Playing with Cases: Rendering Expressive Music with Case-Based Reasoning*

Ramon Lopez de Mantaras and Josep-Lluis Arcos

■ This paper surveys long-term research on the problem of rendering expressive music by means of AI techniques with an emphasis on Case-Based Reasoning. Following a brief overview discussing why people prefer listening to expressive music instead of non-expressive synthesized music, we examine a representative selection of well-known approaches to expressive computer music performance with an emphasis on AI-related approaches. In the main part of the paper we focus on the existing CBR approaches to the problem of synthesizing expressive music, and particularly on TempoExpress, a case-based reasoning system developed at our Institute, for applying musically acceptable tempo transformations to monophonic audio recordings of musical performances. Finally we briefly describe an ongoing extension of our previous work consisting of complementing audio information with information about the gestures of the musician. Music is played through our bodies, therefore capturing the gesture of the performer is a fundamental aspect that has to be taken into account in future expressive music renderings. This paper is based on the "2011 Robert S. Engelmore Memorial Lecture" given by the first author at AAAI/IAAI 2011.

Why is expressive music important?

The simple rendering of a quantized score by a sequencer sounds monotonous and uninteresting. On the other hand, musicians make intentional deviations from the score to convey their own interpretation of the music. These deviations constitute what we call expressiveness and are mostly intended to clarify the musical structure of the composition. This includes the metrical structure (Sloboda, 1983), the phrasing (Gabrielsson, 1987), and harmonic structure (Palmer, 1966). Besides clarifying the structure, expressiveness is also used as a way of communicating affective content (Juslin, 2001; Lindström, 1992; Gabrielsson, 1995).

But, why do we prefer listening to expressive music instead of non-expressive synthesized music? There is a neurological explanation for that: The brain is interested in change. Indeed, auditory neurons, like most neurons in the brain, fire constantly even in silent environments. What really matters, therefore, is not the base firing rate but the changes in firing rate. There are auditory neurons whose firing rate changes only when the sound frequency or the sound intensity increase or decrease. Other neurons react similarly when a sound repeats. Conversely, most of the primary auditory neurons also exhibit what is known as habituation (Baars, 1998) which means that when neurons repeatedly receive the same stimulus their firing rate decreases over time, which means that we deafen to a sound unless it manifests some sort of novelty or renewal in its characteristics. Therefore, it is not surprising that music becomes more interesting when it contains alterations in dynamics, timbre, pitch, rhythm, etc. This pack of alterations might at least partially explain why synthesized music is much less interesting than human-preformed music: A real instrument gives the auditory cortex more stimuli to respond to than synthesized music (Jourdain, 1977). The

* Dedicated to Max Mathews (11/13/1926 – 4/21/2011) for his seminal work on music synthesis

alterations provided by expressive music resources such as changes of timing, loudness, phrasing and improvised ornamentation are an extremely rich source of stimuli to our brains that are absent in the inexpressive, mechanical renderings.

By tickling our neurons, music reaches our hearts. Emotions arise in part through the ups and downs of pitch, dynamics, rhythm, and tension (alteration between consonance and dissonance) in music. Indeed, certain sounds elicit powerful emotions in people possibly as a consequence of evolution because music is built on universal features of human sound processing that have deep evolutionary roots (Trainor, 2008). Mothers in all cultures talk and sing to their infants using a cooing soft voice with high pitch (known as "motherese"). By doing so and introducing melodic and rhythmic variations, mothers help pre-linguistic infants regulate their emotional states (Trainor, 2008).

Synthesizing expressive music with AI

In this section we focus on well known approaches to expressive computer music performance with an emphasis on AI-related approaches. For a complete survey on expressive computer music performance we refer the reader to Kirke and Miranda (2009).

One of the first attempts to address expressiveness in music is that of Johnson (1992). She developed an expert system to determine the tempo and the articulation to be applied when playing Bach's fugues from "The Well-Tempered Clavier". The rules were obtained from two expert human performers. The output gives the base tempo value and a list of performance instructions on notes duration and articulation that should be followed by a human player. The results very much coincide with the instructions given in well known commented editions of "The Well-Tempered Clavier". The main limitation of this system is its lack of generality because it only works well for fugues written on a 4/4 meter. For different meters, the rules should be different. Another obvious consequence of this lack of generality is that the rules are only applicable to Bach fugues.

The work of the KTH group from Stockholm (Friberg, 1995; Friberg et al., 1998, 2000; Bresin, 2001), is one of the best known long term efforts on performance systems. Their current Director Musices system incorporates rules for tempo, dynamic, and articulation transformations constrained to MIDI. These rules are inferred both from theoretical musical knowledge and experimentally by training, specially using the so-called analysis-by-synthesis approach. The rules are divided in three main classes: Differentiation rules, which enhance the differences between scale tones; Grouping rules, which show what tones belong together; and Ensemble rules, that synchronize the various voices in an ensemble.

Canazza et al (1997) developed a system to analyze how the musician's expressive intentions are reflected in the performance. The analysis reveals two different expressive dimensions: one related to the energy (dynamics) and the other one related to the kinetics (rubato) of the piece. The authors also developed a program for generating expressive performances according to these two dimensions.

The work of Dannenberg and Derenyi (1998) is also a good example of articulation transformations using manually constructed rules. They developed a trumpet synthesizer that combines a physical model with a performance model. The goal of the performance model is to generate control information for the physical model by means of a collection of rules manually extracted from the analysis of a collection of controlled recordings of human performance.

Another approach taken for performing tempo and dynamics transformation is the use of neural network techniques. In (Bresin, 1998), a system that combines symbolic decision rules with neural networks is implemented for simulating the style of real piano performers. The outputs of the neural networks express time and loudness deviations. These neural networks extend the standard feed-forward network trained with the back propagation algorithm with feedback connections from the output neurons to the input neurons. The Emotional Flute system (Camurri et al. 2000) also uses artificial neural networks to train the system to play expressively. This system is related and extends Bresin's system in order to deal with a flute and by adding a way of modeling the mood of the performance. They use several neural networks, one for timing, one for loudness and a third one for crescendo and diminuendo at the note level.

There are several very interesting approaches based on Evolutionary Computation (EC). For instance, Ramirez and Hazan used a Genetic Algorithm (GA) to learn a set of regression trees (Ramirez and Hazan, 2005) that emulate a set of human performance actions. They also applied GA to learn performance rules (Ramirez and Hazan, 2007). Zhang and Miranda (2006) also applied a GA to compute timing and dynamics curves for a given melody. These curves are then used to influence the evolution of pulse sets (sets of numbers multiplying tempo and dynamic values in the score) which are unique to each composer. That is, each composer has a unique pattern of amplitude and tempo variations (a unique pulse) running through performances. In (Zhang and Miranda (2007), the authors have proposed a Multiagent System based on the

hypothesis that expressive performance evolves as a result of interaction in the performer's society. That is, each performer agent listens to other performer agents and learns by imitation from those performances that are better than their own. The differences in the performances are computed based on their pulse sets. This social dimension is a very interesting idea because it certainly reflects what human performers actually do.

Most of the systems are limited to two expressive resources such as timing and dynamics, or timing and articulation. This limitation has to do with the fact that it is very difficult to find models general enough to capture the variety present in different performances of the same piece by the same musician and even the variety within a single performance (Kendall and Carterette, 1990). Furthermore, the different expressive resources interact with each other. That is, the models for dynamics alone change when rubato is also taken into account. Obviously, due to this interdependency, the more expressive resources one tries to model, the more difficult is finding the appropriate models.

Widmer et al. (2009) describe a computer program that learns to expressively perform classical piano music. The approach is data intensive and based on statistical learning. Performing music expressively certainly requires high levels of creativity, but the authors take a very pragmatic view to the question of whether their program can be said to be creative or not and claim that "creativity is in the eye of the beholder." In fact, the main goal of the authors is to investigate and better understand music performance as a creative human behavior by means of AI methods. For additional information on approaches to computational creativity, we refer the reader to the special issue of AI Magazine (AI Magazine, 2009) edited by Colton et al. (2009).

CBR approaches to expressive music rendering

The basic principle underpinning Case-Based Reasoning (CBR) is that a new problem can be solved by reusing solutions to past similar problems (Lopez de Mantaras, 2001, 2006a, 2006b). The main advantage of CBR is that a case is a very convenient way of capturing knowledge, specially in weak theory domains, where the relations between causes and effects may not be well understood. To avoid this limitation, we developed a system called SaxEx (Arcos et al., 1998) a computer program capable of synthesizing high quality expressive tenor sax solo performances of jazz ballads based on cases representing human solo performances. As mentioned above, previous rule-based approaches cannot easily deal with many expressive parameters simultaneously because it is too difficult to infer rules general enough to capture the variety present in expressive performances. Besides, the different expressive parameters interact with each other making it even more difficult to find appropriate rules taking into account these interactions.

With CBR, we have shown that it is possible to deal with the five most important expressive parameters: dynamics, rubato, vibrato, articulation, and attack of the notes. To do so, SaxEx uses a case memory containing examples of human performances, analyzed by means of spectral modeling techniques and background musical knowledge. The score of the piece to be performed is also provided to the system. The core of the method is to analyze each input note determining (by means of the background musical knowledge) its role in the musical phrase it belongs to, identify and retrieve (from the case-base of human performances) notes with similar roles, and finally, transform the input note so that its expressive properties (dynamics, rubato, vibrato, articulation, and attack) match those of the most similar retrieved note. Each note in the case base is annotated with its role in the musical phrase it belongs to, as well as with its expressive values. Furthermore, cases do not contain just information on each single note but they include contextual knowledge at the phrase level. Therefore, cases in this system have a complex object-centered representation.

Although limited to monophonic performances, the results convincingly demonstrate that CBR is a very powerful methodology to directly use the knowledge of a human performer that is implicit in her playing examples rather than trying to make this knowledge explicit by means of rules. Some audio results can be listened at http://www.iiia.csic.es/Projects/music/Saxex.html. More recent papers (Arcos and Lopez de Mantaras, 2001; Lopez de Mantaras and Arcos, 2002), describe this system in great detail.

Based on the work on SaxEx, we developed TempoExpress (Grachten et al. 2006), a case-based reasoning system for applying musically acceptable tempo transformations to monophonic audio recordings of musical performances. Existing algorithms are mainly focused on maintaining sound quality of audio recordings, rather than maintaining the musical quality of the audio. However, as demonstrated by H. Honing (2007), humans are able to detect, based only on expressive aspects of the performances, whether audio recordings are original or uniformly time stretched. The next section describes in some detail this system. For a very detailed description we refer the reader to (Grachten et al. 2006).

TempoExpress: A tempo transformation system

TempoExpress has a rich description of the musical expressivity of the performances, that includes not only timing deviations of performed score notes, but also represents more rigorous kinds of expressivity such as note ornamentation, consolidation, and fragmentation. Within the tempo transformation process, the expressivity of the performance is adjusted in such a way that the result sounds natural for the new tempo. A case base of previously performed melodies is used to infer the appropriate expressivity. The problem of changing the tempo of a musical performance is not as trivial as it may seem because it involves a lot of musical knowledge and creative thinking. Indeed, when a musician performs a musical piece at different tempos the performances are not just time-scaled versions of each other (as if the same performance were played back at different speeds). That is, changing the tempo is a problem that cannot be reduced to applying what is known as a Uniform Time Stretching (UTS) transformation to the original tempo. This is so because together with the changes of tempo, variations in musical expression need to be made (Desain and Honing, 1994). Such variations do not only affect the timing of the notes, but can also involve for example the addition or deletion of ornamentations, or the consolidation/fragmentation of notes. Apart from the tempo, other domain specific factors seem to play an important role in the way a melody is performed, such as meter, and phrase structure. Tempo transformation is one of the audio post-processing tasks manually done in audio-labs. Automatizing this process may, therefore, be of industrial interest.

TempoExpress architecture

A schematic view of the system is hown in figure 1. We will focus our explanation on the gray box, that is, the steps involved in modifying the expressive parameters of the performance at the musical level. For a detailed account of the audio analysis and audio synthesis components, we refer the reader to Gómez et al. (2003) and Maestre and Gómez (2005).

Given a score of a phrase, a monophonic audio recording of a saxophone performance of that phrase at a particular source tempo, and a number specifying the desired target tempo, the task of the system is to render the audio recording at the desired target tempo adjusting the expressive parameters of the performance in accordance with the target tempo. In order to apply the CBR process, the first task is to build a phrase input problem specification from the given input data (see figure 1). This is a data structure that contains all the information necessary to define a tempo transformation task for a musical phrase. Besides the given source and target tempos and the input audio performance, the phrase input problem specification requires an abstract description of the melody as well as a description of the expressivity of the input performance. These two extra pieces of information are automatically inferred by the modules Musical Analysis and Performance Annotation (see figure 1).

Figure 1 (Schematic view of TempoExpress) over here

The musical analysis is inferred from the score and derives information about various kinds of structural aspects of the score. In particular, it derives a description of the melodic surface of the phrase, above the note level, in terms of the eight basic Implication-Realization" structures of Narmour (Narmour, 1990; Lopez de Mantaras and Arcos, 2002), and a segmentation of the phrase capturing the grouping of notes within the phrase. The performance annotation is computed by comparing, via the edit-distance, the score and the input performance.

The performance annotation describes the musical behavior of the performer by means of a sequence of performance events that maps the performance to the score. For example, the occurrence of a note that is present in the score but has no counterpart in the audio performance will be represented by a deletion event. Although important, such deletion events are not very common since the majority of score notes are actually performed, be it with alterations in timing and dynamics. This type of event is called transformation event because it establishes a correspondence between the note in the score and the corresponding note in the performance. Once such a correspondence is established, expressive transformations such as onset time, duration and dynamic changes can be derived by calculating the differences of these attributes on a note-to-note basis. Analyzing the corpus of monophonic tenor saxophone recordings of jazz standards that we have used (4256 performed notes), we identified the following types of performance events: Insertion (the occurrence of a performed note that is not present in the score), deletion (the presence of a note in the score that does not occur in the performance), consolidation (multiple notes in the score that are performed as a single note whose duration is approximately the sum of the durations of the multiple corresponding notes in the score), fragmentation (a single note in the score that is performed as multiple notes whose total duration is approximately equal to the duration of the single score note), and ornamentation (the insertion of one or several short notes, not present in the score, to anticipate a score note that is also a performed

note). In order to infer the sequence of performance events, the notes in the performance are matched to the notes in the score using the well-known edit-distance (Levenshtein, 1966).

An example of performance annotation is shown in figure 2. The bars below the staff represent performed notes. The letters represent the performance events ("T" for transformation, "O" for ornamentation, "C" for consolidation, and "D" for deletion).

Figure 2 (Performance annotation of "Body and Soul") over here

Once we have build the phrase input problem, the CBR problem solving cycle can start. The phrase input problem is used to query the case base, whose cases contain the scores of phrases together with twelve performance annotations for each phrase that correspond to audio performances at twelve different tempos. The goal is to retrieve the phrase in the case base with highest similarity to the phrase input problem and reuse the solution. This is done analyzing the differences between the performance annotations at the source and target tempo in the retrieved phrase and adapting (reusing) these differences in order to infer the performance annotation of the phrase input problem at the target tempo. Next we further describe this CBR problem solving process with the help of the example of Figure 3. In particular we explain how a solution is obtained for each segment of each phrase input problem. We do so by briefly explaining the numbered steps, shown in figure 3, one by one.

Figure 3 (Example of case retrieval and reuse for an input segment) over here

The first step is to find the case in the case base that is most similar to the input problem. The similarity is assessed by calculating the edit-distance, at the note level, between the sequence of score notes of the segment input problem and the sequences of score notes of the segments of all the phrases contained in the case base.

In the second step, an optimal alignment between the input problem and the most similar segment, retrieved in step one, is made. This optimal alignment is actually given as a side effect of the computation of the edit-distance in step one.

In the third step, the performance annotations corresponding to the relevant tempos are extracted. That is, the source tempo for the input problem, and the source and target tempo for the retrieved segment, in such a way that the source tempo of the retrieved segment is similar (within a 10 BPM tolerance interval) to the source tempo of the input segment and the target tempo of the retrieved segment is similar to the target tempo given by the user.

The fourth step consists in linking, in the retrieved segment, the performance annotation at the source tempo with the performance annotation at the target tempo. In figure 3 this linking can be seen in the upper part of box 4 and consists in the following three relations: ⟨T → T⟩, ⟨TT →OTT⟩, ⟨C → TT⟩. Besides, the alignment between the input segment and the retrieved segment, given by the edit-distance, is used to determine which performance events from the retrieved segment belong to which performance events of the input segment leading to what we call annotation patterns. In figure 3 we can see the following three annotation patterns: [T, ⟨T → T⟩], [T, ⟨TT →OTT⟩], and [T, ⟨C → TT⟩]. The first pattern reflects a rather simple situation because it involves the same number of notes (one in this case) in the input segment performance at the source tempo as well as in the two performances at different tempos (source and target) of the retrieved segment. This pattern means that a score note of the retrieved segment was played as T at the source tempo and played as T (most probably with some dynamic, duration, and onset deviations) at the target tempo while a melodically similar note of the input segment has been played as T at the source tempo. Based on this, the CBR system infers how to play the input segment note at the target tempo by imitating the dynamic, duration and onset deviations used in the target tempo of the retrieved segment.

The remaining two annotation patterns are a bit more complex because they involve a different number of notes. More concretely we can see that a single note in the input segment corresponds to two notes in the retrieved segment. To deal with these situations, the system employs a set of adaptation rules that are used in the fifth step. Figure 3 shows the two rules that have been respectively applied to these annotation patterns in the fifth step. We will see why the upper rule infers OT based in the case of the annotation pattern [T, ⟨TT →OTT⟩]. Indeed, this annotation pattern indicates that in the retrieved segment two notes were performed as two transformation events at the source tempo but an ornamentation note was added an the target tempo performance. Since the performance of the input segment at the source tempo is T, the application of the rule infers that the performance at the target tempo should be OT. The net result is thus the introduction of an ornamentation note in front.

The lower rule in the fifth step states that the annotation pattern [T, ⟨C → TT⟩], infers F. The motivation for this is that from an acoustic point of view changing a performance from a consolidation event (C) to two transformation events (TT) amounts to changing from one performed note to two performed notes. To reproduce this perceptual effect when the input performance is a single performed note (T), a fragmentation of this note has to be applied.

We have experimentally evaluated the results of TempoExpress on the task of tempo transformation and compared these results with a Uniform Time Stretching (UTS) process (Grachten et al. 2006). A leave-one-

out method was used to evaluate the system over 64 input segments involving a total of 6364 note tempo transformation problems. For each transformation problem, the TempoExpress performance at the target tempo was compared, by means of the edit-distance between performance annotations, to both a UTS-based performance and a human performance also at the target tempo. The conclusion is that TempoExpress is clearly closer (Wilcoxin signed-rank test significance $p < 0.001$) than UTS to the human performance when the target tempo is slower than the source tempo. When the target tempo is faster than the source tempo the improvement is not statistically significant.

Other CBR approaches to expressive music

Other applications of CBR to expressive music are those of Suzuki (2003), and those of Tobudic and Widmer (2003, 2004). Suzuki's Kagurame system (2003), uses examples of expressive performances to generate multiple polyphonic MIDI performances of a given piece with varying musical expression, however they deal only with two expressive parameters due to the limitations of the MIDI representation. Although the task of their system is performance generation rather than transformation, it has some sub-tasks in common with our approach, such as performance to score matching, segmentation of the score, melody comparison for retrieval, and the use of the edit-distance for performance-score alignment.

Tobudic and Widmer (2003) apply instance-based learning (IBL) also to the problem of generating expressive performances. The IBL approach is used to complement a note-level rule-based model with some predictive capability at the higher level of musical phrasing. More concretely, the IBL component recognizes performance patterns, of a concert pianist, at the phrase level and learns how to apply them to new pieces by analogy. The approach produced some interesting results but, as the authors recognize, was not very convincing due to the limitation of using an attribute-value representation for the phrases. Such simple representation cannot take into account relevant structural information of the piece, both at the sub-phrase level and at the inter-phrasal level. In a subsequent paper, Tobudic and Widmer (2004), succeeded in partly overcoming this limitation by using a relational phrase representation.

Adding Gesture

Music is played through our bodies. These body movements may be involved in the sound production or may pursue the goal of enforcing emotional communication. In a recent experiment (Vines et al, 2011) demonstrated the contribution of musician's movements not involved in sound production to enforce musical expressivity. Therefore, capturing the gesture of the performer is another fundamental aspect that has to be taken into account in expressive music renderings.

Gesture capture can be done by adding sensors to instruments becoming "augmented" instruments or "hyper-instruments". Take a traditional instrument, for example a cello, and connect it to a computer through electronic sensors in the neck and in the bow, equip also with sensors the hand that holds the bow and program the computer with a system similar to SaxEx that allows to analyze the way the human interprets the piece, based on the score, on musical knowledge and on the readings of the sensors. The results of such analysis allow the hyper-instrument to play an active role altering aspects such as timbre, tone, rhythm and phrasing as well as generating an accompanying voice. In other words, this yields an instrument that can be its own intelligent accompanist. Tod Machover, from MIT's Media Lab, developed an hyper-cello and the great cello player Yo-Yo Ma premiered a piece, composed by Tod Machover, called "Begin Again Again..." at the Tanglewood Festival several years ago. The hyper-cello is based on the Hyperbow system (Young 2002) initially developed to capture the performance parameters in violin playing. Also related with modeling violin expressivity, inductive logic programming techniques have been applied to learn violin expressive models by combining audio and gestural information (Ramirez et al, 2010).

Gesture analysis has been also conducted in woodwind instrument performers (Wanderley and Depalle, 2004). Their experiments with a clarinet show how some expressive nuances are directly caused by body movements not directly related to sound production. For instance, postural adjustments or upward/downward movements of the instrument influence recorded sound.

Heijink and Meulenbroek (2002) proposed the use of a three-dimensional motion tracking system, Optotrak 3020, to analyze the left hand fingering in a classical guitar. Their experiments demonstrate that, although biomechanical hand constraints play a role when playing, fingering decisions are mainly aimed at producing the desired expressive effect. Norton (2008) is another example of the use of an optical motion caption system based on a capture system by Phase Space Inc., with quite successful results. For a detailed review of existing approaches to gestural acquisition in music we refer the reader to Wanderley and Depalle (2004).

Extending our previous work, we are currently focused on complementing audio information with information of musician gestures. This multimodal approach is very useful when analyzing string instruments where the same notes can be played at different positions or when the analysis of the fingers' movements allows to characterize expressive nuances very difficult to capture with the current audio analysis technology. Our research is focused on the study of guitar expressivity and aims at designing a

system able to model and extend the expressive resources of that instrument (see http://www.iiia.csic.es/guitarLab).

Musician gestures are captured by a sensing system mounted in the guitar fretboard (Guaus et al 2010). The sensors are non-intrusive to the player and track the gestures of the left hand fingers (see Figure 4). The system captures from macro-scale changes (i.e. the presence of finger bars) to micro-scale changes (i.e. vibrato) in player's movements. Specifically, gesture information is used to model expressive articulations such as legatos, appoggiaturas, glissandi, and vibratos. Moreover, preliminary experiments show that gesture information allows to build a deeper fingering model that, in turn, improves note identification and characterization. We are analyzing the use of these expressive resources working with pieces of different styles such as Bach Preludes or Jazz Standards.

Figure 4 (Non intrusive capacitive sensors mounted on the first 10 frets of a nylon strings guitar) over here

Concluding Remarks

In the first part of this paper, we presented a brief overview discussing why we prefer listening to expressive music instead of lifeless synthesized music. Next we have surveyed a representative selection of well-known approaches to expressive computer music performance with an emphasis on AI-related approaches. In the second part of the paper we have focused on the existing CBR approaches to the problem of synthesizing expressive music, and particularly on TempoExpress, a case-based reasoning system developed at our Institute, for applying musically acceptable tempo transformations to monophonic audio recordings of musical performances. Experimental results have shown that the TempoExpress tempo transformations are better than the Uniform Time Stretching (UTS) ones, in the sense that they are closer to human performances when the target tempo is slower than the source tempo. Finally we briefly survey some work on gesture caption and analysis and particularly our current and future work on complementing audio information with information of musician gestures in the case of a study of guitar expressivity. Specifically, gesture information is used to model expressive articulations, appoggiaturas, glissandi, and vibratos. Preliminary experiments show that gesture information allows to build a better fingering model that, in turn, improves note identification and characterization with the aim of extending the expressive modeling of that instrument.

Acknowledgments

We are grateful to Maarten Grachten, Enric Guaus and Xavier Serra for their contributions to the described work during our long-term research on expressive music rendering. This research is partially supported by the Ministry of Science and Innovation under the project NEXT-CBR (TIN2009-13692-C03-01) and the Generalitat de Catalunya AGAUR Grant 2009-SGR-1434.

References

AI Magazine 30(3) 2009. Special issue on Computational Creativity.

Arcos, J.L., Lopez de Mantaras, R., and Serra, X. 1998. SaxEx: A case-based reasoning system for generating expressive musical performances, Journal of New Music Research 27(3): 194-210.

Arcos, J.L., and Lopez de Mantaras, R. 2001. An interactive case-based reasoning approach for generating expressive music, Applied Intelligence 14(1): 115-129.

Baars, B. 1998. A cognitive theory of consciousness. Cambridge University Press, New York.

Bresin, R. 1998. Artificial neural networks based models for automatic performance of musical scores, Journal of New Music Research 27(3): 239-270.

Bresin, R. 2001. Articulation rules for automatic music performance. In Proceedings of the 2001 International Computer Music Conference 2001. San Francisco, Calif.: International Computer Music Association.

Cammuri, A., Dillon, R., and Saron, A. 2000. An experiment on analysis and synthesis of musical expressivity. In Proceedings of the 13th Colloquium on Musical Informatics. L'Aquila, Italy, September.

Canazza, S.; De Poli, G.; Roda, A., and Vidolin, A. 1997. Analysis and synthesis of expressive intention in a clarinet performance. In Proceedings of the 1997 International Computer Music Conference, 113-120. San Francisco, Calif.: International Computer Music Association.

Colton, S., Lopez de Mantaras, R., Stock, O. 2009. Computational creativity: Coming of age. AI Magazine 30(3): 11-14.

Dannenberg, R.B., and Derenyi, I. 1998. Combining instrument and performance models for high quality music synthesis, Journal of New Music Research 27(3): 211-238.

Desain, P., Honing, H. 1994. Does expressive timing in music performance scale proportionally with tempo? Psychological Research, 56:285–292.

Friberg, A. 1995. A quantitative rule system for musical performance. PhD dissertation. KTH, Stockholm.

Friberg, A.; Bresin, R.; Fryden, L., and Sunberg, J. 1998. Musical punctuation on the microlevel: automatic identification and performance of small melodic units. Journal of New Music Research 27(3): 271-292.

Friberg, A.; Sunberg J., and Fryden L. 2000 . Music From Motion: Sound Level Envelopes of Tones Expressing Human Locomotion. Journal on New Music Research, 29(3): 199-210.

Gabrielsson, A. 1987. Once again: The theme from Mozart's piano sonata inAmajor (K. 331).Acomparison of five performances. In Gabrielsson, A., editor, Action and perception in rhythm and music, pages 81–103. Royal Swedish Academy of Music, Stockholm.

Gabrielsson, A. 1995. Expressive intention and performance. In Steinberg, R., editor, Music and the Mind Machine, pages 35–47. Springer-Verlag, Berlin.

Gomez, E., Grachten, M., Amatriain, X., Arcos, J. L. 2003. Melodic characterization of monophonic recordings for expressive tempo transformations. In Proceedings of Stockholm Music Acoustics Conference 2003.

Grachten, M., Arcos, J.L., Lopez de Mantaras, R. 2006. A case based approach to expressivity-aware tempo transformation. Machine Learning 65(2-3): 411-437.

Guaus, E., Ozaslan, T., Palacios, E., Arcos, J.L. 2010. A left hand gesture caption system for guitar based on capacitive sensors," in Proceedings of NIME-2010, pp. 238– 243.

Heijink, H., and Meulenbroek, R.G.J. 2002. On the complexity of classical guitar playing:functional adaptations to task constraints. Journal of Motor Behavior 34:4, 339–351.

Honing, H. 2007. Is expressive timing relational invariant under tempo transformation? Psychology of Music 35(2): 276–285.

Johnson, M.L. 1992. An expert system for the articulation of Bach fugue melodies. In Readings in Computer Generated Music, ed. D.L. Baggi, 41-51. Los Alamitos, Calif.: IEEE Press.

Jourdain, R. 1977. Music, Brain, and Ecstasy. New York: Avon Books.

Juslin, P. 2001. Communicating emotion in music performance: a review and a theoretical framework. In Juslin, P. and Sloboda, J., editors, Music and emotion: theory and research, pages 309–337. Oxford University Press, New York.

Kendall, R.A., and Carterette, E.C. 1990. The communication of musical expression. Music Perception 8(2):129.

Kirke, A., Miranda, E.R. 2009. A Survey of Computer Systems for Expressive Music Performance. ACM Computing Surveys 42(1): 3:1-3:41.

Levenshtein, V.I. 1966. "Binary codes capable of correcting deletions, insertions, and reversals". Soviet Physics Doklady 10(8): 707–710

Lindström, E. 1992. 5 x "oh, my darling clementine". the influence of expressive intention on music performance. Department of Psychology, Uppsala University.

Lopez de Mantaras, R. 2001. Case-Based Reasoning. Machine Learning and its Applications. Lecture Notes in Artificial Intelligence nº 2049, pp. 127-145. Springer-Verlag.

Lopez de Mantaras, R., Arcos, J.L. 2002. AI and Music: From composition to expressive performance. AI Magazine 23(3): 43-57.

Lopez de Mantaras, R., McSherry, D., Bridge, D., Leake, D., Smyth, B., Craw, S., Faltings, B., Maher, M.L., Cox, M., Forbus, K., Keane, M., Aamodt, A., Watson, I. 2006a. Retrieval, Reuse, Revise, and Retention in CBR. Knowledge Engineering Review 20(3): 215-240.

Lopez de Mantaras, R., Perner, P., Cunningham, P. 2006b. Emergent Case-Based Reasoning Applications. Knowledge Engineering Review 20(3):325-328.

Maestre, E., Gomez, E. 2005. Automatic characterization of dynamics and articulation of expressive monophonic recordings. In Proceedings of the 118th Audio Engineering Society Convention, Barcelona.

Narmour, E. 1990. The analysis and cognition of basic melodic structures: The Implication-Realization model. Chicago: University of Chicago Press.

Norton, J. 2008. Motion capture to build a foundation for a computer-controlled instrument by study of classical guitar performance. PhD thesis, Stanford University.

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

Ramirez, R., Hazan, A. 2005. Modeling expressive performance in jazz. In Proceedings of the 18th International Florida ArtificiaI Intelligence Research Society Conference (AI in Music and Art), Clearwater Beach, FL, May 2005, AAAI Press, Menlo Park, CA, 86–91.

Ramirez, R., Hazan, A. 2007. Inducing a generative expressive performance model using a sequentialcovering genetic algorithm. In Proceedings of the Genetic and Evolutionary Computation Conference. London, UK. ACM Press, New York, NY.

Ramirez, R., Perez, A., Kersten, S, Rizo, D., Román, P., Iñesta, J.M. 2010. Modeling Violin Performances Using Inductive Logic Programming. Intelligent Data Analysis, 14:5, 573-585.

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

Suzuki, T. 2003. The second phase development of case based performance rendering system "Kagurame". In Working Notes of the IJCAI-03 Rencon Workshop, pages 23–31.

Tobudic, A., Widmer, G. 2003. Playing Mozart phrase by phrase. In Ashley, K.D. and Bridge, D.G. (eds) Proceedings of the Fifth International Conference on Case-Based Reasoning. Berlin: Springer, 552-566.

Tobudic, A., Widmer, G. 2004. Case-based relational learning of expressive phrasing in classical music. In Proceedings of the 7th European Conference on Case-based Reasoning (ECCBR'04), Madrid.

Trainor, L. 2008. The neural roots of music. Nature 453: 598-599.

Widmer, G., Flossmann, S., Grachten, M. 2009. YQX plays Chopin. AI Magazine 30(3): 35-48.

Vines, B. W., Krumhansl, C. L., Wanderley. M. M. , Dalca, I. M., and Levitin, D. J. 2011. Music to my eyes: Cross-modal interactions in the perception of emotions in musical performance. Cognition 118:157–170.

Wanderley. M. M. and Depalle, P. 2004. Gestural control of sound synthesis. Proceedings of the IEEE, vol. 92:4, 632-644.

Young, D. 2002. The hyperbow controller: Real-time dynamics measurement of violin performance. In Proceedings of New Interfaces for Musical Expression.

Zhang, Q., Miranda, E. R. 2006a. Evolving musical performance profiles using genetic algorithms with structural fitness. In Proceedings of the 8th Annual Conference on Genetic and Evolutionary Computation, Seattle,Washington. J. V. Diggelen,M. A. Wiering, and E. D. D. Jong, Eds. ACM Press, New York, NY, 1833–1840.

Zhang, Q., Miranda, E. R. 2007. Evolving expressive music performance through interaction of artificial agent performers. In Proceedings of the ECAL Workshop on Music and Artificial Life (MusicAL), Lisbon, Portugal.

Ramon Lopez de Mantaras is Research Professor of the Spanish National Research Council (CSIC) and Director of the Artificial Intelligence Research Institute (IIIA). Master of Sciences in Computer Science from the University of California Berkeley, PhD in Physics (Automatic Control) from the University of Toulouse, and PhD in Computer Science from the Technical University of Barcelona. A pioneer of Artificial Intelligence in Spain, with contributions, since 1976, in Pattern Classification, Approximate Reasoning, Expert Systems, Machine Learning, Case-Based Reasoning, Autonomous Robots, and AI & Music. Author of around 250 papers. Invited plenary speaker at numerous international conferences. Former Editor-in-Chief of Artificial Intelligence Communications, current editorial board member of several international journals, and Associate Editor of the Artificial Intelligence Journal. Program committee member in numerious conferences. Program committee co-Chairman of UAI-94 and ECML'00. Conference Chairman of ECAI-06, ECML-07, PKDD-07, and IJCAI-07. ECCAI Fellow. Co-recipient of five best paper awards at international conferences. Recipient of the "City of Barcelona" Research Prize, and the "2011 Association for the Advancement of Artificial Intelligence (AAAI) Robert S. Engelmore Memorial Award". President of the Board of Trustees of IJCAI from 2007 to 2009. Presently working on case-based reasoning, machine learning for autonomous robots and AI applications to music. For additional information please visit: http://www.iiia.csic.es/~mantaras.

Josep Lluís Arcos is a Research Scientist of the Artificial Intelligence Research Institute of the Spanish National Research Council (IIIA-CSIC) where he is member of the Learning Systems Department. Dr. Arcos received the M.S. in Computer Science (1991) and the PhD in Computer Science (1997) from the Technical University of Catalonia, Spain. He also received a M.S. in Sound and Music Technology (1996) from the Pompeu Fabra University, Spain. He is co-author of more than 100 scientific publications an co-recipient of several awards at case-based reasoning conferences and computer music conferences. Presently, he is working on case-based reasoning and machine learning, on self-organization and self-adaptation mechanisms, and on artificial intelligence applications to music.