KUNGL TEKNISKA HÖGSKOLAN

TRITA-TMH 2000:9
ISSN 1104-5787
ISBN 91-7170-643-7

Virtual Virtuosity

Studies in
Automatic Music Performance

av

Roberto Bresin

Akademisk avhandling som med tillstånd av Kungliga Tekniska Högskolan i Stockholm framlägges till offentlig granskning för avläggande av teknologie doktorsexamen fredagen den 1 december kl 16.00 i sal E1, Lindstedtsvägen 3 Entrepl, KTH, Stockholm. Avhandlingen försvaras på engelska.

Tal Musik Hörsel, http://www.speech.kth.se
Kungliga Tekniska Högskolan, http://www.kth.se
Stockholm 2000

Dissertation for the degree of Ph.D. in Music Acoustics to be presented with due permission for public examination and criticism in the Aula Magna E1 at KTH, Stockholm, on December 1, 2000, at 4 p.m.

Respondent:	Roberto Bresin, M.Sc.
Opponent:	Roger Dannenberg, Senior Researcher Carnegie Mellon University
Examination committee:	Doc. Anders Askenfelt, KTH, Stockholm Prof. John Bowers, KTH, Stockholm Patrik Juslin, Ph.D., Uppsala University
Supervisor:	Prof. Johan Sundberg

Front cover: CPE Bach's (1753) instructions for the performance of *staccato* notes and an analysis of *staccato* articulation as reported by Bresin and Widmer (2000).

Speech Music and Hearing, KTH, Stockholm, Sweden, 2000
Printed at Universitetsservice US AB, Stockholm, Sweden

TRITA-TMH 2000:9
ISSN 1104-5787
ISBN 91-7170-643-7

a Christine

Contents

Included parts *vi*

Abstract *vii*

Acknowledgments *viii*

Glossary *ix*

Introduction **1**

Previous research **1**

Method: a virtual performer **3**

Rule-based model 4

Paper I. Artificial neural network based model 4

Paper II. ANNs vs rules: musical punctuation at the micro-level 6

Papers III and IV. Articulation in piano music performance **7**

Legato articulation 8

Staccato articulation 9

Articulation of repetition 9

Rules for articulation in automatic piano performance 9

Analogies with step movements 13

Paper V. Emotional coloring of music performance **14**

Paper VI. Applications **16**

Automatic music performance on the Internet 16

Automatic analysis of the emotional color of music performance 18

Gestural rendering of synthesized sounds 18

Music education 19

Cellular phones 19

MPEG-7 20

Conclusions **20**

References **22**

Appendix I. Electronic appendixes **31**

Appendix 2. Co-authors' roles **32**

Included parts

The dissertation consists of a summary and the following parts:

Paper I	Bresin, R. (1998). Artificial neural networks based models for automatic performance of musical scores. *Journal of New Music Research* 27(3): 239-270
Paper II	Friberg, A., Bresin, R., Frydén, L., & Sundberg, J. (1998). Musical punctuation on the microlevel: Automatic identification and performance of small melodic units. *Journal of New Music Research,* 27 (3):271-292
Paper III	Bresin, R., & Battel, G.U. (forthcoming). Articulation strategies in expressive piano performance. *Journal of New Music Research*
Paper IV	Bresin, R., & Widmer, G. (2000). Production of *staccato* articulation in Mozart sonatas played on a grand piano. Preliminary results. *Speech Music and Hearing Quarterly Progress and Status Report*, Stockholm: KTH, 2000(4):1-6
Paper V	Bresin, R., & Friberg, A. (forthcoming). Emotional coloring of computer controlled music performance. *Computer Music Journal*, 24(4):44-62
Paper VI	Bresin, R., & Friberg, A. (2000). Software tools for musical expression. In I. Zannos (Ed.) *Proceedings of the International Computer Music Conference 2000,* San Francisco: International Computer Music Association, 499-502
Software I	Bresin, R. (1998). "JAPER and PANN: two JAVA applets for music performance." Included in the CD-ROM *MidiShare: Operating System for Musical Applications*, Lyon: National Center of Contemporary Music - GRAME, http://www.grame.fr

The papers will be henceforth referred to by their Roman numerals. Figures and tables will be referred to in the same way as they appear in their respective papers.

Abstract

This dissertation presents research in the field of automatic music performance with a special focus on piano.

A system is proposed for automatic music performance, based on artificial neural networks (ANNs). A complex, ecological-predictive ANN was designed that *listens* to the last played note, *predicts* the performance of the next note, *looks* three notes ahead in the score, and plays the current tone. This system was able to learn a professional pianist's performance style at the structural micro-level. In a listening test, performances by the ANN were judged clearly better than deadpan performances and slightly better than performances obtained with generative rules.

The behavior of an ANN was compared with that of a symbolic rule system with respect to musical punctuation at the micro-level. The rule system mostly gave better results, but some segmentation principles of an expert musician were only generalized by the ANN.

Measurements of professional pianists' performances revealed interesting properties in the articulation of notes marked *staccato* and *legato* in the score. Performances were recorded on a grand piano connected to a computer. *Staccato* was realized by a micropause of about 60% of the inter-onset-interval (IOI) while *legato* was realized by keeping two keys depressed simultaneously; the relative key overlap time was dependent of IOI: the larger the IOI, the shorter the relative overlap. The magnitudes of these effects changed with the pianists' coloring of their performances and with the pitch contour. These regularities were modeled in a set of rules for articulation in automatic piano music performance.

Emotional coloring of performances was realized by means of macro-rules implemented in the Director Musices performance system. These macro-rules are groups of rules that were combined such that they reflected previous observations on musical expression of specific emotions. Six emotions were simulated. A listening test revealed that listeners were able to recognize the intended emotional colorings.

In addition, some possible future applications are discussed in the fields of automatic music performance, music education, automatic music analysis, virtual reality and sound synthesis.

Keywords: music, performance, expression, interpretation, piano, automatic, artificial neural networks, rules, articulation, legato, staccato, emotion, virtual reality, human computer interaction, perception, music education, Director Musices, JAPER, PANN, computer music, MIDI, MidiShare, Disklavier, Bösendorfer, cellular phone, mobile phone, MPEG-7, Java, Lisp

Acknowledgments

I am most grateful to Professor Johan Sundberg who guided and assisted me during the four years of this work with great enthusiasm, trust, and stimulating discussions. Without him this work would not have been possible.

I would also like to express my gratitude to Anders *Fryed* Friberg, with whom I established a fruitful and creative collaboration.

I am indebted to Peta White who cleaned up my English in many papers as well as in this dissertation. My working hours have been gilded by the joy and friendship of the members of the fantastic *Musikakustikgruppen* at the Speech Music and Hearing Department at KTH: those not already mentioned are Anders Askenfelt, Sofia Dahl, Svante Granqvist, Jenny Iwarsson, Erik Jansson, Eric Prame, Sten Ternström, Monica Thomasson. A special thanks goes to Cathrin Dunger for the editing of papers V and VI.

My gratitude also goes to the co-authors and reviewers of the papers included in this dissertation; to Diego Dall'Osto and Patrik Juslin for valuable discussions; and to the anonymous listeners who participated in the listening tests reported in this work.

My *Grazie!* goes to Professors Giovanni De Poli and Alvise Vidolin with whom I started my research at Padua University; to my piano teacher Vincenzina Dorigo Orio; to Roberto Galloni, Scientific Attaché at the Italian Embassy in Stockholm, and to Rosino Risi, director of the Italian Institute of Culture in Stockholm.

Finally, I thank my family for loving support.

This work was supported by the EU Fourth Framework Training and Mobility of Researchers (TMR) Program (Marie Curie scholarship ERBFMBICT950314), and by the Bank of Sweden Tercentenary Foundation.

The preliminary research on *staccato* was made possible by a research visit to the Austrian Research Institute for Artificial Intelligence, Vienna, financed by the START project Y99-INF (Austrian Federal Ministry for Education, Science, and Culture) and the EU Fifth Framework Project HPRN-CT-2000-00115 (MOSART).

Glossary

ANN	Artificial Neural Network
DM	The Director Musices program
DR	Tone Duration
DRO	Offset-to-Onset Duration
HMI	Human-Machine Interaction
IOI	Inter-Onset-Interval
JAPER	The “Java Performer” Applet
KDR	Key Detached Ratio
KDT	Key Detached Time
KOR	Key Overlap Ratio
KOT	Key Overlap Time
PANN	The “Punctuation with Artificial Neural Networks” Java Applet
PC	Personal Computer
SL	Sound level
VR	Virtual reality

Introduction

Since the design of the first computer a tempting perspective has been to replicate human behavior with machines. Nowadays humanoids can walk (Pandy and Anderson 2000), dance (Lim, Ishii, and Takanishi 1999), play the piano, talk, listen and answer (Kato, Ohteru, Shirai, Narita, Sugano, Matsushima, Kobayashi and Fujisawa 1987). Yet, these machines lack the ability to understand and process the emotional states of real humans and to develop and synthesize an emotional state and personality of their own. To overcome this limitation research on music performance seems particularly promising, since music is a universal communication medium, at least within a given cultural context. Music performances mostly are emotionally colored, and hence measurements of performances provide data on the code used for such coloring. Also, interviews with and instructions to performers can supply *a priori* knowledge of their expressive and emotional intentions (Gabrielsson and Juslin 1996; Bresin and Battel 2000; De Poli, Rodà and Vidolin 1998) and formal listening tests can provide information to listeners on the communication of performers' intentions (Juslin 1997c; Bresin and Friberg 2000).

Research on music performance has revealed interesting analogies in the communication of emotions in singing and speech (Sundberg 2000; Bresin and Friberg 2000). Also analogies between body movement patterns and music performance have been noticed with respect to final *ritardandi* (Friberg and Sundberg 1999) and further investigations will probably reveal more analogies of this kind. These observations suggest that research on music performance represents a promising starting point for understanding human behavior.

Studies in music performance have a particular value in our time. The art of performing music is the result of several years of training. At the same time, contemporary information technology offers the possibility of automatic playing of music specially composed for computers or stored in large databases, e.g. on the Internet. In such case, the music is typically played deadpan, i.e., exactly as nominally written in the score, thus implicitly ignoring the value of a living performance and its underlying art and diversity. Objective data on music performance are needed in the defense of humanity's cultural heritage. Research on music performance can also provide expressive tools that traditionally have been hiding in musicians' skill and musical intuition. When explicitly formulated these tools will give the user the possibility to play music files with different expressive coloring.

Results from research in music performance can be used in the development of new applications in a number of contexts, such as music education, human-machine interaction (HMI), the entertainment industry, cellular phone ringing tones, and the synthesizer industry to name a few.

Previous research

The principal vehicle for the communication of musical compositions is the music score in which the composer codifies his intentions. Thus the score implicitly includes a cognitive reference to the composition. However, the information written in the score does not represent an exhaustive description of the composer's intentions. The performer renders each note in the score in terms of intensity, duration and timbre by movements of fingers, arms, feet, mouth, chest, etc. They may result in different performances of the same piece reflecting each performer's culture, mood, skill and intention. These differences also contribute to determining the performing styles of different musicians. The performer could thus be regarded the unifying link between the symbolic description (the musical score) and its interpretation. Analogous situations can be found in speech and dance: the performer of a literary text is free to decide on intonation, accents, pauses etc.; likewise, the ideas of a choreographer is realized in terms of the dancer's personal body movements.

Research on music performance has been quite intense in the XX century, particularly in its last decades (for an extensive overview of this research, see Gabrielsson, 1999). Most of the studies have focused on piano music performance. Seashore (1938) and coworkers at Iowa University conducted measurements of performances on a specially prepared piano and found astonishing differences between score notation and its performance. Shaffer (1982, 1984a, 1984b) analyzed rhythm and timing in piano performance, and later Clarke (1999) wrote an overview of the same aspects. Clarke (1988) outlined some generative principles in music performance, and Sloboda (1983) studied how musical meter is communicated in piano performance. In 1994 Parncutt proposed a theory of meter perception that it is based on the prominence of different pulse trains at different hierarchical levels. Palmer (1989) pointed out that differences between performances of the same score reflect the existence of many sources of possible deviations from a strictly mechanical (henceforth deadpan) performance. Repp (1990, 1992, 1995, 1997, 1998a, 1998b) has presented several statistical and quantitative analyses of piano performances. Clynes (1983) claimed the existence of a "composer pulse" characterizing the timing of the beats in the bar in performances of western classical music. Gabrielsson (1994, 1995) analyzed intention and emotional expression in music performance. Gabrielsson and Juslin (1996) outlined the cues in the code used by performers when communicating different intentions to listeners. These cues were used by Juslin (1997c) for the synthesis of performances in his experiment on perceived emotional expression.

Few authors have proposed models of automatic music performance. Todd (1985, 1992) presented a model of musical expression based on an analysis-by-measurement method. Rule-based systems have been proposed by De Poli and coworkers (De Poli, Irone, and Vidolin 1990) and by Friberg and coworkers (Friberg, Frydén, Bodin, and Sundberg 1991; Sundberg 1993; Friberg 1995a; Friberg, Colombo, Frydén and Sundberg 2000). Also fuzzy logic-based rule systems have been tried out (Bresin, Ghetta and De Poli 1995a, 1995b). Performance systems based on artificial intelligence techniques have also been developed. Widmer (1996, 2000) proposed a machine-learning based system extracting rules from performances. Ishikawa and coworkers developed a system for the performance of classical tonal music; a number of performance rules were extracted from recorded performances by using a multiple regression analysis algorithm (Ishikawa, Aono, Katayose and Inokuchi 2000). Arcos and coworkers (Arcos, López de Mántaras and Serra 1998) developed a case-based reasoning system for the synthesis of expressive musical performances of sampled instruments. Dannenberg and Derenyi (1998) proposed a performance system that generates functions for the control of instruments based on spectral interpolation synthesis.

The present work is organized in four main parts.

In the first part a model for automatic music performance is proposed in terms of artificial neural networks (ANNs) which are related to the performance rule system developed at KTH (Paper I). The automatic detection of punctuation marks in a score was used for a comparison of results produced by the ANN-based system and by the KTH rule-based system (Paper II).

In the second part the analysis-by-measurement method is applied in the design of new rules for articulation in expressive piano performance. Performances of fourteen Mozart piano sonatas played on computer-monitored grand pianos were used in this study (Papers III and IV).

In the third part the possibility of producing emotionally colored performances with the KTH system is presented (Paper V).

In the last part some applications and future developments are proposed (Paper VI, Software I).

Method: a virtual performer

The principal characteristic of an automatic performance system is that it converts a music score into an expressive musical performance typically including time, sound and timbre deviations from a deadpan realization of the score. Mostly, two strategies have been used for the design of performance systems, the analysis-by-synthesis method and the analysis-by-measurement method.

The first method implies that the intuitive, nonverbal knowledge and the experience of an expert musician are translated into performance rules. These rules explicitly describe musically relevant factors. A limitation of this method can be that the rules mainly reflect the musical ideas of specific expert musicians. On the other hand professional musicians' expertise should possess a certain generality, and in some cases rules produced with the analysis-by-synthesis method have been found to have a general character.

Rules based on an analysis-by-measurement method are derived from measurements of real performances usually recorded on audio CDs or played with MIDI-enabled instruments connected to a computer. Often the data are processed statistically, such that the rules reflect typical rather than individual deviations from a deadpan performance, even though individual deviations may be musically highly relevant.

A recent tendency in music performance research is the merging of the two methods (c.f. Gabrielsson 1985). Often one method is used to validate the rules obtained by the other method. Also rules are generally validated with listening tests using expert and non-expert subjects.

Rule-based model

One of the most successful methods for automatic expressive music performance has been the rule-based system developed at KTH in Stockholm. It consists of a generative grammar for music performance that includes approximately thirty rules. These rules, obtained mainly by the analysis-by-synthesis method, have been implemented in the Director Musices (DM) program (Friberg 1995; Friberg, Colombo, Frydén and Sundberg 2000). Rules can be combined so as to produce deviations in note duration and intensity, global tempo and intensity, and also in instrument timbre, provided that the instrument allows such effects. Each note can be processed by several rules, and the expressive deviations produced by the different rules are mostly added.

The DM system has been continuously developed over a long period. Papers III and IV present recent complements concerning *staccato* and *legato* articulation. The associated rules are presented in the chapter "Articulation in Piano Music Performance" below. Another recent development of DM is presented in the "Emotional Coloring of Music Performance" chapter (Paper V). In addition, some future possible applications of DM are described in the "Applications" chapter (Paper VI).

Paper I. Artificial neural network based model

Paper I tested the idea of combining the rule-based DM system with Artificial Neural Networks (ANNs), thus proposing a hybrid system for real-time music performance based on the interaction of symbolic and sub-symbolic rules. The main idea was to develop a real-time system for the simulation of the style of a professional pianist. For this reason the system had to be based on local information and therefore it operates at the micro-level of the musical structure.

The ANN model was based on feed-forward networks trained with the error back-propagation algorithm. The DM rules played an important role in the design of performance ANNs. A crucial aspect in the design of ANNs is the choice and the representation of input and output parameters.

For the modeling of performance ANNs, seven rules were chosen as relevant in the production of local deviations in piano performances (Table I in Paper I). The parameters used in these rules were codified and assigned to different input nodes of an ANN (Figure 2 in Paper I). The ANN was trained to learn the seven DM rules mentioned above. During the training phase, the output nodes were trained with the time and intensity deviations produced by the seven rules; in this sense the ANN model presents a strong relationship with the KTH rule-based system.

In Paper I, the results of a listening test for the validation of this ANN model are reported. Twenty subjects volunteered for the experiment. The quality of deadpan performances was compared with that of performances produced by this ANN, and by the seven DM rules that were used for the training. Two melodies from Mozart's piano sonatas K331 and K281 were used. ANN and rules performances received a significantly higher score relative to deadpan performances, and ANN performances were best overall (Figure 6 in Paper I). This result validated our ANN model.

The next step was the modeling of more complex ANNs, capable of learning the performing style of a professional pianist. During the training phase of this ANN, the output nodes were trained with the time and intensity deviations from an expressive performance that had been produced by a professional pianist playing a synthesizer connected to a personal computer (PC). Different ANN architectures were tested in various experiments. These experiments resulted in the design of complex ANN models that were called the ecological ANN and the ecological-predictive ANN. For each note in the score the former produced loudness variations while the latter generated deviations in duration and inter-onset-interval (IOI) (Figures 12 and 17 in Paper I). These ANNs operate on contexts comprising four and five notes, respectively. Analyses of the behaviors of these ANNs showed that the ANNs learned rules similar to the DM's symbolic rules (pages 263-264 in Paper I). In particular, a duration contrast rule was generalized by the ecological-predictive ANN which possessed the same qualitative behavior as the corresponding DM rule. The ANNs could extrapolate the style of a professional pianist from 16 structurally important tones in a performance of Schumann's *Träumerei*. The deviations produced by the ANNs were quite large since the pianist performed the score with a quite passionate style. The same ANNs were used also to generate the performance of an excerpt of a Mozart piano sonata. Here, it was necessary to introduce a fixed damping for the deviations produced by the ANNs. The resulting performance was judged as musically acceptable in an informal listening test.

DM includes a subset of N rules that are simple in the sense that they do not require any particular processing of the score. Therefore, it was hypothesized that deviations produced by the ANNs could be combined with deviations produced by the N rules. Decision rules and user interaction determined which ANN to use each time. This hybrid-system can be formalized with the following equation:

$$\text{(1)} \qquad \mathrm{Y}_n = \sum_{i=1}^{N} k_i \cdot f_i(\bar{x}_n) + net(\bar{k}, \bar{x}'_n)$$

The first term in equation 1 takes into account the DM rules included in the system; N is the number of simple DM rules, $\bar{x}_n$ represents the vector of the DM rule parameters associated with the n-th note, and the $f_i()$ functions represent the various rules, k_i is a constant used to emphasize the deviation generated by each rule.

The second term in equation 1, $net()$, represents a set of possible performance ANNs. The $\bar{k}$ vector corresponds to the selection, made either by decision rules or by the user, of particular ANNs or $f_i()$, and $\bar{x}'_n$ is a vector representing the ANNs input pattern for the n-th note (Figure 1 in Paper I).

A limitation of the ANN-based model can be the difficulty in the choice of ANN structure, and the difficulty in training the ANN, i.e. in choosing and coding input and output training patterns. A common criticism of ANN models is the difficulty in interpreting the behavior of an ANN. In Paper I it is shown that it is possible to explain the ANNs' behavior in terms of rules and thus a new use of ANNs as a tool for performance analysis is suggested. The analysis of deviations produced by performance ANNs can help to identify the relevant set-up of symbolic rules and thus to give a deterministic explanation of a performer's conscious and subconscious preferences.

A slightly modified version of the ecological-predictive ANN model was successfully used for the development of a virtual flutist at Genoa University, Italy (Dillon 1999; Camurri, Dillon, and Saron 2000).

Paper II. ANNs vs rules: musical punctuation at the micro-level

In Paper II, rules and ANNs are used for accomplishing the same task: the marking of melodic structures in a score by inserting micropauses at boundaries separating melodic gestures. These are small structural units, typically consisting of 1 to 7 notes that are perceived as belonging together. The separation of melodical gestures by micropauses will henceforth be referred to as *punctuation*. Structure segmentation at this level seems to depend more on the performer's choices than on more general principles, such as in the case of marking phrase boundaries. Punctuation was found to be important for the emotional coloring of automatically generated piano performances (Paper V).

A punctuation rule system was constructed by means of the analysis-by-synthesis method. It operates on a context of five notes (Appendix in Paper II). An ecological ANN was designed on the basis of this punctuation rule system, using a context of five notes and information about their pitches, durations, and distance from the root of the prevailing chord (Figure 3 in Paper II). A professional musician, Lars Frydén, made a segmentation of fifty-two melodic excerpts. Half of the analyzed melodies were used for the optimization of the DM punctuation rule and the training of the ANN. The performance of these two alternative systems was then tested on the remaining twenty-six melodies. In five cases the ANN approximated the choices of the expert musician better than the rule system, but in general the DM symbolic rule system yielded better results than the ANN. In most excerpts, the punctuation ANN introduced more segmentation points than the punctuation rule system. However, most of these points were judged to appear in musically acceptable positions in informal listening tests. The performance of the ANN varied between the excerpts. In one excerpt, the ANN's markings matched all of those made by the musician, while the rule system succeeded in identifying fewer in this case (Table 2 in Paper II). This may suggest that punctuation is style-dependent. It is also likely that the ANN generalized punctuation principles not yet implemented in the punctuation rule system. A further analysis of these aspects would be worthwhile.

Different versions of punctuation ANNs were implemented in the JAVA applet PANN (Punctuation with ANN, Software I). In PANN it is possible to control the output of the ANNs in terms of the number of punctuation points in a score.

The PANN system has been applied in the Anima animation program (Lundin 1992; Ungvary, Waters, and Rajka 1992). Here, the micropauses introduced in a score by the PANN were connected to MIDI controls such that each micropause was associated with a particular gesture of a virtual dancer, e.g., a pirouette. The possibility to combine automatic performance of music scores with virtual choreography should be further explored in more depth in the future.

Papers III and IV. Articulation in piano music performance

In the past, few researchers have paid attention to the analysis of articulation in piano performance. This could be due to two main reasons. First, it is difficult to detect the instant when a key is released and when the associated sound has passed the threshold of audibility. Second, a precise measurement of the mechanical movements of piano keys and hammers is possible only in commercial MIDIfied pianos like Disklavier and Bösendorfer, or in pianos provided with various sensors, such as photocells as used by Shaffer (1981) and accelerometers on the hammers and the keys as used by Askenfelt and Jansson (1990).

Mathews (1975) observed that tone overlap was required in order to produce a *legato* effect in tones generated by electroacoustic means. His observation was later corroborated by Purbrick (2000) who pointed out the difficulty in producing expressive performances with computer-controlled synthesizers, and proposed an automatic generation of *legato* articulation in a guitar sound generated with a physical model-based synthesis.

In investigations of articulation in both digital and acoustic piano playing, Repp had professional pianists perform scales and *arpeggi* at different tempi according to a flashing metronome (Repp 1995, 1997, 1998b). He examined both perception and production of *legato* and *staccato* articulation and found that an acoustic overlap was required to produce a *legato* while a micropause was needed to produce a *staccato.* Palmer (1989) reported that in *legato* articulation the IOI between two overlapping notes is a major factor for the amount of overlap. Gabrielsson and Juslin pointed out how articulation, with its variations, is one of the most important and effective cues in communication and perception of emotional character in music performance (Gabrielsson 1994, 1995; Gabrielsson and Juslin 1996).

Two performance databases were used in this study. One consisted of performances by five diploma students who played the Andante movement of W A Mozart's Piano Sonata in G major, KV 545. They were asked to play the piece in nine different performance styles on a Disklavier connected to a PC. The styles were given in terms of adjectives (*bright, dark, heavy, light, hard, soft, passionate, flat,* and *natural,* i.e. in the way preferred by the pianist). The other database consisted of recordings of thirteen Mozart's piano sonatas played by a professional pianist on a Bösendorfer grand piano that was connected to a PC. For the analysis presented in Paper III only the notes played with the right hand were considered.

The data available in the two databases allowed an analysis of articulation focused on the movement of the piano keys and not on the acoustic realization. Apart from the IOI, four parameters have been used here for describing articulation when performed on a piano. The first is the key overlap time (KOT), i.e., the time during which the keys corresponding to two successive notes are depressed simultaneously. The second parameter is the key detach time (KDT), defined as the time during with neither of two keys corresponding to two successive notes is depressed, such that there is a micropause between the tones. The Key Overlap Ratio (KOR) refers to the ratio between the KOT and the IOI. The Key Detached Ratio (KDR) refers to the ratio between the KDT and the IOI. The definitions of these measures are illustrated in Figure 2 in Paper III.

Legato articulation

Statistical analysis of *legato* articulation was conducted on 2237 notes included in the KV 545 database. This analysis revealed that *legato* notes were played with a KOR that depended on the IOI; a larger IOI was associated with a lower KOR and a larger KOT (Table 1 in Paper III). Performances played according to the 9 adjectives gave different values of KOR, higher for *passionate* performances, lower for *flat* performances and intermediate for *natural* performances (Figure 4 in Paper III). These results confirm observations by the pianist Bruno Canino that *legato* articulation is not merely a technicality but must also correspond to an expressive intention (Canino 1997).

A separate analysis conducted only on sixteen-notes revealed that notes in descending melodic patterns are played more *legato* than notes in ascending patterns (Figures 8 and 9 in Paper III). The measured values of KOT are in accordance with data from previous research by Repp (1997), and MacKenzie and Van Eerd (1990). Figure 1 compares their results with the results for *natural* performances from Paper III. The dependence of the IOI is similar in these three investigations, although the magnitude of the KOT was greater in Repp's study and lower in MacKenzie and Van Eerd's study. These differences would be related to the examples and playing styles studied in the

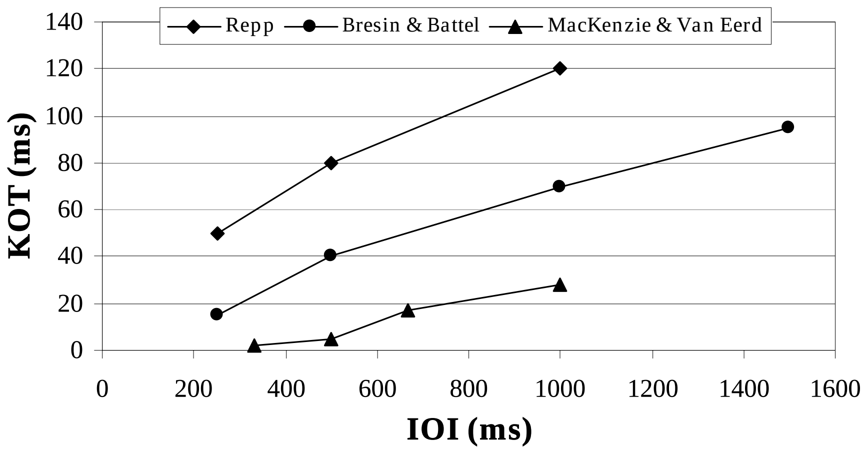

Figure 1. KOT vs IOI reported by Repp (1997), Bresin and Battel (Paper III), and MacKenzie and Van Eerd (1990).

three investigations. Thus, the pianists in Repp's investigation played five-note ascending and descending scales and *arpeggi*, those in the MacKenzie and Van Eerd's study played ascending and descending two-octave C-major scales, while our data refer to a more complex composition that was performed according to the *natural* condition. Also, as mentioned above, the magnitude of the KOT was affected by the performance style.

Staccato articulation

Statistical analyses were performed on both databases. Two main results emerged from the analysis of the 548 notes selected from the KV 545 database (Table 2 in Paper III). First, *staccato* was realized by means of a KDR that was independent of IOI. For *natural* performances it amounted to approximately 60% of the IOI. Second, KDR varied with the performance style, higher (in the range of *staccatissimo)* for *bright* and *light* performances, and lower (in the range of a *mezzostaccato)* for *heavy* performances. These measurements confirmed empirical observations by Carl Philippe Emanuel Bach who, in his "Versuch über die wahre Art das Clavier zu spielen" (1753), wrote that *staccato* notes should be rendered with a duration less than 50% of their nominal duration.

The independence of KDR from IOI was confirmed from the statistical analysis conducted on the second database, i.e. the performances of thirteen Mozart's piano sonatas (Paper IV). Isolated *staccato* notes were performed with highest KDR, and notes in *staccato* sequences were performed with lowest KDR (Figure 2 in Paper IV). It was also found that KDR varied from 61% for *Adagio tempi* to 80% for *Menuetto tempi* (Figure 7 in Paper IV). Pitch contour had also a significant effect on KDR; repeated *staccato* notes were performed with higher KDR than notes in uphill and downhill pitch contours. Moreover, in uphill patterns KDR was higher than in downhill patterns, thus implying longer duration of *staccato* notes in downhill (Figure 6 in Paper IV). This is analogous to what found in Paper III about higher KOR for *legato* notes in downhill patterns (Figure 8 in Paper III).

Articulation of repetition

In the KV 545 database, there were only 2 cases of note repetition. All performances of these notes were selected for analysis. Repeated notes were played in average with a KDR of about 40% of the IOI, well below the *staccato* range. An important result from the statistical analysis was that, unlike notes played *staccato*, the KDR in repeated notes varied with IOI (Table 3 in Paper III). In *heavy* and *natural* performances the average KDT was almost constant across IOIs, with shorter KDT for *heavy* performances. A possible explanation for this is that note repetition is a matter of technicality rather than an expressive means; pianists have to lift their fingers in order to press the same key two times.

Rules for articulation in automatic piano performance

Since the publication of Paper III, new research has been carried out regarding articulation in piano performance (Bresin forthcoming). As the results are relevant to the conclusions inferred from Papers III and IV, a brief summary of this new research will be given here.

A new set of rules for automatic articulation in expressive piano music is presented below. These rules are based on results from statistical analyses conducted on measurements of the performances stored in the two databases mentioned above. The effect of these rules has not yet been tested with a listening test. However, the analysis-by-synthesis method confirmed their importance to the improvement of the quality of performance. The new rules are included in the DM rule system, and therefore all rules are written in Common Lisp language. For coherence with the DM documentation, the term DRO (offset-to-onset duration, also referred to as “off-time duration”) will be used instead of KOT and KDT; a positive DRO corresponds to KDT, and a negative DRO corresponds to KOT.

Score legato articulation rule

DM Lisp function name: Score-legato-art.

Description: this rule produces an overlap of tones, or *legato*. The pseudocode of the rule is presented below.

Affected sound parameter: offset-to-onset duration, `DRO`.

Usage and limitations: the rule can be used to control the quantity of *legato* articulation. It is applied to notes which are marked *legato* in the score, as suggested by the name of the rule. Groups of *legato* notes are marked in the score with the Lisp commands `(LEGATO-START T)` and `(LEGATO-END T)`.

Pseudocode for the *Score Legato Articulation* rule:

if $1 < K <= 5$
  then $DRO \leftarrow (IOI \cdot (0.5 \cdot 10^{-6} \cdot K - 0.11 \cdot 10^{-3}) + 0.01105 \cdot K + 0.16063) \cdot IOI$
  else if $0 < K <= 1$
    then $DRO \leftarrow (IOI \cdot (-4.3 \cdot 10^{-6} \cdot K - 6.6 \cdot 10^{-6}) + 58.533 \cdot 10^{-3} \cdot K + 113.15 \cdot 10^{-3}) \cdot IOI$

where `K` is a weighting parameter determining the magnitude of `DRO` (or KOT).

The `K` values can be associated with the different playing styles corresponding to the adjectives used for the experiment in Paper III:

$K = 5$ ⇒ passionate *legato*
$K = 1$ ⇒ natural *legato*
$K = 0.1$ ⇒ flat *legato*

Score staccato articulation rule

DM Lisp function name: Score-staccato-art.

Description: this rule introduces a micropause after a *staccato* tone. The pseudocode of the rule is presented below.

Affected sound parameter: offset-to-onset duration, *DRO*.

Usage and limitations: the rule can be used to control the quantity of *staccato* articulation. It is applied to notes marked *staccato* in the score, as suggested by the name of the rule. *Staccato* is marked in the score with the Lisp command `(STACCATO T)`. An extra parameter, *Tempo-indication*, can be used to achieve different quantities of *staccato* for different *tempo* indications. The DM command line for the *Score Staccato Articulation* rule is therefore:

```
Score-staccato-art <K> :Tempo-indication <tempo>
```

Pseudocode for the *Score Staccato Articulation* rule:

```
if 1 < K <= 5
   then DRO ← (0.0216·K + 0.643)·IOI
   else if 0 < K <= 1
      DRO ← (0.458·K + 0.207)·IOI
DRO ← pitch-contour · context · Tempo-indication · DRO
```

where K is a weighting parameter determining the magnitude of *DRO* (or KDT), *pitch-contour*, *context* and *Tempo-indication* are three variables realizing the effects due to pitch contour, *staccato* context and *tempo* indication, as presented in Figures 4, 5, 6 and 7 in Paper IV.

The K values are associated with the different playing styles given below corresponding to the adjectives used for the experiment in Paper III:

$K = 5$ ⇒ default *staccatissimo*
$K = 3$ ⇒ light
$K = 1$ ⇒ natural
$K = 0.6$ ⇒ default *staccato*
$K = 0.5$ ⇒ heavy
$K = 0.1$ ⇒ default *mezzostaccato*

Default value for both variables *pitch-contour* and *context* is 1. Their values can be modified by DM, according to the results discussed in Paper IV.

The *Tempo-indication* values are associated with the different *tempi* given below corresponding to those observed in the measurements presented in Paper IV:

Tempo-indication = 1.3 ⇒ *Presto* and *Menuetto*
Tempo-indication = 1.15 ⇒ *Allegro*
Tempo-indication = 1 ⇒ *Adagio* and *Andante*

Articulation of repetition rule

DM Lisp function name: Repetition-art.

Description: the rule inserts a micropause between two consecutive tones with same pitch. The pseudocode of the rule is presented below.

Affected sound parameter: offset-to-onset duration, *DRO*.

Usage and limitations: the rule inserts a micropause between two consecutive tones with the same pitch. An expressive parameter *Expr* can be used to achieve two different kinds of articulation, one with constant *DRO*, the other with *DRO* dependent on IOI. The DM command line for the *Articulation of Repetition* rule is therefore:

```
Repetition-art <K> :Expr constant-dro
```

Pseudocode for the *Repetition of Articulation* rule:

```
If Expr = constant-dro
   then DRO ← 20·K;
   else if Expr = varying-dro
      if K > 1
         then DRO ← (K·(-46·10^-6·IOI – 23.67·10^-3) – 878·10^-6·IOI + 0.98164)·IOI
         else if K <= 1
            then DRO ← (K·(- 532·10^-6·IOI + 0.3592) – 248·10^-6·IOI + 0.3578)·IOI
```

where K is a weighting parameter determining the magnitude of the rule effect.

The *Expr* and K values are associated with the different playing styles given below corresponding to the adjectives used for the experiment in Paper III:

if *Expr* = *constant-dro* **then**:
$K = 1$ ⇒ natural
$K = 0.7$ ⇒ heavy

if *Expr* = *varying-dro* **then**:
$K = 5$ ⇒ dark
$K = 4$ ⇒ soft
$K = 2$ ⇒ passionate
$K = 1$ ⇒ bright
$K = 0.5$ ⇒ flat and light
$K = 0.1$ ⇒ hard

Duration contrast articulation rule

The first version of this rule was presented in Bresin and Friberg (1998). The current, slightly modified version is described here.

DM Lisp function name: Duration-contrast-art.

Description: the rule inserts a micropause between two consecutive tones if the first note is a short one, i.e., if it has duration between 30 and 600 milliseconds, see Table 1.

Affected sound parameter: offset-to-onset duration, *DRO*.

Usage and limitations: this rule can be used for the purpose of articulation, as suggested by its name. It can also be inverted, in the sense that it produces overlap of tones, or *legato*. Thus, the rule can be used to control the type of articulation, ranging from *staccato* to *legato*. It applies to notes which are

Table 1. Relation between tone duration (DR, in ms) and offset-to-onset duration (DRO, in ms) according to the rule Duration Contrast Articulation.

DR	**< 30**	**200**	**400**	**> 600**
DRO	0	-16.5	-10.5	0

marked neither *legato* nor *staccato* in the score, as such notes are processed by the *Score Legato Articulation* and the *Score Staccato Articulation* rules. The rule is not applied to the first tone of tone repetitions.

Analogies with step movements

Paper III presents some analogies between gait patterns during walking and running and how *legato* and *staccato* are achieved in piano performance. Some further comments and figures will be presented that support these analogies.

When walking, a double support phase is created when both feet are on the ground at the same time, thus there is a step overlap time; this phenomenon is similar to *legato* articulation. Figure 2 plots the KOT and the double support phase duration (T_{dsu}) as a function of the IOI and of half of the stride cycle duration ($T_c/2$), respectively. The great inter-subject variation in both walking and *legato* playing, along with biomechanical differences, made quantitative matching impossible. Nevertheless, the tendency to overlap is clearly common to piano playing and walking. Also common is the increase of the overlap with increasing IOI and increasing ($T_c/2$), respectively.

Both jumping and running contain a flight phase, during which neither foot has contact with the ground. This is somewhat similar to *staccato* articulation. In Figure 3 the flight time (T_{air}), and KDT are plotted as a function of half of stride cycle duration (Tc/2) and of IOI. The plots for T_{air} correspond to typical step frequency in running. The plots for KDT represent *mezzostaccato* (KDR = 25%) as defined by Kennedy (1996) and *staccato* performed with different expressive intentions as reported by Bresin and Battel (forthcoming). The similarities suggest that it would be worthwhile to explore the perception of *legato* and *staccato* in formal listening experiments.

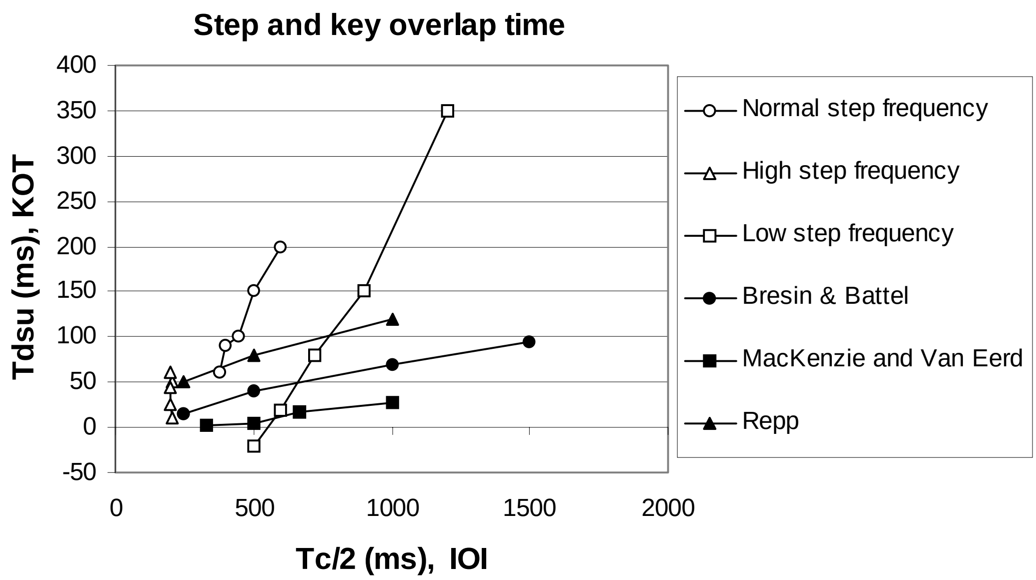

Figure 2. The double support phase (T_{dsu}, filled symbols) and the key overlap time (KOT, open symbols) plotted as function of half of stride cycle duration (Tc/2) and of IOI. The plots for T_{dsu} correspond to walking at step frequency as reported by Nilsson and Thorstensson (1987, 1989). The KOT curves are the same as in Figure 1, reproducing data reported by Repp (1997), Bresin and Battel (forthcoming), MacKenzie and Van Eerd (1990).

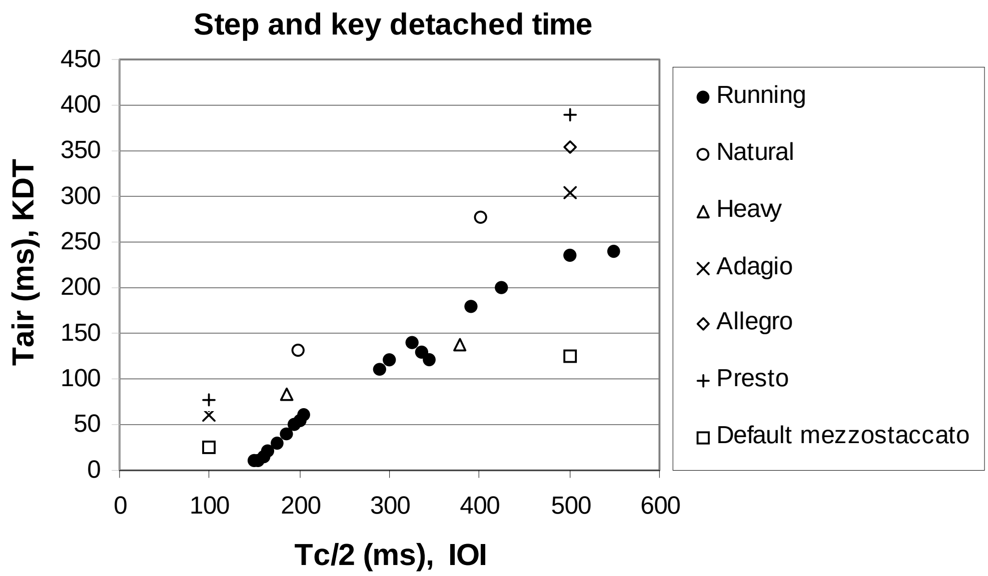

Figure 3. The time when both feet are in the air (T_{air}, filled symbols) and the key detached time (KDT, open symbols) plotted as function of half of stride cycle duration (Tc/2) and of IOI. The plots for T_{air} correspond to normal frequency steps in running (Nilsson and Thorstensson 1987, 1989). The KDT for mezzostaccato (KDR = 25%) is defined in the Oxford Concise Dictionary of Music (Kennedy 1996). The values for the other KDTs are reported in works by Bresin and Battel (forthcoming) and Bresin and Widmer (2000).

Paper V. Emotional coloring of music performance

Recent research in music interpretation has seen a flourishing of new studies on the importance of the emotional component in performance rendering. Alf Gabrielsson and his group in Uppsala have been particularly active in this area (e.g. Gabrielsson 1994, 1995; Gabrielsson and Juslin 1996). They have focused mainly on four of the so-called *basic emotions* (anger, sadness, happiness and fear), sometimes complemented with solemnity and tenderness. They isolated qualitative descriptions of acoustic cues that were important both in the communication and in the perception of the player's expressive intentions.

These cues were used in Paper V for the design of six DM macro-rules, one for each emotion (Table 1 in Paper V). Each macro-rule consisted of a selection of DM rules that were appropriate for the rendering of a specific emotion. Each macro-rule produced performances with a particular emotional coloring.

Performances of two contrasting pieces were produced with all six macro-rules. One piece was the melody line of a Swedish nursery rhyme ("Ekorrn satt i granen", henceforth *Ekorrn*, "The squirrel sat on the fir-tree", composed by Alice Tegnér), written in major tonality (Figure 3 in Paper V). The other was a computer generated piece, by Cope (1992), in minor tonality (henceforth *Mazurka*). This piece was written in an attempt to portray the musical style of Fréderic Chopin (Figure 4 in Paper V). A grand piano sound (Kurzweil sound samples of the Pinnacle Turtle Beach soundboard) was used for the synthesis.

The resulting deviations for the *angry* version of *Ekorrn*, are described in Paper V and shown in Figure 2 in the same paper. It can bee seen that the observations by Gabrielsson and Juslin on the involved cues were quantitatively reproduced. Interestingly, the *Duration Contrast Articulation* rule, described above, introduced small articulation pauses after all comparatively short notes. Thus, this rule produced an equivalent of the "mostly non-*legato* articulation" observed by Gabrielsson and Juslin (Table 1 in Paper V). Relative deviations of IOI and the variation of DRO and sound level for all six performances of each piece are shown in Figures A1 and A2 in Paper V.

These performances, together with their deadpan versions (referred to as *no-expression* in Paper V and in the following), were used in a forced-choice listening test to assess the efficiency of the macro-rules. Twenty subjects of seven different nationalities were asked to classify the performances according to their elicited emotion. The main result from the listening test was that the emotions associated with the DM macro-rules were correctly classified in most cases (Figure 7 in Paper V). The statistical analysis gave a number of interesting results. First, listeners showed an overall tendency to perceive some emotions (*anger*, *sadness* and *tenderness*) more frequently than other emotions. Second, the listeners classified the performances of *Mazurka* as mainly *angry* and *sad*, while the performances of *Ekorrn* were perceived as more *happy* and *tender*. These observations confirm the well-known association of happiness with major mode and sadness with minor mode (Figure 6 in Paper V). Third, there was also a significant influence of the score on the perception of the intended emotion. Thus, the listeners classified the different performances of *Ekorrn* as intended in all cases, and of *Mazurka* in all cases but one, the *tender* version being classified as *sad*. Fourth it was easier to recognize *angry* and *happy* performances of both *Ekorrn* and *Mazurka* (Tables 3a and 3b in Paper V). Finally, both confusion matrixes of Tables 3a and 3b in Paper V show a high degree of symmetry along the main diagonal, thus demonstrating consistency in the listeners' responses.

A principal component analysis of the 17 parameters involved in the macro-rules reduced the number of dimensions of this space to 2 principal factors that explained 61% (Factor 1) and 29% (Factor 2) of the total variance (Figure 8 in Paper V). Factor 1 was closely related to variation of sound pressure level and tempo. Factor 2 was closely related to the articulation and phrasing variables. The principal component analysis revealed an interesting distribution of the six macro-rules in the 2-dimensional space; *tenderness* and *sadness* were placed almost symmetrically to *happiness* and *solemnity*, and *fear* symmetrically to *anger*. This distribution is similar to those presented in previous works on expressive music performance and obtained with different methods (De Poli, Rodà and Vidolin 1998a; Orio and Canazza 1998; Canazza, De Poli, Di Sanzo and Vidolin 1998).

Another interesting result emerging from the principal component analysis was the behavior of the *duration contrast* rule. This rule set-up changes clockwise from the fourth quadrant to the first. Note durations in *tender* and *sad* performances received a very slight contrast (shorter notes were played longer). The contrast was stronger in the *angry* and *happy* performances, and strongest in the *fear* versions (Figure 8 in Paper V).

Analysis of the acoustic cues of the six synthesized emotional performances of each piece show that *angry* and *happy* performances were thus played quicker and louder while *tender, afraid,* and *sad* performances were performed slower and softer relative to a *no-expression* rendering. The *fear* and *sadness* versions have larger standard deviations obtained mainly by exaggerating the duration contrast but also by applying the phrasing rules (Figure 9 in Paper V). An interesting outcome is the absence of performances that are at the same time quicker and softer than a *no-expression* one.

Variations of sound level (SPL) and IOI in the emotionally colored fourteen performances of the two scores were qualitatively similar to those observed in studies of expressiveness in singing and speech (Figure 10 in Paper V). These similarities confirm that it is possible to use similar strategies to express emotions in instrumental music, singing, and speech. An interesting project for the future would therefore to apply DM macro-rules in contexts other than instrumental music.

The results further demonstrate the previously unexplored possibility of rendering emotionally different performances by means of the DM system. The results show that in music performance emotional coloring corresponds to an enhancement of the musical structure; except for mean tempo and loudness, all DM rules are triggered by the structure as represented by the score. It is tempting to draw a parallel with hyper- and hypoarticulation in speech; quality and quantity of vowels and consonants vary with the speaker's emotional state or the intended emotional communication (Lindblom 1990). Yet, the structure of phrases and the meaning of the speech remain unchanged.

The results of Paper V corroborate the observation of Gabrielsson and Juslin (1996), that articulation is relevant to the emotional coloring of a performance. This was observed already by Carl Philippe Emanuel Bach (1753) who wrote "...activity is expressed in general with *staccato* in Allegro and tenderness with *portato* and *legato* in Adagio…". More recent versions of the DM macro-rules for emotional coloring of performances include the new rules for articulation presented above. A new macro-rule (Table 2) has been formulated for *sad,* including the new articulation rules. The effect of this macro-rule is illustrated in the version of Carl Michael Bellman's song *Letter 48* presented in Figure 4. This example in available on the Internet (see the Appendix I in this dissertation for the address).

Paper VI. Applications

There are a number of different possible applications for the research presented in this dissertation. In this section those presented in Paper VI are summarized and some other applications that might be particularly relevant in the near future are described.

Automatic music performance on the Internet

An interesting application of the performance ANNs and of the DM rules is automatic performance of music stored in the already existing large databases on the Internet. These databases contain music that is often stored in a deadpan version.

Some software tools for automatic music performance have already been developed (Bresin and Friberg 1997; Paper VI; Software I). In some of these tools the Java programming language was chosen for several reasons. First, it facilitates the programming of the tools for performing music over the Internet. Also it meets demands on software portability and maintenance. A Java applet can be executed in an Internet browser thus allowing easy

Table 2. DM macro-rule description for the sad performance of Carl Michael Bellman's song Letter 48.

Expressive Cue	**Gabrielsson & Juslin**	**Director Musices**
Tempo	Slow	*Tone Duration* is shortened by 15%
SPL	Moderate or Low	*Sound Level* is increased by 8 dB
Articulation	Legato	Score *Legato* Articulation rule (k = 2.7)
Time deviations & SPL deviations	▪ Moderate ▪ Soft duration contrast ▪ Relative large deviations in timing	▪ Duration Contrast rule (k = -2, amp = 0) ▪ Punctuation rule (k = 2.1) ▪ Phrase Arch rule applied to three phrase levels (level 1, k = 2.7; level 2, k = 1.5; level 3, k = 1.5) ▪ High Loud rule (k = 1)
Final ritardando	Yes	Obtained from the Phrase Arch rule

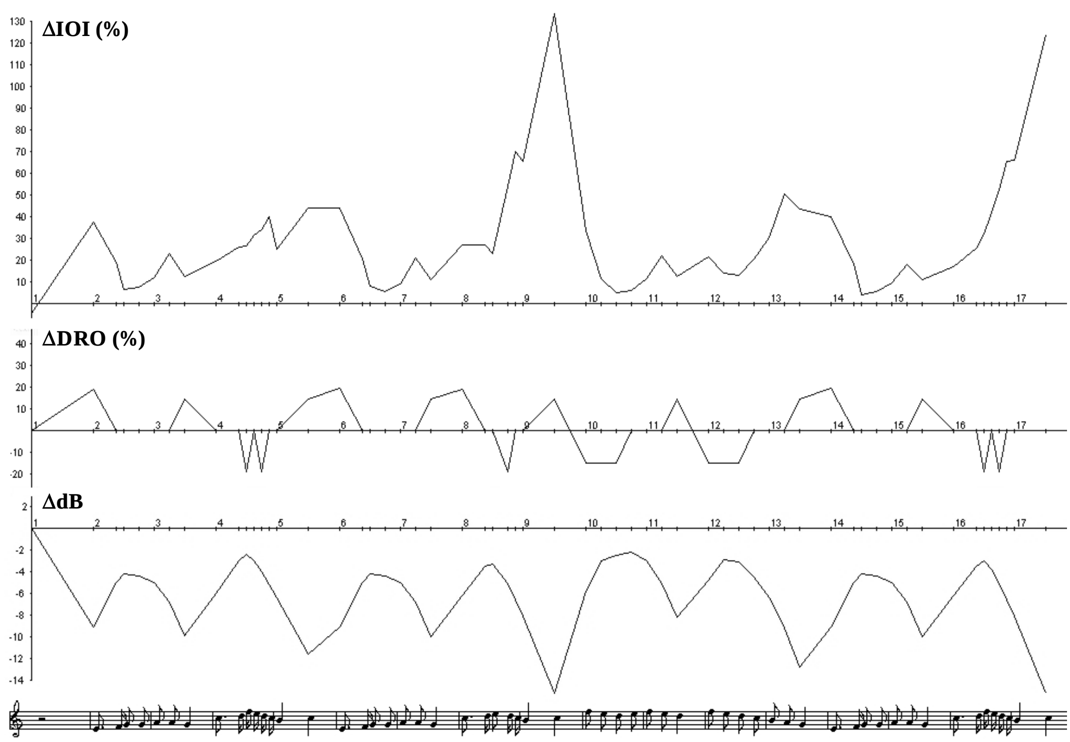

Figure 4. Inter-onset-interval (IOI in %), offset-to-onset duration (DRO in %) and sound level (dB) deviations in the sad version of Carl Michael Bellman's song Letter 48.

interaction with existing music databases and other Internet-based services. In addition, Java code has the advantage of small dimensions and short download time.

Three pieces of software have been developed so far. JAPER, Java performer, is a Java applet implementing a sub-set of the rules included in DM, see Figure 2 in Paper VI. PANN, based on the same structure as JAPER, is an applet implementing punctuation ANNs, as mentioned above (Paper II). JALISPER, a Java-Lisp performance system, is made by a Java client implementing only the user interface, while the performance rules are implemented in a special version of the DM program, written in Lisp, and running as a server. The Midi interface for both JAPER and PANN was developed using the MidiShare operating system (Orlarey and Lequay 1989; Fober 1994; Orlarey 1994). These applets have been successfully tested on both Macintosh and PC platforms, Software I (see Appendix I in this dissertation for the Internet addresses of JAPER and PANN).

Automatic analysis of the emotional color of music performance

Results from the study on the emotional coloring in automatic music performance presented in Paper V could be used for the realization of a system that analyses the emotional content in music performance. The principal component analysis on the parameters used in the DM macro-rules and the acoustic analysis of the effects produced by them (presented above and in Paper V) give an unique correspondence between macro-rules and their effects. Thus, an acoustic analysis of a performance could place it in the two-dimensional space defined by deviations of IOI and of SL (Figure 9 in Paper V). In this way it would be possible to backwards determine a probable DM macro-rule setup representing the performance in the space defined by Factor 1 and Factor 2, emerging from the principal component analysis presented in Paper V. Finally it would be possible to give a description of the emotional content in the performance analyzed in terms of involved DM rules and their parameters.

A system of this type could be used also in the modeling and communication of expressive and emotional content in a collaborative environment involving humans, avatars and robots.

Gestural rendering of synthesized sounds

The control of sound synthesis is a well-known problem. This is particularly true if the sounds are generated with physical modeling techniques that typically need specification of numerous control parameters. Outcomes from studies on automatic music performance can be useful to tackle this problem.

Sound models can be developed that respond to physical gestures. Performance rules could be used to develop control models. These models would produce a musically expressive variation of the control parameters in accordance with the dynamics of the gestures. The sound models, specified by physical descriptions and control models, can be integrated into artifacts that interact with each other and that are accessed by direct manipulation, for instance in virtual reality (VR) applications.

Music education

Advances in research as well as new software tools for the analysis of performance data open up a new area in the field of music education (Friberg and Battel forthcoming). It has been pointed out that relatively little time is dedicated to interpretative aspects of performance (Persson, Pratt, and Robson 1992). DM and the Java tools described above would represent a powerful resource in this connection. For example the models for automatic music performance can be applied to a given piece of music with separate control of each acoustic parameter and each performance cue. The output produced by these models can be quantitatively controlled, visualized on a screen, played back and listened to several times.

This new method of music performance analysis has a number of advantages and possibilities. For example the effects of the performance rules can be exaggerated so that anybody, regardless of musical training, can detect the difference and concentrate on a particular aspect of the performance. The student can compare his actual performance with a model performance, and similarities and diversities can be discussed with the teacher. These possibilities have been tested with promising results by UB Battel at the Venice Music Conservatory (Friberg and Battel, forthcoming).

The analytical comparison between a *natural* performance and performances with particular expressive intentions also seems to possess a potential for music pedagogy. Such a comparison would help students to focus their attention on expressively important aspects of different renderings of a piece of music (Paper III; Paper V; Battel and Fimbianti 1998; De Poli, Rodà, and Vidolin 1998).

Cellular phones

In the “back-yard” area of music performance, e.g., games, answering machines or ringing tones, music is typically performed with a deadpan style. The representation of the score is often similar to a MIDI file. Hence, there seem to be good possibilities to apply the ANN and DM models. For instance, a given melody in a game can be played in a sad or happy way, depending on a particular user or game action (Paper V).

The ringing tones in cellular phones often appear somewhat irritating. The reason is not only their crude sound quality, but also the deadpan performance. Here, better performances would significantly increase the pleasantness of the signal. In particular, enjoyable applications could be developed by using emotionally colored ringing tones; ringing signals corresponding to different emotions could be associated to different telephone groups or numbers. Thus, when a call is arriving, the corresponding ringing tone is played. The possibility to control the ringing tone of the receiver’s phone could be included in cellular phones of the next generation; when a person is calling, an emoticon[1] could be attached to number called, determining how the ringing tone is played in the receiver’s cellular phone.

---

[1] Emoticon: emotional icon. It is used to indicate emotional icons generally used in e-mail messages or in chat systems; :-) and :-(are the most used emoticons.

MPEG-7

Results from the papers presented in this work could also be useful for the development of accessory applications to be included in the new Mpeg-7 standard. For instance, MPEG-7 includes smart Karaoke applications, where the user will sing the melody of a song to retrieve it from a database. An emotional toolbox, capable of both recognizing the emotion of the singer and of translating it back in the Karaoke performance, may improve the human-computer interaction (Ghias, Logan, Chamberlain, and Smith 1995). It seems therefore appropriate to embody a performance system, based on rules or other techniques, in the next MPEG-7 standard, so as to enhance the expressive potentials in interactive systems involving music.

Conclusions

A complex ANN-based model for automatic music performance was presented. It generates real-time sound level and time deviations for each note represented in the input to the ANN. The model operates in a context of five consecutive notes. The design of the ANN was inspired by the symbolic rules implemented in the DM system. It is demonstrated that the ANN-based performance system is able to reproduce the behavior of performance rules as well as the style of a professional pianist. According to the results of formal evaluation tests with expert listeners the quality of performances generated by the ANN model was musically quite acceptable.

The ANN model was also used to produce punctuation in performances, i.e. the insertion of micropauses at structural boundaries. Analyses made by a professional musician were used both for training and for testing the punctuation ANN and for optimizing and testing an alternative, generative rule. In general the rule gave better results but the ANN model could reproduce choices made by the professional musician, which were not produced by the rule. An integration of these two models for music punctuation should be tested in future studies.

Both the DM rule-based model and the ANN-based model lacked the ability to realize *legato* and *staccato* articulation. A specific study of this type of articulation in expressive piano performance is reported. The application of the analysis-by-measurement method outlined the different strategies used by professional pianists in their realization of articulation under different expressive conditions. Articulation was found varying also in different music structure situations. In particular, it seems that in *legato* articulation the KOR decreases with increasing IOI, while in *staccato* the KDR is independent of IOI. It was also observed that repeated tones are performed *mezzostaccato*, at least in *natural* performances. These observations were integrated into a set of new symbolic rules.

Six macro-rules were presented. They are subsets of DM performance rules that appeared important to the simulation of six different emotional expressions; anger, sadness, happiness, fear, solemnity and tenderness. These macro-rules were assessed with a listening test. Participants in this test could recognize all simulated intended emotions. This demonstrated the previously unexplored possibility of DM to produce emotionally colored performances by means of different rule combinations. Since articulation is an important cue in the communication and perception of emotional expression in music performance, the new articulation rules have been integrated in the latest version of macro-rules for the emotional coloring of performances.

Results presented in this dissertation open the possibility to develop several new applications. Thus, applications are proposed in the field of automatic music performance, automatic performance analysis, music education, sound synthesis, virtual reality, cellular phone and human computer interaction. Certainly one of the most fascinating applications is the realization of a performance analyzer. Such an analyzer could be realized by applying the DM system backwards; the emerging rule set-up that could produce an observed performance would allow a deeper and quantitative perspective.

References

All references quoted in the dissertation are listed here. References labeled with (**D**) are quoted in the summary of the dissertation. The references quoted in the papers constituting this dissertation are labeled with their corresponding roman numbers (**I**, **II**, **III**, **IV**, **V**, **VI**).

Ahlbäck, S. (1997). A computer-aided method of analysis of melodic segmentation in monophonic melodies. In *Proceedings of the third triennial ESCOM conference*, Uppsala, 263-268 (**II**)

Arcos, J.L., López de Mántaras, R., & Serra, X. (1998). Saxex: A Case-Based Reasoning System for Generating Expressive Musical Performance. *Journal of New Music Research* 27(3): 194-210 (**D**)

Askenfelt, A., & Jansson, E. (1990). From touch to string vibrations. I: Timing in the grand piano action. *Journal of Acoustical Society of America*, 88(1), 52-63 (**D**, **III**)

Bach, C.P.E. (1753). *Versuch über die wahre Art das Clavier zu spielen*. Lotha Hoffmann-Ebrecht, Berlin. Reprinted in 1957 by Breitkopf and Härtel, Leipzig (**D**, **III**, **IV**, **V**)

Bach, C.P.E. (1949). *Essay on the true art of playing keyboard instruments*. Translated and edited by W. J. Mitchell, New York: W W Norton and Company Inc (original title: Versuch über die wahre Art das Klavier zu spielen, Berlin 1753). (**III**, **IV**)

Baker, M. (1989). A computational approach to musical grouping analysis. *Contemporary Music Review*, 4:311-325 (**II**)

Battel, G.U., & Bresin, R. (1994). Analysis by synthesis in piano performance: a study on the theme of Brahms' Paganini-Variationen. In *Proceedings of Stockholm Music Acoustic Conference 1993,* Stockholm: KTH, 69-73 (**I**)

Battel, G.U., Bresin, R., De Poli, G., & Vidolin, A. (1993). Automatic performance of musical scores by mean of neural nerworks: evaluation with listening tests. In *Proceedings of X Colloquim on Musical Informatics*, Milano: Associazione di Informatica Musicale Italiana, 97-101 (**I**)

Battel, G.U., & Fimbianti, R. (1998). How communicate expressive intentions in piano performance. In A. Argentini & C. Mirolo (Eds.) *Proceedings of the XII Colloqium on Musical Informatics,* Gorizia: Associazione di Informatica Musicale Italiana, 67-70 (**III**, **V**, **VI**)

Battel, G.U., & Fimbianti, R. (1999). Expressive intentions in five pianists' performance. *General Psychology – Psicologia Generale*, 3:277-296 (**III**)

Berry, W. (1989). *Musical structure and performance*. New Haven: Yale university Press (**II**)

Bezooijen, R.A.M.G. van. (1984). *The characteristics and recognizability of vocal espression of emotion*. Dordrecht: Foris. (**V**)

Bresin, R. (1993). MELODIA: a program for performance rules testing, for teaching, and for piano scores performing. In *Proceedings of X CIM*

Colloquium on Musical Informatics, Milano: Associazione di Informatica Musicale Italiana, 325-327 (**I**)

Bresin, R. (1998). Artificial Neural Networks Based Models For Automatic Performance of Musical Scores. *Journal of New Music Research,* 27(3):239-270 (**Paper I**; **D**, **II**, **VI**)

Bresin, R. (1998). JAPER and PANN: two JAVA applets for music performance. Included in the CD-ROM *MidiShare: Operating System for Musical Applications*, Lyon: National Center of Contemporary Music - GRAME, http://www.grame.fr (**Software I**; **D**, **VI**)

Bresin, R. (forthcoming). A model of articulation in expressive piano performance. (**D**)

Bresin, R., & Battel, G.U. (forthcoming). Articulation strategies in expressive piano performance. *Journal of New Music Research* (**Paper III**; **D**, **IV**, **V**)

Bresin, R., De Poli, G., & Ghetta, R. (1995a). A Fuzzy Approach to Performance Rules. In *Proceedings of XI CIM Colloquium on Musical Informatics,* Bologna: Associazione di Informatica Musicale Italiana, 163-168 (**D**, **I**)

Bresin, R., De Poli, G., & Ghetta, R. (1995b). Fuzzy Performance Rules. In A. Friberg & J. Sundberg (Eds.) *Proceedings of the KTH Symposium on "Grammars for music performance",* Stockholm: KTH, 15-36 (**D**, **I**)

Bresin, R., De Poli, G., & Vidolin, A. (1991). A connectionist approach to timing deviation control in musical performance. In *Proceedings of the Secondo Convegno Europeo di Analisi Musicale,* Trento: University of Trento, 635-638 (**I**)

Bresin, R., De Poli, G., & Vidolin, A. (1992). Symbolic and sub-symbolic rule system for real time score performance. In *Proceedings of the International Computer Music Conference 1992,* San Francisco: International Computer Music Association, 211-214 (**I**)

Bresin, R., De Poli, G., & Vidolin, A. (1994). A Neural Networks Based System for Automatic Performance of Musical Scores. In *Proceedings of Stockholm Music Acoustic Conference 1993,* Stockholm: KTH, 74-78 (**I**)

Bresin, R., & Friberg, A. (1997). A multimedia environment for interactive music performance. In *Proceedings of KANSEI - The Technology of Emotion AIMI International Workshop*, Genoa, Italy, 64-67 (**D**)

Bresin, R., & Friberg, A. (1998). Emotional expression in music performance: synthesis and decoding. *Speech Music and Hearing Quarterly Progress and Status Report*, Stockholm: KTH, 1998(4):85-94 (**D**, **III**, **V**)

Bresin, R., & A. Friberg. (2000). Rule-based emotional colouring of music performance. In I. Zannos (Ed.) *Proceedings of the International Computer Music Conference 2000,* San Francisco: International Computer Music Association, 364-367 (**V**, **VI**)

Bresin, R., & Friberg, A. (2000). Software Tools for Musical Expression. In I. Zannos (Ed.) *Proceedings of the International Computer Music Conference 2000,* San Francisco: International Computer Music Association, 499-502 (**Paper VI**; **D**)

Bresin, R., & Friberg, A. (2000). Emotional Coloring of Computer Controlled Music Performance. *Computer Music Journal*, 24(4):44-62 (**Paper V**; **III**)

Bresin, R., & Widmer, G. (2000). Production of *staccato* articulation in Mozart sonatas played on a grand piano. Preliminary results. *Speech Music and Hearing Quarterly Progress and Status Report*, Stockholm: KTH, 2000(4):1-6 (**Paper IV**; **D**)

Camurri, A., Dillon, R., & Saron, A. (2000). An Experiment on Analysis and Synthesis of Musical Expressivity. In M.C. De Amicis (Ed.) *Proceedings of the XIII Colloquio di Informatica Musicale*, L'Aquila: Associazione di Informatica Musicale Italiana, 123-126 (**D**)

Cambouropoulos, E. (1997). Musical rhythm: A formal model for determining local boundaries, accent and metre in a melodic surface. In M. Leman (Ed.) *Music, gestalt & computing studies in systematic and cognitive musicology*. Berlin: Springer-Verlag, 277-293 (**II**)

Canazza, S., De Poli, G., Di Sanzo, G., & Vidolin, A. (1998). Adding expressiveness to automatic musical performance. In A. Argentini & C. Mirolo (Eds.) *Proceedings of the XII Colloqium on Musical Informatics*, Udine: Associazione di Informatica Musicale Italiana, 71-74 (**D**, **V**)

Canazza, S., De Poli, G., Rinaldin, S., & Vidolin, A. (1997). Sonological analysis of clarinet expressivity. In M. Leman (Ed.) *Music, gestalt, and computing: studies in cognitive and systematic musicology*. Berlin, Heidelberg, New York: Springer Verlag, 431-440 (**V)**

Canino, B. (1997). *Vademecum del pianista da camera*. Seconda Edizione. Firenze-Antella: Passigli Editori, 76-81 (**D**)

Carlson, R., Granström, B., & Nord, L. (1992). Experiments with emotive speech: acted utterances and synthesized replicas. In *Proceedings ICSLP 92*, Banff, Alberta, Canada: University of Alberta, 1:671-674 (**V**)

Cecconi, F., & Parisi, D. (1990). Learning to predict the consequences of one's own actions. In R. Eckmiller et al. (Ed.) *Parallel processing in neural systems and computers*. Amsterdam: North-Holland (**I**)

Clarke, E.F. (1988). Generative principles in music performance. In J. Sloboda (Ed.) *Generative Processes in Music*. Oxford: Clarendon Press, 1-26 (**D**)

Clarke, E.F. (1999). Rhythm and timing in music. In D. Deutsch (Ed.) *The psychology of music*. San Diego: Academic Press [second edition], 473-500 (**D**, **IV**)

Clynes, M. (1983). Expressive microstructures in music, linked to living qualities. In J. Sundberg (Ed.) *Studies of music performance*. Stockholm: Royal Swedish Academy of Music, Publication No. 39, 76-181 (**I**)

Cope, D. (1992). Computer modeling of musical intelligence in experiments in musical intelligence. *Computer Music Journal*, 16 (2):69-83. (**D**, **V)**

Dannenberg, R. (1996). The CMU MIDI Toolkit, version 3. Pittsburgh: Carnegie Mellon University, http://www.cs.cmu.edu/~music/cmt/ (**I**)

Dannenberg, R., & Derenyi, I. (1998). Combining Instrument and Performance Models for high-quality Music Synthesis. *Journal of New Music Research,* 27(3):211-238 (**D**)

De Poli, G., Irone, L., & Vidolin, A. (1990). Music score interpretation using a multilevel knowledge base. *Interface (Journal of New Music Research),* 19:137-146 (**D**, **I**)

De Poli, G., Rodà, A., & Vidolin, A. (1998a). A model of dynamic profile variation, depending on expressive intention, in piano performance of classical music. In A. Argentini & C. Mirolo (Eds.) *Proceedings of the XII Colloqium on Musical Informatics,* Udine: Associazione di Informatica Musicale Italiana, 79-82 (**D**, **V**)

De Poli, G., Rodà, A., Vidolin, A. (1998b). Note-by-note Analysis of the Influence of Expressive Intentions and Musical Structure in Violin Performance. *Journal of New Music Research*, 27(3):293-321 (**III**, **VI**)

Dillon, R. (1999*). Un sistema ibrido per la composizione e l'interpretazione musicale.* Master Thesis. Department of Communication, Computer and System Sciences, University of Genoa, Italy (**D**)

Erlbaum, L. (1986). Attractor Dynamics and Parallelism in a Connectionist Sequential Machine. In *Proceedings of the 1986 Cognitive Science Conference,* 531-546 (**I**)

Fober, D. (1994). Real-time Midi data flow on Ethernet and the software architecture of Midishare. In *Proceedings of the International Computer Music Conference 1994*, ICMA: San Francisco, 447-450 (**D**)

Fraisse, P. (1982). Rhythm and tempo. In D. Deutsch (Ed.), *The Psychology of Music,* New York: Academic Press, 149-180 (**IV**)

Friberg, A. (1991). Generative Rules for Music Performance: A Formal Description of a Rule System. *Computer Music Journal*, 15(2):56- 71 (**I**, **II**, **III**, **V**, **VI**)

Friberg, A. (1995a). *A Quantitative Rule System for Musical Expression*. Doctoral dissertation, Royal Institute of Technology, Stockholm, TRITA-TMH 1995:3, ISSN 1104-5787, ISRN KTH / TMH / FR--95 / --SE (**D**, **I**, **II**, **III**, **IV**, **V**)

Friberg, A. (1995b). Matching the rule parameters of Phrase arch to performances of "Träumerei": A preliminary study. In A. Friberg & J. Sundberg (Eds.), *Proceedings of the KTH symposium on Grammars for music performance,* Stockholm: KTH, 37-44 (**II**, **V**, **VI**)

Friberg, A., & Battel, G. U. (forthcoming). Structural Communication: Timing and Dynamics. In R. Parncutt & Gary McPherson (Eds.) *Science and Psychology of Music Performance,* Oxford University Press (**D**, **V**, **VI**)

Friberg, A., Bresin, R., Frydén, L., & Sundberg, J. (1998). Musical punctuation on the microlevel: Automatic identification and performance of small melodic units. *Journal of New Music Research* 27(3):271-292 (**Paper II**; **D**, **I**, **V**, **VI**)

Friberg, A., Colombo, V., Frydén, L., & Sundberg, J. (2000). Generating Musical Performances with Director Musices. *Computer Music Journal*, 24(3):23-29 (**D**, **IV**, **V**, **VI**)

Friberg, A., Frydén, L., Bodin, L., & Sundberg J. (1991). Performance Rules for Computer-Controlled Contemporary Keyboard Music. *Computer Music Journal,* 15(2):49-55 (**I**)

Friberg, A., & Sundberg, J. (1987). How to terminate a phrase. An analysis-by-synthesis experiment on a perceptual aspect of music performance. In A. Gabrielsson (Ed.) *Action and perception in rhythm and music,* Stockholm: Royal Swedish Academy of Music, Publication No. 55:49-56 (**II**)

Friberg, A., & Sundberg, J. (1995). *Proceedings of the KTH Symposium Grammar for Music Performance*. Speech, Music and Hearing Department, Stockholm: KTH (**I**)

Friberg, A., & Sundberg, J. (1999). Does music performance allude to locomotion? A model of final *ritardandi* derived from measurements of stopping runners. *Journal of the Acoustical Society of America*, 105(3):1469-1484. (**D**, **IV**, **V**, **VI**)

Friberg, A., & Sundström, A. (1997). Preferred swing ratio in jazz as a function of tempo. *Speech Music and Hearing Quarterly Progress and Status Report,* Stockholm: KTH, 1997(4):19-28 (**VI**)

Gabrielsson, A. (1985). Interplay between analysis and synthesis in studies of music performance and music experience. *Music Perception,* 3(1):59-86 (**D**, **V**)

Gabrielsson, A. (1987). Once again: The theme of Mozart's piano sonata in A major (K. 331). A comparison of five performances. In A. Gabrielsson (Ed.) *Action and perception in rhythm and music,* Stockholm: Royal Swedish Academy of Music, Publication No. 55:81-103 (**II**)

Gabrielsson, A. (1994). Intention and emotional expression in music performance. In A. Friberg, J. Iwarsson, E. Jansson & J. Sundberg (Eds.) *Proceedings of the Stockholm Music Acoustics Conference 1993,* Stockholm: Royal Swedish Academy of Music, 108-111 (**D**, **V**)

Gabrielsson, A. (1995). Expressive intention and performance. In R. Steinberg (Ed.) *Music and the Mind Machine: the Psychophysiology and the Psychopathology of the Sense of Music.* Berlin, Heidelberg, New York: Springer Verlag, 35-47 (**D**, **V**)

Gabrielsson, A., & Juslin, P. (1996). Emotional expression in music performance: between the performer's intention and the listener's experience. *Psychology of Music,* 24:68-91 (**D**, **IV**, **V**)

Gabrielsson, A. (1999). The Performance of Music. In D. Deutsch (Ed.) *The psychology of music.* San Diego: Academic Press [second edition], 501-602 (**D)**

Gerardi, G. M., & Gerken, L. (1995). The development of affective responses to modality and melodic contour. *Music Perception*, 12(3):279-290 (**V**)

Ghias, A., Logan, J., Chamberlain, D., & Smith, B.C. (1995). Query By Humming - Musical Information Retrieval in an Audio Database. In *Proceedings of ACM Multimedia ´95,* San Francisco, New York: ACM - Association for Computing Machinery, 231-236 (**D**, **V**)

Granqvist, S. (1996). Enhancements to the Visual Analogue Scale, VAS, for listening tests. *Speech Music and Hearing Quarterly Progress and Status Report*, Stockholm: KTH, 1996(4):61-65 (**V**)

Hashimoto, S. (1997). KANSEI as the Third Target of Information Processing and Related Topics in Japan. In *Proceedings of KANSEI - The Technology of Emotion*, Genova: Associazione di Informatica Musicale Italiana, 101-104 (**I**)

House, D. (1990). On the perception of Mood in Speech: Implications for the Hearing Impaired. Lund University, Department of Linguistics. *Working Papers*, 36:99-108 (**V**)

Ishikawa, O., Aono, Y., Katayose, H. and Inokuchi, S. (2000). Extraction of musical performance rule using a modified algorithm of multiple regression analysis. In I. Zannos (Ed.) *Proceedings of the International Computer Music Conference 2000*, San Francisco: International Computer Music Association, 348-351 (**D**)

Juslin, P.N. (1997a). Emotional communication in music performance: a functionalist perspective and some data. *Music Perception*, 14(4):383-418 (**V**)

Juslin, P.N. (1997b). Can results from studies of perceived expression in musical performances be generalized across response formats? *Psychomusicology*, 16:77-101 (**V**)

Juslin, P.N. (1997c). Perceived emotional expression in synthesized performances of a short melody: capturing the listener’s judgment policy. *Musicae Scientiae*, 1(2):225-256 (**D**, **IV**, **V**)

Juslin, P.N., Friberg, A., & Bresin, R. (1999). Towards a Computational Model of Performance Expression: The GERM Model. Paper presented at the *Meeting of the Society for Music Perception and Cognition (SMPC'99)*, Evanston, USA (**VI**)

Kastner, M. P., & Crowder, R. G. (1990). Perception of the major/minor distinction: IV. Emotional connotation in young children. *Music Perception*, 8(2):189-202 (**V**)

Kato, I., Ohteru, S., Shirai, K., Narita, S., Sugano, S., Matsushima, T., Kobayashi, T., Fujisawa, E. (1987). The robot musician 'WABOT-2'. *Robotics*, 3(2):143-55 (**D**)

Kennedy, M. (1996). *Oxford Concise Dictionary of Music*. Oxford: Oxford University Press (**D**, **III**, **IV**)

Kroiss, W. (2000). *Parameteroptimierung für ein Modell des musikalischen Ausdrucks mittels genetischer Algorithmen*. Master's Thesis, Department of Medical Cybernetics and Artificial Intelligence, University of Vienna, Austria. (**IV**)

Kitahara, Y., & Tohkura, Y. (1992). Prosodic control to express emotions for man-machine interaction. *IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences*, 75:155-163 (**V**)

Kotlyar, G.M., & Morozov, V.P. (1976). Acoustical correlates of the emotional content of vocalized speech. Soviet Physics-Acoustics, 22(3):208-211 (**V**)

Langeheinecke, E.J., Schnitzler, H.U., Hirsher-Buhrmester, M., & Behne, K.E. (1999). Emotions in the singing voice: Acoustic cues for joy, fear, anger, and sadness. *Journal of the Acoustical Society of America*, 105(2), Pt. 2, 1331 (**V**)

Langner, J., & Kopiez, R. (1995). Oscillations triggered by Schumann's Träumerei: towards a new method of performance analysis based on a "Theory of oscillating systems" (TOS). In A. Friberg & J. Sundberg (Eds.) *Proceedings of the KTH Symposium on "Grammars for music performance"*, Stockholm: KTH, 45-58 (**I**)

Lerdahl, F., & Jackendoff, R. (1983). *A generative theory of tonal music*. Cambridge, MA: The MIT Press (**II**)

Lim, H., Ishii, A., & Takanishi, A. (1999). Basic emotional walking using a biped humanoid robot. In *Proceedings of IEEE International Conference on Systems, Man, and Cybernetics*, SMC '99, 4:954 –959 (**D**)

Lindblom, B. (1990). Explaining phonetic variation: a sketch of the H&H theory. In Hardcastle & Marchal (Eds.) *Speech production and speech modeling*, Dordrecht: Kluwer, 403-439 (**D**)

Lundin, M. (1992). Move I and Anima II; Movement Description and Movement Animation Programs. In A.W. Smith (Ed.) *Dance and Technology I: Moving Toward the Future. Proceedings of the First Annual Conference,* Madison: University of Wisconsin, WI. 28-1, 29—31 (**D**)

MacKenzie, C.L., & Van Erde, D.L. (1990). Rhythmic Precision in the Performance of Piano Scales: Motor Psychophysics and Motor Programming. In M. Jeannerod (Ed.) *Proceedings of the Thirteenth international Symposium on Attention and Performance*, Hillsdale: Lawrence Erlbaum Associates, Inc., Publishers, 375-408 (**III**)

Mathews, M.V. (1975). How to make a slur. *Journal of Acoustical Society of America*, Vol. 58 (S1), S132 (**D**, **III**)

Meyer, L.B. (1956). *Emotion and Meaning in Music.* Chicago and London: The University of Chicago Press (**V**)

Mozziconacci, S. (1998). *Speech Variability and Emotion: Production and Perception*. Doctoral dissertation. Eindhoven: Technische Universiteit (**V**)

Nilsson, J., & Thorstensson, A. (1987). Adaptability in frequency and amplitude of leg movements during human locomotion at different speeds. *Acta Physiologica Scandinavica*, 129, 107-114 (**D**, **III**)

Nilsson, J., & Thorstensson, A. (1989). Ground reaction forces at different speeds of human walking and running. *Acta Physiologica Scandinavica*, 136, 217-227 (**D**, **III**)

Orio, N., & Canazza, S. (1998). How are expressive deviations related to musical instruments? Analysis of tenor sax and piano performances of "How High the Moon" theme. In Argentini A & Mirolo C (Eds.) *Proceedings of the XII Colloqium on Musical Informatics*, Udine: Associazione di Informatica Musicale Italiana, 75-78 (**D**, **V**)

Orlarey, Y. (1994). Hierarchical Real Time Interapplication Communications. In *Proceedings of the International Computer Music Conference 1991*, San Francisco: International Computer Music Association, 408-415 (**D**, **VI**)

Orlarey, Y., & Lequay, H. (1989). MidiShare: a Real Time multi-task software for Midi applications. In *Proceedings of the International Computer Music Conference 1989*, San Francisco: International Computer Music Association, 234-237 (**D**, **I**, **VI**)

Palmer, C. (1989). Mapping musical thought to musical performance. *Journal of Experimental Psychology: Human Perception and Performance*, 15:331-346 (**D**, **III**)

Pandy, M.G., & Anderson, F.C. (2000). Dynamic Simulation of Human Movement Using Large-Scale Models of the Body. *Phonetica*, 57(2-4):219-228 (**D)**

Parncutt, R. (1994). A perceptual model of pulse salience and metrical accent in musical rythms. *Music Perception*, 11:407-464 (**D**)

Persson, R. S., Pratt, G., & Robson, C. (1992). Motivational and influential components of musical performance: A qualitative analysis. *European Journal for High Ability*, 3:206-217 (**D, VI**)

Purbrick, J. (2000). Automatic Synthesizer Articulation. *Computer Music Journal*, 24(1):20-31 (**D**, **III**)

Repp, B. H. (1990). Patterns of expressive timing in performances of Beethoven minuett by nineteen famous pianists. *Journal of Acoustical Society of America***,** 88(2):622-641 (**D**)

Repp, B. H. (1992). Diversity and commonality in music performance: An analysis of timing microstructure in Schumann's Träumerei. *Journal of Acoustical Society of America*, 92(5):2546-2568 (**D**, **I**, **II**, **V**)

Repp, B. (1995). Acoustics, perception, and production of legato articulation on a digital piano. *Journal of Acoustical Society of America*, 97(6):3862-3874 (**D**, **III**, **IV**)

Repp, B. (1997). Acoustics, perception, and production of legato articulation on a computer-controlled grand piano. *Journal of Acoustical Society of America*, 102(3):1878-1890 (**D**, **III**, **IV**)

Repp, B. H. (1998a). A microcosm of musical expression: I. Quantitative analysis of pianists' timing in the initial measures of Chopin's Etude in E major. *Journal of the Acoustical Society of America*, 104(2):1085-1100 (**D**, **V**)

Repp, B. (1998b). Perception and Production of Staccato articulation on the Piano. Unpublished manuscript, Haskins Laboratories, http://www.haskins.yale.edu/haskins/STAFF/repp.html (**D**, **III**, **IV**)

Rumelhart, D. E., & McClelland J. L. (1986). *Parallel Distributed Processing*, vol. 1 and 2, Cambridge: MIT Press (**I**, **II**)

Seashore, C. E. (1938). *Psychology of Music.* New York: McGraw-Hill (**D**, **I**)

Shaffer, L. (1981). Performances of Chopin, Bach and Bartok: studies in motor programming. *Cognitive Psychology* 13:326-276 (**D**)

Shaffer, L.H. (1982). Rhythm and timing in skill. *Psychological Review*, 89:109-122 (**D**)

Shaffer, L.H. (1984a). Timing in solo and duet piano performances. *Quarterly Journal of Experimental Psychology*, 36A:577-595 (**D**)

Shaffer, L.H. (1984b). Timing in musical performance. In J. Gibbon & L. Allan (Eds.), Timing and time perception [Special issue]. *Annals of the New York Academy of Sciences*, 423:420-428 (**D**)

Sloboda, J.A. (1983). The communication of musical metre in piano performance. *Quarterly Journal of Experimental Psychology*, 35A:337-396 (**D**, **III**, **IV**)

Sundberg, J. (1993). How can music be expressive? *Speech Communication*, 13:239-253 (**D**, **V**)

Sundberg, J. (2000). Emotive transforms. *Phonetica*, 57(2-4): 95-112 (**D**, **V**)

Sundberg, J., Friberg, A. and Frydén, L. (1991a). Common secrets of musicians and listeners – An analysis-by-synthesis of musical performance. In P. Howell, R. West, & I. Cross (Eds.) *Representing musical structure*. London: Academic Press (**II**)

Sundberg, J., Friberg, A., & Frydén, L. (1991b). Threshold and preference Quantities of Rules for Music Performance. *Music Perception*, 9(1):71-92 (**V**)

Sundberg, J., Friberg, A., & Frydén, L. (1989). Rules for automated performance of ensemble music. *Contemporary Music Review*, 3:89-109 (**V**)

Tenney, J., & Polansky, L. (1980). Temporal Gestalt Perception in Music. *Journal of Music Theory*, 24:205-241 (**II**)

Todd, N.P. McAngus. (1985). A model of expressive timing in tonal music. *Music Perception*, 3:33-58 (**D**, **II**)

Todd, N.P. McAngus. (1992). The dynamics of dynamics: A model of musical expression**.** *Journal of Acoustical Society of America*, 91(6):3540-3550 (**D**, **I**, **II**, **IV**)

Ungvary, T., Waters, S., & Rajka, P. (1992). Nuntius – A Computer System for the Interactive Composition and Analysis of Music and Dance. *Leonardo*, 25(1):59-68 (**D**)

Widmer, G. (1996). Learning expressive performance: The structure-level approach. *Journal of New Music Research*, 25(2):179-205 (**D**)

Widmer, G. (2000). Large-scale Induction of Expressive Performance Rules: First Qualitative Results. In I. Zannos (Ed.) *Proceedings of the International Computer Music Conference 2000*, San Francisco: International Computer Music Association, 344-347 (**D**)

Appendix I. Electronic appendixes

Links related to papers I and II

Sound examples of performances by ANNs trained with the style of a pianist as described in Paper I:

http://www.speech.kth.se/music/performance/download

Automatic punctuation with ANNs. The PANN Java applet implementing the punctuation ANNs described in Paper II:

http://www.speech.kth.se/~roberto/pann

KTH performance rules description:

http://www.speech.kth.se/music/performance/

Links related to papers III, IV and V

Sound examples of emotionally different performances used in the listening test described in paper V:

http://www.speech.kth.se/~roberto/emotion

Deadpan and sad versions of Carl Michael Bellman's song *Letter 48*. The *Score Legato Articulation* and the *Score Staccato Articulation* rules were applied in the sad version.

http://www.speech.kth.se/music/performance/germ

Test results of listeners' ability to recognize the emotional color of automatic performances; the tests were carried out at two seminars in Stockholm and Göteborg:

http://www.speech.kth.se/~roberto/emotion/lerici19990310
http://www.speech.kth.se/~roberto/emotion/gbg19990507

Links related to paper VI

The KTH performance rule system, Director Musices (Windows and Mac OS):

http://www.speech.kth.se/music/performance/download

The Melodia performance system (Windows):

http://www.speech.kth.se/music/performance/download

Automatic performance on the Internet. The JAPER Java applet:

http://www.speech.kth.se/~roberto/japer

Automatic punctuation with ANNs. The PANN Java applet:

http://www.speech.kth.se/~roberto/pann

Appendix II. Co-authors' roles

Paper II

Frydén inserted punctuation markings in the 52 melodies that were used for training and testing of both the rule system and the ANN for punctuation. Friberg developed the punctuation rule system, and trained and tested it. Sundberg and Frydén participated in the analysis-by-synthesis procedure in the preliminary assessment of the punctuation rule system. Bresin designed the ANN for music punctuation, trained and tested it. Bresin developed the PANN Java applet for automatic punctuation. Friberg wrote the first draft. Bresin wrote the parts related to the ANN-based model. Sundberg assisted in writing the paper.

Paper III

Bresin carried out the measurements of the articulation in the 45 performances. He conducted the statistical analysis on these measurements, and proposed the two parameters KOR and KDR for the analysis of *legato* and *staccato,* respectively. Battel designed the recordings of the performances with the Disklavier and chose the adjectives used as expressive intentions. Bresin wrote the first draft of the paper.

Paper IV

Widmer did the matching between the notes marked *staccato* in the scores of the 13 Mozart's piano sonatas and the corresponding performed notes, and extracted all the data necessary for the analysis of *staccato* articulation. Bresin carried out the measurements of the *staccato* articulation. He conducted the statistical analysis on these measurements, and wrote the first draft of the paper.

Paper V

Friberg collaborated with Bresin in the design of the macro-rules. Bresin designed the experiment, conducted the listening test and made the subsequent statistical analysis. Bresin made the measurements on the performances. He also wrote the first draft of the paper.

Paper VI

Bresin and Friberg wrote the paper together. Bresin implemented all the software written in Java.