A ROBOT BASED INTERACTIVE COMPOSITION SYSTEM

Jônatas Manzolli[#], Klaus Wassermann[##], Mark Blanchard[##] & Paul Verschure[##]

[#] *Interdisciplinary Nucleus for Studies on Sound Communication,*
Rua da Reitoria, 165, 13036-300
University of Campinas - Brazil.

[##] *Institute of Neuroinformatics,*
ETH-UZ, Switzerland.
Winterthurerstrasse 190, CH-8057
Zurich, Switzerland.
jonatas@nics.unicamp.br, (klausw jmb pfmjv)@ini.phys.ethz.ch

ABSTRACT

This paper describes Roboser (http://www.ini.unizh.ch/~jmb/Roboser.html) an interactive real-world composition system. Roboser consists of a real-world device (robot) which autonomously interacts with its environments and an algorithmic computer-based composition system, CurvaSom. Starting with fragmented sequences of MIDI events, sensory events and internal states of the robot's control model are interfaced with CurvaSom to generated new sound material. These first experiments demonstrate that novelty and change in algorithmic composition can be generated by a complete synthetic system, which interacts with the world.

INTRODUCTION

Presently two main trends can be observed in the development and application of novel technology to generate and compose music. One is the shift from the notion of computer assisted composition to that of Interactive Music Systems (e.g. Rowe, 1993). In this direction special emphasis is placed on the construction of interfaces in order to create novel music instruments (see paradiso97 for an overview). Waisvisz has for instance, developed novel interfaces. He created a gesture interface, The Hands (Krefeld, 1990; Ryan, 1991), which controls sound production using several sensors and keys attached to the players hands. Utilising ultrasound technology and the translation of analogue to MIDI signals, by the so-called SensorLab, movements of the hands, the fingers and the arms are used to control the production of sounds by MIDI devices. Others have used constructed hand controlled musical interfaces using data gloves (Sawada et. all, 1997). Another example can be found in the work of Machover who, in his project Brain Opera, presents humans with a collection of novel interfaces to musical instruments, or hyper instruments in the so called Mind Forest (Paradiso, 1997; Paradiso, 1999). The "Very Nervous System" created by Rokeby, transforms movements into sonic events and was first defined in the middle of the eighties (Rokeby, 1998). Hence, in these approaches the emphasis is on using novel technology to create more accessible interfaces to physical or virtual musical instruments and to find new ways to allow the generation of compositions consisting of both human and computer generated musical material.

A complementary approach aims at developing more advanced systems for the specification and generation of sound material and musical compositions. For instance, at IRCAM, several researchers have developed a range of dedicated solutions in both hardware, e.g. 4X and the ISPW workstation, and software, e.g. jMax (Dechelle et. all, 1998). The goals of this approach move in the direction of synthesis of sonic events, the specification and control of computer generated musical compositions, the transformation of sound material produced by human musicians, and its integration with computer generated material. For instance, jMax, an object-oriented visual-programming environment for real-time musical control, provides a number of DSP implemented functions which give composers the possibility to synthesise and transform sonic events. jMax has, for instance, been used in the ESCHER sound synthesis environment were parametric interpolation of gestures is used to synthesise sounds in real-time (Wanderley et. all 1998).

The Roboser project presents a new trend in the realm of interactive music systems by allowing sensory input and states intend to a real-world system to control an algorithmic composition process. Central elements of Roboser are the distributed real-time neural simulation environment IQR421 (Verschure, 1997) and CurvaSom, an algorithmic musical composition system which synthesises midi events in real-time (Manzolli & Maia, 1998).

INTERACTIVE ALGORHITMIC COMPOSITION ENVIRONMENT

The operations performed by interactive musical systems can in general be described in terms of the sensory interface, the transduction of external events to a central processing stage and a subsequent transformation into the parameters for sound production (**Figure 1**). Each of these stages can be affected using externally set parameters where the interface and production stages allow humans to affect the parameters of a composition system and observe its sonic output. The features of the composition are crucially dependent on the transformations performed by the central processing stage.

An interface allows the transformation of physical events, e.g. robot actions, into some analogue or digital representation, which allows further processing. Transduction transforms these events into the data format and representation of the central processing stage. Central processing defines the parameters of the composition, which are transformed into information to control sound production. At the stage of sound production this digital information is transformed into analogue musical signals via a DSP card or a MIDI port to address MIDI sound modules.

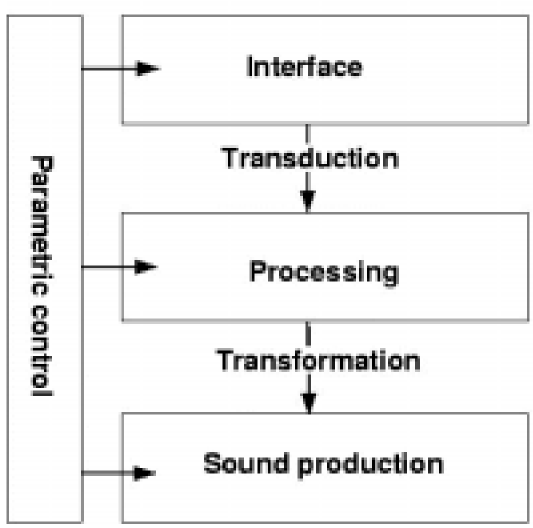

Figure 1 – Diagram of the sound production process in an interactive music system.

Key issues in Interactive Composition Systems research are the properties of the interpretation process that is applied to the events captured by the sensors and the transformation process into the acoustic domain, respectively. Traditionally, research in this domain deal with synthesising pitches and timbres. CurvaSom, however, aims at synthesising a complex composition using a selection of predefined sound sets, taking advantage of the MIDI protocol. It generates musical events in real-time and is organised around a number of internal heuristics for sound generation, so called Sound Functors (Manzolli & Maia, 1998) and a set of parametric sound specifications with which the system is seeded. The original version of CurvaSom had a graphical interface by which a user could change these parameters to produce complex musical events by drawing lines on a computer screen. For Roboser, CurvaSom was adapted to accept input signals from external sensors using IQR421.

The notion of Sound Functors (*F*, see **Figure 2**) is based on the mathematical theory of categories (Manzolli & Maia, 1998). In this approach a compositional process is divided into two components or domains: the control domain and the sonic domain. These domains are defined as mathematical representations of acoustic phenomena. The control domain is a parametric set, containing sub-sets specifying the numerical control of sound features. The sonic domain contains the set of specific sound features, which can be controlled by the parameters of the control domain. The compositional process depends on the relationship between these two domains. These relationships, or morphisms (M1-n), transfer the structure of one domain onto the other.

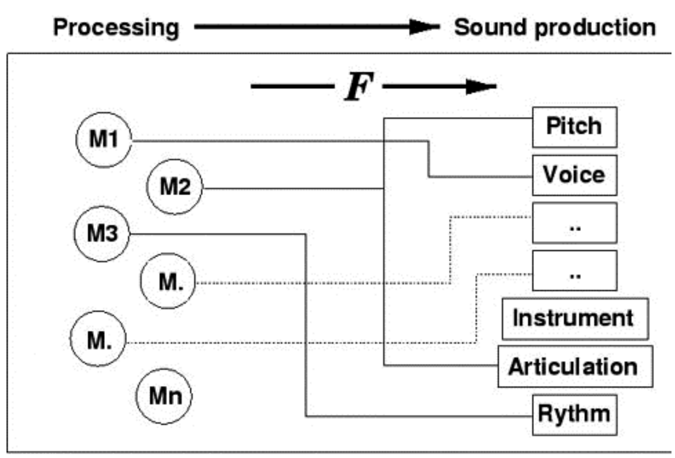

Figure 2 – Diagram of a Sound Functor (*F*).

The sound domain is defined in terms of the MIDI protocol. Sound features controlled by the system are voice, pitch range, articulation, rhythm, dynamic, scales, bars, instruments, etc. Morphism defines processes, which generates a stream of MIDI events. The composition model implemented by CurvaSom assumes that a composer immediately evaluates the sonic results of his or her actions and redirecting the musical trajectory. This model interprets musical composition as a self-organising, interactive process as discussed in (Manzolli, 1996).

TRANSFORMATION OF REAL-WORLD INPUT TO SOUND STRUCTURES

In our general model of Roboser the Gestural Controller and the Mapping Layer are encapsulated in the Processing Level (Figure 3). The data entering the Processing Level originates from a variety of sensors including cameras, Infrared-sensors, microphones and pressure sensors. Indeed, in terms of the type of sensors used, the system is very flexible. The sensory data is processed by IQR421; a program developed to simulate large-scale neural networks in real-time. IQR421 extracts parameters from the sensory input that form the information on which a composition is produced. CurvaSom in turn generates music using the parameters defined by IQR421. Therefore, our approach emphasises the Processing Level of an interactive compositional system creating an organism-like structure in which a combination of algorithmic compositional tools is connected to the real world via a neuronal system, an artificial brain. The result is a system which is not only able to capture and map analogue sensory input to musical output, but also to create and change musical forms, biased by sensory input, in an open-loop manner and in real-time.

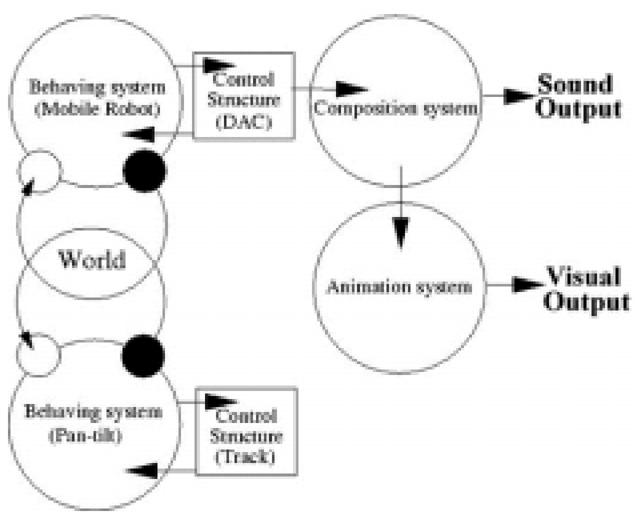

Figure 3 – General diagram of Roboser

THE EXPERIMENT

In the first experiment, presented at the computer fair Orbit '98 (Basel, Switzerland), we used a mobile robot (Khepera, K-team, Lausanne, Switzerland) as a source for sensory data. The environment of the robot comprised multiple interconnected hardware and software systems (see Figure 4). These elements included cameras, lights and a so-called "silicon retina" (Mahowald & Mead, 1991), which provided a rich sensory input for the development of the music. The robot moved around on a clear transparent Perspex panel within a custom-build arena of 70cm diameter. Mounted below the floor was an array of 17 lights; mounted above the arena was a pan-tilt system, which carried both a colour CCD video camera and the "silicon retina". The musical composition was visualised by an animated computer graphic, AMV, which was projected to a wall using a video beamer displaying each voice of Roboser by a distinct 3D object.

Roboser changed its musical output based on the states of a reactive control structure (Verschure & Voegtlin, 1998) implemented in IQR421. The robot displayed a strong preference for light sources, which it would approach, while it would steer away from obstacles. These appetitive and aversive states were translated into changes of the musical composition. For instance, as an expression of the pleasure to find light Roboser would play more pronounced improvisations in its composition, while unpleasant events would induce dramatic changes in voice instrumentation. A colour CCD camera mounted on the pan-tilt system was used to track strongly coloured objects, e.g. visitors. The pan-tilt system was steered so as to keep the preferred colour within the centre of the view. The "silicon retina" responded to rapid changes in the images as the coloured object was tracked and, via a predefined mapping, changed the pattern of lights in the robot arena. These changes in light altered the behaviour of the robot and in turn the development of the composition. The control of the hardware devices was defined within IQR421, an integrated system for simulating large neural circuits in real-time (Verschure, 1997). IQR421 allows processes to be distributed among a network of computers via the TCP/IP protocol and is written in the C programming language. In this exhibit processes ran on four Pentium II PC computers under the Red Hat Linux 5.1 operating system.

The robot would "dance" over a light source until the light was turned off. This would allow the robot to explore the environment searching for another light source. Visitors could actively influence the robot's behaviour by using small torchlights to attract it, or by putting small objects into the arena the robot had to steer around. The combination of sensory events, i.e. lights or collisions, and behavioural states like exploration, avoidance or approach, were used to control the musical composition process.

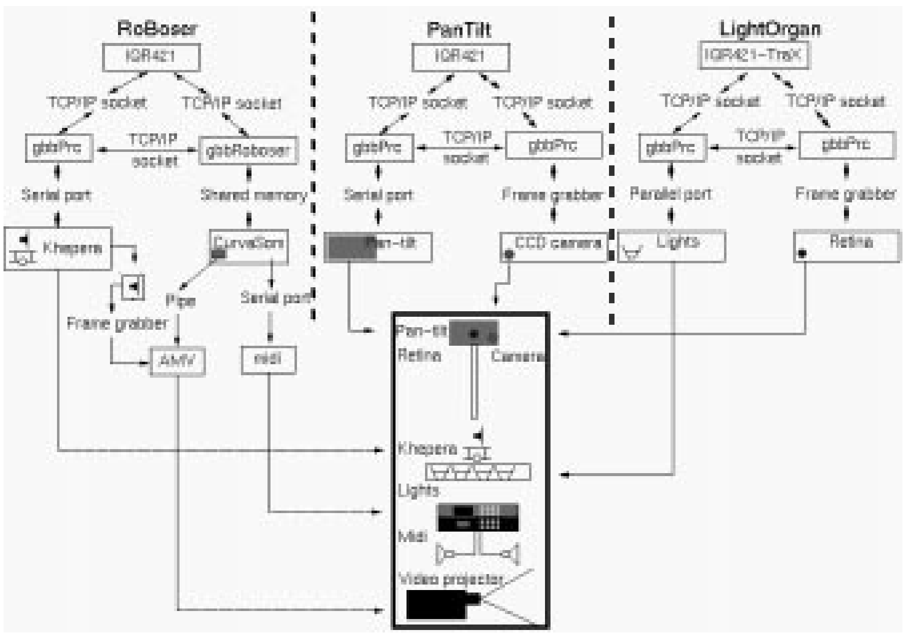

Figure 4 - The environment used for presentation at ORBIT '98, showing interconnected hardware and software systems. See the text for more details.

DISCUSSION

Rapid advances have been made in the technology to synthesise sonic events, and to define, influence, and algorithmically generate musical compositions. The most challenging question at this point in time, however, is whether we can use this technology to *autonomously* generate musical compositions. In this context it has been argued that there is a critical barrier beyond which machines cannot go. For instance, Roads (1996) proposes that the human role in the compositional process is and will remain crucial. His argument is essentially an empirical one: In case machines could be constructed which would be able to display creativity and virtuosity in musical composition, they would already have been built and have replaced humans. An additional argument is based on the observation that it is still a human being that defines the algorithms that perform the composition. This last problem, however, is not special for the area of algorithmic composition and is haunting most of the traditional and contemporary research in artificial intelligence (Searle 1982; Dreyfus 1992; Vershure 1993). It in essence deals with the problem of defining artificial systems that can develop beyond their a priori specification (Vershure, 1998).

We want to evaluate to what extent human intervention in the compositional process is the necessary and sufficient condition for the generation of appealing musical structures. We do agree with the critics of the field, such as Roads, that until now humans have played the central role in the artificial creation of musical compositions and the synthesis of sound. The creativity and aesthetics of the compositions these systems produce still fully relies on human pre-specification. We disagree, however, with the claim that this would demonstrate it in principle impossible for machines to achieve a level of musical composition and generation paralleling ``human'' performance.

We assume that an artefact with the ability to autonomously synthesise novel and complete musical compositions, or an autonomous synthetic musical composition system (ASMCS), will have specific properties. First, an artificial system, which has complex abilities, such as expressed in musical composition, might have to be based on a control structure, e.g. a brain, which is equally complex. Second, this control structure must be able to have, and ideally acquire, knowledge of its task domain and, third, to be able to interact with its environment. Central for such a system is that its behaviour, e.g. sonic output, is an expression of the domain ontology of the system itself and not that of its designers. In its realisation such a system needs to be based on methods and tools which allow the specification of the domain knowledge for musical composition, the specification of complex control structures, and both sensory and effector systems which allow interactions with the real-world.

We do not claim to be close to having developed an autonomous synthetic musical composition system but we realised initial experiments on a system, called Roboser. It combines the three key technologies identified as necessary components of an ASMCS. In particular: CurvaSom an algorithmic musical composition system (Manzolli & Maia, 1998), IQR421, a system which supports the simulation of large scale neuronal systems which can be interfaced to external devices (Vershure, 1997) and a collection of real-world artefacts and sensor technology such as a mobile robot, Khepera (K-team, Lausanne, Switzerland), (Mondada et all, 1993) a motorised pan-tilt system (Directed Perception Burlingame, Ca, USA), and imagers such as standard colour CCD cameras (K-team, Lausanne, Switzerland) and silicon retinae implemented using CMOS based analogue Very Large Scale Integration (aVLSI).

CONCLUSION

Roboser allows an autonomously behaving robot to control an algorithmic composition system. The robot behaves, it senses the world and its control structure interprets these sensory events and transforms them into actions in the world. Roboser transforms the robot's experience of the world into an evolving musical performance. The goal of this paper is to introduce a process, which is considered artistic and creative into a robotic system. In this way the possibility is created of defining autonomous real-world composition systems, which would "create" music without human interference. It could lead to various kinds of applications ranging from entertainment to complex pieces of music. This way of thinking is interesting, not only because it allows us to have dynamic sound behaviour using the robot but also because it opens a research framework that connects Neuroinformatics to Multimedia Sciences, Algorithmic Composition and other forms of Electronic Art.

ACKNOWLEDGMENTS

We thank the Brazilian Research Agencies CNPq and FAPESP for the financial support for the Third Brazilian-International Conference on Cognitive Science, which made this collaboration possible. Part of this project was realised through the support of FAPESP to Manzolli for the Gesture Interface Laboratory in which CurvaSom was developed. The realisation of Roboser at ORBIT98 was possible through support by ETH Zurich, Jecklin Zurich, and K-team Lausanne. Blanchard and Verschure are supported by SPP-SNF Switzerland.

REFERENCES

Dechelle, F., Borghesi, R., DeCecco, M., Maggi, E., Rovan B. & Schnell, N. (1998). jMax: a new JAVA-based editing and control system for real-time musical applications. Proceedings of ICMC: International Computer Music Conference, Ann Arbor, USA, October 1998.

Dreyfus, H.L. (1992). What Computers Still Can't Do A Critique of Artificial Reason. Cambridge Ma.: MIT Press.

Krefeld. , V. (1990). The Hand in the Web: An interview with Michel Waisvisz. Computer Music Journal, 14:2:28-33.

Manzolli, J. (1996). Auto-organização: um paradigma Composicional. Published at Auto-organização: estudos interdisciplinares, org. M.E.Q. Debrun and N. Gonzales and O. Pessoa Jr., Campinas, Brazil: CLE-Unicamp, Vol: 18, pg. 417-435.

Manzolli, J. & Maia Jr., A 1998. Sound Functors Applications. In proceedings of the fifth Brazilian Symposium on Computer Music, XVII Congress of the SBC, UFMG, Belo Horizonte, Brazil, pp. 115-120.

Mahowald, M. (1994). An analog VLSI system for stereoscopic vision. Dordrecht: Kluwer.

Mondada, F. and Franzi, E. and Ienne, P. (1993). Mobile robot miniaturisation: A tool for investigation in control algorithms. Experimental Robotics III: Proceedings of the 3rd International Symposium on Experimental Robotics, Kyoto, Japan, October 28-30, 1993. Berlin: Springer Verlag, pg. 501-513.

Paradiso, J. (1997). Electronic Music Interfaces: New Ways to Play. IEEE Spectrum Magazine, 34:12:18-30, 1997.

Paradiso, J. (1999). The Brain Opera Technology: New instruments and gestural sensors for musical interaction and performance. Journal of New Music Research, 28:2:130-149.

Roads, C. (1996). The Computer Music Tutorial. Cambridge, Ma.: MIT Press.

Rokeby, B. (1998). The construction of experience: Interface as content. Published at Digital Illusions: Entertaining the future with high technology, ed. Association for Computing Machinery, Reading, Ma.: Addison-Wesley.

Rowe, R. (1993). Interactive Music Systems: Machine Listening and Composing. Cambridge, Ma.: MIT Press.

Ryan, J. (1991). Some remarks on musical instrument design at STEIM. Contemporary Music Review, 6:1:3-17.

Sawada, H., Onoe, N. & Shuji, H. (1997). Sounds in Hand - A Sound Modifier Using Datagloves and Twiddle Interface. Proceedings of the International Computer Music Conference, ICMC, 309-312.

Searle, J. (1982). Minds, brains and programs. Behavioral and Brain Sciences, 3:417-424.

Verschure, P. F. M. J. (1993). Formal minds and biological brains. IEEE Expert, 8:5:66-75.

Verschure, P.F.M.J. (1997). Xmorph: A software tool for the synthesis and analysis of neural systems, Technical report Institute of Neuroinformatics, ETH-UZ.

Verschure, P. F. M. J. and Voegtlin, T. (1998) "A bottom-up approach towards the acquisition retention, and expression of sequential representations: Distributed Adaptive Control III", Neural Networks, 11:1531-1549.

Wanderley, M. and Schnell, N. and Rovan, J. (1998). ESCHER - Modeling and Performing Composed Instruments in real-time. Proceedings of IEEE Int. Conference on Systems, Man and Cybernetics San Diego - CA, USA, October 1998.