EXPLORING EXPRESSIVE PERFORMANCE TRAJECTORIES:
SIX FAMOUS PIANISTS PLAY SIX CHOPIN PIECES

Werner Goebl[1], Elias Pampalk[1], and Gerhard Widmer[1,2]

[1] Austrian Research Institute for Artificial Intelligence (ÖFAI), Vienna

[2] Dept. of Medical Cybernetics and Artificial Intelligence (IMKAI), Medical University of Vienna

ABSTRACT

This paper presents an exploratory approach to analyzing large amounts of expressive performance data. Tempo and loudness information was derived semi-automatically from audio recordings of six famous pianists each playing six complete pieces by Chopin. The two-dimensional data was segmented into musically relevant phrases, normalized, and smoothed in various grades. The whole data set was clustered using a novel computational technique (i.e., aligned self-organizing maps) and visualized via an interactive user interface. Detailed cluster-wise statistics across pianists, pieces, and phrases gave insights into individual expressive strategies as well as common performance principles.

1. INTRODUCTION

In the last decades, research on music performance has grown considerably (see amount of references listed in Gabrielsson, 1999, 2003). However, studies in that field either restricted themselves to a few bars of music and to one expressive parameter at a time (mostly timing, e.g., Repp, 1992, 1998), or to a few individual performances (e.g., Sloboda, 1985; Windsor and Clarke, 1997), in order to be able to interpret the vast amounts of expressive data that even a single piano performance yields.

2. AIMS

In this paper we describe an exploratory approach to analyzing large amounts of expressive performance data obtained from audio recordings, i.e., six complete romantic piano pieces played by six famous concert pianists, in order to disclose certain expressive principles of particular performers or expressive constraints of certain phrases determined by the score or by convention. To this end, we used novel computational techniques (i.e., aligned self-organizing maps) to cluster the data. The goal was to explore the expressive tempo-loudness phrase patterns and to determine inherent typicalities of individual performers and certain phrases.

3. METHOD
3.1. Material & data acquisition

The analyzed data were commercially available audio recordings of 3 Nocturnes (op. 15, No. 1 and both op. 27) and 3 Préludes (op. 28, No. 4, 8, and 17) by Frédéric Chopin, played by 6 renowned pianists: Claudio Arrau, Vladimir Ashkenazy, Adam Harasiewicz, Maria João Pires, Maurizio Pollini, and Artur Rubinstein. The 36 performances, more than two hours of music, were "beat-tracked" — that is, the onset times of all performed events (notes or chords) at a particular (low) metrical level (e.g., a sixteenth note) were determined with the aid of a purpose-built computational tool that performs automatic beat-tracking (Dixon, 2001b) and allows interactive and iterative manual correction of the obtained results (see Dixon, 2001a). For each measured onset, an overall loudness value (in sone) was determined from the audio signal to get a rough measure of dynamics.

The six pieces were dissected into small segments of around 1–2 bars length, according to their musical phrase structure (by the first author). All phrases of the six pieces by six performers resulted in over 1600 two-dimensional time series each representing the tempo-loudness performance trajectory of one phrase played by one pianist. The two-dimensional data are arranged visually with the tempo information on the x axis and the loudness information on the y axis (the "Performance Worm," Dixon et al., 2002). The phrase segments had varying length, ranging from 3 to 25 tempo-loudness points or durations from 0.5 to 25.7 seconds. As comparing extremely short phrases (e.g., with 3 data pairs) with extremely long phrases (e.g., 25 data pairs) does not make sense, extreme outliers were removed from the data. Only phrases with a length between 5 and 15 data pairs and durations between 2 and 10 s were included in the experiment. Finally, 1216 phrase segments went into the experiment. In order to be able to compare the phrase segments to each other, they had to have exactly the same number of data pairs. Therefore, all phrase segments were interpolated so that each phrase segment contained 25 data pairs (cubic interpolation).

3.2. Clustering & data normalization

The clustering technique used in this study is designed to give the researcher the opportunity to predefine various potentially interesting sets of input parameters. After computing all combinations of the input parameter sets, their impact on the clustering process can be interactively explored in a graphical user interface. For our 6×6 data set, we defined three different parameters: the type of normalization applied to the data in order to make them comparable, the degree of smoothing, and the weighting between tempo and loudness.

We defined 5 forms of normalization that may be seen at three different levels (see Fig. 1, top left). No normalization was applied at the first level, the second level normalizes by subtracting the mean and the third level normalizes by dividing by the mean. Thus at the second level, we compare absolute changes with each other (in beats per minute or in sone); at the third level relative changes (in percent). For the second and the third level, we normalized either by the mean of a piece (global mean) or by the mean of an individual phrase segment (local mean).

The amount of smoothing applied to the data corresponds to level of detail at which the researcher wants to examine the performance data (Langner and Goebl, 2003). Exploring unsmoothed performance data reveals every single accent or delayed note, while examining smoothed data gives insight into larger-scale performance developments (e.g., at bar level). We chose five different levels of
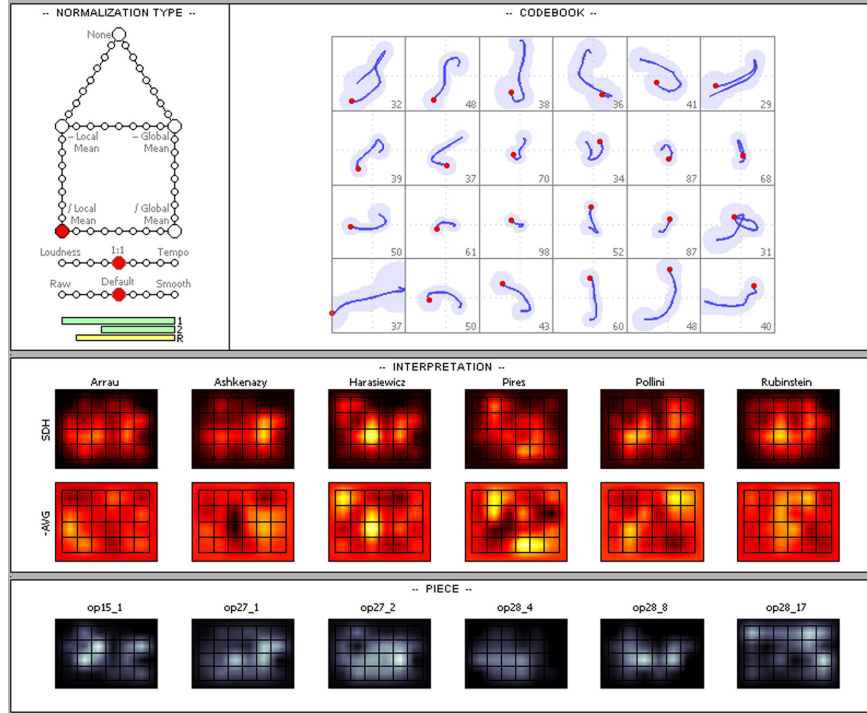
Figure 1: Screen shot of the interactive data viewer. The current settings (see navigation unit, top left) display the data scaled to the local mean, with equal tempo-loudness weighting and medium smoothing (0.75 beats either side). The axes of the codebook (top right) range from 50% to 145% of the local mean tempo on the x axes and 60% to 167% of the local mean loudness on the y axes. A brighter color in the smoothed data histograms correspond to more instances in a cluster.

smoothing: none, and smoothing windows corresponding to mean performed durations of 0.5, 0.75, 1, or 2 beats either side. A smoothing window of 2 beats either side denotes a Gaussian window with the mean performed duration of 4 beats from the left to the right point of inflection (Langner and Goebl, 2003).

The whole data set with all its different parametrizations was input to the aligned self-organizing maps algorithm (aligned-SOM, see Pampalk, 2003). A conventional SOM groups data into a pre-defined number of clusters that are displayed on a two-dimensional map so that all elements of a data cluster are similar to each other and similar clusters are located close to each other on the map (Kohonen, 2001). The iterative SOM algorithm is usually randomly initialized and stopped when a convergence criterion is fulfilled. The aligned-SOM algorithm takes various potentially interesting parametrizations of the same data set as input (defined by the researcher). It calculates for each parametrization a SOM that is explicitly forced to form its clusters at the same locations as the adjacent SOM with similar parameters. At the end, the user can continuously vary input parameters (in our case normalization coefficients, smoothing window, or tempo-loudness weighting) and study their influence on the clustering process by examining the gradual changes in the aligned maps.

3.3. Visualization

The results are visualized as an interactive HTML page (Pampalk et al., 2003). A screenshot is displayed in Figure 1. The display is structured in three parts. The navigation unit (located in the upper-left corner) controls the 5 normalization forms (the corners of the "house"), the tempo-loudness weighting, and the amount of

smoothing. The user controls the display by moving the mouse over the house circles. On the right, a two-dimensional map of obtained clusters is displayed (the codebook), each with its prototype (mean) performance trajectory, its variance (shading), and the number of contained phrase segments. Underneath, it shows frequency distributions over the codebook by performer (first two rows) and by piece (third row). They are visualized as smoothed data histograms (SDH, Pampalk et al., 2002). To elaborate differences between the six pianists' SDHs, we also show their SDHs after subtracting the average SDH (second row). This part of the display is of particular interest, because it shows whether a pianist uses a certain performance pattern particularly often or seldom.

In order to further explore which phrase segments were included in one particular cluster, we extended the user interface with a so-called "cluster inspector." It displays all performance segments of that specific cluster preceded by histograms by pianists, pieces, and phrases (see Figure 2). The user can then click on each phrase segment and listen to the music.

4. RESULTS & DISCUSSION

This novel approach to exploring expressive performance properties yields a variety of interesting results. Due to the limited space here, we have to restrict ourselves to the most striking ones. We invite the reader to follow the results described in this section on the interactive web interface.[1]

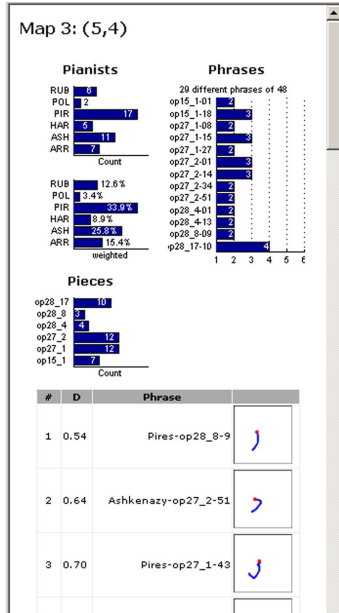
[1] http://www.oefai.at/~werner.goebl/icmpc8/

Figure 2: Screenshot of the "cluster inspector" displaying basic statistics of the cluster in the fifth column (5) in the fourth row (4) of the codebook as shown in Fig. 1. Histograms are shown for pianists (also weighted by the distance *D* from the prototype), pieces, and phrases.

4.1. Unnormalized data comparison

Examining the codebook of the unnormalized data (peak of the house), it becomes apparent that the segments of a particular piece fall typically into certain clusters that are different from those of other pieces. Each piece contains typical tempo and loudness progressions and ranges determined by the score that dominate the clustering process. In the two pieces that have a contrasting middle section (op. 15 No. 1 and op. 27 No. 1), the piece-wise data histograms clearly show two separate bright areas representing respectively the softer and slower outer parts and the louder and faster middle sections.

Although this unnormalized data basically clusters along piece boundaries, certain expressive strategies stand out that are characteristic of individual pianists. A very obvious example is Pollini playing op. 28 No. 17. He dominates (45%) the bottom-right cluster (column 6, row 4) that contains primarily trajectories with a clear acceleration-deceleration pattern within a quite narrow loudness range.

This unnormalized view reflects also all intrinsic recording properties, especially the volume level of each recording. When inspecting solely the loudness dimension (the tempo-loudness weighting control slided to the left), Arrau and Harasiewicz show a considerable lack of very soft phrase segments that three other pianists (Pires, Pollini, Rubinstein) represent strongly. It is hard to tell whether this effect is due to recording level or to particular expressive strategies.

4.2. Normalized data comparison

To be able to compare phrase segments with different basic tempo and loudness, we normalized the data in various ways as described above. Due to space limitations we focus especially on one normalization type, namely dividing by the local mean (lower-left corner
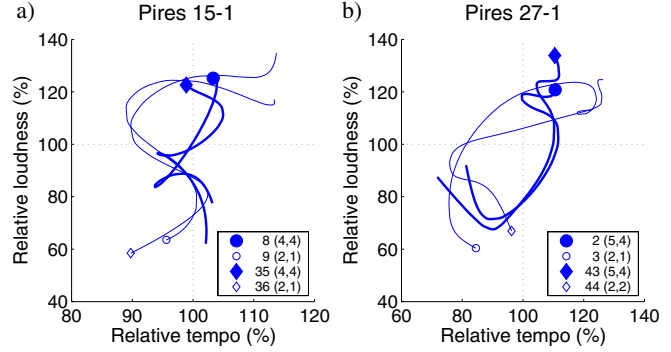

Figure 3: Pairs of consecutive phrase segments played by Pires, where each first segment depicts an upward, opening tendency (bold texture) and each second a downward, closing shape (light). (a) op. 15 No. 1, phrase 8–9 and 35–36. (b) op. 27 No. 1, phrase 2–3 and 43–44 (theme and its recurrence). The legends indicate the cluster (column,row) where each phrase segment can be found (see Fig. 1). The second segments do not continue exactly where the first ended, because each one is scaled to its local mean.
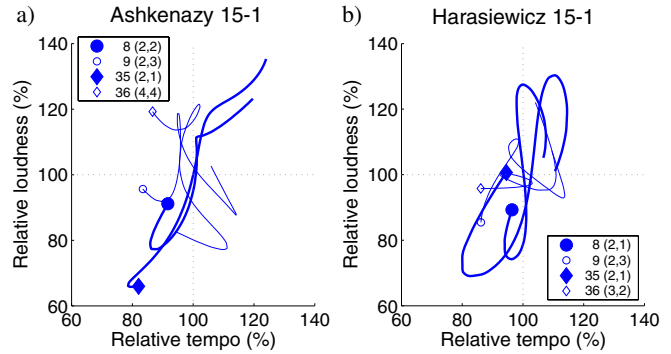

Figure 4: Excerpt from op. 15 No. 1, phrases 8–9 (bars 15–18) and 35–36 (bars 63–66), performed by Ashkenazy (a) and by Harasiewicz (b). The clusters that contained the individual phrases are specified in the legends.

of the navigation house; Figure 1 shows this parametrization), thus comparing deviations relative to the local mean (in percent).

The most striking observation from this view is the apparently antagonistic expressive strategies of Pires and Pollini. Pollini's SDH exhibits peaks where Pires' has minima and vice versa (Fig. 1). Pires' SDH has two very bright areas: one at the center-bottom (4–5,4) and the other on the top-left side (2,1–2). As a typical phrase shape is characterized by an initial acceleration and loudness increase towards the middle and a slowing down and decrescendo towards its end (e.g., Todd, 1992), these 4 clusters can be seen as two parts of one single phrase. The first would be the opening part (4–5,4, see also Fig. 2) with an acceleration and crescendo, the other the closing part with a movement towards bottom-left corner of the panel (ritardando, diminuendo). In Fig. 3, four examples of such consecutive segments, forming one single larger-scale phrase and performed by Pires are shown.

Fig. 3a shows phrases 8–9 (bars 15–18) of op. 15 No. 1 and parallel section in the repetition of the first part (phrases 35–36, bars 63–66). Pires plays these four bars always under one arch with the apex in the middle. This clearly sets her apart from the other pianists, who follow quite opposite strategies: e.g., Ashkenazy plays the first two bars in a gradual diminuendo and ritardando and builds up loudness in the second two bars (Fig.4a). Another strategy for
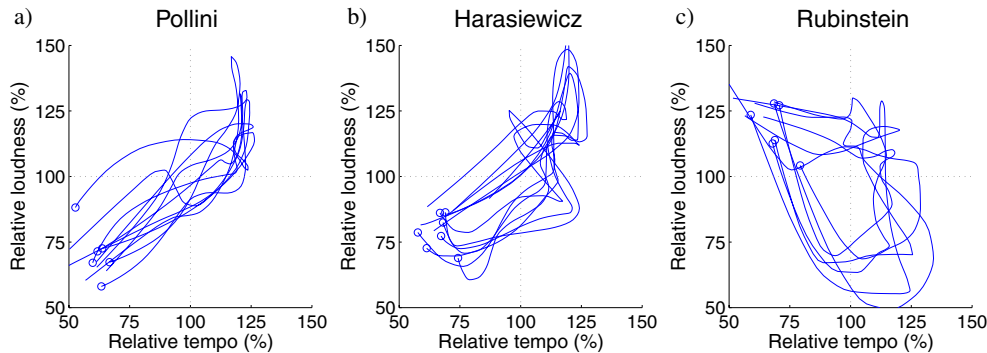
Figure 5: The Coda section from op. 28 No. 17, phrases 35–42 (bars 65–81), played by Pollini (a), Harasiewicz (b), and Rubinstein (c).

this excerpt is shared by Harasiewicz, Pollini, and Rubinstein (see as an example Harasiewicz, Fig.4b), who place the apex of the first two bars at the third beat of the first bar and close the first phrase with a strong decrescendo and diminuendo. The upwards curve at the end of the first phrase is due to the accented high note of the third bar. The second part of that excerpt is performed in a gradual retarding descent only interrupted by the little ornament in the fourth bar (bar 18 and 66, respectively).

Another example of Pires' tendency of phrasing over two segments is depicted in Fig. 3b. This example is a 4-bar excerpt of the theme and its recurrence of op. 27 No. 1 (bars 3–6 and 86–89, respectively). With only a short hesitation in the beginning of the second bar, she plays towards the fourth bar in order to relax there extensively. This interpretation is quite different from that of her colleagues. Arrau (both) and Ashkenazy, Pollini, and Rubinstein (one each) show phrases from cluster 6,2 as their performance from the first phrase (a clear dynamic apex at the second bar). Harasiewicz (both), Pollini, and Rubinstein (one each) have phrases from cluster 5,2 at the first two bars — a cluster with a similar dynamic apex at the second bar as cluster 6,2, but with far less temporal hesitation. The second two bars of this example is typically realized by phrases from cluster 2,4 (Arrau, Harasiewicz, and Pollini) — a clearly retarding phrase shape.

As another example of different expressive strategies, the coda section from op. 28 No. 17 has to be mentioned (phrases 35–42 or bars 65–81). The main theme is repeated here in a very soft (*pp sotto voce*) and somehow distant atmosphere with a *sforzato* bass at the beginning of each two-bar phrase throughout the whole section. Three pianists show a typical two-bar phrasing strategy (Pollini, Harasiewicz, and Rubinstein, see Fig. 5) that repeats with only a few exceptions through the whole section. Interestingly, each pianist has his phrase segments for this section in one particular cluster: Pollini (6,1),[2] Harasiewicz (1,1),[3] and Rubinstein (5,1).[4] Pollini's and Harasiewicz's shapes (Fig. 5a and b) show both a diagonal, "Todd-like" (accelerando–ritardando and louder–softer pattern, Todd, 1992) trajectory. Harasiewicz's shapes include typically a small loop on the top-right side, a result of a faster decrescendo than the temporal descent towards the end of the phrase. Somehow contrary appear Rubinstein's shapes (Fig. 5c) which depict a clockwise rotation. This shape is due to Rubinstein's extremely literal realization of the score. He played the bass tones very strongly while the actual melody remains very soft and in the background. This strategy places the dynamic apex at the beginning of each two-bar segment, while the two others had it in the middle. The other

pianists (Arrau, Ashkenazy, and Pires) followed mixed strategies. Ashkenazy phrases this section clearly over 8 bars (65–72, and later even longer, 72–84).

Apart from the above reported diversities, a considerable number of common phrase shapes were observed as well. Two clusters (containing fewer phrase segments than the expected $1215/24 = 50.6$) have to be mentioned: 4,1 and 1,4. Two examples of phrases are given in Figure 6a and b in which all 6 pianists followed a similar strategy that caused their phrases to be arranged in the same cluster. To illustrate artifacts of the clustering behavior, Figure 6c shows a phrase that all 6 pianists played extremely similarly, but due to particular constraints of that phrase. It depicts 6 performances of bars 23–24 from op. 28 No. 4 (all cluster 3,1), containing only two chords and a long rest in between. This is an extreme case in which specifications from the score dominated the shape of the trajectories so that possible individual performance characteristics did not become apparent.

4.3. Problems and shortcomings

Although our present approach, which focusses on timing and loudness, captured essential expressive information about 36 complete performances and used novel clustering techniques to reduce complexity of the data, some potential shortcomings need to be discussed. First, only overall loudness was measured from the sound file (cf. Repp, 1999), disregarding the loudness of individual voices. This measure depends strongly on the texture of the music. For example, events with a melody note will be considerably louder than those with accompaniment only, which is in fact reflecting constraints from the score rather than properties of a particular performance. Second, performance information was determined at a defined track level, a procedure that sometimes disregarded potentially important events in some pieces (e.g., op. 28 No. 8 was tracked in quarter notes, thus ignoring 7 onsets between each tracked onset). As a third and very common problem, we mention the measurement error here. Earlier studies revealed that it lies within a range of ± 10 ms — sufficiently precise for the present purpose (Goebl and Dixon, 2001).

The fourth, probably most influential factor is data interpolation for comparison reasons. This processing step is necessary to compare phrases of varying length. However, outliers in the data (e.g., a local lengthening of one single note) in combination with interpolation can disassociate the trajectory from the actual performance. In some cases the trajectory may exhibit a strong ritardando that can't be perceptually found in the performance, because it stemmed from a single delayed event that is not perceived as a ritardando. The input variable smoothing condition makes the cluster prototypes smaller with growing smoothing window; still the reported main effects (e.g., Pires–Pollini contrast) remain present.

---

[2]Pollini: phrases 35–37 and 39–42 in cluster 6,1; phrase 38 in 1,1.

[3]Harasiewicz: phrases 36–42 in cluster 1,1; phrase 35 in 6,1.

[4]Rubinstein: phrases 35, 36, 38–40, and 42 in cluster 5,1; phrases 37 and 41 in 3,1. Also there, the shapes are dominated by the loud bass tone.
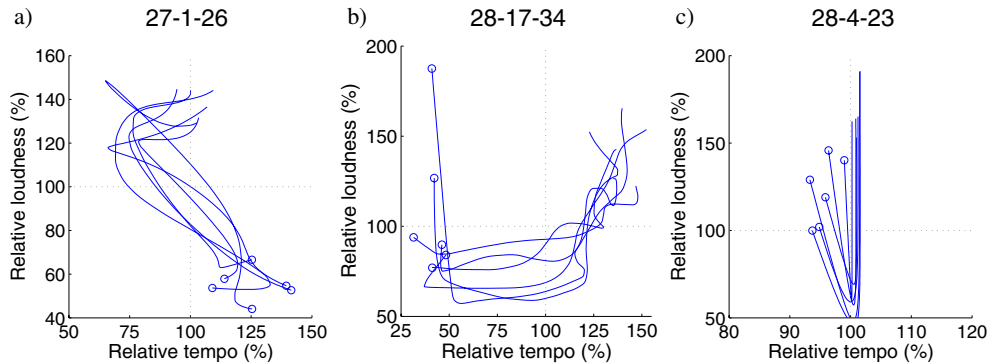
Figure 6: Commonalities between all six pianists. (a) The left panel shows all six performances of phrase 26 from op. 27 No. 1 (the descent from a *fff* apex and a written *ritenuto*, while speeding up to a new and softer *agitato*). They were all from cluster 3,1. (b) The middle panel displays phrase 34 from op. 28 No. 17 (cluster 1,4), the transition to the coda with a notated decrescendo bar. (c) Right panel: Artifact from op. 28 No. 4 phrase 23 (bar 23–24). This phrase contains only two chords and a long rest in between.

The present two-dimensional data representation simultaneously uses information from only two performance parameters, though essential ones. It completely disregards information on articulation and pedalling, as well as information about the score. In trying to understand the "worm" shapes while listening to the music, sometimes the perception of tempo and loudness progression gets confounded with the perception of pitch and melody; i.e., it is hard to listen only to the two displayed parameters totally independently of other variables. We consider incorporating other score and performance information for future research.

5. CONCLUSION

We reported on an exploratory approach to analyzing a large corpus of expressive tempo and loudness data derived from professional audio recordings of more than two hours of romantic piano music. It revealed both diversities and commonalities among performers. The advantage of our approach is that it deals with large amounts of data and it reduces their complexity and visualizes them via an interactive user interface. Nevertheless, it is a quite complex approach, because the researcher still has to verify "manually" whether observed effects are musically relevant, or if they are simply artifacts such as some of those described above.

REFERENCES

Dixon, S. E. (2001a). An interactive beat tracking and visualisation system. In Schloss, A., Dannenberg, R., and Driessen, P., editors, *Proc. of the 2001 ICMC*, pages 215–218. Int. Comp. Mus. Assoc., San Fransico.

Dixon, S. E. (2001b). Automatic extraction of tempo and beat from expressive performances. *J. New Music Res.*, 30(1):39–58.

Dixon, S. E., Goebl, W., and Widmer, G. (2002). The Performance Worm: Real time visualisation based on Langner's representation. In Nordahl, M., editor, *Proc. of the 2002 ICMC, Göteborg, Sweden*, pages 361–364. Int. Comp. Mus. Assoc., San Fransisco.

Gabrielsson, A. (1999). Music Performance. In Deutsch, D., editor, *Psychology of Music*, pages 501–602. Academic Press, San Diego, 2nd edition.

Gabrielsson, A. (2003). Music performance research at the millenium. *Psych. Mus.*, 31(3):221–272.

Goebl, W. and Dixon, S. E. (2001). Analyses of tempo classes in performances of Mozart piano sonatas. In Lappalainen, H., editor, *Proc. of the 7th Int. Symposium on Systematic and Comparative Musicology, 3rd Int. Conf. on Cognitive Musicology, August 16–19, 2001*, pages 65–76. University of Jyväskylä, Jyväskylä, Finland.

Kohonen, T. (2001). *Self-Organizing Maps*. Springer, Berlin, Germany, 3rd edition.

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

Pampalk, E. (2003). Aligned self-organizing maps. In *Proc. of the Workshop on Self-Organizing Maps, September 11–14, 2003*, pages 185–190. Kyushu Inst. of Technology, Ktakyushu, Japan.

Pampalk, E., Goebl, W., and Widmer, G. (2003). Visualizing changes in the inherent structure of data for exploratory feature selection. In *Proc. of the 9th ACM SIGKDD Int. Conf. on Knowledge Discovery and Data Mining*, pages 157–166. ACM, Washington DC.

Pampalk, E., Rauber, A., and Merkl, D. (2002). Using smoothed data histograms for cluster visualization in self-organizing maps. In Dorronsoro, J. R., editor, *Proc. of the Int. Conf. on Artificial Neural Networks (ICANN'02), Madrid,* pages 871–876. Springer, Berlin.

Repp, B. H. (1992). Diversity and commonality in music performance: An analysis of timing microstructure in Schumann's "Träumerei". *J. Acoust. Soc. Am.*, 92(5):2546–2568.

Repp, B. H. (1998). A microcosm of musical expression. I. Quantitative analysis of pianists' timing in the initial measures of Chopin's Etude in E major. *J. Acoust. Soc. Am.*, 104(2):1085–1100.

Repp, B. H. (1999). A microcosm of musical expression: II. Quantitative analysis of pianists' dynamics in the initial measures of Chopin's Etude in E major. *J. Acoust. Soc. Am.*, 105(3):1972–1988.

Sloboda, J. A. (1985). Expressive skill in two pianists: Metrical communication in real and simulated performances. *Canad. J. Exp. Psychol.*, 39(2):273–293.

Todd, N. P. M. (1992). The dynamics of dynamics: A model of musical expression. *J. Acoust. Soc. Am.*, 91(6):3540–3550.

Windsor, W. L. and Clarke, E. F. (1997). Expressive timing and dynamics in real and artificial musical performances: Using an algorithm as an analytical tool. *Music Percept.*, 15(2):127–152.