Expert Systems with Applications 38 (2011) 155–160

Contents lists available at ScienceDirect

Expert Systems with Applications

journal homepage: www.elsevier.com/locate/eswa

A state of the art on computational music performance

Miguel Delgado, Waldo Fajardo, Miguel Molina-Solana *

Department of Computer Science and Artificial Intelligence, Universidad de Granada, Daniel Saucedo Aranda s/n, 18071 Granada, Spain

ARTICLE INFO

Keywords:
Computational music
Expressive performance
Machine learning

ABSTRACT

Musical expressivity can be defined as the deviation from a musical standard when a score is performed by a musician. This deviation is made in terms of intrinsic note attributes like pitch, timbre, timing and dynamics. The advances in computational power capabilities and digital sound synthesis have allowed real-time control of synthesized sounds. Expressive control becomes then an area of great interest in the sound and music computing field. Musical expressivity can be approached from different perspectives. One approach is the musicological analysis of music and the study of the different stylistic schools. This approach provides a valuable understanding about musical expressivity. Another perspective is the computational modelling of music performance by means of automatic analysis of recordings. It is known that music performance is a complex activity that involves complementary aspects from other disciplines such as psychology and acoustics. It requires creativity and eventually, some manual abilities, being a hard task even for humans. Therefore, using machines appears as a very interesting and fascinating issue. In this paper, we present an overall view of the works many researchers have done so far in the field of expressive music performance, with special attention to the computational approach.

1. Introduction

Imagine the scene. You switch on your hi-fi system, select *Clair de Lune* by Claude Debussy and Sergei Rachmaninoff as performer. After that, you hit the 'Play' button, sit on your favourite armchair and enjoy the music. The situation sounds perfectly normal … but for a small detail: Rachmaninoff never recorded Debussy's *Clair de Lune*!

To listen to a performance we need a performer. So far, this role has always been assumed by humans but, why can't the hi-fi system (more generally, a computer) be the performer and play the music as it was Rachmaninoff himself? All it needs is enough knowledge of how to play.

As Widmer, Dixon, Goebl, Pampalk, and Tobudic (2003) stated, when skilled musicians play a piece of music, they do not do it mechanically, with constant tempo or loudness, exactly as written in the printed music score. Rather, they speed up at some places, slow down at others and stress certain notes. The most important parameters available to a performer are timing (tempo variations) and dynamics (loudness variations). The way these parameters 'should be' varied during the performance is not precisely specified in the printed score. So that, it is performer's duty to use them properly.

It is a fact that student musicians spend more time practicing than almost any other activity. Weekly music lessons, endless scales, nightly rehearsals and recitals for friends and family are commonplace in their lives. Hours of practicing will help them learn to interpret a piece of music as the composer envisioned it, as well as to develop their own signature sound – one that is unique to each of them. In other words, what makes a piece of music come alive is also what distinguishes great artists from each other.

Other questions arise at this point: how should those expressive resources be employed? What is that which makes Rachmaninoff an outstanding pianist? And those simple questions, which many people have asked for many years, do not have still a clear answer from musicologists. Even when those questions will eventually find an acceptable answer, another will be posed: can a computer take advantage of that knowledge, being able to substitute a famous performer? As we will see in this paper, many attempts have been made and several computational models have been proposed during the last century to do so.

This work is organized as follows: first of all, Section 2 describes what is musical performance and its parameters, and how they can be used to distinguish between performers; Section 3 presents some works where computers were used for extracting information about those parameters from the music itself, for representing that knowledge in a computer, and for applying it to generate new

* Corresponding author.
E-mail addresses: mdelgado@ugr.es (M. Delgado), aragorn@ugr.es (W. Fajardo), miguelmolina@ugr.es (M. Molina-Solana).

0957-4174/$ - see front matter © 2010 Elsevier Ltd. All rights reserved.
doi:10.1016/j.eswa.2010.06.033

performances; Section 4 introduces some difficulties to be faced in the future when using computers in this domain; and Section 5 summarizes the work.

2. Music performance

Most people would judge the literal execution of a musical score to be significantly less interesting than a performance of that piece by even a moderately skilled musician. Why is that so? Because what we hear is not a literal rendition of the score. Of course, the principal vehicle for the communication of musical compositions is the music score in which the composer codifies his intentions. However, the information written in the score does not represent an exhaustive description of the composer's intentions. It carries information such as the rhythmical and melodic structure of a certain piece, but there is not yet a notation able to describe precisely the timing and timbre characteristics of the sound.

When speaking, we use several voice resources such as changing velocity, tone or loudness. All this effects are not explicitly in the text we are reading. In fact, when several people read a text, resulting sounds are not the same, even though words in the sheet remain unchanged. So does in music. In the same way that in a written poem there is no explicit notation for how to pronounce, in musical scores there is also such a lack of information. This comparison is actually quite appropriate because former research on music performance has revealed interesting analogies in the communication of emotions in singing and speech (Bresin & Friberg, 2000; Sundberg, 2000).

Performing is a crucial activity in music. In many kinds of music the performer acts as a kind of mediator: a mediator between composer and listener, between written score and musical sound. It is the performer who renders each note in the score in terms of intensity, duration and timbre by movements of fingers, arms or mouth. This results in different performances of the same piece reflecting each performer's culture, mood, skill and intention. These variances also contribute to determining the performing styles of different musicians. So that, the music we hear has two main sources: the score and the performance, and they both need from the other.

Briefly, Widmer and Goebl (2004) define *expressive music performance* as ‘‘the deliberate shaping of the music by the performer, in the moment of playing, by means of continuous variations of parameters such as timing, loudness or articulation”. Changes in tempo (timing) are non-linear warping of the regular grid of beats that defines time in a score. It is also possible to change only the duration of certain notes. Changes in loudness (or dynamics) are modifications of the intensity of notes with respect to the others and to the general energy of the fragment in consideration. Articulation consists in varying the gap between contiguous notes by, for instance, making the first one shorter or overlapping it with the next.

Music performance is a deep human activity which requires emotional, cognitive and artistic aptitudes. At the same time, it is also a complex task involving physical, acoustic, physiological, psychological, social and artistic aspects. Several factors determine the rendition of a musical piece. One of the most obvious is the physical condition of the performer. Not in vain, performer's mood, health and fatigue play a crucial role in the process of playing an instrument. Some studies (see those from Gabrielsson (1995) and from Rigg (1964)) have shown major variations in renditions by the same performer when he is in different moods.

Manual abilities are also an important point that is especially visible when comparing a beginner with an expert. With practice, musicians can improve their velocity and precision, reducing the amount of unintended deviations with respect to the score (commonly known as errors). Other factors that affect the rendition are the location where it takes place and the instrument being used. The acoustics of the place are important because they establish the sounds that can be made. So does the instrument, which has an evident influence on the character of the work.

Because the conventional score is quite inadequate to describe the complexity of a musical performance, and since the literal synthesis of notes from a score is flat and unappealing, there is an opportunity for learning systems that can automatically produce compelling expressive variations. Hence, methods for automatically ‘‘bringing life” to musical scores become useful and interesting. Research in this field ranges from studies aimed at understanding expressive performance to attempts at modelling aspects of performance in a formal, quantitative and predictive way, so that a computer might be able to perform them.

2.1. Functions of expressivity

Since the very first moment that some deviations exist in the way of playing a score, we can ask for the motives of their existence. Two main aims can be identified in a first sight.

In first place, expressivity is used as an instrument for communicating emotions. Meyer (1956) stated that meaning (be it emotional or aesthetic) arises in music when expectations raised by the music are not realized. It was Rigg's paper (Rigg, 1964) one of the pioneer works which tackled the relation between emotions and musical structure. Some interesting and typical regularities found throughout the years were described there: solemn music tend to be slow, low pitched, and without irregularities; happy music is fast, major mode and high pitched.

Gabrielsson (1995) and Lindström (2006) studied the relation between motional intentions and musical microstructure (for instance, tempo deviations, changes in intensity or articulations). Canazza, Poli, Drioli, Rodà, and Vidolin (2000) studied how physical parameters in musical recordings (tone, articulations or global tempo) were affected by the modification of performer's expressive intentions. In their experiments, the performer was asked to express, by her rendition of the musical score, sensorial concepts such as ‘bright’, ‘light’ or ‘dark’. The sonological analysis of the recordings made it possible to relate certain values to given concepts (e.g., a ‘light rendition’ was found to be in fast tempo, with shortened note durations and soft attacks).

The significance of various performance parameters in the identification of emotional qualities of a performance has been tested in synthesis experiments. Automatic performances were obtained by setting certain expressive cues to greater or lesser values and, in formal listening tests, listeners were able to recognize and identify the intended emotions. In the computer program developed by Canazza et al. expressiveness was applied both to a ‘neutral’ performance played by a musician with no intended emotion, and to a computer-generated ‘deadpan’ performance. Juslin (1997), on the other hand, manually adjusted the values of some previously identified cues by means of ‘‘appropriate settings on a Roland JX1 synthesizer that was MIDI-controlled” by a Synclavier III.

For more information regarding research in musical performance, including the role expressivity plays in the communication of emotions, Gabrielsson's work (Gabrielsson, 2003) might be consulted.

In second place, expressivity clarifies the musical structure, understanding within this term the metrical structure, phrasing and harmonic structure. In the work by Sloboda (1983), one could observe that performers tend to play louder and more legato the notes at the beginning of measures. It was also reported that the more expert the pianist was, the more frequent those resources were employed and the easier to transcribe the music for the audience.

Musical structure has its influence on the expressivity of performances too. It has been discovered that the beginning and the end of phrases tend be slower than the rest. For instance, Todd (1989) proposed a model to predict the final *rubato* in musical works.

Harmonic progressions in a work also have an influence on the expressivity of its renditions. In particular, Palmer (1996) demonstrated that melodic expectation —the degree in which an expected note is finally realized— was related to the energy with which notes are played.

3. Computational music performance achievements

Advances in digital sound synthesis and in computational power have enabled real-time control of synthesized sounds. Expressive control of these becomes then a relevant area of research in the *Sound and Music Computing*[1] field. Empirical research on expressive music performance has its origin in the 1930s, with the pioneering work by Seashore (1938). After a 40-years gap, the topic experienced a real renaissance in the 1970s, and music performance research is now highly productive. A comprehensive overview of this research can be found in Gabrielsson (2003).

As said before, research in musical performance has a multidisciplinar character, with studies that veer from understanding expressive behaviour to modelling aspects of renditions in a formal quantitative and predictive way. Historically, research in expressive music performance has focused on finding general principles underlying the types of expressive 'deviations' from the musical score (e.g., in terms of timing, dynamics and phrasing) that are a sign of expressive interpretation. Works by Poli (2004) and Widmer and Goebl (2004) contains recent overviews on expressive performance modelling.

Three different research strategies can be distinguished: (1) acoustic and statistical analysis of performances by real musicians –the so-called analysis-by-measurement method; (2) making use of interviews with expert musicians to help translate their expertise into performance rules –the so-called analysis-by-synthesis method; and (3) inductive machine learning techniques applied to large databases of performances.

Studies by several research teams around the world have shown that there are significant regularities that can be uncovered in these ways, and computational models of expressive performance (of mostly classical music) have proved to be capable of producing truly musical results. These achievements are currently inspiring new research into more comprehensive computational models of music performance and also ambitious application scenarios.

One of the issues in this area is the representation of the way certain performers play by just analyzing some of their renditions (i.e., study the individual style of famous musicians). That information would enable us to identify a performer by only listening to their rendition. These studies are difficult because the same professional musician can perform the same score in very different ways (compare several commercial recordings by Sergei Rachmaninoff or Vladimir Horowitz). Recently, new methods have been developed for the recognition of music performers and their style. Among them, the most relevant are the fitting of performance parameters in rule-based performance models, and the application of machine learning methods for the identification of performing style of musicians. Recent results of specialized experiments show surprising artist recognition rates (for instance, see those from Saunders, Hardoon, Shawe-Taylor, & Widmer, 2008; or Molina-Solana, Arcos, & Gomez, 2008).

So far, music performance research has been mainly concerned with describing detailed performance variations in relation to musical structure. However, there has recently been a shift towards high-level musical descriptors for characterizing and controlling music performance, especially with respect to emotional characteristics. For example, it has been shown that it is possible to generate different emotional expressions of the same score by manipulating rule parameters in systems for automatic music performance (Bresin & Friberg, 2000).

Interactive control of musical expressivity is traditionally a conductor's task. Several attempts have been made to control the tempo and dynamics of a computer-played score with some kind of gesture input device. For example, Friberg (2006) describes a method for interactively controlling, in real-time, a system of performance rules which contains models for phrasing, micro-level timing, articulation and intonation. With such systems, high-level expressive control can be achieved. Dynamically controlled music in computer games is another important future application.

Recently, some efforts have been made in the direction of visualizing expressive aspects of music performance. Langner and Goebl (2003) have developed a method for visualizing expressive performances in a tempo-loudness space: expressive deviations leave a trace on the computer screen in the same way as a worm does when it moves, producing a sort of 'fingerprint' of the performance. This method has been recently extended by Grachten, Goebl, Flossmann, and Widmer (2009). This and other recent methods of visualization can be used for the development of new multi-modal interfaces for expressive communication, in which expressivity embedded in audio is converted into visual representation, facilitating new applications in music research, music education and Human–Computer Interaction, as well as in artistic contexts. A visual display of expressive audio may also be desirable in environments where audio display is difficult or must be avoided, or in applications for hearing-impaired people.

For many years, research in Human–Computer Interaction in general and in sound and music computing in particular was dedicated to the investigation of mainly 'rational' abstract aspects. In the last ten years, however, a great number of studies have emerged which focus on emotional processes and social interaction in situated or ecological environments. The broad concept of 'expressive gesture', including music, human movement and visual (e.g., computer animated) gesture, is the object of much contemporary research.

3.1. Data acquisition

In this interdisciplinary research field, the obtention of information on musical expressivity can be approached from different perspectives. One approach is the musicological analysis of music and the study of the different stylistic schools. This approach provides a valuable understanding about musical expressivity.

Another perspective is the computational modelling of music performance by means of automatic analysis of recordings. This sound analysis perspective can be raised by the (studio specific) recording of several performers where several expressive resources are emphasized. That information can be gathered by using augmented instruments (i.e., instruments provided with sensors of pressure or movement). Proceeding this way, the data on obtains is very precise, but it is necessary a complex setup and those special instrument are anything but cheap. Furthermore, getting the performers is a difficult task and many times even impossible (e.g. dead performers).

An alternative approach is to directly use commercial recordings for the analysis of expressivity, extracting all the relevant data from the audio signals themselves. This approach has several advantages: there are tons of recordings available (and often some performers have several ones); and the performances are 'real' and gather the decisions taken by the performers without any external

[1] http://smcnetwork.org.

influence. Nevertheless, working with commercial recordings has some important drawbacks too: some information (consider, for instance, the bow speed in a violin) cannot be easily gained from the audio; these recordings do not come from a controlled scenario and the sound analysis may become more difficult.

Computers are important in both approaches, because they allow us to store and to process all the gathered data. This information is huge in size and it is impossible to deal with it in a manual way.

3.2. Computational models for artistic music performance

The use of computational music performance models in artistic contexts (e.g., interactive performances) raises a number of issues that have so far only partially been faced. The concept of a creative activity being predictable and the notion of a direct 'quasi-causal' relation between the musical score and a performance are both problematic. The unpredictable intentionality of the artist and the expectations and reactions of listeners are neglected in current music performance models. Surprise and unpredictability are crucial aspects in an active experience such as a live performance. Models considering such aspects should take account of variables such as performance context, artistic intentions, personal experiences and listeners' expectations.

In the past, this problem has been tackled by using machine learning techniques. For instance, Juslin, Friberg, and Bresin (2002) described the main sources of expressivity in musical renditions and expressed the necessity of integrating some of this aspects in a common model they started to sketch.

Ramírez, Maestre, Pertusa, Gómez, and Serra (2007) proposed a model for identifying saxophonists from the way of playing by using very precise information about deviations in parameters such as pitch, duration and loudness. They measure those deviations both in inter and intra note level.

De Mántaras and Arcos (2002) studied the expressivity of several AI-based systems for music composition. They compared this expressivity with the one that exists in human recordings. Moreover, they introduced *SAXEX*, a system capable of generating expressive performances of jazz ballads by using examples from human performers and a case-based reasoner.

Hong, on the other hand, studied how musical expressivity is affected by tempo and dynamics variations (Hong, 2003). He employed cello recordings for the experiments. He extended previous work by Todd (1992), by applying new musical ideas from the 20th century to Todd's model.

Dovey (1995) proposed an attempt to use inductive logic in order to determine the rules that pianist Sergei Rachmaninoff may have used in their performances with an augmented piano. The aim was to extract general rules (in the form of universal predicates) about each note's duration, tempo and pressure. All that information was obtained from the way of playing the piano.

The group led by Gerhard Widmer has worked in the automatic identification of pianists. In Widmer et al. (2003), they studied how to measure several aspects of performances by applying machine learning techniques; whereas in another work (Stamatatos & Widmer, 2005), they proposed a set of simple features that could serve to represent performer's expressivity from a rendered musical work.

Moreover, in a recent paper, Saunders et al. (2008) represent musical performances as string of symbols from an alphabet. Those symbols contain information about changes in timing and energy within the song. After that, they use *Support Vector Machines* to identify the performer in new recordings.

Sapp's work is also an interesting proposal, as it represents musical renditions by means of sketches which are based on the correlation between time and energy (Sapp, 2007).

Most of the modelling attempts in performance research, try to capture common performance principles, that is, they focus on commonalities between performances and performers. However, the ultimate goal of this kind of research and of many of the works is not the automatic style replication or the creation of artificial performers, but to use computers to teach us more about the elusive artistic activity of expressive music performance. While it is satisfying to see that the computer is indeed capable of extracting information from performance measurements that seems to capture aspects of individual style, this can only be a first step. In order to get real insight, we will need learning algorithms that, unlike nearest-neighbour methods, produce interpretable models.

Although it may sound odd, there are concrete attempts at elaborating computational models of expressive performance to a level of complexity where they are able to compete with human performers. The Rendering Contest (Rencon)[2] (Hiraga, Bresin, Hirata, & Katayose, 2004) is an annual event first launched in 2002. It tries to bring together scientist from all over the world for a competition of artificially created performances. It uses an human judge to evaluate music performances automatically generated by computers. Participants are asked to generate a rendition of a musical work by using a predictive level. In a wider sense, we can somehow see this paradigm as an expressive performance *Turing test*.[3] In other words, the best systems are those than manage to generate performances which sounds indistinguishable from human ones.

As can be seen, music performance is an interesting research topic which enables the study of human's emotions, intelligence and creativity. These are precisely the issues Marvin Minsky referred to when he wrote about music as a human activity (Minsky, 1992).

3.3. Automatic music performance

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.

Many authors have proposed models of automatic music performance. Todd (1992) presented a model of musical expression based on an analysis-by-measurement method. Rule-based

[2] http://www.renconmusic.org.

[3] The Turing test is a proposal for a test of a machine's ability to demonstrate intelligence. Described by Alan Turing in the 1950 paper "Computing Machinery and Intelligence", it proceeds as follows: a human judge engages in a natural language conversation with one human and one machine, each of which try to appear human. All participants are placed in isolated locations. If the judge cannot reliably tell who the machine and the human are, the machine is said to have passed the test.

systems have been proposed by Zanon and Poli (2003), Friberg (1991) and Friberg, Colombo, Frydén, and Sundberg (2000).

Performance systems based on artificial intelligence techniques have been developed too. Widmer (2003) proposed a machine learning based system extracting rules from performances. Ishikawa, Aono, Katayose, and Inokuchi (2000) 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. Arcos, de Mántaras, and Serra (1998) developed a case-based reasoning system for the synthesis of expressive musical performances of sampled instruments. Delgado, Fajardo, and Molina-Solana (2009) developed a multi-agent approach to music composition and generation.

4. Future challenges

Since the literal synthesis of notes from a score is bland and unappealing, there is an opportunity for learning systems that can automatically produce compelling expressive variations. The problem of synthesizing expressive performance is as exciting as challenging. Music performance is one of the many activities that trained people do very well without knowing exactly how they do it. This is, precisely, one of the main problems to be faced because there is no model that accurately tells us how to perform.

When referring to artistic domains, it is hardly possible to find a 'correct' model whose predictions always correspond with what humans do and what they think is acceptable. We cannot forget that evaluation in these domains is often subjective and heavily-dependent on who is speaking.

Many aspects are involved within expressive performance and it is almost impossible to use them all. Moreover, there are some parameters and dimensions which are commonly considered as non-relevant but that, in fact, might be. Only a portion of the whole problem is tackled by current techniques. One future challenge is to address the problem by using as much dimensions as possible. It could also be possible that some important patterns are hidden and we haven't still discovered them.

Moreover, to obtain very precise data about all those parameters is a challenging problem that cannot still be done in a automatic way. Annotating all this information is a very time-consuming task and requires a lot of effort from several humans. Early systematic investigations in the field have dealt with this problem either by reducing the length of the music (to just some seconds) or by controlling the size of the collections.

Recent approaches try to avoid this task by the use of some statistical learning techniques and by focusing in a more abstract representation of the real notes and their values. Statistical musicology has not historically received much attention, but it is increasing its popularity as problems are getting more and more complex, and the amount of available data grows, even though collect large amount of quantitative data is a really hard task. Temperley (2007) tackles musical perception from a probabilistic perspective in his recent book *Music and Probability*. Apart of proposing a Bayesian network model, the author carries out an interesting survey of works that use statistical tools to solve problems in the *Sound and Music Computing* area.

Despite some successes in computational performance modelling, current models are extremely limited and simplistic regarding the complex phenomenon of musical expression. It remains an intellectual and scientific challenge to probe the limits of formal modelling and rational characterization. Clearly, it is strictly impossible to arrive at complete predictive models of such complex human phenomena. Nevertheless, work towards this goal can advance our understanding and appreciation of the complexity of artistic behaviours. Understanding music performance will require a combination of approaches and disciplines, such as musicology, AI and machine learning, psychology and cognitive science.

For cognitive neuroscience, discovering the mechanisms which govern the understanding of music performance is a first-class problem. Different brain areas are involved in the recognition of different performance features. Knowledge of these can be an important aid to formal modelling and rational characterization of higher order processing, such as the perceptual differentiation between human-like and mechanical performances. Since music making and appreciation is found in all cultures, the results could be extended to the formalization of more general cognitive principles.

Finally but not least, it is the problem of the individuality of each work. Even though there is a huge amount of available data, every song is different from the rest. Hence, it would not be adequate just to apply the way of playing Beethoven's Ninth Symphony to Brahms' Symphonies. A deep study of the work is needed in order to understand the author, the context and the music. One should always keep in mind that artistic performance is far from being predictable.

5. Conclusions

At this point, the question in the beginning of the paper strikes again: can the computer play like a human? This work has tried to offer a comprehensive overview of the current research that is going on in the field of computational expressive music performance. As shown, there is still plenty of room for new research in the area, and the field is currently very active. We have shown the problems been faced as well as the most promising directions for further work.

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. In such case, the music is typically played exactly as nominally written in the score, thus implicitly ignoring the value of a living performance and its underlying art and diversity.

As seen, research on music performance ranges from studies aimed at understanding expressive performance to attempts at modelling aspects of performance in a formal, quantitative and predictive way. This research can 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 colouring.

Even though we are sceptical about a machine completely replacing a human performer, we are sure that this technology will be available in a not very far future for certain tasks. Scenes like the one in the beginning of this paper will not be science-fiction anymore and it is only a matter of time that they will become commonplace. We have also shown that there are currently some attempts in this direction, like the Rencon contest.

We strongly believe that it is time for computer science to work in the music domain. This research will make a great impact in both the arts and sciences. Not in vain, music is more than an interesting and, somehow, odd domain; it is part of our human essence.

Acknowledgements

This research has been partially supported by the Spanish Ministry of Education and Science under the project TIN2006–15041-C04–01. M. Molina-Solana is also supported by FPU Grant AP2007–02119.

References

Arcos, J. L., de Mántaras, R. L., & Serra, X. (1998). Generating expressive musical performances with SaxEx. *Journal of New Music Research, 27*(3), 194–210.

Bresin, R., & Friberg, A. (2000). Emotional coloring of computer-controlled music performances. *Computer Music Journal, 24*(4), 44–63.

Canazza, S., Poli, G. D., Drioli, C., Rodà, A., & Vidolin, A. (2000). Audio morphing different expressive intentions for multimedia systems. *IEEE MultiMedia, 7*(3), 79–83.

Delgado, M., Fajardo, W., & Molina-Solana, M. (2009). INMAMUSYS: Intelligent multiagent music system. *Expert Systems with Applications, 36*(3), 4574–4580.

De Mántaras, R. L., & Arcos, J. L. (2002). AI and music: From composition to expressive performance. *AI Magazine, 23*(3), 43–57.

Dovey, M. J. (1995). Analysis of Rachmaninoff's piano performances using inductive logic programming. In *Proceedings of the eighth european conference on machine learning (ECML95)* (pp. 279–282). London, UK: Springer-Verlag.

Friberg, A. (1991). Generative rules for music performance: A formal description of a rule system. *Computer Music Journal, 15*(2), 56–71.

Friberg, A. (2006). pDM: An expressive sequencer with real-time control of the KTH music performance rules movements. *Computer Music Journal, 30*(1), 56–71.

Friberg, A., Colombo, V., Frydén, L., & Sundberg, J. (2000). Generating musical performances with director musices. *Computer Music Journal, 24*(3), 23–29.

Gabrielsson, A. (1995). *Music and the mind machine. Expressive intention and performance*. Berlin: Springer [pp. 35–47].

Gabrielsson, A. (2003). Music performance research at the millennium. *The Psychology of Music, 31*(3), 221–272.

Grachten, M., Goebl, W., Flossmann, S., & Widmer, G. (2009). Phase-plane representation and visualization of gestural structure in expressive timing. *Journal New Music Research, 38*(2), 183–195.

Hiraga, R., Bresin, R., Hirata, K., & Katayose, H. (2004). Rencon 2004: Turing test for musical expression. In *Proceedings of the 2004 conference on new interfaces for musical expression (NIME04)* (pp. 120–123). Singapore: National University of Singapore.

Hong, J.-L. (2003). Investigating expressive timing and dynamics in recorded cello. *Psychology of Music, 31*(3), 340–352.

Ishikawa, O., Aono, Y., Katayose, H., & Inokuchi, S. (2000). Extraction of musical performance rules using a modified algorithm of multiple regression analysis. In *Proceedings of the 2000 international computer music conference, San Francisco* (pp. 348–351).

Juslin, P. N. (1997). Perceived emotional expression in synthesized performances of a short melody: Capturing the listeners judgment policy. *Musicae Scientiae, 1*(2), 225–256.

Juslin, P. N., Friberg, A., & Bresin, R. (2002). Toward a computational model of expression in performance: The GERM model. *Musicae Scientiae*(special issue), 63–122.

Langner, J., & Goebl, W. (2003). Visualizing expressive performance in tempo-loudness space. *Computer Music Journal, 27*(4), 69–83.

Lindström, E. (2006). Impact of melodic organization on perceived structure and emotional expression in music. *Musicae Scientiae, 10*, 85–117.

Meyer, L. (1956). *Emotion and meaning in music*. Chicago, IL: University of Chicago Press.

Minsky, M. (1992). *Machine models of music. Music mind and meaning*. Cambridge, MA, USA: MIT Press.

Molina-Solana, M., Arcos, J. L., & Gomez, E. (2008). Using expressive trends for identifying violin performers. In *Proceedings of ninth international conference on music information retrieval (ISMIR2008)* (pp. 495–500).

Palmer, C. (1996). Anatomy of a performance: Sources of musical expression. *Music Perception, 13*(3), 433–453.

Poli, G. D. (2004). Methodologies for expressiveness modelling of and for music performance. *Journal of New Music Research, 33*(3), 189–202.

Ramírez, R., Maestre, E., Pertusa, A., Gómez, E., & Serra, X. (2007). Performance-based interpreter identification in saxophone audio recordings. *IEEE Transactions on Circuits and Systems for Video Technology, 17*(3), 356–364.

Rigg, M. G. (1964). The mood effects of music: A comparison of data from former investigators. *Journal of Psychology, 58*, 427–438.

Sapp, C. (2007). Comparative analysis of multiple musical performances. In *Proceedings of eighth international conference on music information retrieval (ISMIR 2007) Vienna, Austria* (pp. 497–500).

Saunders, C., Hardoon, D., Shawe-Taylor, J., & Widmer, G. (2008). Using string kernels to identify famous performers from their playing style. *Intelligent Data Analysis, 12*(4), 425–440.

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

Sloboda, J. A. (1983). The communication of musical metre in piano performance. *Quarterly Journal of Experimental Psychology, 35*, 377–396.

Stamatatos, E., & Widmer, G. (2005). Automatic identification of music performers with learning ensembles. *Artificial Intelligence, 165*(1), 37–56.

Sundberg, J. (2000). Emotive transforms. *Phonetica, 57*, 95–112.

Temperley, D. (2007). *Music and probability*. The MIT Press.

Todd, N. P. (1989). A computational model of rubato. *Contemporary Music Review, 3*(1), 69–88.

Todd, N. P. (1992). The dynamics of dynamics: A model of musical expression. *Journal of the Acoustical Society of America, 91*(6), 3540–3550.

Widmer, G. (2003). Discovering simple rules in complex data: A meta-learning algorithm and some surprising musical discoveries. *Artificial Intelligence, 146*(2), 129–148.

Widmer, G., Dixon, S., Goebl, W., Pampalk, E., & Tobudic, A. (2003). In search of the Horowitz factor. *AI Magazine, 24*(3), 111–130.

Widmer, G., & Goebl, W. (2004). Computational models of expressive music performance: The state of the art. *Journal of New Music Research, 33*(3), 203–216.

Zanon, P., & Poli, G. D. (2003). Estimation of parameters in rule systems for expressive rendering in musical performance. *Computer Music Journal, 27*(1), 29–46.