AN ECOLOGICAL SIMULATION OF AGENTS WITH SELF-ORGANIZING VOWEL SYSTEMS

Kaspar de Geus, Jeroen Kools
{KGEUS, JKOOLS}@SCIENCE.UVA.NL

Abstract: This paper presents a simulation of a world where agents evolve a vowel system through interaction with other agents. These populations interact with each other through imitation games which are played between agents that are *close* to each other. We will show that depending on location vowel system will differ between agents and between population clusters.

1. Introduction

The evolution of language is affected by many different kinds of factors, ranging from physiological and social to environmental and historical influences. Likewise, the emergence and individual acquisition of language and sound systems is a very complex process. Within certain contexts it has been shown that social and ecological parameters (Fought *et al.*, 2004; Holman *et al.,* 2007; Trudgill, 2004), genetic factors (Dediu & Ladd, 2007) and (for animals) acoustic properties of the habitat (e.g. Sugiura *et al*., 2006) act upon these linguistic phenomena. For example, Trudgill suggests that community size and isolation have a direct effect on specific properties of phoneme inventories in these communities, based on empirical evidence in Austronesian languages. Nevertheless, such evidence is scarce and not very reliable without theoretical verification by means of a computational model.

However, the more computationally inclined strains of research in evolution of language have not given much consideration to some of these factors. Linguistic simulations have shown that vowel systems resembling ones found in humans can emerge in changing groups of agents, when they interact in so-called imitation games (De Boer 1998, De Boer & Vogt, 1999). Other models show that social status plays a role in language change and diversity (Nettle, 1998) and also that there are relations between geographical distance and similarity of languages (Holman *et al*., 2007). But ecology remains a largely unexplored factor in these models.

When we say ecology, we envision a system of physical properties of a spatial environment that either increase or decrease a population's chances for survival and spread. In reality, ecological factors are expressed in geographical features such as seas, deserts and mountain ranges, or in variations in soil fertility and abundance of other resources. The former are all obstacles to migrations, causing isolation, while the latter constrain or facilitate population growth in the area. Some existing models do have a spatial representation of the world, but these are generally featureless rectangular lattices with evenly distributed populations (Nettle 1998, Barr 2004). This paper will investigate the emergence of vowel systems in a model that does represent said factors, and look into their effects on resulting vowel systems. More specifically, we will use such a model to try to verify the suggestions made by Trudgill and Holman.

2. Background

Agent-based simulations have a history already spanning several decades in disciplines as diverse as economics, sociology and biology (e.g. Reynolds, 1987). Since such systems have proven to be well-suited for the study of complex, group-level behavior, it is no surprise that agent-based modeling has found its way into linguistics. Topics in computational linguistics that have been studied with this paradigm include language evolution, acquisition and diversity, evolution of meaning, evolution of grammar, et cetera. Luc Steels (1997) introduced the notion of *imitation games* to this context. Bart de Boer, in turn, wrote a series of papers on the subject of vowel systems emerging in groups of agents interacting in this way.

An imitation game is a specific sequence of interactions between two agents, with one agent imitating an action of the other and drawing conclusions from the imitation's success or failure. This principle is ultimately derived from one of the *language games* described by Wittgenstein (1953/1973).

3. The model

The simulation consists of four basic, modular parts: a world model, an agent model, a vowel space model and an interaction scheme. On top of that sits the top level program that controls the simulation. Time in the simulation elapses in discrete steps called *epochs*. Every epoch, the agent population changes and communicates. We will now discuss the details of each component.

Firstly the world itself is a two-dimensional grid. Every tile has two ecological parameters and is home to a number of agents. The agents themselves, however, have a more precise, continuous location within their tile. A tile has a value representing its *habitability,* which is a number between 0 and 1 that indicates how 'good' it is for an agent to live there. Each tile also has a *traversability* value between 0 and 1 which indicates how easy it is to cross this square. This limits the agent's movement and communication range.

The agents have the ability to move, communicate and of course reproduce and die. They are born to one parent, inheriting nothing but their location. Agents have an age, and as they get older their learning rate slowly declines. Their memory is a two-dimensional space storing vowel locations. A vowel's xy location represents the first and effective second formant in Hertz. The vowel space is bound to the parameters in (1) with $200 \le f_1 \le 900$ to restrict it to vowels spoken by human beings[1]:

$$f_2 = \begin{Bmatrix} 1200 + 6f_1 \ge f_2 \ge 6600 - 21f_1 & ,if \quad 200 \le f_1 \le 300 \\ 3750 - 2{,}5f_1 \ge f_2 \ge -150 + 1{,}5f_1 & ,if \quad 300 < f_1 \le 900 \end{Bmatrix} \quad (1)$$

This constraint results in the vowel space shown in figure 1. Every agent starts with an empty vowel space which is gradually filled by means of interaction.

This is where the imitation games, as described by De Boer (2000), come in. First an agent is chosen randomly to speak one of his vowels, this will be the *initiator*. If he has no vowel he will randomly create one. Noise is added to the initiator's vowel and an agent close to him, named the *imitator,* will be chosen to listen to this vowel. Algorithm `chooseAgent` is used to choose the second agent.

```
Algorithm chooseAgent(FirstAgent, SpeakingDistance,
TravData):
    Input: An agent FirstAgent, an integer
    SpeakingDistance and an array TravData containing
    numbers between 0 and 1 representing the traversability
    of blocks in the world.
    Output: An agent SecondAgent that is found near the
    FirstAgent, or FirstAgent if not successful in 50 tries.
⇒  TryCount = 0
⇒  While SecondAgent isnot FirstAgent AND TryCount < 50
    do
      • X ← FirstAgent.GetX() + (a random number at a
        Gaussian distance with mean 0 and variance
        SpeakingDistance) *
        TravData(FirstAgent.GetX(),FirstAgent.GetY())
        • Y ←FirstAgent.GetY()  + (a random number at a
          Gaussian distance with mean 0 and variance
```

[1] Formant values from a large number of vowels from different languages, gathered from widely available sources on the internet, all fall within this area.

```
          SpeakingDistance) *
          TravData(FirstAgent.GetX(), FirstAgent.GetY())
      •   If there is one or more agents at this
          location X,Y then
             ▪  SecondAgent is a random agent at this
                location
⇒  If TryCount == 50 then
      •   SecondAgent ← FirstAgent
⇒  return SecondAgent
```

Here, GetX() and GetY() return an agent's x or y location, while SpeakingDistance is a constant. TravData(x, y) returns the traversability of the tile (x,y). So, a location near the initiator is chosen depending on the traversability of the tile and the speaking distance of the agent. The imitator is selected from the inhabitants of this location. If no suitable imitator is found in 50 tries, the initiator will speak to itself.

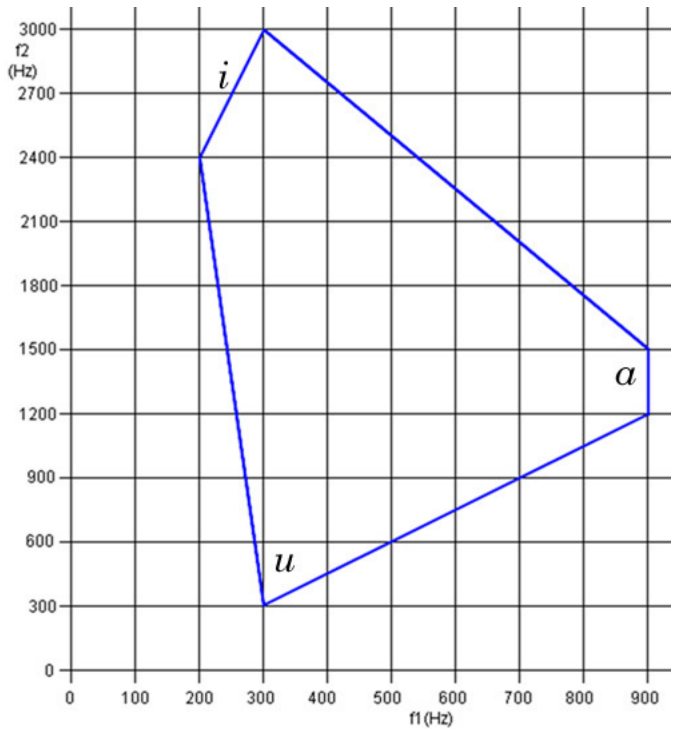

Figure 1: *The area within the thick line is the vowel space used in our experiments.*

The *imitator* then finds a vowel in his repertoire that is closest to the one just heard from the *initiator*. If he does not have any vowel yet, he will copy the vowel heard. Again adding noise, the vowel is returned. Now the *initiator* checks if this returned vowel is closest to the initial one. If it is, the success rate of both vowels is increased and the imitator's vowel is moved closer to the initiator's vowel (the amount is a percentage of the distance dependent on the parameter "Learning rate"). If it is not a success, then there are two possibilities. If the vowel's success rate is below 50%, the vowel is moved closer tot the initiator's vowel (just like when the imitation game is a success), but if the success rate is higher, it is probably a good vowel, so a new vowel is placed where the spoken vowel of the *initiator* was heard.

Apart from the imitation games there is also another way for the vowel space to change. This is called the vowel space update and occurs once every epoch. This is done as in De Boer (2000) to keep the agents vowel repertoires clean from bad vowels. Vowels that have been used >5 times, but have a success rate below 70% cannot be very good, so they are removed. Furthermore if two vowels become too similar (they are closer together than the noise level) they will be merged into one vowel. This is done by deleting the one with the lowest success rate and adding the use/success count to the more successful one. Finally there is a small chance (0.2%) that a new vowel is added randomly.

Next, let's consider the agent model and its functions. Every epoch each agent has a chance to die, move and reproduce, in that order. Whether they do depends on the conditions in their part of the world and how crowded it is.

An agent moves, or migrates, a small distance $\vec{m}$ in a random direction depending on the traversability t of their location. The distance between their old location $L_{x,y}$ and new location $M_{x,y}$ is a normally distributed function $\varphi(0,t)$ as formalized in (2).

$$\begin{aligned} \vec{m} &= \left(L_x, L_y, \sqrt{(X \sim \varphi)^2 + (Y \sim \varphi)^2}\right) \\ M_{x,y} &= L_{x,y} + \vec{m} \end{aligned} \qquad (2)$$

Death, on the other hand, is a looming threat that awaits all agents eventually. An agents chance of dying grows with his age a, and is higher in areas with high population/population capacity rates R, but lower in friendly environments with high habitability h, as in (3).

$$P_{death} = (P_{base} - c_1 h + c_2 R) \cdot c_3^{\,a} \qquad (3)$$

With the constants we used[2] this formula results in agents reaching ages of up to 50-60 epochs in healthy environments, and on average around 30 epochs in

[2] The specific constants can be found in the code available at http://science.uva.nl/~jkools

crowded circumstances. The probability that an agent produces a child P_{child} depends on the location's habitability *h* according to (4). Only agents between certain ages[3] can get children. Also, if the agent is the sole inhabitant of the tile, his chance to give birth is increased by two thirds.

$$P_{child} = c_4 \cdot (1.5 - h) \qquad (4)$$

After the agents have completed this phase of the current epoch the imitation games are played. For every agent in the world three imitation games are played, but the agents are chosen randomly, so not every agent plays the same amount of games every epoch. When the imitation games are done the vowel space of every agent gets updated. After that a new epoch starts and everything repeats.

The next section will further elaborate on the parameters that varied in our experiments. The experiments serve two basic goals: Finding values that result in the most realistic results, and study the effect of the parameters on the resulting vowel spaces when changed individually.

4. Results and Experiments

4.1. *Parameters*

In the following paragraph we will show what parameters are important to check, what tests we have run to find the best values for these parameters and what the resulting outcome is.

The most important parameters are the following:

- The noise level
- The speaking distance
- The learning rate
- The minimal learning rate
- The amount of people in the initial population
- The different maps

The noise level determines the amount of noise that is added to vowels that are spoken before they are heard. Noise represents inaccuracies in the signal due to imperfect hearing, pronunciation, other sounds, etc. The noisy vowel is computed as a normally distributed variable with the mean being the original vowel and the variance being half of the noise level. The noise level is also used to determine if two vowels should merge (as explained in section 3).

[3] Between ages 18 and 50 in this simulation

The speaking distance is a parameter which helps determine the distance an agent will travel to talk to someone. It is not solely responsible for the distance, because the traversability of the terrain is also a factor to consider (people living in tough terrain do not walk all day just for a chat). How the actual distance is computed is explained in Algorithm `chooseAgent`.

The learning rate λ determines how much an agent will change his own vowel towards the one he heard from another agent. The amount of change is dependent on the parameters chosen and the age of the agent. It is calculated every epoch for every agent by:

$$\lambda^* = \lambda + \alpha \cdot (\lambda_{min} - \lambda) \qquad (5)$$

Where α is a step size parameter which influences how fast the agent becomes less adaptive (in this system 0.5), and λ_{min} is the minimal amount of learning the agent will always do. λ is initialized with the global learning rate parameter. When a vowel is moved nearer to another vowel it will take λ as a percentage of the difference in location and move the vowel that much closer. This parameter represents the decrease in learning also visible in humans and will make new agents adapt their vowels very fast to their older peers, who will also be able to learn something from their younger peers, but much slower.

On the simulation's very first epoch, two initial groups of agents (Age: 20, Vocal space: empty) are placed in the world. They are distributed between two preselected areas on the left and the right half of the world. Another important issue is which maps to use for the habitability and traversability values. The map also determines the size of the world.

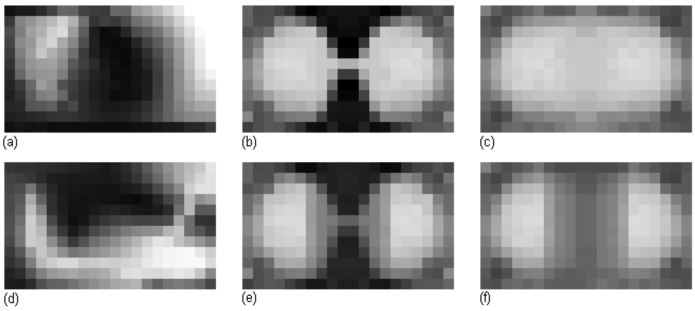

Figure 2: *The maps we used in our experiments. Maps a-c are habitability maps, d-f are the corresponding traversability maps. If not specified otherwise, we used the default maps (a,d).*

4.2. *Results*

We ran a series of tests with varying parameters and maps to see the effects on the resulting vowel spaces. The maps are shown in figure 2. We ran 3 simulations with varying noise level, using the following settings:

Noise Level	Speaking Dist.	Learning Rate	Min. Learn. Rate	Initial pop.
150/300/450	3	30	5	20

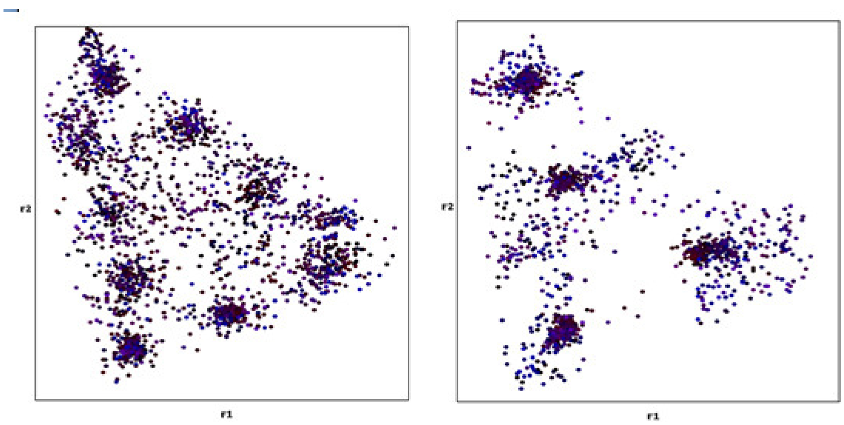

Figure 3: *Resulting vowel space after 500 epochs (± 600.000 imitation games). High noise level on the left, medium noise on the right.*

With the noise level at 150Hz the resulting vowel spaces had 8-9 vowels on average, and it was hard to differentiate between then because a lot of agents had vowels that were not shared. A small example can be seen in figure 3. With the noise level set to 450Hz the resulting vowel spaces tended to end with only two or three different vowel clusters. When using a noise level of 300Hz, the most realistic vowel spaces emerged with an average of four or five vowels after 500 epochs as can be seen in figure 3.

For the Speaking Distance parameter test we used the following settings:

Noise Level	Speaking Dist	Learning Rate	Min. Learn. Rate	Primordial Pop.
300	1/3/5	30	5	20

With speaking distance set to one, there appears to be very little resemblance between the average vowel spaces of the different hemispheres. Due to the short speaking distance agents probably had almost no contact with agents more than one grid block away, which caused the rift between the hemispheres.

With Speaking Distance set to five agents were not really hindered by the barriers we put up to semi-divide the two hemispheres. The result was similar vowel systems on both sides.

Finally with a Speaking Distance of three there was still some obvious influence from one hemisphere's agents to the other, but there were also still some differences in vowel systems.

For the learning rate parameters test we used the following settings:

Noise Level	Speaking Dist	Learning Rate	Min. Learn. Rate	Primordial Pop.
250	2	15/30/60	3/5/10	20

For this test the parameters noise level and Speaking Distance were already changed to the values the earlier tests showed are more realistic, but nonetheless the only thing that was derived from this test is that visually it is not easy to spot the influence the learning rate has on the vowel systems. More tests could be run and a mathematical comparison between vowel systems from different tests could clarify its influence. See the paragraph 'Conclusion and future work' for more details.

For the amount of primordial population parameters test we used the following settings:

Noise Level	Speaking Dist	Learning Rate	Min. Learn. Rate	Primordial Pop.
250	2	30	5	5/20/40

The only difference this made was how fast the population grew and how fast the vowel systems found equilibrium.

Another series of experiments were done with different maps using the following settings:

Noise Level	Speaking Dist	Learning Rate	Min. Learn. Rate	Primordial Pop.
250	2	30	5	20

In this experiment we chose not to include statistical data comparisons, although the implementation would allow for it. Time constraints compelled us to keep to visual comparison of the data. Statistical evidence is therefore not included in this paper. However there were some obvious trends visible in the resulting images and tendencies and speculation will be addressed in the conclusion.

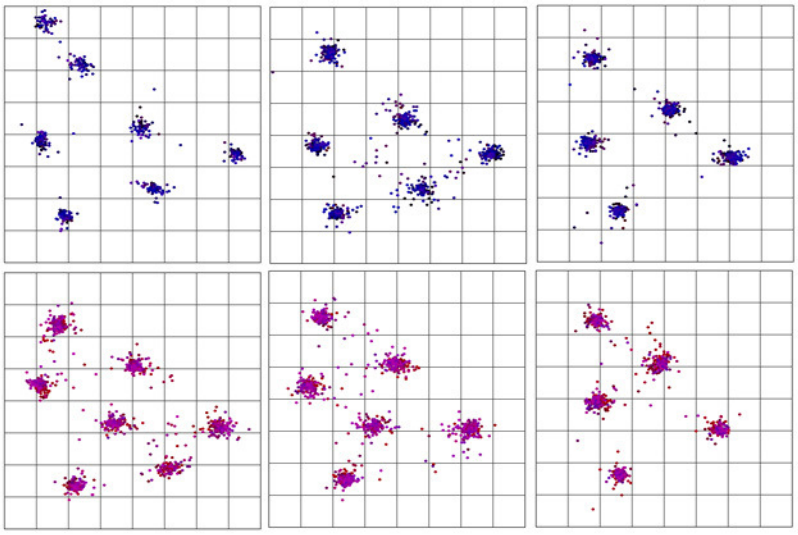

Figure 4: *The top and bottom charts show 2 populations in isolated areas after 150 (left), 250 (middle) and 350 (right) epochs. Slowly, a common language emerges*

The population size nicely adheres to what might be expected in a situation of

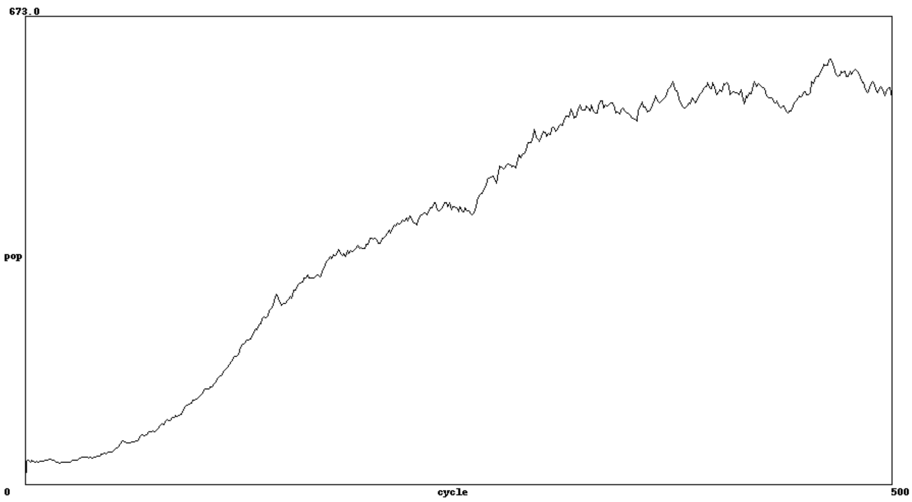

Figure 5: *Population over time in a representative simulation*

limited growth, as illustrated by figure 5. The trend resembles an s-curve or a logistic function, known to population ecologists as the Verhulst-Pearl function

(Kingsland, 1985/1995). This observation strengthens our claim that our model is an ecological one.

5. Conclusion and future work

A typical midway simulation with colored dots for vowels depending on the location of the corresponding agent (see figure 6) shows a number of interesting trends. The global languages tend to converge towards more and more similar vowels, even with relatively high degrees of isolation. Local, and gradual variation in the exact location of vowels, however, persists.

In our simulation, Trudgill's statement that high degrees of language contact may lead to larger phoneme inventories seems only to hold for a limited amount of time after two 'languages' first come into contact. In the end, the number of vowel clusters will return to a value dependent on noise level and 'speaking distance' only. We did not find any evidence that smaller communities develop smaller vowel inventories. In fact, the contrary appeared to be true. This might however be a result of a world model that is too static.

What really appears clearly in our data is that globally (i.e. between separate populations), differences are found in the number of vowels and actual presence of a cluster in a rough location, while local variations consist in variations of the location of a vowel. Such local 'dialects' occur both with sharp and with gradual transitions. A tentative but interesting conclusion is that the degree of global differences seems to be mostly dependent on isolation and not on distance, and vice versa for local variance.

The amount of tests we were able to run, as well as the amount of analysis we could perform on the resulting data were severely limited due to time restrictions. For this reason alone, more research is warranted. But we also dare to think that we have provided several clues for future investigation.

While we observed some interesting effects of ecological parameters on the evolution of vowel systems, there is much that could be improved on for realism. We suspect that changing environments and fluctuating degrees of isolation have had serious repercussions on historical evolution of language. Famines, wars, economic collapses etc. cause a historically important loss of infrastructure and interregional communication that is not represented in our model at all. Additionally, migrations should perhaps be modeled to occur in groups and be directed to more desirable locations, instead of randomly as we have done. Agents could have a tendency to physically group with speakers of the same language, too.

Another related possibility would be to adapt the model in such a way that one could enforce isolation for a certain duration, so that separated populations have time to mature and develop more distinct vowel systems before coming in contact with one another.

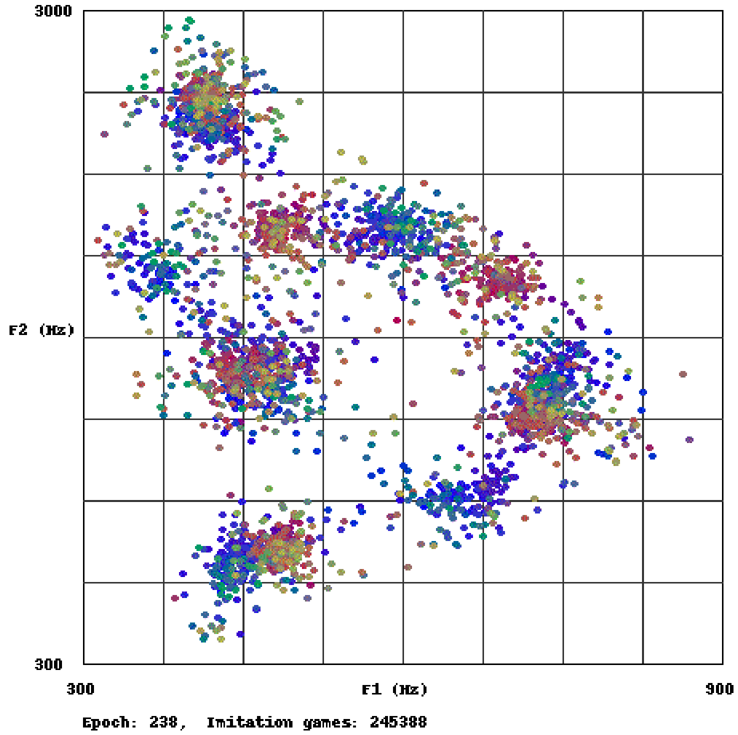

Figure 6: *Local and global differences.*

We would also like to point out that the experiment comparing simulations set on different maps are possible not as accurate as they should have been, because we did not take into account that adding an obstacle in the world not only increases isolation, but also decreases the world's maximum population and thus effective community size. This means that this experiment does not exclusively compare the effect of differing degrees of isolation. This might be addressed by normalizing the habitability values before use so they all add up to a constant value.

Moreover, the data we have collected could use a more sophisticated, more statistically based analysis. We were able to observe notable mechanisms and typical behavior, but without proper investigation of many repeated simulations, the significance of the observed phenomena is unclear.

References

Barr, D. (2004). Establishing conventional communication systems: Is common knowledge necessary? *Cognitive Science* 28 (2004) 937–962

De Boer, B. (1998). Emergence of Sound Systems through Self-Organization. In Han La Poutre and Jaap van den Herik (eds.) *NAIC '98 Proceedings*, Amsterdam 37-46

De Boer, B. (2000). Self-organization in vowel systems. *Journal of Phonetics* 27, 1-25

De Boer, B. & Vogt, P. (1999). Emergence of speech sounds in changing populations. In *ECAL'99 Advances in artificial life* (Floreano, D., Nicoud, J.-D. & Mondada, F., eds.), Lecture Notes in Artificial Intelligence 1674, 664-673

Fought, J. et al (2004). Sonority and Climate in a World Sample of Languages: Findings and Prospects. *Cross-Cultural Research* 38, 27-51.

Holman, E. et al. (2007). On the relation between structural diversity and geographical distance among languages: Observations and computer simulations. *Linguistic Typology* 11, 393-421

Kingsland, S. (1995). *Modeling Nature: Episodes in the History of Population Ecology*. 2nd edition, University of Chicago Press, 98-126.

Nettle, D. (1998). Using Social Impact Theory to simulate language change. *Lingua* 108, pp. 95-117

Nettle, D. (2007). Language and genes: A new perspective on the origins of human cultural diversity. *PNAS* 104(26), 10755-10756

Sugiura, H. et al. (2006). Sound transmission in the habitats of Japanese macaques and its possible effect on population differences in coo calls. *Behaviour* 143, 993-1012

Steels, L. (1997). *Synthesising the Origins of Language and Meaning Using Co-Evolution, Self-Organisation and Level Formation*. In: Evolution of Human Language, Hurford, J (ed.), Edinburgh Univ. Press, Edinburgh

Trudgill, P. (2004). Linguistic and social typology: The Austronesian migrations and phoneme inventories. *Linguistic Typology* 8, 305–320.

Wittgenstein, L. (1973). *Philosophical Investigations* (3rd ed.). New York: Prentice Hall, §7