Proceedings of the 11th International Conference on Music Perception and Cognition (ICMPC 11). Seattle, Washington, USA.
S.M. Demorest, S.J. Morrison, P.S. Campbell (Eds)

THE MAGALOFF CORPUS: AN EMPIRICAL ERROR STUDY

Sebastian Flossmann[1], Werner Goebl[2], and Gerhard Widmer[1, 3]

[1]Department of Computational Perception, Johannes Kepler University, Linz, Austria
[2]Institute of Musical Acoustics, University of Music and Performing Arts Vienna, Austria
[3]The Austrian Research Institute for Artificial Intelligence (OFAI)

ABSTRACT

Musicians at all levels of proficiency must deal with performance errors and have to find strategies for avoiding them. As their level of skill increases, errors occur less frequently, tend to be more subtle, and mostly go unnoticed by the audience. The phenomena of performance errors have been investigated before. However, precise performance data is difficult to acquire, as the automatic extraction of information related to timing and dynamics from audio data is still not possible at the required level of precision. Hence, most studies focus on data gathered on computer-controlled pianos under laboratory conditions. We present a study conducted on a unique corpus of precisely measured performances: the complete works for solo piano by Chopin, performed on stage by the Russian pianist Nikita Magaloff. The data was recorded on a Bösendorfer SE computer-controlled grand piano in a series of public recitals in Vienna in 1989. In this first large-scale error study, we examine Magaloff's performances from qualitative, quantitative, and perceptual standpoints.

1. INTRODUCTION

Musicians at all levels of proficiency must deal with performance errors and have to find strategies for avoiding them. As their level of skill increases, errors occur less frequently and also seem to mostly go unnoticed by the audience. Many factors contribute to producing an error, among them technical deficiencies, lack of concentration, and poor memorization. Which of the errors the audience in fact notices depends not only on the perceptual salience of the error but also greatly on the listeners' musical abilities and acquaintance with the played piece. We investigate the production side of performance errors: How many errors are in fact made; when and under what circumstances do they occur; and what is the most likely cause.

The main problem in studying this phenomenon in an empirical way is the acquisition of data that is both representative and precisely measured. While audio recordings abound, extracting information related to timing, dynamics, and articulation automatically is still not possible at the level of precision required for large-scale music performance studies. Current techniques for audio transcription focus on extracting pitches and their respective onsets. Although the overall precision is promising, parts with low intensity or extensive use of the sustain pedal are still not sufficiently well recognized. Extracting the dynamics of individual notes from audio recordings is virtually impossible. On the whole this makes audio recordings unusable for studying performance errors.

The study presented in this paper has been conducted on a unique corpus of precisely measured performances that were recorded by a highly skilled artist in a concert situation: the complete works for solo piano by Chopin, performed publically by the Russian pianist Nikita Magaloff on a Bösendorfer SE computer-controlled grand piano. This first large-scale error study, though still incomplete in many respects, gives a first impression of Magaloff's performance errors from qualitative, quantitative and perceptual angles.

The work closest to our analysis is (Repp, 1996), which focuses on performance errors and their perceptual salience under laboratory conditions. After a short rehearsal period, ten graduate piano students played four short piano pieces repeatedly on a Yamaha Disklavier. All errors were identified and classified in the recorded MIDI data. In a listening experiment, musicians, partly acquainted with the pieces in question, would then try to detect the errors. From the low number of actually detected errors, the author concluded that most errors are perceptually inconspicuous and only a very small fraction is likely to be noticed by a concert audience. Focusing on the production rather than on the perceptual aspects, (Palmer & van de Sande, 1993) investigated how performance errors may shed light on the way performers memorize and organize music in their memory.

2. THE DATA CORPUS

This work is part of a series of music performance studies centred on a unique resource of precisely measured performance data: The *Magaloff Corpus*. The Russian pianist Nikita Magaloff (1912-1991) was best known for his performance cycles of Chopin's entire works for solo piano. One of the final cycles took place at the Vienna Konzerthaus in spring 1989: Magaloff performed Chopin's entire work for solo piano in strictly chronological order in six public appearances. The concerts were performed and recorded on a Bösendorfer computer-controlled grand piano. The recorded data comprises over 10 hours of continuous playing, over 150 pieces or more than 320.00 performed notes, precisely documenting the temporal and dynamic information for each played note. Tables 1 and 2 show an overview of the data. A more detailed description of the corpus and several other studies conducted on the data can be found in (Flossmann, Goebl, Grachten, Niedermayer, & Widmer, 2010).

ISBN: 1 876346 62 0 © 2010 ICMPC 11

Category	Pieces	Score Notes	Played Notes	Matches	Insertions	Omissions	Substitutions
Ballads	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
Scherzos	4	21951	22633	20849	1369	707	376
Sonatas	12	38971	40450	37015	1651	1498	731
Waltzes	8	18651	18876	18178	461	675	237

Table 1: The Magaloff Corpus by piece category. The generic category *Pieces* includes: Introduction & 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.

3. DATA PREPARATION AND ERROR IDENTIFICATION

In order to judge if a note was played correctly (and which notes were not played), the performance and the score must be aligned: each score note must be matched to its counterpart in the performance or be marked as an *omission*; each performance note must be matched to its counterpart in the score or be marked as an *insertion.* We first digitized and converted all score sheets[1] into a machine-readable MusicXML[2] representation, using the optical music recognition software SharpEye[3]. The recorded MIDI data were then matched to the symbolic scores using an automatic matcher based on the edit-distance paradigm (see (Grachten, 2006) for details on the technique), followed by laborious manual corrections.

Of course, we cannot be certain which edition of the scores Magaloff used or learned the pieces from. In fact, for some pieces it is obvious that he used a score edition different from ours. Sequences of consecutive insertion and/or omission errors without intermediate matched notes mark passages where this is the case. Such passages were excluded from subsequent analyses. The range of the excluded passages varies from short figurations that were repeated too many or too few times, e.g., in the last bars of Nocturne Op.9 No.2, to tens of bars in the Sonata Op.4 Mv.1 in which the performed music differs considerably from our edition of the score.

Pieces/Movements	155
Playing Time	10h7m52s
Score Notes	328.800
Performed Notes	335.542
Matched Notes	307.900
Inserted Notes / Insertion Rate	12.325 / 3.67%
Omitted Notes / Omission Rate	11.506 / 3.5%
Substituted Notes / Substitution Rate	5.105 / 1.55%
Performance notes matched to trills	5923
Grace Notes Matched	4289
Grace Notes omitted	449

Table 2: An Overview over the Magaloff Corpus.

All remaining notes were categorized as *match*, *insertion* (a performed note that has no matching score note), *omission* (a score

[1] We used the Henle Urtext Edition with the exception of the Rondos (Op. 1, 5 & 16) and the Sonata Op. 4 for which we were forced to use the obsolete Paderewski Editions.

[2] See http://www.recordare.com/xml.html

[3] See http://www.visiv.co.uk

note that was not played), or *substitution* (a score note that was performed with a different pitch). In case a performance/score note pair could be interpreted as both a substitution and a combination of insertion and omission, we favoured the latter.

Table 2 gives an overview of the data in the Magaloff 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, and were hence excluded from the data. Trills constitute many-to-one matches of several performance notes to a single score note. As a trill is counted as one match operation, the number of performance notes matched to a trill has to be accounted for separately when counting the performance notes in the corpus. Accordingly, the complete number of performed notes is composed of the number of matches, substitutions, insertions, matched grace notes, and performance notes matched to trills. The complete number of score notes is composed of the number of matches, substitutions, omissions, and matched and omitted grace notes.

4. STATISTICS AND DISCUSSION

In the following section, we describe several investigations of Magaloff's performance errors: first, a purely quantitative assessment of the error counts; second, considerations addressing the issue of perceptual salience; and third, preliminary qualitative investigations into two specific error patterns found repeatedly in the performances.

4.1 Quantitative Aspects

Table 2 shows the number of errors and error rates found in the Magaloff corpus: 3.67% of all performed notes are insertions, 3.5% of all score notes were omitted, and 1.55% of all matched notes were played at a wrong pitch. This exceeds the percentages reported in (Repp, 1996) (1.08% insertions, 1.64% omissions, and 0.26% substitutions). Looking only at the Chopin piece Repp used in his study (Prelude Op.28/15), we encounter error rates that are more similar: 0.72%/1.58%/0.52% (Magaloff) vs. 0.98%/1.48%/ 0.21% (Repp).

Table 1 gives an overview of the categories of pieces with their respective sizes and error numbers. Assessing the pieces by category, the Scherzos 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 also the category with the lowest percentage of deletion errors (2.22%), while Etudes and Polonaises exhibit the highest percentage of deletions (above 4.5%).

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 played per time unit (3 seconds), the greater the number of errors. This holds to a varying degree for all kinds of errors. Overall, the corpus exhibits correlation coefficients between note density and frequency of (a) insertion errors, (b) deletion errors and (c) substitution errors of 0.39, 0.26 and 0.61, respectively. Table 3 shows the correlation coefficients

Category	IR[%]	$r_{i\text{-}nd}$	OR[%]	$r_{o\text{-}nd}$	SR[%]	$r_{s\text{-}nd}$
Ballads	**4.95**	**0.63**	2.55	**0.73**	1.29	**0.55**
Etudes	3.95	0.05	**4.12**	0.39	1.37	0.34
Impromptus	1.31	-0.01	2.20	-0.05	0.87	0.39
Mazurkas	2.41	0.29	3.57	0.47	1.01	0.39
Nocturnes	2.10	0.42	2.82	**0.64**	0.98	0.46
Pieces	4.22	0.02	3.76	0.41	**2.32**	0.38
Polonaises	**5.65**	**0.53**	**4.35**	0.40	1.60	0.13
Preludes	3.37	0.27	3.15	-0.22	1.60	**0.66**
Rondos	1.79	0.35	2.53	0.31	**2.52**	0.37
Scherzos	**6.06**	0.25	3.23	0.43	1.72	0.11
Sonatas	4.19	**0.59**	**3.86**	0.58	**1.89**	**0.70**
Waltzes	2.44	0.35	3.62	**0.76**	1.27	0.51

Table 3: Error rates (Insertion Rate (IR), Omission Rate (OR), Substitution Rate (SR)) and corresponding correlations with note density ($r_{i\text{-}nd}$, $r_{o\text{-}nd}$, and $r_{s\text{-}nd}$); the three highest values per column in bold font

of error frequency and note density for the respective categories of pieces. The Ballads and Polonaises both show a high error percentage as well as a high correlation of error frequency and note density. This may indicate that these are technically particularly demanding.

4.2 Perceptual Aspects

One of the main hypotheses of (Repp, 1996) is that skilled musicians avoid errors that are obvious. Whether an error is conspicuous is closely related to several factors:

- How loud was an added note played in relation to the other notes in the vicinity?
- How well does an insertion/substitution note fit into the harmonic context, or how important was an omitted note for the harmony?
- Is the error located in a melody or an inner voice, and how many simultaneous voices surround it?

We compared the loudness of each insertion note with the average loudness of all correct notes in the immediate vertical vicinity (notes with the same onset). The loudness of the majority (65%) of erroneously inserted notes amounts to 40%-100% of the loudness of the surrounding notes (their numbers follow approximately a uniform distribution in that range). 16% are louder than the average note with the same onset, 19% are inserted at less than 40% of the average loudness.

A very important aspect of the perceptual salience of an inserted/substituted note is how well it fits into the local harmonic context. Considering the vast number of errors, an assessment by listening, as done in (Repp, 1996), is not feasible. Instead, we estimated the consonance of an insertion/substitution error with respect to the local harmony. (Temperley, 2007) derived key profiles for major and minor scales from the *Essen Folksong Collection* that rate the probability of occurrence of pitch classes within the context of a given harmony. We used these profiles to determine automatically the most likely local harmony given the pitches that were identified as correctly played. To judge the consonance of an erroneous note within an estimated local harmony, we used the key-profiles proposed by Krumhansl and Kessler (Krumhansl & Kessler, 1982). The profiles were established via probe-tone experiments and rate how well a pitch class fits into a harmonic context. The normalized values range from 0.0534 for the least consonant (minor second upwards from the tonic) to 0.1522 for the most consonant (tonic) pitch classes in a scale. If we assume that pitches with a value below 0.0603, the value of the tritone in the major scale, are perceived as harmonically inappropriate[4], then 46% of all insertions and 44% of all substitutions are not compatible with the local harmony.

Localising the errors addresses the third facet mentioned above. The omission rate in the melody voices is 1% as compared to 4.1% for voices not belonging to the melodic part of the score[5]. Table 4 shows the error rates by staff and surrounding musical texture: The staff-wise insertion (omission) rates (columns *All* in the table) were calculated relative to the overall number of performed notes (score notes) in the respective staff. To assess errors with respect to the surrounding musical texture, we defined an onset to be *mono-voiced* (relative to a staff) if there is only one score note present, and *multi-voiced* otherwise. The texture-specific insertion (omission) rates (columns *Multi* and *Mono*) are calculated relative to all performance notes (score notes) present in the respective texture and staff. Overall, in the upper staff, insertions are more likely to occur than omissions and vice versa in the lower one. For omission errors the difference between multi-voiced and mono-voiced situations is obvious: A single note is less likely to be omitted, particularly if the note is located in the right hand. Insertion errors are slightly more likely in multi-voice situations, but the difference is less striking.

	IR[%]			OR[%]		
	Multi	Mono	All	Multi	Mono	All
Upper Staff	4.67	3.24	4.00	4.01	0.87	2.55
Lower Staff	3.65	3.42	3.55	5.71	2.33	4.34

Table 4: Insertion and Omission Rates by staff and surrounding musical texture.

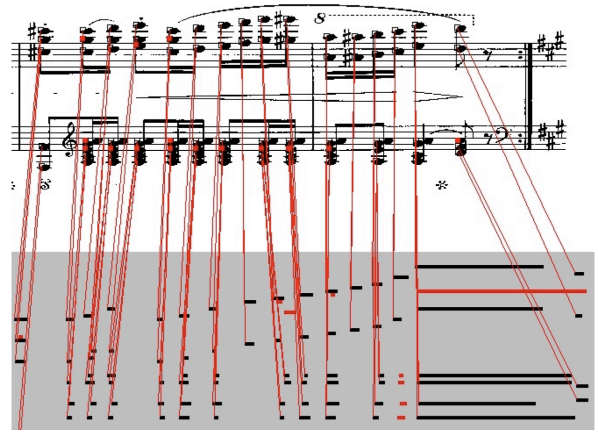

Figure 1: Chopin Polonaise, Op.40/1, Bars 79-80; Magaloff's performance (lower panel) matched to the score (upper panel); red squares in the score identify omissions, red performance notes are insertions. (upper panel).

4.3 Qualitative Aspects

An error pattern recurring very often throughout the corpus is the complete omission of an inner voice. Figure 1 shows two bars from the Polonaise in A major, Op. 40 No. 1: In the left hand, the "e" is left out throughout the whole sequence of chords. The accompaniment consists of a repetition of the same A^7 chord at a medium tempo and can be considered technically not particularly difficult. Since the "e" is not the most important voice of the A^7 chord, the omission can probably be attributed to a memory problem.

Other examples of an omitted inner voice can be found very frequently in the data. However, the perceptual salience and probable cause of the omissions vary. Bar 16 of the Etude Op. 25 No. 10 (the first bar shown in Figure 2) shows an omission of the inner voices in both the left and the right hand. Compared to the Polonaise, the voice is perceptually much more prominent. Clearly, this omission cannot be attributed to memory. It is much more likely that Magaloff simplified the passage technically.

Parallel octaves, as found most prominently in the Etude Op. 25 No.10, seem to pose a particular technical problem to Magaloff. In the second bar shown in Figure 1, the inner voice in both the right and the left hand (half notes f and g) should only be struck once and held throughout the passage, which makes the passage

[4] For a major scale those are: minor second, minor third, tritone, minor sixth, and minor seventh; for a minor scale those are: minor second, major third, tritone, and major sixth.

[5] We assume that the highest pitch in the upper staff at any given time is the melody voice of the piece. In the case of Chopin, this very simple heuristic is correct often enough (though not always) to be justifiable. We simply do not have the resources to manually identify and label all melody notes in Chopin's complete piano works.

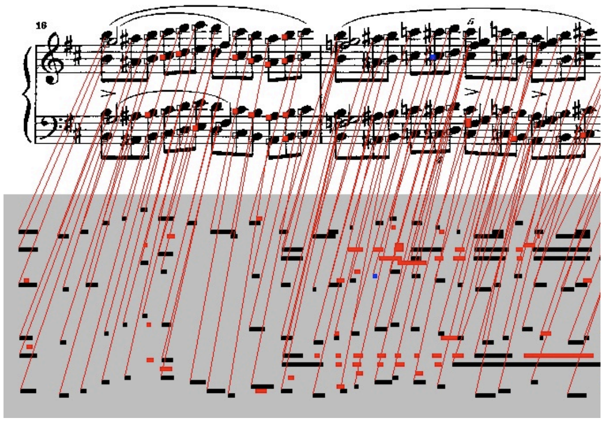

Figure 2: Chopin Etude Op.25/10, Bars 16-17; Magaloff's performance (lower panel) matched to the score.

considerably more difficult. Magaloff repeats the inner voices with almost every onset (see the inserted notes in the lower panel of Figure 2), which conserves the harmonic content but simplifies the passage. Other instances of parallel octaves, e.g., in the Nocturne Op. 48 No.1, bars 46-48, exhibit an increased amount of insertion notes within the octaves. On the whole, only a few instances of parallel octaves were performed with a small number of errors, indicating a particular technical deficiency (which may also be related to Magaloff's age at the time of the recordings).

5. CONCLUSION AND FUTURE WORK

This study forms part of a series of empirical investigations into a unique collection of a highly skilled artist's live concert performances. The findings corroborate in part previous studies conducted under laboratory conditions. Preliminary analyses of the inserted and substituted notes from a harmonic point of view, however, suggest that the errors are more salient than initially assumed. Impressions from informal listening experiments of a reproduction of the data on a Bösendorfer CEUS (the new computer-controlled concert grand by Bösendorfer) confirm this.

If we may make a somewhat speculative comment here, the fact that Magaloff did not reduce his performance tempi even at age 77 (see (Flossmann, Goebl, & Widmer, 2009) for a more detailed inspection of tempo in Magaloff's Etudes) and that his performances display relatively high error rates might indicate that Magaloff's aim was realising his musical ideas of Chopin's work rather than at error-free performances.

The corpus offers a unique view into the wide range of possible performance errors. For a further, more detailed analysis it will be necessary to build a fine-grained typology of performance errors. Identifying relations between characteristics of the score and specific error patterns may help to determine the technical complexity of a piece. Insights into those connections can be helpful for educational purposes, e.g., automatic performance assessment and practice supervision.

6. ACKNOWLEDGEMENTS

This work is funded by the Austrian National Research Fund FWF via grants P19349-N15 and Z159 ("Wittgenstein Award").

7. REFERENCES

Flossmann, S., Goebl, W., & Widmer, G. (2009). Maintaining skill across the life span: Magaloff's entire Chopin at age 77. *Proceedings of the International Symposium on Performance Science 2009*. Utrecht, The Netherlands: European Association of Conservatoires (AEC).

Flossmann, S., Goebl, W., Grachten, M., Niedermayer, B., & Widmer, G. (2010). The Magaloff Project: An Interim Report. *Journal of New Music Research (to appear)*.

Grachten, M. (2006). Expressivity-Aware Tempo Transformations of Music Performances Using Case Based Reasoning. *PhD Thesis*.

Krumhansl, C. L., & Kessler, E. J. (1982). Tracing the dynamic changes in perceived tonal organization in a spatioal representation of musical keys. *Psychological Review, 89*, 334-368.

Palmer, C., & van de Sande, C. (1995). Range of Planning in Music Performance. *Journal of Experimental Psychology: Human Perception and Performance, 21* (5), 947-962.

Palmer, C., & van de Sande, C. (1993). Units of Knowledge in Music Performance. *Journal of Experimental Psychology, 19* (2), 457-470.

Repp, B. H. (1996). The Art Of Inaccuracy: Why Pianist's Errors Are Difficult To Hear. *Music Perception, 14* (2), 161-148.

Temperley, D. (2007). *Music and Probability*. Cambridge, MA: The MIT Press.

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