REPORT ON THE FIRST AIM CONFERENCE
SANKT AUGUSTIN, GERMANY
SEPTEMBER 1988

PETER DESAIN AND HENKJAN HONING

THE INSTITUTE of the GMD (Gesellschaft für Mathematik und Datenverarbeitung) in the neighbourhood of the small village Sankt Augustin (near Bonn, Germany) was the host of the first workshop on Artificial Intelligence and Music, organized by Christoph Lischka of the GMD. This institute comprises modern research facilities on different domains of computer science plus a real castle: Schloss Birlinghoven, all surrounded by hills and woodland. What the Schloss had in style it lacked in acoustics; the speakers at this conference had to insert a brief silence after each sentence for the reverberation to die out. This article will give some of our personal impressions of the workshop. They are grouped in research on environments for music composition (Cope, Potard, Rahn, Blevis, Modler), connectionist approaches to music (Lindster, Leman, Desain and Honing), expert systems for music (Camurri and Zaccaria, Widmer), and
the representation of music (Balaban, Böcker and Mahling). The latter could be found in all lectures more or less implicitly.

ENvironments for Music Composition

The last ten years show more and more attention to the design and realization of workstations for music. Availability of higher, more abstract programming languages (such as Lisp) and powerful microcomputers made the design of a system for music composition more realistic (e.g. Schottsteadt 1983, Rodet and Cointe 1984). New developments in computer science, especially in AI research, excited a lot of new ideas in designing systems for music.

David Cope from the University of California, Santa Cruz was the first speaker at the workshop with his talk about experiments in musical intelligence (EMI). The EMI composition system is centered around what he calls “nonlinear linguistic-based composition.” Cope started working on the project in 1984, and published a number of articles on EMI (e.g. Cope 1987). At the AIM workshop he presented an overview of the system, its current state, and several sound examples. EMI is meant to be an interactive partner for the composer. Starting with a motive, a full work of implied intervals and related parameters is created based on linguistic processes inherent in the original germ. David Cope prefers to compose with hierarchical structures instead of flat strings of notes (which he calls linear). The formalism used to describe the elaboration of an initial germ to the complete piece is “Augmented Transition Networks,” a way of specifying a context-sensitive grammar. Cope is convinced that music can be described in a linguistic way. He debugged the system by checking its linguistic behaviour—generating haikus—ensuring that the grammar is linguistically correct. Cope played a tape with some Mozart imitations made with EMI which were quite impressive but of course do not guarantee the usefulness of this linguistic approach in other musics. Cope pointed out that he should not be looked upon as someone doing AI research, but as a composer and musician inspired by the methods and ideas developed in AI.

Eli Blevis from Kingston presented a paper on the generalization of event sequences in music. It is part of the larger project done at Queens University called a Composition Analysis/generation Language for Music (CALM), a continuation of ideas developed in the CDS system (Hamel et al. 1987), in which the same team was involved. CALM uses ideas from new trends and styles in programming: functional, declarative, and typed languages in particular. It is written in the Nial programming language, combining ideas from Lisp and Prolog (Blevis et al. 1988). Blevis proposed
a method for generalizing sequences of note events. In musical terminology, one can think of the "generalize" operation as one that "gathers" musical objects considered to be similar into a class abstraction. The proposed generalization algorithm at the moment has some restrictions (e.g. length of event lists, number of parameters) but they will be loosened in future research. Generalization is now mostly used for analysis but is also very useful in the generation of music: CALM's initial goal is to support both directions.

John Rahn from the University of Washington (reading his notes from a napkin) explained his work on representation of musical objects. The aim of the research is to establish a high-level programming kernel that is independent of the kind of synthesis hardware available on current workstations. It should also facilitate the communication between different programs, like composition systems, music notation packages, MIDI sequencers, and so on. The kernel is written in a subset of any LISP that has lexical scope. Rahn's programming leans towards an object-oriented approach: the musical objects (notes) have high-level properties associated with them, such as their musical context and so forth. Besides the explanation of the kernel, some attention was given to possible workstation configurations (equipped with the "vaporware board") that could make use of current trends in parallel computing that seem suitable for sound synthesis.

Yves Potard from IRCAM—the Institute for Research and Coordination in Acoustics and Music directed by Pierre Boulez in Paris—gave a presentation on yet another proposition for a computer-nusic workstation. Without referring to the attempts made elsewhere, he claimed that an off-the-shelf SUN (UNIX machine) plus array processor (Mercury) would fit the needs of most people working in the field. Although the design goals presented focussed mainly on standardization and transportability, Potard assumed a wide use of MIDI-LISP and LeLisp (the French LISP Dialect without lexical scoping), on top of that an idiosyncratic object-oriented extension, and above that a home-made window system. However, transportability within IRCAM between MAC II and SUN is indeed assured in this way. A possible application running on this workstation could be pre-Formes. The question of what this all had to do with AI (IA for the French) was answered by the counter question: "What is AI really?"

Paul Modler from the Technical University of Berlin explained the ideas and the development of the CAMP system. In this workstation, using the ATARI, an integrated software environment is evisaged that will support real-time interaction and processing in FORMULA (extension of FORTH), has network capabilities, is open to future kinds of equipment (signal processors), and even has the possibility of programming in a LISP dialect. It remained unclear what was actually realized, but surely a good low-cost, networked computer music workstation would be of benefit in classrooms, for example.
EXPERT SYSTEMS FOR MUSIC

Perhaps the most popular field of AI research of the last twenty years is expert systems. Expert systems entail a considerable amount of knowledge that is domain-specific. This usually results in large systems with a lot of rules specific to a certain domain. More recent research introduces learning to make more efficient use of these large knowledge databases.

Gerhard Widmer presented a knowledge-based approach to machine learning of tonal music. A learning-apprentice system is under construction that deals with the problem of writing a second voice (the counterpoint) when the first (the cantus firmus) is given. In a learning-apprentice system the system has complete knowledge of the domain, so the system knows all the rules and constraints that are used in evaluating a solution (like the rule of forbidden parallel fifths). At first sight there is not much to be learned for a system with complete knowledge, but this is not the case. How to reach the goal in an intelligent way is not known, because static, normative knowledge does not entail the operative kind. Only a brute-force search of the gigantic solution space can be undertaken, which will finally result in a good, or the best solution. A human expert will have a completely different approach and will use all kinds of heuristic rules and shortcuts to reach a solution in a reasonable time. A learning-apprentice system deduces these kinds of rules from observing a human expert at work, only falling back on “deep” knowledge when everything else fails. Of course certain representations of deep knowledge will facilitate this kind of learning better than others. One of the aims of the research is characterizing this dependency.

A. Camurri and R. Zaccaria from the University of Genoa presented their Key-music system for knowledge representation of music and music composition. Examples were given of their prototype expert system named Jam. It has the expertise of a jazz musician in terms of functional harmony and basic melodic and rhythmic patterns. The system can “improvise” on a given harmonic track, according to user’s defined constraints and suggestions. The examples were refreshing after all the Harmony expert-systems applications that have been made in the last decade.

The main ideas used in designing the system are those of Petri-net theory and a frame-based system for knowledge representation. The latter are used for describing the general definition of musical timed processes, instances of which are the actual individual music processes. Output is in both MIDI and CMUSIC format. Further research will be done in allowing a user to retrieve information from the knowledge system. The system should also be able to manage concurrent plans and multiple solutions to a problem.
Connectionism proposes a new model that has some characteristics which traditional AI models were lacking. Connectionist models are characterized by their robustness and flexibility (Rumelhart and McClelland 1986). They consist of a large number of simple elements with their own activation level, connected with each other in a complex network. These cells excite or inhibit each other via their connections. After the network is given a starting state it can converge to an equilibrium.

Marc Leman from Ghent University concentrated in his presentation on the use of massive parallel processing and connectionist models for dealing with time-varying data such as musical information. After a review of the existing types of networks (Hopfield, Boltzmann, and so forth), he showed how a feedback of the processed output to the input of a network can provide the historic context that is needed when processing new incoming data. He elaborated also upon the implementation of this network in a Transputer configuration that made actual simulation and experimentation possible. Most of the experimental work was done with binary signals (ones and zeros). The implementation of these networks for the high-level, symbolic constructs introduced in the beginning of his talk still remains to be done.

Marc Leman gave Christiane Linster, who worked at the GMD institute in Sankt Augustin, a headstart in her talk about the analysis of rhythms with neural networks. She uses a standard three-layer network with backpropagation that was trained with simple metrical rhythms at its input and tree structures of the same pieces at its output. These would correspond to the groupings of beamed notes. After about fifty training sessions the network was able to produce a metrical structure of new, not yet presented, rhythms at the output. Each of the training sessions took a lot of processing time (around a day on a LISP machine).

Two ways of encoding the input note-durations were tried. The first was a categorical one, with a set of inputs per event (one for each allowed duration of a half, quarter, dotted-half note, and so on). The second one was a time-grid representation with an input activated if a note started on the corresponding point in time. The first representation was claimed to be superior. The representation of the tree at the output was highly symbolic. A specific output could acquire the value of a left bracket, a right bracket, or signify an element of the input rhythm itself. Although neural nets are very flexible, it seems that the performance depends largely on the chosen encoding of the input and output, and little is known about how to choose good representations.

The authors of this report presented a paper on the use of a connectionist model for quantization, that converges from nonmetrical performed data to
a metrical equilibrium. The model uses three kinds of cells. The basic cell has an activation value equal to the played or heard duration, and the interaction cell is connected bidirectionally to two basic cells. The interaction cell steers the basic cells that it is connected to toward integer multiples of each other, but only if they are already near such a state. Sum cells are also postulated that adjust themselves to the sum of the activation levels of the subsequent basic cells they are connected to. In this way they represent the longer time intervals generated by a sequence of notes. These cells are also interconnected by interaction cells, so that they also tend to stabilize on an integer division of each other and while doing so, they steer the basic cells toward a metrical score. The network can be rather sparse, allowing only interaction of subsequent or hierarchical time intervals. The models were tested in simulation. A typical net of around ten basic cells with a total number of around a hundred cells will stabilize in about forty iterations in which deviations from the metrical score from 0 to 30% are reduced to 0.1%. This seems quite promising. Further research will concentrate on the characterization of the limits of these models, and evaluate their computational requirements, their psychological plausibility, and the possibilities of real-time processing.

Representations of Music

Böcker and Mahling presented a proposal for A MUSic EDitor for visualizing Harmonic relationships, abbreviated Amused. Amused is an alternative to the commercially available editors of musical scores in that it provides the normally lacking level of a more conceptual description of the score. Three aspects of knowledge are supported by the system: the graphical representation; secondly, its constraints (such as possible locations of the symbol in the score); and finally, aspects independent of the visual representation, e.g. the function of a tone in a chord. Most of the Common Music Notation editors (e.g. Professional Composer, Score) support the first two. The last one is hard to find in commercially available systems. Böcker and Mahling’s proposal has a lot in common with other proposals by researchers in the AI field (e.g. Steels 1986): an object-oriented description of the music that supports both a graphical representation and a knowledge representation of musical objects and their relations.

Mira Balaban of Ben-Gurion University, Israel presented a paper with the title “The Cross-Fertilization between Music and AI.” She presented a view of music pieces as possibly multiple, hierarchical structures organized along the time scale: a representation language to make structured descriptions of music. This representation language is extended with an attribute
mechanism that can characterize the structured objects and provide information about them. Balaban proposed the standardization of this language, because all known systems would fit in. In this language the simplest element is considered the occurrence of a sound: an elementary structured musical piece. A larger form of music is described as a composite structured musical piece, a collection of "time stamped" musical structures. The time stamp of a composite piece is given relative to the beginning of the whole piece. An elementary music can be described as a durationless sound object: a description of sound stripped from its temporal characteristic, as when and how long it occurs. Initial development is on a SUN machine with LISP, with plans to port it to a PC environment.

**Conclusion**

The importance of representation was implicit in most talks. The representation of temporal knowledge is especially important and needs more attention in research. All this is reflected in the growing field in AI research of knowledge representation. The workshop also showed the two main directions in AI: the Connectionist approach versus GOFAI (Good Old Fashioned AI), implicit versus explicit representation of knowledge.

The proceedings of this workshop will be published by Springer Verlag in 1989. Many of the speakers also presented work at the AAAI workshop on AI and music (Proceedings can be ordered from Steve Taglio, 445 Burgess Drive, Menlo Park, CA).
REFERENCES


