On Computational Models of the Evolution of Music: From the Origins of Musical Taste to the Emergence of Grammars

Eduardo Reck Miranda, Simon Kirby and Peter M. Todd

Evolutionary computing is a powerful tool for studying the origins and evolution of music. In this case, music is studied as an adaptive complex dynamic system and its origins and evolution are studied in the context of the cultural conventions that may emerge under a number of constraints (e.g. psychological, physiological and ecological). This paper introduces three case studies of evolutionary modelling of music. It begins with a model for studying the role of mating-selective pressure in the evolution of musical taste. Here the agents evolve "courting tunes" in a society of "male" composers and "female" critics. Next, a mimetic model is introduced to study the evolution of musical expectation in a community of autonomous agents furnished with a vocal synthesizer, a hearing system and memory. Finally, an iterated learning model is proposed for studying the evolution of compositional grammars. In this case, the agents evolve grammars for composing music to express a set of emotions.

KEYWORDS: origins of music, evolution of musical taste, imitation, sensory-motor mapping, evolution of grammar

I. Introduction

Evolutionary computing (EC) may have varied applications in music. Perhaps the most interesting application is for the study of the circumstances and mechanisms whereby musical cultures might originate and evolve in artificially created worlds inhabited by virtual communities of software agents. In this case, music is studied as an adaptive complex dynamic system; its origins and evolution are studied in the context of the cultural conventions that may emerge under a number of constraints, including psychological, physiological and ecological constraints. Music thus emerges from the overall behaviour of interacting autonomous elements.

A better understanding of the fundamental mechanisms of musical origins and evolution is of great importance for musicians looking for hitherto unexplored ways to create new musical works. As with the fields of acoustics (Rossing 1990), psychoacoustics (Howard and Angus 1996) and artificial intelligence (Balaban *et al.* 1992; Miranda 2000), which have greatly contributed to our understanding of music, EC has the potential to reveal a new perspective from which music can be studied.

The pursuit of the origins of music is not new; philosophers throughout the ages have addressed this problem. As an example, we cite the book *Music and the*

Contemporary Music Review
ISSN 0749-4467 print/ISSN 1477-2256 online © 2003 Taylor & Francis Ltd
http://www.tandf.co.uk/journals
DOI: 10.1080/0749446032000150915

Origins of Language, by Downing Thomas (1995) as an excellent review of the theories purported by philosophers of the French Enlightenment. And, more recently, *The Origins of Music*, edited by Nils Wallin *et al.* (2000), collates a series of chapters written by top contemporary academics. With the exception of one chapter (Todd 2000), however, none of these thinkers sought theoretical validation through computer modelling. Although we are aware that musicology does not need such support to make sense, we do think, however, that computer simulation can be useful for developing and demonstrating specific musical theories. The questions that have been addressed by EC-based musicology overlap with those considered in evolutionary linguistics (Cangelosi and Parisi 2001; Christiansen and Kirby 2003): "What functional theories of its evolutionary origins makes sense?", "How do learning and evolved components interact to shape the musical culture that develops over time?" and "What are the dynamics of the spread of musical memes through a population?", to cite but three.

The paper begins by introducing a model for studying the role of mate-choice as a selective pressure in the evolution of musical taste. Next, a mimetic model is introduced for studying the evolution of musical expectation in a community of autonomous agents furnished with a vocal synthesizer, a hearing apparatus and a memory device. Finally, an iterated learning model is proposed for studying the evolution of compositional grammars. In this case, the agents evolve grammars for composing music to express a set of emotions.

II. Mate-choice and the Origins of Musical Taste

Peter Todd and Gregory Werner (1999) proposed a model for studying the role of mating-selective pressure in the origins of musical taste. Inspired by the notion that some species of birds use songs to attract a partner for mating, the model employs mating-selective pressure to foster the evolution of fit composers of courting tunes. The model co-evolves "male" composers, who play simple musical tunes, along with "female" critics, who judge these tunes and decide with whom to mate in order to produce the next generation of composers and critics.

Each composer holds a tune of thirty-two musical notes from a set of twenty-four different notes spanning two octaves. The critics encode a Markov chain that rates the transitions from one note to another in a heard tune. The chain is a 24 × 24 matrix, where each entry represents the female's expectation of the probability of one pitch following another in a song. Given these expectations, a critic can decide how well she likes a particular tune in one of a few ways. When she listens to a composer, she considers the transition from the previous pitch to the current pitch for each note of the tune, gives each transition a score based on her transition table, and adds those scores to come up with her final evaluation of the tune. Each critic listens to the tunes of a certain number of composers who are randomly selected, and all critics hear the same number of composers. After listening to all the composers in her courting choir, the critic selects as her mate the composer who produces the tune with the highest score. In this selective process all critics will have exactly one mate, but a composer may have a range of mates from none to many, depending on whether his tune is unpopular with everyone, or if he has a song that is universally liked by the critics. Each critic has one child per generation created via crossover and mutation with her chosen mate. This child will have a mix of the musical traits and preferences encoded in its mother and father. The sex of the child is randomly determined and a third of the population is removed at random after a mating session in order not to reach a population overflow.

From the many different scoring methods proposed for judging the tunes, the one that seems to produce the most interesting results is the method whereby critics enjoy being surprised. Here the critic listens to each transition in the tune individually, computes how much she expected the transition, and subtracts this value from the probability that she attached to the transition she most expected to hear. For example, if a critic has a value 0.8 stored in her Markov chain for the A–E transition, whenever she hears a note A in a tune, she would expect a note E to follow it 80 per cent of the time. If she hears an A–C transition, then this transition will be taken as a surprise because it violates the A–E expectation. A score is calculated for all the transitions in the tune and the final sum registers how much surprise the critic experienced; that is, how much she likes the tune. What is interesting here is that this does not result in the composers generating random tunes all over the place. It turns out that in order to get a high surprise score, a tune must first build up expectations, by making transitions to notes that have highly anticipated notes following them, and then violate these expectations, by not using the highly anticipated note. Thus there is constant tension between doing what is expected and what is unexpected in each tune, but only highly surprising tunes are rewarded (figure 1).

The composers are initiated with random tunes and the critics with Markov tables set with probabilities calculated from a collection of folk-tune melodies. Overall, this model has shown that the selection of co-evolving male composers who generate *surprising tunes*, and female critics who assess these tunes according to their preferences, can lead to the evolution of tunes and the maintenance and continual turnover of tune diversity over time.

This model is remarkable in the sense that it demonstrates how a Darwinian system with a survival imperative can initiate the evolution of coherent repertoires of melodies in a community of software agents. There is, however, a puzzling fundamental question that has not been addressed in this model: Where do the

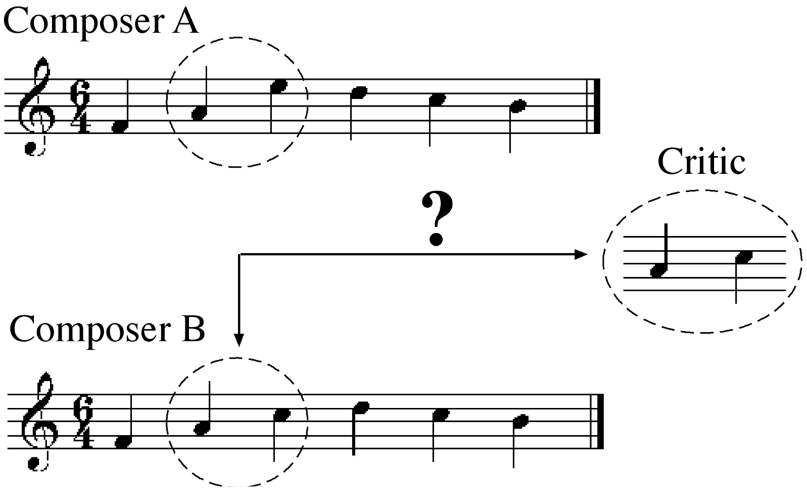


Figure 1
The critic selects composer B because it produces the most surprising tune.

expectations of the female critics come from? Currently, the model initializes their reference tables with coefficients computed from samples of existing folk-tune melodies. Would it be possible to evolve such expectations from scratch? A model that may provide support for addressing this question is introduced next.

III. Imitation and the Origins of Musical Expectation

Eduardo Miranda (2002a) proposed a *mimetic model* to demonstrate that a small community of interactive distributed agents furnished with appropriate motor, auditory and cognitive skills can evolve a shared repertoire of melodies (or tunes) from scratch, after a period of spontaneous creation, adjustment and memory reinforcement. In this case, expectation is defined as a sensory-motor problem, whereby agents evolve vectors of motor control parameters to produce imitations of heard tunes. The agents thus expect to hear pitch sequences that correspond to their evolved motor vectors.

Tunes are not coded in the genes of the agents and the agents do not reproduce or die, as in the previous models, but the agents are programmed with two motivations: (a) to form a repertoire of tunes in their memories; and (b) to foster social bonding. Both motivations are complementary because, in order to be sociable, an agent must form a repertoire that is similar to the repertoire of its peers. Sociability is therefore assessed in terms of the similarity of the agents' repertoires. In addition to the ability to produce and hear sounds, the agents are born with a basic instinct: to *imitate* what they hear.

The agents are equipped with a voice synthesizer, a hearing apparatus, a memory device and an enacting script. The voice synthesizer is essentially implemented as a physical model of the human vocal mechanism (Boersman 1993; Miranda 2002b). The agents need to compute three vectors of parameters in order to produce tunes: lung pressure, the width of the glottis, and the length and tension of the vocal chords – *lung_pressure*(*n*), *interarytenoid*(*n*) and *cricothyroid*(*n*) respectively. As for the hearing apparatus, it employs short-term, autocorrelation-based analysis to extract the pitch contour of a spoken signal. The algorithm features a parameter that regulates the resolution of the hearing apparatus, by controlling the resolution of the short-term autocorrelation analysis (Miranda 2001), defining the sensitivity of the auditory perception of the agents.

The agent's memory stores its repertoire of tunes and other data such as probabilities, thresholds and reinforcement parameters. An agent processes and stores tunes in terms of synthesis and analysis parameters. They have a dual representation of tunes in their memories: a *motor map* (synthesis) and a *perceptual representation* (analysis). The motor representation is in terms of vectors of motor (i.e. synthesis) parameters and the perceptual representation is in terms of an abstract scheme we designed for representing melodic contour derived from auditory analyses; refer to Appendix I.

Imitation is defined as the task of hearing a tune and activating the motor system to reproduce it (figure 2). When we say that the agents should evolve a shared repertoire of tunes, we mean that the perceptual representation in the memory of the agents of the community should be identical, but the motor representation may be different. An important presupposition in this model is that the action of singing tunes involves the activation of certain vocal motor mechanisms in specific ways. The recognition of tunes here therefore requires knowledge of the activation of the

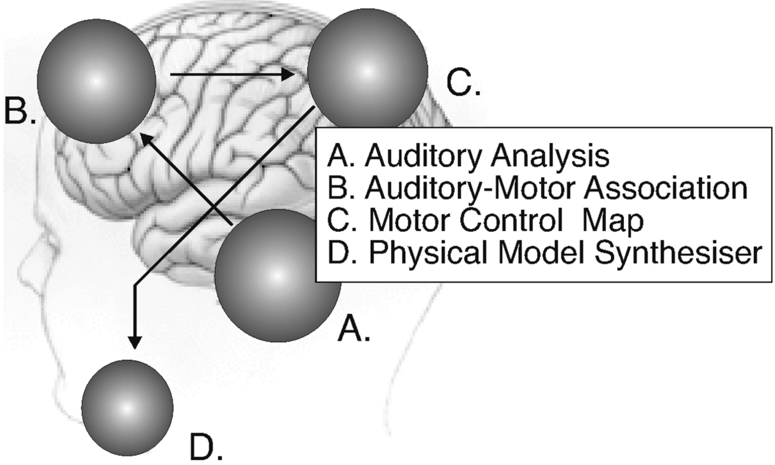


Figure 2
Imitation is defined as the task of hearing a tune and activating controls of the vocal synthesizer in order to reproduce it.

right motor parameters (i.e. synthesis parameter values) in order to reproduce the tune in question.

III.i. The Enacting Script

The enacting script provides the agent with knowledge of how to behave during the interactions: the agent must know what to do when another agent produces a tune, how to assess an imitation, when to remain quiet, and so forth. The enacting script does not evolve in the present model; all agents are alike in this respect. Also, all agents have identical synthesis and listening apparatus. At each round, each of the agents in a pair from the community plays one of two different roles: the *agent-player* or the *agent-imitator*; the main algorithm is given in Appendix II. In short, the agent-player starts the interaction by producing a tune p_1, randomly chosen from its repertoire. The agent-imitator then analyses the tune p_1, searches for a similar tune in its repertoire (p_2) and produces it. The agent-player in turn analyses the tune p_2 and checks if its repertoire holds no other tune, p_n, that is more perceptibly similar to p_2 than p_1 is. If it finds another tune, p_n, that is more perceptibly similar to p_2 than p_1 is, then the imitation is unsatisfactory, otherwise it is satisfactory. If it is satisfactory, then the agent-imitator will reinforce the existence of p_2 in its memory. If unsatisfactory, the agent has to choose between two potential courses of action. If it finds out that p_2 is a weak tune (i.e. low past success rate) in its memory, because it has not received enough reinforcement in the past, then it will try to modify its representation of p_2 slightly, as an attempt to further approximate it to p_1. It is hoped that this approximation will give the tune a better chance of success if it is used again in another round. But if p_2 is a strong tune (i.e. good past success rate), then the agent will leave p_2 untouched (because it has been successfully used in previous imitations and a few other agents in the community probably know it too), will create a new tune that is similar to p_1, and will include it in its repertoire. Before terminating the round, both agents perform final updates. First, they scan their repertoire and merge those tunes that are considered to be perceptibly identical to each other. Also, at the end of each round, both agents have a certain probability, P_b, of undertaking a spring-cleaning to get rid of weak tunes; those tunes that have not been sufficiently reinforced are forgotten. Finally, at the end of each round, the agent-imitator has a certain probability, P_a, of adding a new randomly created tune to its repertoire.

III.ii. A Typical Example

The graph in figure 3 shows a typical example of the evolution of the average repertoire of a community of five agents, with snapshots taken after every 100 interactions over a total of 5000. Note that the agents quickly increase their repertoire to an average of between six and eight tunes per agent. At about 4000 interactions, more tunes appear, but at a lower rate. The general tendency is to settle quickly into a repertoire of a certain size, which occasionally increases at lower rates. The pressure to increase the repertoire is mostly due to the probability P_a of creating a new random tune, combined with the rate of new inclusions due to imitation failures. In this case the repertoire settled to eight tunes between 1600 and 4000 interactions.

The graph in figure 4 plots the imitation success rate of the community, measured at every 100 interactions. At approximately 1800 interactions, the imitation rate goes back up to 100 per cent. Then, occasional periods of lower success rate occur owing to the appearance of new random tunes. Although the repertoire tends to increase with time, the success rate stays consistently high. This is good evidence that the community does manage to foster social bonding.

Figure 5a portrays the perceptual memory of a randomly selected agent, after

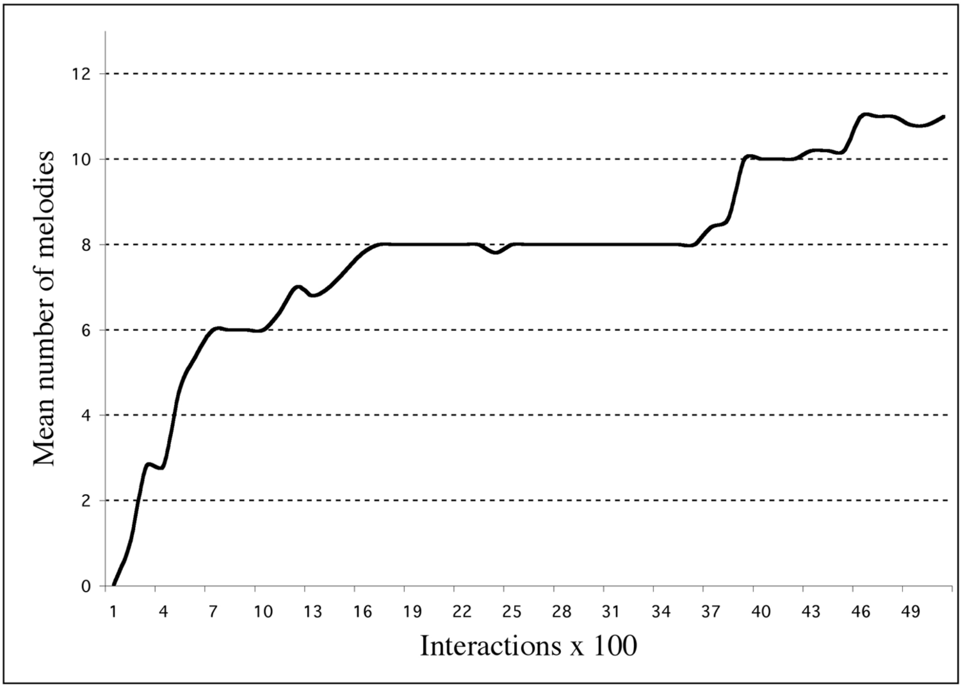


Figure 3
The evolution of the average size of the repertoire of tunes of the whole community.

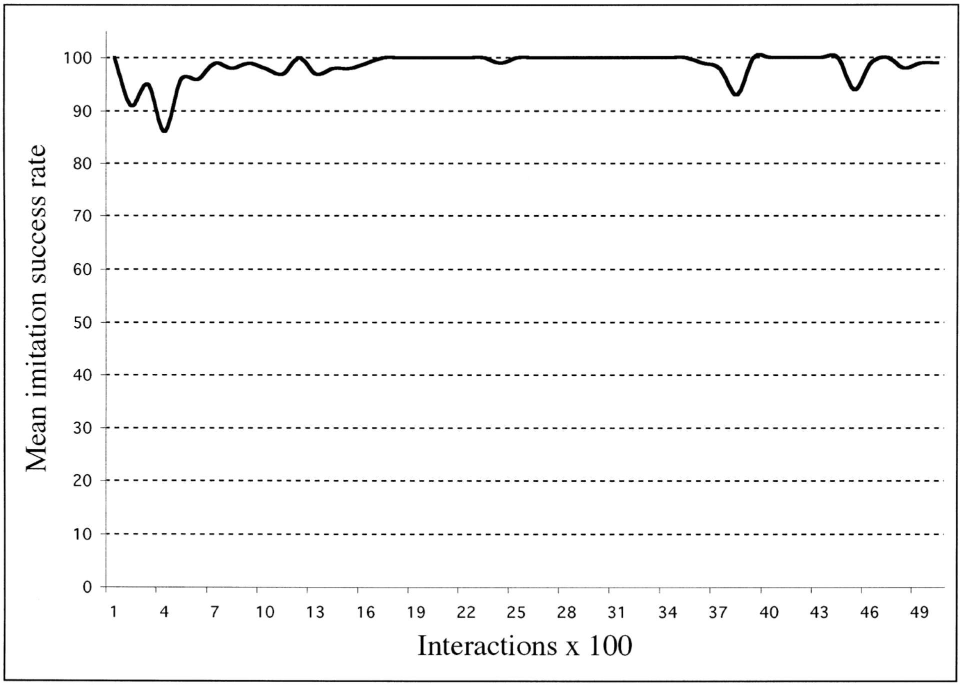


Figure 4
The imitation success rate over time.

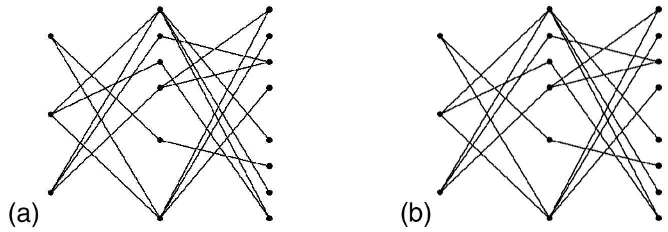


Figure 5
The perceptual memory of the agents.
For the sake of clarity, the background metrics and labels of the graph are not shown; refer to Appendix I.

5000 interactions. The repertoire of all five agents plotted on top of each other is shown in figure 5b. The latter demonstrates the agent whose memory is plotted in the former shares identical tunes with the whole community.

What makes this model interesting is that it does not assume the existence of a one-to-one mapping between perception and production. The agents learn by themselves how to correlate perception parameters (analysis) with production (synthesis) ones and they do not necessarily need to build the same motor representations for what is considered to be perceptibly identical (figure 6). The repertoire of tunes emerges from the interactions of the agents, and there is no global procedure supervising or regulating them; the actions of each agent are based solely upon their own evolving expectations.

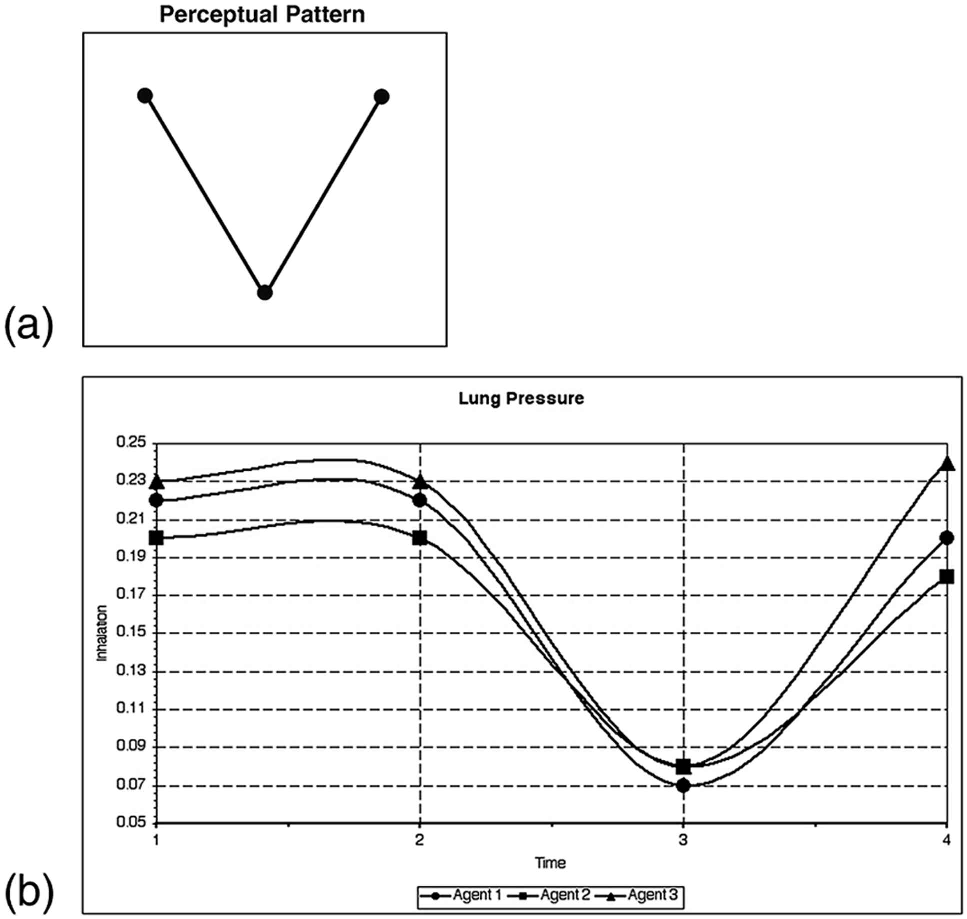


Figure 6
(a) One of the perceptual patterns from figure 5b and its corresponding motor control vectors developed by three different agents; (b) the *lung_pressure* vector.

The agents develop a close link between perceptual and motor representations, which allows them to enact the tune even if they cannot fully recognize it. In fact, the agents always think that they can fully recognize everything they hear. In order to produce an imitation, an agent will use the motor vectors that best match its perception. It is the other agent who will assess the imitation based on its own expectations. Musical expectation here is a social convention but it is grounded on the nature of their sensory-motor apparatus.

Both models so far deal with short tunes. But how about dealing with larger musical compositions? Surely, our brain does not represent every musical piece we know as explicitly as has so far been suggested. Musical intelligence certainly requires the ability to abstract rules about music, which in turn enables us to process larger and complex structures.

The following section presents a model whereby the agents evolve musical rules by a process called iterated learning. Complementary issues such as *emotion* and *semantics* are also addressed.

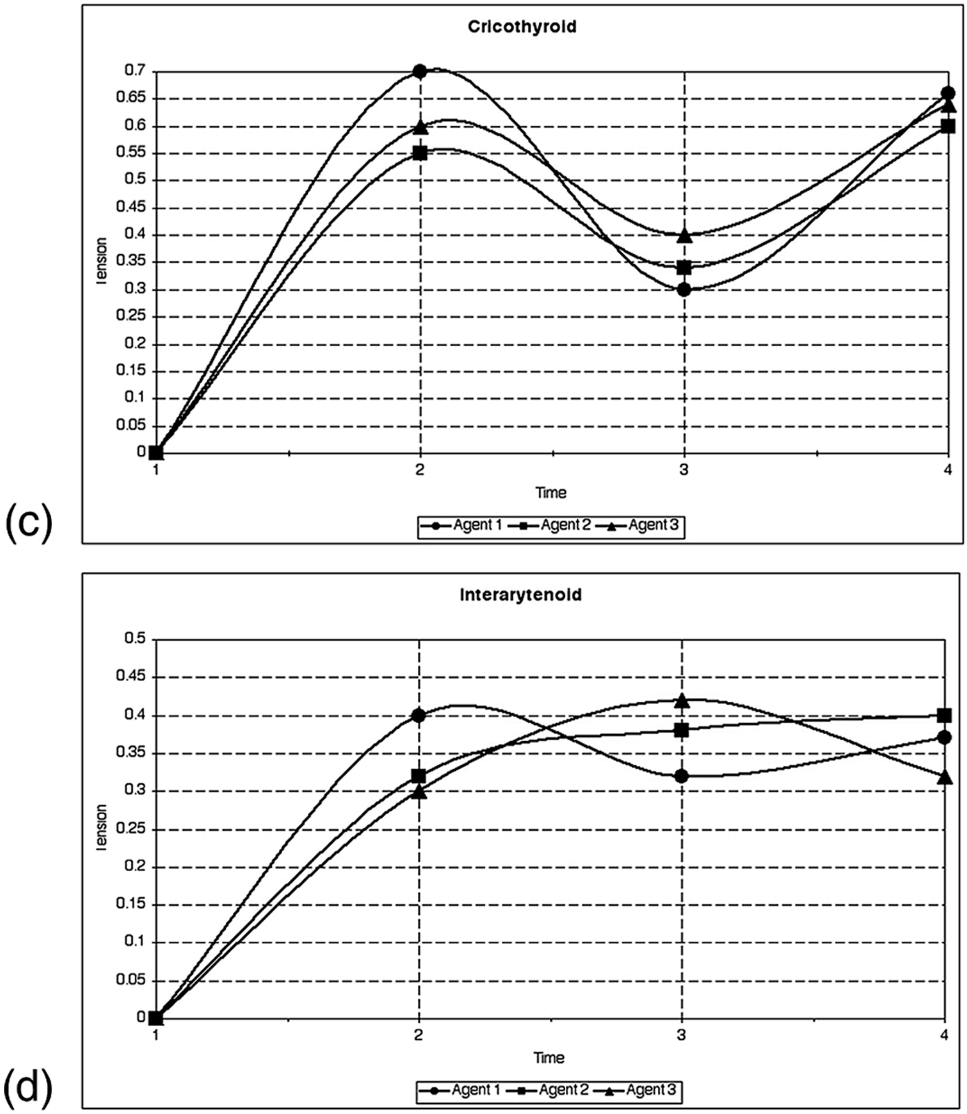


Figure 6
(Continued.)
(c) The *cricothyroid* vector; and (d) the *interarytenoid* vector.

IV. Cultural Transmission of Emotions and the Emergence of Musical Grammar

Simon Kirby and Eduardo Miranda have designed a model to study the emergence of musical grammars after Simon Kirby's successful simulations in the realm of the evolutionary linguistics (Kirby 2001, 2002). Like language, music is unique not only for its syntactic structure, but also in the way it evolves and preserves structure over time. In language, information about the mapping between meaning and signals is transmitted from generation to generation through a repeated cycle of use, observation and learning. Music is similarly culturally transmitted, with the difference that musical meaning is not so tightly coupled with its signals as linguistic meaning is. Leaving aside the mating-selective pressure and mimetic endowment of the previous models, the following paragraphs explore the idea that musical structure may emerge from the dynamics arising from this cycle of use, observation and learning. To this end, we have extended the Iterated Learning Model (ILM; Kirby 2001) to account for the cultural transmission of musical behaviour. There are four components in the ILM:

1. *A signal space.* In this paper, the signals take the form of musical sections, made up of sequences of riffs. (Riffs are short musical passages, or clichés, that are usually repeated various times in a piece of music.)
2. *A semantic space*. Musical meanings here are made up of emotions and moods, which can be combined to form more complex hierarchical (semantic) structures.
3. *Agent-teachers/-adults*. The adult agents use grammars to convey emotions and moods with music.
4. *Agent-learners/-children*. The learners induce grammars from their observation of the adults' musical behaviour.

In the simulations reported below, only one adult and one child were employed at any one time, although over the length of a simulation, many hundreds of agents will be involved. Each cycle of the ILM involves the adult agent being given a set of randomly chosen elements of the semantic space (i.e. meanings) to produce signals for. The resulting meaning/signal pairs form training data for the child agent. After learning, the child becomes a new adult agent, the previous adult is removed and a new child is introduced. This cycle of performance, observation and learning is repeated hundreds of times, and the internal representations of the agents are logged in addition to the musical behaviour of each adult.

Every agent is born knowing nothing about the musical culture it is born into. They must learn the rules for themselves by observing the performances of adults. We initialize the simulation without any musical culture at all, so the initial generations of adults must produce purely random riff sequences. After the adult has produced a number of pieces, it is removed and the child becomes the new adult. A new child is born into the simulation and the cycle is repeated.

Despite the fact that the simulation starts with entirely "non-musical" agents whose only means of expression is random sequences of notes, we quickly see a musical culture emerging in the simulations. Over many generations this culture becomes more and more complex – a structured system of expression evolves. The following sections go into more details of the set-up of the model and results.

IV.i. The Signal Space

The signal space, or symbol alphabet, of real musical cultures is largely determined by the musical instruments available. In this model, there is only one instrument available: the flute. The agents play a physical model of a flute implemented after one of the 9000-year-old bone flutes found in a Neolithic site in China.[1] The flute has seven holes, roughly corresponding to the notes A5, B5, C6, D6, E6, F♯6 and A6. Each symbol of the musical alphabet here aggregates two pieces of information: note and duration. There are six different duration values – very short note, short note, medium duration note, long note and very long note – represented as vs, s, m, l and vl, respectively. The musical alphabet thus comprises thirty-five different symbols (7 notes × 5 durations) as follows: a5vs, a5s, a5m, . . ., b5l, b5vl, and so on.

IV.ii. The Meaning Space

Musical systems all over the world have evolved riffs that, combined, form larger musical compositions. Examples of this can be found in musical styles as diverse as Indian drumming, Brazilian samba, Japanese *gagaku* and western pop music (Reck 1997). For instance, the traditional music of Iran uses a collection of over 200 skeletal melodies and phrases (*gusheh*), which a performer uses as a basis for improvisation. These riffs often have arbitrary labels (e.g. after mythology, cosmology, Gods, etc.) and some cultures may associate them with emotions. For instance, in western music, a riff in a minor scale is often associated with sadness; a fast rhythmic riff is normally associated with happiness, and so forth. It is possible the humans are biologically programmed to respond to sound patterns in specific ways, but we can safely assume that these associations are prominently cultural.

Meaning is fundamentally represented in this model as a combination of *riffs* and *emotion*s. Here, the agents are programmed to evolve nine riffs named after the names of the stars of the constellation *Eridanus* – one of the original forty-eight constellations first drawn by Ptolemy: *Achernar*, *Cursa*, *Zaurak*, *Rana*, *Azha*, *Acamar*, *Beid*, *Keid*, *Angetenar*.[2] A repertoire of twenty-four emotions distributed into eight different groups (table 1) has been defined after the work of psychologist Kate Hevner (1936).

A riff is associated with a particular emotion but it is not a composition per se. A composition is a combination of riffs. A composition, therefore, can evoke a number of different emotions.

For the simulations reported here, the semantic representations are of the following forms:

1. A combination of two riffs render a certain emotion: *emotion*(*riff*, *riff*). Example: satisfying (zaurak, rana).
2. A composite emotion can be defined recursively as follows: *emotion*(*riff*, (*emotion*(*riff*, *riff*))). Example: dreamy(azha, satisfying(zaurak, rana)).

A combination of two emotion-structures renders a certain mood. This gives us the highest-level semantic structure:

Table 1
Emotions are classified into eight different groups

Group 1	Spiritual, lofty, awe-inspiring, dignified, sacred, solemn, sober, serious
Group 2	Pathetic, doleful, sad, mournful, tragic, melancholy, frustrated, depressing, heavy, dark
Group 3	Dreamy, yielding, tender, sentimental, longing, yearning, pleading, plaintive
Group 4	Lyrical, leisurely, satisfying, serene, tranquil, quiet, soothing
Group 5	Humorous, playful, whimsical, fanciful, quaint, sprightly, delicate, light, graceful
Group 6	Merry, joyous, gay, happy, cheerful, bright
Group 7	Exhilarated, soaring, triumphant, dramatic, passionate, sensational, agitated, exciting, impetuous, restless
Group 8	Vigorous, robust, emphatic, ponderous, majestic, exalting

1. A combination of two emotion-structures renders a certain mood: *mood*(*emotion-structure, emotion-structure*). Example relaxing(satisfying(zaurak, rana), sad(azha, beid)).

There can be eight different moods, each associated to one of the groups in table 1: *religious*, *gloomy*, *romantic*, *relaxing*, *satirical*, *festive*, *passionate* and *martial*, respectively. In the simulation, a set of these mood-structures is generated at random for the adult agents to express. The learners listen to each of the compositions and it is these training data from which their own musical competence is built. In general with ILMs, we can instil upon the agents arbitrary preferences, which act to prune instances from these training data. In other words, we can define selection criteria that determine whether a particular performance is listened to by a learner or not. For the moment, we simply define a preference for mood-structures whose emotions are compatible with the mood being expressed. Consider that table 1 is circular, in the sense that group 1 follows group 8. Emotions are compatible with a mood if the groups they are in are within five rows of table 1; that is, two rows above and two below the row of the mood in question. For instance, the emotion *whimsical* is highly compatible with the mood *satirical* (group 5). It would be compatible with mood *passionate* (group 7; two rows below) but not with mood *gloomy* (group 2; three rows above). An important line of future research will be to combine these kinds of arbitrary preference with the evolutionary models by Peter Todd and Gregory Werner, discussed in section II.

Clearly, we could design experiments using a large range of different representations of musical meaning and musical preference. The important point is that the meanings have some kind of internal structure. It is access to this internal structure of emotions and moods that enables the formation of structure in the agents' behaviour.

IV.iii. The Learning Model

The algorithm of the learning model is identical to the one used for Simon Kirby's language-evolution simulations (Kirby 2001, 2002), so we will not go into great detail here. Essentially, an individual agent's competence for musical expression takes the form of a context-free grammar (Nijholt 1980) with category labels that have attached semantic expressions. It is simplest to demonstrate this with a couple of examples.

First, let us consider the knowledge of an agent that has listened to a few random

Table 2
Knowledge of an agent that has listened to a few random compositions

Input	
lyrical(achernar, rana)	b5m e6vl d6vl
lyrical(achernar, cursa)	b5m e6vs c6l
quaint(keid, bright(rana, cursa))	b5l fs6vs b5l
Grammar	
C/lyrical(achernar, rana) ?	b5m e6vl d6vl
C/lyrical(achernar, cursa) ?	b5m e6vs c6l
C/quaint(keid, bright(rana, cursa)) ?	b5l fs6vs b5l

Table 3
Generalization of the *achernar* riff

C/lyrical(x, rana) ?	A/x e6vl d6vl
C/lyrical(x, cursa) ?	A/x e6vs c6l
A/achernar ?	b5m

compositions (table 2). For simplicity, we are dealing only with emotion-structures here; the complete mood compositions simply combine two of these.

In the case of table 2, the learner is simply storing a memory of each composition heard. The learning algorithm works by storing experience in this way but additionally looking for ways in which multiple rules can be compressed into smaller sets of rules by generalization. For example, the first two rules in table 2 could lead to a generalization that the *achernar* riff could be expressed using a single note b5m. The corresponding grammar fragment could look as in table 3.

Each grammar rule in table 2 consists of a category, a meaning, and a sequence of notes. The category C defines a complete composition, so this grammar simply stores each of the three compositions the learner has heard. The grammar fragment now in table 3 has a new category, arbitrarily named "A". The right-hand sides of the C rules both refer to this category, rather than the sequence of notes directly. In addition, a variable "x" is introduced to stand in for the meaning of this component of the composition.

In this way, the learner can exploit any regularity in the music it hears to generate grammars that it, in turn, will use to produce music. When the music is unstructured, and the grammars are simply lists of compositions for specific meanings, the learner will only be able to reproduce those specific compositions. When the music has structure, however, the learners may be able to generalize and produce compositions for meanings that they have never heard expressed before. In other words, agents with structured grammars are able to express creatively, as opposed to simply randomly.

The key result from our simulations is that this structured, creative, expressive musical culture can emerge spontaneously out of initial unstructured, limited and random behaviour.

IV.iv. Running the Simulation

The simulation run proceeds as follows:

1. Start with one learner and one adult performer each with empty grammars.
2. Choose a meaning at random.
3. Get adult to produce composition for that meaning. The agent may need to invent random note-sequence of up to three notes if its grammar does not have a way of producing a composition for that meaning.
4. Relay the meaning-composition pair to learner.
5. Repeat steps 2–4 one hundred times with simple moods (no embedding), and then one hundred times with moods containing nested emotions of depth 1.
6. Delete adult performer.
7. Make learner be the new performer.
8. Introduce a new learner (with no initial grammar).
9. Repeat steps 2–8 indefinitely.

In figure 7, the results from one run of the simulation are plotted. In every simulation we see a movement from an initial random stage to a later stage where the music has a great deal of structure. This structure is apparent not only in the music that is produced, but also by examination of the grammars; refer to Appendix III.

Figure 8 shows two samples of the music produced early on in the simulation. There is a marked contrast between this unstructured system and the structured musical culture that has emerged by the end of the simulation, shown in figure 9.[3]

Every run of the simulation gives a different result, but the general movement from unstructured to structured systems is always evident. In some cases, the agents converge on a *recursive* musical system like the ones shown in bold in Appendix III. This system has limitless expressivity – the agents could, if they were prompted, express an unlimited range of arbitrarily complex meanings as music. This is not coded into the simulation, but arises because of the dynamics of cultural transmission.

The learners are constantly seeking out generalizations in their input. Once a generalization is induced, it will tend to propagate itself because it will, by definition, be used for more than one meaning. In order for any part of the musical culture to survive from one generation to the next, it has to be apparent to each learner in the randomly chosen 200 compositions each learner hears. A composition that is only used for one meaning and is not related to any other composition can only be transmitted if the learner hears that composition in its input. Musical structure, in the form of increasingly general grammar rules, results in a more stable musical culture. The learners no longer need to learn each composition as an isolated, memorized piece of knowledge. Instead, the learners can induce rules and regularities that they can then use to create new compositions that they themselves have never heard, yet still reflect the norms and systematic nature of the culture in which they were born.

V. Conclusion and Further Work

EC allows for the study of music as an adaptive complex dynamic system. In this context, the origins and evolution of music can be studied using computer models and simulations, whereby music emerges from the overall behaviour of the interacting agents. This paper introduced three case studies where interacting agents evolve repertoires of tunes and compositional grammars. In these case studies, the authors addressed fundamental issues concerning the origins of musical taste and expectation, and the expression of emotions and moods in music.

The case studies presented in this paper are clearly indications that musicology can benefit enormously from EC. The degree of sophistication and plausibility of EC-based musicology is proportional to the degree of complexity and the range of questions that can be realistically addressed. The natural progression for the work presented in this paper is the definition of a framework to combine these and possibly other models and simulations to form more complex and realistic scenarios.

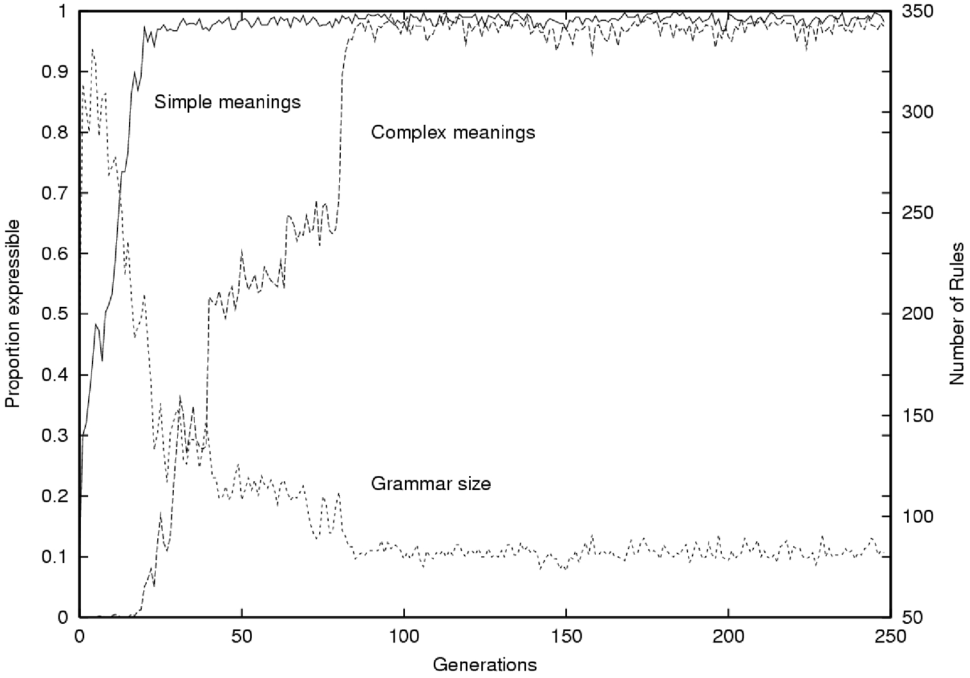


Figure 7
A run of the simulation showing cultural evolution of musical structure.
As the musical culture evolves more and more emotions become expressible and, at the same time, the size of the grammar goes down. Ultimately, the agents are able to express arbitrarily complex emotions by using a condensed, structured, recursive grammar.

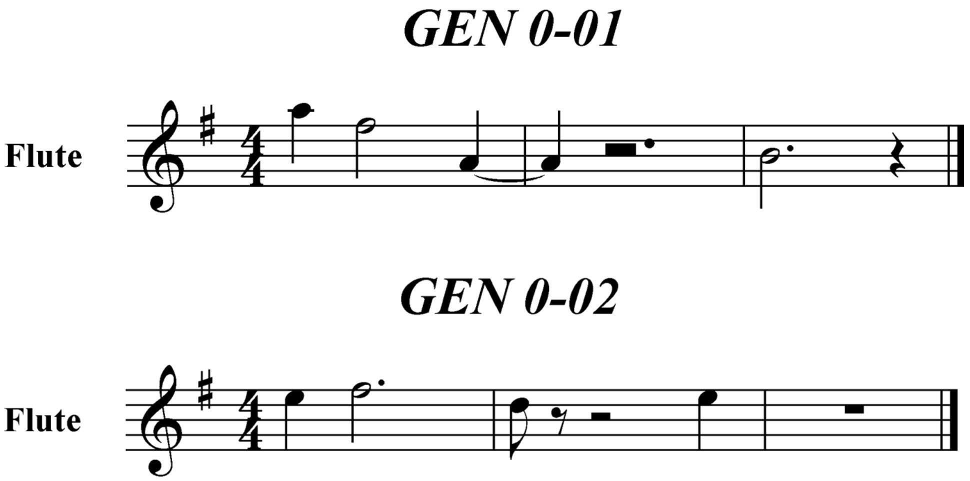
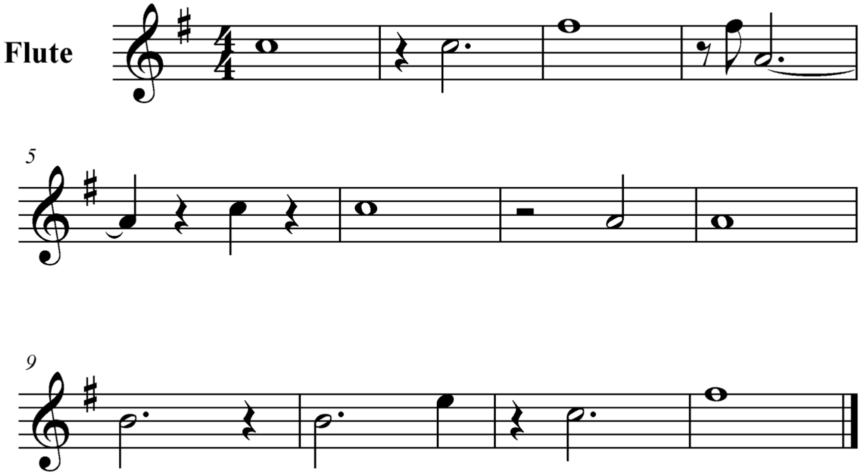


Figure 8
Two compositions by an agent early in the simulation.
The compositions *GEN 0-01* and *GEN 0-02* correspond to religious(exalting(azha, keid), majestic(keid, azha)) and religious(tragic(azha, beid), plaintive(azha, achernar)), respectively.

GEN 248-01



GEN 248-02

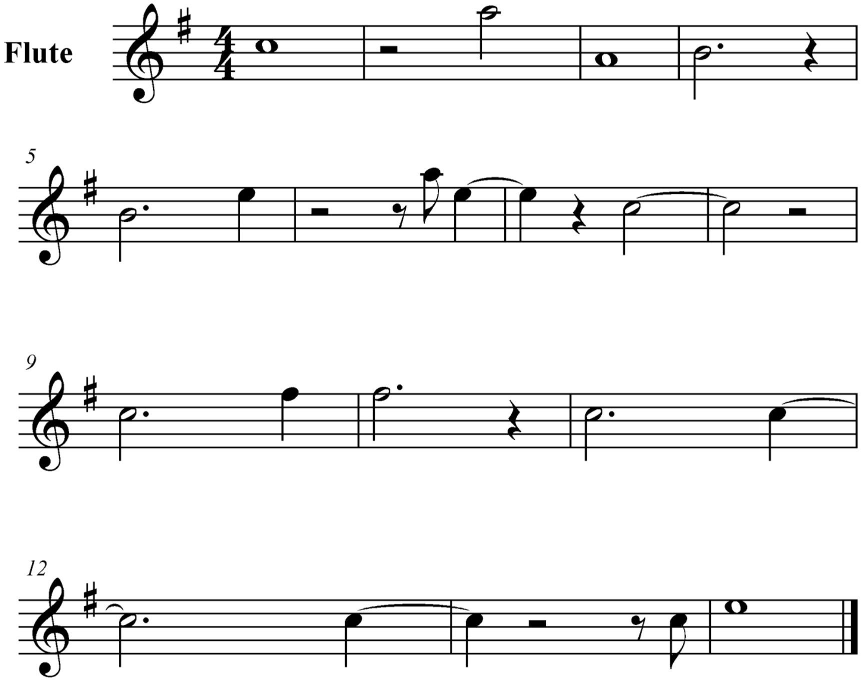


Figure 9
Two compositions produced by an agent late in the simulations.
The compositions *GEN 248-01* and *GEN 248-02* correspond to religious(tender(angetena, achernar), melancholy(achernar, acamar)) and religious(melancholy(keid, azha), quiet(cursa, achernar)), respectively.

References

Balaban, M., Ebcioglu, K. and Laske, O. (eds) (1992) *Understanding Music with AI*. Cambridge, MA: MIT Press.

Boersma, P. (1993) "Articulatory synthesizers for the simulations of consonants". In *Proceedings of Eurospeech'93*, pp. 1907–1910. Berlin, Germany.

Cangelosi, A. and Parisi, D. (eds) (2001) *Simulating the Evolution of Language*. London: Springer-Verlag.

Christiansen, M. H. and Kirby, S. (eds) (2003) *Language Evolution: The States of the Art*. Oxford: Oxford University Press.

Hevner, K. (1936) "Experimental studies of the elements of expression in music". *American Journal of Psychology* 48, 246–268.

Howard, D. M. and Angus, J. (1996) *Acoustics and Psychoacoustics*. Oxford: Focal Press.

Kirby, S. (2001) "Spontaneous evolution of linguistic structure: an iterated learning model of the emergence of regularity and irregularity". *IEEE Transactions on Evolutionary Computation* 5(2), 102–110.

Kirby, S. (2002) "Learning, bottlenecks and the evolution of recursive syntax". In *Linguistic Evolution Through Language Acquisition: Formal and Computational Models*, ed. T. Briscoe, pp. 173–203. Cambridge: Cambridge University Press.

Miranda, E. R. (ed.) (2000) *Readings in Music and Artificial Intellligence*. Amsterdam: Harwood Academic Publishers.

Miranda, E. R. (2001) "Synthesising prosody with variable resolution". In *AES Convention Paper 5332*. New York: Audio Engineering Society, Inc.

Miranda, E. R. (2002a) "Mimetic model of intonation". In *Music and Artificial Intelligence – Second International Conference ICMAI 2002*, Lecture Notes on Artificial Intelligence 2445, ed. Christina Anagnostopoulou, Miguel Ferrand and Alan Smaill, pp. 107–118. Berlin: Springer-Verlag.

Miranda, E. R. (2002b) *Software Synthesis: Sound Design and Programming*, 2nd edn. Oxford: Focal Press.

Nijholt, A. (1980) *Context-Free Grammars: Covers, Normal Forms, and Parsing*. Berlin: Springer-Verlag.

Reck, D. (1997) *The Music of the Whole Earth*. New York: Da Capo Press.

Rossing, T. D. (1990) *The Science of Sound*, 2nd edn. Reading, MA: Addison-Wesley.

Thomas, D. A. (1995) *Music and the Origins of Language*. Cambridge: Cambridge University Press.

Todd, P. M. (2000) "Simulating the evolution of musical behavior". In *The Origins of Music*, ed. Nils Wallin, Bjorn Merker and Steven Brown, pp. 361–388. Cambridge, MA: MIT Press.

Todd, P. M. and Werner, G. M. (1999) "Frankensteinian methods for evolutionary music composition". In *Musical Networks: Parallel Distributed Perception and Performance*, ed. Neil Griffith and Peter M. Todd, pp. 313–339. Cambridge, MA: MIT Press/Bradford Books.

Wallin, N. J., Merker, B. and Brown, S. (eds) (2000) *The Origins of Music*. Cambridge, MA: MIT Press.

Notes

[1] See BBC News: http://news.bbc.co.uk/1/hi/sci/tech/454594.stm (accessed 14/10/03).

[2] See Peoria Astronomical Society: http://www.astronomical.org/constellations/eri.html (accessed 14/10/03).

[3] The musical notation in figures 8 and 9 are approximations of the pitches and durations of the actual sounds. The notes should be played two octaves higher than notated.

Appendix I. Abstract Representation of Melodic Contour

A melodic unit (MU) is represented as a graph whose vertices stand for initial (or relative) pitch points and pitch movements, and the edges represent a directional path. Whilst the first vertex must have one outbound edge, the last one must have only one incoming edge. All vertices in between must have one incoming and one outbound edge each. Vertices can be of two types, initial pitch points (referred to as *p-ini*) and pitch movements (referred to as *p-mov*) as follows (figure 10):

p-ini = {SM, SL, SH}
p-mov = {VLSU, LSU, MSU, SSU, RSB, SSD, MSD, LSD, VLSD}

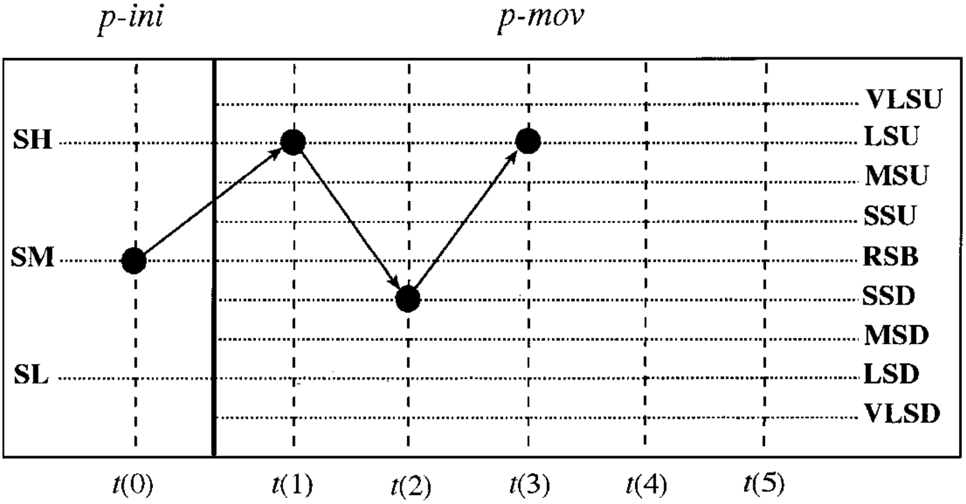


Figure 10
The representation of a melodic unit.

where

SM = start MU in the middle register
SL = start MU in the lower register
SH = start MU in the higher register

and

VLSU = very large step up
LSU = large step up
MSU = medium step up
SSU = small step up
RSB = remain at the same band
SSD = small step down
MSD = medium step down
LSD = large step down
VLSD = very large step down

An MU will invariably start with a *p-ini*, followed by one or more *p-movs*. It is assumed that an MU can start at three different voice registers: low (SL), middle (SM) and high (SH). Then, from this initial point the next pitch might jump or step up or down, and so forth.

It is important to note that labels or absolute pitch values are not relevant here because this scheme is intended to represent abstract melodic contours rather than a sequence of musical notes drawn from a specific tuning system.

Appendix II. The Main Algorithm of the Enacting Script

Agent-player (AP)	*Agent-imitator (AI)*

{ IF *repertoire(AP)* not empty:
pick motor control for p_d;
produce p_d;
ELSE
generate random motor control for p_d;
add p_d to *repertoire(AP);*
produce p_d; }
	{ analyse p_d }
	{ build perceptual representation; }
	{ IF *rep(AI)* not empty
	i_n = most perceptually similar to p_d;
	ELSE
	generate random motor control for i_n;
	add i_n to *repertoire(AI);*
	produce i_n; }

{ analyse i_n; }
{ build perceptual representation; }
{ p_n = most perceptually similar to i_n; }
{ IF $p_n = p_d$
send positive feedback to AI;
reinforce p_d in *repertoire(AP);*
ELSE
send negative feedback to AI; }
	{ IF feedback = positive
	approximate i_n to p_d perceptually;
	generate appropriate motor control;
	reinforce i_n in *repertoire(AI);* }
	
	{ IF feedback = negative
	IF i_n scores good H_T;
	execute *add_new_similar(snd);*
	ELSE
	Modify motor representation of i_n towards p_d; }

{ execute *final_updates(AP);* }	{ execute *final_updates(AI);* }

The *add_new_similar()* function works as follows: the agent produces a number of random intonations and then it picks the one that is perceptually most similar to p_d to include in the repertoire.

Appendix III. The Evolution of Grammar: From Random Compositions to the Emergence of Structure

The following is a fragment of the grammar of one of the adult agents early in a simulation run. As with the example in table 2, a complete composition is described by context-free grammar rules with the category C. In this case, the agent's grammar is almost entirely a list of idiosyncratic compositions for each meaning that that agent has heard. There is no consistent structure, nor is the agent able to generalise to new meanings with recourse to random invention.

C/x(cursa,azha) → d6s A/x
C/exhilarated(acamar,joyous(zaurak,cursa)) → c6m a5l c6m
C/fanciful(achernar,keid) → fs6vs
C/fanciful(keid,cursa) → fs6m e6m
C/gay(achernar,zaurak) → d6m a6s
C/gay(zaurak,beid) → a6s
C/humorous(beid,leisurely(keid,acamar)) → d6vs b5vs
C/humorous(keid,achernar) → a6m a6m c6vl
C/humorous(zaurak,graceful(cursa,keid)) → a6l c6s b5l
C/joyous(acamar,quaint(cursa,zaurak)) → a6s b5s
C/joyous(zaurak,lyrical(cursa,azha)) → b5s d6vl b5s
C/leisurely(azha,achernar) → e6m b5m
C/light(rana,keid) → fs6vl d6vs

C/lyrical(achernar,rana) → d6l e6vl d6vl
C/lyrical(azha,cursa) → fs6s a6l
C/lyrical(beid,keid) → e6vs
C/plaintive(azha,achernar) → e6s
C/quaint(cursa,x) → d6vl B/x
C/quaint(keid,bright(rana,achernar)) → b5l fs6vs b5l
C/restless(rana,robust(zaurak,achernar)) → d6vl b5l
C/robust(keid,beid) → fs6vs c6s
C/satisfying(cursa,angetena) → a6m e6vs a5vl
C/sensational(achernar,cursa) → b5m e6vs c6l
C/sentimental(zaurak,light(angetena,zaurak)) –> d6vs c6vs
C/soaring(keid,beid) → fs6m a5vs
C/tender(cursa,achernar) → fs6l a5vs a6s
C/vigorous(angetena,keid) → a5vl a5vs
C/whimsical(acamar,cursa) → fs6m a5l
C/whimsical(keid,melancholy(acamar,azha)) → c6m c6m
A/joyous → a5s a5vs
A/longing → b5s fs6vl
B/achernar → fs6vs a5m
B/playful(beid,achernar) → e6vl a6l
. . .

The situation is very different many generations later in the same simulation. Here, there are only 4 top-level rules. Two of these are very general indeed in that they refer to the new categories A, B, and D which act like a dictionary of riffs, moods and emotions. Of particular interest here are the two A rules in bold. These are actually recursive, and serve to define a potentially infinite set of hierarchical emotion-structures. It is because of this, that the musical culture in this simulation has the potential for unlimited expressivity.

C/x(y,z) → c6vl A/z B/x A/y
C/x(y,z) → c6vl A/z a5vl b5l D/x A/y
. . .
C/sober(x,beid) → c6vl fs6l a6l fs6l c6vl A/x
C/exhilarated(x,passionate(achernar,angetena)) → c6vl fs6vl a5vs c6s a5l c6vl c6s b5vs c6vl A/x
. . .
A/x(y,z) → fs6vl B/x E/y A/z b5vs
A/x(y,z) → fs6vl a5vl b5l D/x E/y A/z b5vs
. . .
A/acamar → a5m
A/achernar → c6l fs6vl
A/angetena → c6s
. . .
B/agitated → a5vs fs6s b5l
B/awe_inspiring → d6m c6l e6vs
B/bright → a5vs d6vs b5m
. . .
D/dark → d6vl a5s

D/melancholy → b5l e6s
D/passionate → e6l a6l
. . .
E/acamar → e6vs a6s d6vl
E/achernar → a5l c6vl
E/angetena → b5vl