Combining AI Techniques to Perform Expressive Music by Imitation

Josep Lluís Arcos and Ramon López de Mántaras
IIIA, Artificial Intelligence Research Institute
CSIC, Spanish Council for Scientific Research
Campus UAB, 08193 Bellaterra, Catalonia, Spain.
{arcos, mantaras}@iiia.csic.es, http://www.iiia.csic.es

Abstract

In this brief paper we describe several extensions and improvements of a previously reported system (Arcos, López de Mántaras, & Serra 1998) capable of generating expressive music by imitating human performances. The system is based on Case-Based Reasoning (CBR) and Fuzzy techniques.

Introduction

One of the major difficulties in the automatic generation of music is to endow the resulting piece with the expressiveness that characterizes human performers. Following musical rules, no mater how sophisticated and complete they are, is not enough to achieve expression, and indeed computer music usually sounds monotonous and mechanical. The main problem is to grasp the performers personal touch, that is, the knowledge brought about when performing a score. A large part of this knowledge is implicit and very difficult to verbalize. For this reason, AI approaches based on declarative knowledge representations are very useful to model musical knowledge an indeed we represent such knowledge declaratively in our system, however they have serious limitations in grasping performance knowledge. An alternative approach, much closer to the observation imitation - experimentation process observed in human performers, is that of directly using the performance knowledge implicit in examples of human performers and let the system imitate these performances. To achieve this, we have developped the *SaxEx*, a case-based reasoning system capable of generating expressive performances of melodies based on examples of human performances. CBR is indeed an appropriate methodology to solve problems by means of examples of already solved similar problems.

In the next section we describe the system and in particular the fuzzy set-based extension of thereuse step. Then, we briefly mention some relevant related work and,finally, we give some conclusions.

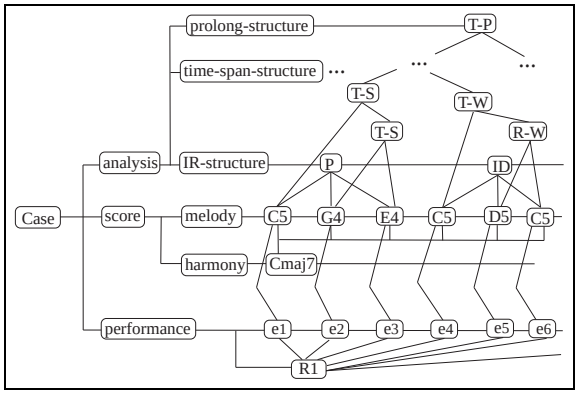

Figure 1: Overall structure of the beginning of an 'All of me' case.

System description

The problem-solving task of the system is to infer, via imitation, and using its case-based reasoning capability, a set of expressive transformations to be applied to every note of an inexpressive musical phrase given as input. To achieve this, it uses a case memory containing human performances and background musical knowledge, namely Narmours theory of musical perception (Narmour 1990) and Lerdahl & Jackendoffs GTTM (Lerdahl & Jackendoff 1993). The score, containing both melodic and harmonic information, is also given.

Modeling musical knowledge

Problems solved by *SaxEx*, and stored in its memory, are represented as complex structured cases embodying three different kinds of musical knowledge (see Figure 1): (1) concepts related to the score of the phrase such as notes and chords, (2) concepts related to background musical theories such as implication/realization (IR) structures and GTTM's timespan reduction nodes, and (3) concepts related to the performance of musical phrases.

A score is represented by a melody, embodying a sequence of notes, and a harmony, embodying a sequence of chords. Each note holds in turn a set of features such as its pitch (C5, G4, etc), its position with respect to the beginning of the phrase, its duration, a reference to its underlying harmony, and a reference to the next note of the phrase. Chords hold also a set of features such as name (Cmaj7, E7, etc), position, duration, and a reference to the next chord.

The musical analysis representation embodies structures of the phrase automatically inferred by *SaxEx* from the score using IR and GTTM background musical knowledge. The analysis structure of a melody is represented by a process-structure (embodying a sequence of IR basic structures), a time-span-reduction structure (embodying a tree describing metrical relations), and a prolongational-reduction structure (embodying a tree describing tensing and relaxing relations among notes). Moreover, a note holds the metrical-strength feature, inferred using GTTM theory, expressing the note's relative metrical importance into the phrase.

The information about the expressive performances contained in the examples of the case memory is represented by a sequence of *affective regions* and a sequence of *events*, one for each note, (extracted using the SMS sound analysis capabilities), as explained below.

Affective regions group (sub)-sequences of notes with common affective expressivity. Specifically, an affective region holds knowledge describing the following affective dimensions: *tender-aggressive*, *sad-joyful*, and *calm-restless*. These affective dimensions are described using five ordered qualitative values expressed by linguistic labels as follows: the middle label represents no predominance (for instance, neither tender nor aggressive), lower and upper labels represent, respectively predominance in one direction (for example, absolutely calm is described with the lowest label). For instance, a jazz ballad can start very tender and calm and continue very tender but more restless. Such different nuances are represented in *SaxEx* by means of different affective regions.

The expressive transformations to be decided and applied by the system affect the following expressive parameters: dynamics, rubato, vibrato, articulation, and attack. Except for the attack, the notes in the human performed musical phrases are qualified using the SMS (Spectral Modeling and Synthesis) system (Serra *et al.* 1997), by means of five different ordered values. For example, for dynamics the values are: very low, low, medium, high and very high and they are automatically computed relative to the average loudness of the inexpressive input phrase. The same idea is used

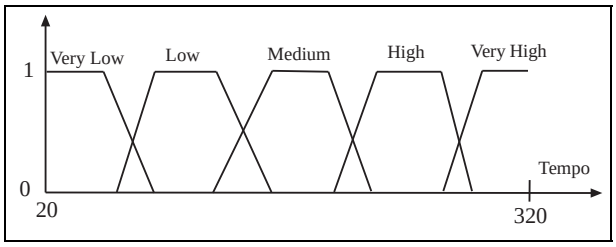

Figure 2: Linguistic fuzzy values for rubato expressive parameter.

for rubato, vibrato (very little vibrato to very high vibrato) and articulation (very legato to very staccato). In the previous system these values where mere syntactic labels but in the improved system, the meanings of these values are modeled by means of fuzzy sets such as those shown in figure 2 for Rubato. We will explain below the advantage of this extension. For the attack we have just two situations: reaching the pitch from a lower pitch or increasing the noise component of the sound.

The SaxEx CBR Task

The task of *SaxEx* is to infer a set of expressive transformations to be applied to every note of an inexpressive phrase given as input. To achieve this, *SaxEx* uses a CBR problem solver, a case memory of expressive performances, and background musical knowledge. Transformations concern the dynamics, rubato, vibrato, articulation, and attack of each note in the inexpressive phrase. The cases stored in the episodic memory of *SaxEx* contain knowledge about the expressive transformations performed by a human player given specific labels for affective dimensions.

For each note in the phrase, the following subtask decomposition (Figure 3) is performed by the CBR problem solving method implemented in Noos:

- *Retrieve*: The goal of the retrieve task is to choose, from the memory of cases (pieces played expressively), the set of precedent notes—the cases—most similar for every note of the problem phrase. Specifically, the following subtask decomposition is applied to each note of the problem phrase:
  - *Identify*: its goal is to build retrieval perspectives (explained in the next subsection) using the affective values specified by the user and the musical background knowledge integrated in the system (retrieval perspectives are described in Subsection). These perspectives guide the retrieval process by focusing it on the most relevant aspects of the current problem, and will be used either in the *search* or in the *select* subtasks.

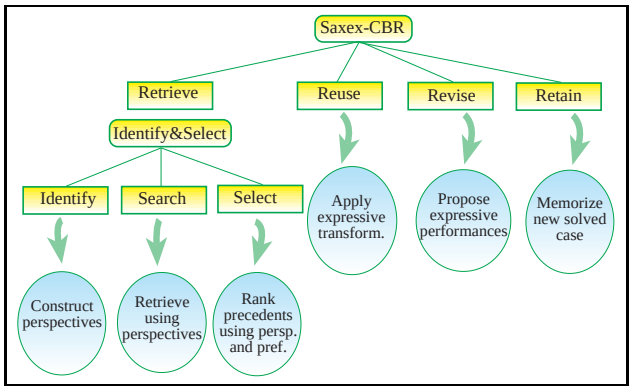

Figure 3: Task decomposition of the *SaxEx* CBR method.

  - *Search*: its goal is to search cases in the case memory using Noos retrieval methods and some previously constructed perspective(s).
  - *Select*: its goal is to rank the retrieved cases using Noos preference methods. The collection of *SaxEx* default preference methods use criteria such as similarity in duration of notes, harmonic stability, or melodic directions.

- *Reuse*: its goal is to choose, from the set of most similar notes previously retrieved, a set of expressive transformations to be applied to the current note. The default strategy of *SaxEx* is the following: the first criterion used is to adapt the transformations of the most similar note. When several notes are considered equally similar, the transformations are computed using a fuzzy combination (see section 'The use of fuzzy techniques ...'). The user can, however, select alternative criteria, not involving this fuzzy combination such as majority rule, minority rule, etc. When the retrieval task is not able to retrieve similar precedent cases for a given note, no expressive transformations are applied to that note and the situation is notified in the revision task. Nevertheless, using the current *SaxEx* case base, the retrieval perspectives allways retrieved at least one precedent in the experiments performed.

- *Revise*: its goal is to present to the user a set of alternative expressive performances for the problem phrase. Users can tune the expressive transformations applied to each note and can indicate which performances they prefer.

- *Retain*: the incorporation of the new solved problem to the memory of cases is performed automatically in *Noos* from the selection performed by the user in the *revise* task. These solved problems will be available for the reasoning process when solving future problems. Only positive feedback is given. That is, only those examples that the user judges as good expressive interpretations are actually retained.

In previous versions of *SaxEx* the CBR task was fixed. That is, the collection of retrieval perspectives, their combination, the collection of reuse criteria, and the storage of solved cases were pre-designed and the user didn't participate in the reasoning process. Moreover, the *retain* subtask was not present because it is mainly a subtask that requires an interaction with the user.

Now, in the current version of *SaxEx* we have improved the CBR method by incorporating the user in the reasoning process (Arcos & López de Mántaras 2001). This new capability allows users to influence the solutions proposed by *SaxEx* in order to satisfy their interests or personal style. The user can interact with *SaxEx* in the four main CBR subtasks. This new functionality requires that the use and combination of the two basic mechanisms—perspectives and preferences— in the Retrieve and Reuse subtasks must be parameterizable and dynamically modifiable.

Retrieval perspectives

Retrieval perspectives are built by the *identify* subtask and can be used either by the *search* or the *select* subtask. Perspectives used by the *search* subtask will act as filters. Perspectives used by the *select* subtask will act only as a preference. Retrieval perspectives are built based on user requirements and background musical knowledge. Retrieval perspectives provide partial information about the relevance of a given musical aspect. After these perspectives are established, they have to be combined in a specific way according to the importance (preference) that they have.

Retrieval perspectives are of two different types: based on the affective intention that the user wants to obtain in the output expressive sound or based on musical knowledge.

1) *Affective labels* are used to determine the following declarative bias: we are interested in notes with affective labels similar to the affective labels required in the current problem by the user.

As an example, let us assume that we declare we are interested in forcing *SaxEx* to generate a calm and very tender performance of the problem phrase. Based on this bias, *SaxEx* will build a perspective specifying as relevant to the current problem the notes from cases that belong first to "calm and very tender" affective regions (most preferred), or "calm and tender" affective regions, or "very calm and very tender" affective regions (both less preferred).

When this perspective is used in the *Search* subtask, *SaxEx* will search in the memory of cases for notes that satisfy this criterion. When this perspective is used in the *Select* subtask, *SaxEx* will rank the previously retrieved cases using this criterion.

2) *Musical knowledge* gives three sets of declarative retrieval biases: first, biases based on Narmour's implication/realization model; second, biases based on Lerdahl and Jackendoff's generative theory; and third, biases based on Jazz theory and general music knowledge.

Regarding Narmour's implication/realization model, *SaxEx* incorporates the following three perspectives:

- The "*role in IR structure*" criterion determines as relevant the role that a given note plays in an implication/realization structure. That is, the kind of IR structure it belongs to and its position (`first-note`, `inner-note`, or `last-note`). Examples of IR basic structures are the `P` process (a melodic pattern describing a sequence of at least three notes with similar intervals and the same ascending or descending registral direction) and the `ID` process (a sequence of at least three notes with the same intervals and different registral directions), among others. For instance, this retrieval perspective can specify biases such as "look for notes that are the `first-note` of a `P` process".

- The "*Melodic Direction*" criterion determines as relevant the kind of melodic direction in an implication/realization structure: `ascendant`, `descendant`, or `duplication`. This criterion is used for adding a preference among notes with the same IR role.

- The "*Durational Cumulation*" criterion determines as relevant the presence—in a IR structure—of a note in the last position with a duration significally higher than the others. This characteristic emphasizes the end of a IR structure. This criterion is used—as the previous—for adding a preference among notes with the same IR role and same melodic direction.

Regarding Lerdahl and Jackendoff's GTTM theory, *SaxEx* incorporates the following three perspectives:

- The "*Metrical Strength*" criterion determines as relevant the importance of a note with respect to the metrical structure of the piece. The metrical structure assigns a weight to each note according to the beat in which it is played. That is, the metrical weight of notes played in strong beats are higher than the metrical weight of notes played in weak beats. For instance, the metrical strength bias determines as similar the notes played at the beginning of subphrases since the metrical weight is the same.

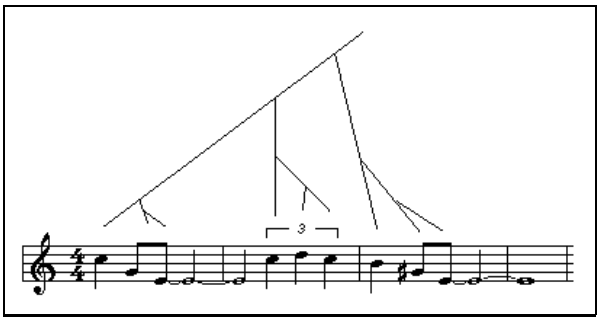

Figure 4: Example of a Time-Span Tree for the beginning of the 'All of me' ballad.

- The "*role in the Time-Span Reduction Tree*" criterion determines as relevant the structural importance of a given note according to the role that the note plays in the analysis Time-Span Reduction Tree.

  Time-Span Reduction Trees are built bottom-up and hold two components: a segmentation into hierarchically organized rhythmic units and a binary tree that represents the relative structural importance of the notes within those units. There are two kinds of nodes in the tree: left-elaboration nodes and right-elaboration nodes.

  Since the Time-Span Reduction Tree is a tree with high depth, we are only taking into account the two last levels. That is, given a note this perspective focuses on the kind of leaf the note belongs (left or right leaf) and on the kind of node the leaf belongs (left-elaboration or right-elaboration node).

  For instance, in the 'All of me' ballad (see Figure 4) the first quarter note of the second bar (`C`) belongs to a left leaf in a right-elaboration node because the following two notes (`D` and `C`) elaborate the first note. In turn, these two notes belong to a left-elaboration (sub)node because second note (`D`) elaborates the third (`C`).

- The "*role in the Prolongational Reduction Tree*" criterion determines as relevant the structural importance of a given note according to the role that the note plays in the Prolongational Reduction Tree. Prolongational Reduction Trees are binary trees built top-down and represent the hierarchical patterns of tension and relaxation among groups of notes. There are two basic kinds of nodes in the tree (tensing nodes and relaxing nodes) with three modes

of branch chaining: *strong prolongation* in which events repeat maintaining sonority (e.g., notes of the same chord); *weak prolongation* in which events repeat in an altered form (e.g., from `I` chord to `I6` chord); and *jump* in which two completely different events are connected (e.g., from `I` chord to `V` chord).

As in the previous perspective we are only taking into account the two last levels of the tree. That is, given a note this perspective focuses on the kind of leaf the note belongs (left or right leaf), on the kind of node the leaf belongs (tensing or relaxing node), and the kind of connection of the node (strong, weak, or jump).

Finally, regarding perspectives based on jazz theory and general music knowledge, *SaxEx* incorporates the following two:

- The "*Harmonic Stability*" criterion determines as relevant the role of a given note according to the underlying harmony. Since *SaxEx* is focused on generating expressive music in the context of jazz ballads, the general harmonic theory has been specialized taking harmonic concepts from jazz theory. The Harmonic Stability criterion takes into account in the following two aspects: the position of the note within its underlying chord (e.g., first, third, seventh, ...); and the role of the note in the chord progression it belongs.
- The "*Note Duration*" criterion determines as relevant the duration of a note. That is, given a specific situation, the set of expressive transformations applied to a note will differ depending on whether the note has a long or a short duration.

The use of fuzzy techniques in the Reuse step

Having modeled the linguistic values of the expressive parameters by means of fuzzy sets, allows us to apply a fuzzy combination operator to these values of the retrieved notes in the reuse step. The following example describes this combination operation.

Let us assume that the system has retrieved two similar notes whose fuzzy values for the rubato are, respectively, 72 and 190, The system first computes the maximum degree of membership of each one of these two values with respect to the five linguistic values characterizing the *rubato* shown in figure 2. The maximum membership value of 72 corresponds to the fuzzy value *low* and is 0.90 (see figure 5) and that of 190 correponds to *medium* and is 0.70. Next, it computes a combined fuzzy membership function, based on these two values. This combination consists on the fuzzy disjunction of

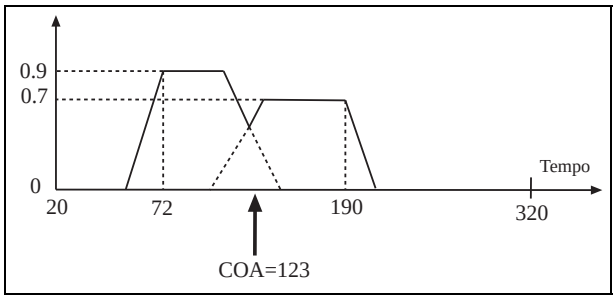

Figure 5: Fuzzy combination and defuzzification of rubato value.

the fuzzy membership functions *low* and *medium* truncated, respectively, by the 0.90 and 0.70 membership degrees. That is:

$$Max(min(0.90, f_{low}), min(0.70, f_{medium}))$$

The result is shown in figure 5. Finally *defuzzifies* this result by computing the COA (Center of Area) of the combined function (Klir & Yuan 1995). The defuzzification step gives the precise value for the tempo to be applied to the initially inexpressive note, in this example the obtained result is 123. An analogous process is applied to the other expressive parameters. The advantage of such fuzzy combination is that the resulting expression takes into account the contribution of all the retrieved similar notes whereas with criteria such as *minority rule*, *majority rule* etc. this is not the case. For example, if the system retrieves three notes from the expressive examples, and two of them had been played with low rubato and the third with medium rubato, the majority rule dictates that the inexpressive note should be played with *low rubato*. This conclusion is mapped into an a priori fixed value that is lower than the average rubato of the inexpressive input piece. It is worth noticing that each time the system concludes *low rubato* for several inexpressive notes, these note will be played with the same rubato even if the retrieved similar notes were different (*very low* would be mapped into a value much lower than the average rubato, *high* would be mapped into a value higher than the average and *very high* into a value much higher than the average and the same procedure applies to the other expressive parameters such as dynamics, vibrato and legato). With the fuzzy extension, the system is capable of increasing the variety of its performances because, after defuzzification, the final value for each expressive parameter is computed and this computation does not depend only on the linguistic value (low, etc.) of the retrieved similar notes but also on the membership degree of the actual numerical values that

are used to truncate the membership functions as explained above, therefore the final value will not be the same unless, of course, the precedent retrieved notes is actually the same note.

The system is connected to the SMS (4) software for sound analysis and synthesis based on spectral modeling as pre and post processor. This allows to actually listen to the obtained results. These results clearly show that a computer system can play expressively. In our experiments, we have used Real Book jazz ballads.

Related work

Previous work on the analysis and synthesis of musical expression has addressed the study of at most two expressive parameters such as rubato and vibrato (Clynes 1995; Desain & Honing 1995; Honing 1995), rubato and dynamics (Widmer 1996; Bresin 1998) or rubato and articulation (Johnson 1992). Concerning instrument modeling, the work of Dannenberg and Derenyi (Dannenberg & Derenyi 1998) is an important step towards high-quality synthesis of wind instrument performances. Other work such as in (De Poli, Rodà, & Vidolin 1998; Friberg *et al.* 1998) has focalized on the study of how musicians expressive intentions influence performers. To the best of our knowledge, the only previous works using learning techniques to generate expressive performances are those of Widmer (Widmer 1996), who uses explanation-based techniques to learn rules for dynamics and rubato using a MIDI keyboard, and Bressin (Bresin 1998), who trains an artificial neural network to simulate a human pianist also using MIDI. In our work we deal with five expressive parameters in the context of a very expressive non-MIDI instrument (tenor sax). Furthermore, ours was the first attempt to use Case-based Reasoning techniques. The use of CBR techniques was also done later by (Suzuki, Tokunaga, & Tanaka 1999) but dealing only with rubato and dynamics for MIDI instruments.

Conclusions

We have briefly described a new improved version of our *SaxEx* system. The added interactivity improves the usability of the system and the use of fuzzy techniques in the reuse step increases the performance variety of the system. Some ideas for further work include further experimentation with a larger set of tunes as well as allowing the system to add ornamental notes and not to play some of the notes, that is moving a small step towards adding improvising capabilities to the system.

Acknowledgements

The research reported in this paper is partly supported by the ESPRIT LTR 25500-COMRIS *Co-Habited Mixed-Reality Information Spaces* project. We also acknowledge the support of ROLAND Electronics de España S.A. to our AI & Music project.

References

Arcos, J. L., and López de Mántaras, R. 2001. An interactive case-based reasoning approach for generating expressive music. *Journal of Applied Intelligence*. In press.

Arcos, J. L.; López de Mántaras, 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.

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

Clynes, M. 1995. Microstructural musical linguistics: composers' pulses are liked most by the best musicians. *Cognition* 55:269–310.

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

De Poli, G.; Rodà, A.; and Vidolin, A. 1998. 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.

Desain, P., and Honing, H. 1995. Computational models of beat induction: the rule-based approach. In *Proceedings of IJCAI'95 Workshop on AI and Music*, 1–10.

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.

Honing, H. 1995. The vibrato problem, comparing two solutions. *Computer Music Journal* 19 (3):32–49.

Johnson, M. 1992. An expert system for the articulation of Bach fugue melodies. In Baggi, D., ed., *Readings in Computer-Generated Music*. IEEE Computes Society Press. 41–51.

Klir, G., and Yuan, B. 1995. *Fuzzy Sets and Fuzzy Logic*. Prentice Hall.

Lerdahl, F., and Jackendoff, R. 1993. An overview of hierarchical structure in music. In Schwanaver, S. M., and Levitt, D. A., eds., *Machine Models of Music*.

The MIT Press. 289–312. Reproduced from Music Perception.

Narmour, E. 1990. *The Analysis and cognition of basic melodic structures : the implication-realization model.* University of Chicago Press.

Serra, X.; Bonada, J.; Herrera, P.; and Loureiro, R. 1997. Integrating complementary spectral methods in the design of a musical synthesizer. In *Proceedings of the ICMC'97*, 152–159. San Francisco: International Computer Music Asociation.

Suzuki, T.; tokunaga, T.; and Tanaka, H. 1999. A case-based approach to the generation of musical expression. In *Proceedings of IJCAI'99.*

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