an overall loudness, the partial loudness of the 25 critical bands was summed up. A temporal resolution is given by the model, but the time distance of the calculated output values can be chosen as optional parameters.

Figures 1a—c show time-loudness graphs resulting from three different methods of calculation: using a section from Honegger's Pacific 231 (measures 109—113), it can be seen that the „worst“ loudness curve is produced by using the decibel scale (Figure 1a). It is nearly impossible to see a connection between this „wriggle-curve“ and the musical events in the score. In Figure 1b the loudness curve is calculated by using the frequency-dependent loudness scale of sone, according to the model of Zwicker (Zwicker & Fastl 1990). Obviously, a much better onset detection of musical events is possible as a result of a better distinction between noise and signal. Figure 1c shows a loudness curve from the same section by using the frequency-dependent loudness scaling of Zwicker with an additional weighting of the high frequencies. In correspondence to the score, the rhythm of the trumpets and clarinets can be seen in bar 109 (see the score in Figure 2 and listen to the sound example CD track 1). Every onset in trumpet and clarinet is detectable, both the enhanced psychoacoustically based loudness curve (Figure 1b) and a manipulated version of it (Figure 1c) are necessary to provide an oscillation analysis with complex acoustical signals like orchestral music.
Figure 1a—c. Three loudness curves belonging to the measures 109—113, first beat of Honegger's Pacific 231 (conducted by Jean Martinon). Figure 1a represents the sound pressure level in decibel, Figure 1b the loudness in sone calculated according to the model of Zwicker (Zwicker & Fastl 1990), Figure 1c is a curve resulting from a Zwicker-like procedure but weighting the high frequencies much more strong. The three graphs show increasing aptitude for detection of trumpet and clarinet onsets. Compared to the score, these onsets are almost not detectable in Figure 1a, but clearly visible in Figure 1c.
RESULTS AND DISCUSSION

Figure 3 shows the properties of an oscillogram: The y-axis of the upper graph shows in each row one of the 120 layered oscillators of the analytical model. Numbers indicate their oscillation frequencies in Hertz. On the Figure's right side, note symbols indicate the metrical layers which correspond to the oscillator's periods. These duration symbols differ from piece to piece and depend upon the absolute tempo and the durations contained. The x-axis is a time-axis. In the upper area of Figure 3, we can see how the intensity of each oscillator's activation – triggered by an external rhythmic signal – is represented by different degrees of black shading. The graph's default color is white. In the box below the oscillogram, the input signal's loudness curve is shown in a time (seconds) x intensity (decibel) diagram.

Figure 3 represents the oscillations triggered by a simple (synthesized) 4/4-beat with M.M. = 120 in our model. The oscillogram indicates that the highest intensity of a rhythmic oscillation was triggered in the vicinity of 2 Hz, corresponding to the input's two beats per second. At the same time, an activation on the half note and the whole note level should also be noted. It is interesting to compare this multi-activation to the results of so-called tapping-experiments, because as Parnccut (1994) showed, a subject's tapping to a given pulse can occur on different metrical levels. In the vicinity of 4 Hz we can see a surprising effect: The 2 Hz-pulse seems to produce something like a „rhythmic overtone“ in our perception – however, with a much lower intensity than the pulse actually has.

Figure 3. Oscillogram (upper graph) and loudness curve (lower graph) belonging to quarter-notes played in a 4/4-time-signature by a drumcomputer. The three metrical levels of the musical input correspond to three layers of activated oscillators indicated by black shading in the oscillogram.
Another example using a synthesised rhythm (same conga-sound) is shown in Figure 4: A 4/4-beat combined with an accelerando. In this Figure, the frequencies of the oscillators of the three main layers of time perception rose in correspondence to the accelerating tempo. In the previous Figures, the three curves move parallel and it would be interesting to know if the existence of non-parallel or even opposing directions could be demonstrated.

The first example of non-parallel tempo shaping was taken from a well-known source: the Viennese Waltz. Bengtsson & Gabrielsson (1983) investigated the timing of the quarter-notes as played by qualified musicians and established that the first quarter-note of every bar was considerably shortened, whereas the second was considerably lengthened. Sound example CD track 2 contains an idiomatic waltz rhythm played by a drum computer programmed with data obtained from the cited study.

We believe that this synthetic version contains something of the feeling of a Viennese Waltz. When looking at the loudness progression (Figure 5, lower graph), a shortening and lengthening of the interonset intervals and the strong intensity accent on the first quarter-note – due to 28 additional Midi-velocities – can be seen. The oscillogram in Figure 5 shows a dark band below 1 Hz, this corresponds to the level of bars; the tempo here remains constant. Oscillations belonging to the quarter-notes are located above 2 Hz; slight deviations from an average tempo are evident. These deviations in the oscillogram are weaker than might be expected from the timing data (the activation peaks are hardly displaced at all). Owing to the rapid succession of fluctuations in the tempo, the method produces something akin to a „smoothing“ – which agrees with the listeners' aural impression and does not register any pronounced ritardandi or accelerandi.

Figure 4. Oscillogram (upper graph) and loudness curve (lower graph) belonging to quarter-notes played in a 4/4-time-signature by a drumcomputer with an additional accelerando at the end. This accelerando leads to a parallel upward movement of the dark bands in the oscillogram, which means that the activation changes to higher oscillation frequencies.
Figure 5. Oscillogram (upper graph) and loudness curve (lower graph) belonging to quarter-notes played in a 3/4-time-signature with Viennese Waltz timing by a drum-computer. The dark band below 1 Hz in the oscillogram remains absolutely constant, no tempo changes occur at the level of the bars, whereas the tempo deviations at the level of quarter-notes lead to corresponding deviations in the oscillators' activation above 2 Hz.

The next example of the application of an oscillogram is a simple rhythm in a 7/8-time-signature performed by a human performer on a conga. In comparison to the waltz, the oscillogram (Figure 6) reveals the opposite here: a substantially constant progression at the upper level (eighth-notes), changes of tempo, on the other hand, in lower layers. The drops in tempo here belong to the groups of three eighth-notes at the end of each 7/8-bar.

Figure 6. Oscillogram (upper graph) and loudness curve (lower graph) belonging to eighth-notes played in a 7/8-time-signature by a human performer. In the oscillogram we can see a nearby constant tempo at the level of the eighth-notes but tempo drops at a lower level. These tempo drops belong to the asymmetric accentuation in the succession of eighth-notes which are grouped into 2+2+3 each bar, the tempo, with respect to these accented notes, slows down with the occurrence of the group of three.
In many cases, and especially when based on a live performance, the oscillogram does not reveal a completely clear picture. Oscillations were obtained which did not correspond to any of the values occurring in the music. It could be possible to implement an algorithm which suppresses all oscillations, whose frequencies have no simple integer ratios to the main oscillations; such an algorithm would result in a much „cleaner“ oscillogram. However, it is not certain whether and to what degree such a suppression would be desirable. In some cases though, results were obtained in which these somewhat strange oscillations could be called upon to explain certain puzzling musical phenomena. Therefore, for the moment, the oscillogram was left as is.

The last example (Figure 7, sound example CD track 1) was taken from the orchestral piece Pacific 231 by Arthur Honegger. The onset detection in this example was a rather difficult task and, therefore the sophisticated method described above was used.

On the right side of Figure 7, three primary oscillation bands, corresponding to the eighth-, quarter- and half-note levels, can be seen. The important event occurs at measure 118: an acceleration was detected on the upper level, but a deceleration on the lower two levels. The tempo speeds up and slows down simultaneously (!). This process is an integral part of the composition: In the score corresponding to this section (see Figure 2) Honegger changed from eighth-notes to triplets, while reducing the underlying tempo from quarter-note M.M. = 152 to 144 at the same time (incidentally, the conductor does not perform this exactly, slowing down from M.M. = 169 to 127).
GENERAL DISCUSSION

We believe that the significance of the multilayer make-up, the hierarchical or non-hierarchical structuring in various levels, has become clear in the realm of tempo shaping. A method was introduced which enables us to analytically approach complex relationships and present the results in an unambiguous manner.

A perspective for future applications of our method shall be demonstrated with the phenomenon of body movements in a performance. Body movements and gestures are in close relation to musical timing. Our neuro-psychological oscillation model implies that body movements can be triggered by musical events. As Delalande (1988) observed from a semiotic approach, it is certain that a performer’s gestures are not accidental; gestures are in close relation to the musical structure and are used to draw the observer’s attention to important structural aspects. However, until now there has been no procedure to analyse a musician’s movements under the perspective of body-oscillations. In the future, we hope to investigate the correlation between oscillations triggered by acoustical and visual stimuli by implementing video-taped body movements (e.g. the forwards and backwards movements of the head) as input for oscillation analysis.

The last area we would like to discuss is the strength of our theoretical approach in explaining production and perception using oscillations. In the player’s memory the expressive deviations – for example from a metronomic timing – in his or her performance are stored as a set of numbers, representing the so-called „function of state“, which can be thought and used as a set of commands for oscillators. This „function of state“ enables the economical control and execution of the performance of a complete composition because it is possible to store a musical process developing in time with a minimal set of data which remain invariant. On the listener’s side the perceived performance triggers oscillations which are stored again, which means that remembering a performance reactivates a function of state of oscillations caused by the perception of the piece. These imaginary oscillations could explain why it is possible to compare different interpretations of the same piece: by comparing patterns of oscillations. From this perspective, production and perception of expressiveness in music are essentially one entity. This being the case, a good musician is always his/her best listener!

ACKNOWLEDGMENT

The software for frequency-dependent onset detection in orchestral music has been written by Bernhard Feiten and Markus Spitzer (both Technical University of Berlin).

REFERENCES


