

Visualizing Expressive Performance in Tempo-Loudness Space

Author(s): Jörg Langner and Werner Goebel

Source: *Computer Music Journal*, Winter, 2003, Vol. 27, No. 4 (Winter, 2003), pp. 69-83

Published by: The MIT Press

Stable URL: <https://www.jstor.org/stable/3681902>

JSTOR is a not-for-profit service that helps scholars, researchers, and students discover, use, and build upon a wide range of content in a trusted digital archive. We use information technology and tools to increase productivity and facilitate new forms of scholarship. For more information about JSTOR, please contact support@jstor.org.

Your use of the JSTOR archive indicates your acceptance of the Terms & Conditions of Use, available at <https://about.jstor.org/terms>



The MIT Press is collaborating with JSTOR to digitize, preserve and extend access to *Computer Music Journal*

JSTOR

Visualizing Expressive Performance in Tempo–Loudness Space

The previous decades of performance research have yielded a large number of very detailed studies analyzing various parameters of expressive music performance (see Palmer 1997 and Gabrielsson 1999 for an overview). A special focus was given to expressive piano performance, because the expressive parameters are relatively few (timing, dynamics, and articulation, including pedaling) and comparatively easy to obtain. The majority of performance studies concentrated on one of these parameters exclusively, and in most of these cases, this parameter was expressive timing.

In our everyday experience, we never listen to one of these parameters in isolation as it is analyzed in performance research. Certainly, the listener's attention can be guided sometimes more to one particular parameter (e.g., the forced stable tempo in a Prokofieff Toccata or the staccato–legato alternation in a Mozart Allegro), but generally the aesthetic impression of a performance results from an integrated perception of all performance parameters and is influenced by other factors like body movements and the socio-cultural background of a performer or a performance as well. It can be presumed that the different performance parameters influence and depend on each other in various and intricate ways. (For example, Todd 1992 and Juslin, Friberg, and Bresin 2002 provide modeling-based approaches.) Novel research methods could help us to analyze expressive music performances in a more holistic way to tackle these questions.

Another problem of performance analysis is the enormously large amounts of information the researcher must deal with, even when investigating, for example, only the timing of a few bars of a sin-

gle piece. In general, it remains unclear whether the expressive deviations measured are due to deliberate expressive strategies, musical structure, motor noise, imprecision of the performer, or even measurement errors.

In the present article, we develop an integrated analysis technique in which tempo and loudness are processed and displayed at the same time. Both the tempo and loudness curves are smoothed with a window size corresponding ideally to the length of a bar. These two performance parameters are then displayed in a two-dimensional performance space on a computer screen: a dot moves in synchrony with the sound of the performance. The trajectory of its tail describes geometric shapes that are intrinsically different for different performances. Such an animated display seems to be a useful visualization tool for performance research. The simultaneous display of tempo and loudness allows us to study interactions between these two parameters by themselves or with respect to properties of the musical score.

The behavior of the algorithm and insights provided by this type of display are illustrated with performances of two musical excerpts by Chopin and Schubert. In the first case study, two expert performances and a professional recording by Maurizio Pollini are compared; in the second case study, an algorithmic performance according to a basic performance model is contrasted by Alfred Brendel's performance of the same excerpt. These two excerpts were chosen because articulation is constant throughout the whole excerpt (legato), and analysis can concentrate on tempo and dynamics.

Method

Our visualization requires two main steps in processing. The first step involves data acquisition ei-

ther from performances made on special recording instruments such as MIDI grand pianos or directly from conventional audio recordings (i.e., commercial compact discs). Second, the gathered data must be reduced (smoothed) over a certain time window corresponding to a certain granularity of display.

Timing Data

The timing information of expressive performances in MIDI format has the advantage of having each onset clearly defined, although the precision of some computer-monitored pianos is not much higher than obtaining timing data from audio recordings. (For a Yamaha Disklavier, see Goebel and Bresin 2001). Yet, each performed onset must be matched to a symbolic score of a given piece so that the onsets of the *track level* can be automatically determined (i.e., score-performance matching; see Heijink et al. 2000 and Widmer 2001). The track level is a unit of score time (e.g., quarter note, eighth note) that defines the resolution at which tempo changes are measured. The track level is usually faster than the beat as indicated through the time signature. For example, in the Chopin *Etude* it is the sixteenth note. From this, the tempo curves (in beats per minute relative to the notated beat) are computed.

Timing information from audio recordings was obtained by using an interactive software tool for automatic beat detection (Dixon 2001a, 2001b). The software analyzes audio data, finding the onsets of as many of the musical notes as possible, and proposes a possible beat track by displaying the beats as vertical lines over the amplitude envelope of the audio signal. The user has the opportunity to adjust false track times. The system also provides audio feedback in the form of a percussion track (playing at the track times) mixed with the original sound. After correcting some of the errors made by the system, the remainder of the piece can be automatically retracked, taking into account the corrections made by the user. When the user considers the track times to be correct, the track times can be saved on disk in a text format. With this tool, timing data of audio recordings can be gathered

relatively quickly. The typical error of ± 20 msec is sufficiently precise for the present purpose (see also Goebel and Dixon 2001).

Loudness Data

For both MIDI and audio data sources, the loudness information was always taken from a recording of the MIDI file or the audio file itself, respectively. It would have been very difficult to model an overall loudness curve based only on MIDI information. A MATLAB implementation of Zwicker's loudness model (Zwicker and Fastl 2001) was used to convert the audio file (in pulse-code modulated WAV format) into its loudness envelope in sones (Pamplak, Rauber, and Merkl 2002). First, the audio signal was converted into the frequency domain and bundled into critical-bands according to the Bark scale. After determining spectral and temporal masking effects, the loudness sensation (sones) was computed from the equal-loudness levels (phons), which in turn were calculated from the sound pressure levels in decibels (dB SPL). The loudness envelope was sampled at 11.6-msec intervals according to the window size and sampling rate used (1,024 at 44,100 samples per second with 50% overlap). A similar implementation was used in earlier studies (Langner et al. 2000; Langner 2002; Langner and Goebel 2002). The advantage of using a loudness measure instead of sound level is discussed by Langner (2002, pp. 29–30).

From the loudness envelope, for each tracked point in time one loudness value was taken for further data processing. This single loudness value for each track time was taken as the maximum value in a window half an inter-track interval before and after the corresponding track time. Because the track grid could miss important loud notes in some cases that did not coincide with a track time, this windowing procedure accounted for that and would have taken the neighboring louder note as the loudness value of that particular track. This procedure is particularly important for low track rates.

Data Reduction

Both tempo and loudness data were smoothed using overlapping Gaussian windows. We refer to the

Table 1. Window sizes used for smoothing the performances of the Chopin *Etude*

Performer	Window Size (sec)
Pianist 09	2.486 (bar)
Pianist 18	2.896 (bar)
M. Pollini	3.212 (bar)
	1.606 (quarter note = half bar)

The window sizes correspond to the mode duration of a performed bar or to the mode duration of a quarter note, respectively. The mode duration was the most often occurring inter-onset interval (quantized to 10 msec; see Goebel and Dixon 2001).

window size as being the time from the left to the right point of inflection (turning point) of the Gaussian window (in sec) corresponding to two standard deviations (i.e., 2σ). A smoothed data point y at a certain time t is determined as in Equation 1, where $x(t)$ is the unsmoothed data with the sampling period F (frame period in sec); k was set to the integer nearest to $3\sigma/F$, since for larger values the exponential term is vanishingly small.

$$y(t) = \frac{\sum_{i=-k}^k x(t + iF)e^{-(iF)^2/2\sigma^2}}{\sum_{i=-k}^k e^{-(iF)^2/2\sigma^2}} \quad (1)$$

For the current study, we usually took a window size corresponding to the average performed duration of a bar, resulting typically in a window size of around 3 sec (see Tables 1 and 2). The choice of window size is arbitrary and can be set by the investigator. For example, a window size corresponding to the length of a quarter note will bring out more local phenomena of a performance, and a smoothing window on a four-bar level will show only very long-term developments.

Two-Dimensional Display

The smoothed data is displayed in a two-dimensional space of tempo (x axis) versus loudness (y axis). This visualization is conceptualized to work as an animation over time: a red dot moves in synchrony with the music, leaving behind it a trajectory. To elaborate the impression of time (the

third dimension in the display) the trajectory of the initial red dot fades out and decreases in size over time. This is meant to evoke an impression of a three-dimensional virtual space in which the trajectory moves towards the viewer.

The current dot of the display can show high-level score information, such as the current bar number. To indicate some types of structural properties of the score, the current dot is enlarged and changed in color at phrase boundaries.

Snapshots of this visualization technique as well as the animations are implemented in the MATLAB environment. The animations were saved as QuickTime movies using a routine freely available on the Internet (Slaney 1999). The frame rate was chosen to be 0.1 sec (i.e., 10 frames per sec). The examples discussed below can be downloaded as QuickTime videos from www.oefai.at/~werner/ animations.

Case Study I: Chopin's E-Major *Etude*

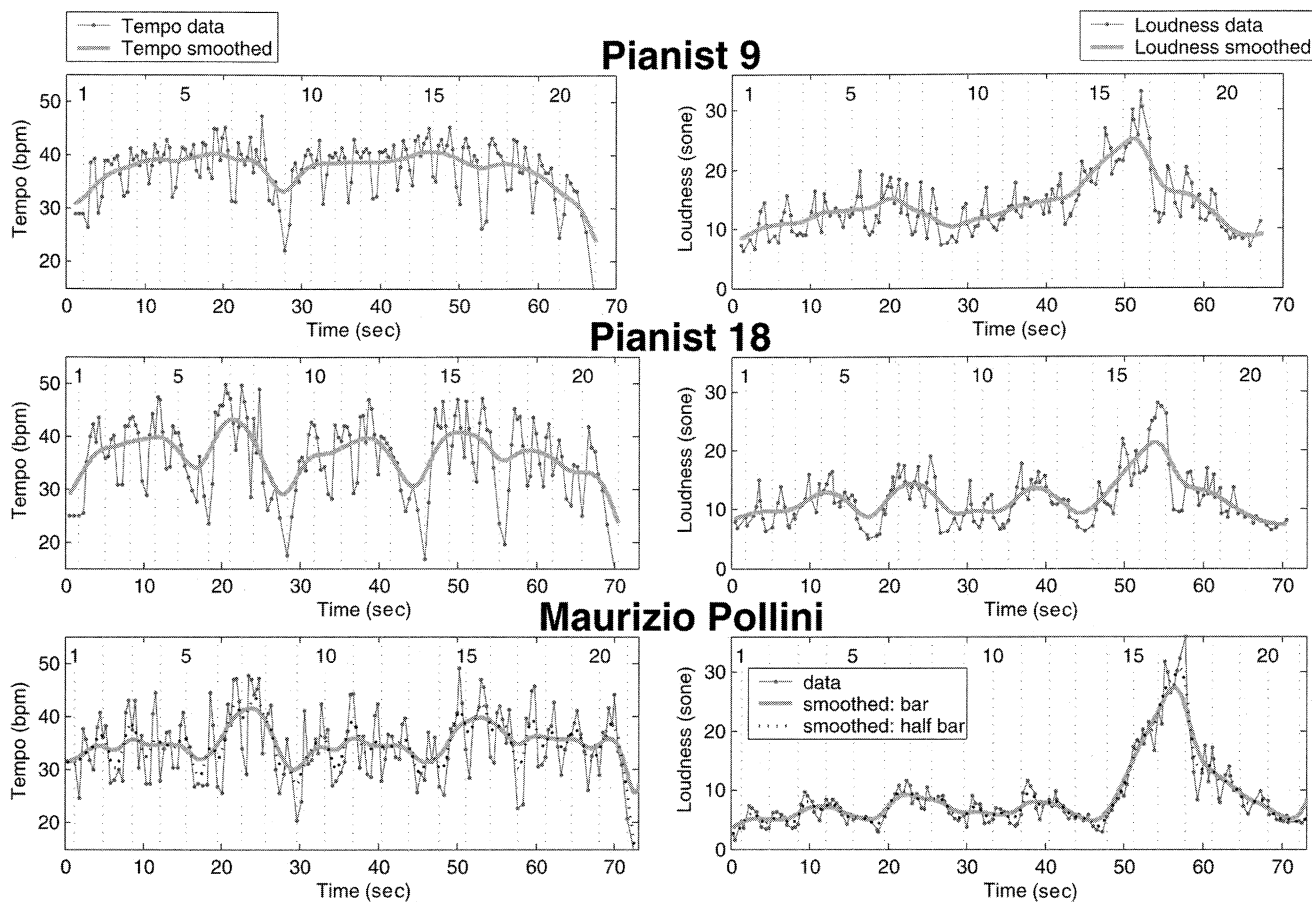
To exemplify the properties of the visualization technique, several performances of the initial 21 bars of Chopin's *Etude*, Op. 10, No. 3 were chosen. Two of them were recordings on a Bösendorfer SE290 computer-controlled grand piano made for a previous study (Pianists 9 and 18 from Goebel 2001; the timing information comes in a MIDI-like format, and the pianists were asked to perform only until bar 21), and a professional recording by Maurizio Pollini (1985). The Chopin *Etude* has a very homogenous texture. Sixteenth notes are omnipresent; therefore, the track level was set to the sixteenth-note level. In Figure 1, the raw tempo and loudness curves on the sixteenth-note level of the three performances of the *Etude* are plotted against time. The thick lines represent the smoothed data curves with a smoothing window corresponding to an average performed bar, or two quarter notes. The exact window sizes are printed in Table 1. Figure 2 shows a musical score of the excerpt.

The display of tempo and loudness curves in Figure 1 illustrates the second of the two main steps in data processing as mentioned above: data

Figure 1. Tempo and loudness curves of the initial 21 bars of the Chopin Etude in E Major, Op. 10, No. 3, as played by Pianist 9 (top), Pianist 18 (middle), and Maurizio Pollini (bottom). The figure also

shows tempo curves (left) and loudness curves (right) on the sixteenth note level (lines with dot markers) and smoothed on the bar level (thick line) against time in sec. The faint vertical lines indicate bar

lines with the corresponding numbering on top. In the bottom panels (Pollini), the data were also smoothed on the half-bar level (quarter-note level, dotted lines).



reduction via smoothing. These smoothed curves are the basis for the animations. In Figure 3, snapshots of such animations are shown. The instant where the animation was stopped is either shortly after the beginning of bar 14 (left column) or shortly before the beginning of bar 21 (right column). The thicker black disks indicate bar lines.

Many interesting observations emerge from that combined display that reveal both the similarities and the differences between the performances. The most fundamental commonality between the three performances is that the expressive trajectory tends to go to the lower left side of the space at phrase boundaries at the beginnings of the bars 6, 9, and 14, which corresponds to a slowing down and a decrescendo towards the end of phrases. An exception is Pianist 9, who ignores the phrase boundaries at bars 6 and 14. In the snapshots of Figure 3, the

beginnings of bar 6 are hidden by the trajectories. However, the reported effects can be seen in Figure 1 or with the QuickTime movies available online. Another striking similarity between the pianists is that at the beginning of a phrase, the performers mostly increase the tempo first and then the loudness. This leads repeatedly to the phenomenon that the tempo apex occurs before the loudness apex, as can be seen in all three performances (see Figure 3, left column). Moreover, we can state a certain tendency towards a counterclockwise movement of the expressive trajectories, as found in all three performances.

On the other hand, the differences between the performances are striking at first glance. As mentioned above, Pianist 9 is not concerned much with slowing down at phrase boundaries, except at the end of bar 8. It happens that he starts the large cres-

cendo from measure 14 onward already in the middle of the space, whereas both Pianist 18 and Mr. Pollini begin this development closer to the lower-left corner.

The visually prominent loops resulting from the increase of measures 14 onward also reveal counterclockwise shapes for all three pianists. These trajectories reflect the strong crescendo and the *ritenuto* at the apex of that phrase in bar 17, as indicated in the score. The *ritenuto* at the apex of Pianist 18 is strongest, but Pianist 18 then speeds up again while decreasing volume, a tendency that is only weakly found in Mr. Pollini's performance and in Pianist 9's performance.

The artistry of Mr. Pollini comes out when the concept of his interpretation is taken into consideration: throughout the first 14 bars, he remains very soft and avoids larger tempo changes to spare his expressive energies for the coming outburst. Another difference between the famous pianist and the two others is that Pollini does not slow down too much at the end of the section. He planned, of course, to play the whole *Etude* and not only the first 21 measures, as the other pianists did (see also Figure 1).

The shape of a trajectory is very much dependent on the window size of the smoothing window applied to the data. With the average performed bar duration as the size of the Gaussian window between the two turning points, we eliminate all fluctuations within a bar. Only larger developments of tempo and dynamics remain in the display. Figure 4 displays the same performance of Maurizio Pollini with the smoothing window shrunk by a factor of 2, so that it corresponds to the quarter-note level (half bar; see also Figure 1 bottom panels, dotted line). Many more loops appear, and the extent of used performance space is larger. For example, the late *ff* chord in bar 17 and its delayed succeeding note (as plotted in Figure 1, left bottom panel) clearly appear in that display, but this very local phenomenon was smoothed out at bar-level windowing. The general question arises whether we could speak of a slowing down when one or two notes appear delayed rather than of a local expressive deviation.

It is also important to point out that smoothing introduces some artifacts too. Smoothing can

change the location and extent of turning points and peaks in data. All three examples in Figure 3 suggest that the performed dynamic peak does not coincide with the notated climax in bar 17. This is true for Pianists 9 and 18, but not for Mr. Pollini, as can be verified in Figure 1. Mr. Pollini plays the *ff* chord loudest while the two others do not. The quarter note smoothing condition (Figure 1, lower left panel) shifts Mr. Pollini's dynamic peak towards the correct position, but the performance trajectory still suggests a decrescendo before the climax on bar 17. When changes in data do not occur symmetrically, smoothing will shift the turning point. Therefore, analysis with this visualization technique should be accompanied by conventional data display for detailed analysis.

Case Study II: Schubert's G-Flat-Major *Impromptu*

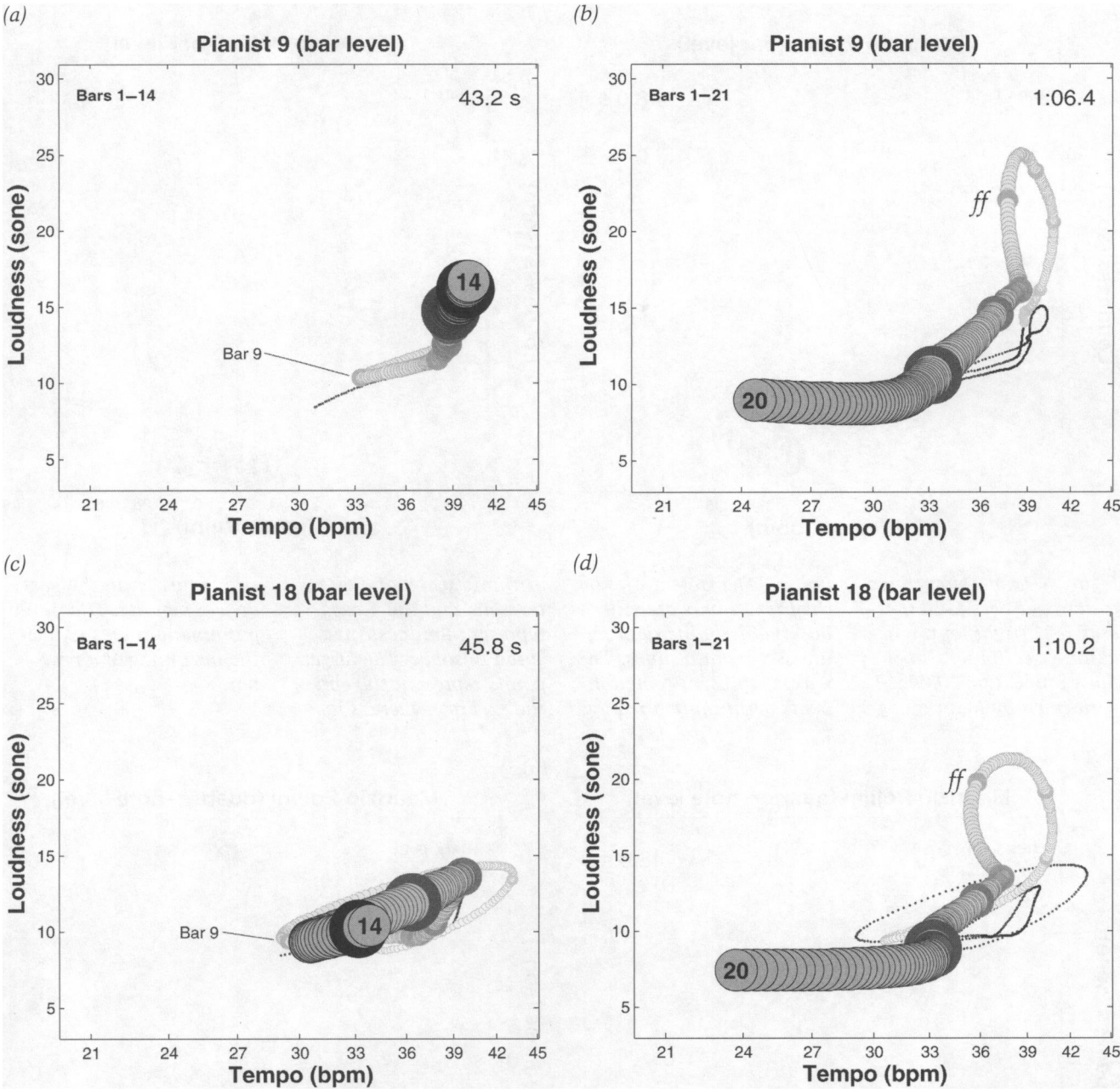
The main strength of the introduced display is to elucidate relations between tempo and dynamics. Exactly this relationship was modeled by Neil Todd in a very simple way: "[T]he faster the louder, the slower the softer" (Todd 1992). Windsor and Clarke (1997) used this model to test how much of the expressive variation of a real performance could be explained and what remains still unexplained. They used Schubert's G-flat-major *Impromptu* (D. 899, No. 3) as a test piece, as shown in Figure 5. The only input to the Todd model is the grouping structure of the musical score according to Lerdahl and Jackendoff (1983). Windsor and Clarke (1997) produced their preferred algorithmic performance using different parameters for timing and dynamics and called it the *hybrid performance*. It was fitted to a human performance by varying the parameters of the Todd model on five different phrase levels by trial and error. The chosen values are AP(1,1,1,2,4) for timing and AP(4,8,8,1,1) for dynamics. (The five different numbers refer to the model parameters, beginning with the highest phrase levels (8-bar, 4-bar, 2-bar, 1-bar, half-bar). Thus, in this hybrid performance, timing is monitored by the smallest phrase units (half-bar level), whereas the dynamics contours are shaped more by higher phrase levels (especially 4- and 2-bar phrase

Figure 2. Frédéric Chopin:
Etude in E Major, Op. 10,
No. 3, measures 1–21. The
score was prepared with
computer software by the
authors after the Paderew-
ski Edition.

Lento ma non troppo (♩ = 100)

The score is for Frédéric Chopin's Etude in E Major, Op. 10, No. 3, measures 1–21. It is in E major (three sharps) and 2/4 time. The tempo is marked *Lento ma non troppo* with a quarter note equal to 100 beats per minute. The score is written for piano and consists of five systems of two staves each. The first system (measures 1–4) begins with a *legato* marking and a piano (*p*) dynamic. The second system (measures 5–8) includes a *cresc.* (crescendo) marking. The third system (measures 9–12) continues the *cresc.* and includes a *riten.* (ritardando) marking. The fourth system (measures 13–16) features a *stretto* marking, a *cresc.*, and a *con forza* instruction. The fifth system (measures 17–21) includes a *ff* (fortissimo) dynamic, a *ten.* (tenuto) marking, a *sempre legato* instruction, a *dim.* (diminuendo) marking, and a *rallent.* (rallentando) marking. The piece concludes with a *pp* (pianissimo) dynamic and a *ten.* marking. The score is prepared with computer software after the Paderewski Edition.

Figure 3. Expression trajectories over bars 1–14 (left column) and 1–21 (right column) of the beginning section of Chopin’s Etude, Op. 10, No. 3, performed by Pianist 9 (top), Pianist 18 (middle), and Maurizio Pollini (bottom). The trajectories of the first 14 bars still can be seen in the right panel snapshots as very faint lines. The larger points represent the end of the excerpt, whereas in-



structure; see Windsor and Clarke 1997, p. 141). The exact smoothing values we used are shown in Table 2. In Figure 6, the first 16 bars of this algorithmic performance is contrasted to a professional performance by Alfred Brendel (1997). The hybrid perfor-

mance shows basically a diagonal line that reflects the “faster-louder” rule proposed by Todd. However, it is not a perfect diagonal as would have been expected for the following two reasons. First, in this hybrid performance, different parameters for timing and intensity were chosen. Second, the lin-

Figure 3. Continued.

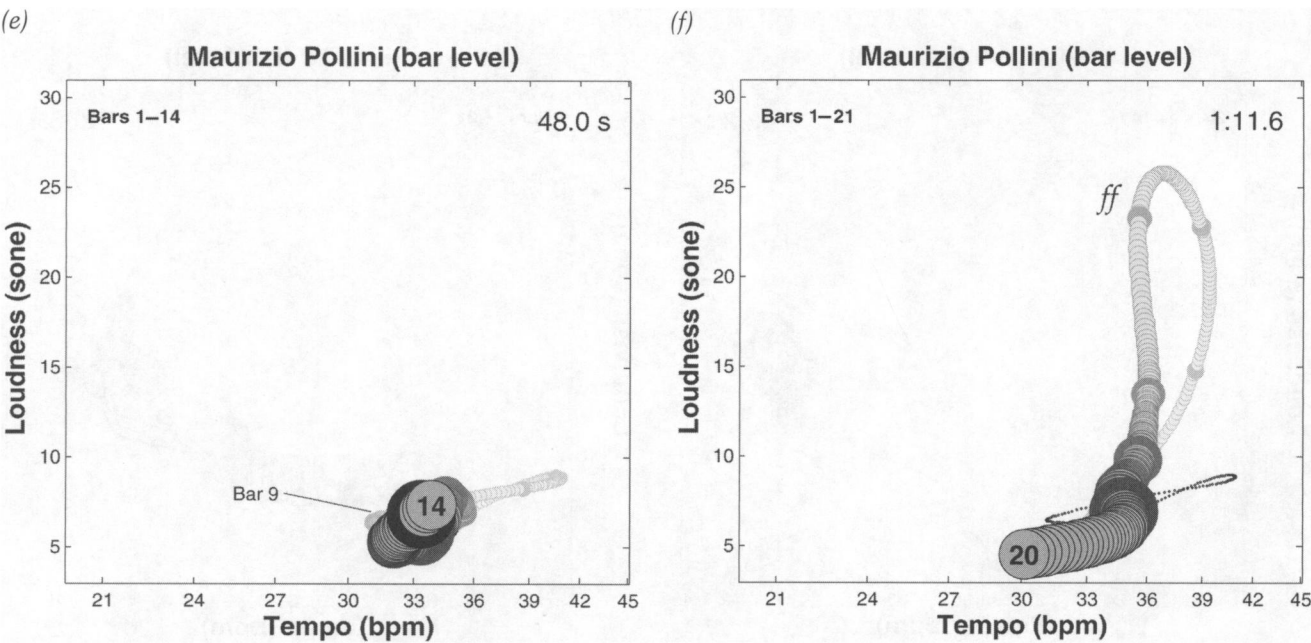
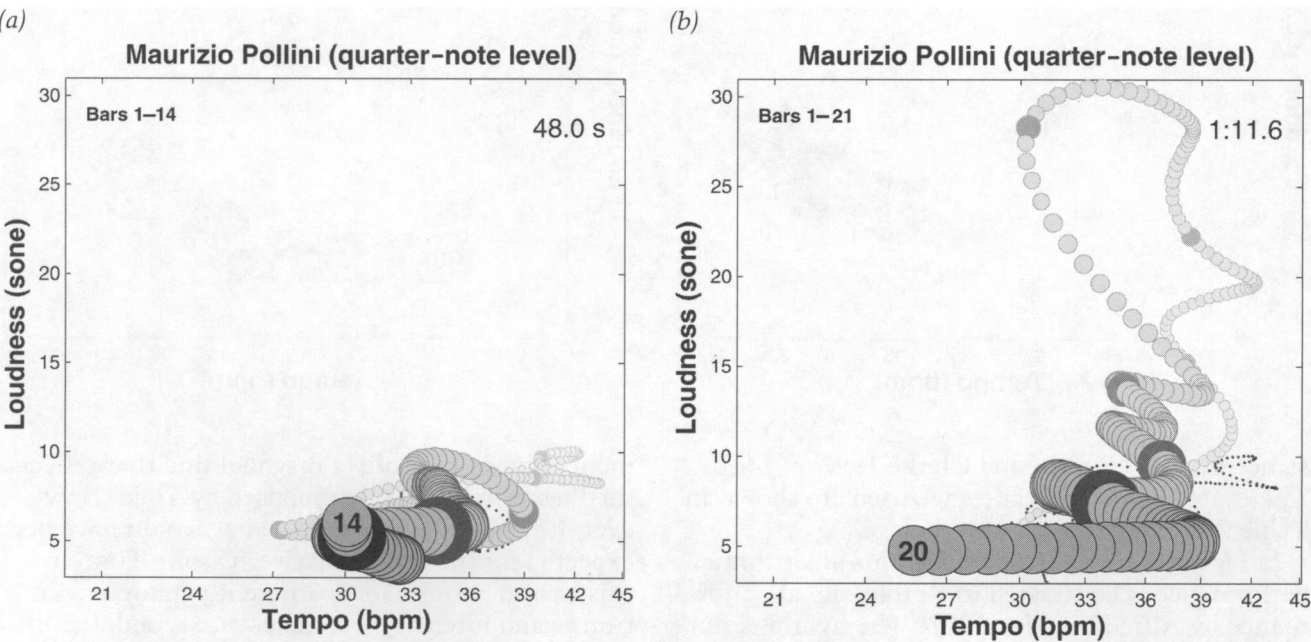


Figure 4. Expression trajectories over bars 1–14 (left) and 1–21 (right) of the beginning section of Chopin's Etude, Op. 10, No. 3, performed by Maurizio Pollini. The trajectories of the first 14 bars are still observable in the right figure as very faint lines. The x axes represent tempo in beats per minute (on a logarithmic scale) of a quarter note, and the y axes represent loudness measured in sones. The larger points represent the end of the excerpt, whereas in-



ear relationship between timing and dynamics in terms of MIDI velocity units is not linear when dynamics are expressed in terms of peak loudness units (in sones), as measured from the audio file recorded from the MIDI file using a standard software synthesizer. Still, the difference between the two performances is striking: Mr. Brendel divides the first eight bars into two phrases of equal length (Figure 6, upper right panel). Then, he spans a larger arch over the second eight bars (bars 9–16), where he only holds back his tempo increase briefly approximately at the middle of bar 10. On the other hand, Todd's model responds obediently to the symmetrical grouping structure. The asymmetric phrase concept of Mr. Brendel's interpretation becomes even more lucid with two-bar level smoothing (Figure 6, bottom panels). At this level, almost no other trajectory movement can be found in the Todd performance than a strict diagonal one. (To remark that Mr. Brendel mysteriously immortalized his initials A. B. in the two-bar smoothing would not belong in a scientific discourse!)

This comparison of a performance model to a professional performance was conducted not to prove Todd's model too simple, which the model was obviously planned to be, but to demonstrate the behavior of the two-dimensional display on real and artificial data and its strength of highlighting relations between the two performance parameters displayed.

Conclusions and Future Work

A novel approach for an integrated visualization of two performance parameters—tempo and loudness—was introduced, and its properties and behavior were demonstrated with examples from several expressive performances. This approach involves data acquisition from audio recordings and MIDI files, data reduction by continuous smoothing over a certain time window, and a two-dimensional display that can be viewed on a computer screen as an animation or as a snapshot for a particular excerpt of the performance.

The display shows both the tempo and the dynamic shaping of a performance and elucidates the

interaction between these two parameters. This visualization is not only useful for scientific research, but it is also intuitively comprehensible for musicians and audiences. It may therefore serve as a visible link between performance research and performance practice.

The number of performances analyzed so far is comparably small. However, two of the tendencies observed with this method are so striking that we decided to present them as preliminary results. First, the pianists of the investigated performances tend to approach the climax of a phrase by increasing tempo first and loudness slightly later. Second, they shape tempo and loudness within a phrase in a way that makes the expression trajectories move counterclockwise. We found this phenomenon also at smaller time windows, for example at the eighth-note level for the Chopin *Etude* for all three performances to a striking degree. We observed similar tendencies for performances of compositions by Chopin, Bach, and Beethoven as well; detailed analyses will be reported in future studies. These insights seem to be new for performance research and had not been expected, although similar tendencies can be found in the data plots of earlier performance studies (e.g., Gabrielsson 1987).

The findings of the present study imply some preliminary hypotheses that remain subject to further investigation. Pianists tend to increase loudness at tempo maxima, to decrease tempo at loudness maxima, to decrease loudness at tempo minima, and to increase tempo at loudness minima. This observation possibly mirrors a deep principle of shaping tempo and loudness. Further investigation will help to clarify the validity and scope of this principle.

Smoothing leads to a reduction of the amount and complexity of data. However, another advantage of smoothing is that with the choice of the window size, a specific granularity of data display can be determined. Expressive microstructure (e.g., the systematic variations in the Siciliano rhythm reported by Gabrielsson 1987) is the lowest level on which timing can be studied. Larger performance trends and developments will become clear at higher levels of abstraction.

The abstraction level in the performance display can be set by the smoothing window size. Multi-

Figure 5. Franz Schubert's
Impromptu in G-flat Ma-
jor, D. 899, No. 3, first
16 bars. The score was pre-
pared with computer soft-
ware by the authors after
Editio Musica Budapest.

Andante

pp legato

p

cresc.

p

cresc.

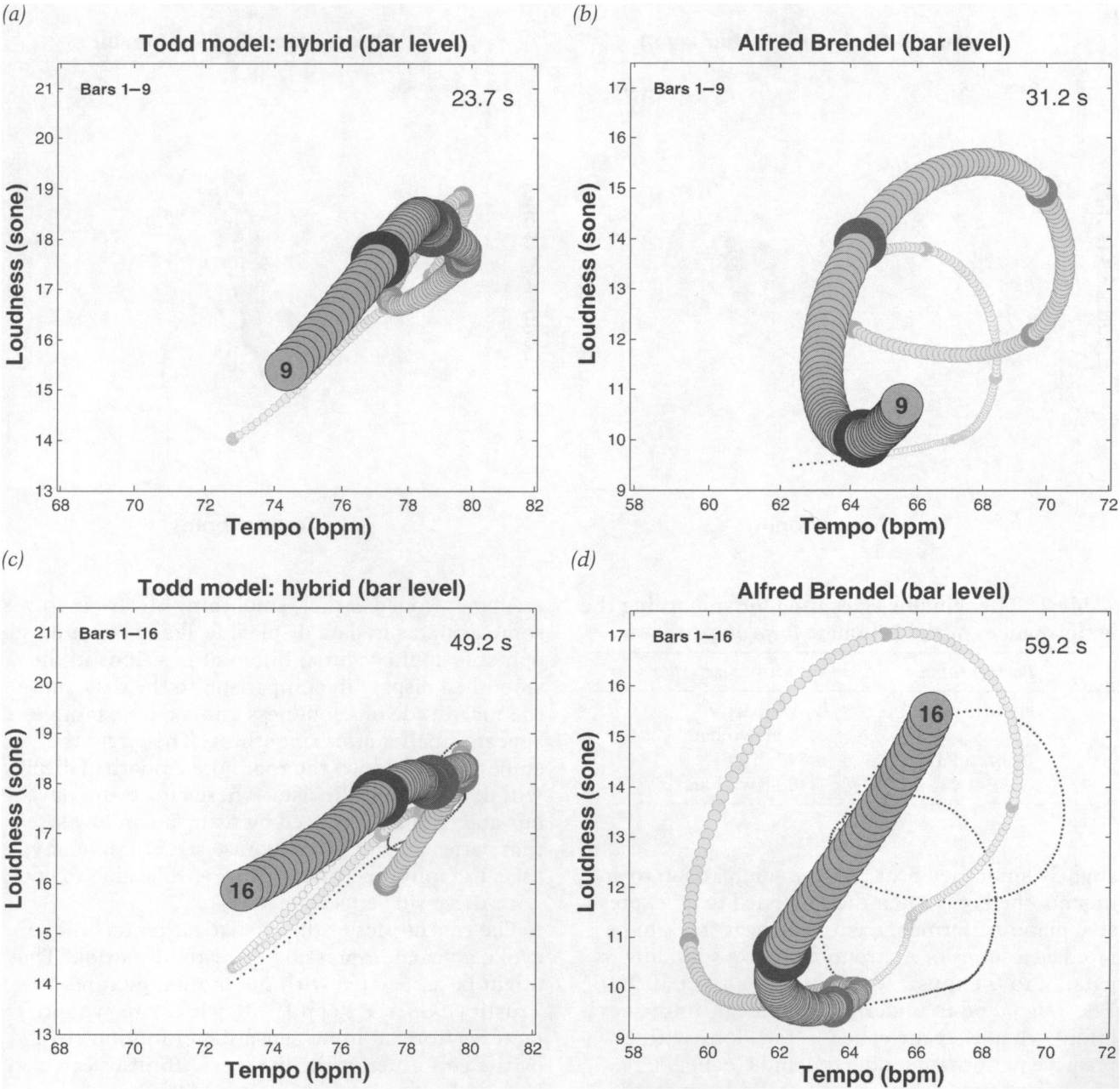
pp

dim.

cresc.

...

Figure 6. The first 16 bars of Franz Schubert's Impromptu in G-flat major, D. 899, No. 3, as "performed" by the Todd model (left column) and Alfred Brendel (right column). The x axis displays the tempo in beats per minute of a half note (because the time signature of that piece is 4/2). The axes of the figures have the same scale but show different ranges. In the top panels, the snapshot stopped shortly after the beginning of bar 9; in the middle and bottom panels, the snapshot stopped immediately after the beginning of bar 16. The upper two pairs of figures display data smoothed at the bar level. In the bottom panels, 16 bars of the same performances were smoothed at the two-bar level.



level displays of performance data try to avoid the choice for a particular time horizon. Examples for this type of display are *Dynagrams* (Langner et al. 2000) or *Oscillograms* (Langner, Kopiez, and Feiten 1998; Langner 2002). However, they only deal with a single performance parameter at a time.

The smoothing procedure could be supported by a perceptual hypothesis: the perceived tempo of an expressive performance is more stable than the tempo resulting from the played note onsets measured. This hypothesis was supported by recent research. In a preference task, listeners appreciated a

Figure 6. Continued.

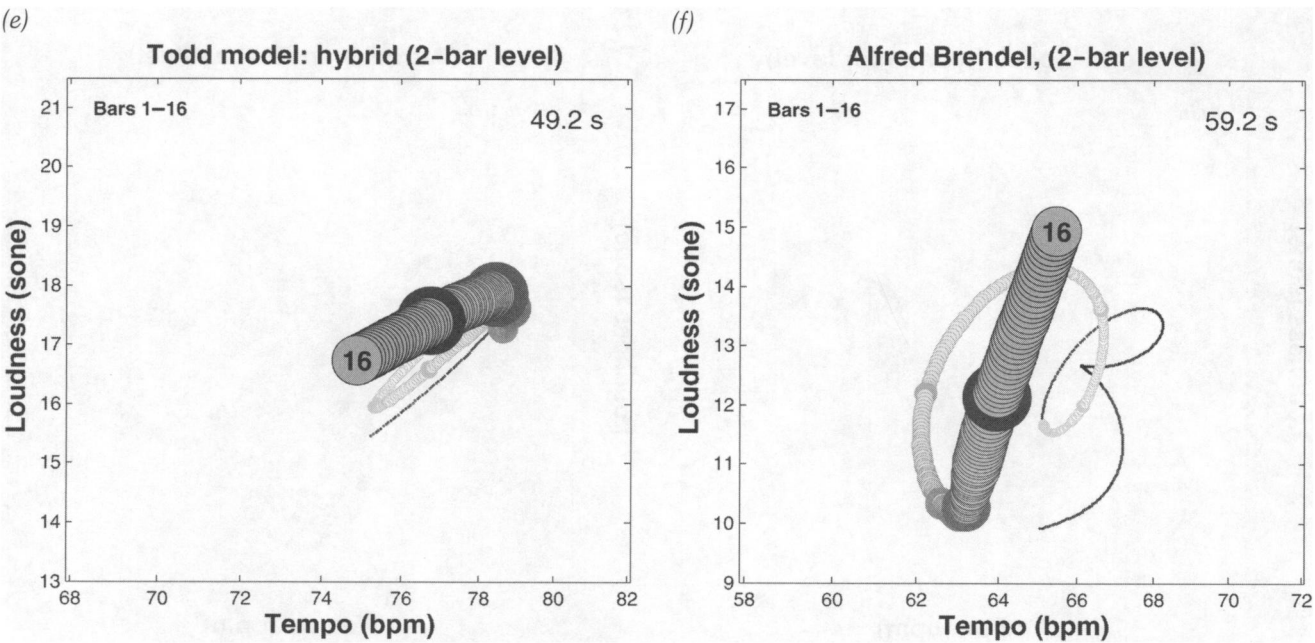


Table 2. The window sizes used for smoothing the performances of the Schubert *Impromptu*

Performance	Window Size (sec)
Hybrid	2.896 (bar)
	5.797 (two bars)
A. Brendel	3.690 (bar)
	7.380 (two bars)

slightly smoothed beat track in comparison to an unsmoothed one when they listened to an expressive piano performance with the beat track to be rated as a series of metronome clicks sounding in parallel to the music (Cambouropoulos et al. 2001). They also tend to underestimate local timing deviations when they are asked to tap along with expressive performances (Dixon and Goebel 2002). These findings, although still preliminary, could support the smoothing from a perceptual point of view. Still, it remains to be investigated what smoothing windows (type and size) correspond best to a perceptual or mental representation of an expressive performance and whether the same integrating mechanisms apply to the continuous perception of dynamics.

As mentioned earlier, smoothing always involves some artifacts in data display. Peaks of larger developments could occur at different positions in the smoothed display in comparison to the data, and the magnitude of a loudness climax, for example, appears smaller after smoothing. The larger the smoothing window, the more the smoothed display will deviate from the data. These effects are obvious and could be reduced by using a window size that varies with musical phrase structure. In any case, the informed investigator will be able to cope with these side effects.

The trajectories of this visualization technique evoke a visual impression of gestural motion. They might be associated with the motion pictures by Truslit (1938; see Repp 1993), which are synoptic pictures from an inner sensation of motion either by the performer or the listener. Affinities can also be found to shapes coming from Manfred Clynes's sentograph (e.g., Clynes 1983). Similarities of both Truslit and Clynes to the expression trajectories may be seen as merely incidental: they both reflect some perceptual properties of expressive music performance. Our two-dimensional display, to the contrary, picture measured performance data only. These similarities possibly reflect a deeper connec-

tion. What they do have in common is an affinity for the phenomenon of motion, which is very meaningful for both the production and the perception of music.

We envisage a range of future applications of this technique. Besides the described analysis of large numbers of performances by famous musicians, it could also be used in music teaching to clarify certain high level developments of students's performances. Especially for this purpose, a real-time implementation would be very helpful. A first implementation of a real-time system was done by Simon Dixon (Dixon, Goebel, and Widmer 2002), who used a real-time tempo-tracking algorithm that processes audio data directly without using any higher-level knowledge of the musical structure.

The idea of the two-dimensional space could also be reversed, using it as an interactive control of music performance. The user moves the head of the trajectory using the computer mouse and the performance unfolds according to the mouse movement in the space. A first prototype of such a performance control system was implemented in Java by students at the Austrian Research Institute for Artificial Intelligence in Vienna. A likewise implementation was developed using a different two-dimensional control space as an user interface for controlling morphing between different emotional states in synthesized expressive performances (Canazza et al. 2000). Their control space is called a *perceptual expressive space* and spans a space between adjectives like "hard," "soft," "light," and "heavy."

Further work is needed to understand the meaning of certain repeating shapes of the trajectories. The analyzed performances could be rated by expert listeners in great detail; that is, they could be asked to indicate which sections they like and which they dislike. These qualitative data could be used to evaluate the meaning of the corresponding shapes of the expressive trajectories (Langner and Goebel in preparation). Furthermore, it would be interesting to compare the performance trajectories with emotional responses of musically trained listeners, for example by using Emery Schubert's method of two-dimensional emotional space (Schubert 1996, 1999). Another approach could be to

quantitatively analyze the two-dimensional time series as such with modern artificial intelligence methods (using self-organizing maps; see Pampalk, Widmer, and Chan in press). Such a method allows one to pinpoint pianist-specific performance characteristics over large data sets and to name intrinsic peculiarities of particular famous pianists.

Acknowledgments

This research was in part supported by project Y99-INF, sponsored by the Austrian Federal Ministry of Education, Science, and Culture (BMBWK) in the form of a START Research Prize. The Austrian Research Institute for Artificial Intelligence acknowledges basic financial support from the Austrian Federal Ministry for Education, Science, and Culture as well as from the Austrian Ministry for Transport, Innovation, and Technology. Thanks are due to Luke Windsor and Eric Clarke for providing the Todd hybrid performance MIDI file of the Schubert *Impromptu*, and to Simon Dixon, Elias Pampalk, Giampiero Salvi, Asmir Tobudic, and Gerhard Widmer for providing generous help during work on this article. Finally, we are grateful to Roberto Bresin and an anonymous reviewer for their comments on an earlier draft.

References

- Brendel, A. 1997. (Recorded 1972.) *Schubert: The Complete Impromptus; Moment Musicaux*. London: Philips Classics, 456 061–2.
- Cambouropoulos, E., et al. 2001. "Human Preferences for Tempo Smoothness." In H. Lappalainen, ed. *Proceedings of the VII International Symposium on Systematic and Comparative Musicology, III International Conference on Cognitive Musicology*. Jyväskylä, Finland: University of Jyväskylä, pp. 18–26.
- Canazza, S., et al. 2000. "Audio Morphing Different Expressive Intentions for Multimedia Systems." *IEEE Multimedia* 7(3):79–83.
- Clynes, M. 1983. "Expressive Microstructure in Music, Linked to Living Qualities." In J. Sundberg, ed. *Studies of Music Performance*. Stockholm: Royal Swedish Academy of Music, pp. 76–181.

- Dixon, S. E. 2001a. "Automatic Extraction of Tempo and Beat from Expressive Performances." *Journal of New Music Research* 30(1):39–58.
- Dixon, S. E. 2001b. "An Interactive Beat Tracking and Visualisation System." *Proceedings of the 2001 International Computer Music Conference*. San Francisco: International Computer Music Association, pp. 215–218.
- Dixon, S. E., and W. Goebel. 2002. "Pinpointing the Beat: Tapping to Expressive Performances." In C. Stevens, D. Burnham, G. McPherson, E. Schubert and J. Renwick, eds. *Proceedings of the 7th International Conference on Music Perception and Cognition, Sydney 200*. Adelaide, Australia: Causal Productions, pp. 617–620.
- Dixon, S. E., W. Goebel, and G. Widmer. 2002. "Real Time Tracking and Visualisation of Musical Expression." In C. Anagnostopoulou, M. Ferrand, and A. Smaill, eds. *Proceedings of the 2nd International Conference on Music and Artificial Intelligence*. Berlin: Springer, pp. 58–68.
- Gabrielsson, A. 1987. "Once Again: The Theme from Mozart's Piano Sonata in a Major (K.331)." In A. Gabrielsson, ed. *Action and Perception in Rhythm and Music*. Stockholm: Royal Swedish Academy of Music, pp. 81–103.
- Gabrielsson, A. 1999. "Music Performance." In D. Deutsch, ed. *Psychology of Music*, 2nd ed. San Diego: Academic Press, pp. 501–602.
- Goebel, W. 2001. "Melody Lead in Piano Performance: Expressive Device or Artifact?" *Journal of the Acoustical Society of America* 110(1):563–572.
- Goebel, W., and R. Bresin. 2001. "Are Computer-Controlled Pianos a Reliable Tool in Music Performance Research? Recording and Reproduction Precision of a Yamaha Disklavier Grand Piano." In C. L. Buyoli and R. Loureiro, eds. *Workshop on Current Research Directions in Computer Music*. Barcelona, Spain: Audiovisual Institute, Pompeu Fabra University, pp. 45–50.
- Goebel, W., and S. E. Dixon. 2001. "Analyses of Tempo Classes in Performances of Mozart Piano Sonatas." In H. Lappalainen, ed. *Proceedings of the VII International Symposium on Systematic and Comparative Musicology, III International Conference on Cognitive Musicology*. Jyväskylä, Finland: University of Jyväskylä, pp. 65–76.
- Heijink, H., et al. 2000. "Make Me a Match: An Evaluation of Different Approaches to Score-Performance Matching." *Computer Music Journal* 24(1):43–56.
- Juslin, P. N., A. Friberg, and R. Bresin. 2002. "Toward a Computational Model of Expression in Performance: The GERM Model." *Musicae Scientiae Special Issue* 2001–2002:63–122.
- Langner, J. 2002. *Musikalischer Rhythmus und Oszillation. Eine theoretische und empirische Erkundung*. Frankfurt: P. Lang.
- Langner, J., and W. Goebel. 2002. "Representing Expressive Performance in Tempo-Loudness Space." *Escom 10th Anniversary Conference on Musical Creativity*. Liège, Belgium: Université de Liège
- Langner, J., and W. Goebel. In preparation. "Aesthetic Performance Evaluations in the Context of Performance Research: An Assessment of Two-Dimensional Performance Trajectories."
- Langner, J., R. Kopiez, and B. Feiten. 1998. "Perception and Representation of Multiple Tempo Hierarchies in Musical Performance and Composition: Perspectives from a New Theoretical Approach." In R. Kopiez and W. Auhagen, eds. *Controlling Creative Processes in Music*. Frankfurt: P. Lang, pp. 13–35.
- Langner, J., et al. 2000. "Real-Time Analysis of Dynamic Shaping." In C. Woods et al., eds. *Proceedings of the 6th International Conference on Music Perception and Cognition*. Keele, UK: Keele University, Department of Psychology, pp. 452–455.
- Lerdahl, F., and R. Jackendoff. 1983. *A Generative Theory of Tonal Music*. Cambridge, Massachusetts: MIT Press.
- Palmer, C. 1997. "Music Performance." *Annual Review of Psychology* 48:115–138.
- Pampalk, E., A. Rauber, and D. Merkl. 2002. "Content-Based Organization and Visualization of Music Archives." *Proceedings of the 10th ACM International Conference on Multimedia*. Juan les Pins, France: ACM, pp. 570–579.
- Pampalk, E., G. Widmer, and A. Chan. In press. "A New Approach to Hierarchical Clustering and Structuring of Data with Self-Organizing Maps." *Intelligent Data Analysis Journal* 8(2).
- Pollini, M. 1985. (Recorded 1972.) *Chopin: Etudes*. Hamburg: Deutsche Grammophon Gesellschaft 413 794–2.
- Repp, B. H. 1993. "Music as Motion: A Synopsis of Alexander Truslit's 'Gestaltung und Bewegung in der Musik'." *Psychology of Music* 21:48–72.
- Schubert, E. 1996. "Measuring Temporal Emotional Response to Music Using the Two-Dimensional Emotion Space." In B. Pennycook and E. Costa-Giomi, eds. *Proceedings of the 4th International Conference for Music Perception and Cognition*. Montreal: Faculty of Music, McGill University, pp. 263–268.
- Schubert, E. 1999. "Measuring Emotion Continuously: Validity and Reliability of the Two-Dimensional

-
- Emotion-Space." *Australian Journal of Psychology* 51(3):154–165.
- Slaney, M. 1999. "Quicktime Tools for MATLAB." *Interval Technical Report* 1999–066. Palo Alto, California : Interval Research Corporation.
- Todd, N. P. M. 1992. "The Dynamics of Dynamics: A Model of Musical Expression." *Journal of the Acoustical Society of America* 91(6):3540–3550.
- Truslit, A. 1938. *Gestaltung und Bewegung in der Musik*. Berlin-Lichtenfelde: Chr. Fiedrich Vieweg.
- Widmer, G. 2001. "Using AI and Machine Learning to Study Expressive Music Performance: Project Survey and First Report." *AI Communications* 14(3):149–162.
- Windsor, W. L., and E. F. Clarke. 1997. "Expressive Timing and Dynamics in Real and Artificial Musical Performances: Using an Algorithm as an Analytical Tool." *Music Perception* 15(2):127–152.
- Zwicker, E., and H. Fastl. 2001. *Psychoacoustics: Facts and Models*, 2nd edition. Berlin: Springer-Verlag.