Affect-Driven Generation of Expressive Musical Performances

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

Abstract

We present an extension of an existing system, called *SaxEx*, capable of generating expressive musical performances based on Case-Based Reasoning (CBR) techniques. In our previous work, *SaxEx* did not take into account the possibility of using affective labels to guide the CBR task. This paper discusses the introduction of such affective knowledge to improve the retrieval capabilities of the system. Three affective dimensions are considered—tender-aggressive, sad-joyful, and calm-restless—that allow the user to declaratively instruct the system to perform according to any combination of five qualitative values along these three dimensions.

Introduction

The problem with the automatic generation of expressive musical performances is that human performers use musical knowledge that is not explicitly noted in musical scores. Moreover, this knowledge is difficult to verbalize and therefore AI approaches based on declatarive knowledge representations have serious limitations. An alternative approach is that of directly using the knowledge implicit in examples from recordings of human performances.

Previous work has addressed this problem mainly by means of MIDI instruments with the unavoidable limitations regarding expressivity. Our goal is the generation of expressive musical performances in the context of instruments with rich and continuous expressive capabilities (like wind instruments). We have developed *SaxEx*(Arcos, López de Mántaras, & Serra 1998), a case-based reasoning system for generating expressive performances of melodies based on examples of human performances. Case-Based Reasoning (CBR) (Aamodt & Plaza 1994) is an approach to problem solving and learning where new problems are solved using similar previously solved problems. The two basic mechanisms used by CBR are (i) the retrieval of solved problems (also called precedents or cases) using some similarity criteria and (ii) the adaptation of the solutions applied in the precedents to the new problem. Case-based reasoning techniques are appropriate on problems where many examples of solved problems can be obtained—like in our case where multiple examples can be easily obtained from recordings of human performances.

Sound analysis and synthesis techniques based on spectrum models like Spectral Modeling Synthesis (SMS) (Serra 1997; Serra *et al.* 1997) are useful for the extraction of high level parameters from real sounds, their transformation and the synthesis of a modified version of the original. *SaxEx* uses SMS in order to extract basic information related to several expressive parameters such as dynamics, rubato, vibrato, and articulation. The SMS synthesis procedure allows *SaxEx* the generation of new expressive interpretations (new sound files).

SaxEx incorporates background musical knowledge based on Narmour's implication/realization (IR) model (Narmour 1990) and Lerdahl and Jackendoff's generative theory of tonal music (GTTM) (Lerdahl & Jackendoff 1993). These theories of musical perception and musical understanding are the basis of the computational model of musical knowledge of the system.

SaxEx is implemented in *Noos* (Arcos & Plaza 1997; 1996), a reflective object-centered representation language designed to support knowledge modeling of problem solving and learning.

In our previous work on *SaxEx* (Arcos, López de Mántaras, & Serra 1998) we had not taken into account the possibility of exploiting the affective aspects of music to guide the retrieval step of the CBR process. In this paper, we discuss the introduction of labels of affective nature (such as "calm", "tender", "aggressive", etc.) as a declarative bias in the Identify and Search subtasks of the Retrieval task (see Figure 2).

Background

In this section, we briefly present some of the elements underlying *SaxEx* which are necessary to understand the system.

SMS

Sound analysis and synthesis techniques based on spectrum models like Spectral Modeling and Synthesis (SMS) are useful for the extraction of high level parameters from real sound files, their transformation, and the synthesis of a modified version of these sound files. *SaxEx* uses SMS in order to extract basic information related to several expressive parameters such as dynamics, rubato, vibrato, and articulation. The SMS synthesis procedure allows the generation of expressive reinterpretations by appropriately transforming an inexpressive sound file.

The SMS approach to spectral analysis is based on a decomposing a sound into sinusoids plus a spectral residual. From the sinusoidal plus the residual representation we can extract high level attributes such as attack and release times, formant structure, vibrato, and average pitch and amplitude, when the sound is a note or a monophonic phrase of an instrument. These attributes can be modified and added back to the spectral representation without loss of sound quality.

This sound analysis and synthesis system is ideal as a preprocessor, giving to *SaxEx* high level musical parameters, and as a post-processor, adding the transformations specified by the case-based reasoning system to the inexpressive original sound.

Noos

SaxEx is implemented in *Noos* (Arcos & Plaza 1997; 1996), a reflective object-centered representation language designed to support knowledge modeling of problem solving and learning. Modeling a problem in Noos requires the specification of three different types of knowledge: domain knowledge, problem solving knowledge, and metalevel knowledge.

Domain knowledge specifies a set of concepts, a set of relations among concepts, and problem data that are relevant for an application. Concepts and relations define the domain ontology of an application. For instance, the domain ontology of *SaxEx* is composed by concepts such as notes, chords, analysis structures, and expressive parameters. Problem data, described using the domain ontology, define specific situations (specific problems) that have to be solved. For instance, specific inexpressive musical phrases to be transformed into expressive ones.

Problem solving knowledge specifies the set of tasks to be solved in an application. For instance, the main task of *SaxEx* is to infer a sequence of expressive transformations for a given musical phrase. Methods model different ways of solving tasks. Methods can be elementary or can be decomposed into subtasks. These new (sub)tasks may be achieved by other methods. A method defines an execution order over subtasks and an specific combination of the results of the subtasks in order to solve the task it performs. For a given task, there can be multiple alternative methods that may solve the task in different situations. This recursive decomposition of a task into subtasks by means of a method is called task/method decomposition.

The metalevel of Noos incorporates, among other types of (meta-)knowledge, Preferences, used by *SaxEx* to rank cases, and Perspectives, used in the retrieval task. *Preferences* model decision making criteria about sets of alternatives present in domain knowledge and problem solving knowledge. For instance, preference knowledge can be used to model criteria for ranking some precedent cases over other precedent cases for a task in a specific situation. *Perspectives* (Arcos & López de Mántaras 1997), constitute a mechanism to describe declarative biases for case retrieval in structured and complex representations of cases. They provide a flexible and dynamical retrieval mechanism and are used by *SaxEx* to make decisions about the relevant aspects of a problem. *SaxEx* incorporates two types of declarative biases in the perspectives. On the one hand, metalevel knowledge to assess similarities among scores using the analysis structures built upon the IR and GTTM musical models. On the other hand, (metalevel) knowledge to detect affective intention in performances and to assess similarities among them.

Once a problem is solved, Noos automatically memorizes (stores and indexes) that problem. The collection of problems that a system has solved is called the episodic memory of Noos. The problems solved by Noos are accessible and retrievable. This introspection capability of Noos is the basic building block for integrating learning, and specifically CBR, into Noos.

Musical Models

SaxEx incorporates two theories of musical perception and musical understanding that constitute the background musical knowledge of the system: Narmour's implication/realization (IR) model (Narmour 1990) and Lerdahl and Jackendoff's generative theory of tonal music (GTTM) (Lerdahl & Jackendoff 1993).

Narmour's implication/realization model proposes a theory of cognition of melodies based on eight basic structures. These structures characterize patterns of melodic implications that constitute the basic units of the listener perception. Other parameters such as metric, duration, and rhythmic patterns emphasize or inhibit the perception of these melodic implications. The use of the IR model provides a musical analysis based on the structure of the melodic surface.

Examples of IR basic structures are the `P` process

(a melodic pattern describing a sequence of at least three notes with similar intervals and same ascending or descending registral direction) and the `ID` process (a sequence of at least three notes with same intervals and different registral directions).

On the other hand, Lerdahl and Jackendoff's generative theory of tonal music (GTTM) offers a complementary approach to understanding melodies based on a hierarchical structure of musical cognition. GTTM proposes four types of hierarchical structures associated with a piece.

Examples of GTTM analysis structures are `prolongational-reduction`—a hierarchical structure describing tension-relaxation relationships among groups of notes—and `time-span-reduction`—a hierarchical structure describing the relative structural importance of notes within the heard rhythmic units of a phrase. Both are tree structures that are directly represented in *Noos* because of the tree-data representation capabilities of the language.

Our goal in using both, IR and GTTM models, is to take advantage of combining the IR analysis of melodic surface with the GTTM structural analysis of the melody. These are two complementary views of melodies that influence the execution of a performance.

Affective Descriptions

The use of affective adjectives to characterize different aspects of musical performance has a long tradition. In baroque music, each piece or movement had an "affect" associated with it that was intended to have "the soul exert control over the body and fill it with passions that were strongly expressed" (Cyr 1992). Many lists of affective adjectives have been proposed by different theorists, e.g., Castiglioni, Galilei, Rousseau, Quantz, Mattheson, or more recently Cooke (Cooke 1959). The main problems with the use of affective adjectives for musical purposes are that their meaning might vary over time, they are highly subjective and usually redundant or overlapping, and it is very difficult to assess what are the relationships between different labels. In order to avoid these problems, we decided not to use isolated adjectives, but rather to rank affective intentions along three orthogonal dimensions: tender-aggressive, sad-joyful, and calm-restless. To come out with these dimensions, we drew inspiration from the experiments conducted by (Canazza & Orio 1997), where sonological analysis of jazz recordings and the listeners' perception of them showed that a broad set of affective adjectives (16 in the experiments reported there) could be clustered into a few main dimensions. In addition, these dimensions relate to semantic notions, such as activity, tension versus relaxation, brightness, etc., although a one-to-one correlation cannot be neatly established.

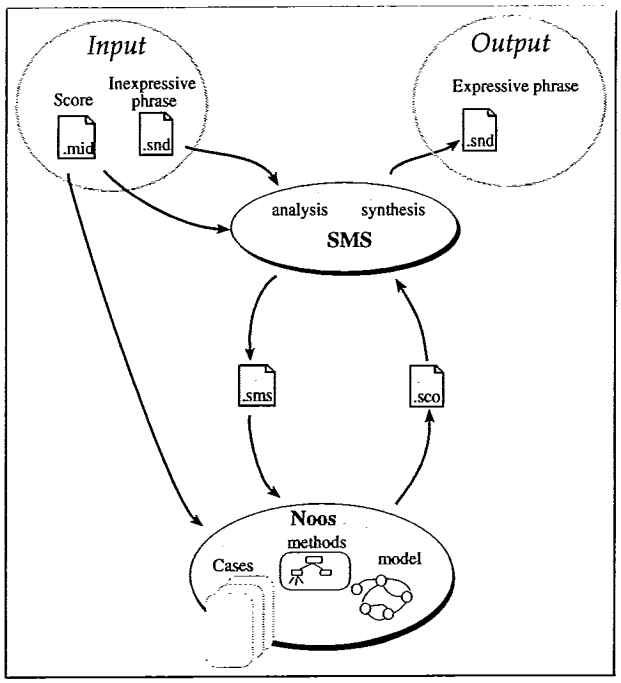

Figure 1: General view of *SaxEx* blocks.

SaxEx

An input for *SaxEx* is a musical phrase described by means of its musical score (a MIDI file), a sound, and specific qualitative labels along affective dimensions. Affective labels can be partially specified, i.e. the user does not have to provide labels for every dimension.

The score contains the melodic and the harmonic information of the musical phrase. The sound contains the recording of an inexpressive interpretation of the musical phrase played by a musician. Values for affective dimensions will guide the search in the memory of cases. The output of the system is a new sound file, obtained by transformations of the original sound, and containing an expressive performance of the same phrase.

Solving a problem in *SaxEx* involves three phases: the analysis phase, the reasoning phase, and the synthesis phase (see Figure 1). Analysis and synthesis phases are implemented using SMS sound analysis and synthesis techniques. The reasoning phase is performed using CBR techniques and implemented in Noos and is the main focus of this paper.

The development of *SaxEx* involved the elaboration of two main model: the domain model and the problem-solving models. The domain model contains the concepts and structures relevant for representing

musical knowledge. The problem-solving model consists mainly of a problem solving method for inferring a sequence of expressive transformations for a given musical phrase. Problems to be solved by *SaxEx* are represented as complex structured cases embodying three different kinds of musical knowledge: (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 structures and GTTM's time-span reduction nodes, and (3) concepts related to the performance of musical phrases. Affective labels belong to this third type.

Modeling Musical Knowledge

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 inferred using Narmour's 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). 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 *events* (extracted using the SMS sound analysis capabilities) and a sequence of *affective regions*, as explained below.

There is an *event* for each note within the phrase embodying information about expressive parameters applied to that note. Specifically, an event holds information about dynamics, rubato, vibrato, articulation, and attack. These expressive parameters are described using qualitative labels as follows:

- Changes in dynamics are described relative to the average loudness of the phrase by means of a set of five ordered labels. The middle label represents average loudness and lower and upper labels represent respectively increasing or decreasing degrees of loudness.
- Changes in rubato are described relative to the average tempo also by means of a set of five ordered labels. Analogously to dynamics, qualitative labels about rubato cover the range from a strong accelerando to a strong ritardando.
- The vibrato level is described using two parameters: frequency and amplitude. Both parameters are described using five qualitative labels from no-vibrato to highest-vibrato.
- The articulation between notes is described using again a set of five ordered labels covering the range from legato to staccato.
- Finally, *SaxEx* considers two possibilities regarding note attack: (1) reaching the pitch of a note starting from a lower pitch, and (2) increasing the noise component of the sound. These two possibilities were chosen because they are characteristic of saxophone playing but additional possibilities can be introduced without altering the system.

Affective regions group (sub)-sequences of events 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 qualitative labels as follows. The middle label represents no predominance (e.g. neither tender nor aggressive), lower and upper labels represent, respectively predominance in one direction (e.g. absolutely calm is described with the lowest label).

The SaxEx 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 problem. 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. Affective knowledge is the basis for guiding the CBR problem solver.

For each note in the phrase, the following subtask decomposition (Figure 2) is performed by the case-based 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 notes—the cases—most similar to

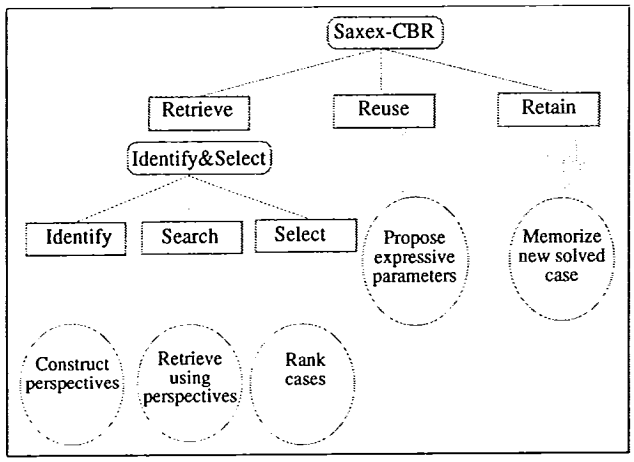

Figure 2: Task decomposition of the *SaxEx* CBR method.

the current one—the problem. This task is decomposed in three subtasks:

- *Identify*: its goal is to build retrieval perspectives using the affective values specified by the user and the musical background knowledge integrated in the system. Affective labels are used to determine a first declarative retrieval bias: we are interested in notes with affective labels close to affective labels required in the current problem. Musical knowledge gives two possible declarative retrieval biases: a first bias based on Narmour's implication/realization model, and a second bias based on Lerdahl and Jackendoff's generative theory. These perspectives guide the retrieval process by focusing it on the most relevant aspects of the current problem.
- *Search*: its goal is to search cases in the case memory using Noos retrieval methods and some previously constructed Perspective(s). As an example, let us assume that, by means of a Perspective, we declare that we are interested in notes belonging to calm and very tender affective regions. Then, the Search subtask will search for notes in the expressive performances that, following this criterion, belong to either 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).
- *Select*: its goal is to rank the retrieved cases using Noos preference methods. 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 more similar notes previously selected, a set of expressive transformations to be applied to the current note. The first criterion used is to adapt the transformations of the most similar note. When several notes are considered equally similar, the transformations are selected according to the majority rule. Finally, in case of a tie, one of them is selected randomly.

- *Retain*: the incorporation of the new solved problem to the memory of cases is performed automatically in Noos. All solved problems will be available for the reasoning process in future problems.

Experiments

We are studying the issue of musical expression in the context of tenor saxophone interpretations. We have done several recordings of a tenor sax performer playing several Jazz standard ballads ('All of me', 'Autumn leaves', 'Misty', and 'My one and only love') with different degrees of expressiveness, including an inexpressive interpretation of each piece. These recordings did not take into account the expression of affects. They were analyzed, using the SMS spectral modeling techniques, to extract basic information related to the expressive parameters.

Two sets of experiments had been conducted previously combining the different Jazz ballads recorded. The first set of experiments consisted in using examples of three different expressive performances of twenty note phrases of a piece in order to generate an expressive reinterpretation of another inexpressive phrase of the same piece. This group of experiments revealed that *SaxEx* identifies clearly the relevant cases even though the new phrase introduces small variations with respect to the phrases existing in the memory of precedent cases.

The second set of experiments was intended to use examples of expressive performances of some pieces in order to generate expressive reinterpretations of different inexpressive pieces. More concretely, we worked with three different expressive performances of a piece having about fifty notes in order to generate expressive reinterpretations of about thirty-note inexpressive phrases of a different piece. This second group of experiments revealed that the use of perspectives in the retrieval step allows to identify situations such as long notes, ascending or descending melodic lines, etc. Such situations are also usually identified by a human performer.

We are now in the process of conducting a new set of experiments taking into account the expression of affects. In the previous *SaxEx* experiments the performer was only required to play several versions of the same ballad introducing different expressive resources but without being forced to give any emotional intention. The categorization according to affective labels

has been done *a posteriori* by us, after a careful analysis of the recordings. Each musical phrase (of eight bars) has been divided into affective regions and labeled according to the averaged labeling of several listeners. This division allows us to track the evolution of the affective intention that the musician introduces in a phrase.

The idea underlying these experiments is to be able to ask the system to perform expressively any musical phrase according to a specific affective label or combination of them.

Prospect and future work

The integration of affective labels allows to improve the performance of *SaxEx* in several ways. From the perspective of users, a more friendly interaction with the system is possible. On the one hand, users can work in a more intuitive way, without needing formal musical knowledge. On the other hand, it is possible to generate a wider range of expressive intentions by combining affective labels in multiple ways.

Affective knowledge also enhances the reasoning of the system. In particular, affective labels constitute an additional criterion to discriminate among the several candidate performances of a same phrase.

The experiments we are currently carrying on were designed using already existing recordings that had been made without the purpose of communicating affects. As a next step, we plan to incorporate into the system additional recordings in which the performer will be required to play according to affective labels. This will allows us to obtain a richer case memory and to better assess how the affect that the musician intends to communicate is perceived by the listeners. This will also ease the task of relating affective labels with expressive parameters—done by hand in the current experiments. This analysis could be used in the future to have *SaxEx* learn automatically associations of labels and the use of expressive parameters for situations appearing recurrently in the cases. Finally, it would be interesting to discriminate situations where expressive variations are used because of the logical structure of the score, as opposed to situations where these variations come from the affective intentions of the musician.

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

Aamodt, A., and Plaza, E. 1994. Case-based reasoning: Foundational issues, methodological variations, and system approaches. *Artificial Intelligence Communications* 7(1):39–59.

Arcos, J. L., and López de Mántaras, R. 1997. Perspectives: a declarative bias mechanism for case retrieval. In Leake, D., and Plaza, E., eds., *Case-Based Reasoning. Research and Development*, number 1266 in Lecture Notes in Artificial Intelligence. Springer-Verlag. 279–290.

Arcos, J. L., and Plaza, E. 1996. Inference and reflection in the object-centered representation language Noos. *Journal of Future Generation Computer Systems* 12:173–188.

Arcos, J. L., and Plaza, E. 1997. Noos: an integrated framework for problem solving and learning. In *Knowledge Engineering: Methods and Languages.*

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.* In Press.

Canazza, S., and Orio, N. 1997. How are the players perceived by listeners: analysis of "how high the moon" theme. In *International workshop Kansei Technology of Emotion (AIMI'97).*

Cooke, D. 1959. *The Language of Music.* New York: Oxford University Press.

Cyr, M. 1992. *Performing Baroque Music.* Portland, Oregon: Amadeus Press.

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.

Serra, X. 1997. Musical sound modeling with sinusoids plus noise. In Roads, C.; Pope, S. T.; Picialli, A.; and De Poli, G., eds., *Musical Signal Processing.* Swets and Zeitlinger Publishers. 91–122.