Proceedings of the 12th International Conference on Music Perception and Cognition and the 8th Triennial Conference of the European Society for the Cognitive Sciences of Music, July 23-28, 2012, Thessaloniki, Greece
Cambouropoulos E., Tsougras C., Mavromatis P., Pastiadis K. (Editors)

Musical Agreement via Social Dynamics Can Self-Organize a Closed Community of Music: A Computational Model

İsmet Adnan Öztürel,[*1] Cem Bozşahin[#2]

[*]*Cognitive Science Department, Middle East Technical University, Ankara, Turkey*
[#]*Cognitive Science Department, Middle East Technical University, Ankara, Turkey*
[1]adnan@ii.metu.edu.tr, [2]bozsahin@metu.edu.tr

ABSTRACT

This study aims to model social dynamics of an idealized closed musical society to investigate whether a musical agreement in terms of shared musical expectations can be attained without external intervention or centralized control. Our model implements a multi-agent simulation, where identical agents, which have their own private two dimensional transition matrix that defines their expectations on all possible bi-gram note transitions, are involved in round-based pairwise interactions. Throughout an interaction two agents are randomly chosen from the population, one as the performer and the other as the listener. Performers compose a fixed length melodic line by successively appending their most expected note sequences recursively by using sounds from a finite inventory. Listeners assess this melody to determine the success of the interaction by evaluating how familiar they are to the bi-gram transitions that they hear. According to success the interacting parties perform updates on their transition matrices. All agents start with a flat transition matrix, and the simulation ends when they converge on a state of agreement. We have found that 30 out of 144 possible bi-grams, 74 out of 1728 possible tri-grams, and 7 out of 20736 four-grams emerged as agreements, although only bi-grams are communicated. The findings signify that melodic building blocks for the modeled society are self-organizing, given the limited bi-gram expectations of individuals, and that convergence trends are dependent on simulation parameters.

I. INTRODUCTION

Evolutionary explanations are fruitful as they provide holistic views of societies and cultures instead of focusing on the behavioral regularities of individuals that arise upon the phenomena that is being investigated. They can be used to interpret emergence and evolution of well-structured symbolic systems which have substantial social functions, like music.

Particularly, within the domain of music cognition previous generative and combinatory theories (like Schenkerian Analysis, Generative Theory of Tonal Music and Combinatory Categorial Grammars) were interested in explaining only well structured musical pieces of tonal culture (Forte & Gilbert, 1982; Lerdahl & Jackendoff, 1996; Steedman 1984). Some other social aspects of music cognition need to be studied conveniently. Research questions like "*How do shared sound systems emerge?", "How do hierarchical systems like modality, tonality and their alikes evolve with a musical culture?"* or "*Does population dynamics play a crucial role in evolution and emergence of musical conventions?*" still remains unanswered. Considering these questions, it may be proposed that social dynamics of a musical culture may influence compositional routines of its own.

By any means, compositional grouping is for sure not random in any musical culture. Musical systems can be broadly formalized over the processes undertaken by the composer to generate a musical piece, in correlation with listeners' effort to resolve overall dependencies between the musical events within that piece to form a mental representation of what is heard. Accordingly, minimal agreement is required to bridge the compositional grammar adopted by the composer to generate and organize musical events and the listening grammar used by the listener to parse the composed piece (Lerdahl, 1988). From the listeners' perspective, compositional rules are not directly accessible, if not explicitly presented. However, listeners can reconstruct organizational rules between the musical events if they have a familiarity with the structural organization of the heard piece. Taking this into account, for a musical piece to be successfully parsed by auditors, composers must construct a structural organization within the composition based on a shared musical grammar which embraces both compositional and listening grammars.

Specifically, common and widely spread musical conventions among a culture form the natural grammar of music for that society. Natural grammars of music outline the boundaries for compositional and listening grammars that can be generated in a specific culture. Musical conventions can be exemplified with commonly used harmonic structures, melodic and rhythmic movements. These are not ossified, rather they are dynamically subject to change depending on time and culture in which they are in use. Besides, new musical styles emerge throughout time within a society and they impose an expansion in the set of musical conventions.

Keeping all these in mind, it can be inferred that musical systems cannot be grasped by only modeling cognitive abilities of individuals of a specific culture. Music is highly dependent on social interactions and cultural know-how. Therefore, a broader understanding on how musical conventions emerge and evolve can only be investigated in a model that can fulfill all these preliminary assumptions about musical systems.

In correlation, this paper will present a computational model of musical interactions and agents which are captured as a complex dynamical system. Computational CAS models are capable of capturing overall behavior of dynamical non-deterministic systems with respect to interactions of their individual constituent components, thus this lane of research is promising for studying music as a social tool and it can reveal intriguing facts about music cognition.

Briefly, the scope is of this paper is narrowed down only to study how a social consensus on musical expectations may be attained in a model of closed musical community. Correspondingly, agent's compositional preferences and their aesthetic assessments of the songs, which are exchanged among them, are only grounded to their musical expectations.

In accordance, it is aimed to explore how much of emergence of culturally dependent musical structures (such as commonly used melodic lines) can be explained with these minimal assumptions. In the following section, relevant models of the literature will be briefly overviewed in correlation with the essential methodology of CAS. Our model and results will be then be presented and discussed.

II. BACKGROUND

Dynamic multi-agent systems can be classified as CAS if large number of local interactions creates an adaptive behavior. Collective adaptive behavior gives cause for system complexity. Interactions between the micro-level constituents, which are generally called the agents, engender a structural reorganization on the system for it to reach a state that may promote a specific macro-level functional behavior. Self-organization in CAS is a never-ending process. Hence, widely spread self-adjusting interactions make the system behavior non-linear so that the system exposes a state far from optimality in a given time (Holland, 1992). Within the literature artificial distributed multi-agent simulations of CAS are being used for testing the plausibility of hypotheses related with emergence and evolution of linguistic and musical conventions, since they can capture social dynamics on macro-level.

Computational models of evolutionary linguistics try to model language as a social tool with adaptive complex dynamical systems. The field of research is often called semiotic dynamics as it investigates how a population of agents generate a structural organization on the way to create commonly shared social contexts or semiotic systems that involves social conventions which are essential for cooperative action (Steels & Kaplan, 1999). In correspondence, computational evolutionary musicology literature, which emanate from semiotic dynamics as the historical successor, focus on computational modeling of emergence and evolution of musical conventions (Miranda & Todd, 2007). Both lines of research investigate emergent behaviors of societies where complex local interactions between individuals effect global organization.

In this context, models of computational evolutionary musicology are non-situated, meaning that musical interactions between the agents do not relate them on an explicit representation of their environment. That is to say, musical structures do not indicate some state of affairs. Therefore, musical signal lacks indexicality. Besides, music can be represented as a sequential organization of musical events in time. To represent sequentiality, signals are composed as collections of successive musical events (in most cases notes or sounds).

In a simplistic way, models that are going to be presented in this section implement iterative rounds of pairwise (or groupwise) musical interactions. An instance of interaction encloses certain assumptions to regulate how agents engage in a mutual activity. They are designed to explore whether they would have specific bearings on the population behavior.

Particularly, agents are computational abstractions which are usually implemented as robots or computer programs. Agents communicate over signals. They are psychological competent, meaning that they are capable of producing signals to externally transmit to the peers so as to perform and they are capable to listen an external signal input. Therefore, they have sensori-motor apparatus which can be used as detectors and effectors. Moreover, they must be knowledgeable of what a signal designates and how to compose a signal from lower-level components. They should know how to interact with each other and how to assess a perceived signal that is externalized by a peer, so that they have a list of classifiers that specifies their possible actions in a given state. Usually, agents have a memory which represents their knowledge. Though they are not aware of their peers mental states or memory.

Respectively, an agent takes following actions while interacting with a peer:

1. Collect signals either from its environment or from another agent through the detectors/sensory apparatus.
2. Input signals are assessed with the classifiers.
3. Memory is updated depending on the rules of the interaction.
4. If it is entailed an action or an output signal is produced with the effectors/motor apparatus.

At this stage, it could be argued that foregoing formalization of musical interactions cannot capture some other substantial aspects of music cognition. For instance, non-disputably musical interactions culminate with an emotional response on the listener on individual basis. However, models of computational evolutionary musicology exclude such aspects with an abstraction. Taking all these into consideration, rest of this section will briefly review three of the models that were influential in designing ours.

A. Emergence of Shared Repertoire of Sounds

Composers and listeners must share some common knowledge on musical conventions to complete a successful musical interaction. Miranda (2002) proposes that the primary requirement for a society to bootstrap a shared musical lexicon is to attain a state where its individuals' knowledge on musical conventions must be sufficiently similar. It has been argued that the primary aim of a musical agent must be to reorganize its musical knowledge by interacting with other members of the community to have a common background. Accordingly, this effort can be named as sociability or social bonding. In other words, an agent becomes accepted in a society if it can produce pieces that can be parsed by others and if that agent can parse the pieces composed in accordance with the conventions of that society (Miranda, 2002; Miranda & Drouet, 2005; Miranda, 2008).

A simulation is designed by Miranda (2002) to capture the effects of above mentioned sociability hypothesis on organization of the social structure of a musical society. Basically, the aim of each agent is to successfully mimic the heard signal which is created by a composer counterpart. Simulation models a population of agents that are capable of playing the role of both performer and imitator. In each round a musical interaction occurs between two randomly chosen agents where one is the performer and the other is the imitator.

Agents of this simulation are robot implementations and the musical signals (sounds) shared between them are real world acoustic signals. To process a sound, agents use a two-fold representation scheme. Each one of them is equipped

with two separate lexicons to store motor and perceptual representations of a sound. Moreover, they are capable of remembering how many times they were successful in imitating a specific sound.

To hear a sound agents use a hearing apparatus which converts the acoustic signal to its perceptual representation. Perceptual representation is the rough estimations of the pitch, loudness and duration of a sound which can be calculated by the hearing apparatus from the heard acoustic signal. If an agent wants to play a specific sound it first chooses the perceptual representation from its own lexicon and then uses the corresponding motor representation to articulate. Motor representation consists of the parameters fundamental frequency, amplitude and duration. Vocal synthesizer of an agent uses these values to synthesize the sound that is intended to be generated.

Overall, as it is sketched out by Miranda (2002) in an interaction a random performer and imitator is chosen among the population. Performer plays a sound from its lexicon by choosing a perceptual representation from the lexicon and articulating its corresponding motor representation. If its lexicon is empty, then a random sound is generated. Imitator extracts a perceptual representation from the heard sound. After the extraction, imitator searches for a similar perceptual representation in its lexicon. When a match is found corresponding motor representation for the most similar sound is articulated. A randomly generated sound is played back as the imitation when its lexicon is empty. Performer listens and assesses whether the imitation is sufficiently similar to the sound it played. If it is similar enough imitator is informed that its imitation was successful. Otherwise, imitation is ranked as unsuccessful and the imitator is informed accordingly. Both performer and imitator update their lexicon regarding the success of the imitation. If the original performance and its imitation are similar, both agents reinforce and increase the amount of successes gained by using that sound. Sounds, which are not used in a successful imitation for a specific amount of time, are forgotten.

To refine the assessment procedure, it can be stated that different agents can have different motor representations for a particular sound. However, two sounds are classified as the same if their perceptual representations overlap. Therefore, an imitation is successful if interacting agents can come up with a perceptual match rather than agreement of the sensory representations.

Whenever an agent imitates a sound from its lexicon it slightly alters its motor representation on performance. To this end, alteration of the motor representation on real-time articulation confronts the spread of new intonations to the society. With the help of reformative updates on the lexicon after successful imitations population dynamically reaches to a state of agreement of lexicons. Eventually within this agreement state imitations are observed to be successful which assures social bonding between the individuals.

B. Evolution of Musical Behavior

Werner and Todd (1997) argue that evolution of music is a consequence of selective pressure in that community. They propose that effects of different compositional routines on a society can be reformulated and analyzed as a co-evolutionary mating problem through musical interactions. That is to say, a population of agents from two opposing sex, namely males and females, are modeled. Males undertook the role of being musical performers, whereas females were listeners. Agents are knowledgeable about a closed lexicon of notes. Males have their own songs as a genotype which is a sequential arrangement notes (an abstraction of a melodic line). Each female have a transition table that encodes musical expectations of their own. To clarify, transition tables contain information about how probable an antecedent-consequent note sequence is for the listener. Females use it to evaluate the coherency between the heard musical signal and their expectations.

Each round of interaction models a mating process. After individual interactions breeding occurs between a female and a male counterpart, which is chosen by the female, to create a child with merged compositional preferences of its parents. The aim of each female is to choose the most eligible candidate for them to mate. Briefly, within a round each female interacts with a predefined number of performers that are randomly selected among males in order to choose a mate. Each selected male plays its song to the female and the female listener evaluates these songs and chooses the highest scored performer as the mate.

With this model Werner and Todd (1997) examines whether the songs of this society evolves in subsequent generations contrastingly depending on the evaluation methodology of females. To put it in another way, females' preference scheme can be to look for the most familiar song or to the one that has the most surprising elements. Accordingly, Werner and Todd (1997) claim that a surprise seeking assessment methodology can produce plausible musical diversity while attaining a co-evolutionary trend in male songs.

As a follow up study, Bown and Wiggins (2005) altered above-mentioned model with two assumptions. In the first place, Bown and Wiggins (2005) was interested in topological distribution of the interacting agents. Therefore, interactions are constrained by allowing listeners to listen performers which are credibly close to them. Such a limited listening space assumption on performer selection contrasts with Werner and Todd (1997)'s free and random listener selection. Secondly, Bown and Wiggins (2005) does not try to capture cultural genetic evolution of musical signals. Rather, their main aim is to explore dynamic spatial organization of the agents in a limited social space. Therefore, agents are considered to be musically skilled to both perform and listen, though they do not have a sex that dictates a role on them.

Respectively, in Bown and Wiggins (2005) every agent in the population, even the performers, has a transition matrix like listeners of Werner and Todd (1997). This transition table defines an agent's compositional preference if it is performing, whereas it is a tool to assess the heard signal that encodes agent's musical expectations if it is listening. It is still presumed that all agents are aware of a global closed set of lexicon consisting of notes. Each agent has a random starting position in space. Listeners score performers' songs by summing up the individual expectation values for each transition. According to the result of the evaluation listener moves closer to or farther away from the performer. Hence, if the song has a high score listener moves close to that performer so that their chance to interact in upcoming rounds

increases. The adverse scenario applies if the interaction is not successful.

The most intriguing outcome of this model is the formation of stable musical subcultures via spatial clustering. Close and relatively smaller clusters seem to affect each other so that they can merge or change their position in space within time. However, relatively large clusters have their own isolated mainstream expectation trends. They are robust when compared to the smaller ones. Eventually, Bown and Wiggins (2005) shows how diversification of spatial musical expectations emerges and evolve within a musical culture where distinct subcultures have an influence on each other.

III. MODEL

The model that is going to be presented in this section primarily aims to capture a society of musical agents with random musical expectations at the outset which can attain a state where musical expectations of individuals will allow them to compose melodic lines that will be pleasing when listened by peers. It is assumed that for performers to be appreciated by listeners their compositions must satisfy expectations of the audience. That is to say, assumptions about agents and interactions are chosen in a way that an agreement on a commonly shared musical preference scheme can become observable with freely interacting agents.

In this scope, musical expectations will pretty much be the same with what Meyer (1957) has proposed. Individual expectation schemes profile agents' familiarity and foresight for certain patterns of successive musical events. It is presumed that agents anticipate for a specific precursor after each musical event they hear on the go. They get surprised if the precursor is not the one that they were expecting.

In what follows, we will first present a formal description of our model in order to overlay how agents are implemented and how they interact with each other. Successively, we will list our predictions and model based preliminary assumptions. Finally, we will conclude this section as we elaborate on the experiments that we have designed to test our predictions.

A. Agents and Interactions

Our model is a computational abstraction of a population of musical agents $A = \{a_1, ..., a_i\}$, where population size $N_A = i$. Agents are capable of playing the roles of both performer and listener. Each and every agent is equipped with a transition table T which defines their musical expectations. Transition table is a two-dimensional matrix where each dimension has the size of the lexicon N_L. To demonstrate, assume that the lexicon only consists an octave of pitches, so that $N_L = 12$. In this case, the lexicon would be $L = \{C, C\#, D, D\#, E, F, F\#, G, G\#, A, A\#, B\}$. Then T will be a 12 x 12 matrix where rows will represent all possible antecedent notes and columns will represent all possible consequent notes, such as:

$$T = \begin{bmatrix} \alpha_{C,C} & \alpha_{C,C\#} & \cdots & \alpha_{C,A\#} & \alpha_{C,B} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \alpha_{B,C} & \alpha_{B,C\#} & \cdots & \alpha_{B,A\#} & \alpha_{B,B} \end{bmatrix}$$

$$\text{where } 0 \leq \alpha_{n_i,n_j} \leq 1 \text{ for } \forall n_i, n_j \in L$$

Within the transition table a cell will define how much an agent expects a specific antecedent-consequent note pair to occur successively. For instance, the value of $\alpha_{D,F\#}$ will give us the amount of expectation of an agent to hear an F# after it hears an instance of D within a signal. All agents start the game with a flat transition table, that is all α in T has the value 0 at $t = 0$. Every agent's musical expectations are dynamically shaped with readjusting modifications on their transition tables with respect to the success of the interactions that they get involved throughout the simulation.

Agents of our model can be characterized with their table of musical expectations T, as their aim in each interaction is to evaluate whether a musical piece is pleasant enough in terms of satisfying their expectations on bi-gram level. It should be kept in mind that our transition table implementation only allows agents to devise bi-gram expectations (i.e. agents can have a specific expectation for the note pair C-G to occur successively, but not for any longer n-gram sequences like C-G-F-C). Therefore, evaluation of the heard melodic line will be carried out over individual successive bi-grams.

The rules of each round of interaction is as follows:

1. Performer a_x and listener a_y is selected randomly from the population to interact, where $x \neq y$ and $a_x, a_y \in A$.
2. Performer a_x composes a song S with predefined length N_S by using its transition table T_x and plays it to the listener a_y.
3. Listener a_y evaluates S by using its transition table T_y and conveys the success of the interaction to its counterpart.
4. Participants a_x and a_y modify their transition tables T_x and T_y.

To compose a melodic line agents complete the following steps:

i. An empty song template $S' = [\,]$ is created.

ii. A random note $n_k \in L$ is selected and placed in the template as the first note. At the end of this stage template with $N_{S'} = 1$ looks like $S' = [n_k]$, where $S'[1] = n_k$.

iii. Rest of the song is recursively built in $N_S - 1$ iterations. In each recursion performer takes the last note $S'[N_{S'}]$ from the template and checks its transition table T_x for the most expected successor. This search is carried out with λ function which is defined as follows:

$$\lambda(S'[N_{S'}]) = max(\alpha_{S'[N_{S'}],n_i}) \quad for \quad \forall n_i \in L$$

λ retrieves the most expected consequent for a given antecedent. If there are more than one successor notes with the same expectation value then one of them is randomly chosen. In each iteration $N_{S'}$ increases by one and composition is completed when $N_{S'} = N_S$.

To evaluate a song success score is calculated in an additive fashion. To be clear, a musical signal represents one complete melodic line, therefore its pleasantness for the listener could be fixed on summation of listeners expectation values for each transition that they have encountered sequentially. Notably, agents use the following local scoring policy for evaluation:

$$score = \sum_{k=1}^{N_S-1} \alpha_{S[i],S[i+1]}$$

For the interaction to be successful agents must be familiar with antecedent-consequent note pairs that they hear in the song to some extend. This measure of familiarity is fixed on a predefined threshold θ. Agents use the evaluation function ϵ to assess a song, as it is presented in below:

$$\epsilon(score) = \begin{cases} successful & if \quad \theta * (N_S - 1) \leq score \\ unsuccessful & if \quad score < \theta * (N_S - 1) \end{cases}$$

Overall score for the song is calculated by adding the listener's expectation values for all transitions. However, evaluation is completed over the whole song score. Therefore, each transition will have an impact on the success of the song, but they will not be decisive for the success individually.

Finally, listeners always increase expectation values for all transitions that they hear in the song without taking success into consideration in order to modify their transition tables. This is because, it is assumed that listeners familiarity for antecedent-consequent pairs increase as they encounter them in a song. However, performers modify their transition table with regard to the listeners' evaluation. Performers increment expectation values for all transitions if the interaction was successful; otherwise they are decreased.

It is crucial to mention that transition tables are also used for composition other than evaluation. Thus, if an agent is performing its table defines its compositional preferences. In accordance, performers update their tables to make further use of the antecedent-consequent pairs that they gained success in previous interactions in their upcoming compositions. On contrary, decrease in expectation values after unsuccessful interactions help them to avoid using those note pairs in future.

The amount of this modification on expectation values is predefined by learning rate τ for both increment and decrement. Agents use $\mu: T \rightarrow T'$ function for table updates and it is formally defined as follows:

$$\mu(T,S) = T' = \begin{cases} \alpha_{S[i],S[i+1]} = \alpha_{S[i],S[i+1]} + \tau \; for \; \forall i, \; 0 < i < N_S & if \; successful \\ & or \; listener \\ \alpha_{S[i],S[i+1]} = \alpha_{S[i],S[i+1]} - \tau \; for \; \forall i, \; 0 < i < N_S & otherwise \end{cases}$$

Notably, it can be observed from the above definition that the agents do not employ inhibition while performing table updates. For instance, if there are two C-G pairs in a song that ends up as a successful interaction, performer and listener increases $\alpha_{C,G}$ in their transition tables by 2τ. However, values for $\alpha_{C,ni}$ where $\forall n_i \in L, \; n_i \neq G$ will not be inhibited. Therefore, T is not a pure probability table, that is to say expectation values in a column will not always add up to 1.0. In this respect, an $\alpha_{ni,nj}$ will give us an expectation value, but not a probability, for defining how often an agent expects an n_i,n_j pair. This expectation is merely dependent on how many times it heard that specific pair as a listener or how many times it used it as a performer in a successful interaction.

B. Model Predictions and Assumptions

In a state of agreement a population can only have specific expectations on bi-gram level (as they only predict for successor notes), but throughout this agreement signals which are pleasant may have significant ossified n-gram melodic lines in them. Hence, it is predicted that if the population can agree on shared expectations their composition can employ significant usage of melodic progressions with length more than two. This prediction stems from the sequentiality of the musical signal. In order to clarify, a musical piece is presumed to be the sequential arrangement of atomic units that are notes from the lexicon set. Within this context, in a possible agreement state population will compromise on particular bi-grams as socially shared musical building blocks. Accordingly, when these come together sequentially they may form lengthier significant melodic lines. However, our model will not entail them to be hierarchically or categorically ordered as our agents will not be capable of representing the relationships between these building blocks.

In accordance, specialized model specific assumptions for the problem that we are interested in can be listed as:

• **Identical Agents:** Every agent is cognitively and psychologically capable of composing musical pieces and perceiving them. The channels that are used for performance and audition are separable. Moreover, each agent's musical expectations are unique. Agents are not able to directly access to others expectations. They can make limited estimations about a peer's musical expectations over its composition when they interact. Agents are not bestowed with a representation of the global system behavior and none of them have a direct impact on it. Therefore, system control is decentralized so that only local interactions can superimpose an aggregate system-wide behavior.

• **Closed Lexicon:** Agents are aware of a stable and closed lexicon of sounds. Songs can only be composed with this set of sounds which is an abstraction of musical notes generally ranging over several octaves. For instance, if lexicon is set to be two octaves, then it will consist all pitches in between C2-B3. This assumption is required to ensure that all agents are aware of all possible musical pitches which can be included in a composition. Computationally it is required to presume that such a global closed lexicon exists, since agents enhance their compositional preferences in accordance with their expectations over all possible bi-grams that can be built upon this set while they are interacting.

• **Closed Community:** Throughout the simulation population size will be kept constant. Interacting agents will not be replaced with new ones at any stage; hence there will not be any disturbance on overall population behavior. Briefly, it is aimed to investigate significant consequences of agent interactions (such as emergence of significant compositional preferences or reusable musical patterns) while population is reaching to an agreement.

• **Free Interactions:** Every agent is equally probable to interact with each other. Any kind of topological distribution is omitted. Participants of an interaction are randomly chosen among the population and any agent can be chosen to participate in an interaction. Such a free interaction assumption assures every agent to equally effect others musical expectations.

• **Consecutive Interactions:** Agents cannot concurrently interact with each other. In fact, parallel interactions are prevented as a consequence of assumption of free interactions.

Concurrent interactions can cause conflicting memory updates as interacting parties are randomly chosen. Contradictions on how to perform knowledge updates will be computationally restrictive and it is avoided as it can result in abnormalities on population dynamics.

• **No Boredom:** Agents cannot get bored with extremely familiar signals. In other words, interactions are successful even if entire antecedent-consequent note pairs of a song are the ones which are expected by the listener. This assumption is required if an absolute convergence of a shared musical expectations is awaited. In our model dynamic restructuring of individual preferences is expected to be manifested throughout the learning phase before convergence.

C. Experiments

With the model that is described above two interrelated experiments are conducted to investigate the overall behavior of the population with respect to various simulation parameters. These experiments are:

1. Test for agreement of expectation tables.

The simulation is run for a base case with 50 agents (N_A), where the lexicon size was two octaves (N_L) to examine whether agents can converge on a shared musical preference. In this specific case agents composed musical signals of size 32 (N_S), and their learning rate (τ) was 0.05. Successively, population dynamics is investigated for varying number of agents (N_A = 25, 50, 75), lexicon size (N_L = 12, 60, 120), signal length (N_S = 16, 24, 32, 64), learning rate (τ = 0.025, 0.05, 0.075, 0.1) and success threshold (θ = 0.3, 0.4, 0.5, 0.6) independently.

2. Test for emergence of reusable units.

Once more the simulation is run for N_A = 25, N_L = 12, N_S = 12, τ = 0.05, θ = 0.5, but throughout this run a global transition table (*GT*) is calculated for each round of interaction. This global table represents overall average of the bi-gram expectations for the whole population. *GT* can be defined as follows:

$$T[i][j] = \alpha_{n_i,n_j} \quad for \quad \forall n_i, n_j \in L$$

$$GT[i][j] = \frac{\sum_{k=1}^{k=N_A} T_k[i][j]}{N_A}$$

In this case, *GT* is examined particularly to investigate whether all agents can form a consensus on expectations for specific note pairs. In this experiment a chi-square ($\chi 2$) significance test is applied to a corpus of musical signals to find significant collocations. Corpus (C_x) is the collection of signals that are exchanged between interacting agents after overall success rate *S(t)* exceeds *x*. In addition, *S(t)* is the number of average successful interactions throughout the simulation

In this test three different corpora ($C_{0.0}$, $C_{0.5}$ and $C_{0.9}$) are generated for *S(t)* = 0.0, 0.5, 0.9. $C_{0.0}$ consists all signals that are composed from the beginning of the game. $C_{0.5}$ and $C_{0.9}$ includes signals after *S(t)* = 0.5 and 0.9 respectively. It could be presumed that an agreement is formed after *S(t)* = 0.9, because principally it is guaranteed for this setup that agents were interacting successfully for at least 40,000 rounds to reach this success rate. So, $C_{0.9}$ supposedly only includes signals of the agreement state. $\chi 2$ is applied to all these three corpora to observe the dynamic nature of consensus formation.

$\chi 2$ values for each possible bi-gram that can be formed by using the notes of the lexicon makes it possible to test whether they significantly co-occur successively throughout the agreement stage. In fact $\chi 2$ significance test can be extended to n-grams of any length, though it is crucial to note that the degrees of freedom changes when applying $\chi 2$ to find out significant collocations for varying lengths. In particular, for $\chi 2$ test where n is the length of the significant collocation that is being tested there is $(n-1)^2$ degrees of freedom (i.e. for tri-grams there is 4 degrees of freedom and 4-grams there is 9 degrees of freedom). In our experiment we apply $\chi 2$ test on the corpus for all possible collocations of bi-grams, tri-grams and 4-grams. That is to say, it is examined whether any significant musical pattern of length three or four can emerge from agents' limited bi-gram musical expectations. This test is carried out in correlation with dynamic evolution of global musical expectations that is defined by *GT*.

IV. RESULTS

A. Population Dynamics

In the beginning of the game agents do not have specialized preferences for composition and assessment. After successful interactions they perform memory updates in order to build their own private tables of musical expectations. Accordingly, growth of the success rate will signify convergence on a state of agreement on global musical expectations as agents start to interact successfully more often. In fact, an investigation of individual interactions will be barely useful to study population dynamics since large number of random and free interactions result in an unmanageable stochasticity. In this respect, throughout this section success rate *S(t)* and time of convergence will be our primary measures to interpret model performance. Notably, all results that are going to be presented in this section are averaged over 10 runs for each test case.

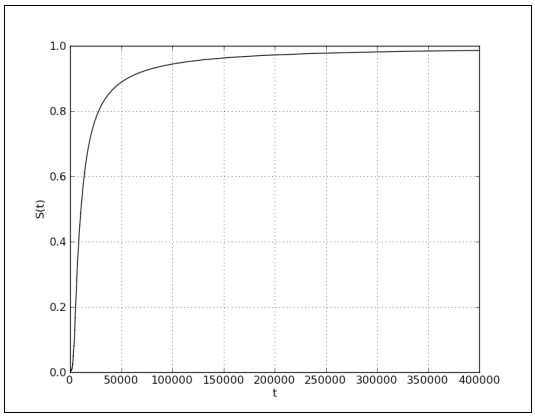

Figure 1. Success rate of interactions and convergence on a shared table of expectations. Simulation is run for N_A = 50, N_L = 24, N_S = 32, τ = 0.05, Θ = 0.5.

Figure 1 presents change of *S(t)* within time for a baseline simulation that tests whether agents can ever come to a state of agreement when convenient conditions are provided. It can be observed that success rate increases rapidly at the outset and it grows steadily till *S(t)* = 0.9. Within this phase, most of the learning takes place. After *S(t)* = 0.9 rate of increase in

success rate decreases and the curve flattens since learning is brought to a completion.

This distinctive success rate curve denotes a decisive minimal agreement as shared musical expectation scheme becomes spread among the population. However, time of convergence is heavily dependent on population and agent characteristics, such as population size, learning rate of the agents, success threshold, number of notes used for composing and length of the signal.

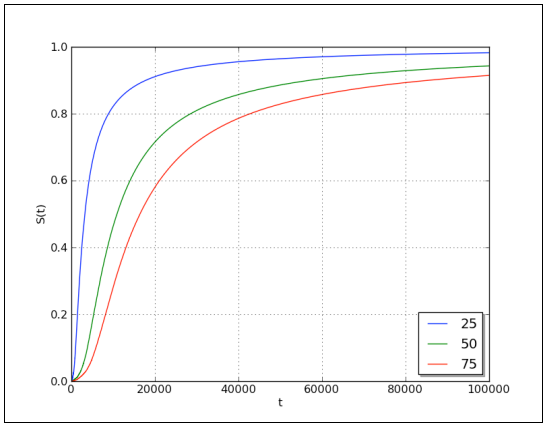

Figure 2. Effect of population size N_A on convergence. Simulation is run for N_L = 24, N_S = 32, τ = 0.05, Θ = 0.5. Convergence trends for N_A = 25, 50 and 75 and presented.

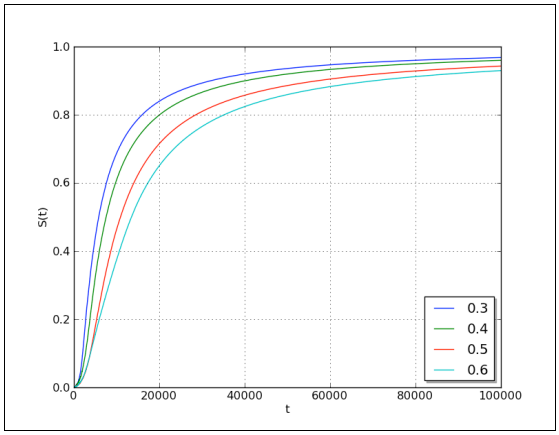

Figure 3. Effect of success threshold Θ on convergence. Simulation is run for N_A = 50, N_L = 24, N_S = 32, τ = 0.05. Convergence trends for Θ = 0.3, 0,4, 0.5 and 0,6 are presented.

Effects of population size on time of convergence can be observed in Figure 2, particularly for N_A = 25, 50 and 75. For larger populations agreement comes late. Therefore, increasing the population size will also delay convergence since dominating bi-gram expectations has to be spread to the individuals of the population to attain an agreement. It could be deduced that all agents must participate in a considerable amount of interaction to learn what others expectations are for a global transition table to become observable. As a consequence, greater number of interactions is required for the members of larger communities to interchange their preferences on winning note pairs.

Agents' learning is dependent on two independent factors that are success threshold θ and learning rate τ. To evaluate a song, raw sum of expectations for each bi-gram of that song must exceed a specific success threshold. In consequence, θ defines a lower boundary for a song to be pleasant. In other words, θ determines how much performers and listeners transition tables should overlap for the interaction to be successful. In Figure 3 varying $S(t)$ curves are presented for increasing success thresholds 0.3, 0.4, 0.5, 0.6, and it could be observed that time of convergence increases with increasing θ. This is because listeners look for higher number of expected transitions in the song to classify the song acceptably familiar when θ is large.

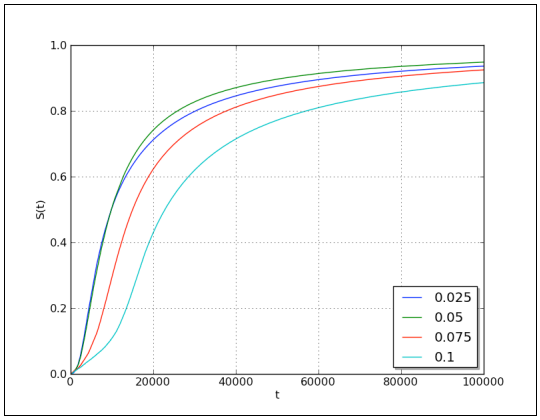

Figure 4. Effect of learning rate τ on convergence. Simulation is run for N_A = 50, N_L = 24, N_S = 32, Θ = 0.5. Convergence trends for τ = 0.025, 0.05, 0.075 and 0.1 are presented.

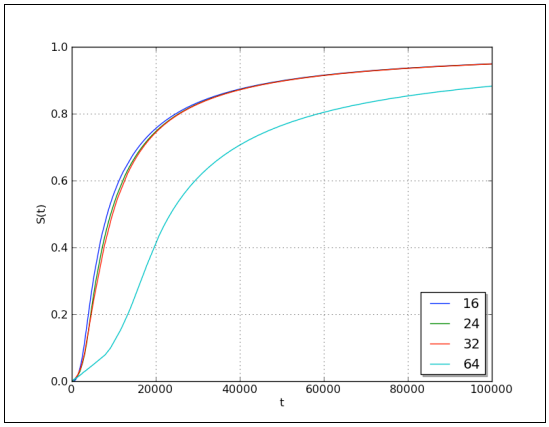

Figure 5. Effect of signal length N_S on convergence. Simulation is run for N_A = 50, N_L = 24, τ = 0.05, Θ = 0.5. Convergence trends for N_S = 16, 24, 32 and 64 are presented.

From Figure 4 it can be observed that time of convergence significantly depends on τ. System's learning is optimal at around τ = 0.5. Within the range $0.2 < \tau < 0.7$ system tends to converge rapidly. However, for greater or smaller values of τ time of convergence increases. In particular, τ determines how fine agents search the state space. So, for both considerably small and large learning rates this search is not optimal, thus performance is affected negatively. When τ is fairly small increase in expectation values for antecedent-consequent pairs of successful interactions are negligibly small so that the population cannot bring out winning bi-grams promptly. In a similar fashion, if τ is larger than the aforementioned boundary agreement comes late since expectation values for bi-grams that brings success drastically alters after memory

updates throughout the learning phase.

In Figure 5 alteration of *S(t)* is presented with respect to signal lengths N_S = 16, 24, 32 and 64. Notably, change in signal length does not affect convergence up to a hard boundary. From the figure it can be observed that *S(t)* curves overlap for N_S = 12, 24 and 32. However, when signal length grows significantly larger (such as N_S = 64), *S(t)* drastically drops. Besides, *S(t)* for N_S = 64 can not even catch up success rates of N_S = 12, 24 and 32. Lengthier signals consist relatively larger amounts of transitions to be evaluated. Consequently, it is more likely for an antecedent-consequent pair to be involved in a composition more than once for larger N_S. Hence, when signal size increases expectation values for winning bi-grams that are involved in performers compositions are intensely modified. Therefore, population could not easily settle on a dominating set of bi-grams. Indeed, an adverse effect should be expected for shorter signals. However, from Figure 5 it could be deduced that performance does not always improve for short signals. Arguably this is because, impact of other independent parameters such as τ and θ supervenes the impact of N_S on learning rate.

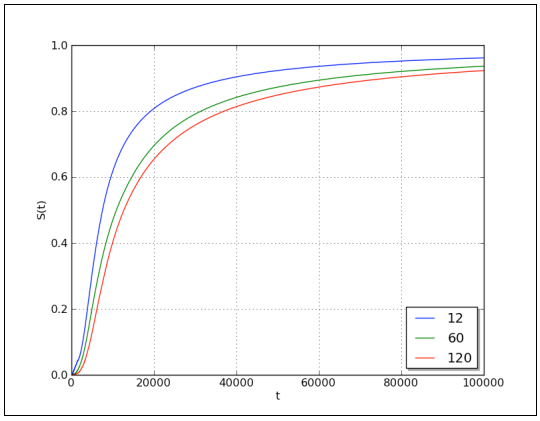

Figure 6. Effect of lexicon size N_L on convergence. Simulation is run for N_A = 50, N_S = 32, τ = 0.05, Θ = 0.5. Convergence trends for N_L = 12, 60 and 120 presented.

Finally, Figure 6 presents how lexicon size N_L effects convergence. Lexicon size determines the size of the state space. If agents are allowed to use greater number of notes in their compositions, the amount of all possible bi-grams that can be produced from the lexicon grows exponentially. As long as the state space grows the time required for the population to form an agreement in one of the attractor states increases. Consequently, it can be observed from Figure 6 that an increase in lexicon size lags convergence.

B. Self-Organization and Emergence of Reusable Units

In this section, a representative run of the simulation will be examined to present dynamic self-organization of the population. In Figure 7 it can be observed that population convergences on a global table of expectations roughly at 50,000. At the end of 200,000 rounds *S(t)* converges to 1.0. The global transition table at this point is presented in Table 1.

In the global transition table (*GT*) there are thirty antecedent-consequent pairs, which have significantly high expectation values (i.e. C-C, D#-E, etc.). These note pairs are the ones which are commonly agreed on by the population at exactly t = 200,000. However, when we perform a $\chi 2$ to a corpus of signals for $S(t) \geq 0.9$ test yields thirty-two bi-grams that significantly appear successively throughout the agreement state. For instance, with a quick comparison between Tables 1 and 2-(a) it could be seen that B-G and B-B bi-grams do not have high expectation values in *GT*, whereas they appear to be significant according to the $\chi 2$ test. This dissimilarity arises from self-organizing nature of the system.

To be clear, dominating bi-grams are not deterministically predefined. Interactions between the agents would result in alterations in the set of winning bi-grams throughout the game. In other words, winning bi-grams can lose their significance or adversely a non-significant bi-gram can become a winning pair dynamically. This spontaneous restructuring is continuously carried on. For instance, Tables 1 and 2-(a) show that B-G and B-B were the winning pairs at early stages of agreement, however they lost their significance later on.

Self-organization becomes prominent when we perform $\chi 2$ test to $C_{0.0}$ and $C_{0.5}$. For $C_{0.0}$ (notably all signals of the game included in this corpus), $\chi 2$ test shows that 140 of the all possible 144 bi-gram collocations were significant. This means that the population nearly searched for all states throughout the game. Successively, for $C_{0.5}$ there are only 44 bi-grams, so it can be deduced that individuals of the population converge on an attractor state by fine-tuning their expectation tables to narrow down the set of significant bi-grams. Set of winning pairs can be different for each distinct run, however convergence on a specific set of bi-grams is being attained outright.

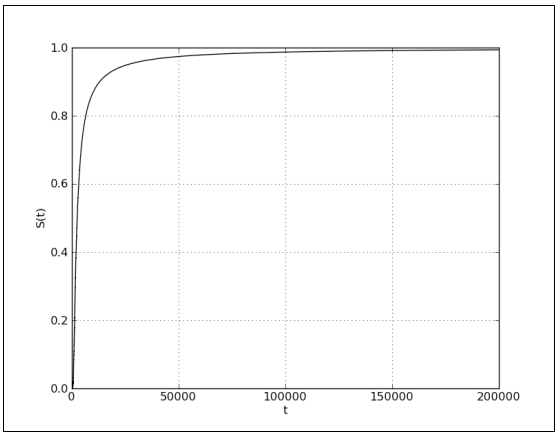

Figure 7. S(t) curve for N_A = 25, N_L = 12, N_S = 12, τ = 0.05, Θ = 0.5.

As a consequence, all winning antecedent-consequent note pairs become bi-gram building blocks of a musical signal. In Tables 2-(b) and 2-(c) it can be concluded that some fragments of signal of length three and four can be significantly observed by applying $\chi 2$ test on $C_{0.9}$. There are 74 observable tri-grams and 7 4-grams that are significantly used in melodic lines through the agreement. Notably, there is a vast difference between the number of significant tri-grams and 4-grams. As the length increases the number of significant n-gram melodic lines decrease. This trend could be grounded on the learning trend of the population. All these n-gram note sequences are composed of sequential arrangements of some of the winning bi-grams, thus it could be stated that these

n-gram sequences are the commonly shared pseudo melodic lines of the population. Hence, winning bi-grams become the reusable musical units for the population to compose lengthier structures.

Table 1. Global Transition Table after 200,000 interactions for N_A = 25, N_L = 12, N_S = 12, τ = 0.05, Θ = 0.5. Bold values indicate the bi-grams, which are currently agreed on by the population.

	c	c#	d	d#	e	f	f#	g	g#	a	a#	b
c	**1.000**	0.012	0.020	0.026	0.038	0.012	0.022	0.012	0.018	0.032	0.054	0.024
c#	0.032	**1.000**	0.048	0.130	0.024	0.012	0.048	0.054	0.088	0.116	0.006	0.012
d	0.012	0.022	**1.000**	0.074	0.044	0.060	0.020	0.014	0.012	0.150	0.022	0.024
d#	**1.000**	0.094	0.046	0.088	**1.000**	0.012	0.008	0.016	0.018	**1.000**	0.008	0.098
e	**1.000**	0.008	0.016	**1.000**	**1.000**	**1.000**	0.048	0.056	0.018	0.028	**1.000**	0.010
f	**1.000**	0.026	**0.996**	0.006	**1.000**	0.050	**1.000**	**1.000**	0.016	0.018	0.022	**1.000**
f#	0.022	0.008	0.040	0.018	0.010	0.100	**1.000**	0.044	0.150	0.030	0.008	0.014
g	**1.000**	0.012	0.012	0.004	0.012	**1.000**	0.014	0.072	0.076	0.016	0.074	0.028
g#	**1.000**	0.032	0.024	0.040	0.012	0.014	**1.000**	0.026	0.078	0.018	**1.000**	0.022
a	**1.000**	0.014	0.022	0.020	**1.000**	0.030	0.046	0.010	**1.000**	0.054	**1.000**	0.016
a#	**1.000**	0.056	0.070	0.018	0.040	0.014	0.012	0.010	0.054	0.012	**1.000**	0.044
b	**1.000**	0.056	0.008	0.072	0.484	0.012	0.006	0.504	0.032	0.012	0.018	0.096

Table 2. (a) Significant bi-gram collocations found with χ^2 on $C_{0.9}$. (b) Significant tri-gram collocations found with χ^2 on $C_{0.9}$. (c) Significant 4-gram collocations found with χ^2 on $C_{0.9}$. Simulation is run for N_A = 25, N_L = 12, N_S = 12, τ = 0.05, Θ = 0.5.

(a) Bi-grams	(b) Tri-grams		(c) 4-grams
C-C	C-C-C	F-F#-F#	D#-E-F-B
C#-C#	C#-C#-C#	F-G-C	E-E-F-B
D-D	D-D-D	F-G-F	F-E-F-B
D#-C	D#-C-C	F-B-C	F-G-F-B
D#-E	D#-E-C	F-B-E	A-E-F-B
E-A	D#-E-D#	F-B-G	B-G-F-B
E-C	D#-E-E	F#-F#-F#	B-B-B-B
E-D#	D#-E-F	G-C-C	
E-E	D#-E-A#	G-F-C	
E-F	D#-A-C	G-F-D	
E-A#	D#-A-E	G-F-E	
F-C	D#-A-G#	G-F-F#	
F-D	D#-A-A#	G-F-G	
F-E	E-C-C	G-F-B	
F-F#	E-D#-C	G#-C-C	
F-G	E-D#-E	G#-F#-F#	
F-B	E-D#-A	G#-A#-C	
F#-F#	E-E-C	G#-A#-A#	
G-C	E-E-D#	A-C-C	
G-F	E-E-E	A-E-C	
G#-A#	E-E-F	A-E-D#	
G#-C	E-E-A#	A-E-E	
G#-F#	E-F-C	A-E-A#	
A-C	E-F-D	A-G#-C	
A-E	E-F-E	A-E-F	
A-G#	E-F-F#	A-G#-F#	
A-A#	E-F-G	A-G#-A#	
A#-C	E-F-B	A-A#-C	
A#-A#	E-A#-C	A-A#-A#	
B-C	E-A#-A#	A#-C-C	
B-G	F-C-C	A#-A#-C	
B-B	F-D-D	A#-A#-A#	
	F-E-C	B-C-C	
	F-E-D#	B-E-A#	
	F-E-E	B-G-C	
	F-E-F	B-G-F	
	F-E-A#	B-B-B	

To put it differently, all possible signals with length N_S that can be generated from the lexicon L create the state space for the agents. As the population dynamically self-organize to reach a consensus on bi-gram musical expectations they agree on a set of winning antecedent-consequent pairs. This self-organizing behavior can also be described as a search problem where interactions and consecutive memory updates help the population to settle on an attractor state. Notably, within this state performing agents have a favor to use winning bi-grams in their compositions. Since $N_S > 2$ compositions within this attractor state involves melodic patterns that are greater than length two which can be classified as pleasant by the listeners. This is because, throughout the agreement phase performing agents append the winning bi-grams sequentially to compose a song.

V. CONCLUSION

In this paper, we have presented a computational model of a multi-agent musical society which can capture social dynamics of musical agreement in terms of shared musical expectations. We have found that a closed community of agents can converge on a global musical expectations scheme without any external intervention and centralized control when specific baseline conditions are provided. These conditions can be characterized with simulation parameters such as population size, learning rate of the agents, success threshold, lexicon size and signal length.

Our method of modeling has proven to be successful to investigate how musical structures change in time within a culture with pairwise interactions of the involved agents. Overall, it is presented that a closed community can attain a state where it has its own specialized musical expectations. The change in cultural know-how of compositional preferences and aesthetic evaluation of a song can be modeled in a self-organizing system as a continuously evolving dynamic phenomenon. Moreover, it is concluded that building blocks of a musical piece can emerge as a result of the sequential organization while agents converge on the shared expectation scheme.

The model and the findings are novel with respect to previous research of cognitive musicology. However, it has been presented that emergence of musical conventions could be studied in a model in which musical agents are only acting in accordance with their musical expectations. Within this context, emergence of socially shared musical conventions such as harmonic and melodic progressions and rhythmic movements might be worked out over the structural characteristics of a musical piece like we did.

Particularly, it should be kept in mind that dynamics of real world musical interactions are most likely different from this computational model. We are abstracting the musical signal in a way that we are only representing its constituents while leaving out the whole auditory experience. Therefore, aforementioned findings may not be always fully applicable for real world musical system.

Briefly, our formalization provides a broad framework, which can be extended in various ways. Herein, we will conclude with some of these possible proposals for future research.

Our agents are using their transition tables/expectations to compose and listen. However, they are not capable of working out the relationships between the constituents of a song. Tonal categories create the hierarchical organization in a musical piece. In a simplistic way, modality, tonality and any other hierarchical system is based on how tones are related with each other. Thus, agents could be modified in a way that they

could track how often several tones come together to find out the relationship between them.

In our model a constant learning rate is used for agents to perform table updates both for incrementing expectation values after successful interactions and decrementing after unsuccessful ones. An experiment on differing increment and decrement rates (possibly non-equal increment and decrement rates) might cause intriguing impacts on convergence trends.

Moreover, the number of winning bi-grams, which are agreed on, might be bound to simulation parameters. It might be valuable to examine whether the set of winning antecedent-consequent pairs depend on simulation parameters.

Finally, our model is suitable for studying other sequencing tasks in a broader sense. For instance, agents and interactions could be modified to tackle similar problems from the domain of evolutionary linguistics. Emergence and evolution of phonemes is such a sequencing task which is eminent to study emergence of spoken linguistic communication. With a modified version of our model, which will intend to capture a phoneme sequencing task, it can be studied whether the modeled society emerges distinctive patterns of phoneme sequences on the way to attain a social agreement.

REFERENCES

Bown, O., & Wiggins, G. A. (2005). Modelling Musical Behaviour in a Cultural-Evolutionary System. *In Proceedings of the IJCAI* (Vol. 5).

Broek, E. M. F. Van Den, & Todd, P. M. (2009). Evolution of Rhythm as an Indicator of Mate Quality. *Musicae Scientiae*, 369-386.

Brown, S. (2000). The 'Musilanguage' Model of Music Evolution. In N. L. Wallin, B. Merker, & S. Brown (Eds.), *The Origins of Music*. MIT Press.

Cohen, J. E. (1962). Information Theory and Music. *Behavioral Science*, 7(2), 137-163.

Coutinho, E., Gimenes, M., Martins, J. M., & Miranda, E. R. (2005). Computational Musicology: An Artificial Life Approach. In *Proceedings of 2005 Portuguese Conference on Artificial Intelligence*. IEEE.

Forte, A., & Gilbert, S. (1982). *Introduction to Schenkerian Analysis*. Norton: New York.

Gong, T., Zhang, Q., & Wu, H. (2005). Music Evolution in a Complex System of Interacting Agents. In Proceedings of 2005 IEEE Congress on Evolutionary Computation (Vol. 1-3). IEEE.

Holland, J. (1992). Complex Adaptive Systems. Daedalus, 121(1), 17-30.

Holland, J. (2005). Language Acquisition as a Complex Adaptive System. In J. Minnett & W.Y. Wang (Eds.), Language Acquisition, Change and Emergence: Essays in Evolutionary Linguistics. City University of Hong Kong Press.

Holland, J. (2006). Studying Complex Adaptive Systems. *Journal of Systems Science and Complexity*, 19(1), 1-8.

Lerdahl, F. (1988). Cognitive Constraints on Compositional Systems. In J. A. Sloboda (Ed.), *Generative Process in Music: The Psychology of Performance, Improvisation, and Composition* (1st ed.). Oxford University Press.

Lerdahl, F. (2009). Genesis and Architecture of the GTTM Project. *Music Perception*, 26(3), 187-194.

Lerdahl, F., & Jackendoff, R. (1996). *A Generative Theory of Tonal* Music. MIT Press.

Lerdahl, F., & Krumhansl, C. L. (2007). Modeling Tonal Tension. *Music Perception*, 24(4), 329-366.

Martins, J. M., & Miranda, E. R. (2006). A Connectionist Architecture for the Evolution of Rhythms. In *Proceedings of Applications of Evolutionary Computing* (Vol. 3907). Springer Press.

Meyer, L. B. (1957). Meaning in Music and Information Theory. *The Journal of Aesthetics and Art Criticism*, 15(4), 412-424.

Miranda, E. R. (2002). Emergent Sound Repertoires in Virtual Societies. *Computer Music Journal*, 26(2), 77-90.

Miranda, E. R. (2008). Emergent Songs by Social Robots. *Journal of Experimental & Theoretical Artificial Intelligence*, 20(4), 319-334.

Miranda, E. R., & Drouet, E. (2005). Evolution of Musical Lexicons by Singing Robots. In P. S. Dowland & S. M. Furnell (Eds.), *Advances in Networks, Computing and Communications* (Vol. 4). University of Plymouth Press.

Miranda, E. R., Kirke, A., & Zhang, Q. (2010). Artificial Evolution of Expressive Performance of Music: An Imitative Multi-Agent Systems Approach. *Computer Music Journal*, 34, 80-96.

Miranda, E. R., & Todd, P. M. (2007). Computational Evolutionary Musicology. In E. R. Miranda & J. A. Biles (Eds.), *Evolutionary Computer Music* (1st ed.). Springer Press.

Steedman, M. J. (1984). A Generative Grammar for Jazz Chord Sequences. *Music Perception*, 2(1), 52-77.

Steedman, M. J. (1996). The Blues and the Abstract Truth: Music and Mental Models. In *Mental Models in Cognitive Science* (p. 305-318). Mahwah, NJ: Erlbaum.

Steedman, M. J., & Baldridge, J. (2009). Combinatory Categorial Grammar. *Nontransformational Syntax: A Guide to Current Models*. Oxford: Blackwell.

Wallin, N. L., Merker, B., & Brown, S. (1999). *The Origins of Music*. MIT Press.

Werner, G. M., & Todd, P. M. (1997). Too Many Love Songs: Sexual Selection and the Evolution of Communication. In P. Husbands & I. Harvey (Eds.), *Fourth European Conference on Artificial Life* (p. 434-443). MIT Press.