

The Magaloff Project: An Interim Report

Sebastian Flossmann¹, Werner Goebel², Maarten Grachten³, Bernhard Niedermayer¹, and Gerhard Widmer^{1,4}

¹Department of Computational Perception, Johannes-Kepler-University,
Linz

²Institute of Musical Acoustics, University of Music and Performing Arts
Vienna

³Institute of Psychoacoustics and Electronic Music, Ghent University

⁴Austrian Research Institute for Artificial Intelligence (OFAI), Vienna

Abstract

One of the main difficulties in studying expression in musical performance is the acquisition of data. While audio recordings abound, automatically extracting precise information related to timing, dynamics, and articulation is still not possible at the level of precision required for large-scale music performance studies. In 1989, the Russian pianist Nikita Magaloff performed essentially the entire works for solo piano by Frédéric Chopin on a Bösendorfer SE, a computer-controlled grand piano that precisely measures every key and pedal action by the performer. In this paper, we describe the process and the tools for the preparation of this collection, which comprises hundreds of thousands of notes. We then move on to presenting the results of initial exploratory studies of the expressive content of the data, specifically effects of performer age, performance errors, between-hand asynchronies, and tempo rubato. We also report preliminary results of a systematic study of the shaping of particular rhythmic passages, using the notion of phase-plane trajectories. Finally, we briefly describe how the Magaloff data were used to train a performance rendering system that won the 2008 Rencon International Performance Rendering Contest.

1 Introduction

By now there is a substantial history of quantitative, computer-based music performance research (e.g., Clarke and Windsor (2000); Clarke (1985); Gabrielsson (1999, 2003); Goebel (2001); Honing (2003); Palmer (1989, 1996a,b); Repp (1995, 1992); Repp et al. (2002); Widmer et al. (2003); Widmer and Goebel (2004); Windsor et al. (2006), to name but a few). The main difficulty is the acquisition of representative data – preferably a large

amount of precise information by high-class artists under concert (not laboratory) conditions. The Magaloff project is centered around such a resource: The Magaloff Chopin Corpus, recordings of the Russian pianist Nikita Magaloff publicly performing the complete works for solo piano by Frédéric Chopin on stage, at the Vienna Konzerthaus in 1989. The collection meets all of the above mentioned criteria: It comprises over 150 pieces, over 10 hours of playing time, over 330,000 played notes. Having been performed and recorded on a Bösendorfer SE computer-controlled grand piano, the precise measurements of timing, loudness, etc. of each played note along with the pedal movements are available. To the best of our knowledge, as such it is the first precisely documented comprehensive collection of the complete works of a composer performed by a single artist as well as the largest collection of performances of a single artist available for performance research. By special permission of Magaloff's widow we are allowed to use this data for our research.

For further use of the corpus, it is necessary to annotate the raw MIDI data with the corresponding score information. That includes converting the music score sheets into a machine readable format and aligning the score with the performance (see section 3). The result constitutes the *Magaloff Corpus*, the empirical foundation and first milestone of our project. Based on this data, we seek new insights into the performance strategies applied by an accomplished concert pianist. In Section 4, we describe several research strands that are currently pursued and present some first preliminary results related to several aspects of performance: the effects of age and how Magaloff copes with it; the phenomenon of performance errors; the use of between-hand asynchronies as an expressive device, and especially tempo rubato. We also describe first results of a systematic study of the temporal shaping of particular rhythmic passage, using the notion of phase-plane trajectories. In Section 5, finally, we discuss the use of the Magaloff Corpus as training data for a performance rendering system that won the 2008 Rencon International Performance Rendering Contest in Japan.

2 Nikita Magaloff

2.1 Biographical Remarks

Nikita Magaloff, born on February 21, 1912, in St. Petersburg, was a Russian pianist. As his family was friendly with musicians like Sergei Rachmaninov, Sergei Prokofiev and Alexander Siloti, he grew up in a very musical environment. In 1918, the family first moved to Finland and then to Paris soon after (1922), where Nikita Magaloff started studying piano with Isidore Philipp, graduating from the Conservatoire in 1929 (Cella and Magaloff, 1995).

Magaloff started his professional career mainly in Germany and France, often appearing together with the violinists József Szigeti (whose daughter Irène he later married) and Arthur Grumiaux, and the cellist Pierre Fournier. In 1949, he took over Dinu Lipatti's piano class at the Geneva Conservatoire where he continued teaching until 1960. His pupils include Jean-Marc Luisada, Maria Tipo, Sergio Calligaris, Michel Dalberto and Martha

Argerich.

Magaloff is especially known for his performances of the complete works of Frédéric Chopin, which he usually presented live in a cycle of six recitals. The first ever recording of the complete works of Chopin was made by Magaloff in the years 1954–1958 for Decca. He repeated this for Philips in 1975. Other than that, only a few studio recordings by Magaloff exist. Nikita Magaloff died in on 26 December 1992, at the age of 80 in Vevey, in the Canton Vaud in Switzerland (Cella and Magaloff, 1995).

2.2 Magaloff’s Vienna Concerts in 1989

Between 1932 and 1991, Magaloff appeared in 36 concerts in the *Wiener Konzerthaus*, one of Vienna’s most illustrious concert venues – 24 solo concerts, 10 concerts as orchestra soloist, 2 chamber recitals together with József Szigeti.¹ In 1989, he started one of his famous Chopin cycles in which he would play all Chopin’s works for solo piano that were published in the composer’s lifetime, essentially Op. 1 to Op. 64, in ascending order. Each of the six concerts was concluded with an encore from the posthumously published work of the composer. The concerts took place between January 16 and May 17, 1989, in the *Mozartsaal* of the *Wiener Konzerthaus*. At the time of the concerts, Magaloff was already 77 years old. Daily newspapers commenting on the concerts praise both his technique and his unsentimental, distant way of playing (Sinkovicz, 1989; Stadler, 1989). Table 1 lists the programs of the six concerts.

Although the technology had only been invented a short time before (first prototype in 1983, official release 1985 (Moog and Rhea, 1990)), all six concerts were played and recorded on a Bösendorfer SE, precisely capturing every single keystroke and pedal movement.² This was probably the first time the new Bösendorfer SE was used to such an extent. The collected data is most likely the most comprehensive corpus ever recorded from one performer. In 1999, we received written and exclusive permission by Irène Magaloff, Nikita Magaloff’s widow, to use the data for our research.

3 Preparation of the Corpus

The recorded symbolic performance data requires careful preparation to become accessible for further investigations. Without any reference to the score, nothing can be said about how specific elements were realised. A lengthened eighth note and a shortened quarter note may account for the same amount of performed time, the former probably being part of a slower passage in the same piece. Without any information about the notated duration

¹Information available through the program archive of the *Wiener Konzerthaus*, <http://konzerthaus.at/archiv/datenbanksuche>

²Each note on- and offset is captured with a temporal resolution of 1.25ms. The velocity of the hammer at impact is converted and mapped to 128 midi loudness values. See Goebel and Bresin (2003) for details.

Date	Played
16 Jan	Rondo Op. 1; Piano Sonata No. 1 Op. 4; Rondo Op. 5; 4 Mazurkas Op. 6; 5 Mazurkas Op. 7; 3 Nocturnes Op. 9; 12 Etudes Op. 10. <i>Encore:</i> Fantaisie-Impromptus Op. posth. 66.
19 Jan	Variations Op. 12; 3 Nocturnes Op. 15; Rondo Op. 16; 4 Mazurkas Op. 17; Grande Valse Op. 18; Bolero Op. 19; Scherzo No.1 Op. 20; Ballade No. 1 Op. 23; 12 Etudes Op. 25. <i>Encore:</i> Variations “Souvenir de Paganini” (posth.)
15 Mar	2 Polonaises Op. 26; 2 Nocturnes Op. 27; 24 Preludes Op. 28; Impromptu No.1 Op. 29; 4 Mazurkas Op. 30; Scherzo No.2 Op. 31. <i>Encore:</i> Waltz in E minor (posth.)
10 Apr	2 Nocturnes Op. 32; 4 Mazurkas Op. 33; 3 Waltzes Op. 34; Piano Sonata No.2 Op. 35; Impromptu No.2 Op. 36; 2 Nocturnes Op. 37; Ballade No.2 Op. 38; Scherzo No.3 Op. 39; 2 Polonaises Op. 40; 4 Mazurkas Op. 41; Waltz Op. 42; Tarantella Op. 43. <i>Encore:</i> Waltz Eb-Major (posth.)
13 Apr	Polonaise Op. 44; Prelude Op. 45; Allegro de Concert Op. 46; Ballade No.3 Op. 47; 2 Nocturnes Op. 48; Fantaisie Op. 49; Impromptu No.3 Op. 51; 3 Mazurkas Op. 50; Polonaise Op. 53; Scherzo No.4 Op. 54. <i>Encore:</i> Ecossaises Op. posth. 72 No.3.
17 May	2 Nocturnes Op. 55; 3 Mazurkas Op. 56; Berceuse Op. 57; Piano Sonata No.3 Op. 58; 3 Mazurkas Op. 59; Barcarolle Op. 60; Polonaise-Fantaisie Op. 61; 2 Nocturnes Op. 62; 3 Mazurkas Op. 63; 3 Waltzes Op. 64. <i>Encore:</i> Waltz Op. posth. 69 No. 1

Table 1: The Magaloff *Konzerthaus* Concerts 1989.

of the note, no assumption can be made about what kind of modification the performer applied to the note.

In the following we describe the steps we undertook to provide the score information for all performed notes – a rather demanding challenge, that took more or less a whole person year. We need the final state for the corpus to be a piecewise list of all performed notes aligned with their counterparts in the score. For this we first need symbolic, computer-readable representations of all scores, which then are aligned to the MIDI data representing Magaloff’s performances. Given the nature of Chopin’s music – high note density, high degree of expressive tempo variation – automatic matching will be error-prone and accordingly, intensive manual correction of the alignment is required. As the most intuitive way to view a score is the music score itself, the easiest access for manually inspecting and correcting an alignment is to display the score page and the piano roll representation of the performance (MIDI) joined together by the alignment. This requires a score representation that contains not only information pertaining to the musical content of the piece but also to the geometrical location of each and every element on the original printed score.

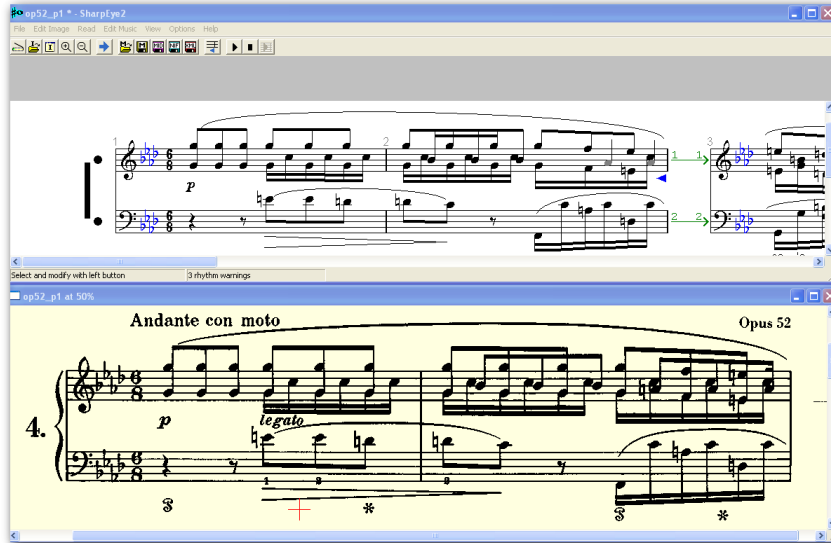


Figure 1: The SharpEye OMR software showing the printed score (lower panel) and the result of the recognition software (upper panel).

The format most suitable for our needs is musicXML (Recordare, 2003). MusicXML is intended to describe all information – musical content, expressive annotations, editorial information – contained in a score and is also very commonly used in *optical music recognition (OMR) software*. As it is text-based and human readable, it is easy to extend the format with the geometrical information we need.

3.1 From Printed Score to Extended MusicXML

The first step in digitising the score is to scan the sheet music. As we have no information as to which score editions Magaloff used, we used the Henle Urtext Editions (Zimmermann, 2004) with the exceptions of the Sonata Op. 4 and the Rondos Op. 1, Op. 5 and Op. 16, that Henle does not provide; in these cases we were forced to use the obsolete Paderewski editions (Paderewski, 1999, 2006). The 930 pages of sheet music were scanned in greyscale with a resolution of 300 dpi. The commercial OMR software *SharpEye*³ was used to extract the musical content from the scanned sheets. Figure 1 shows a screenshot of the program working on Chopin’s Ballade Op. 52.

The example illustrates several problems in the recognition process: the middle voice starting in the beginning of the second bar (B \flat 4) is misinterpreted as a series of sixteenth notes instead of eighths, which is easy to miss both when reviewing the score as well as listening to a mechanical MIDI rendering. The middle voice in the second half of the bar could not be read from the scan and has to be added manually. To emphasise a melody voice or to clarify a situation where voices cross, a note may have two stems with different

³see <http://www.visiv.co.uk>



Figure 2: A multivoice situation where a rest has to be added so that the middle voice starting is placed on the correct symbolic onset (*left*: score image, *right*: SharpEye Interface)

durations. In the case shown in figure 1, the sixteenth notes G4 starting in the first measure on beat 4 can be interpreted as expressive annotation or interpretative advice rather than actual note content. Keeping the ones with the shortest duration, the duplicated notes had to be removed, as they would bias the error statistics we carry out on the performances (see 4.2). Other common problems include *8va* lines (dashed lines indicating that certain notes actually have to be played one octave higher or lower) that are not recognised by SharpEye, bars spanning more than one line, and certain n -tuplets of notes. Especially rhythmically complex situations with different independent voices can lead to problems in the conversion. Figure 2 shows such a situation: A sixteenth rest has to be added so that SharpEye places the B4 on the correct onset. Thus, intensive inspection and extensive manual corrections have to be made. The graphical alignment software discussed in Section 3.2 provides for manual post-correction of those.

The choice of SharpEye was also motivated by the fact that, while SharpEye exports the results in a musicXML format which originally does not store the geometric location of the elements on the page, it also provides access to the intermediate, internal representation of the analysed page. This information is stored in *mro* files, SharpEye’s native file format. In *mro* files, all recognised elements are described graphically rather than musically: notes are stored with their position relative to the staff rather than with a musical interpretation of the note that takes the clef into account. Figure 3 shows the same chord represented in the two different formats. A custom-made script was used to extract the geometrical position of the note elements from the *mro* file and add the information to the corresponding elements in the musicXML file, thus linking the musicXML file with its original sheet music image.

3.2 Score-Performance Matching and Graphical Correction

Score-Performance Matching is the process of aligning the score and a performance of a musical piece in such a way that for each note of the score the corresponding performed note is marked and vice versa. Each score note is either marked as *matched* or *omitted* if the score note was not played, and each performed note is marked as either *matched* or *inserted* if the played note has no counterpart in the score. With the exception of trills and some other ornaments, this constitutes a one-to-one matching situation of score and performance.

<pre> <note> <pitch> <step>G</step> <octave>5</octave> </pitch> <duration>840</duration> <voice>1</voice> <type>eighth</type> <stem>up</stem> <staff>1</staff> ... </note> <note> <chord/> <pitch> <step>G</step> <octave>4</octave> </pitch> <duration>840</duration> <voice>1</voice> <type>eighth</type> <stem>up</stem> <staff>1</staff> ... </note> </pre>	<pre> chord { <attributes of the chord and the beam> flagposn -56,232 headend 2 notes { note { shape Solid staveoffset 0 p -5 accid None accid_dc 0 normalside True } note { shape Solid staveoffset 0 p 2 accid None accid_dc 0 normalside True } } } </pre>
--	---

Figure 3: A chord in the musicXML format (left panel) and its counterpart in the SharpEye *mro* format (right panel).

Several matching strategies are mentioned and evaluated in the literature (Heijink et al., 2000; Raphael, 2006), ranging from straight-forward matching to dynamic time warping or Hidden Markov Models. We use the edit-distance paradigm that was initially invented for string comparisons (Wagner and Fischer, 1974) and has been used in different music computing applications (Dannenberg, 1984; Pardo and Birmingham, 2002). Grachten (2006) offers more detailed information on edit-distance-based matching as a score-performance alignment algorithm. Since the edit-distance assumes a strict order of the elements in the sequences to be aligned, it is not directly applicable to polyphonic music. To solve this problem, we represent polyphonic music as sequences of *homophonic slices* (Pickens, 2001), by segmenting the polyphonic music at each note onset and offset. The segments, represented as the set of pitches at that time interval, have a strict order, and can therefore be aligned using the edit-distance. A series of edit operations – insertion, omission, match and trill operations in our case – then constitute the alignment between the two sequences. Each of the applied operations comes at a cost (the better the operation fits in a specific situation, the lower the cost), the sum of which is minimised over the two sequences – score and performance.

Due to the complexity of the music and the highly expressive tempo and timing variations in the performances, the automatic score-performance matching is very error-prone. As the number of notes is vast, the interface for correcting and adjusting the alignment has to be intuitive and efficient. Extending the musicXML by geometric information from the scanning process allows for an application displaying the original score sheet in an interactive way: each click on the note elements in the image can be related to the cor-

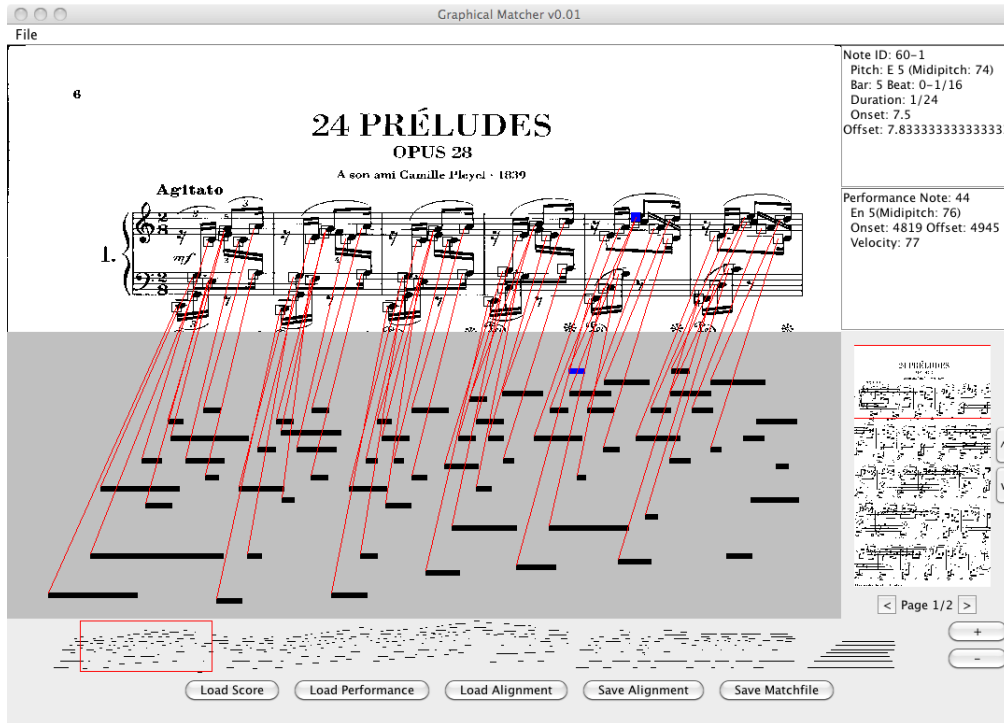


Figure 4: jGraphMatch: a Software tool for display and manual correction of score-performance alignments.

responding entry in the musicXML score. A combined display of this interactive score and the performance as a piano roll provides easy access to inspecting and modifying the alignment. Figure 4 shows a screenshot of the platform independent Java-Application we developed.

One problem with the matching was that in some pieces there are differences between our version of the score and the version performed by Magaloff: this ranged from small discrepancies where, e.g., Magaloff repeats a group of notes more often than written in the score (e.g., in the Nocturne Op. 9 No. 3, bar 111), to several skipped measures (e.g., Waltz Op. 18, where he omitted bars 85 to 116), to major differences that probably are the result of a different edition being used by Magaloff (e.g., in the Sonata Op. 4 Mv. 1, bars 82 to 91, where the notes he plays are completely different from what is written in the score). In the error analysis presented in Section 4.2 below, we will not count these as performance errors, and we also do not count these cases as insertions or omissions in the overview table 2.

3.3 Statistical Overview

Table 2 gives a summary of the complete corpus. Grace notes and trills are mentioned separately: Grace notes do not have a nominal duration defined by the score. Therefore they cannot contribute to discussions of temporal aspects of the performance. As a con-

Pieces/Movements	155
Score Pages	930
Score Notes	328.800
Performed Notes	335.542
Playing Time	10h 7m 52s
Matched Notes	318.112
Inserted Notes	12.325
Omitted Notes	11.506
Substituted Notes	5.105
Matched Grace Notes	4289
Omitted Grace Notes	449
Trill Notes	5923

Table 2: Overview of the Magaloff Corpus.

sequence we normally exclude those from the data. Trills constitute many-to-one matches of several performance notes to a single score note. When counting the performance notes in the corpus, the number of performance notes matched to a trill have to be accounted for. Accordingly, the complete number of performed notes is composed of the number of matches, substitutions, insertions, matched grace notes, and trill notes. The complete number of score notes is composed of the number of matches, substitutions, omissions, and matched and omitted grace notes.

Table 3 shows the note and matching statistics according to piece categories. The generic category *Pieces* includes: Introduction and Variations Op. 12, Bolero Op. 19, Tarantella Op. 43, Allegro de Concert Op. 46, Fantaisie Op. 49, Berceuse Op. 57, Barcarolle Op. 60, and Polonaise-Fantaisie Op. 61. The encores were not included in the corpus.

4 Exploratory Intra-Artist Research

This section describes a number of initial studies we performed on the data in order to explore characteristics of Magaloff’s playing style. We view these as first steps into investigating the art of a world-class pianist based on data with unprecedented precision.

4.1 Performer Age

One of the remarkable aspects of Magaloff’s Chopin concerts in 1989 is the age at which he undertook this formidable task: he was 77 years old.⁴ Performing on stage up to old ages is not exceptional among renowned pianists: Backhaus played his last concert at 85, Horowitz at 84, Arrau at 88. The enormous demands posed by performing publicly

⁴At age 77, Alfred Brendel performed one solo program and one Mozart Concerto for his last season in 2008

Category	Pieces	Score	Played	Matches	Insertions	Omissions	Substitutions
Ballades	4	19511	20223	18971	1001	496	251
Etudes	24	40894	40863	38684	1615	1681	561
Impromptus	3	7216	7310	7150	96	159	64
Mazurkas	41	47312	47043	45260	1129	1669	470
Nocturnes	19	31109	32016	30943	671	873	302
Pieces	7	39759	41068	38249	1728	1487	916
Polonaises	7	27873	28301	26232	1597	1189	436
Preludes	25	20067	20239	19234	683	631	321
Rondos	3	18250	18331	17347	324	441	440
Scherzi	4	21951	22633	20849	1369	707	376
Sonatas	12	38971	40450	37015	1651	1498	731
Waltzes	8	18651	18876	18178	461	675	237

Table 3: Overview by piece category

include: motor skills, memory, physical endurance, and stress factors (Williamon, 2004). A psychological theory of human life-span development identifies three factors that are supposed to be mainly responsible for “successful ageing”: Selection, Optimisation, and Compensation (SOC model (Baltes and Baltes, 1990)). Applied to piano performance, this would imply that older pianists play a smaller repertoire (selection), practice these few pieces more (optimisation), and hide technical deficiencies by reducing the tempo of fast passages while maintaining tempo contrasts between fast and slow passages (compensation) (Vitouch, 2005). In Flossmann et al. (2009a), we tested whether Magaloff actually used strategies identified in the SOC model.

The first aspect of the SOC model, *selection*, seems not to be supported in this case: Magaloff performed the entire piano works by Chopin within four months.⁵ We cannot make a statement about *optimisation* processes due to our lack of information about his practice regime before and during the concert period.

Regarding possible *compensation* strategies, we studied Magaloff’s performance tempi in the context of other recordings – on the études only, to keep the effort manageable. We analysed selected recordings of Chopin’s études by several renowned pianists, including an earlier recording by Magaloff at the age of 63. These audio recordings, a total of 289 performances of 18 études by 16 performers⁶, were semi-automatically beat-tracked using the software *Beatroot* (Dixon, 2001, 2007) to determine a tempo value.⁷

⁵Of course, Magaloff’s repertoire might have been broader in younger years, which would then indicate otherwise. A systematic comparison of earlier concert seasons and all concerts in 1989 would provide further insights into that particular aspect.

⁶Arrau (recorded 1956), Ashkenazy (1975), Backhaus (1928), Biret (1990), Cortot (1934), Gavrillov (1985), Giusiano (2006), Harasiewicz (1961), Lortie (1986), Lugansky (1999), Magaloff (1975), Magaloff (1989), Pollini (1972), Schirmer (2003), Shaboyan (2007), and Sokolov (1985).

⁷A basic tempo value was estimated by the mode value, the most frequent bin of an inter-beat interval

Compared to these performances, Magaloff’s Op. 10 etudes are on average 1.2% slower, the Op. 25 études 5.6% slower than the average performance. Compared with the metronome markings in the Henle editions, 12 out of 18 of Magaloff’s performances are within a 10% range, three pieces more than 5% slower, three pieces more than 5% faster. Comparing Magaloff’s recordings at the age of 63 and 77, the tempi vary to a surprising degree, but no systematic tempo decrease in the latter could be found. On the contrary, in 12 pieces out of 18, the recording at age 77 is faster, sometimes to a considerable degree (up to 17% in Op. 10 No. 10). On the whole, Magaloff’s performances do not suggest a correlation between age and tempo, while the tempi of the other pianists’ recordings show a slight age effect (with piecewise correlations between pianist age and tempo ranging from -0.66 to 0.51 , with an average of -0.17).⁸

As an exemplary piece containing tempo contrasts, we examined the Nocturne Op. 15 No. 1 (*Andante cantabile*), which contains a technically demanding middle section (*con fuoco*). The tempo values of performances by 14 other pianists, including Argerich, Rubinstein and Pollini, show a significant correlation between the age of the performer at the time of the recording, and the tempo of the middle section (the older, the slower). The tempo ratios between the contrasting sections, however, showed no overall age effect, confirming Vitouch’s interpretation of the SOC model (Vitouch, 2005). Magaloff’s performance of the Nocturne does not fall into this pattern: he played faster than the youngest of the performers while keeping a comparable tempo ratio.

Thus, our analysis of Magaloff’s tempi does not point to any compensation processes, which were indeed found with other pianists. In sum, Magaloff’s Chopin does not seem to corroborate the SOC model.

4.2 Error Analysis

Performance errors occur at all levels of proficiency. Studies have been conducted under laboratory conditions and give first insights into the phenomenon (e.g. Palmer and van de Sande (1993, 1995); Repp (1996)). However, confirming these results under real concert conditions has been difficult so far. In Flossmann et al. (2009a) and Flossmann et al. (2010) we analyse Magaloff’s performance errors, put them into context of both performance and score and test whether the findings corroborate previous studies.

As can be derived from Table 2 the Magaloff performances contain 3.67% insertion errors, 3.50% omission errors and 1.55% substitution errors. This exceeds the percentages Repp found (1.48%, 0.98%, and 0.21%, respectively (Repp, 1996)), but looking only at the particular piece used by Repp (Prelude Op. 28 No. 15) the error percentages are similar (0.72%, 1.58% and 0.52%, respectively). Among the piece categories, the Scherzi and Polonaises stand out in terms of insertion errors (above 5%), the Rondos and Impromptus constitute the low-insertion categories (insertion rate below 2.0%). The Impromptus are

histogram with a bin size of 4% of the mean inter-beat interval.

⁸These considerations are based on the underlying assumption that the difficulty of a piece increases with the tempo. This is not universally true. However, for the pieces in question – the fast pieces of the Études – the assumption seems warranted.

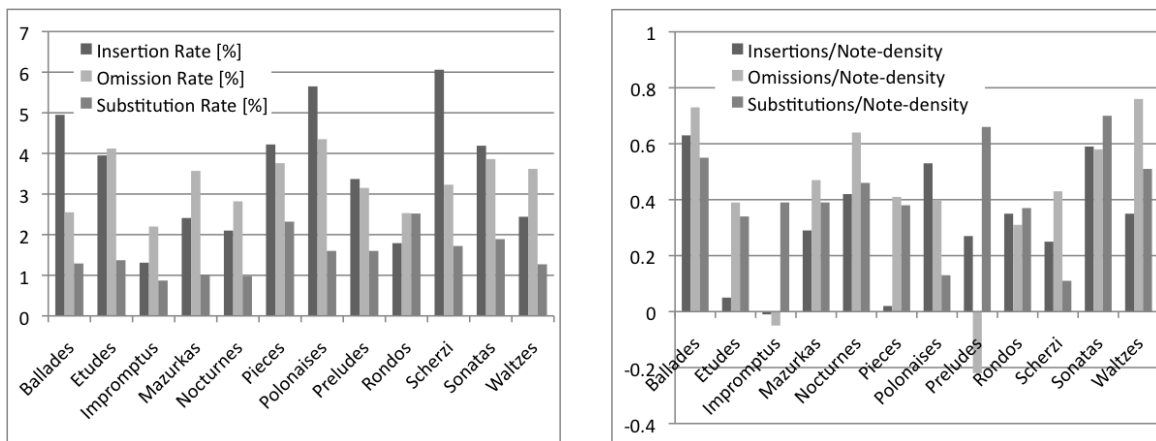


Figure 5: *Left panel:* Error percentages by piece category. *Right panel:* Correlation coefficients between note-density and error rate by piece category.

also the category with the lowest percentage of omission errors (2.20%), while *Études* and *Polonaises* exhibit the highest percentage of omission (above 4%).

Considering the errors in the context of the general tempo of a piece, we found that a high note density goes along with a higher error frequency (the more notes per time unit, the more errors). This holds to a varying degree for all kinds of errors: overall the corpus exhibits correlation coefficients between note density and frequency of insertion errors, omission errors and substitution errors of 0.39, 0.26 and 0.61, respectively. Figure 5 shows the error rates and correlation coefficients of error frequency and note density for the respective categories of pieces. The *Ballades* and *Polonaises* show both high error percentages as well as a high correlation of error frequency and note density, suggesting that these are technically particularly demanding.

The perceptual discernibility of an insertion or a substitution error is closely related to how loud the wrong note was played in proportion to the other notes in the vicinity and how well the note fits into the harmonic context (Repp, 1996). Viewing the insertion notes in the corpus in their vertical and horizontal context reveals that the majority of notes are inserted with at most 70% of the loudness of the adjacent (horizontal and vertical) notes. An analysis of the harmonic appropriateness of the insertion and substitution notes in their context, however, suggests that the errors are perceptually more conspicuous than assumed: 40% of the respective errors are not compatible with the local harmony.

Our findings in this live performance data mostly corroborate Repp’s findings under laboratory conditions (Repp, 1996): the percentage of errors in melody voices is lower than in non-melody voices (omission rates of 1% (melody voices) and 4.1% (non-melody voices)), and the majority of insertion errors are of low intensity compared to their immediate neighbourhood. The error frequency is related to a varying degree to the note density, depending on the technical demands of the actual piece.

If we may make a somewhat speculative comment here, the fact that Magaloff did not reduce his performance tempi even at age 77 (see Section 4.1) and that his performances

display relatively high error rates might be taken as an indication that Magaloff aimed at realising his musical ideas of Chopin’s work rather than at error-free performances. Further analyses will try to establish connections between score characteristics and certain error patterns.

4.3 Between-hand Asynchronies

Temporal offsets between the members of musical ensembles have been reported to yield specific characteristics that might reflect expressive intentions of the performers; e.g., the principal player in wind or string trios precedes the others by several tens of milliseconds (Rasch, 1979), and soloists in jazz performances have been shown to synchronise with the rhythm section at offbeats (Friberg and Sundström, 2002). As the hands of a pianist are capable of producing different musical parts independently, the temporal asynchronies between the hands may be an expressive means for the pianist. In Goebel et al. (2010), we examined the between-hand asynchronies in the Magaloff corpus. The asynchronies were computed automatically over the entire corpus based on staff information contained in the score, assuming that overall the right hand played the upper staff and the left hand the lower. For the analysis of this phenomenon we excluded all onsets marked in the score as *arpeggiated*; in these cases temporal deviations are prescribed by the score rather than being part of the interpretation. The main results of this study (Goebel et al., 2010) are reported briefly in the following.

The analysis of over 160,000 nominally simultaneous events revealed tempo effects: slower pieces were played by Magaloff with larger asynchronies than faster pieces. Figure 6 (left panel) shows the correspondence between event rate and asynchrony. Moreover, pieces with chordal texture were more synchronous than pieces with melodic textures. Subsequent analyses focussed on specific kinds of between-hand asynchronies: bass anticipations and occurrences of “tempo rubato in the earlier meaning” (Hudson, 1994).

As *bass anticipations* we consider events where a bass note precedes the other voices by more than 50 ms. They can be clearly perceived due to their large asynchronies and can be considered to be expressive decisions by the performer. Magaloff’s performances contain a considerable number of these bass anticipations (about 1% of all simultaneous events). Again, higher proportions are found in slower pieces.

The “*tempo rubato in the earlier meaning*” refers to particular situations in which the right hand deviates temporally from a stable timing grid established by the left hand (Hudson, 1994). Chopin, in particular, recommended to his students this earlier type of rubato as opposed to the later type that refers to a parallel slowing down and speeding up of all parts of the music (today referred to as expressive timing). We automatically identify sequences where Magaloff apparently employed an “earlier tempo rubato” by searching for *out-of-sync regions* in the pieces. An out-of-sync region is defined as a sequence of consecutive asynchronies that are larger than the typical perceptual threshold (30ms) and that comprises more events than the average event rate of that piece. On average, 1.8 such regions were found per piece (283 in total) with particularly high counts in the Nocturnes – a genre within Chopin’s music that leaves most room for letting the melody move freely

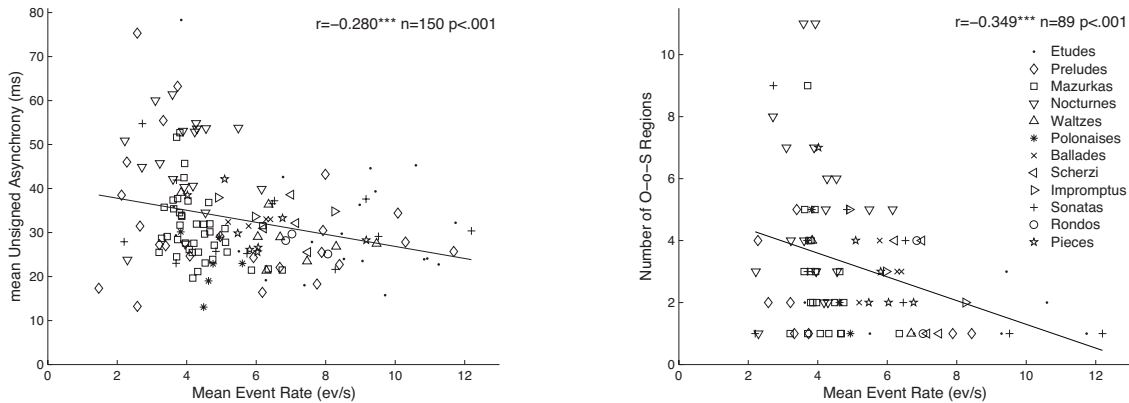


Figure 6: *Left panel:* Absolute asynchronies plotted against the mean event rate by piece category. *Right panel:* The number of out-of-sync regions plotted against the mean event rate plotted.

above the accompaniment. Figure 6 shows the correspondence between event density and number of “earlier tempo rubato” sequences.

Finally, an attempt was made to predict Magaloff’s asynchronies on the basis of a set of mostly local score features using a probabilistic learning model. Between-hand asynchronies in some individual pieces could be predicted quite well (Étude Op. 25 No. 11 or Impromptu Op. 29), but generally the prediction results were poor. It might be that a more complex representation of the score might be required to explain and predict between-hand asynchronies, which potentially contain a range of expressive intentions in Magaloff’s Chopin (Goebel et al., 2010).

4.4 Phase-plane Representations for Visual Analysis of Timing

In this section, we illustrate how *phase-plane* representations of timing data provide a tool for exploring and understanding various aspects of the data. The phase-plane representations is a visualisation tool common in physics and dynamic systems theory. It was introduced in the context of music performance research (Grachten et al., 2008; Ramsay, 2003; Grachten et al., 2009), mainly because of its emphasis on dynamic aspects of the data. This is of particular relevance for the analysis of “expressive gestures”, which are (at least partially) manifest as fluctuations in timing and loudness of performed music.

A phase-plane plot of expressive timing displays measured tempo data as a trajectory in a two-dimensional space – the *state-space* –, where the horizontal axis represents tempo and the vertical axis represents the first derivative of tempo. Passages of constant tempo do not cause any motion through the state-space, but changes in tempo lead to (typically curved) clockwise trajectories, where accelerandi correspond to motion through the first and second quadrants, and ritardandi to motion through the third and fourth quadrants.

The phase-plane representation of empirical data is part of a larger methodology known

as *functional data analysis* (Ramsay and Silverman, 2005). The core of this methodology is the construction of a suitable functional approximation of measured data points, which are assumed to be measurements of some continuous process. In our case the functional approximation is done using linear combinations of piecewise polynomial curves (B-splines) to fit the data. The fitting process uses a least-squares criterion that includes a penalty term for roughness – thus higher penalties lead to smoother curves. Phase-plane plots of data are obtained by plotting the fitted function against its derivative. Derivatives can easily be computed due to the piece-wise polynomial form. More details can be found in Grachten et al. (2009).

The method for computing phase-planes used here differs slightly from the one presented in Grachten et al. (2009). Rather than approximating tempo data, which are derived from measured onset times in the performance, we fit the measurements directly as a score-performance time-map. This method is more robust in the sense that the fitted function is less susceptible to *overshoot* due to fitting with low roughness-penalties. The IOI curve is obtained by taking the first derivative of the function fitted to the score-performance time-map. Rather than converting the inter-onset interval (IOI) curve into a tempo curve for phase-plane display, we compute the phase-plane trajectory by taking the negative logarithm of the IOI curve, where the IOI values are divided by the average IOI value over the region of interest. The resulting curve is very similar conceptually to a tempo curve (that is, greater values imply a faster tempo), with the difference that the scale is logarithmic. Thus, a value of 1 corresponds to doubling the nominal tempo, and a value of -1 to half the nominal tempo.

A data set like the Magaloff corpus offers a unique opportunity to study how expressive patterns relate to musical structure. Typical corpora of music performances are less suited for such studies, since they tend to contain performances by many pianists, but for a relatively small amount of musical material. Since the Magaloff corpus contains virtually all of Chopin’s piano works, there is an abundance of musical material.

As a preliminary study of Magaloff’s style of expressive timing, we investigate timing patterns corresponding to particular rhythmical contexts throughout the corpus. The first step is to select rhythmical contexts that occur frequently in different pieces. We then compare the expressive timing data corresponding to all instances of those rhythmical contexts to see whether the rhythmical contexts can be characterized by a typical timing pattern. We restrict a rhythmical context to be of fixed length, namely one measure. The context is uniquely determined by its time signature and the onset times of the left hand (relative to the start of the measure). Table 4 shows two such rhythmical contexts and their occurrences in the corpus. Note that the patterns are both regular divisions of the measure into 16 equal parts, the only difference being that one pattern has a 2/2 time signature, and the other 4/4.

Figure 7 shows the phase-plane trajectories corresponding to the patterns A and B. In order to avoid clutter not all instances of both patterns have been drawn in the plots. Instead, we show the average trajectory (bold line), together with the average trajectories of four clusters within the set of trajectories (thin lines), in order to give an impression of

	Pattern A	Pattern B
Time signature:	$\frac{2}{2}$	$\frac{4}{4}$
Onset Pattern:	$0, \frac{1}{8}, \frac{2}{8}, \frac{3}{8}, \frac{4}{8}, \frac{5}{8}, \frac{6}{8}, \frac{7}{8},$ $1, 1\frac{1}{8}, 1\frac{2}{8}, 1\frac{3}{8}, 1\frac{4}{8}, 1\frac{5}{8}, 1\frac{6}{8}, 1\frac{7}{8}$	$0, \frac{1}{16}, \frac{2}{16}, \frac{3}{16}, 1, 1\frac{1}{16}, 1\frac{2}{16}, 1\frac{3}{16},$ $2, 2\frac{1}{16}, 2\frac{2}{16}, 2\frac{3}{16}, 3, 3\frac{1}{16}, 3\frac{2}{16}, 3\frac{3}{16}$
Occurrences:	Op. 10, No. 12, <i>Étude</i> (46 times) Op. 25, No. 6, <i>Étude</i> (4 times) Op. 28, No. 16, <i>Prelude</i> (2 times) Op. 28, No. 3, <i>Prelude</i> (26 times) Op. 46, <i>Allegro de Conc.</i> (10 times)	Op. 10, No. 4, <i>Étude</i> (26 times) Op. 10, No. 8, <i>Étude</i> (17 times) Op. 16, <i>Rondo</i> (7 times) Op. 25, No. 1, <i>Étude</i> (4 times) Op. 46, <i>Allegro de Conc.</i> (4 times) Op. 58, Mv. 1, <i>Sonata</i> (29 times) Op. 62, No. 2, <i>Nocturne</i> (8 times)

Table 4: Two rhythmical contexts, and their occurrences in the Magaloff corpus.

the variability within a pattern.⁹

Both patterns show roughly circular trajectories, indicating a speeding up in the first half of the measure and a slowing down in the second half of the measure. Although both patterns are quite similar, as might be expected based on the similarity of the rhythmical contexts, there are also two clear distinctions. Firstly, the absolute sizes of the (averaged) trajectories differ between patterns A and B. Pattern A shows larger trajectories than pattern B, implying greater fluctuation of tempo. Secondly, pattern B shows an embedded cyclic form halfway through the trajectories. This corresponds to a brief slowing down and speeding up in the middle of the measure, and suggests that the weak metrical emphasis on the third beat is accentuated by a slight lengthening. This accentuation is completely absent in pattern A.

A last notable aspect of the plots is that the trajectories are not completely circular. There is a slight, but apparently systematic, discrepancy between beginning and end points. Although this might be an artifact of averaging, it is likely that the rhythmical contexts are themselves part of a larger context that has a characteristic timing pattern, since such patterns often span more than one measure.

4.5 Towards Comprehensive Inter-Artist Investigations

While the Magaloff corpus allows us to analyse one pianist’s playing style in great depth and with high precision, even more insights can result from comparing Magaloff’s style to the style of other pianists. This, of course, would require more such corpora of symbolic data. Unfortunately, in most cases the only available resource for data by other pianists are audio recordings.

Manually annotating a large number of audio recordings is beyond the limits of our resources. Therefore a longer-term goal is to develop a system that can extract symbolic

⁹Note that the clustering was not done for any analytical purpose, only to summarize the trajectory data succinctly.

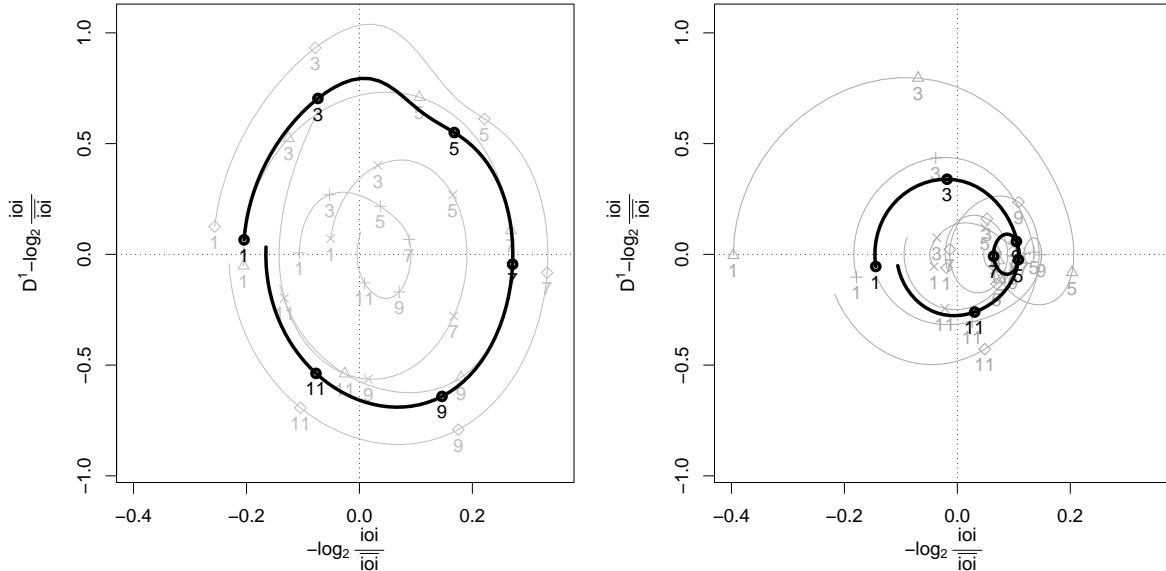


Figure 7: Phase-plane trajectories for the timing of two rhythmic patterns (left: pattern A; right: pattern B). Overall average trajectories are displayed in bold lines, cluster-average trajectories in thin lines. The uneven beats are numbered and marked by symbols along the trajectories.

data from audio recordings. Since the score that a performance is based on can be assumed to be known in most cases, the prime task is to identify each score note’s position within the audio recording – a problem known as *audio-to-score alignment*. Based on this step further performance parameters, like loudness, timbre characteristics, etc., can be estimated.

Audio-to-score alignment has been an issue in computational music research for more than ten years. By now two competing approaches as well as numerous variations and improvements have been established. One technique is to use the Dynamic Time Warping algorithm in order to align feature sequences computed from audio as well as the score (Hu and Dannenberg, 2005). The other method is to build statistical, graphical models, which can not only embed the temporal order of note events but also additional a priori knowledge, like relative note durations (Raphael, 2006)

While a lot of recent work has focused on real-time aspects of audio-to-score alignment, in the present context accuracy is much more crucial. We have recently introduced a refinement method which was able to extract onset times more accurately than the human threshold of recognition for about 40% of the notes within our test set (Niedermayer, 2009). Although this result is encouraging, it clearly shows that manual post-processing is still required in order to create accurate data.

Given more reliable automatic annotations the prospect is to build corpora of symbolic performance data for other pianists with a manageable amount of manual post-processing. Within the course of this long-term objective the Magaloff corpus will play several roles: (1) the existing transcriptions can serve as ground truth data for the quantitative evaluation

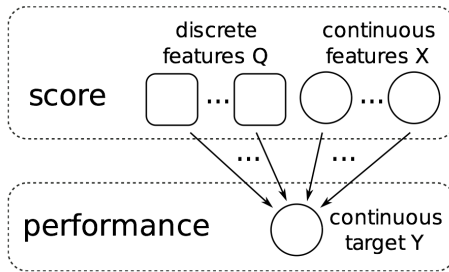


Figure 8: YQX: the probabilistic model

of any alignment system; (2) manual annotations (like separation of the score into melody-line, bass-line, etc.) within the Magaloff corpus can be transferred to other performances by means of alignment; and (3) it serves as a basis for inter-artist performance analysis once symbolic data describing other artists’s performances have been generated.

5 The Magaloff Corpus as Training Data for Expressive Performance Rendering

A data corpus of this dimension and precision is not only interesting for what it shows about the pianist that created it. The detailed annotation of score information for each played note makes the corpus a valuable asset as ground-truth data for various data driven music processing tasks. One such task is *Expressive Performance Rendering* – the problem of automatically generating a performance of a given musical score that sounds as “human” and “natural” as possible. To this end, first a model of the score or certain structural elements and musical elements of the score is calculated. The score model is then projected onto performance trajectories (for timing, dynamics etc.) by a predictive model that is usually learned from a large dataset of expressive performances.

In Widmer et al. (2009) and Flossmann et al. (2009b) we give a detailed description of our performance rendering system *YQX*. The core of the system is a probabilistic model that captures dependencies between score and performance characteristics, and learns to predict expressive timing, dynamics, and articulation. Given a musical score, the system predicts the most likely performance as seen in the database it was trained on – in our case, the Magaloff Corpus. As the prediction is done note by note for the melody voice of the piece¹⁰, the system computes a characterisation of all melody notes through a number of features that describe some aspects of the local context of each melody note. The features – both discrete and continuous variables – include among others: the pitch interval to the next note, the rhythmic and harmonic context, and the distance to the nearest point of musical closure according to an Implication-Realization analysis (Narmour, 1990) of the

¹⁰We assume that the highest pitch at any given time is the melody voice of the piece. This very simple heuristic is certainly not always true, but in the case of Chopin is correct often enough to be justifiable.

melodic content.¹¹ See Flossmann et al. (2009b) for further detail on the score features.

For each melody note three performance characteristics are extracted from the corpus describing the tempo, dynamics and articulation. The dependencies of score characteristics and performance characteristics are modelled through conditional probability distributions as depicted in Figure 8: for each configuration of discrete features we train a model that relates the continuous features to the observables. Hence, predicting tempo, dynamics and articulation for a melody note basically means answering the following question: given a specific score situation what are the most likely performance parameters found in the data corpus.

The predicted sequences are then projected onto a mechanical MIDI representation of the score in question, rendering an expressive version of the piece. A crucial issue is the trade-off between the specificity of the description of score context on the one hand, and the availability of training examples on the other. Using unspecific score context descriptions it may be impossible to narrow down an appropriate range of performance feature values per score context. On the other hand, using too specific descriptions of score context it is hard to *reliably* infer performance feature values, due to the small number of instances per score context. By enhancing the learning algorithm to optimise the predicted values over the complete piece instead of just choosing the locally most appropriate, we managed to slightly improve the results (Flossmann et al., 2009b).

Judging the expressivity of the generated performances in terms of how “human” or “natural” they sound is a highly subjective task. The only scientific environment for comparing different models according to such criteria is the annual *Performance Rendering Contest RENCON* (Hashida, 2008), which offers a platform for presenting and evaluating, via listener ratings, state-of-the-art performance modelling systems. In the RENCON 2008, two pieces specifically composed for the contest had to be rendered autonomously, one piece supposedly being Mozart-like, the other Chopin-like. Awards were given for expressivity (RENCON Award, by audience), technical sophistication of the system (RENCON Technical Award, by the committee) and for “affecting the composer most” (RENCON Murao Award, by T. Murao)¹². Trained on Magaloff’s Chopin Corpus, YQX won all three of these.¹³ See Widmer et al. (2009) for more information on this, and www.cp.jku.at/projects/yqx for videos of YQX performing live at the RENCON contest.

6 Conclusion

The goal of this article was to give the readers a broad introduction to, and a current status report on, a large-scale piano performance research project that is based on an exceptional

¹¹According to Narmour’s theory, “musical closure” is achieved when the melodic progression arouses no further expectations in the listener’s mind. The emerging segmentation of the score is comparable to a crude phrase structure analysis.

¹²see <http://www.renconmusic.org/icmpc2008/autonomous.htm>

¹³We only have access to the audience evaluation scores for the RENCON Award. There, YQX scored a total of 628 points, compared to 515 points scored by the second-ranked system.

corpus of empirical data. Dealing with data sets of this size raises a number of practical (and in some cases also conceptual) problems, which we tried to briefly illustrate here. The Magaloff corpus provides us with unique opportunities for studying a wide range of piano performance questions in great detail; the specific studies presented above are only first steps in a much longer-term research endeavour. While we cannot make the Magaloff corpus publicly available, due to the restricted, exclusive usage rights associated with it, we do hope that the experimental results based on it will contribute new insights to music performance research, and we hope to be able to at least make available to the research community some of the software tools we are developing for this exciting endeavour.

Acknowledgements

We hereby want to express our gratitude to Mme Irène Magaloff for her generous permission to use this unique resource for our research. This work is funded by the Austrian National Research Fund FWF via grants P19349-N15 and Z159 (“Wittgenstein Award”). The Austrian Research Institute for Artificial Intelligence acknowledges financial support from the Austrian Federal Ministries BMWF and BMVIT.

References

- Baltes, P. B. and Baltes, M. M. (1990). Psychological perspectives on successful aging: The model of selective optimization with compensation. In Baltes, P. B. and Baltes, M. M., editors, *Successful Aging*, pages 1–34. Cambridge University Press., Cambridge.
- Cella, F. and Magaloff, I. (1995). *Nikita Magaloff*. Nuove Edizione, Milano.
- Clarke, E. and Windsor, W. (2000). Real and simulated expression: A listening study. *Music Perception*, 17(3):277–313.
- Clarke, E. F. (1985). Some aspects of rhythm and expression in performances of Erik Satie’s “Gnossienne No. 5”. *Music Perception*, 2:299–328.
- Dannenberg, R. (1984). An on-line algorithm for real-time accompaniment. In *Proceedings of the 1984 International Computer Music Conference*. International Computer Music Association.
- Dixon, S. (2001). Automatic extraction of tempo and beat from expressive performances. *Journal of New Music Research*, 30(1):39–58.
- Dixon, S. (2007). Evaluation of the audio beat tracking system BeatRoot. *Journal of New Music Research*, 36:39–50.
- Flossmann, S., Goebel, W., and Widmer, G. (2009a). Maintaining skill across the life span: Magaloff’s entire chopin at age 77. In *Proceedings of the International Symposium on*

- Performance Science 2009*, Utrecht, The Netherlands. European Association of Conservatoires (AEC).
- Flossmann, S., Goebel, W., and Widmer, G. (2010). The Magaloff corpus: An Empirical Error Study. In *Proceedings of the 10th International Conference on Music Perception and Cognition (ICMPC) (to appear)*.
- Flossmann, S., Grachten, M., and Widmer, G. (2009b). Expressive performance rendering: Introducing performance context. In *Proceedings of the SMC 2009 – 6th Sound and Music Computing Conference*, pages 155–160.
- Friberg, A. and Sundström, A. (2002). Swing ratios and ensemble timing in Jazz performance: Evidence for a common rhythmic pattern. *Music Perception*, 19(3):333–349.
- Gabrielsson, A. (1999). Music performance. In Deutsch, D., editor, *Psychology of Music*, pages 501–602. Academic Press, San Diego, CA, 2nd edition.
- Gabrielsson, A. (2003). Music performance research at the millenium. *Psychology of Music*, 31:221–272.
- Goebel, W. (2001). Melody lead in piano performance: Expressive device or artifact? *Journal of the Acoustical Society of America*, 110:563–572.
- Goebel, W. and Bresin, R. (2003). Measurement and reproduction accuracy of computer-controlled grand pianos. *Journal of the Acoustical Society of America*, 114(4):2273–2283.
- Goebel, W., Flossmann, S., and Widmer, G. (2010). Investigations into between-hand synchronisation in magaloff’s chopin. *Computer Music Journal (in press)*, 34(3).
- Grachten, M. (2006). *Expressivity-Aware Tempo Transformations of Music Performances Using Case Based Reasoning*. PhD thesis, Department of Technology, Pompeu Fabra University.
- Grachten, M., Goebel, W., Flossmann, S., and Widmer, G. (2008). Phase-plane visualizations of gestural structure in expressive timing. In *Proceedings of the fourth Conference on Interdisciplinary Musicology*, Thessaloniki, Greece.
- Grachten, M., Goebel, W., Flossmann, S., and Widmer, G. (2009). Phase-plane representation and visualization of gestural structure in expressive timing. *Journal of New Music Research*, 38(2):183–195.
- Hashida, M. (2008). RENCON - performance rendering contest for computer systems. <http://www.renconmusic.org/>.
- Heijink, H., Desain, P., Honing, H., and Windsor, L. (2000). Make me a match: An evaluation of different approaches to score—performance matching. *Computer Music Journal*, 24(1):43–56.

- Honing, H. (2003). The final ritard: On music, motion, and kinematic models. *Computer Music Journal*, 27(3):66–72.
- Hu, N. and Dannenberg, R. B. (2005). A bootstrap method for training an accurate audio segmenter. In *Proceedings of the 6th International Conference on Music Information Retrieval*, pages 223–229, London, UK.
- Hudson, R. (1994). *Stolen Time: The History of Tempo Rubato*. Clarendon Press, Oxford.
- Moog, R. A. and Rhea, T. L. (1990). Evolution of the keyboard interface: The bosendorfer 290 se recording piano and the moog multiply-touch-sensitive keyboards. *Computer Music Journal*, 14(2):52–60.
- Narmour, E. (1990). *The Analysis and Cognition of Basic Melodic Structures: The Implication-Realization Model*. University of Chicago Press.
- Niedermayer, B. (2009). Improving accuracy of polyphonic music-to-score alignment. In *Proceedings of the 10th International Conference on Music Information Retrieval ISMIR (2009)*, Kobe, Japan.
- Paderewski, I. J. (1999). *Chopin Complete Works, XII Rondos*. The Fryderyk Chopin Institute, Warsaw, Poland, 14 edition.
- Paderewski, I. J. (2006). *Chopin Complete Works, VI Sonatas*. The Fryderyk Chopin Institute, Warsaw, Poland, 17 edition.
- Palmer, C. (1989). Mapping musical thought to musical performance. *Journal of Experimental Psychology: Human Perception and Performance*, 15:331–346.
- Palmer, C. (1996a). Anatomy of a performance: Sources of musical expression. *Music Perception*, 13(433-453).
- Palmer, C. (1996b). On the assignment of structure in music performance. *Music Perception*, 14:23–56.
- Palmer, C. and van de Sande, C. (1993). Units of knowledge in music performance. *Journal of Experimental Psychology*, 19(2):457–470.
- Palmer, C. and van de Sande, C. (1995). Range of planning in music performance. *Journal of Experimental Psychology: Human Perception and Performance*, 21(5):947–962.
- Pardo, B. and Birmingham, W. (2002). Improved score following for acoustic performances. In *International Computer Music Conference (ICMC)*, Goteborg, Sweden. The International Computer Music Association.
- Pickens, J. (2001). Feature selection for polyphonic music retrieval. In *Proceedings of the 24th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 428–429, New Orleans, Louisiana, USA.

- Ramsay, J. and Silverman, B. (2005). *Functional Data Analysis (2nd ed)*. Springer.
- Ramsay, J. O. (2003). Functional data analysis of continuous judgments in music cognition. FDA Workshop during the International Meeting of the Psychometric Society. Presentation.
- Raphael, C. (2006). Aligning music audio with symbolic scores using a hybrid graphical model. *Machine Learning*, 65(2–3):389–409.
- Rasch, R. A. (1979). Synchronization in performed ensemble music. *Acustica*, 43:121–131.
- Recordare (2003). MusicXML definition. <http://www.recordare.com/xml.html>.
- Repp, B. (1995). Expressive timing in Schumann’s “Träumerei”: An analysis of performances by graduate student pianists. *Journal of the Acoustical Society of America*, 98:2413–2427.
- Repp, B. (1996). The art of inaccuracy: Why pianists’ errors are difficult to hear. *Music Perception*, 14(2):161–184.
- Repp, B. H. (1992). Diversity and commonality in music performance - an analysis of timing microstructure in Schumann’s “Träumerei”. *Journal of the Acoustical Society of America*, 92(5):2546–2568.
- Repp, B. H., Windsor, W. L., and Desain, P. (2002). Effects of tempo and timing of simple musical rhythms. *Music Perception*, 19(4):565–594.
- Sinkovicz, W. (1989). Chopin für Magaloff Fans. *Die Presse*, Mittwoch, 18. Jänner [Wednesday, January 18].
- Stadler, P. (1989). Chopin-Marathon. *Der Standard*, 18. Jänner [January 18].
- Vitouch, O. (2005). Erwerb musikalischer expertise [Acquisition of musical expertise]. In Stoffer, T. H. and Oerter, R., editors, *Allgemeine Musikpsychologie (Enzyklopädie der Psychologie)*, volume D/VII/1, pages 657–715. Hogrefe, Göttingen, Germany.
- Wagner, R. A. and Fischer, M. J. (1974). The string-to-string correction problem. *Journal of the Association of Computing Machinery*, 21(1):168–173.
- Widmer, G., Dixon, S., Goebel, W., Pampalk, E., and Tobudic, A. (2003). In search of the Horowitz factor. *AI Magazine*, 24(3):111–130.
- Widmer, G., Flossmann, S., and Grachten, M. (2009). YQX plays Chopin. *AI Magazine*.
- Widmer, G. and Goebel, W. (2004). Computational models of expressive music performance: The state of the art. *Journal of New Music Research*, 33(3):203–216.
- Williamon, A. (2004). *Musical Excellence*. Oxford University Press., Oxford.

Windsor, W. L., Desain, P., Penel, A., and Borkent, M. (2006). A structurally guided method for the decomposition of expression in music performance. *Journal of the Acoustical Society of America*, 119(2):1182–1193.

Zimmermann, E. (1976-2004). *Chopin Komplette Werke, Urtext*. G. Henle Verlag, Munich, Germany.