A COMPUTATIONAL MODEL OF SOCIAL DYNAMICS OF MUSICAL AGREEMENT

A THESIS SUBMITTED TO
THE GRADUATE SCHOOL OF INFORMATICS INSTITUTE
OF
MIDDLE EAST TECHNICAL UNIVERSITY

BY

İSMET ADNAN ÖZTÜREL

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF
MASTER OF SCIENCE
IN
THE DEPARTMENT OF COGNITIVE SCIENCE

SEPTEMBER 2011

Approval of the Graduate School of Informatics

Prof. Dr. Nazife Baykal
Director

I certify that this thesis satisfies all the requirements as a thesis for the degree of Master of Science.

Prof. Dr. Deniz Zeyrek
Head of Department

This is to certify that we have read this thesis and that in our opinion it is fully adequate, in scope and quality, as a thesis for the degree of Master of Science.

Assoc. Prof. Dr. H. Cem Bozşahin
Supervisor

Examining Committee Members

Assoc. Prof. Dr. Ferda Nur Alpaslan	(METU, CENG)	
Assoc. Prof. Dr. H. Cem Bozşahin	(METU, COGS)	
Dr. Cengiz Acartürk	(METU, COGS)	
Dr. Murat Perit Çakır	(METU, COGS)	
Assoc. Prof. Dr. Ü. Erkan Mumcuoğlu	(METU, IS)	

I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.

Name, Last Name: **İSMET ADNAN ÖZTÜREL**

Signature :

ABSTRACT

A COMPUTATIONAL MODEL OF SOCIAL DYNAMICS OF MUSICAL AGREEMENT

Öztürel, İsmet Adnan
M.Sc., Department of Cognitive Science
Supervisor : Assoc. Prof. Dr. H. Cem Bozşahin

September 2011, 59 pages

Semiotic dynamics and computational evolutionary musicology literature investigate emergence and evolution of linguistic and musical conventions by using computational multi-agent complex adaptive system models. This thesis proposes a new computational evolutionary musicology model, by altering previous models of familiarity based musical interactions that try to capture evolution of songs as a co-evolutionary process through mate selection. The proposed modified familiarity game models a closed community of agents, where individuals of the society interact with each other just by using their musical expectations. With this novel methodology, it is found that constituent agents can form a musical agreement by agreeing on a shared bi-gram musical expectation scheme. This convergence is attained in a self-organizing fashion and throughout this process significant usage of n-gram melodic lines become observable. Furthermore, modified familiarity game dynamics are investigated and it is concluded that convergence trends are dependent on simulation parameters.

Keywords: computational evolutionary musicology, complex adaptive systems, musical expectation, emergence, machine learning

ÖZ

MÜZİKAL ANLAŞMANIN TOPLUMSAL DİNAMİKLERİNİN SAYISAL BİR MODELİ

Öztürel, İsmet Adnan
Yüksek Lisans, Bilişsel Bilimler Bölümü
Tez Yöneticisi : Doç. Dr. H. Cem Bozşahin

Eylül 2011, 59 sayfa

İşaretbilimsel dinamikler ve hesapsal evrimsel müzikoloji literatürü dilbilimsel ve müzikal konvansiyonların emerjansı ve evrimini hesapsal kompleks adaptif sistem modelleri ile incelemektedir. Bu çalışma yeni bir hesapsal evrimsel müzikoloji modeli önermektedir. Önerilen model daha önceki müzikal benzerlik temelli etkileşimlerle gerçekleşen şarkıların eş seçimi ile beraber evrimini inceleyen modellerin geliştirilmiş halidir. Değişitirilmiş müzikal benzerlik oyunu topluluktaki bireylerin sadece müzikal beklentilerini kullanarak etkileşimde bulundukları ajan temelli kapalı bir popülasyonu modellemeyi amaçlamaktadır. Bu yeni modelle, ajanların iki nota uzunluğundaki müzikal beklentiler üzerine bir anlaşma oluşturabildikleri gözlemlenmektedir. Bu bağlamda, müzikal anlaşma sistemin kendi kendini organize etmesiyle ortaya çıkmaktadır ve bu süreç içerisinde şarkılarda iki notadan uzun bazı melodik yapıların belirgin kullanımı gözlemlenmiştir. Buna ek olarak, önerilen oyununun dinamikleri incelenmiş ve anlaşmanın oluşumunun simülasyon parametrelerine bağlı olduğu sonucuna varılmıştır.

Anahtar Kelimeler: hesapsal evrimsel müzikoloji, kompleks adaptif sistemler, müzikal beklenti, emerjans, makineli öğrenme

To my family ...

ACKNOWLEDGMENTS

I would like to express my deepest gratitude to my advisor Dr. Cem Bozşahin for his continued interest and kind support. I am sure that this thesis would not be completed without his guidance and his willingness to share his invaluable insights.

I thank members of METU Informatics Institute for all the enjoyable moments within last three years. I especially owe a lot to Sibel Gülnar and Necla Işıklar for their assistance in completing all the paperwork without a hitch.

Last but not least, I also would like to credit Lale and Necil for their patience and for their everlasting effort in backing me up, Ayşe Naz Helvacı for her endless encouragement, Kerem Eryılmaz for all the guiding academic discussions that we carried on, Fatih Ömruuzun for his accompaniment and Emre Özyetiş for his never-ending commitment on guiding me through my life in every aspect.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

Introduction

Both language and music has a substantial social function. Individuals in a community continuously interact with each other and these interactions significantly influence how linguistic and musical conventions evolve within that society over time. In this regard, this thesis will focus on social dynamics of musical agreement on musical expectations and emergence of melodic conventions, by computationally modeling an idealized musical community and its constituent agents' interactions.

Music has hierarchical structures, just like language. By adopting the *organized sound* definition it can be defined as the sequential arrangement of sounds as building blocks. When constituent sounds are grouped together for compositional purposes relationships between them create higher-level concepts such as harmony, melody and rhythm. Moreover, music is another well structured symbolic system, other than language, which can reveal valuable information about cognitive processes. Production and perception of musical pieces require advanced cognitive abilities to structurally process hierarchical organization of individual tones in combination with each other within organized time slices. Such an organization is dependent on changes in pitch, duration, articulation and timbre of each and every tone in sequence.

One of the well accepted conceptualization of music as a rule system, in which rules translate musical structures into listeners' experience, was first proposed by Lerdahl and Jackendoff in Generative Theory of Tonal Music (GTTM) (Lerdahl & Jackendoff, 1996). It is based on psychologically plausible structures, where organization of musical elements are investigated in a multi-layered representational medium. Varying modes of abstraction separately realize each layer of representation and underlying assumptions are claimed to be empirically testable.

Representational layers in GTTM are constructed over the notions of hierarchies and associations (Lerdahl, 2009). Briefly, hierarchies capture the structural organization of fundamental musical objects, whereas associations work out the similarities between these musical objects. Therefore, theory of associations, or namely theory of similarity, and theory of hierarchies are strictly bound to each other.

GTTM is composed of four distinct hierarchical structure layers, which covers all tonal compositions (Lerdahl & Krumhansl, 2007). Grouping structure explains apparent perceptual segmentation of music into subgroups. Metrical structure deals with rhythmic segmentation of a group. In accordance, time-span reduction hierarchically positions rhythmic segments of metrical structures according to their importance. Finally, prolongation reduction generates hierarchies of tension relaxation patterns. In order to comprehensively study complex structures of music all these layers of explanation must come together to provide an essential basis. In fact, this four layered approach stems from Jackendoff's conceptualization of language. As it is proposed in Jackendoff (2002), a simple linguistic utterance also consists four generative layers, namely phonological structure, syntactic structure, semantic/conceptual structure and spatial structure. Accordingly, all these distinct structural layers has to be connected by using interface rules.

On the other hand, Combinatory Categorial Grammars (CCG) can provide a pure formalization for higher-level structures such as harmonic movements over chord sequencing instead of examining them in distinct layers, as it allows bi-direction resolution of the rules. CCGs are distinguishable from other formalisms as they adopt functional composition and type-raising schemes (Steedman, 1996). Functional composition allows us to follow up an incremental derivation, where the syntactic type of the constituent formed throughout the derivation is determined according to all the chords encountered so far. Furthermore, functional type raising allows us to perform a function application in both forward and backward directions. When applied to music, CCG can capture the harmonic disposition rules successfully.

However, just like its precursors such as Schenkerian Analyis, GTTM and similar generative and combinatory theories of music are just interested in explaining only well-structured musical pieces of tonal culture (Forte & Gilbert, 1982; Lerdahl & Jackendoff, 1996; Steedman, 1984). Still some other social aspects of music cognition needs to be studied conveniently. Research questions like "*How shared sound systems emerge?*", "*How does 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).[1] From the listeners perspective, compositional rules are not directly accessible, if not explicitly presented. However, organizational rules between the musical events can be reconstructed by the listener if listener has a familiarity with the structural organization of the heard piece.[2] 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.[3] 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 used. Besides, new musical styles emerge throughout time within a society and they impose an expansion in the set of musical conventions.

---

[1] Terms *compositional grammar* and *listening grammar* was first introduced by Fred Lerdahl in his essay "Cognitive Constraints on Compositional Systems" in 1989. *Compositional grammar* is the set of rules used by the composer, which regulates the organization and sequencing of musical events. On contrary, *listening grammar* can be defined as the intuitive constraints of the auditor, which is used for generating a structural description of the heard music from the raw acoustic signal.

[2] For instance, Fred Lerdahl pinpoints some examples of the early serialism that are structurally hidden to the listener because an agreement between the compositional and listening grammars could not be attained.

[3] Culturally shared musical conventions are referred as *natural music grammars* in contrast with *artificial music grammars*, which are invented and superimposed on the society on purpose by the composer. For instance, within the following 20 years after the birth of electro-acoustic music in late 1940s some composers started practising methods of algorithmic composition. Structural organization of these pieces can be considered as examples of artificial music grammar products for the European society, which has an extensive background on western tonality.

Keeping all these in mind, it can be inferred that musical systems can not be grasped by only modeling cognitive abilities of individuals of a specific culture. Similar to language, music is also 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.

Complex adaptive systems (CAS) and their computational models have proven to be successful for investigating learning abilities of dynamical non-deterministic systems. Computational CAS models are capable of capturing overall behavior of the system with respect to interactions of its individual constituent components. Within the literature essential methodology of computational modeling of CAS have also been adopted by social linguistics domain to research emergence and evolution of form, meaning, form-meaning association and hierarchical grammatical structures (Steels, 2000). Successively, computational models on emergence of shared sounds systems and evolution of musical structures have ensued this research by modifying the same methodology for it to be applicable to musical contexts, where interactions between the constituents did not involve any indexicality and semantic content (Miranda & Todd, 2007). Overall, this line of research is promising for studying language and music as a social tool and it can reveal intriguing facts about linguistic communication and music cognition.

In correlation, this thesis will present a model of musical interactions and agents which are captured as a complex dynamical system. Briefly, the scope 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.

The organization of this thesis is as follows:

Chapter 2, aims to provide a detailed overview of CAS. Subsequently, two of the essential features, namely emergence and self-organization, are reviewed. In correlation, last section of this chapter presents a methodology for designing computational models of CAS.

Chapter 3, focuses on computation models of evolutionary linguistics and computational evolutionary musicology. A domain specific methodology is overlaid and current literature on emergence and evolution of linguistic and musical conventions is elaborated.

Chapter 4, presents the model and relevant motives behind this study. Moreover, empirical results of the experiments that were conducted are also given in this chapter.

Chapter 5, presents a discussion of major contributions to the field and possible future research over the findings.

CHAPTER 2

Background on Complex Adaptive Systems

Complex adaptive systems (CAS) are nonlinear dynamical multi-agent systems that are highly adaptive as constituent parts learn while they interact (Holland, 2006). Many recent problems in domains like economics, biology, linguistics and musicology can be investigated by modeling them as CAS.

Within this chapter shared characteristics of CAS will be presented in Section 2.1. Two of these distinctive characteristics, namely emergence and self-organization, are often mistakenly used interchangeably within the literature. To clarify this dichotomy Section 2.2 will present essential conditions that are required for emergence and self-organization to arise. Notably, throughout this review working definitions for these concepts will be borrowed from De Wolf and Holvoet (2005). Finally, Section 2.3 will present the essential methodology to computationally model complex adative systems. This formalization will be excerpted from Holland (2006).

2.1 Common Properties

Dynamic multi-agent systems can be classified as CAS if large number of local interactions create an adaptive behavior. Collective adaptive behavior give 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. Thus, 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). Under such complexity, studying unexpectedly generated macro-level effects instead of distinct micro-level interactions becomes favorable.

Briefly, shared characteristics of these systems can be identified as (Holland, 1992):

- **Evolution:** For the individual constituents to survive in the environment that they are acting on, they exhibit an evolutionary behavior. Evolutionary process can be attributed to their eager in adaptation. Within the literature evolutionary trend is grounded in agents ability to learn.

- **Aggregate Behavior:** Discrete interactions between the agents can impose an unexpected emergent aggregate system-wide behavior. In this sense, aggregation does not connote a straightforward add up of the behavior of the parts, rather it is the property that supervenes on macro-level over coherent local interactions. Motives of this aggregate behavior could not always be traced back in micro-level interactions. However, it can effect local interactions once it is spread to the system.

- **Anticipation:** CAS exhibit an expectation for interventions. Moreover, they have the ability to respond to any perturbation. Anticipation of the system assures flexibility to be susceptible for reorganization.

- **Simultaneous Interaction:** Local interactions between the agents involves exchange of signals.[1] Signal processing is a conditional assessment procedure. Agents receive signals as inputs and if that signal satisfies an internal condition, either an action is performed on the environment or an output signal is produced and transmitted to the corresponding interacting parties. In correlation, agents can simultaneously create more than one action or signal as an output if more than one internal condition is satisfied throughout the assessment of the input signal. Moreover, signal exchanging and interactions are concurrent to ensure parallelism.

- **Self-Similarity:** Interacting agents use a common decision making policy to generate or process signals. However, agents does not always have to take the same learning path. Therefore, within CAS the essential medium for agents to interact is successfully attained with self-similarity, while conserving individual differences regarding

[1] In real world complex adaptive systems signals can come in any form as long as they carry necessary information to complete an interaction. For instance, if we take immune system as a complex adaptive system, then exchange of proteins between antibodies will be their fundamental method of signaling.

their past experiences. Furthermore, self-similarity is also necessary for aggregate behavior to emerge. That is to say, a certain behavior which is adopted by a certain agent can only be spread to the system if other agents are capable of performing the same behavior.

2.2 Emergence and Self-Organization

2.2.1 Emergence

The concept of parts summing up to form a greater whole is deeply rooted in a great deal of research after twentieth century. Philosophical foundation of emergence can be traced back to Aristotle as he claimed "*the totality is not, as it were, a mere heap, but the whole is something besides the parts*" to describe that an organism possesses all the required internal principles in its parts and growth is a process of transformation over these principles to reach a state of perfection (Clayton, 2006). Contemporary understanding of the concept originated with G.H. Lewes's scientific research. Lewes first pointed out emergent behavior in chemical reactions by stating that observable resultants may be created by a series of chemical reactions, though underlying steps of the chemical process to produce these resultants may not be identifiable (De Wolf & Holvoet, 2005).

Thereafter, throughout the twentieth century theories about emergence disunite and group under two categories, that are weak and strong emergence. Specifically, *strong emergence*, or as it is also called 'ontological emergence', asserts that new causal processes related with the emergent property, which are distinct from the causal processes that gave rise to it, are created upon the origination of it. To put it in different words, emergence of the high-level phenomenon creates its own truths that are not deducible from the truths of the low-level domain which gave rise to it (Chalmers, 2006). On the contrary, *weak emergence*, or 'epistemological emergence', suggests that the causal processes which creates the emergent property is also applicable for the emergent property itself. In this case, emergent phenomenon and the truths of the high-level domain are just unexpected with respect to the truths of the low-level domain (Chalmers, 2006).

Conceptual disparity between strong and weak emergenism is concerned with the measure of emergence that is observed throughout the process. A strongly emergent phenomenon might also be identified as weakly emergent considering that it can arise unexpectedly while its truths are not deducible from the truths of low-level domain. Yet, an instance of weak emergence can not always be a case of strong emergence, because being unexpected does not always make certain that emergent phenomenon has its own causal processes independent of the low-level domain.

To be clear, strongly emergent qualities can be explained by only establishing high and low level domains as different levels of nature. In real world systems strong emergence can only occur if and only if physical laws governing the low-level gives rise to a property which can not be explained over physical causal relationships. David Chalmers as a philosopher, who has an extensive research on consciousness, argues that strongly emergent phenomenon can be exemplified with conscious experience (Chalmers, 2006). Substantially, what he proposes is that given the same two physical low-level systems in two possibly different worlds, a strongly emergent phenomenon like consciousness would not necessarily emerge in both of these worlds. To put it in another way, it is not possible to emerge consciousness over a replica of physical states of a conscious individual. That is to say, strongly emergent phenomenon supervenes low-level physical facts, however there is an explanatory gap between the high-level and low-level domain so that the facts about the emergent property are not deducible from the facts of physical causality.

As it is claimed by Chalmers (2006) consciousness is the only real world example that is found to be strongly emergent. All other emergent real world systems are weakly emergent, which means supervenient emergent property has downward causal power on the low-level physical domain and they are reducible to the interactions of physical agencies. In fact, even consciousness is a doubtful example for strong emergence and it poses open-ended problems, since there is not a unique definition for conscious experience. From a relativist point of view if we accept that consciousness comes in different degrees, then everything must be conscious to some extend (Minsky, 2006). For instance, it could be assumed that even atoms have a little degree of consciousness and so human consciousness could be reduced to that. Hence, with a relativist explanation consciousness could asserted to be an instance of weak emergence. Accordingly, consciousness can only be strongly emergent if we presuppose that only humans are conscious. For this reason, for the rest of this section we will leave strong emergenism aside and focus on weak emergence to elaborate the concept accordingly. In that respect, a working definition for emergence could be drawn as (De Wolf & Holvoet, 2005) :

"*A system exhibits emergence when there are coherent emergents at the macro-level that dynamically arise from the interactions between the parts at the micro-level. Such emergents are novel with respect to the individual parts of the system.*"

All in all, essential conditions for emergence to rise is as follows:

- **Micro-Macro Effect:** The working definition mentions about macro-level products, namely emergents. Emergents are the structural, functional or behavioral outcomes in macro-level as a result of micro-level interactions between the agents. So to speak, for emergence to occur local phenomenon must effect global behavior.
- **Radical Novelty:** Agents should not have an embedded behavioral bias, which can impose a certain macro-level effect. Emergent property should be novel in regard to microscopic interactions and it can only arise as an effect of collective behavior of constituents of the system.
- **Coherence:** A behavior is emergent if it is identifiable in the system. Therefore, local interactions must come to a state of global coherence so that the relationship between the parts could represent a unity. Strictly speaking, emergent behavior can only disperse to the system if parts are logically consistent.
- **Interacting Parts:** Constituent parts must interact in micro-level for emergents to arise in macro-level. Accordingly, interaction is a fundamental prerequisite for emergence.
- **Dynamical:** Dynamicity is an obvious requirement for emergence. Emergents can only originate in course of time. Emergent property can not be attained instantaneously. Behavior of the parts should first dynamically and coherently coincide.
- **Decentralized Control:** Macro-level behavior can not be centrally controlled. Behavior of parts can be controlled in isolation, but none of them can directly impose a certain change of behavior in macro-level. This is because, agents can not have a representation of the macro-level behavior.
- **Two-Way Link:** Between micro and macro levels there needs to be a bidirectional causal relationship. As a consequence of micro-macro effect, local interactions cause higher-level emergent behavior. On the other hand, a certain behavior which emerges on macro-level can also causally effect the behavior of its constituent parts.

- **Robustness and Flexibility:** Emergent behavior is relatively persistent. Functional deficiency of micro-level agents does not massively effect it. System preserves it in case of replacement or failure of individual agents. Therefore, emergence is flexible and fault tolerant with respect to sudden local behavioral alterations.

2.2.2 Self-Organization

Just like emergence, the notion of self-organization is not novel. The concept was first entitled in Second World War after advancements in cybernetics (De Wolf & Holvoet, 2005). Generally, self-organization can be described as the adaptive behavior of the system to attain a future state of increased structure or order compared to the starting state. The driving force to reorganize the structure of the system can be extrinsic, thus a self-organizing system is open for external input.

The concept could be formally defined as (De Wolf & Holvoet, 2005) :

> "*Self-organization is a dynamical and adaptive process where systems acquire and maintain structure themselves, without external control.*"

Accordingly, necessary and sufficient conditions for self-organization to arise in a dynamic system is as follows:

- **Increase in Order:** Self-organizing systems strive to increase the order of the system to reach a state in which the system can perform a specific function. In other words, system gradually arranges its parts to reach an objective state, which is called the attractor, among its state space. Specifically, the attractor state should not be the one which assures maximal order. Redundantly ordered systems can also foster lack of useful functional behavior, just like chaotic systems.

- **Autonomy:** Increase in order is not sufficient for a system to be self-organizing. Agents must also be autonomous, meaning that their behavior should not be governed by external regulations or centrally controlling agents. Accordingly, imposing such global rules on the system would lead the behavior of the organization in a specific direction on macro-level, therefore spontaneous micro-level interactions of the agents would not be the primary driving force to readjust the system structure. However, it should not be deduced that self-organizing systems are closed to external interventions. Intrusion of external input is permissible, nevertheless they should not have any regulatory information in them. Strictly speaking, agents of self-organizing systems are close to any type of controlling data. For instance, agents can acquire perceptual data, which is explanatory about the environment, yet that should not involve any controlling impact on overall system behavior.

- **Robustness:** Within this context, robustness refers to agents' ability to alter system behavior in terms of adaptability while keeping their structural characteristics constant. System must have an adaptive character with respect to changes.

- **Dynamical:** A self-organizing system must be capable of spontaneously reacting to disturbances as its agents autonomously organize their behavior. Self-organization is not instantaneous nor imposed. Change of state to increase system order is a continuum.

2.2.3 Emergence vs. Self-Organization

Emergence and self-organization might be used to describe the overall behavior of a dynamical multi-agent system. Both phenomena occur dynamically as a result of interacting agents, which means that both processes are dependent on time. Moreover, they are both robust with respect to systems durability on adopting the aggregate behavior and its agents capability to learn and adapt to changes. However, as it can be inferred from previously presented definitions emergence is the process which promotes an unexpected system-wide behavior on macro-level as a result of micro-level interactions of the constituting agents, whereas self-organization is systems ability to adapt to a certain condition by internal reorganization. Nevertheless, they are not the same phenomenon and they do not necessarily be observed together. Herein, it would be appropriate to use illustrative examples to point out the dissimilarity. The examples that are going to be presented below may not be the ones that are solely emergent or self-organizing, since most of the complex adaptive systems employ both emergent and self-organizing character together. Still, emergent or self-organizing nature of the below presented examples are explicitly observable.

Perceptrons, particularly simple feed-forward neural networks, could be representative for emergence. Perceptrons consist of logical units as building blocks that are functionally separable, in such a way that they do not have an internal representation of the whole network.

Nonetheless, extensive exposure to a specific set of data can effectuate a specific behavior on the entire network. In particular, perceptrons adequacy in learning and solving the XOR problem is well-known. Without training on XOR data set, parts could not perform a coherent behavior. Throughout the training, logical units interact with each other in direction of input to output units iteratively in the absence of a central controlling unit. As a consequence of these recurring interactions, perceptron can emerge a computational competence to XOR a given input data. In this sense, throughout its learning perceptrons structure is not reformed with external interventions, for this reason emergence of its computational adequacy can be classified as an instance of weak emergence.

One of the most prominent example for self-organization is the immune system. Immune system consists of simultaneously active components called antibodies. Functionally, antibodies need to discriminate offensive intruders called antigens and take preventive action against them. However, system do not employ a global predefined list of threats and precautions, rather it dynamically adapts. Throughout the development, immune system dynamically gets better and better in performing this task. For instance, when a specific antigen of an illness, let's say flu viruses, intrudes to the body the amount of antibodies that are required to surmount flu is dynamically specified by the system, if its historical memory contains information about the required preventive action for that specific type of antigen. Historical memory is built up in time as parts of the system generally learn how to combat with an unknown intruder after the first encounter by interacting with each other. Such an internal organization is continuously carried on to attain the attractor state of keeping the organism disease free.

2.3 Computational Modeling: Classifier Systems

Due to non-linear nature of adaptive process in CAS, data-driven or deterministic models are not suitable for analyzing internal dynamics of these systems. CAS can be studied in computational exploratory models which facilitate computer based programs or simulations. At the outset, a model of interacting agents must be built up by using specific assumptions. Subsequently, significant alternations in system behavior that become apparent as a consequence of agent interactions can be observed by executing these models as computer simulations. Such an approach is exploratory, since we are testing whether the underlying assumptions about agents and interactions that are embedded in our model cause any interesting overall effect on the system.

Holland (2006) specifies *classifier systems* as the essential formalism that can be used to devise exploratory computational models. A classifier system to define agents must involve classifiers, reservoirs, detectors, effectors and signals. Each of these principle components will be elaborated below.

As it is mentioned earlier in Section 2.1 agents communicate over signals. Throughout an interaction agents receive a signal from its counterparts and they assess that signal by using a conditional evaluation scheme. Meaning that, every agent adopt a set of IF-THEN structured decision making rules. This decisive rule set is called the set of *classifiers*. Indeed, if an input satisfies the conditional IF part of a classifier from the set of classifiers agents compose an output signal by using the corresponding THEN block. Specifically, if an input signal satisfies more than one classifier from the set, then the output signal may be generated in accordance with all the classifiers that it satisfies.

Set of classifiers of an agent determine its expectations while interacting. That is to say an agent's needs are satisfied as long as it can satisfy its classifiers so that it can increase its performance. It could be deduced that an agent's needs are dependent on the set of rules that it is knowledgeable about. In this respect, set of *reservoirs* is used to keep track of how much an agent satisfies its needs. If it can satisfy a specific classifier, corresponding reservoir for that classifier from the set of reservoirs can be raised. Otherwise, reservoirs diminish over time.

Agents use *detectors* to retrieve environmental data. Data taken as an input by using detectors are turned into signals. Detectors can come as perceptual channels, which can collect information about environment in different modalities (such as visual, auditory, etc). In addition, signals are received over detectors when interacting with another agent. Likewise, *effectors* are used by agents to act on the environment which they are surrounded by. That is, effectors are the action inducers for agents to cause a change in the environment. Just like detectors agents can effect the environment in various modes. For instance, effectors may be used to move in space or to change the place of an object.

Signals are stored in the *list of signals*. Members of this list can either be created by the agent itself (i.e. it can be created as a notifier signal when reservoirs are about to be emptied) or received externally by using detectors. If a signal is gathered externally through detectors, first it has to satisfy a classifier to be inserted into the list of signals. Hence, list of signals define internal states of the agents. Accordingly, all possible actions that an agent can take in a given state is determined by the collection of signals and classifiers it has.

The set of rules that an agent is aware of is not static. Rules in the classifiers set compete with each other. Classifiers are associated with a strength value to measure how much each and every classifier contributes to the overall performance of the agent. Classifiers could be strengthened if they induce a change in the environment with effectors, or if they are repeatedly satisfied by the input signal which is collected through detectors. Even more, agents can come up with new rules and classifiers. Though, rule discovery is not a random process. For instance, new rules can be generated by slightly altering strong classifiers. By doing so agents can discover alternate states within the state space.

To this end, interactions between the agents are iterative and possibly concurrent cycles of a classifier system. However, interaction designs may vary depending on the system and problem that we are dealing with. In accordance, following chapter will thoroughly investigate similar computational models that deal with emergence and evolution of linguistic and musical conventions.

CHAPTER 3

Background on Emergence and Evolution of Language and Music

As it is presented in previous section, exploratory simulations are convenient tools for constructing multi-agent systems that may present emergent and self-organizing behavior, where constituents are fully equipped to cooperatively act on a shared environment. Similar artificial distributed multi-agent simulations 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).[1] Both lines of research investigate emergent behaviors of societies, where complex local interactions between individuals effect global organization. Underlying methodology for this investigation closely resembles the classifier systems approach.

This chapter will present a detailed review of models of these two domains. Section 3.1 will overlay essential methodological components in correlation with classifier systems. Following sections will detail some of the noteworthy findings of the domain to make the method-

[1] Accordingly, a great deal of research, which are classified under computational evolutionary musicology, investigates practical automatized music generation methods over evolutionary models. These models will be left out of the scope for this chapter and only theoretical frameworks will be reviewed.

ological approach clearer. Section 3.2 will present how shared sound systems can emerge. Section 3.3 will focus on evolution of musical behavior. Finally, Section 3.4 will elaborate how shared vocabularies can emerge.

3.1 Methodological Issues

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

Within this context, agents of a population is a computational abstraction, that are usually implemented as robots. An agent can both play roles of speaker and hearer or composer and listener in different instances of interactions depending on the model. Members of the population are alike to assure self-similarity. Following premises would be descriptive to outline the characteristics of an agent:

- **Psychological Competence**: Agents communicate over signals. An agent must be capable of producing signals to externally transmit to the peers that it is interacting with. Moreover, it must also be capable to listen an external signal input. Therefore, agents must have sensori-motor apparatus which can be used as detectors and effectors.
- **Cognitive Competence**: An agent can conceptualize specific linguistic or musical skills depending on the game that it is engaged in. To be more precise, it must be knowledgeable of what a signal designates and how to compose a signal from lower-level components. Furthermore, it should define a list of signals that can serve as a memory.
- **Enacting Script:** Agents should know how to interact with each other and how to asses a perceived signal that is externalized by a peer. Meaning that agents should have a list of classifiers that specifies its possible actions in a given state.
- **Discrete Memory**: An agent is not aware of its peers mental states or memory. Particularly, if it is playing the role of being a hearer, then it can only be informed about the

externalized signal that is produced by its match.

Respectively, an agent takes following actions through an interaction:

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

While designing a game in this context, it could be assumed that musical pieces do not carry a referential semantic content in contrary to linguistic utterances.[2] This dissimilarity evokes a substantial variation between the models of language and music.

Rules of language games relate agents to their environment so that they can agree on linguistic conventions which are descriptive about their environment. Hence, language game models involve a non-restrictive environment, which can accommodate numerous different states as agents interact through it. Moreover, communicative signals are token type and they generally represent a word.

Constrastively, models of computational evolutionary musicology are non-situated, meaning that musical games 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, former definition of music presented in Chapter 1 emphasizes 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).

It could be argued that foregoing formalization of musical interactions can not capture some other substantial aspects of music cognition. For instance, non-disputably a musical interaction culminate with an emotional response on the listener on individual basis. However,

[2] In fact, debates about affective meaning of music are of long-standing (Wiggins, 1998). To simplify, hereafter musical affect will be reduced only to the expectations of subsequent harmonic or melodic structures that are induced on the listener.

models of computational evolutionary musicology exclude such aspects with an abstraction. Precisely, it is just questioned how much about emergence and evolution of musical conventions can be revealed within a fully abstracted musical society. So it could be deduced that these models, are relevant with real world musical interactions as much as computational evolutionary linguistics models are relevant with natural languages.

Taking all these into consideration, rest of this section will unfold various types of language and musical games.

3.1.1 Language Games

The term *language game* was first used by Wittgenstein to characterize linguistic communication (Wittgenstein, 1953). Wittgenstein construed linguistic meaning over the use of linguistic tokens. In depth, use of a sentence or even a word involves an act of using signs. Therefore, for a successful linguistic communication to take place all participants must engage in a shared activity that should guarantee that they league together in a common ground, which he depicts as a game-like activity (Biletzki & Matar, 2011).[3] Intentions of the utterer and the meaning conveyed by that specific utterance can only grasped if hearer gets involved in a so-called 'language game' with the speaker. Wittgenstein does not provide a clear cut definition for the game. However, he outlines the basic characteristics of language games and linguistic communication with the following two assertions, which will also serve as a basis to understand the underlying assumptions of models of computational evolutionary linguistics:

- **Form of Life:** Specifically, language is meaningful when used within a social context. Wittgenstein refers to the collection of social contexts that have its own set of rule applications as "forms of life". By conceptualizing linguistic communication over social contexts, it is presumed that an utterance can have different meanings in different societies or contextual settling. This also represents the dynamic nature of language, as forms of life is constantly subjected to change.

- **Rule Following:** Linguistic communication is a rule-following activity. A rule is an abstract entity, which comes with its set of possible applications. However, rules do not

[3] See *PI2* in Wittgenstein's Philosophical Investigations for a detailed description of the builders' language-game, which presents an idealized minimal linguistic communication that takes place between two participants to achieve a common aim.

give clear descriptions for the corresponding real world actions. An utterance conveys meaning if it satisfies expectations that arise with the corresponding rule application for that utterance in a given form of life. Accordingly, one should be aware of the socially shared rules and their applications to perform linguistic communication which fulfills his/her intentions.

Within the literature of computational modeling of language as a complex adaptive system, various derivatives of language games are explored to investigate emergence and evolution of phonetics, semantics and grammar by adopting the above mentioned assumptions to conceptualize linguistic communication as an ever changing dynamical process (Steels, 2000). Correspondingly, a language game model can be outlined as :

1. A pair of speaker and hearer is selected randomly among the population.
2. Both parties attend to a topic, which could be anything related with the environment (e.g. an object). Topic must be perceivable by both parties.
3. Speaker transmits a feature about the topic to the hearer.
4. Hearer assesses whether it has that feature associated with the topic in its memory.
5. Both parties update their memory in accordance with the result of this assessment.

Well studied representative examples of language games are:

- **Discrimination Game:** Discrimination games focus on creation of shared meaning repertoires (Steels, 1996b). Briefly, speaker attends to a portion of the perceptual space and shares this selection with the hearer as the context. Within the context an object or a perceptual entity is chosen as the topic by the speaker. Speaker transmits a word that represents a feature about the topic from its memory to the hearer. The word which is externalized by the speaker should supposedly be distinguishing for the topic from the context. Hearer tries to identify the topic in the context by using the word provided by the speaker. Both parties are informed about the success of this identification and they perform memory updates accordingly. Throughout the game agents form categorization trees. Each category denotes a specific feature set and words are members of these feature sets. For instance, *color* of an object is internally represented as a category and words like *red, blue, green* would be categorized as color features. While interacting agents adjusts these category trees to carve out a finer representation of the perceptual space.

- **Naming Game:** Both speaker and hearer attends to an object from the environment. Speaker utters a name referring to that object and hearer checks whether it has encountered a use of that specific name denoting that object. Participating agents try to agree on a name for the objects of the environment through iterative interactions. Naming game models aim to show that agents can converge on shared vocabularies for separately identifiable objects (Baronchelli et al., 2005; Baronchelli, Felici, et al., 2006; Baronchelli, Dall'Asta, et al., 2006; De Vylder & Tuyls, 2006; Steels, 1996a; De Vylder & Tuyls, 2006). Game also allow a topological investigation over the population structure (Lenaerts et al., 2005).

3.1.2 Musical Games[4]

In correlation with language games, a broad generalization of musical games could be drawn as:

1. Performer(s) and listener(s) are selected randomly among the population.
2. Either an imitation or familiarity based interaction occurs between performer(s) and listener(s).
3. Success of the interaction is assessed by one of the participating parties depending on the type of the game.
4. According to the result of this assessment musical knowledge of participants are updated.

Assessment and knowledge updating methods can differ according to the type of the musical interaction. Within the literature most common evaluation and updating methods can be exemplified with the following games:

[4] Although previous studies on emergence and evolution of musical conventions does not name their models of musical interactions as *musical games*, the term will be introduced here for the first time by taking the methodological similarity between models of musical interactions and linguistic communication into consideration. For the rest of this thesis we will refer musical interactions as musical games.

- **Imitation Game:** Listener tries to imitate the musical signal generated by the performer. Performer listens the imitation and it compares the imitation with the original signal by evaluating how similar they are. Imitation games can be used to bootstrap shared repertoires of sounds among musical agents (Miranda, 2002; Miranda & Drouet, 2005; Miranda, 2008). Conventional performance dynamics such as certain forms of pitch, duration and amplitude alternations of a sound can also emerge and evolve in a society via imitation (Miranda et al., 2010). Moreover, there are also computational evolutionary linguistics models which implement imitation game to investigate how shared sound systems like vowels and consonants emerge (De Boer, 1997).

- **Familiarity Game:** Listener assesses the musical signal created by the performer by considering how familiar it is with the organizational structure of the signal with respect to its past experiences that have shaped its musical expectations. Accordingly, quality of a performer can be evaluated by assessing whether its songs are pleasant enough in the way of satisfying listener's expectations. Current literature on familiarity games are based on the assumption that music is a co-evolutionary process. The game is commonly implemented to see how musical behavior and expectations evolve in a population through pair selection with mate quality from generation to generation (Werner & Todd, 1997; Bown & Wiggins, 2005; Gong et al., 2005). Recent studies also show that novel rhythmic forms can emerge in cultural-evolutionary systems where mating is carried on through a mate quality assessment (Broek & Todd, 2009). In parallel, the model that is going to be presented in Chapter 4 will investigate whether a closed community of agents can agree on a globally shared musical expectation scheme on the way to emerge melodic conventions by playing a modified familiarity game. Distinctively, instead of working out evolution of musical expectations over generations as a co-evolutionary process like models of Werner and Todd (1997) and Bown and Wiggins (2005), the model which is going to be presented in the upcoming chapter will just focus on social dynamics of musical agreement of expectations within one generation.

3.2 Imitation Game: Emergence of a Shared Repertoire of Sounds

As discussed earlier in Chapter 1, 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*. That is to say, 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).

An imitation game 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, or specifically an imitation game occurs between two randomly chosen agents, where one is the performer and the other is the imitator. Before going any further in describing imitation game dynamics in details, it would be handy to elaborate psychological and cognitive capabilities of agents used in this model to apprehend how they compose and perceive musical signals throughout an interaction.

Agents of this simulation are robot implementations and the musical signal (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 use these values to synthesize the sound that is intended to be generated.

Overall, within the imitation game as sketched out by Miranda (2002), an interaction goes as follows:

1. A random performer and imitator is chosen among the population.

2. Performer plays a sound from its lexicon by choosing a perceptual representation from the lexicon and articulating its corresponded motor representation. If its lexicon is empty, then a random sound is generated.

3. Imitator extracts a perceptual representation from the heard sound. To extract pitch and loudness of the heard sound, a short sample is taken from the head of the signal and then rest of the signal is overviewed whether it has recurring fragments that match with the sample. If there is a match, periodicity and frequency of the signal can be calculated by using the time interval between the sample and matching fragments. Sampling and comparison procedure is repeated for varying window sizes. This extraction methodology is named as efficient short-term auto-correlation method (Boersma, 1993). Duration of the signal is just estimated from the raw input. 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.

4. 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.

5. Both performer and imitator updates their lexicon regarding the success of the imitation. If the sound and its imitation is 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.

3.3 Familiarity Game: Evolution of Musical Behavior

Music carries an information content, which is embodied in its structure. Meyer (1957) proposed that the information content of a musical piece can be correlated with the measure of fulfilled *musical expectations*, which arise dynamically all through a real-time exposure to the musical structure itself. In general, musical expectations originate above the relationship between antecedent and consequent musical events, which are following each other successively, within a temporal frame as a Markov process. In other words, if a consequent event which follows a specific antecedent is assorted to be unexpected then information content of that structure increases. Notably, musical expectations of an individual are shaped with their previous musical experiences. That is to say, musical expectations may not be the same for every individual, since they will not have the same amount of familiarity for each structure they hear. Therefore, it could be claimed that information content of a specific musical structure is individually dependent on auditors familiarity to that structure.

Accordingly, musical agents can revise their compositional routines to appeal a wider range of listener populace. Musical structures with abundant information content may be enjoyed by a wider range of audience, which can embed them into that culture over time. Known and commonly used transitions between musical events can cause so called boredom on the listener, yet presumably listeners can also get distracted if the structure is totally unfamiliar. For this reason, compositional practice to equate familiarity and surprise rests on a fine balance. Diversity of musical conventions in a culture can be an indication for the pursuit of finding surprising elements in music. Therefore, it is admissible to predict that listeners tend to seek for some constrained amount of surprise within the structure.

Werner and Todd (1997) argues 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, which encodes musical expectations of their own. To clarify, transition table contains information about how probable an antecedent-consequent note sequence is. 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. In one round of musical interaction all females engage in a familiarity game, to choose a mate. From one of the females perspective this game is:

1. Predefined number of performers are randomly selected among males.
2. Each selected male plays its song to the listener.
3. 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. A formal clarification for these assessment methodologies is presented below.

Let $A = \{M, F\}$ be the set of agents, which consists two different groups of agents and they are all aware of a global closed lexicon of notes $L = \{n_1, ..., n_i\}$, where $N_L = i$.

- **Performers** $M = \{p_1, ..., p_j\}$ is the set of male agents with size $N_M = j$. Each performer

p has a genotype $G = [N_1, ..., N_m]$ an array of notes, where $N \in L$, $N_{G_i} = m$ and $N_{G_i} > N_L$.

- **Listeners** $F = \{l_1, ..., l_k\}$ is the set of female agents with size $N_F = k$. Each listener l has a corresponding transition table T, which encodes the musical expectations of that listener. T is a $N_L * N_L$ sized two dimensional array, such as:

$$T = \begin{bmatrix} \alpha_{1,1} & \cdots & \alpha_{1,N_L} \\ \vdots & \ddots & \vdots \\ \alpha_{1,N_L} & \cdots & \alpha_{N_L,N_L} \end{bmatrix}$$

In T, $\alpha_{x,y}$ holds the value for the probability of n_x followed by n_y with in the heard signal, according to the listeners expectations. For instance, if listener l_3 has the value of $\alpha_{5,12} = 0.01$ in T_3, then it could be considered as l_3 does not expect to hear an instance of note n_{12} after n_5. Whereas, if this value was $\alpha_{5,12} = 0.99$, after l_3 hears n_5 in the signal it would most probably expect a n_{12} to succeed it. In this case, if it turns out that a note other than n_{12} is the successor, l_3 would be surprised.

Regarding this formalization when a performer p_i plays its song G_i the listener l_j, it could be evaluated by using following scoring policies:

- **Local Score:** is the score given by the listener to a heard melodic signal, by summing up the individual expectancy values for each transition in real-time, as it is presented in (3.1).

$$local\ score = \sum_{k=1}^{N_{G_i}-1} \alpha_{G_i[k],G_i[k+1]} \tag{3.1}$$

- **Global Score:** is the overall score of a song when it is evaluated as a whole. To compute the global score, listener listens the whole song. Consequently, the number of each possible transition is calculated and a transition table for that individual song is created such as:

$$T_{G_i} = \begin{bmatrix} \alpha_{1,1} & \cdots & \alpha_{1,N_L} \\ \vdots & \ddots & \vdots \\ \alpha_{1,N_L} & \cdots & \alpha_{N_L,N_L} \end{bmatrix}$$

To demonstrate, $\alpha_{x,y}$ in T_{G_i} is the probability of n_x followed by n_y in G_i. In fact, probability values of T_{G_i} should not be confused with listeners l_j's musical expectations T_j. These are only song specific distribution values. Subsequently, the global score of the song is determined regarding the similarity between song specific probability distribution and listeners expectation matrix, as it is presented in (3.2).

$$global\ score = T_j - T_{G_i} \tag{3.2}$$

- **Surprise Score:** is the summation of the amount of surprise raised by each transition between the notes within the heard piece, as it is presented in (3.3). To compute it, let *max()* be the function that would return the probability for the most expected consequent note for a specific antecedent. This would be used by the agent to find the probability of its expectation for the upcoming note, when a note is heard in real-time. In correlation, the amount of surprise while listening a transition, is the difference of the probability of the most expected transition and the probability of the actual antecedent-consequent pair.

$$surprise\ score = \sum_{k=1}^{N_{G_i}-1} max(G_i[k]) - \alpha_{G_i[k],G_i[k+1]} \tag{3.3}$$

Among these three evaluation schemes global and local scoring policies resulted in a similar evolutionary trend for the songs of the modeled society. In short, these two scoring policies can not generate a diversity in the musical signal over generations, therefore they are not adequate to explain assorted forms of real world musical systems. Accordingly, Werner and Todd (1997) claims that 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 just 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 competent to both perform and listen, though they do not have a sex which dictates a role on them.

Respectively, in Bown and Wiggins (2005) every agent in the population, even the performers, has a transition matrix just like listeners of Werner and Todd (1997). This transition table defines an agents compositional preferences if it is performing, whereas it is a tool to asses the heard signal that encodes agents 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 starting position in space. Briefly, any agent a_i, where $a_i \in A$ will have a position vector $coord_i =< x, y >$ defining its place with respect to the origin $< 0, 0 >$. Hence one round of musical interaction in Bown and Wiggins (2005) occurs as:

1. A listener is selected randomly upon the population .
2. Predefined number of agents are chosen from the population, which are the ones most proximate to the listener, as the set of performers. Listener attends to performers. Among this set of performers a semi-random selection is made to choose the actual performer.
3. Performer plays its song to the listener by using its transition table.
4. Listener assess the song using the local scoring policy of Werner and Todd (1997). However, the local familiarity scoring is delimited with a lower and upper boundary. This assumption depicts that agents can get bored with completely familiar signals and they can also dislike highly unfamiliar structures.
5. 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, which increases their chance to interact in upcoming rounds. 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 effect 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 emerge and evolve within a musical culture, where distinct subcultures have an influence on each other.

3.4 Naming Game: Emergence of a Shared Lexicon

For agents to cooperatively act on the environment, they must agree on a shared form-meaning mapping. That is to say, individuals of a population can only cooperate if they can conceive what a name refers to when it is uttered by a peer. To this end, naming game provides the essential model for a virtual society of agents to reach a consensus on a common vocabulary of nouns.

Traditional naming game models a population of agents $A = \{a_1, ..., a_n\}$ with size $N_A = n$. Each and every agent can equally perceive the environment, which contains a set of objects $O = \{o_1, ..., o_m\}$, where $N_O = m$. Agents can be described over their own private lexicons, which defines an inventory of a set of word-object pairings $I = \{\sigma_1, ...\sigma_m\}$. Each word-object pair σ is the set of words that an agent knows to name an object at a given time. For instance, for the object $o_k \in O$, an agent can have $\sigma_k = \{w_1, ..., w_i\}$. Every agent starts the game with an empty inventory, where for all $i \leq m$ the set of word-object pairs $\sigma_i = \emptyset$ in I.

Iteratively in each episode of communication (Baronchelli et al., 2005):

1. A random speaker a_x and hearer a_y is chosen for $x \neq y$ and $x, y \leq n$.
2. Both agents attend to an object o_i for $1 \leq i \leq m$.
3. Speaker transmits a name $w_j \in \sigma_i$ to the hearer from its inventory I_x. If $\sigma_i = \emptyset$ in I_x, it creates a random name.
4. Hearer processes the uttered name. If $w_j \in \sigma_i$ in I_y then communication is a success, else it is a failure.
5. Both parties make final modifications on their inventory. If communication was successful then both parties delete all the words from σ_i in their inventories except w_j, which is agreed on. Else, only hearer a_y adds w_j to σ_i in I_y.

Naming game dynamics can also be studied on a simplified minimal version, where there is only one object. In this case, a personal lexicon of an agent can be reduced to a set of words, such as $I = \{w_1, w_2, ..., w_q\}$ at a given time. Specifically, individuals start the game by inventing new words for the objects as their lexicons are empty at the outset. Therefore, number of words globally present in the society gradually increases up till a maxima, where-after convergence starts. Throughout this initial phase interactions are usually unsuccessful. From this point on total number of words start to decrease exposing an increase on the success rate till the end of the game. At the end, all agents converge on a shared inventory, so that $I_1 = I_2 = ... = I_n$. This lexical agreement occurs at ${N_A}^{1.5}$ (Baronchelli et al., 2008). At a given instance throughout the game agents can have more than one word in their lexicon to name an object, hence synonymy is unavoidable. After all, the winning word-meaning pair for each object might differ over distinct games, but eventually population decisively converges on a shared lexicon in each game.

CHAPTER 4

Methodology and Empirical Work

4.1 Model

4.1.1 Overall Description and Predictions

The model that is going to be presented in this section is a modified version of the familiarity game. Our primary aim is to model 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 pieces that will be pleasing when listened by peers. 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 just 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 an agent's 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. Moreover, in our model musical expectations are just restricted with pairs of notes (bi-grams) in a melodic line. That is to say, agents will be designed to have expectations only for the forthcoming note in sequence and nothing further than that. This restriction is imposed by agents' memory implementation, which defines how an agents musical expectation is computationally represented and it will be elaborated in Section 4.1.3.

Specifically, a population of agents with their own private expectations is going to be modeled. In each round of interaction, two agents will engage in a familiarity game, where one of them will perform a song and other will listen and assess the song. Songs will be composed in variable predefined lengths, which are greater than two notes, depending on the simulation setup. Interactions will be successful if performers can compose a song that will be familiar to the listeners to some degree. It is assumed that for performers to be appreciated by listeners their compositions must satisfy expectations of the audience. However, while interacting agents will not be able to directly retrieve each others expectations. Therefore, performers will compose songs according to their expectations and see whether their musical preferences suit with the agents that they are interacting with. In other words, while performing agents will compose by successively appending their most expected sequences in a row, so that the composition will represent their own expectations. To this end, convergence can occur if individuals of the population form a consensus on which antecedent-consequent pairs to be used in songs. At this state of agreement, it could also be claimed that agents share a unified compositional routine.

In accordance, 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. In a state of agreement 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. 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 structures/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. Nevertheless, even a possible social consensus on musical expectations would be a novel contribution to the literature.

4.1.2 Background Assumptions

Our model adopts all of the required assumptions for emergence and self-organization to rise. These are previously explained in Section 2.2. Herein, additional 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. Meaning that agents use different detector and effector channels to interact. Moreover, each agents musical expectancies are unique. Agents are not able to directly access to others expectancies. They can just make limited estimations about a peers 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. As a consequence, system control is decentralized, so that only local interactions can superimpose an aggregate system-wide behavior. Furthermore, assumption of identical agents is also required for attaining coherency and consistency among the members of the population.

- **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 overall 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, as 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. Furthermore, closed lexicon assumption will be promising to exclude all indispensable inquires about how agents can come to state of agreement where they will have a socially shared set of sounds that are eligible to be used in musical contexts. As it is presented in previous chapter, imitation game models will be relevant to provide essential elucidation on how shared sound systems emerge. Therefore, at this point our model will be simplified by assuming that consensus on a globally shared lexicon of sounds is taken granted at the outset.

- **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. Regarding this assumption, our model differentiates from models of Werner and Todd (1997) and Bown and Wiggins (2005). Within this context, cultural evolution of any kind will be left out of scope. Briefly, it is just aimed to investigate significant consequences of agent interactions (such as emergence of significant compositional preferences or compositional/reusable musical patterns) while population is reaching to an agreement of shared musical preferences in a closed community. The only drawback of closed community assumption is that it will practically not allow us to put robustness of any emergent behavior that may become observable on a test.

- **Free Interactions:** Each an every agent is equally probable to interact with each other. Any kind of topological distribution is omitted. Participants of an interaction is 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. By doing so, model stochasticity will be increased considerably, which will hinder a detailed investigation on the effect of local interactions on global behavior. Yet, our focus will be on macro-level instead of locality.

- **Consecutive Interactions:** Agents can not 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 (i.e. How can an agent compose a song if it is chosen to be a listener and a performer for two distinct interactions at the same time? Would it compose before it listens? Or would it listen the performer first and do the necessary memory updates before composing?). Such a contradiction 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 can not 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 expectancy is awaited. Boredom can be introduced to our model to observe continuous dynamic evolution of global musical preferences of a society. In fact, as long as boredom causes unsuccessful interactions, alteration of globally shared musical expectancy schemes will be everlasting. However, in our model dynamic restructuring of individual preferences is expected to be manifested throughout the learning phase before convergence.

4.1.3 Agents and Interactions

Modified familiarity game models 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 $N_L = 12$. 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:

Let $L = \{C, C\#, D, D\#, E, F, F\#, G, G\#, A, A\#, B\}$

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

where $0 \leq \alpha_{n_i,n_j} \leq 1$ 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 expectancy 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 agents musical expectancies 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.

Accordingly, agents of this model can be characterized with their table of musical expectancies as their aim in each interaction is to evaluate whether a musical piece is pleasant enough in terms of satisfying their expectancy on bi-gram level. At this point, 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.

In each round of interaction a modified familiarity game is played. The rules of this game 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.

   Agents complete following steps to compose:

   (a) An empty song template $S' = []$ is created.

   (b) 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 $S'[1] = n_k$.

   (c) 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 in (4.1). λ retrieves the most expected consequent for a given antecedent. If there are more than one successors with the same expectancy value then one of them is randomly chosen. In each iteration $N_{S'}$ increases by one and composition is completed when $N_{S'} = N_S$.

$$\lambda(S'[N_{S'}]) = max(\alpha_{S'[N_{S'}],n_i}) \quad for \quad \forall n_i \in L \tag{4.1}$$

3. Listener a_y evaluates S and conveys the success of the interaction to its counterpart.

   (a) To evaluate a song agents use a local scoring policy, which is defined in (4.2). Notably, success score is calculated in an additive fashion. In fact, the score could also be calculated globally for a song, if the song had hierarchically organized sections such as musical sentences or partitions. In that case, each musical section would have its own score and the global song score would be the sum of the scores of each section. However, within our model a musical signal represents just one complete melodic line, therefore its pleasantness for the listener could be fixed on summation of listeners expectancy values for each transition that they encountered sequentially.

$$score = \sum_{k=1}^{N_S - 1} \alpha_{S[i],S[i+1]} \tag{4.2}$$

(b) 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 (4.3). Overall score for the song is calculated by adding the listeners expectancy 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.

$$\epsilon(score) = \begin{cases} successful & if \quad \theta * (N_S - 1) \leq score \\ unsuccessful & if \quad score < \theta * (N_S - 1) \end{cases} \tag{4.3}$$

4. Participants a_x and a_y modify their transition tables T_x and T_y.

Listener a_y always increases expectancy values for all transitions that it heard without taking success into consideration. This is because, listeners familiarity for these antecedent-consequent pairs increase as they encounter them in a song. However, performer a_y modifies its table with regard to the listeners evaluation. Expectancy values for all transitions are incremented if 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 expectancy values after unsuccessful interactions help them to avoid using those note pairs in future. Amount of this modification is predefined by learning rate τ for both increment and decrement. More formally, agents use $\mu : T \mapsto T'$ function for table updates. Notably, it can be observed from (4.4) that there is no 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 $2 * \tau$. However, values for α_{C,n_i}, 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 expectancy values in a column will not always add up to 1.0. In this respect, an α_{n_i,n_j} will just give us an expectancy value 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.

$$\mu(T,S) = T' = \begin{cases} \alpha_{S[i],S[i+1]} = \alpha_{S[i],S[i+1]} + \tau \ for \ \forall i, \ 0 < i < N_S & if \ successful \\ & or \ listener \\ \alpha_{S[i],S[i+1]} = \alpha_{S[i],S[i+1]} - \tau \ for \ \forall i, \ 0 < i < N_S & otherwise \end{cases} \tag{4.4}$$

As an example, Figures 4.1, 4.2 and 4.3 present an illustrative modified familiarity game interaction to clarify composition, evaluation and memory update policies of interacting agents. Let's assume that $N_S = 10$, $\theta = 0.5$, $\tau = 0.05$ and L covers only one octave (specifically notes within the range of C to B) for this particular case.

Performing agent a_x will first create an empty song template S' with length 10 to compose as it is shown in Figure 4.1. Therefore, whole composition process will take $N_S - 1 = 9$ iterations. First note will be chosen randomly and it appears to be a C. To append the second note performing agent finds the most expected successor for C from its transition table T_x. In this example $T_x[C][G]$ has the highest value in row $T_x[C]$, for that reason G is appended to S' as the second note. Remarkably, it can be observed from Figure 4.1 that both A# and B can follow F in two different instances. This is because values for $T_x[F][A]$ and $T_x[F][B]$ are equal and maximal in $T_x[F]$ throughout this interaction. Accordingly, successive note for F is chosen randomly among the set of most expected successors. Table look up is carried on iteratively by the performer for $N_S - 2$ iterations, since S' is filled with notes. When composition is completed it is played to listener a_y for an evaluation as the musical signal S .

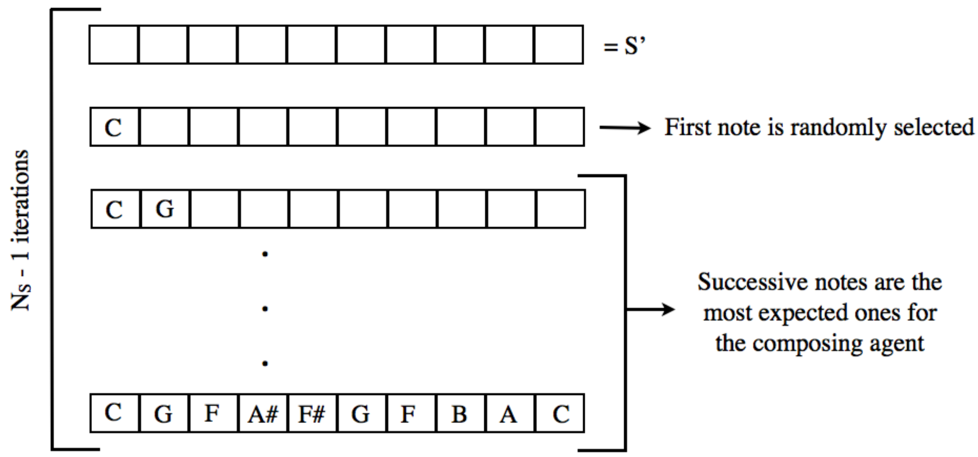

Figure 4.1: **(Model - Agents & Interactions)** An illustrative case of how agents compose a song.

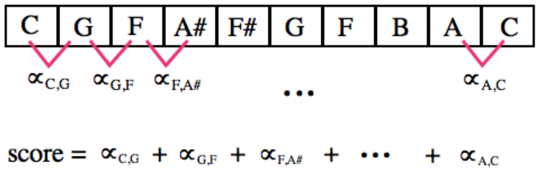

Figure 4.2: **(Model - Agents & Interactions)** An illustrative case of how agents evaluate a song.

To evaluate S listener a_y first calculates a song score as it is presented in Figure 4.2. Listener sums up expectation values for each note transition by using its transition table T_y. In our case, score would be $score = T_y[C][G] + T_y[G][F] + ... + T_y[A][C]$. In consequence, $score$ is used to evaluate pleasantness of S for a_y by checking whether it is above the success threshold θ. In our specific case, there are 9 transitions in the song and $score$ must be greater than 4.5 as $\theta = 0.5$. If $score > 4.5$ interaction would be successful and otherwise it will be unsuccessful. After a_y completes its evaluation it conveys the success to a_x to perform memory updates.

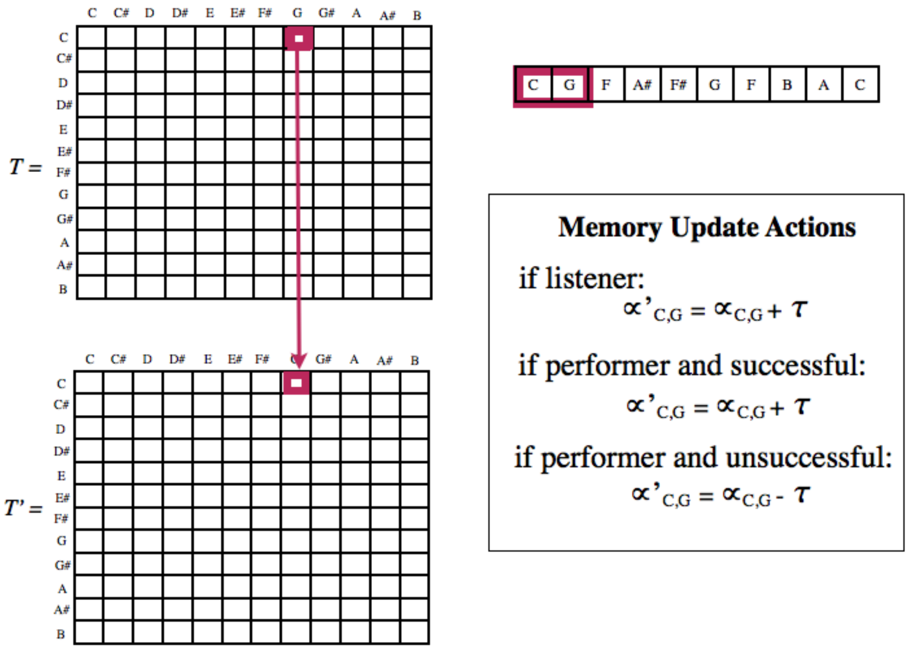

Figure 4.3: **(Model - Agents & Interactions)** An illustrative case of how agents preform memory updates on their table of musical expectations according to the success of the game.

Updates on the transition table of musical expectations are straightforward for both a_x and a_y. Specifically, interacting parties adjust their tables T_x and T_y for each note transition of S iteratively to attain T'_x and T'_y. As it is shown in Figure 4.3 listeners always add τ on the expectancy values of transitions in T_y. However, performers either decrement expectancy values for each transition with τ if the interaction was unsuccessful or increment them if the evaluation was successful.

4.1.4 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, modified familiarity game 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:

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

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

   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 (χ^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. S(t) can be calculated as:

   $$S(t) = \frac{Total\ Number\ of\ Successes\ at\ Round\ t}{t}$$

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. χ^2 is applied to all these three to observe the dynamic nature of consensus formation.

χ^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 χ^2 significance test can be extended to n-grams of any length. However, in our experiment we apply χ^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.

4.1.5 Implementation

Both simulation and analyzer software are implemented with *Python 2.7*, particularly for data collection and data analysis. Specifically, transition tables are implemented with scientific arrays of *NumPy* module from *SciPy* library. All other data structures are from standard *Python 2.7* libraries. Figures, which are going to be presented in Section 4.2 are plotted with standardized *matplotlib* library. Experiments and analyses are conducted on a dual-core Linux machine.

4.2 Empirical Results

4.2.1 Game 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. As it is mentioned in the previous section, success rate is the number of average successful interactions throughout the simulation. 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.

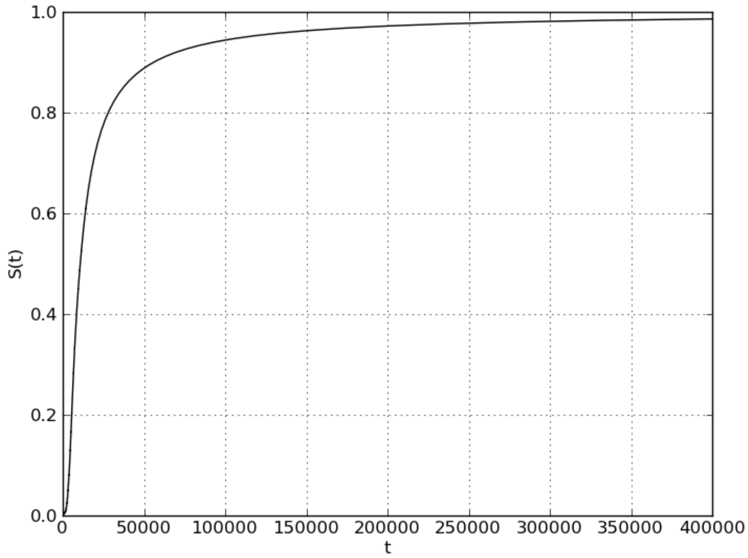

Figure 4.4: **(FG - Game Dynamics)** Success rate of interactions and convergence on a shared table of expectations. Simulation is run for $N_A = 50$, $N_L = 24$, $N_S = 32$, $\tau = 0.05$ and $\theta = 0.5$. Success rates are averaged for 10 runs.

Figure 4.4 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 S(t) increases rapidly just 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 S(t) decreases and S(t) curve flattens since learning is brought to a completion. This distinctive success rate curve denotes to 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.

In Figure 4.5, effect of population size on time of convergence can be observed, particularly

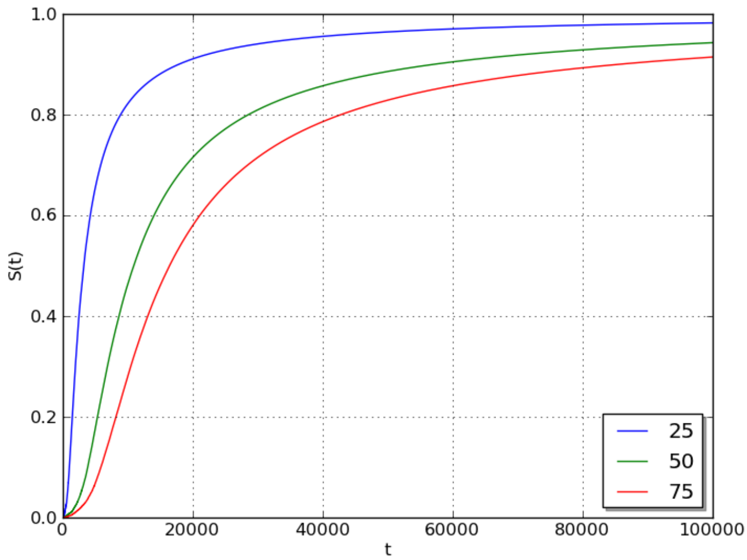

Figure 4.5: **(FG - Game Dynamics)** Effect of population size N_A on convergence. Simulation is run for $N_L = 24$, $N_S = 32$, $\tau = 0.05$ and $\theta = 0.5$. Convergence trends for $N_A = 25$, 50 and 75 are presented. Success rates are averaged for 10 runs. Time of convergence increases as population size grows.

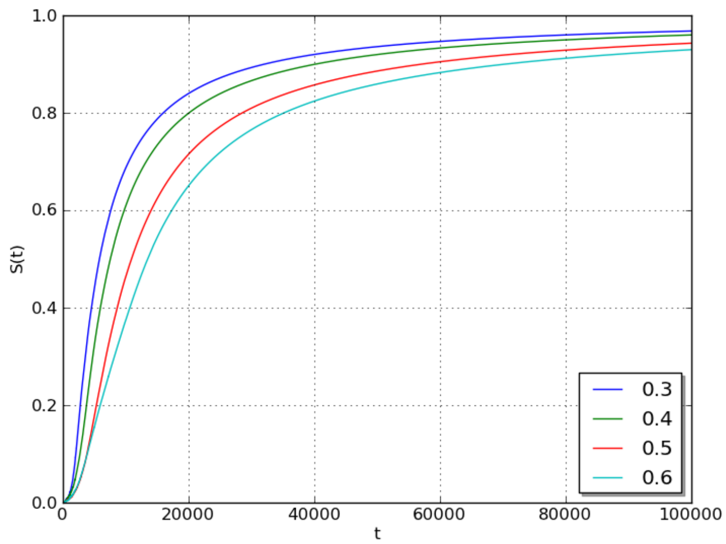

Figure 4.6: **(FG - Game Dynamics)** Effect of success threshold θ on convergence. Simulation is run for $N_A = 50$, $N_L = 24$, $N_S = 32$ and $\tau = 0.05$. Convergence trends for $\theta = 0.3$, 0.4, 0.5 and 0.6 are presented. Success rates are averaged for 10 runs. Time of convergence increases with increasing θ.

for $N_A = 25$, 50 and 75. For larger populations agreement comes late. Therefore, increasing the population size will also delay convergence. This is because, dominating bi-gram expectancies 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 are required for the members of larger communities to interchange their preferences on winning note pairs.

Agent's 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 θ. 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 4.6 varying S(t) curves are presented for increasing success thresholds 0.3, 0.4, 0.5, 0.6. 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.

In Figure 4.7 it can be seen that time of convergence depends on τ. Interestingly, system's learning is optimal at around $\tau = 0.5$. Within the range $0.2 < \tau < 0.7$ systems 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 can not 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 bring success drastically alters after memory updates throughout the learning phase.

In Figure 4.8 alteration of S(t) is presented with respect to signal lengths $N_S = 16$, 24, 32 and 64. Notably, change in N_S does not affect convergence upto a hard boundary. From the figure, it can be observed that S(t) curves overlap for $N_S = 12$, 24 and 32. However, when N_S 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, its more likely

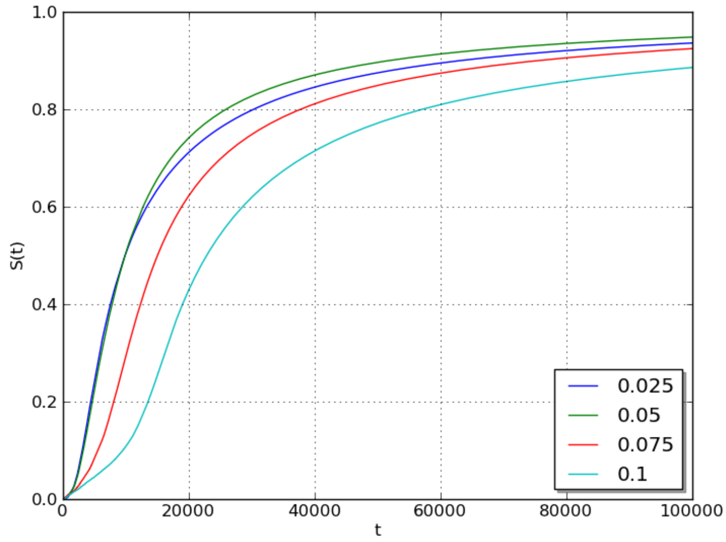

Figure 4.7: **(FG - Game Dynamics)** Effect of success learning rate τ on convergence. Simulation is run for $N_A = 50$, $N_L = 24$, $N_S = 32$ and $\theta = 0.05$. Convergence trends for $\tau = 0.025$, 0.05, 0.075 and 0.1 are presented. Success rates are averaged for 10 runs. Time of convergence increases with increasing τ, when τ is greater than 0.5.

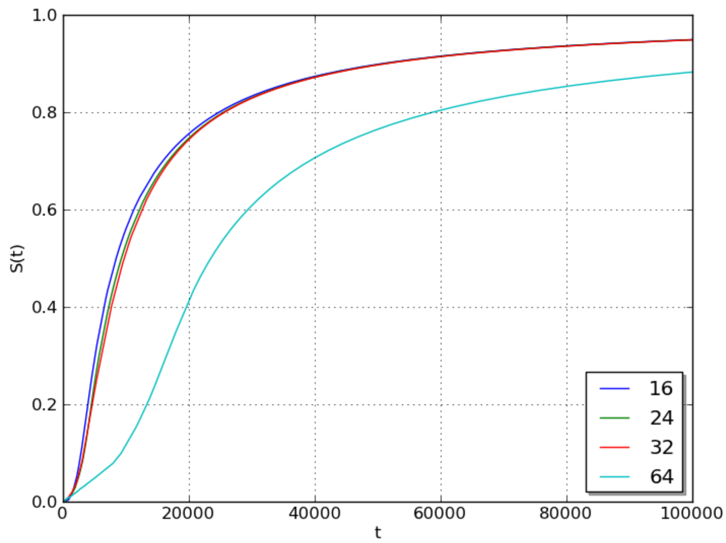

Figure 4.8: **(FG - Game Dynamics)** Effect of signal length N_S on convergence. Simulation is run for $N_A = 50$, $N_L = 24$, $\tau = 0.05$ and $\theta = 0.5$. Convergence trends for $N_S = 16$, 24, 32 and 64 are presented. Success rates are averaged for 10 runs. Agreement trends are similar for short signals with $N_S = 16$, 24, 32. However, success rate drastically drops exposing a delay in agreement when relatively longer signals such as $N_S = 64$ are exchanged.

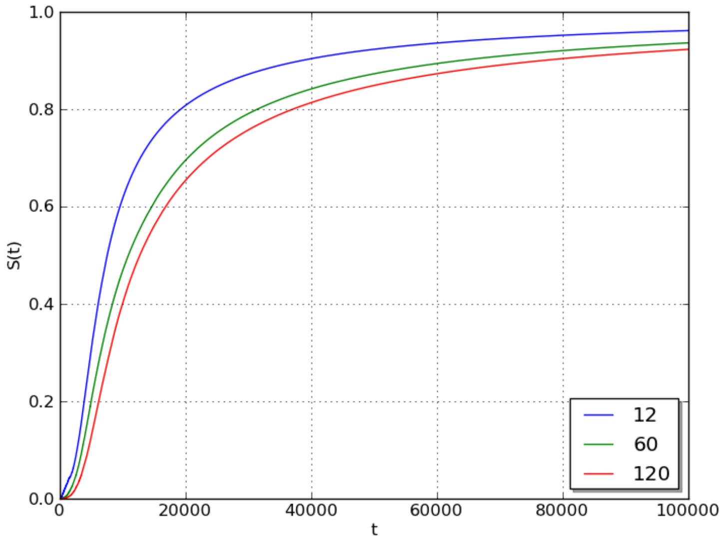

Figure 4.9: **(FG - Game Dynamics)** Effect of lexicon size N_L on convergence. Simulation is run for $N_A = 50$, $N_S = 32$, $\tau = 0.05$ and $\theta = 0.5$. Convergence trends for $N_L = 12$, 60 and 120 are presented. Success rates are averaged for 10 runs. Time of convergence proportionally increases with increasing size of the lexicon.

for an antecedent-consequent pair to be involved in a composition more than once for larger N_S. Hence, when signal size increases expectancy values for winning bi-grams which 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 4.8 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.

Finally, Figure 4.9 presents how lexicon size N_L effects convergence. N_L 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. Accordingly, as long as the state space grows the time required for the population to form an agreement in one of the attractor states increase. Consequently, it can be observed from Figure 4.9 that an increase in lexicon size lags convergence.

4.2.2 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 4.10 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 4.1.

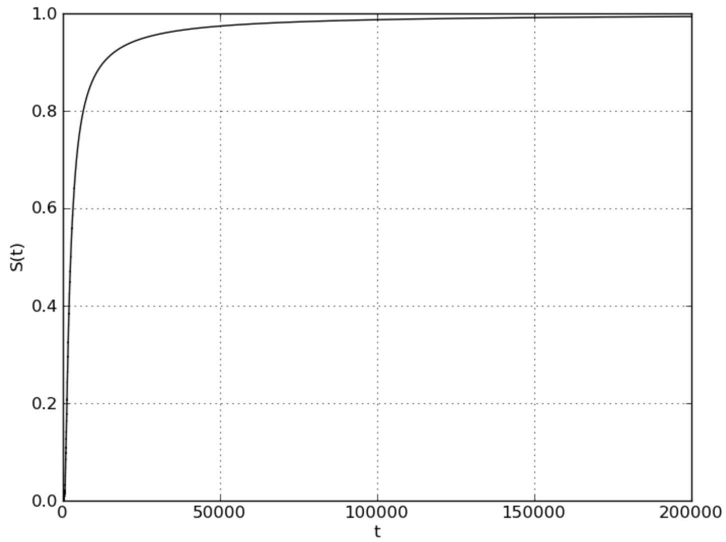

Figure 4.10: **(FG - Self-Organization & Emergence)** S(t) curve for N_A = 25, N_L = 12, N_S = 12, τ = 0.05, θ = 0.5.

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 χ^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 4.1 and 4.2-(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 χ^2 test. This dissimilarity arise from self-organizing nature of the system.

To be clear, dominating bi-grams are not deterministicly 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

Table 4.1: **(FG - Self-Organization & Emergence)** Global Transition Table after 200,000 interactions for $N_A = 25$, $N_L = 12$, $N_S = 12$, $\tau = 0.05$, $\theta = 0.5$. 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

carried on. For instance, Tables 4.1 and 4.2-(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 χ^2 test to $C_{0.0}$ and $C_{0.5}$. For $C_{0.0}$ (notably all signals of the game included in this corpus), χ^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. Accordingly, 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.

As a consequence, all winning antecedent-consequent note pairs become bi-gram building blocks of a musical signal. In Tables 4.2-(b) and 4.2-(c) it can be concluded that some fragments of signal of length three and four can be significantly observed by applying χ^2 test on $C_{0.9}$. There are 74 observable tri-gram and 7 4-grams that are significantly used in composition 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. Arguably, this trend could be grounded on the learning trend of the population. Moreover, all these n-gram note sequences are composed of sequential arrangement of some of the winning bi-grams. 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.

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 just have a favor to use winning bi-grams in their compositions. However, 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 just append the winning bi-grams sequentially to compose a song.

Table 4.2: **(FG - Self-Organization & Emergence)** (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-G-C	D#-E-F-B
C#-C#	C#-C#-C#	F-G-F	E-E-F-B
D-D	D-D-D	F-B-C	F-E-F-B
D#-C	D#-C-C	F-B-E	F-G-F-B
D#-E	D#-E-C	F-B-G	A-E-F-B
E-A	D#-E-D#	F#-F#-F#	B-G-F-B
E-C	D#-E-E	G-C-C	B-B-B-B
E-D#	D#-E-F	G-F-C	
E-E	D#-E-A#	G-F-D	
E-F	D#-A-C	G-F-E	
E-A#	D#-A-E	G-F-F#	
F-C	D#-A-G#	G-F-G	
F-D	D#-A-A#	G-F-B	
F-E	E-C-C	G#-C-C	
F-F#	E-D#-C	G#-F#-F#	
F-G	E-D#-E	G#-A#-C	
F-B	E-D#-A	G#-A#-A#	
F#-F#	E-E-C	A-C-C	
G-C	E-E-D#	A-E-C	
G-F	E-E-E	A-E-D#	
G#-A#	E-E-F	A-E-E	
G#-C	E-E-A#	A-E-A#	
G#-F#	E-F-C	A-G#-C	
A-C	E-F-D	A-E-F	
A-E	E-F-E	A-G#-F#	
A-G#	E-F-F#	A-G#-A#	
A-A#	E-F-G	A-A#-C	
A#-C	E-F-B	A-A#-A#	
A#-A#	E-A#-C	A#-C-C	
B-C	E-A#-A#	A#-A#-C	
B-G	F-C-C	A#-A#-A#	
B-B	F-D-D	B-C-C	
	F-E-C	B-E-A#	
	F-E-D#	B-G-C	
	F-E-E	B-G-F	
	F-E-F	B-B-B	
	F-E-A#		
	F-F#-F#		

CHAPTER 5

Conclusion and Discussions

In this thesis, we have presented a computational model of a multi-agent musical society, where agents played a modified musical familiarity game. The simulation 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. It can be concluded that performance of the model is inversely proportional with the population size, success threshold and lexicon size. Performance is also dependent on agent's learning rate. Significantly small and large learning rates have a negative impact on time of convergence. Therefore, learning rate should be optimal for faster convergence. Population can only form a consensus when the signal length is chosen to be smaller than a specific hard boundary, otherwise agents cannot agree on shared conventions.

Accordingly, our method of modeling has proven to be successful to investigate how musical structures change in time within a culture just 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 computational evolutionary musicology, thus its performance and robustness could not be compared and contrasted with its alikes. 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. In fact, the assumption that music does not carry any semantic content distinguishes our model from models of evolutionary linguistics. Within this context, emergence of socially shared musical conventions such as harmonic and melodic progressions and rhythmic movements can only 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 might most probably be different from this computational model. This is because, 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 applicable for real world musical system.

Briefly, our formalization provides a broad framework, which can be extended in various ways. Some of these possible proposals for future research can be listed as:

- Agents can be modified to construct a population which can emerge hierarchical and categorical structures. Our agents are just 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. For instance, in western tonality notes are positioned according to their distances to a central tone. Therefore, compositional grouping is accomplished over these relationships. In correlation, our agents are not restricted with such rules while they are composing. For instance, agents could be modified in a way that they could track how often several tones come together to find out the relationship between them.

- Agents could get bored with extremely familiar signals. By doing so, an absolute convergence will not be possible, rather the system will continuously evolve through time. Effects of various different agent characteristics on evolution of songs could become observable in such a setup.

- 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 interesting impacts on convergence trends.

- The number of winning bi-grams, which are agreed on, might be bound to simulation parameters. It might be intriguing to examine whether the set of winning antecedent-consequent pairs dependent on population size, learning rate, success threshold, lexicon size or signal length.

- As it is mentioned earlier, our model is suitable for studying other sequencing tasks in a broader sense. For instance, agents and interactions presented in the previous chapter could be modified to tackle problems from the domain of evolutionary linguistics. Emergence and evolution of phonemes is such a sequencing task, which is eminent to study emergence and evolution of spoken linguistic communication. One of the proposed methods to capture social dynamics of emergence of phonemes is to model a society in which interacting agents compose and assess signals that are formed from sequencing of phonemes. Such a model would be a mixture of the naming game and modified familiarity game. An environment which consists objects should be modeled and interacting agents should try to agree on a set of sequential arrangement of phonemes, where each sequence will denote an object. In fact, agents will have a lexicon of phonemes, which will be subjected to alterations depending on the success of the interactions that an agent gets involved to. With such a model, it can be observed whether the modeled society emerges distinctive patterns of phoneme sequencing on the way to attain an agreement.

References

Baronchelli, A., Dall'Asta, L., Barrat, A., & Loreto, V. (2005). Strategies for Fast Convergence in Semiotic Dynamics. *Arxiv Preprint Physics/0511201*.

Baronchelli, A., Dall'Asta, L., Barrat, A., & Loreto, V. (2006). Topology-Induced Coarsening in Language Games. *Physical Review*, *73*(1), 2-5.

Baronchelli, A., Felici, M., Loreto, V., Caglioti, E., & Steels, L. (2006). Sharp Transition Towards Shared Vocabularies in Multi-Agent Systems. *Journal of Statistical Mechanics: Theory and Experiment*.

Baronchelli, A., Loreto, V., & Steels, L. (2008). In-depth Analysis of the Naming Game Dynamics: The Homogeneous Mixing Case. *International Journal of Modern Physics*, *19*(5), 785-812.

Batali, J. (1998). Computational Simulations of the Emergence of Grammar. In *Approaches to the Evolution of Language* (p. 405-426). Cambridge University Press.

Beckner, C., Ellis, N. C., Blythe, R., Holland, J., Bybee, J., Ke, J., et al. (2009). Language Is a Complex Adaptive System: Position Paper. *Language Learning*, *59*, 1-26.

Biletzki, A., & Matar, A. (2011). Ludwig Wittgenstein. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Summer 2011 ed.). `http://plato.stanford.edu/archives/sum2011/entries/wittgenstein/`. (Online; accessed 11-July-2011)

Boersma, P. (1993). Accurate Short-Term Analysis of the Fundamental Frequency and the Harmonics-to-Noise Ratio of a Sampled Sound. In *University of AmsterdamIFA Proceedings* (Vol. 17, p. 97-110).

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* (p. 271-301). MIT Press.

Chalmers, D. J. (2006). Strong and Weak Emergence. In P. Clayton & P. Davies (Eds.), *The Re-Emergence of Emergence.* Oxford University Press.

Clayton, P. (2006). Conceptual Foundations of Emergence Theory. In P. Clayton & P. Davies (Eds.), *The Re-Emergence of Emergence.* Oxford University 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* (p. 85-93). IEEE.

Dall'Asta, L., Baronchelli, A., Barrat, A., & Loreto, V. (2006). Nonequilibrium Dynamics of Language Games on Complex Networks. *Physical Review*, *74*(3).

De Beule, J. (2008). The Emergence of Compositionality, Hierarchy and Recursion in Peer-to-Peer Interactions. In *The Evolution of Language: Proceedings of the 7th International Conference* (p. 75-82).

De Boer, B. (1997). Self-Organisation in Vowel Systems Through Imitation. In P. Husbands & I. Harvey (Eds.), *Proceedings of the fourth European Conference on Artificial Life.* Cambridge Ma.

De Vylder, B., & Tuyls, K. (2006). How to Reach Linguistic Consensus: A Proof of Convergence for the Naming Game. *Journal of Theoretical Biology*, *242*(4), 818-831.

De Wolf, T. (2007). *Analysing and Engineering Self-Organising Applications.* Unpublished doctoral dissertation, Katholieke Universiteit Leuven.

De Wolf, T., & Holvoet, T. (2005). Emergence versus Self-Organisation: Different Concepts But Promising When Combined. In *Engineering Self-Organising Systems: Methodologies and Applications* (Vol. 3464, p. 1-15). Springer-Verlag Berlin.

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, p. 1111-1118). IEEE.

Halley, J. D., & Winkler, D. A. (2008). Classification of Emergence and Its Relation to Self-organization. *Complexity*, *13*(5), 10-15.

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.

Jackendoff, R. (2002). *Foundations of Language: Brain, Meaning, Grammar, Evolution*. Oxford: Oxford University Press.

Kirby, S. (2002). Natural Language From Artificial Life. *Artificial Life*, *8*(2), 185-215.

Kirby, S., & Hurford, J. R. (2002). The Emergence of Linguistic Structure. *Simulating the Evolution of Language*, 121-147.

Lansing, J. (2003). Complex Adaptive Systems. *Annual Review of Anthropology*, *32*, 183-204.

Lenaerts, T., Jansen, B., Tuyls, K., & De Vylder, B. (2005). The Evolutionary Language Game: An Orthogonal Approach. *Journal of Theoretical Biology*, *235*(4), 566-582.

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., p. 231-260). 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.

Livingstone, S. R., & Thompson, W. F. (2009). The Emergence of Music from the Theory of Mind. *Musicae Scientiae*(Sp. Iss. SI), 83-115.

Martins, J. M., & Miranda, E. R. (2006). A Connectionist Architecture for the Evolution of Rhythms. In *Proceedings of Applications of Evolutionary Computing* (Vol. 3907, p. 696-706). Springer Press.

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

Minsky, M. (2006). *The Emotion Machine*. Simon & Schuster.

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, p. 233-246). 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., p. 218-249). Springer Press.

Mithen, S. (2006). *The Singing Neanderthals: The Origins of Music, Language, Mind and Body*. Harvard University 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.

Steels, L. (1996a). Emergent Adaptive Lexicons. *From Animals to Animats*, *4*, 562-567.

Steels, L. (1996b). Perceptually Grounded Meaning Creation. In *Proceedings of the Second International Conference on Multi-Agent Systems* (p. 338-344).

Steels, L. (2000). Language as a Complex Adaptive System. *Lecture Notes in Computer Science. Parallel Problem Solving from Nature*, PPSN - VI.

Steels, L. (2003). Evolving Grounded Communication for Robots. *Trends in Cognitive Sciences*, *7*(7), 308-312.

Steels, L., & Kaplan, F. (1999). Collective Learning and Semiotic Dynamics. *Lecture Notes in Computer Science*, 679-688.

Titchener, J. M., & Broyles, M. E. (1973). Meyer, Meaning, and Music. *The Journal of Aesthetics and Art Criticism*, *32*(1), 17-25.

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.

Wiggins, G. A. (1998). Music, Syntax, and the Meaning of "Meaning". In *Proceedings of the First Symposium on Music and Computers* (p. 18-23). Corfu, Greece: Ionian University.

Wittgenstein, L. (1922). *Tractatus Logico-Philosophicus*. London: Routledge.

Wittgenstein, L. (1953). *Philosophical Investigations*. Oxford: Blackwell.