Imitating Human Performances to Automatically Generate Expressive Jazz Ballads

Dolores Cañamero; Josep Lluis Arcos; Ramon López de Mántaras
Artificial Intelligence Research Institute (IIIA)
Spanish Council for Scientific Research (CSIC)
Campus UAB, E–08193 Bellaterra, Barcelona, Spain
lola, arcos, mantaras@iiia.csic.es

Abstract

One of the main problems 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 tacit, difficult to verbalize, and therefore it must be acquired through a process of observation, imitation, and experimentation. For this reason, AI approaches based on declarative knowledge representations have serious limitations. An alternative approach is that of directly using the knowledge implicit in examples from recordings of human performances. In this paper, we describe a case-based reasoning system that generates expressive musical performances based on examples of expressive human performances.

1 Introduction

One of the major difficulties in the automatic generation of music is to endow the resulting piece with the expressivity that characterizes human performances. Following musical rules, whatever sophisticated and complete they are, is not enough to achieve this expressivity, and indeed music generated in this way usually sounds monotonous and mechanical. The main problem here is to grasp the performer's "personal touch", the knowledge brough about when *interpreting* a score and that is absent from it. This knowledge concerns not only "technical" features (use of musical resources) but also the affective aspects implicit in music. A large part of this knowledge is tacit and therefore very difficult to generalize and verbalize, although it is not inaccessible. Humans acquire it through a long process of observation, imitation, and experimentation (Dowling and Harwood 1986). For this reason, AI approaches based on declarative knowledge representations have serious limitations. An alternative approach, much closer to the observation-imitation-experimentation process observed in humans, is that of directly using the knowledge implicit in examples from recordings of human performances.

In order to achieve that we have developed SaxEx (Arcos et al. 1998b), a case-based reasoning (CBR) system for generating expressive performances of melodies based on examples of human performances (for the moment we have limited ourselves to jazz ballads). CBR is an approach to problem solving and learning where problems are solved using previously solved problems which are considered to be similar enough according to some criteria. The two basic mechanisms used in CBR are (1) the retrieval of solved problems (also called precedents or cases) using some similarity measure and (2) the adaptation of the solutions applied in the precedents to the new problem. CBR is appropriate for problems where (a) many examples of solved problems can be obtained—like in our case where multiple examples can be easily obtained from recordings of human performances; and (b) a large part of the knowledge involved in the solution of problems is tacit, difficult to verbalize and generalize.

SaxEx allows the user to control the degree and type of expressivity desired in the output by means of qualitative affective labels along three orthogonal affective dimensions (tender-aggressive, sad-joyful, and calm-restless). This enables the user to ask the system to perform a phrase according to a specific affective label or a combination of them.

2 The SaXex System

An input for SaxEx is a score (a MIDI file), which provides the melodic and the harmonic information of a musical phrase, a sound file containing an inexpresive performance of the phrase, and specific qualitative labels along affective dimensions through which the user indicates the desired output. 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 through imitation of the expressive resources of expressive performances of different phrases, and containing an expressive performance of the same phrase.

Solving a problem in SaxEx, i.e. generating an ex-

pressive performance that meets the specifications of the user, involves three phases: analysis, reasoning, and synthesis. Analysis and synthesis are performed using SMS, a sound analysis and synthesis technique based on spectrum models. The reasoning phase is performed using CBR, and is the main focus of this paper.

SaxEx uses Spectral Modeling and Synthesis (SMS) (Serra 1997) to extract high level parameters from a real sound file (containing an inexpressive performance), to transform them according to the specifications provided by the CBR module, and to synthesize a modified version of the original sound file. SMS allows to extract basic information related to several expressive parameters—dynamics, rubato, vibrato, articulation, and attack. SMS is thus ideal to be used as both a preprocessor and a postprocessor module in conjunction with the CBR system.

The SaxEx problem solver is implemented in Noos (Arcos and Plaza 1997), a reflective object-centered representation language designed to support problem solving and learning, and in particular CBR. Modeling a problem in Noos requires the specification of three different types of knowledge: domain knowledge (concepts, relations among them, and problem data), problem solving knowledge (tasks that must be solved in order to solve the problem and methods to perform these tasks), and metalevel knowledge. The metalevel of Noos incorporates, among other types of (meta-)knowledge: (a) Preferences, decision-making criteria used by SaxEx to rank cases that provide alternative solutions to the problem at hand; and (b) Perspectives, a mechanism to describe declarative biases for case retrieval in structured and complex representations of cases, used in the retrieval task 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 background musical theories integrated into the system. On the other hand, (metalevel) knowledge to detect affective intention in performances and to assess similarities among them.

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. A score is represented by a melody, embodying a sequence of notes, and a harmony, embodying a sequence of chords. Each note and chord hold a set of features such as name, pitch (for notes only; e.g. C5, G4), position with respect to the beginning of the phrase, duration, a reference to the underlying harmony, and a reference to the next note/chord of the phrase.

(2) Concepts related to background theories of musical perception and understanding used to analyze the score and divide it into meaningful chunks. We use two complementary models: Narmour's (1990) implication/realization (IR) model to analize melodic surface, and Lerdahl and Jackendoff's (1993) generative theory of tonal music (GTTM) to analize the hierarchical structure of the melody. These are two complementary views of melodies that influence the execution of a performance.

(3) Information concerning the performance of the musical phrases contained in the examples of the case base, and which is of two kinds: concepts related to the execution of expressive parameters, represented by a sequence of events, and concepts related with the affective character of the performance represented by a sequence of affective regions. There is an event for each note within the phrase embodying information about expressive parameters applied to that note—dynamics, rubato, vibrato, articulation, and attack—described using qualitative labels. Affective regions group (sub)-sequences of events with common affective expressivity. Specifically, an affective region holds information describing the following orthogonal affective dimensions (see Arcos et al. 1998a for details) by means of five qualitative labels for each dimension: tender-aggressive, sad-joyful, and calm-restless. This division into affective regions allows us to track the evolution of the affective intention that the musician introduces in a phrase. In addition, these affective dimensions relate to semantic notions, such as activity, tension versus relaxation, brightness, etc., although a one-to-one correlation cannot be neatly established.

3 Generating Expressive Performances

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—called episodic memory—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 is performed by the CBR problem solving method:

(1) Retrieve: The selection, from the memory of cases (pieces played expressively), of the set of notes—the cases—most similar to the current one—the problem. This task is decomposed in three subtasks:

1.1) Identify or 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. Perspectives guide the retrieval process by focusing it on the most relevant aspects of the current problem.

1.2) Search 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).

1.3) Select: the ranking of retrieved cases using preference methods, which use criteria such as similarity in duration of notes, harmonic stability, or melodic directions.

(2) 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.

(3) Retain: the incorporation (indexing and storage) 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.

Different sets of experiments have been performed, both without and with the use of affective labels, using several recordings of a tenor sax performer playing standard jazz ballads ('All of me', 'Autumn leaves', 'Misty', and 'My one and only love') with different degrees of expressiveness, and are described elsewhere (Arcos et al. 1998a 1998b). For our purposes here, the main conclusions that we can draw from them are:

- SaxEx successfully identifies the relevant cases even though the phrase given as problem introduces small variations with respect to the phrases contained in the memory of cases.
- The use of declarative biases (perspectives) concerning the expressive parameters of the performance in the retrieval step of the CBR process allows to identify situations such as long notes, ascending or descending melodic lines, etc., which are also identified by human performers.
- The introduction of affective labels as additional declarative biases (perspectives) improves both the problem-solving process and the quality of the solutions. Among other things, the performances generated show a more coherent use of expressive resources, and as a consequence they are perceived by the listeners as being as natural human performances.

4 Discussion

Although the experiments tell us different things about the technical adequacy of our different design choices, the best way to evaluate the system is by listening to the output it produces. All listeners agree that the expressive performances generated by SaxEx sound very natural and human-like. This is due to several characteristics of our system.

First, the use of real sound examples allows SaxEx to grasp a large part of the tacit knowledge brought about when performing music through imitation of these examples, and to integrate it in the problems (inexpresive phrases) to be solved.

Second, the introduction of affective labels provides an intuitive interaction mechanism that allows users (experts and non-experts alike) to experiment with the system in a natural way and helps them to better understand how the different expressive resources are/can be used. In this sense, SaxEx can be used both as a pedagogical tool and as an experimentation tool for musicians.

Finally, the combination of musical models used by the system provide a means to experiment with highly especialized knowledge in a creative way. The deep musical knowledge contributed by these models allows to detect the relevant aspects of the examples and to combine them in sensible ways. This distinguishes SaxEx outputs from a mere (mechanical) copy process, and gives rise to one of the main features of real imitation processes: creative adaptation or "personal touch".

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 support from ROLAND Electronics de Espana S.A. to our AI Music project.

5 References

Arcos, J.L and Plaza, E. 1997. Noos: An Integrated Framework for Problem Solving and Learning. In Proceedings of the KEML'97 Workshop on Knowledge Engineering Methods and Languages. The Open University, Milton Keynes, UK, January 1997.

Arcos, J.L., Canamero, D., Lopez de Mantaras, R. 1998a. Affect-Driven Generation of Expressive Musical Performances. In Emotional and Intelligent: The Tangled Knot of Cognition, Papers from the 1998 AAAI Fall Symposium, Technical Report FS-98-03. Menlo Park, CA: The AAAI Press, 1-6.

Arcos, J.L., Lopez de Mantaras, R., and Serra, X. 1998b. SaxEx: A Case-Based Reasoning System for Generating Expressive Musical Performances. Journal of New Music Research (In press).

Dowling, W.J. and Hardwood, D.L. 1986. Music Cognition. London: Academic Press.

Lerdahl, F. and Jackendoff, R. 1993. An Overview of Hierarchical Structure in Music. In S.M. Schwanaver and D.A. Levitt (eds), Machine Models of Music. Cambridge, MA: The MIT Press, 289-312.

Narmour, E. 1990. The Analysis and Cognition of Basic Melodic Structures: The Implication-Realization Model. University of Chicago Press.

Serra, X. 1997. Musical Sound Modeling with Sinusoids plus Noise. In C. Roads, S.T. Pope, A. Picialli, and G. De Poli (eds), Musical Signal Processing. Swets and Zeitlinger Publishers, 91-122.