The Artificial Life Route to the Origins of Music

Eduardo Reck Miranda
SONY Computer Science Laboratory Paris
6 rue Amyot
75005 Paris - France
miranda@csl.sony.fr

1 Introduction

In the editorial of the first issue of the journal *Artificial Life and Robotics*, John Casti draws our attention to a shift that is occurring in scientific research methodology which will certainly play a key role in the Arts and Sciences of the twenty-first century: Casti observes that Science is increasingly shifting its attention from *matter* to *information* (Casti, 1997). He observes that contemporary Science is no longer primarily concerned with the study of the material composition of nature, but rather with the study of the functional characteristics of its tangled systems, most notably the interaction between the components of a system, the interconnection of different systems and the emergence of global behaviour. Thanks to computer technology, scientists can now create surrogate artificial worlds to perform very complex experiments that would otherwise be impossible to perform. These artificial worlds are analogous to the more familiar laboratories (test tubes, retorts, etc.) used by scientists for centuries to investigate the structure of atoms, chemical components, cells, and so forth; the main difference is that these artificial worlds are used to simulate natural phenomena in order to study them in terms of the functional activity carried by patterns of information. The emergence of fields such as Artificial Life (Alife), for example, is a natural consequence of this shift of paradigms.

Alife is a discipline that studies natural living systems by simulating their biological phenomena *in silico* (Langton and Shimohara, 1997; Olson, 1997). The attempt to mimic biological phenomena on computers is proving to be a viable route for a better theoretical understanding of living organisms, let alone the practical applications of biological principles for technology (robotics, nanotechnology, etc.) and medicine. Because Alife is dealing with very complex phenomena, its development has fostered the creation of a pool of research tools for studying complexity. It is interesting though, that these tools are also proving to be useful in fields other than Biology, most notably Social Sciences (Epstein and Axtell, 1996) and Linguistics (Steels, 1997b). In this essay we propose that the Alife paradigm also has great potential for Musicology.

We are investigating how musical forms may originate and evolve in artificially created worlds inhabited by virtual communities of musicians and listeners. Origins and evolution are not, however, studied here in the context of genetic mutations, but rather in the context of cultural conventions. We consider that variations in musical styles, for example, result from the emergence of new rules and/or from the shifting of existing conventions for *music making* (the term music making is used here to mean both creating music and listening to music). Musical styles maintain their organisation within a cultural framework and yet they are highly dynamic; they are constantly evolving and adapting to new cultural situations (Reck, 1997). Whilst the criteria for natural selection in Biology are chiefly based upon physical fitness and reproductive capability, in Music these criteria will depend upon on the effects of the new rules on the music making experience.

In this context, musical forms can be studied as live organisms and as such Musicology can benefit hugely from Alife s research paradigms. Particularly interesting paradigms that have emerged recently include *genetic algorithms* (Koza, 1992), *cellular automata* (Ermentrout and Edelstein-Keshet, 1993; Wolfram, 1994), *evolutionary modelling* (Epstein and Axtell, 1996) and *autonomous robots* (Steels and Brooks, 1995), to cite but a few. Indeed, several musicological investigations have already begun to look closely at these developments, notably by composers interested in the organisation principles that underlie Alife-based pattern-generation algorithms (Degazzio, in this book; McAlpine *et al.*, 1999; Miranda, 1993; 1994). We have come to believe, however, that the great potential of Alife for Musicology has yet to be explored: *evolutionary modelling*. Our hypothesis is that the mechanisms for generating complexity commonly found in biological systems may also explain the spontaneous origins and evolution of musical forms.

The first half of this essay sets the background scenario of our research. The elements of this scenario include Wittgenstein, talking chimpanzees, babbling babies and a couple of philosophers of the French Enlightenment. As we shed light onto the various points of this rather eclectic scenario, we hope to clarify various fundamental issues that underlie our work. Here we focus mainly on the importance of the interplay between language and music, and the reasons for taking on board the Alife paradigm. Next we introduce the notion of evolutionary modelling and discuss its fundamental mechanisms. Then we present two experiments whose results substantiate our concluding comments.

2 Language games and the origins of meaning

In *Philosophical Investigations*, Wittgenstein (1963) proposes the notion of language games: simple linguistic plots specifically designed to illustrate particular points he wanted to make. At the start of his book, Wittgenstein proposes a simple language game as follows: imagine a language that is meant to serve for communication between a builder and his assistant. The builder is constructing a castle with building-stones: there are blocks, pillars, slabs and beams. The assistant has to pass him the stones, and in the order in which the builder needs them. For this purpose they use a language consisting of the words: "block", "pillar", "slab" and "beam". The builder calls them out and the assistant brings the respective material he has learned to bring, in compliance to such-and-such a call. When the builder says "block", for example, he causes an action to take place; the success of this action is conditioned to the fact that the assistant passes a block to the builder.

As far as the builder is concerned, it is sufficient that the assistant has learned only these four words without learning the all of the English language. The point that Wittgenstein wanted to illustrate with this simple game is that, in principle, the assistant could have learned the language only by associating words with actions and not with labelled pictures of different stones that may or may not appear "catalogued" in his mind. The indication that the assistant has learned the meaning of the words is given when he performs the right actions. To put it simply, Wittgenstein's point is that words (and sentences, for that matter) have meaning only when they have a role in specific contexts within the web of human activities.

Although Wittgenstein was not primarily concerned with the philosophy of language himself, his notions of *language games* and of the *action-based origins of meaning* have made a great impact on our research.

3 Can Sarah build castles in Spain?

Scientists' efforts to teach language to chimpanzees have made tremendous progress since the 1960s. Chimpanzees such as Sarah, Nim Chimpski, Washoe, and Kanzi (Premak, 1976; Terrace, 1979; Gardner and Gardner, 1989; Savage-Rumbaugh and Lewin, 1994) caught the imagination of the world for having learned to play rather sophisticated language games with their tutors. Washoe, for example, was reported to be able to communicate with her teacher by means of simple gesticulation sequences made from a repertoire of over one hundred signals (similar to the signals of the American sign language devised for the deaf). More recently, Kanzi has been featured in the news media interacting with a person on the telephone, using a talking keyboard (Bindon, 1998). These apes have demonstrated their capacity to learn some primitive form of human communication, including evidence that they can spontaneously create new "meaningful" combinations of signs and that one chimpanzee can indeed pick up the language simply by watching their peer chimpanzees using it.

Even though some may still argue that these chimpanzees have simply learned to play with signs in order to please their teachers and get some form of reward, these experiments address important issues concerning the development of language in early humans, let alone the pedagogic techniques for teaching apes that were developed; e.g., Primack's *reinforcement principle* (Primack, 1963). Perhaps the most important issue that has been addressed here is the orthodox linguistic scholar's belief that there is a missing link, that is, an unbridgeable divide between humans and the rest of the animal kingdom. These experiments with apes strongly suggest that the difference may rather be a matter of a gradation of linguistic capacity; that is, we humans are better at learning to speak, at using symbols, at acquiring a grammar and at using it to generate complex sentences, and so on.

In this context, language is better seen as a cultural phenomenon that emerges from sociological interactions, rather than as a ready-made feature of the infant's brain. Humans learn a language by engaging in countless language games, for all kinds of situations. As the games and situations become increasingly complex, children develop a myriad of multi-layered networks of utterances and meanings, and engage in increasingly advanced forms of linguistic understanding. Terms that are learned in more rudimentary games become part of higher-level contexts, and vice-versa. A child that learns to play a "building" game by putting Lego bricks together in order to build a toy castle, for example, will soon build other structures spontaneously and will probably also experience the concept of a castle in another context. This child will then develop the capacity to transfer the meaning from specific games to other contexts, and so forth. Linguistic maturity then is reached when one is capable of making sense of metaphorical settings; for instance, the English expression "to build castles in Spain" denotes a dream that is very unlikely to be realised.

In principle, chimpanzees could learn to build a castle, but not in metaphorical Spain. The point we want to raise here is that even though scientists manage to make further progress in teaching language to apes, we believe that this sort of "apeish" linguistic ability is not sufficient to study the intellectual capacity found in humans.

4 Babbling is music

But what are the limitations of apes? Why do we have such amazing intellectual capacity? Attempts to answer these questions abound in literature: apes are not intelligent enough because their brains are too small (Deacon, 1997); the vocal tracts of monkeys were not designed for speech (Lieberman, 1984); humans developed language because our sophisticated sensory-motor ability allows for better management of time and space (Calvin, 1983); chimpanzees do not have the "language organ" (Chomsky, 1980); to cite but a few. These are by all means plausible hypotheses, but they certainly do not work in isolation; these features must have coevolved. For instance, as the brain of Neanderthal man increased in size, a resonating chamber emerged in the mouth and the position of the larynx was lowered, and our sensory-motor system was improved, all resulting in a suitable physical platform for spoken language. We believe, however, that there is an important component of this coevolution phenomenon that has been largely ignored by our fellow researchers: *music*.

Music is one of the most intriguing phenomena of the human kind. Our sensitivity to timing and imitation, our tendency for imposing order on auditory information, our ability to categorise sound, our tendency to recognise and imitate sound patterning, and so on, are all unique to humans (Storr, 1993). These are essentially musical abilities that form an ideal platform for language development. We believe that the bipartite functioning of our hearing system is the crux of this musical platform: babies are generally more sensitive to *rhythmic patterns* in the left ear and to *pitch* and *timbre* in the right ear (Best, 1988).

As a rough scenario to illustrate our hypothesis, let us consider the fact that babies have a unique disposition to extract little chunks of sounds from the undulating patterns of speech and to imitate these sounds (Jusczyk, 1997). In order to extract these little chunks of sounds, the left ear seeks timing cues in the signal; e.g., starting and ending of longer periodical signals (e.g., vowels) demarcated by short non-periodical bursts of energy (e.g., consonants). The task of the right ear is to listen for sound colouration cues such as the tone of the voice and the melodic contour of the utterances. Indeed, babies do seem to respond better to clearly articulated syllables and exaggerated melodic contours. In the early months, the only response babies can make to heard utterances is to babble; this is because they are still learning how to play with the raw building materials of language. (Virtually all human babies babble; even deaf children seem to babble with their hands.) During this learning process babies develop their neural and muscular apparatus, firstly by learning how to recognise utterances and then by trying to imitate them. Preferences for listening to and imitating the sounds of the mother tongue begin to emerge due to positive feedback responses which babies receive from their interlocutors. Sooner or later children start to form lexicons, to make sentences, shape grammars, and to engage in increasingly sophisticated linguistic experiences.

We by no means assume that people are born with a ready-made music machine. What we are suggesting here is that our musical predisposition forms the basis for the development of the sophisticated discrimination and categorisation machinery that are very important for language. But our musicality, so to speak, develops together with language as we grow up.

5 Early philosophy does makes sense

The notion that our linguistic capacity is closely related to our ability to both make and appreciate music was prominent in Enlightenment thinking in the eighteenth century (Thomas, 1995). Reflections on how primordial utterances, cries and vocalisations would have evolved into language naturally brought musical considerations within the scope of the writings of philosophers such as Condillac and Rousseau.

Condillac in his *Essay on the Origins of Human Knowledge* depicted the earliest spoken language as being composed of action-orientated vocal inflections such as warnings, cries for help, shouts of joy, etc. (Thomas, 1995). But most interestingly, Condillac proposed that these inflexions were accompanied by variations in pitch and timbre. In short, he suggested that the early hominids did not prioritise the invention of different "words", but tended to produce the same form of utterance at different tones in order to express different things; presumably by varying pitch, loudness and duration. Condillac thus suggested that primordial languages did not have consonants but vowel-like intonations. The prosody of earlier languages must have sounded like a kind of primitive song (Arbo, 1998).

Rousseau also purported the idea that language is derived from natural sounds produced by our vocal organs. For Rousseau, however, song and speech have a common ground: *passion*. In the beginning, vocal utterances expressed primarily feelings (e.g., "I am sad."), whilst gestures were normally preferred to express rational thoughts (e.g., "Go hunting, I am hungry!"). Rousseau agrees with Condillac that primeval spoken languages must have sounded like melodies of vowel-like utterances, but Rousseau has an interesting story for the emergence of consonants: as hominids dealings with one another grew in complexity, spoken language needed to become less passionate and more precise. In his *Essay on the Origins of Language* Rousseau argues that language was motivated by the increasing necessity for social bonding (Thomas, 1995). Within this bonding process, the amount of tone variations decreased, giving rise to the appearance of consonants. New articulations needed to be formed, and consequently grammatical rules for making sequences of utterances soon emerged. For Rousseau, modern languages (such as his own mother tongue French) no longer spoke to the heart alone, but also to reason. As language followed the path of logical argumentation, those melodic aspects of the primordial utterances evolved into music instead. Music thus developed from the sounds of passionate speech.

Although dating back to the eighteenth century, these two philosophers conjectures do continue to make sense to a certain extent. For instance, some non-Western languages, such as Chinese, have mechanisms in which multiple intonations of the same word convey different meanings. This is not to say that Chinese is primitive or less rational, but rather to indicate that the Chinese language took a different route. More recently, researchers found out that the structure of Neanderthal man s vocal tract could not produce plosive consonants such as /k/ and /g/ but it could produce almost all vowels, except /i/, /u/ and /a/ (Lieberman, 1998). Interesting though is that the most difficult vowels for the Neanderthal man, that is /i/, /u/ and /a/, are the ones that appear in an average of 85% of more than 300 recently catalogued human languages (Maddieson, 1984; Ladefoged and Maddieson, 1996).

Since Condillac and Rousseau, the relationship between the origins of language and music has hardly been systematically addressed again. Their writings seem to have been overshadowed by the Romanticism that prevailed in Europe in the nineteenth century, most notably in music; influential philosophers who were interested in music often associated its origins with the mystical, the ineffable and the hidden (Monelle, 1992).

Indeed, in the 1880s, the prestigious Linguistic Society of Paris sought to impose a ban on the theme of the origins due to a wave of wild unsupported writings that appeared at the time.

6 Evolutionary modelling

Thanks to the evolutionary modelling techniques that emerged from Alife research, the Enlightenment s quest for gaining a better understanding of the origins of language is now being brought back into play with a promising agenda (Noble and Cliff, 1996; Steels, 1997a; Arita and Koyama, 1998) and music will undoubtedly play a key role in this research. Furthermore, the investigation of the role of music in the evolution of language inevitably urges musicologists in the pursuit of the *origins of music* itself.

Music can be modelled as an adaptive system of sounds used by a number of individuals, or *distributed agents* in computer science jargon, engaged in a collective music making experience; some may only listen to the sounds ("audience") while others may be fully engaged in the generative process ("musicians"). Our hypothesis is that music emerges from the interaction of the agents when they engage in such music making experiences. There is no "global" supervision taking place and the agents do not have direct access to the musical knowledge of the other agents, apart from hearing what they actually do during the interactions. This model could be metaphorically compared to a jam session in a jazz club, where people, who have not necessarily met before, can join in and play, or simply watch. A contrasting scenario would be an orchestral concert where musicians follow a common score under the direction of a conductor; in this case we would say that there is very little room for music evolution.

On the opening pages of the first issue of *Evolution of Communication Journal*, Luc Steels proposes four fundamental mechanisms for studying the origins of language: *evolution*, *coevolution*, *self-organisation* and *level formation* (Steels, 1997b). As these mechanisms also seem to work well for the study of other cultural phenomena, we have taken them as the starting point of our research into the origins of music.

Before we discuss these mechanisms, it is important to clarify the contexts in which the term evolution appears in our research. In Natural History, evolution is frequently associated with the idea of transition from an inferior species to a superior one. It is also generally assumed that this transition is accompanied by an increase in the species' complexity. The Darwinian argument that humans originated from anthropoid apes is a typical example of this approach to evolution. Indeed, this approach made a serious impact on nineteenth century Cultural Anthropology, which maintained that human kind has become increasingly sophisticated during the linear course of history. The legacy of this argument is the general popular belief that the Stone Age, for example, represents far less sophistication than the Iron Age (Ullmann, 1983).

This approach to evolution suffers from taking for granted that one can "measure" evolution entirely based on materialistic criteria. For example, one should not take for granted that European classical music is more sophisticated than African drumming solely based on the technological sophistication of the instrument used. The rudimentary technological development of most non-European societies (e.g., Pygmies) gave rise to very sophisticated social systems and complex religious rituals. A remote community living in an environment where prey and crops abound would not be under pressure to develop sophisticated hunting weaponry or irrigation technology, for instance. These people would probably give priority to the creation of a belief system in which a religious ritual would have to be performed occasionally in order maintain the richness of their habitat.

In the context of our research, however, the term evolution is associated with the idea of improvement. Coevolution then is used to denote the concept of transition from one state of affairs to another, and is not therefore necessarily associated with the idea of improvement.

6.1 Evolution

Evolution occurs when a transformation process creates variants of some type of information. Normally there is a mechanism which favours the best transformations and discards those that are considered inferior, according to certain criteria. For example, the criterion in Biology might be fitness for survival, whilst in Linguistics the selection of an utterance could involve a compromise between the effortless use of the speaker s articulatory mechanism and the degree of understanding by the listener.

In Music, the selection of sound objects is based upon psychoacoustics (e.g., the interplay between repetition and variation) and physical criteria (e.g., the capability of the musical instrument(s) available). These selection criteria are the crux of the design of realistic evolutionary models for Music.

A plausible methodology for studying evolutionary mechanisms in Linguistics has been proposed by Steels and co-workers (Boer, 1997; Steels and Vogt, 1997). Inspired by Wittgenstein’s concept of language games, they have devised a number of different games for investigating different processes of language formation: phonology, meaning, syntax, lexicon, and so forth. An example of a phonological experiment of seminal importance for our research will be introduced later.

6.2 Coevolution

By coevolution we mean the interaction of various contiguous "evolutions". Evolutionary criteria are not fixed, but rather they are affected by an environment that is also in a state of flux. Whilst evolution tends to drive a system towards the improvement of particular aspects, coevolution tends to push the whole system towards greater complexity. For example, the evolution of musical styles should normally be studied in close association with the evolution of musical instruments, and vice-versa.

As an example, consider the case of the keyboard class of instruments: the evolution of the piano is popularly associated with the settlement of the equal-temperament tuning system (a tuning system in which all notes of the scale are equally separated by exactly half a tone) and with the increasing use of expressive dynamics in compositions; the piano can produce a much wider band of variations in dynamics than the harpsichord, from extremely loud to extremely quiet notes. On the one hand, the equal-temperament system alleviated the problem of tuning different instruments in an ensemble but, on the other hand, it increased the scope for composers to explore much more complex harmonic structures in their music; e.g., the modulation to different tonal keys within the same piece (Campbell and Greated, 1987).

6.3 Self-organisation

The notion of self-organisation is closely related to the notion of coevolution. The emergence of sound lexicons and listening contexts leads to the formation of organisational principles. These organisational principles or musical conventions, so to speak, arise without central control, but from the local interaction of distributed agents.

The origins of coherence in distributed systems with many interacting agents is a vast research topic. To put it in simple terms, three ingredients are needed for self-organisation to take place in a system: (i) a set of possible variations, (ii) random fluctuations and (iii) a feedback mechanism. Random fluctuations in the system will eventually strengthen some of the fluctuations because of the feedback mechanism: the more a fluctuation is strengthened, the more predominant it becomes. As an example, imagine a musical game as follows: a group of virtual agents engage in a drumming session but none of the agents is an experienced musician; i.e., they have no musical training. Each agent can bring in a different percussion instrument but an agent must never have played its instrument before. They all agree that they should start playing sounds simultaneously in any way they wish. To begin with, they will certainly produce a highly disorganised mass of rhythms; in this case we say that the system is in equilibrium. The set of possible variations in this system is the set of all noises and rhythms that can be produced by the instruments. Next, imagine a situation in which at some point an agent A makes a very distinct sound pattern that catches the attention of a fellow agent B. The fellow agent then attempts to imitate it. The imitation may not be exact (for example, because agent B s instrument is different from agent A s one), but agent A recognises it as an imitation of the pattern and instinctively reproduces the original pattern again; that is, agent A gives a positive feedback to agent B. The other agents will probably be keen to imitate this pattern as well and variations will certainly start to emerge. After a while, the pattern that was originated by agent A, plus its variations, should become successful conventions. The next time these agents engage in a jam session they will certainly remember these and other patterns, and will probably play them during the session. The more the patterns are played, the more conventional they become. When not engaged in jam sessions, some agents may even consider adapting their instrument to better produce those patterns, whilst others will spend a great deal of effort practising ways to produce them.

The musical plot described above would equally apply to studying the emergence of spoken language, for example. In this case the agents would be engaged in linguistic communication. The need for explicit communication of rational meanings would exert pressure for the formation of different types of syntactical conventions. The intention of the social interaction defines the rules of the game: for example, in a linguistic situation, it may be better that the agents engage in a one-to-one interaction rather than in a simultaneous one. Also, the nature of the coevolutionary phenomena may differ according the intention of the game. For instance, an utterance that is difficult to produce due to the limitations of the vocal tract will have low priority for becoming part of a language, whereas in music such a limitation would be much relaxed because the agents could, for example, modify their musical instruments or even create new ones.

Most cultural phenomena seem to follow identical self-organising principles in some way; see for example an interesting account for the origins of the flamenco song style in a paper by Washabaugh (1995) that appeared in *Journal of Musicological Research*.

6.4 Level formation

Level formation is a consequence of coevolution and self-organisation. By level formation we mean the formation of higher-level abstractions, such as both semantics and syntactical conventions in language, for example.

Suppose that at some point in the musical scenario described above, agents start remembering rhythmic patterns in terms of repeated short sequences grouped together as units; for example, shorter sounds may be grouped by similarity in duration and proximity in time, and repeated patterns may be grouped as units based on their parallel structure. This figurative conceptualisation of rhythm should then yield more abstract conceptualisations such as metric rules and sense of hierarchical functionality; for example, the concept of strong beat versus weak beat (Bamberger, 1991). Multiple rhythmic conventions for groupings and hierarchical organisations would soon start to emerge in the community of agents.

7 The experiments: towards a new paradigm for Musicology

Researchers at the Sony Computer Science Laboratory in Paris are currently developing a framework to study the emerge of language and music in artificial worlds. A toolkit called *Babel* has been developed for implementing simulations based upon the notion of adaptive language games (McIntyre, 1998; Steels, 1997b) and a system for evolutionary experiments in artificial musical worlds is on the drawing board.

A number of language games to study the basic mechanisms of evolutionary modelling, discussed earlier, have been implemented in Babel. These games are played according to a number of rules which define the intention of the game: what can be said, who says what, what agents do with the information, and so forth. The *discrimination games*, for example, are aimed at the study of level formation; that is, the study of the emergence of meanings in a vocabulary. In these games, robots (the agents) furnished with a camera try to categorise the objects they see in the environment by identifying distinctive features: shapes, colours, position, and so forth. A shared vocabulary of meaningful words emerges when agents try to teach each other the objects they have identified (Steels, 1997d).

We also have done a number of pioneering experiments using cellular automata to study the *self-organisation of musical forms* (Miranda, 1993; 1994) and the *dynamic evolution of synthesised sounds* (Miranda, 1995a; 1995b; 1998). The cellular automata investigations and a language game are discussed below.

7.1 Exploring the musical potential of digital life-forms

Cellular automata (CA) are computer modelling techniques widely used to model systems in which space and time are discrete, and quantities take on a finite set of discrete values.

Cellular automata were originally introduced in the sixties by von Neumann and Ulan as a model of a biological self-reproduction. They wanted to know if it would be possible for an abstract machine to reproduce; that is, to automatically construct a copy of itself. Their model consisted of a two-dimensional grid of cells, each cell of which had a number of states, representing the components out of which they built the self-reproducing machine. Controlled completely by a set of rules designed by its creators, the machine would extend an arm into a virgin portion of the grid, then slowly scan it back and forth, creating a copy of itself -reproducing the patterns of cells at another location in the grid (Cood, 1968). Since then cellular automata have been repeatedly reintroduced and applied to a considerable variety of purposes. Many interesting algorithms have been developed during the past thirty years.

In general, CA are implemented as a regular array or matrix of variables called cells. Each cell may assume values from a finite set of integers and each value is normally associated with a colour. The functioning of a cellular automaton is displayed on the computer screen as a sequence of changing patterns of tiny coloured cells, according to the tick of an imaginary clock, like an animated film. At each tick of the clock, the values of all cells change simultaneously, according to a set of transition rules that takes into account the values of their neighbourhood.

Figure 1 illustrates a very simple CA: it consists of an array of 12 cells and each cell can value either 0 or 1, represented by the colours white or black, respectively. At each tick of the clock, the values of all 12 cells change simultaneously according to a set of rules that determines a new value for each cell. In this case, the rules are based upon the values of its two neighbours. For example, if a cell is equal to 0 and if both neighbours are equal to 1, then this cell continues equal to zero in the next stage. More sophisticated CA configurations use a matrix of cells that can assume values other than 0 and 1 (and consequently, colours other than black and white). In these cases, the transition rules normally computes four or eight neighbours.

Figure 1:
A very simple CA. The right hand figure displays the colours associated to cell values; in this case 0=white and 1=black.

[FIGURE 1]

The main objective of our experiments here was to explore the potential of CA for modelling music creativity. As in these experiments we were primarily interested in composition, our main motivation was to investigate whether specific CA algorithms would be suitable for generating musical material. We have designed two programs for the experiments: *Chaosynth* and *CAMUS*. Whereas in Chaosynth we employed CA to control a sound synthesiser (Miranda, 1995a; 1995b; 1998), in CAMUS the CA were programmed to generate musical structures; e.g., sequences of chords, melodies, and so on (Miranda, 1993; 1994; McAlpine *et al.,* 1999). We briefly describe CAMUS below, focusing on only one aspect of its functioning: the use of Conway s *Game of Life* CA (Wilson, 1998); the reader is invited to consult the references for more information about these systems.

Figure 2:
A typical Game of Life sequence of three frames.

[FIGURE 2]

The *Game of Life* works as follows: from one tick of the clock to the next, the cells arranged in a matrix can be either alive (i.e., black) or dead (i.e., white), according to the following rules devised by Conway:

(i) if a cell is dead at time t, it comes alive at time t+1 if it 3 out of its 8 neighbours alive;

(ii) if a cell is alive at time t, it comes dead at time t+1 if it has fewer than 2 or more than 3 neighbours alive;

The rules are applied simultaneously to all cells of the lattice. An initial configuration of live cells may either grow interminably, fall into cyclic patterns, or eventually die off (Figure 2). In our implementation of the *Game of Life*, however, we allowed for the specification other rules, beyond Conway's original rule.

In order to map the behaviour of CAMUS onto music, we devised a Cartesian model to represent a triplet of notes (Figure 3). The model has two dimensions, where the horizontal coordinate represents the first interval of the triplet and the vertical coordinate represents its second interval.

Figure 3:
CAMUS uses a Cartesian model in to represent a triplet of notes.

[FIGURE 3]

To begin a musical session, the *Game of Life* is set up with a starting configuration and set to run. Each screen of the CA (Figure 2) produces a number of triplets (Figure 4). At each time step, the co-ordinates of each live cell are analysed and used to determine a triplet which will be played. In the case of Figure 4, for example, the cell at coordinates (5, 5) is alive, and thus constitutes a set of three notes. The co-ordinates (5, 5) describe the distance between the notes: assume that there is a given fundamental note (it is not important here to understand where this note comes from), then the second note will be five semitones above the fundamental, and the third note will be ten semitones (i.e., 5+5) above the fundamental.

Once the triplet for each cell has been determined, the states of the neighbouring cells in the *Game of Life* are used to calculate the starting time and the duration of each note, according to a set of temporal codes which we metaphorically called as *DNA codes* (Figure 5). These codes determine the temporal shape of each triplet; the actual values for the trigger and duration parameters are calculated in seconds according to other user specified parameters.

We observed that CA can generate interesting musical sequences. Indeed, a number of professional compositions were composed using CAMUS-generated material; e.g., the second movement of the author s string quartet *Wee Batucada Scotica* was almost entirely generated by the system (Miranda, 1998). Most encouraging was that CAMUS produced music in a certain Game of Life style which often sounded interesting, even pleasant.

Also, Chaosynth produces sounds that tend to exhibit a surprisingly great sense of natural movement and flow, and yet most of these sounds cannot be found in the real acoustic world. This is strong evidence that both *abstract musical forms* and *musical sounds* might indeed share similar self-organisation principles with cellular automata models, despite of the arbitrariness of our technique to map the behaviour of the CA onto music and sound simulations.

Figure 4:
Each screen of the Game of Life automaton produces a number of triplets.

[FIGURE 4]

Figure 5:
CAMUS uses temporal codes to determine the rhythm of the music.

[FIGURE 5]

7.2 Simulating the emergence of phonological systems

The results of the CA experiments are encouraging from a pragmatic compositional point of view, but they still do not shed much light on the problem of origins. In simple terms, the CAMUS model generates music in a particular style but it does not elicit the mechanism of its own origins. In order to address this problem we need to apply a modelling paradigm where phenomena, in our case musical styles, might emerge autonomously. Hence the potential of *evolutionary modelling*.

The *imitation game* is one such evolutionary modelling paradigms which has been successfully used to study the emergence of phonological systems in virtual communities of agents (Boer, 1997a). The model for the phonological experiments is composed of a virtual community of agents furnished with a speech synthesiser, an artificial ear and a memory mechanism. (These experiments were originally designed and performed by Bart de Boer, a visiting researcher from the Artificial Intelligence Laboratory of the Free University of Brussels.)

The *speech synthesiser* requires three articulatory parameters to produce sounds: a) the position of the highest part of the tongue in the front to back dimension, b) the vertical distance between the highest part of the tongue and the roof of the mouth and c) the position of the lips (or rounding of the lips). The spectral contour of speech-like sounds has the appearance of a pattern of hills and valleys technically called formants (Miranda, 1998). Speech sounds normally have four or five prominent formants whose central frequencies are crucial for the ear to distinguish between different sounds (Figure 6). The agent s *artificial ear* is able to depict the central frequencies of the formants of a sound. The *memory mechanism* stores the repertoire of sounds that the agent knows.

This knowledge is stored in terms of articulatory parameter values associated with central formant frequency values. The memory also holds other types of information; for example, for each sound of their repertoire, the agents keep a record of how many times it was used and how many times it lead to a successful interaction.

To begin with, the agents do not have any repertoire in memory. The objective of the game is to build their own repertoire of sounds by imitating each other. The only constraints of the game are imposed by the physiology of their speech and hearing apparatus. The game proceeds as follows: two agents are randomly selected for a round; one will play the role of *initiator* and the other the role of *imitator*. The initiator produces a sound, either selected from its own repertoire or invented from scratch. The imitator in turn compares this sound with the ones it already knows. In order to make the comparisons the agents use a Euclidean distance formula to measure the distance between two signals (Boer, 1997b). The imitator should then select from its own repertoire the sound that is most similar to the sound it heard from the initiator and it should produce it. If the imitator does not yet have any sound in its repertoire, then it tries to guess a similar sound using a certain pre-defined approximation algorithm. Next, the initiator compares the imitator s sound with the one it produced originally. If the imitator s sound is similar to the initiator s sound then the game is a success, otherwise it is a failure. This result is communicated to the imitator by means of a non-verbal feedback.

Figure 6:
Formants are crucial for distinguishing different speech sounds.
The horizontal axis represents frequencies in Hz and
the vertical axis represents the amplitudes of the formants.

[FIGURE 6]

Immediately after the game, both agents update their memories according to the result of the game. To put it in simple terms, if the game was successful, then the agents increase the success counters for the sounds in question. If the game was a failure, then the imitator tries to shift its sound closer to the sound it heard, hoping that next time it will result in a better imitation. Sometimes sounds that have not scored successfully for a long while can be deleted from the repertoire. In other occasions, sounds that are too close to each other in the repertoire are merged. After each round, agents always add one new random sound each to their repertoire.

Due to the simplicity of the agents speech synthesiser and of their hearing model, these imitation games work only with vowel-like sounds. Nevertheless, the results of the game are strikingly impressive. After about 20,000 of runs involving no more than 20 agents, coherent vowel systems do emerge. Furthermore, these systems share the same characteristics of the vowel systems of real human languages. Figure 7 portrays a typical result from these games. All vowels of all agents involved in the game (e.g., 20) are plotted on top of each other; the vertical axis corresponds to the first formant s centre frequencies (F1) and the horizontal corresponds to the weighted sum of the second, third and fourth formants centre frequencies (F2). The frequency values are ploted using the Bark frequency scale (Boer, 1997b). Note that the agent s vowels form clusters around formant regions that correspond vowels.

The results from this phonological experiment are very encouraging because they also suggest that musical forms may indeed emerge in a community of virtual agents furnished with sufficient cognitive and sound production capacities. We are now gathering more evidence to corroborate our hypothesis that language and music systems share similar evolutionary principles. Although the phonological experiments were not explicitly designed for a musical context, the results implicitly indicate the emergence a musical pitch systems from imitation games.

Figure 7:
The vowel systems of twenty agents after 25,000 imitation games.

[FIGURE 7]

Figure 8:
A note system derived from formant centre frequency values.

[FIGURE 8]

In a recent paper (Miranda, 1999) we have demonstrated how a pitch system for musical composition can be naturally derived from the formant centre frequency values of a given vowel set (Figure 8). Figure 8 portrays a note system corresponding to the format centre frequency values for the first three formants (F1, F2, and F3) of five vowels (/a/, /e/, /i/, /o/ and /u/) as sung by two types of synthesised voices (M for male-like tones and F for female-like tones). In this system, the notes that would theoretically resonate better for a particular vowel are indicated at the bottom line (F0); they are one octave below the first formant value. As an experiment, we synthesised a great number of melodies for the words of a Latin mass (e.g., Benedictus qui venit in nomine Domini , Kyrie eleison , etc.) using randomly generated notes. We observed that those melodies containing a great number of syllables whose vowels matched their respective F0 notes (e.g., Be = 196 Hz dic = 174.61 Hz, etc.) tended to sound more natural than those melodies with very few matchings. But we also observed that occasionally, those melodies containing specific configurations of matching vowels with non-matching ones tended to form coherent musical prosodies. This observation needs more empirical validation, but we suspect that this may be evidence that mechanisms for the formation of simple musical discourse, based upon the notion of consonance (good resonance) versus dissonance (poor resonance), may emerge naturally from the interactions in a community of virtual agents.

8 Conclusion

Musicology can benefit hugely from Alife research. Alife researchers have developed a number of tools for studying complexity, some of which are proving to be very useful for a variety of fields other than Biology.

As we are interested in studying the origins of music, Alife paradigms such as cellular automata and evolutionary modelling have been very influential in our research. We are currently using evolutionary modelling techniques to investigate how musical forms may originate and evolve in artificially created worlds inhabited by virtual communities of musicians and listeners.

In this essay we have glanced at historical and philosophical concepts of great importance for our work, with emphasis on the notion of language games, action-based meaning formation in language, the evolution of human intelligence and some eighteenth century writings on the origins of language and music. In this overview we have stressed the importance of the interplay between language and music in our research; we have suggested that, for humans, our musical abilities are crucial for the development of our linguistic abilities, and vice-versa. We then introduced the fundamentals of evolutionary modelling, followed by an introduction to the pioneering experiments that are being performed at the Sony Computer Science Laboratory in Paris.

By gaining a better understanding of the origins of music we hope to not only enrich our know-how for the design of better computer music systems and interfaces, but also to broaden our understanding of human intelligence and to contribute to the shaping of the musical praxis of the twenty-first century.

References

Arbo, A., La trace du son. Expression et intervalle chez Condillac , *L espace: Musique/Philosophie*, Chouvel, J-M. and Solomos, M. (eds.), Paris: Editions L Harmattan, 1998.

Arita, T. and Koyama, Y., "Evolution of Linguistic Diversity in Simple Communication System", *Artificial Life*, Vol. 4, pp. 109-124, 1998.

Bamberguer, J., *The Mind Behind the Musical Ear: How Children Develop Musical Intelligence*, Cambridge (MA): Harvard University Press, 1991.

Best, C. T., "The emergence of cerebral asymetries in early human developement: A literature review and a neuroembryological model", *Brain Lateralization in Children*, 5. Molfese, D. L. and Segalowitz, S. J. (eds.), New York: The Guildford Press, 1988.

Bindon, J., *Can Chimps Talk?*, http://www.as.ua.edu/ant/bindon/ant101/syllabus/chmp_tlk/chmp_tlk.htm, 1998.

Boer, B. de, Generating vowel systems in a population of agents , *Fourth European*

Conference on Artificial Life - Brighton, Husbands, P. and Harvey, I. (eds.), Cambridge (MA): The MIT Press, 1997a.
Boer, B. de, *A Second Report on Emergent Phonology*, AI-MEMO 97-04, Vrije Universiteit Brussel, AI-Lab, Brussels, 1997b.
Calvin, W. H., "A Stone's Throw and its Launch Window: Timing Precision and its Implication for Language and Hominid Brains", *Journal of Theoretical Biology*, No. 104, pp. 121-135, 1983.
Capmbell, M. and Greated, C., *The Musician s Guide to Acoustics*, London/Melbourne: Dent & Sons Ltd, 1987.
Casti, J., "Information versus matter", *Artificial Life and Robotics*, 1(1), 1997.
Chomsky, N., "Rules and Representation", *Brain and Behavior Sciences*, Vol. 3, pp. 1-15, 1980.
Cood, E. F., *Cellular Automata*, London: Academic Press, 1968.
Deacon, T., *The symbolic species*, New York: W. Norton and Co., 1997.
Epstein, J. M. and Axtell, R., *Growing Artificial Societies*, Cambridge (MA): The MIT Press, 1996.
Ermentrout, G. B. and Edelstein-Keshet, L., "Cellular Automata Approaches to Biological Modeling", *Journal of Theoretical Biology*, No. 160, 1993.
Gardner, R. A. and Gardner, B. T., "A Cross-Fostering Laboratory", *Teaching Sign Language to Chimpanzees*, R. A. Gardner et al. (eds.), pp. 1-28. Albany: State University of New York Press, 1989.
Jusczyk, P. W., *The discovery of spoken language*, Cambridge (MA): The MIT Press, 1997.
Koza, J., *Genetic Programming: On the Programming of Machines by Means of Natural Selection*, Cambridge (MA): The MIT Press, 1992.
Ladefoged, P. and Maddieson, I., *The Sounds of the World s Languages*, Oxford (UK): Blackwell, 1996.
Langton, C. G. and Shimohara, K. (eds.), *Artificial Live V: Proceedings of the Fifth International Workshop on the Synthesis and Simulation of Living Systems,* Cambridge (MA): The MIT Press, 1997.
Lieberman, P., *The biology and evolution of language*, Cambridge (MA): Harvard University Press, 1984.
Lieberman, P. *Eve Spoke*, London: Picador/Macmillan, 1998.
Maddieson, I., *Patterns of Sounds*, Cambridge (UK): Cambridge University Press, 1984.
McAlpine, K., Miranda, E. R. and Hoggar, S., "Composing Music with Algorithms: A Case Study System", *Computer Music Journal*, 23(2), 1999.
McIntyre, A., "Babel: a testbed for research in origins of language*", Proceedings of COLING-ACL 1998*, Montreal, 1998.
Miranda, E. R., Cellular Automata Music: An Interdisciplinary Project , *Interface / Journal of New Music Research*, 22(1), pp.3-21, 1993.
Miranda, E. R., "Music composition using cellular automata", *Languages of Design*, Vol. 2, 1994.
Miranda, E. R., "Granular Synthesis of Sounds by Means of a Cellular Automata",

Leonardo, 28(4), 1995a.
Miranda, E. R., Cellular Automata Synthesis of Acoustic Particles , *Supercomputer*, No. 56, 1995b.
Miranda, E. R., *Computer Sound Synthesis for the Electronic Musician*, Oxford (UK): Focal Press, 1998.
Miranda, E. R., *Quarteto de Cordas: Wee Batucada Scotica*, musical score and parts, Porto Alegre: Edi es Musicais Goldberg, 1998.
Miranda, E. R., The role of speech synthesis in *Requiem per una veu perduda* , *Organised Sound*, 3(3), 1999.
Monelle, R., *Linguistics and Semiotics in Music*, Reading (UK): Harwood Academic Publishers, 1992.
Olson, E. T., "The Ontological Basis of Strong Artificial Life", *Artificial Life*, 3(1), pp. 29-39, 1997.
Premack, D, "Rate differential reinforcement in monkey manipulation", *Journal of the Experimental Analysis of Behaviour*, No. 6, 1963.
Premack, D., *Why Chimps Can Read*. New York: Harper & Row, 1976.
Reck, D., *Music of the Whole Earth*, New York: Da Capo Press, 1997.
Storr, A., *Music and the Mind*, London: Harper Collins, 1993.
Terrace, H. S., *Nim*, New York: Alfred A. Knopf, 1979.
Thomas, D. A., *Music and the origins of language*, Cambridge (UK): Cambridge University Press, 1995.
Ullmann, R. A., *Antropologia Cultural*, Porto Alegre: Escola Superior de Teologia S o Louren o de Brindes, 1983.
Savage-Rumbaugh, S. and Lewin, R., *Kanzi: The Ape at the Brink of the Human Mind*, New York: John Wiley & Sons, 1994.
Steels, L. and Brooks, R., (eds.) *The Artificial Life Route to Artificial Intelligence: Building Embodied Situated Agents*, Hillsdale (NJ): Lawrence Erlbaum Associates, 1995.
Steels, L. and Vogt, P., Grounding adaptive language games in robotic agents , *Fourth European Conference on Artificial Life - Brighton*, Husbands, P. and Harvey, I. (eds.), Cambridge (MA): The MIT Press, 1997.
Steels, L., "The Spontaneous Self-Organization of an Adaptive Language*", Machine Intelligence 15*, Muggleton, S. (ed.), Oxford: Oxford University Press, 1997a.
Steels, L., "The Synthetic Modeling of Language Origins", *Evolution of Communication Journal*, 1(1), 1997b.
Steels, L., "Synthesising the origins of language and meaning using co-evolution, self-organisation and level formation", *Evolution of Human Language*, Hutford, J. (ed.), Edinburgh: Edinburgh University Press, 1997c.
Steels, L., Constructing and Sharing Perceptual Distinction , *Proceedings of the European Conference on Machine Learning*, Someren, M. and Widmer, G. (eds.), Berlin: Springer-Verlag, 1997d.
Washabaugh, W., "The Politics of Passion: Flamenco, Power, and the Body", *Journal of Musicological Research*, Vol. 15, pp. 85-112, 1995.
Wilson, G. The life and times of cellular automata , *New Scientist*, 8 October 1998.

Wittgenstein, L., *Philosophical Investigations*, Oxford (UK): Basil Blackwell, 1963.
Wolfram, S., *Cellular Automata and Complexity*, Reading (MA): Addison-Wesley, 1994.