Secret Agents: Synthetic Approaches to Language Evolution

Emily Thomforde

January 16, 2004

1 Language Evolution

1.1 Introduction

Extensive work in the field of synthetic ethology demonstrates that multi-agent simulation is a useful tool for exploring a multitude of population-based phenomena, including economics, epidemiology, and language evolution. From a structured canon of local interactions and low-level rules, complex behaviors can emerge in a wide range of situations. This notion of 'emergence' is central to evolutionary studies, biological or otherwise. This paper addresses the current models of language evolution as put forward by contemporary researchers and attempts to codify the rules of language modeling. It seeks to modify the work done by Bart de Boer [7, 6, 8, 10] in self-organizing vowel systems by modifying certain principles of his simulation. With a few algorithmic changes, I hope to isolate the clustering mechanism at the heart of the simulation, so that it may be used in modeling self-organizing systems other than vowel inventories.

Chapter 1 provides an introduction to the study of language evolution, then outlines the current research in the field and synthesizes the central principles of synthetic ethology as demonstrated by these systems. Chapter 2 describes the implementational details of reproducing de Boer's simulation and discusses the results. Chapter 3 explains the simplified model in terms of algorithmic changes and organizational consequences. Chapter 4 discusses what can be learned from these experiments.

1.2 What is Language Evolution?

The study of Language Evolution deals with explaining how human language came about from non-language [25]. It is an problem central to the understanding of humanity and the human brain. In many ways, it is very similar to the problem of biological evolution. Much time and energy has been devoted to investigating the origins of humanity, and this enterprise is integral to our understanding of our species. The same is true of language; it is one of the many aspects of humanity that separates us from the apes, and the one people most often point to in order to illustrate our civility. The first question we must ask is “Where does language come from?” and that is easily answered by Chomsky,“From the Language Organ.” But then we must ask “Where did the Language Organ come from?” and that is a harder question to answer.

We know from studying the brain that there is no single physical structure devoted to language [25], so it is not an easy phenomenon to explain. It seems that most structures involved in the production and processing of language are used for other tasks as well, suggesting that language was not the goal of the human brain, but rather a by-product of its evolution.

Robert Worden sets out four criteria for a theory of language evolution [29]:

> (a) Evolution. It should be consistent with he constraints of the theory of evolution given the fossil evidence of human evolution and the selection pressures on our ancestors.
>
> (b) Language use. It should agree about what we know about the uses of language- including the range of meanings language can express, the speed and robustness with which we use it, and the facts of language learning, structure, and diversity.
>
> (c) Neurophysiology and anatomy. It should agree with what we know from PET scans, lesion data, and other sources about the locations of language processing in the brain.
>
> (d) Computation. It should give a working model of how language computations are done in the brain- how we represent language meanings, how we convert word sounds into those meanings when understanding language, how we convert in the reverse direction to generate language, and how we learn a language.

Ideally, researchers strive for a comprehensive theory that encompasses all of Worden’s elements. This list is daunting, however, because there are still considerable gaps in our knowledge of each.

1.3 How is Language Evolution Studied?

The study of Language Evolution is markedly different from that of biological evolution. Archaeologists have the distinct advantage of being able to mine data from rock and rubble. Because speech does not make for good fossils, researchers in Language Evolution must devise craftier methods of acquiring evidence. Several approaches have been attempted.

First, one can study linguistic and extra-linguistic communication in non-human primates. Because we share many of the same biological and neurological structures, there are certain similarities in our capacities for language [27]. However, several divergences occur between the physiology of humans and other primates that offer us insight as to the differences between human biological and social imperatives and those of our closest evolutionary relative. Using these divergences as a starting point, we can try to figure out which structures are linked to language, then attempt to trace the route that biological evolution may have taken to get this result.

Second, we can examine human language behavior. Much can be understood by studying the way in which children acquire language. It can give us insight into the basic processes at work in a simplified arena and help us to distinguish what is inherent from what is learned [27]. Additionally, studying the effects of different types of aphasia can provide details about the structures of the brain used in language production and perception.

But language is not solely a property of an individual. It is instead a complex and dynamic system owned by a population [25]. When socialization comes into the picture, we must set aside Chomsky's paradigm of the ideal speaker [4] and delve into the third option: the study of population. When we start to explore the effects of population dynamics and the propagation of language through a group, what we encounter is a problem of modeling [18]. This can be approximated with mathematical equations, and has been done so since the 1960s. For example, Liljencrants and Lindblom optimized the acoustic distribution of a vowel inventory by optimizing an abstract energy function [16]. While math is a useful tool for making generalizations about a pattern, it lacks the power to robustly model population-wide behaviors. Several variables must be fixed and there is no room for randomness. Mathematical models are powerless to explain the actual cognitive processes at work in any case because they make large assumptions about the uniformity of individuals in a population [25].

In order to more accurately model the interplay among language users in a dynamic population, synthetic implementations may be employed. It is here that the computer simulation becomes invaluable. Because it is generally accepted that the structures of language constitute an emergent property of human populations [7], it is permissible to model small slices of those populations with groups of locally-interacting artificial language agents. These agents are programmed with the human brain in mind, but in implementation represent highly simplified, highly specialized heuristics that attempt to model human cognition. The conclusions drawn from the simulations can then give us limited, but informative, insight into the actual workings of language evolution. Similar methodology is recycled in studies of epidemiology and economics, where population dynamics are again at the center of the theory [18].

The increased processing power and memory capacity of contemporary computing allows us to fruitfully investigate the origins of language by means of computer simulation. The motivation for this manner of investigation, as explained by de Boer [8] is that it is insufficient to describe a language capacity as 'innate.' We must then explain how that innateness evolved, and the only way we may do so and also justify the explanatory power of our results is to model the evolution of language in a population of simulated, autonomous language agents [25].

1.4 Agent-Based Simulation

Agent-based simulation is a computational approach to the study of populations. 'Agents' are small programs that can interact with each other and their environment according to a set of rules. The purpose of agent-based simulation is to study complex behaviors that emerge from simple rules and local interactions. A particular subfield has come to be recognized with respect to language modeling. The majority of the work has been done in language evolution [30, 22, 5, 13] but some simulations in language change have been attempted [12, 20]. The difference between the two fields is that language evolution tries to build new structures from scratch, where language change starts from some point after language has emerged.

Curran and O'Riordan outline three theoretical approaches to the synthetic modeling of language evolution: Genetic Evolution, Adaptation and Self-organization, and Genetic Assimilation [5]. 'Genetic Evolution' is the most Chomskian approach, assuming that organisms inherently possess the structures necessary for complex systems to arise, a language acquisition device (LAD). Mutation introduces enough variation into the system for properties to evolve over generations.

The second approach is labeled 'Adaptation and Self-organization.' It is based on the assumption that interactions within the population are sufficient to evolve a complex system such as language, and that no LAD exists. The language community is motivated by positive feedback loops, in that a developed language leads to improved communication, which itself reinforces the language. Novelty is introduced through communication errors during language contact. There is no genetic transmission involved; each successive generation learns the current state of the language from scratch, introducing innovation into the language as it does so.

Finally, 'Genetic Assimilation' combines the first two approaches. Under this theory, organisms in a population are motivated by positive feedback loops, success in one iteration increasing the chance of future success, but do not start from scratch each generation. Throughout their lifetimes, they learn through adaptation and then pass their learned traits genetically to their offspring. The strategies most conducive to survival are propagated through the population, effectively creating a race of organisms with an increasing genetic predisposition to language. Thus the LAD is affected both by adaptation on the level of the individual and evolution over generations.

Regardless of the theoretical approach, certain properties of agent-based systems tend to remain constant. All simulations involve agents, a set of rules for language contact, and a learning algorithm [23]. Most are enacted in a finite spatial environment. Most introduce some element of stochasticity, or random error, into the language community [24]. Some specify the replacement of agents over time [8]. These similarities represent the beginning of paradigm. As years pass and more synthetic models are built, the field to converges on a set of rules that future simulations must follow.

Once we can model real-world behavior, we have an arena for studying the response of a language-evolving system to different stimuli. For example, in a working system of twenty agents that play imitation games to develop a lexicon, what happens if there are two hundred agents, or if the possible word space is cut in half? In this way, we are not claiming that language evolution occurred exactly as in our simulation, merely that a system that behaves similarly may give us insight into certain principles of our own behavior [25]. With this goal in mind, there is the imperative to remain faithful to the observed properties of human communication and Worden's criteria for an informative theory.

1.5 Issues in Simulation

1.5.1 Emergence and Self-organization

The evolution of human language is what we may term an emergent process [25]. Though definitions of emergence may be incomplete, it can be said that it describes large-scale patterns manifested as a result of small-scale processes. In grouping language evolution in the class of emergent patterns, we can apply the principles already developed for emergence. First, there is a marked lack of central control. Interactions between individuals are local. Agents do not have access to the inner workings of each other. No agent may directly change the state of another. Agents interact according to a series of local rules [18]. Thus, any population-wide patterns that emerge are the result of processes taking place on the level of the individual [25].

Browman and Goldstein define 'self-organization' as "the spontaneous emergence of order" [2]. The term is often used to describe systems such as market prices, population flux, and phoneme inventories, where multiple internal processes are in effect that determine the form of the system.

1.5.2 Synthetic Ethology

Ethology is the study of the behavior of organism populations in a specific environment. Synthetic ethology brings this study into the domain of computation, where we are able to finely control conditions for experimentation [18]. Synthetic ethology is a powerful tool because it exploits our knowledge of population dynamics to grant us insight into their effects. de Boer says that once we get a simulation that accurately predicts things we can attest in the real world, then that simulation becomes an arena for studying aspects of the phenomenon and its response to different changes in circumstance, environment, etc. [6]. Thus, we have created 'in silico,' in the words of Epstein and Axtell, a method of studying the evolution of language [11]. In theory, any phenomenon that is a property of a group can be modeled and studied with the use of agent-based simulation. Computer simulation gives us insight into the dynamics of interacting language agents. It helps to turn a thought experiment into a physical, manipulable system.

1.5.3 Sufficient v. Necessary

Agent-based simulations seek to illustrate sufficient conditions for the observed behavior of a population. However, they do not claim to prove that these conditions constitute the exact situation under which the behavior developed. This uncertainty is at the heart of the difference between 'sufficient' and 'necessary conditions. Consider a simulation that could be said to successfully evolve language. This shows that the circumstances under which it does so are 'sufficient' for successful language evolution. This is not to say they are 'necessary.' A completely different set of conditions may exist under which language may also successfully emerge. If we are able to determine a large enough set of sufficient conditions, we will be able to study them in combination, find patterns, highlight the central principles, and gain a greater understanding of what conditions may have been necessary to evolve language in humans. More importantly, it provides important evidence for ruling out certain scenarios that could not possibly have resulted in the relevant behavior.

1.6 Approaches to Language Simulation

There have been several approaches to the computational modeling of language evolution over the past decade. From these, a specific canon of principles have emerged, under which subsequent researchers must operate. They include a number of restrictions on the implementation of agents, the rules for language interaction, and sources of variation within the simulation. To codify these principles, five very different approaches to language modeling are considered with respect to Curran and O'Riordan's theories of language evolution.

One of the most extensive and renowned is the 'Talking Heads' simulation of Luc Steels et al. (1998) [24]. A population of robotic agents is shown to evolve a lexicon through iterated naming games with other agents. Learning is motivated by positive feedback loops incurred during communication. New agents are given no genetic advantage and must learn the language just as the previous generation did. There are multiple sources of stochasticity in language contact, stemming from mishearing, imperfect memory access, and a physical component introducing inaccuracy in perception. Steels' agents possess inherent structures that allow them to store form-meaning pairs and make decisions about their utility.

A slightly modified system is employed by de Boer in his vowel organization simulation [6, 7, 8]. Like Steels', it is based primarily in the Adaptation and Self-organization approach. Agents interact and form a vowel inventory without any sort of central control. Motivation is in the form of a positive feedback loop based on the success of iterative imitation games between agents. Each successive generation of agents must start from scratch, not inheriting any sort of genetic predisposition from parents. However, agents possess a limited LAD in the form of inherent structures for analyzing and storing vowels. This is genetic in the sense that these structure are given to each agent at birth, but does not follow the Chomskian model because these structures are not subject to crossover and mutation over generations.

Browman and Goldstein (2000) attempt to model the self-organization of phonetic gestures in an agent-based context [2]. The simulation contains only two agents, endowed with the ability produce, hear, analyze, and remember sound gestures. They do not exist in the context of a spatial environment. There is no replacement over time. Agents have a limited ability to recover gestures from sound signals and all communication takes place in a noisy channel, introducing stochasticity into the simulation. Language interaction is based on imitating signals; both agents take turns speaking and imitating. Browman and Goldstein's simulation takes elements from both the Adaptation and Self-organization and Genetic Evolution approaches, but does not achieve a synthesis as complete as Genetic Assimilation. There are only two agents and no replacement, so no genetic change is taking place. However, there are very specific structures in place for language processing. Thus, this system does not fit well into any of Curran and O'Riordan's theories of language evolution.

Walshe (2001) takes a different approach to language evolution, attempting to model signalling between socially related agents [28]. His 'Linguanas' exist in parent-child relationships where communication is based on the sending of instructions from one to the other. A lexicon is developed based on the success of each signal in helping to increase the status of the agent pairs. Like Steels', Walshe's simulation best fits into the Adaptation and Self-Organization school. Language interaction is motivated by a well-structured reward system. While control is exercised by the environment and adult agents, there is no centralized control of the language. Linguanas possess standard inherent structure necessary for processing linguistic data, but do not pass them on to offspring.

In contrast, Cangelosi and Parisi (1996) [1] take the 'Genetic Assimilation' approach in their simulations. Their system treats language as a series of binary strings that describe the properties of 'mushrooms' that agents seek to gather. Interactions between agents aim to disseminate information about edible and poisonous varieties, and the success of these dictates the composition of a population-wide lexicon. Agents learn through positive feedback loops, but then pass on these learned structures, the trained neural networks, to their children. This effectively models the increased genetic predisposition to language over time. Also different from the other three models is the treatment of the LAD. In using a neural network to represent agents' brains, Cangelosi and Parisi are giving them an acquisition device that is not as specifically tuned to language as, for example, de Boer's phoneme processing system.

All of these simulations attempt to show that language can evolve from a population motivated only by the need for communication. However, as much as they try to abstract away from a LAD, they endow their agents with certain structures designed for language processing. Arguably, these are equivalent to those of the human brain and are therefore allowable in a decentralized model of language evolution [25], but they do not seem consistent with the theory of Adaptation and Self-organization, using elements of Chomskian language theory. In his description of synthetic models, Steels allows general structures to be in place, but not tuned to language capabilities [25]. He also characterizes the Adaptation and Self-organization theory as based on “a system of elements with only local interactions but strong positive feedback loops” [23]. However, both his own experiment and de Boer’s allow for random interaction between agents, as they are not based in a spatial environment. In this respect, neither model conforms fully to the precepts of Adaptation and Self-organization. Even though they give agents an inherent disposition to language, without generational inheritance of learned properties, they likewise do not conform to Genetic Assimilation. Walshe, as well as Browman and Goldstein, eschew both approaches with his lack of inheritance and heavily encoded LAD.

Only Cangelosi and Parisi, who abstract away from language as far as possible, represent a well-structured implementation of a single language evolution theory. Their neural nets comply with Steels’ specification for a general-purpose processing device. They conform to the precepts of both language adaption and genetic inheritance, and typify the study of synthetic ethology [18].

However disparate these models may seem, each employs what can be considered the universal principles of multi-agent language simulation. All of the systems are based on agents with a limited view of the world. These agents can change their states based on interactions with fellow agents, but do not have access to the internal states of their peers. Steels characterizes these principles as ‘autonomy’ and ‘distributedness’ [23]. He would also have ideal simulations include ‘groundedness’ as a property of agents as well, necessitating robotic elements and situating agents in a real-world environment, a necessary part of the evolution of language [26]. Agents are endowed with a small set of domain-specific abilities.

Secondly, each experiment contains sources of stochasticity. This property is unique to agent-based simulation. Mathematical modeling has no way of introducing random error into equations.

The interactions of the agents are structures and goal-driven. Rules are simple and unambiguous, and limited to a small domain. There is always some motivation encoded in the rules of interaction. Often it is imitation, sometimes an imperative to maximize a fitness function. However the interactions are implemented, there is always some inherent directive that agents must follow.

2 Simulation

de Boer’s simulation seeks to illustrate how vowel inventories self-organize. As described in Chapter 1, it is based in an ‘Adaptation and Self-organization approach,’ but also contains structures specifically tuned to language processing. This chapter will outline the nature of human vowel systems and describe the process of reproducing de Boer’s experiments.

2.1 Human Vowel Systems

Human vowel systems are largely similar across languages. There are sweeping generalizations to be made about their size, distribution, and acoustic prop-

Figure 1: Mid-sagittal representation of vocal tract, with tongue positions for /i/ and /u/ corresponding to the vowel trapezoid. [17]

erties. The range of sounds humans can produce is restricted by physiology. Vowel sounds have certain articulatory features: backness, height, and rounding. /i/ is a high front unrounded vowel; /o/ is mid, back, and rounded. In addition to feature specification, vowels are also thought of in terms of their acoustic properties. Certain frequencies in a vowel signal are naturally amplified; these are known as formants, and correspond roughly to tongue position. The first formant, called F1, at the lowest frequency, is inversely proportional to vowel height, the second proportional to backness [15]. Thus, by graphing formant frequencies onto a two-dimensional space, one can visually represent a vowel inventory that roughly corresponds to the place of articulation in the mouth. Figure 1 shows how tongue position relates to the trapezoidal vowel space. When the body of the tongue is high and front, the vowel produced, /i/, is located at the top left corner of the trapezoid. Formulas are often applied to the formant frequencies to achieve a direct mapping to tongue position (see section 2.5).

2.2 How Vowels Work

It has long been hypothesized that human vowel systems are self-organizing, meaning that they emerge out of physiological constraints that are independent of the particular language spoken. The two driving forces in this system are widely thought to be production and perception. Production constraints limit the sounds speakers can efficiently pronounce, coaxing vowel systems into a trapezoidal shape corresponding to the available articulatory space. The constraints of perception deal with ease of differentiation; individuals must be able to distinguish sounds in noisy environments, which causes the vowels in any given inventory to be as far apart as possible. Liljencrants and Lindblom [16] first approached this as an optimization problem, applying a mathematical formula to the vowel inventory of an individual. Bart de Boer detected a flaw in this approach, namely that it did not apply to a population, as proscribed by the newly developing principles of synthetic ethology [7]. de Boer turned the problem into one of multi-agent simulation, endowing agents with certain abilities and rules to follow, and achieving the same result as Liljencrants and Lindblom in the absence of central control and direct optimization. This paper aims to show that the underlying concepts of these two approaches are equivalent, by first replicating de Boer's experiments, then removing all of the extraneous linguistic details of the simulation. The resulting system contains only the bare essentials of a clustering system, independent of sound or language.

2.3 Implementation

de Boer ran a series of simulations that aimed to demonstrate the self-organization of vowel systems motivated by minimization of production and perception costs [7, 6, 8, 10]. His results accurately reflect the range of attested human vowel systems and his simulation is therefore considered to be equivalent to the mechanism that is responsible for vowel self-organization. The reproduced results, with slight modifications, will be called Version 1. Version 1 of the simulation is implemented in Objective C, using the Swarm platform. The implementation details follow.

2.3.1 Agents

The simulation contains a stable population of twenty agents, each endowed with the ability to produce speech sounds, perceive speech sounds, and store them in their memories. Each agent follows the same rules of interaction and adapts its own internal state accordingly. No agent has access to the internal states of others, and there is no central control.

At the beginning of the simulation, all agents start with empty vowel inventories. An agent may adapt its inventory by adding or deleting vowels, and by adjusting their prototypes by small values.

2.3.2 Vowels

A simulated vowel consists of four parts: the articulatory prototype, the acoustic prototype, and scores for success and use. The articulatory prototype is a pair of values between 0 and 1 representing vowel height and backness. These values are continuous, where a 0 in the first position indicates a low vowel, 0.5 a mid vowel, and 1 a high vowel. A 0 in the backness position corresponds to a high vowel, 0.5 mid, and 1 back. Therefore, an articulatory prototype of 1,0 corresponds to the vowel /i/ and 0.5,1 to /o/. Because the values are continuous, agents have the ability to pronounce any sound in the possible vowel space. de Boer's vowel implementation used three values in the articulatory prototype, corresponding to features of height, backness, and rounding. In Version 1, I have reduced the articulatory space by eliminating the third feature: rounding.

The acoustic prototype corresponds to the formant values for the vowel described by the articulatory prototype. de Boer used four formant values, then graphed F1 against F2', a combination of F2, F3, and F4. Formant values are calculated using de Boer's formula, interpolated from Maeda [19], but adjusted for the lack of a third dimension: rounding.

$$
\begin{aligned}
F_1 &= ((-392h^2 + 596h - 146)b^2 \\
&\quad +(348h^2 + -494h + 141)b \\
&\quad +(340h^2 + -796h + 708)) \\
F_2 &= ((-1200h^2 + 1320h + 188)b^2 \\
&\quad +(1864h^2 + -2644h - 561)b \\
&\quad +(-670h^2 + 1355h + 1517))
\end{aligned}
$$

In these experiments, only two formants are used, simplifying calculations and decreasing runtime. This modification is justifiable in human perception because humans are shown to hear vowels with two formants as identical to those with three or more [3]. As a result, no modifications are necessary to combine formants; all vowels are strictly two-dimensional.

In addition to the two prototypes, vowels also keep track of the number of times they have been used (both in speaking and listening) and the number of times they have been successful. These success and use scores are incremented according to the outcome of each imitation game and are used to make decisions about the quality of the corresponding vowel. Sounds with low success-to-use ratios will be removed from the inventory in favor of those with higher ratios.

When a vowel is produced, noise is added to the signal. This is one source of stochasticity in the simulation, and corresponds nicely to the noisy channels speakers and listeners encounter in the real world. Noise is a parameter of the simulation, meaning that it can be changed for different runs. The noise level varies informatively within a range of 0 to 0.25. When a vowel is pronounced, noise is added by multiplying both formant values by a random value in the range (1-n, 1+n). Because noise is multiplicative instead of additive, signals of higher frequencies will be more severely distorted than those of lower frequencies. However, this corresponds to human perception in that higher frequencies within the human speech range are more difficult to distinguish [21].

Vowels are distinguished acoustically by a weighted Euclidean distance measure, D, where

$$D = \sqrt{(F_1^a - F_1^b)^2 + \lambda(F_2^a - F_2^b)^2}$$

λ is the relative weight of F2, where lambda = 0.3 for all experiments [9]. This is related to human perception in that formants 2 and above are less acoustically salient [9]. This has the effect of increasing the degree of backness differentiation within individuals' vowel inventory.

Traditionally, the vowel space is graphed in Hertz, on a logarithmic or semi-logarithmic scale. Using Bark as units eliminates this need; it directly maps Hertz to Bark, which uses a logarithmic function for high frequencies and a linear function for low ones. This paper uses negative Bark for displaying all vowel inventories. Thus, a high front vowel with formant frequencies at $F1 =$

Speaker: A	Listener: B
If ($V = \emptyset$) Add random vowel to V Pick random vowel v from V $u_v \leftarrow u_v + 1$ Produce signal $A_1 : A_1 \leftarrow noise(ac_v)$	

	Receive signal A_1
	If ($V = \emptyset$)
	  Find phoneme (v_{new}, A_1)
	  $V \leftarrow V \cup v_{new}$
	Calculate v_{rec}:
	$v_{rec} \in V \wedge \neg\exists v_2 : (v_2 \in V \wedge D(A_1, ac_{v2})$
	  $< D(A_1, ac_{vrec}))$
	Produce signal $A_2 : A_2 = noise(ac_{vrec})$

Receive signal A_2
Calculate v_{rec}:
$v_{rec} \in V \wedge \neg\exists v_2 : (v_2 \in V \wedge D(A_1, ac_{v2})$
  $< D(A_1, ac_{vrec}))$
If ($v_{rec} = v$)
  Send non-verbal feedback:*success*
  $s_v \leftarrow s_v + 1$
Else
  Send non-verbal feedback:*failure*	

	Receive non-verbal feedback. **Update V according to feedback**
Do other updates of V.	**Do other updates of V.**

Figure 2: Language Interaction Algorithm

$250Hz$ and $F2 = 2200Hz$ will occur on the Bark scale at $F1 = -6Bark$ and $F2 = -7Bark$. Lower vowels will have lower F1 values and farther back vowels will have higher F2 values.

$$Bark = \begin{cases} \frac{\ln(\frac{Hertz}{271.32})}{0.1719} + 2 & Hertz > 271.32 \\ \frac{Hertz - 51}{110} & Hertz \leq 271.32 \end{cases}$$

Agents do computations on formant frequencies in Hertz; Bark is used only for neatly displaying results.

2.3.3 Language Interaction

Agents interact in an iterative series of what de Boer calls ‘imitation games.’ Figure 2 describes the interaction algorithm in pseudo-code, where s_v and u_v represent the success and use scores of a vowel, respectively, S is the synthesizer function, and D is the Euclidean distance between two vowels. This algorithm is explained in detail below.

At each time step, two agents are selected at random from the population: one speaker and one listener. The speaker (A) selects a vowel at random from his inventory; if the inventory is empty, he makes up a new vowel by randomly selecting values for an articulatory prototype. A then pronounces this vowel, adding noise to the formants within the parametric range.

The listener (B) then perceives these formants and compares them to those of the vowels in his own inventory. If B's inventory is empty, he finds an acoustically similar vowel by talking to himself. He then finds the closest vowel he knows, as measured by the Euclidean distance D. B produces this vowel, adding noise.

A hears B's vowel and compares it to the original one he spoke. If the two are acoustically close enough, meaning that they can be confused in the presence of the specified noise levels, the game is a success. If the agents' vowels are not the same, the game fails. A then communicates the outcome of the game to B nonverbally. de Boer justifies this process with speculation about the nature of child language acquisition: Facial expressions or other types of nonverbal feedback can give individuals information about the success of communication [9]. Both agents then update their internal states based on the outcome of the game.

If an imitation game was successful, both agents increment the success scores of the corresponding vowel. B makes adjustments to the prototype by shifting the articulatory prototype by small increments and seeing if the resulting acoustic signal is a better imitation of the original. If the game was a failure, B analyzes the failed vowel in terms of its previous success rate. If the vowel has a success/use ratio less than a certain threshold (0.7), it is probably a bad vowel and is removed from B's inventory. If the success/use ratio is greater than the threshold, some effort is made to adapt by shifting it closer to the original signal (see Figure 3).

Other actions the agents perform are described in Tables 2 and 3. **Do other updates of V** is executed by both agents after the completion of each imitation game. Agents remove the vowels that have low success/use ratios, merge vowels that can be confused in the presence of noise, either by their acoustic or articulatory prototypes, and add new vowels to their inventories. **Merge** acts on a pair of vowels, simply deleting the vowel with the lowest success/use ratio and adding those success and use scores to the more successful vowel.

Find phoneme sets the articulatory prototype to 0.5 for all features, then gets the corresponding acoustic prototype by calling the synthesizer function and calls **Shift closer**. **Shift closer** tests each of the articulatory neighbors of a vowel, determined by adjusting both the height and backness values of the prototype by ± 0.1. If any of these is closer to the desired signal, the agent abandons the original signal in favor of the improved one. This process is repeated in **Find phoneme** until a satisfactory vowel is found.

Shift closer(v, A)	**Do other updates of V**

$v_{best} \leftarrow v$
For (all articulatory neighbors $v_{neigh} of v$) do:
    If $(D(ac_{vneigh}, A) < D(ac_{vrec}, A))$
        $v_{best} \leftarrow v_{neigh}$
$v \leftarrow v_{best}$
	For $(\forall v \in V)$
	    If $(s_v/u_v < 0.7 \wedge u_v > 5)$
	        $V \leftarrow V - v$
	For $(\forall v_1 \in V)$
	    For $(\forall v_2 : (v_2 \in V \wedge v_2 \neq v_1))$
	        If $(D(ac_{v1}, ac_{v2} <$acoustic threshold)
	            Merge(v_1, v_2, V)
	        If $(D(ar_{v1}, ar_{v2}) <$articulatory threshold)
	            Merge(v_1, v_2, V)
	Add new vowel to V with small probability

Merge$(v_1, v2, V)$	**Find phoneme**(v_{new}, A)	**Update according to feedback**

If $(s_{v1}/u_{v1} < s_{v2}/u_{v2})$
    $s_{v2} \leftarrow s_{v2} + s_{v1}$
    $u_{v2} \leftarrow u_{v2} + u_{v1}$
    $V \leftarrow V - v_1$
Else
    $s_{v1} \leftarrow s_{v1} + s_{v2}$
    $u_{v1} \leftarrow u_{v1} + u_{v2}$
    $V \leftarrow V - v_2$
	$ar_v \leftarrow (0.5, 0.5)$
	$ac_v \leftarrow S(ar_v)$
	$s_v \leftarrow 0$
	$u_v \leftarrow 0$
	Do
	    $v_{new} \leftarrow v$
	    Shift closer(v_{new}, A)
	Until$(v = v_{new})$
		$u_{vrec} \leftarrow u_{vrec} + 1$
		If (feedback signal = *success)*
		    Shift closer(v_{rec}, A_1)
		$s_{vrec} \leftarrow s_{vrec} + 1$
		Else
		    If$(u_{vrec}/s_{vrec} > threshold)$
		        Find phoneme(v_{new}, A_1)
		$V \leftarrow V \cup v_{new}$
		    Else
		        Shift closer(v_{rec}, A_1)

Figure 3: Internal Agent Algorithms

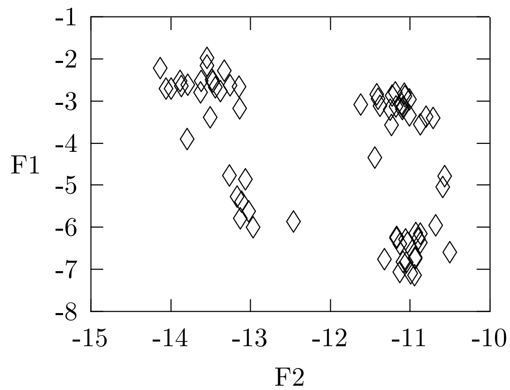

Figure 4: Year 5000, Noise 0.15

2.4 Results

The results of the Version 1 are manifested as a graph of vowels in two-dimensional space, with F2 on the x-axis and F1 on the y-axis. The vowel inventories of all twenty agents are superimposed. Because of the relationship between formant frequencies and articulatory features, this graph can be imagined as a map of the oral cavity. It is important to note that the vowel space of the results is trapezoidal, to correspond with the articulatory constraints set by the synthesizing function.

Figure 4 displays the results of Version 1 after 5000 imitation games, or years, in the presence of 15 percent noise. The clustering effect is quite clear. This run of the simulation produced a six-vowel system, roughly symmetrical and evenly spaced. The points that are not part of clusters in the left-hand side of the figure represent sounds newly introduced to the system. They have not had enough time to propagate throughout the population. Depending on the small variations and stochastic behavior of the system, they may catch on and become part of the inventory, or they may be absorbed into a previously existing cluster, or deleted altogether. This behavior is characteristic of agent-based simulation. While small variations and local events are subject to randomization, the system as a whole exhibits deterministic behavior on a larger scale.

2.4.1 Time

Because the agent interactions are continually subject to stochasticity, it is particularly important to observe the behavior of the system over time. Figure 5 shows the progression of a population's vowel inventory over a number of time steps. After 20 games most agents have had a chance to speak at least once. There has not been enough opportunity for interaction for any system-wide behavior to emerge. At 500 games, clusters are beginning to develop.

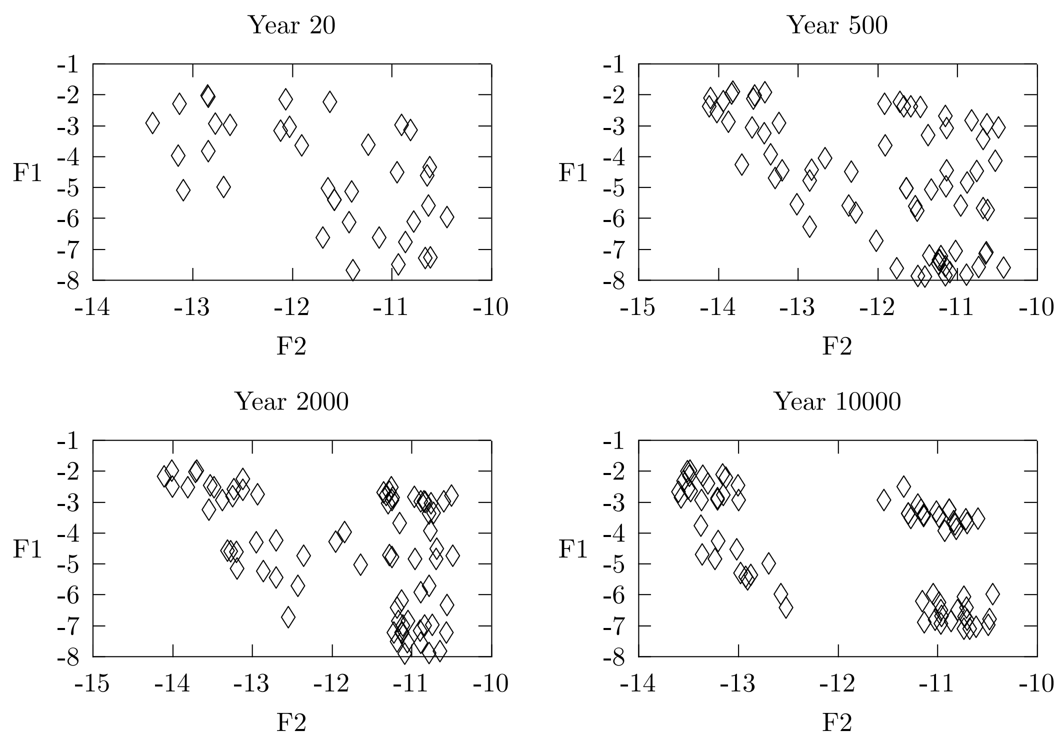

Figure 5: Time-series: Simulation output at years 20, 500, 2000, 10000 for 0.15 noise.

There are one or two vowels that have taken hold; nearly every agent in the simulation has them. Others are newly invented or starting to cluster. At 2000 games the interaction level is nearing sufficient for inventories to span the entire available vowel space. At this point, the system is robust to the creation of new vowels; they are quickly absorbed by nearby clusters. The population inventory is stable, but not static. 10,000 years into the simulation the inventory is much the same as it was at 2000. However, some clusters have shifted slightly and most are packed more tightly together.

2.4.2 Noise

Version 1 behaves as expected under differing noise levels. As signals become noisier, agents must pay more attention to differentiating between adjacent vowels. Therefore a population evolves fewer clusters that are further apart. At low noise levels vowel clusters are more numerous and closer together. The concept of 'noise' is a delicate one, and will be discussed further in Chapter 4.

2.4.3 Variation

Version 1 is equivalent to de Boer's system, as it evolves the same type behavior over under the same circumstances. Due to the stochasticity inherent to both simulations, no two runs of the program will be the same, but they will exhibit certain similar properties. All five graphs show vowel inventories after 10,000 games. However, each is the result of a different series of events. The choices of vowels to use, the outcomes of individual imitation games, and the addition of new vowels by agents are different in each run, but produce similar results. Each inventory is spread maximally through the acoustic space and while each shows slightly different degrees of clustering, they have achieved clustering in the same way.

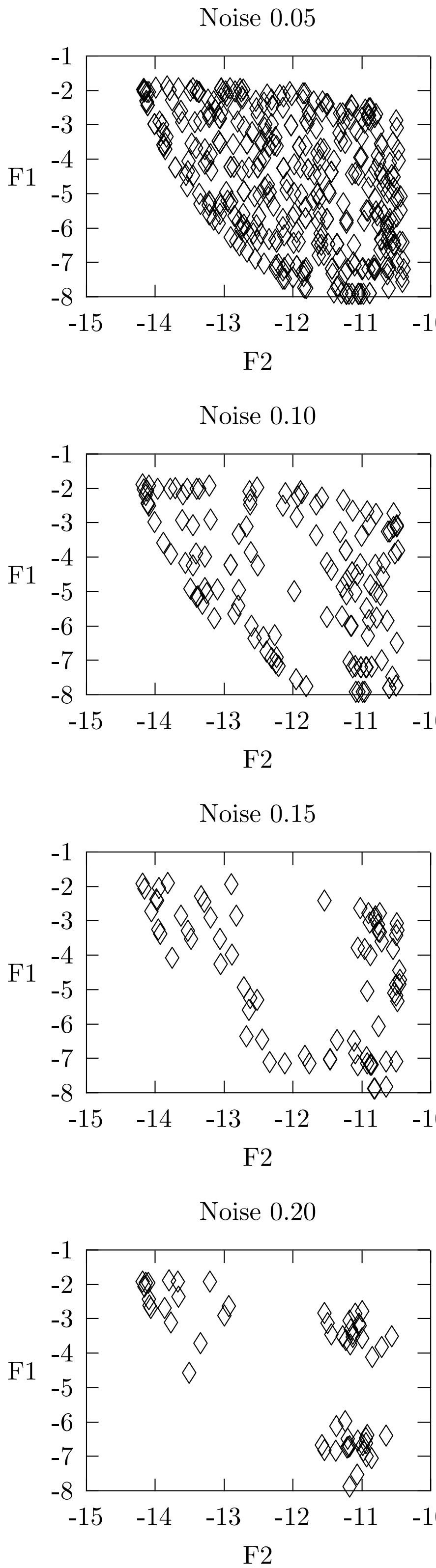

Figure 6: Noise-series: Simulation output at year 10000 for 0.05, 0.10, 0.15, 0.20 noise.

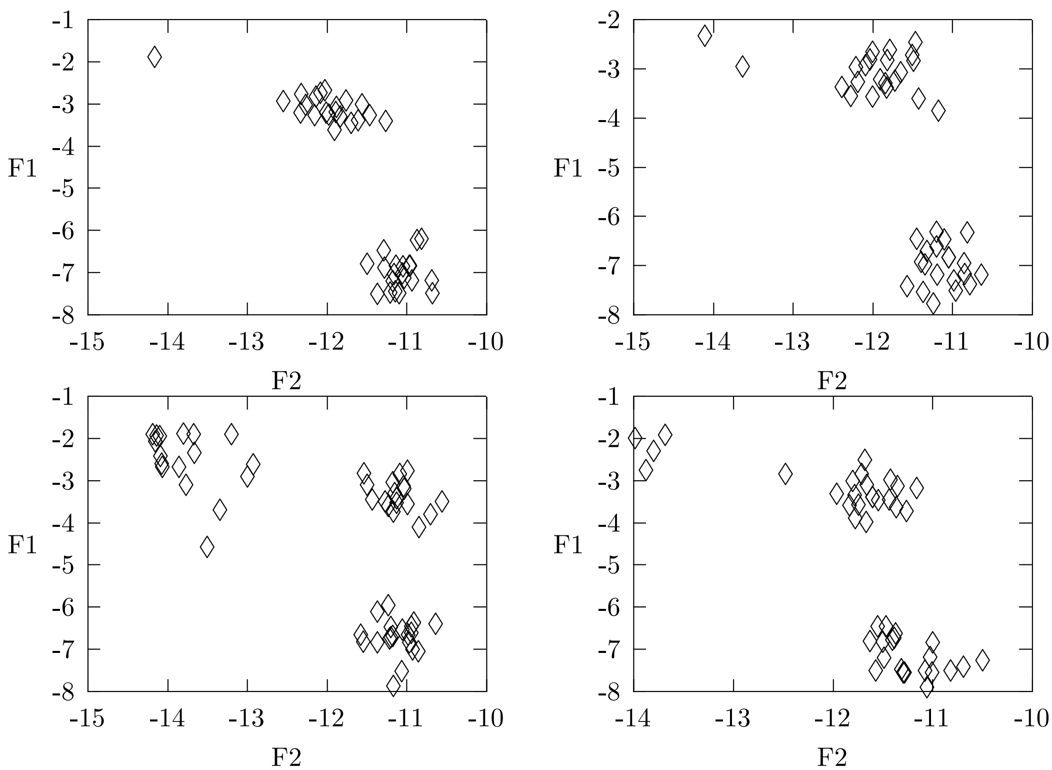

Figure 7: Results of four separate simulation runs, year 10,000, noise 0.2.

Shift closer(v, A)	***Do other updates of V**

$v_{best} \leftarrow v$
For (all acoustic neighbors $v_{neigh} of v$) do:
   If $(D(ac_{vneigh}, A) < D(ac_{vrec}, A))$
      $v_{best} \leftarrow v_{neigh}$
$v \leftarrow v_{best}$
	For $(\forall v \in V)$
	   If $(s_v/u_v < 0.7 \wedge u_v > 5)$
	      $V \leftarrow V - v$
	For $(\forall v_1 \in V)$
	   For $(\forall v_2 : (v_2 \in V \wedge v_2 \neq v_1))$
	      If $(D(ac_{v1}, ac_{v2} <$acoustic threshold)
	         Merge(v_1, v_2, V)
	Add new vowel to V with small probability

Merge$(v_1, v2, V)$	**Find phoneme**(v_{new}, A)	**Update according to feedback**

If $(s_{v1}/u_{v1} < s_{v2}/u_{v2})$
   $s_{v2} \leftarrow s_{v2} + s_{v1}$
   $u_{v2} \leftarrow u_{v2} + u_{v1}$
   $V \leftarrow V - v_1$
Else
   $s_{v1} \leftarrow s_{v1} + s_{v2}$
   $u_{v1} \leftarrow u_{v1} + u_{v2}$
   $V \leftarrow V - v_2$
	
	$ac_v \leftarrow (1000, 5000)$
	$s_v \leftarrow 0$
	$u_v \leftarrow 0$
	Do
	   $v_{new} \leftarrow v$
	   Shift closer(v_{new}, A)
	Until$(v = v_{new})$
		$u_{vrec} \leftarrow u_{vrec} + 1$
		If (feedback signal = *success)*
		   Shift closer (v_{rec}, A_1)
		$s_{vrec} \leftarrow s_{vrec} + 1$
		Else
		   If$(u_{vrec}/s_{vrec} > threshold)$
		      ***Find phoneme**(v_{new}, A_1)
		$V \leftarrow V \cup v_{new}$
		   Else
		      Shift closer(v_{rec}, A_1)

Figure 8: Modified Agent Actions (changes marked with *)

3 Adaptation

The synthetic model is demonstrated to output the same results as attested human behavior, and can therefore be considered an equivalent system. This is excellent evidence to support the long-standing belief that vowel inventories are self-organizing, as well as subject to costs of production and perception. However, I contend that a more basic mechanism is at work in these experiments. The process that causes vowels to cluster can be isolated and used in a range of applications. The simulation depends on agents passing 'vowels' among themselves, where 'vowels' are represented as abstract objects consisting of an articulatory prototype, an acoustic prototype, and scores for success and use. While this is a conceivable solution to how human brains store vowels, it is an unnecessary embellishment to the simulation. I aim to demonstrate that the same behavior can be produced by a more simplified system, and one that abstracts so far away from actual speech that it can be considered domain-independent.

3.1 Implementation

What was a vowel in Version 1 is now expressed as a 'signal' in Version 2. A signal consists of two numerical values, x and y, which are continuous within a specified range, and scores for success and use. Agents pass these signals in an

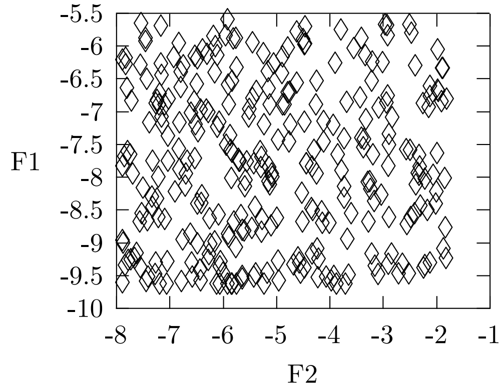

Figure 9: Results of Version 2, year 500, noise 0.05.

imitation game format identical to that of Version 1 (see table 1). All mention of prototypes is eliminated. Signals are shifted closer by adding or subtracting a small value from their numerical x and y elements. The synthesizer function is no longer needed to map articulatory prototypes to acoustic signals.

The major implementational differences occur in **Shift closer**, **Find phoneme**, and **Do other updates of V** (see Figure 8). **Shift closer** now finds acoustic neighbors by adding or subtracting a small value (usually 50) in Hertz. **Find phoneme** starts by setting the acoustic, not articulatory, prototype to a middle value and doesn't run the synthesizer. In **Do other updates of V**, all merging takes place on acoustic distance; the articulatory merge clause is eliminated.

The results of Version 2 are very similar to Version 1. They still exhibit clustering and maximally span the available vowel space. One notable difference is in the shape of the vowel space. Because the acoustic synthesizer no longer constricts the signals to possible human sounds, the trapezoid is lost, and the vowel space is square. Compare Figure 9 to the first graph in Figure 6.

This behavior is entirely predictable. The synthesizer function de Boer uses to map articulatory prototypes onto acoustic signals effectively restricts the phoneme space to the canonical trapezoid. This element of the simulation serves solely to make the resulting vowels look like those of attested human systems. Any other synthesizer function would produce a shape proscribed by the equations it contained. A direct mapping produces a square.

3.2 Results

The results of Version 2 are very similar to Version 1. They still exhibit clustering and maximally span the available vowel space. One notable difference is in the shape of the vowel space. Because the acoustic synthesizer no longer constricts the signals to possible human sounds, the trapezoid is lost, and the vowel space is square. Compare Figure 9 to the first graph in Figure 6.

This behavior is entirely predictable. The synthesizer function de Boer uses

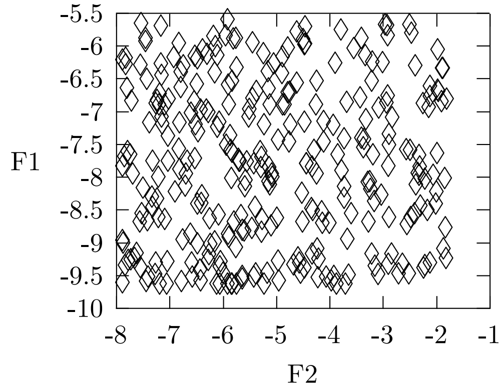

Figure 10: Results of Version 2, year 500, noise 0.05.

to map articulatory prototypes onto acoustic signals effectively restricts the phoneme space to the canonical trapezoid. This element of the simulation serves solely to make the resulting vowels look like those of attested human systems. Any other synthesizer function would produce a shape proscribed by the equations it contained. A direct mapping produces a square (see Figure 9). Figure 10 shows the vowel inventories at the same time steps for two different sets of synthesizer equations. The trapezoid on the left represents the original formulas described in 2.3.2. The triangle on the right shows the effects of changing those formulas to the following:

$$\begin{aligned} F_1 &= 3000h^2 \\ F_2 &= 1400bh \end{aligned}$$

3.2.1 Time

Despite the simplification of the simulation, clustering still takes place, if possibly over a longer period of time due to the increased vowel space. Version 2 exhibits the same patterns of convergence through time as does Version 1, as shown in Figure 10. Again, phoneme clusters develop after a sufficient number of imitation games, and shift slowly as more vowels are added, imitated, and merged. The only behavioral difference between Versions 1 and 2, aside from the shape and size of the signal space,

3.2.2 Noise

Noise-series results (Figure 11) show similar properties to Version 1. As the noise level increases, the number of distinct clusters decreases. At all noise levels, the population takes advantage of the largest available vowel space, and clusters are evenly distributed throughout the space.

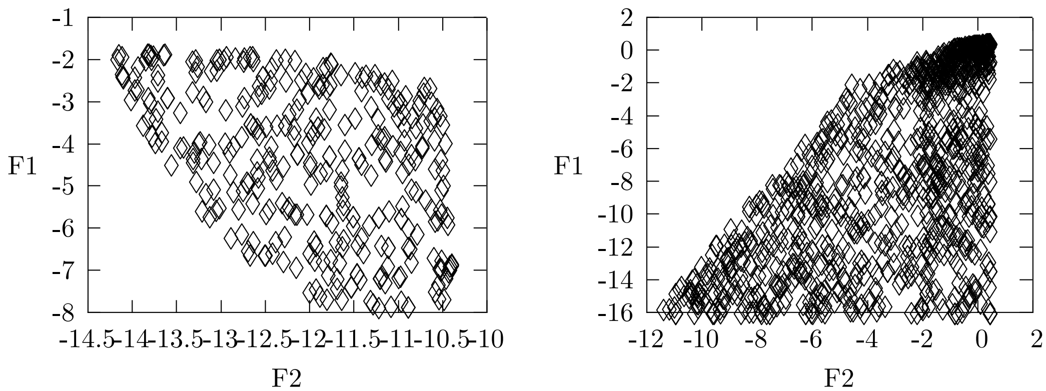

Figure 11: Vowel spaces at year 500 for two different synthesizer functions

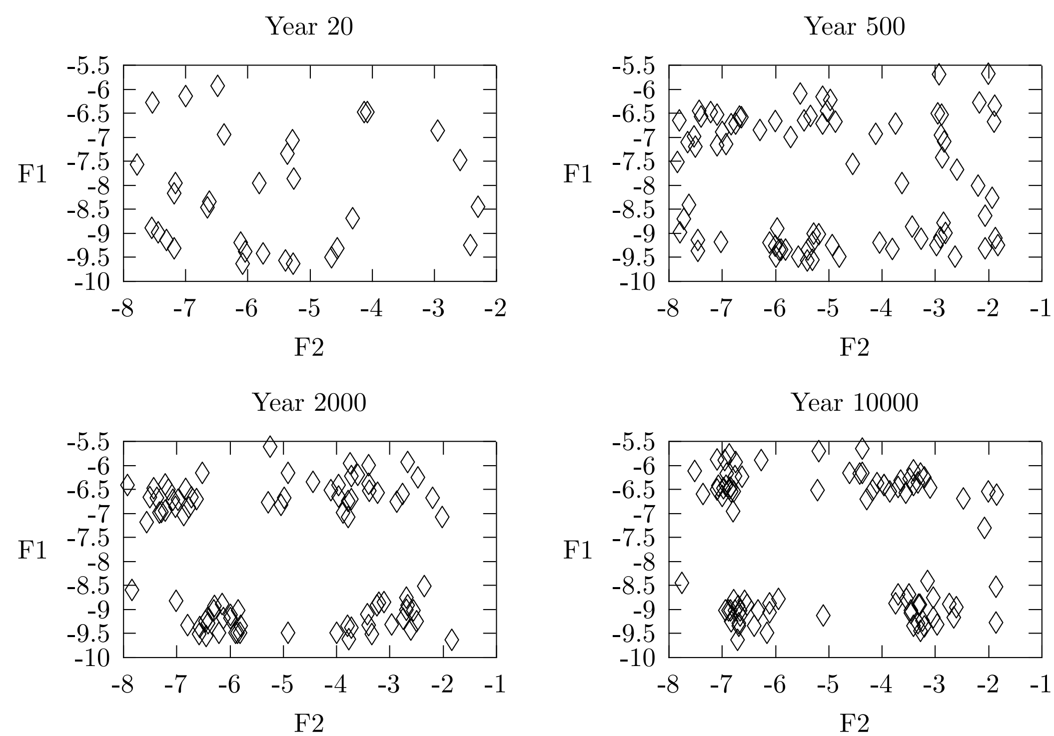

Figure 12: Time-series: Simulation output at year 20, 500, 2000, and 10000 for 0.15 noise.

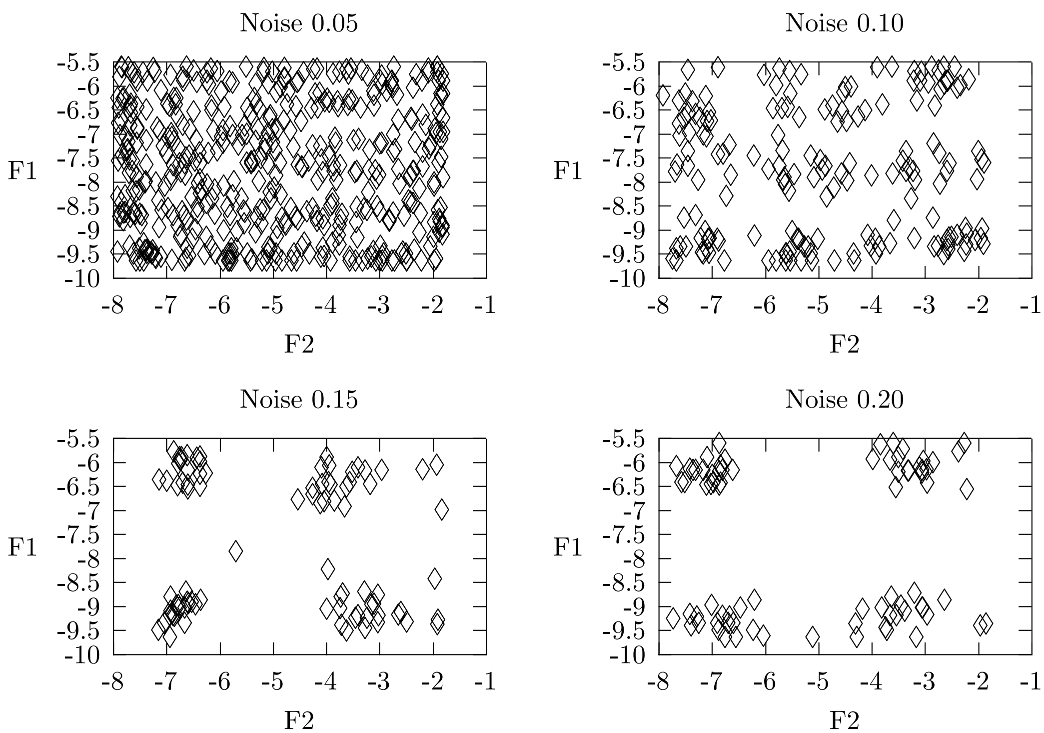

Figure 13: Noise-series: Simulation output at year 10000 for 0.05, 0.10, 0.15, 0.20 noise.

3.3 Analysis

Because this simplified system exhibits equivalent clustering behavior to de Boer's, we can conclude that none of the linguistic features that were eliminated were necessary for a system to exhibit clustering. This conclusion is informative when considering synthetic approaches to any self-organizing system. The consequences thereof will be discussed in the following chapter.

4 Conclusion

4.1 The clustering mechanism

de Boer shows that the behavior of human vowel systems can be reproduced using agent-based simulation. By modifying the algorithm to eliminate several complicating factors, this experiment has shown equivalent behavior in a simplified system. This demonstrates that the articulatory prototype, the synthesizer function, the number of formants, and all processes that rely on them, directly or indirectly, are unnecessary to the clustering mechanism that drives the system. This is not to say that the human brain does not store articulatory prototypes, or that the vocal tract does not produce an acoustic signal dependent on its shape, but only that the emergence of clusters does not depend on them.

Instead, the clustering mechanism is based in the imitation games. The synthesizer function is responsible for producing a realistic vowel trapezoid. Though artificial, this is allowable because it restricts the behavior of the system in the same way as attested human systems are physiologically restricted. Because the function effectively imitates the mapping of articulatory features to acoustic signals in the human vocal tract, it does not deter from the simulation, but rather casts it in an understandable domain. This is in keeping with Worden's second principle of language evolution theory. Limiting the acoustic space exploits what is known about how language works, and also sets the model in the arena of human physiology [14]. The same is true for formant information. A system could cluster based on any number of formants, but humans can distinguish four (to varying degrees), and therefore any system that attempts to mimic human behavior should do the same.

4.2 Optimization

Liljencrants and Lindblom's mathematical approach uses explicit optimization of a production and perception function for one speaker's vowel inventory. de Boer's agent-based system casts the same process as distributed optimization over a population of speakers. Effectively, the only change he made was taking the action of optimization out of central control and into the hands of the agents. This is misleading, though. While there is no central system controlling the form of the vowel inventory, the agents are all given the same imperative to imitate. This is common in the field of synthetic ethology, where motivation is encoded in the rules for interaction. Because all agents have the same range of behaviors, the simulation does run under a limited form of central control.

de Boer rejected the mathematical optimization approach for its lack of realism. However, all he has done is to perform the same optimization over a number of agents in a stochastic domain. His simulation assumes that minimization of production and perception costs is precisely the force at work, effectively reproducing Liljencrants and Lindblom's formula in a distributed arena. He has, however, achieved an implementation of such optimization in the domain of synthetic ethology.

4.3 Production and Perception

All models of vowel self-organization discussed in this paper rely on the assumption that individuals want to minimize the cost of signal production and perception. That is to say, speakers strive to use the least amount of energy to pronounce sounds, and listeners do the same in differentiating them. This is manifested in the agent behaviors of merging vowels that take too much energy to differentiate, discarding vowels that prove unsuccessful, and randomly introducing new vowels. The two processes work in different directions; sounds that take little effort to produce are more likely to be difficult to perceive. This model has been widely hypothesized, becoming the dominant theory of phonemic organization. Liljencrants and Lindblom explicitly assign production and perception costs to individual signals, then optimize a function to find the minimum expenditure of both in combination [16]. de Boer distributes this process, allowing agents to merge vowels that take too much effort to differentiate, while adding new vowels to take advantage of the available acoustic space. This may be seen in the resulting vowel inventories as a tendency to evolve tight clusters, while spreading them as far apart as possible. Because human vowel systems are observed to display said properties, we may conclude that our underlying assumptions about production and perception are correct, or at least equivalent to human cognitive processes.

4.4 Simulation deficiencies

Several of the language simulations discussed in chapter 1 set out an explicit task for the agents to perform. In fact, it can be considered one of the features of agent-based simulation. In limiting the domain to only the relevant processes, agent behavior is forced in a predetermined direction. By encoding an affinity for imitation in the agents' behavior, we are adding not just a specialized language processing device, but a low-level behavioral algorithm designed to evolve population-level patterns. While this is consistent with other experiments in the same field, it violates the basic principles of all three of Curran and O'Riordan's theories of language evolution.

Another imperfection in the simulation concerns stochasticity in implementation. Though their interactions are assigned randomly, over time all agents get to talk to each other. Because the random numbers that choose game players are uniformly distributed, this is equivalent to letting all agents talk to all others. A modification that may make the simulation more realistic is to restrict interaction to pairs of physical neighbors. As agents move around a physical environment, they could only speak to agents in adjacent positions. Not only would this more closely emulate the behavior of human populations, it would be truer to the precepts of synthetic ethology, that stress the importance of studying agents within an environment.

4.5 Noise

The single most influential variable in both Version 1 and 2 is noise. It dictates the diameter of the clusters, as well as their distance from each other. It is easy to think of this variable in terms of the actual acoustic noise we can hear, but it would be unrealistic to apply this feature to human vowel systems. Clearly, those languages with only five vowels did not evolve in an environment with more ambient noise than those with richer inventories. Instead, 'noise' must be defined as some process independent of vowel self-organization, whatever it may be. If a language is consonant-rich, fewer vowels are needed in order to effect a large enough lexeme space for the necessary meanings. However, if it has very few consonants, a small vowel inventory would lead to either a high level of lexeme overlap or ambiguity, or a need for very long words to avoid such ambiguity. This is only one possibility for the presence of 'noise' in a vowel system. Probably it is erroneous, but it is helpful in avoiding the dependence of the variable on what we generally term 'noise.'

4.6 Consequences for human behavior

The difference between sufficient and necessary conditions was discussed in 1.5.3. It is here that such a distinction becomes important. We have shown equivalence in the clustering mechanisms of human vowel inventories and the synthetic models. Drawing a parallel is easy, but explaining the actual computations taking place in the human brain is much harder. The relevant question, then, concerns the application of the demonstrated clustering mechanism to actual language evolution. In other words, did vowels self-organize as a result of iterated imitation games? It is entirely possible. There is currently no evidence to the contrary, so we may conclude that imitation is a candidate explanation for the cognitive process behind the emergence of clustered phoneme inventories. This approaches Worden's fourth criterion for a theory of language evolution [29]: it suggests a theory of human language computation.

In considering a population-level vowel inventory, we necessarily accept that clusters exist. It would be unreasonable to assume that all language users produce or store exactly the same acoustic signal. Instead, we acknowledge that the signals for two speakers must be sufficiently similar to be identified as the same sound. Thus, charting the vowel inventories of a population of speakers on top of one another will result in clusters much the same as the Version 1 output.

4.7 Consequences for further research

Quite possibly, there are still elements of Version 2 that may be simplified, and a clustering mechanism retained. Nevertheless, the simulation has clearly been moved outside the realm of vowels. With small adjustments, the language interaction algorithms may be embellished to represent a wide range of behaviors that exhibit the same type of clustering. Candidate systems may encompass self-organizing population-wide behaviors in lexeme assignment, market prices, traffic jams, or housing patterns. The clustering mechanism may be what is at the heart of self-organization. Then again, though these are sufficient conditions for the emergence of human-like vowel inventories, we cannot claim they are necessary. However, they do seem to be equivalent, and are therefore the clustering mechanism derived in Version 2 is a very powerful tool for studying self-organizing systems.

Note: All original program code discussed in this paper is available online at http://www.sccs.swarthmore.edu/users/04/ethomfo1/thesis/

References

[1] C. Angelo and D. Parisi. The emergence of a language in an evolving population of neural networks. Technical report, National Research Council, Rome, 1996. NSAL–96004.

[2] C. Browman and L. Goldstein. Competing constraints on intergestural coordination and self-organization of phonological structures. *Bulletin de la Communication Parlee*, 5:25–34, 2000.

[3] R. Carlson, B. Granstrom, and G. Fant. Some studies concerning perception of isolated vowels. *Speech Transmission Laboratory-Quarterly Progress and Status Report*, 2-3:19–35, 1970.

[4] N. Chomsky. *Language and Mind*. Harper and Row, New York, 1972.

[5] D. Curran and C. O'Riordan. Language evolution in artificial systems. Technical report, Department of IT., NUI, Galway, 2002.

[6] B. de Boer. Generating vowel systems in a population of agents. In *Fourth European conference on Artificial Life*, pages 503–510. MIT Press, Cambridge, MA, 1997.

[7] B. de Boer. Self organisation in vowel systems through imitation. In *Computational Phonology, Third Meeting of the ACL SIGPHON*, 1997.

[8] B. de Boer. Emergence of sound systems through self-organisation. In H. La Poutre' and J. van den Herik, editors, *NAIC '98 Proceedings*, pages 37–46, Amsterdam, 1998.

[9] B. de Boer. Self-organisation in vowel systems, 1999. PhD Thesis, Vrije Universiteit Brussel AI-lab.

[10] B. de Boer. Emergence of vowel systems through self-organisation. *AI Communications*, 13(1):27–40, 2000.

[11] J.M. Epstein and R.L. Axtell. *Growing Artificial Societies: Social Science from the Bottom Up.* Brookings Institution, 1996.

[12] K. Harrison, M. Dras, and B. Kapicioglu. Emergent behavior in phonological pattern change. In *Artificial Life VIII.* MIT Press, 2003.

[13] E. Hutchins and B. Hazlehurst. How to invent a lexicon: The development of shared symbols in interaction. In N. Gilbert and R. Conte, editors, *Artificial Societies: The Computer Simulation of Social Life*, pages 157–189. UCL Press: London, 1995.

[14] K. Kohler. The development of sound systems in human language. In J. Hurford, M. Studdert-Kennedy, and C. Knight, editors, *Approaches to the Evolution of Language*, pages 265–278. Cambridge University Press, 1998.

[15] P. Ladefoged. *A Course in Phonetics.* Harcourt Brace, 2001.

[16] L. Liljencrants and B. Lindblom. Numerical simulations of vowel quality systems. *Language*, 48:839–862, 1972.

[17] I. MacKay. *Phonetics: The Science of Speech Production.* Little and Brown, Boston, 1987.

[18] B. MacLennan. Synthetic ethology: An approach to the study of communication. In C. Langton, C. Taylor, J. Farmer, and S. Rasmussen, editors, *Artificial Life II*, pages 631–658. Addison-Wesley, Redwood City, CA, 1992.

[19] S. Maeda. Compensatory articulation during speech: Evidence from the analysis and synthesis of vocal tract shapes using an articulatory model. In W.J. Hardcastle and A. Marchal, editors, *Speech Production and Speech Modelling*, pages 131–149. Kluwer, 1986.

[20] T. Satterfield. The socio-genetic solution: A new look at language genesis through swarm modeling. Technical report, Center for the Study of Complex Systems, University of Michigan, 2000. CSCS-2000-007.

[21] M. Schroder, S. Atal, and J. L. Hall. Objective measure of certain speech signal degradations based on masking properties of human auditory perception. In B. Lindblom and S. Ohman, editors, *Frontiers of speech communciation and research*, pages 217–229, London, 1979. Academic Press.

[22] K. Smith, H. Brighton, and S. Kirby. Complex systems in language evolution: the cultural emergence of compositional structure. *Advances in Complex Systems*, forthcoming.

[23] L. Steels. The synthetic modeling of language origins. *Evolution of Communication*, 1:1–34, 1997.

[24] L. Steels. The origins of syntax in visually grounded robotic agents. *Artificial Intelligence*, 103(1-2):133–156, 1998.

[25] L. Steels. The puzzle of language evolution. *Kognitionswissenschaft*, 8(4):143–150, 2000.

[26] L. Steels and P. Vogt. Grounding adaptive language games in robotic agents. In C. Husbands and I. Harvey, editors, *Proceedings of the Fourth European Conference on Artificial Life*, Cambridge MA and London, 1997. MIT Press.

[27] I. Ulbaek. The origin of language and cognition. In J. Hurford, M. Studdert-Kennedy, and C. Knight, editors, *Approaches to the Evolution of Language*, pages 30–43. Cambridge University Press, 1998.

[28] R. Walshe. The origin of the speeches: Language evolution through collaborative reinforcement learning. *Lecture Notes in Computer Science*, 2190:239, 2001.

[29] R. Worden. The evolution of language from social intelligence. In J. Hurford, M. Studdert-Kennedy, and C. Knight, editors, *Approaches to the Evolution of Language*, pages 148–166. Cambridge University Press, 1998.

[30] W. Zuidema. Emergent syntax: The unremitting value of computational modeling for understanding the origins of complex language. *Lecture Notes in Computer Science*, 2159:641, 2001.