Generating Vowel Systems in a Population of Agents

Bart de Boer
VUB AI Laboratory
Pleinlaan 2
1050 BRUSSEL, BELGIUM
`bartb@arti.vub.ac.be`

Abstract:

In the sound systems of human languages remarkable universals are found. These universals can be explained by innate mechanisms, or by their function in human speech. This paper presents a functional explanation of certain universals of vowel systems using Alife-techniques.

It is based on language-like interactions between members of a population of individual agents. The agents start out empty, but have a “drive” to make (vowel) sounds to each other and to imitate these sounds. Through repeated “imitation games” and through modifications of their own sound system, based on the outcome of the imitation games, the agents reach coherence. The sound systems that arise have properties that are similar to those of human vowel systems.

Keywords: language origins, cultural evolution, phonological universals

1. Introduction

Alife techniques have been used to aid many fields of science, such as biology, ethology and psychology. In this paper alife techniques are applied to the field of linguistics. They are used to provide a functional explanation for a number of properties of the sound systems humans use for communication.

The human vocal tract is capable of producing an amazing number of different speech sounds. The UCLA Phonological Segment Inventory Database (UPSID, described in [14]) recognises 921 different speech sounds—652 consonants and 269 vowels—found in 451 languages. Still, the number of speech sounds (phonemes) used by any individual language is quite limited. According to Maddieson [14] the average lies between 20 and 37. The maximum number of phonemes of any language in the UPSID is 141 for the Khoisan language !Xũ, the minimum is 11 for the East-Papuan language Rotokas and the South-American language Múra-Pirahã [7,14]. Also, a number of remarkable universals can be found in the sound systems of languages. Some sounds, such as [a], [p] or [m] are much more frequent than others, such as [ɤ][*], [t̪ɬ’][†] or [ɴ][‡]. Also the structure of sound systems is not random. If a language uses a certain voiced consonant (e.g. [d]), it will usually have the corresponding unvoiced consonant (e.g. [t]) as well. With vowels it is the same: a language will rather have a system consisting of, for example [i, e, a, o, u] than of [y, æ, a, ə, ɔ]. Apparently languages have symmetrical systems with sounds that are spread evenly, rather than random systems.

The traditional explanation for these phenomena is that humans use a number of *distinctive features* [8] for building up the system of speech sounds they use. A distinctive feature is a (usually binary) feature of speech sounds that can cause a difference in meaning between two words. An example is the voicing of consonants in English. The difference between the words “bed” and “bet” is that the last consonants of these words are voiced and voiceless, respectively. This causes a change in meaning, and therefore [voice] is considered a distinctive feature in English.

It is generally assumed [4,8] that these distinctive features are innate. According to this theory, all humans are born with the same set of distinctive features. When learning their mother tongue, they choose the set of distinctive features that this language uses, as well as the settings of the features for the individual sounds in the language. With the right set of distinctive features, the theory is quite able to predict the regularities that are found. It can also predict the sequence in which these sounds are learned. Furthermore it is able to explain that some sounds are rarer than other sounds by assuming that certain features and certain values for features are more *marked* than others.

Still, there are a number of fundamental problems with this theory. First of all, most proposed feature sets are not able to account for all the sounds that are found in the world’s languages. Ladefoged and Maddieson [9, ch. 11] write: “The great variety of data that we have presented shows that the construction of an adequate theory of universal features is much more complex than hitherto thought.” However, even if a feature-based theory would have sufficient features to account for all possible speech sounds, it would still not be able to account for the subtle, but important differences between sounds in different languages and dialects that every speaker of such a language uses and recognises. An example is the difference between English *coo*, French *cou* (neck), German *Kuh* (cow) and Dutch *koe* (cow), all of which would be described as a high back voiceless consonant, followed by a high back rounded vowel. Also the distinctive feature theory does not explain where distinctive features come from in the first place. There is a danger of circularity in deducing features from observations of regu-

---

[*] Mid back unrounded vowel, found e.g. in Vietnamese.

[†] Dental ejective lateral affricate, found e.g in Navajo.

[‡] Uvular nasal, found e.g. in Japanese.

larities in language and then proposing these features as explanations of these regularities. As Lindblom et al. write: “...postulating segments and features as primitive categories of linguistic theory should be rejected...” [11, p. 187]

Another approach to explaining the structure of sound systems of human language is a functional one. Sound systems are explained by assuming that they are based on minimal articulatory and cognitive effort and maximal perceptual contrast. Especially in the area of vowel systems, this approach has been particularly successful. Liljencrants and Lindblom [10], Lindblom [13], Carré and Mrayati [3] and Boë et al. [2] showed, using computer simulations, that vowel systems can be explained by a maximisation of the acoustical contrast, while at the same time minimising the articulatory gestures that are needed. Observations of consonant systems of a wide range of languages [12] have obtained evidence that the same mechanisms are operating there. However no computer simulations to investigate these observations have been done yet because of the more complex articulatory and perceptual characteristics of consonants (for a simulation of simple syllables see Lindblom et al. [11]).

In the computer simulations of Liljencrants and Lindblom [10], Lindblom [13], and of Boë et al. [2], it is assumed that one can assign an energy function to vowel systems. This function has higher energy for systems with their vowels closer together and for systems that need more articulatory gestures. One then minimises this energy function for a given number of vowels.

Unfortunately, these computer simulations do not provide us with a mechanism that explains how this process takes place in human language. The only way in which vowel systems can change in human languages is by the interactions between—and the actions of—the users of the language. As no speaker has control over the language as a whole, this process must be considered an emergent property of language use. We can observe that a minimisation of the “energy” of vowel systems does take place in human language. However, we do not yet know by which actions of the individual speakers this minimisation is caused.

An attempt to model changing vowel systems in a population of communicating agents has been made by Berrah et al. [1], Glotin [5] and Glotin and Laboissiere [6]. They use an approach that combines learning and a technique which the authors call a “pseudo-genetic algorithm” [5, sect 4.4]. Their agents communicate using randomly initialised vowel systems with a fixed number of vowels. They communicate, change their vowel systems in a way that depends on the difference between their own vowels and those of the agent they spoke with. Then they calculate a fitness function that depends on the articulatory efforts they made. After a while a new generation of agents is calculated by procreating the fittest individuals, using selection and crossover. Their system produces vowel systems that look like human vowel systems. However, their system is not quite comparable to human speech communities, because of a number of assumptions they have made. First of all, their agents do not really learn a vowel system from scratch. The number of vowels, as well as the initial position of the vowels is coded in the agents’ genes. The authors use this mechanism to efficiently explore the space of possible vowel systems. However, by precoding the number and position of the vowels, the authors disregard the process by which children acquire speech sounds from scratch. This process, however, is probably an important factor in determining the possible and stable shapes of sound systems. Simplifying this process away might therefore be an oversimplification. Their assumption also prevents the agents from adding or removing sounds from their vowel systems. Also, calculating a new population of agents requires that the *internal* states of the agents be crossed with each other. Glotin and Laboissiere[6] are aware that this is not realistic, but they nevertheless use it for exploring the possible vowel systems. A strong point of their is that it uses a very good speech synthesis model. Unfortunately, the computational complexity of this model makes it unsuitable for long simulations with lots of agents.

This paper proposes a system in which a population of agents learns vowel systems by observing and trying to imitate each other’s speech sounds. The individual agents produce and perceive sounds under constraints that are meant to be similar to human ones. They manipulate their own sound systems in order to maximise the success in imitating the other agents. The system is based on Steels’ ideas about the origins of language [16].

In the next two sections the architecture of the agents (section 2) and their interactions (section 3) are described. Also their relation to Steels’ theory is described in somewhat more detail. In section 4 the results of a number of experiments are presented and in section 5 these results are discussed and related to other work in this area. Possible future work with the system is also suggested.

2. The Agents

Each agent in the system has its own list of vowels. This list is initially empty, and will be filled as the agents engage in interactions with other agents. The vowels are represented by the three main parameters that are used for describing vowels: tongue position, tongue height and lip rounding. The three parameters can have any value between zero and one. For tongue position, zero means front, and one means back. For tongue height, zero means low and one means high, and for lip rounding, zero means unrounded and one

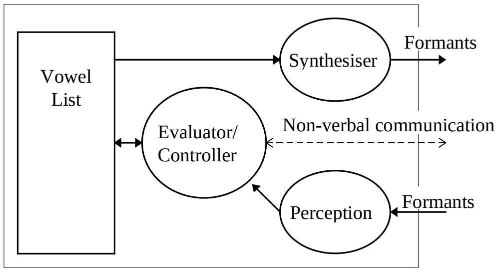

Figure 1: Agent architecture

means rounded. The agents are able to produce any "simple" vowel. The system is completely language-independent. No bias towards the vowel system of any language is present in the agents.

The vowels that are present in the agents are produced by a synthesiser and are recognised by a perception unit. A special control unit regulates the actions of the agents and the evaluation of vowels. The internal architecture of an agent is illustrated in figure 1.

The synthesiser is a simple articulatory synthesiser that is based on a second order interpolation of a number of artificially synthesised vowels. The input of the synthesiser consists of the three articulatory parameters and the output consists of the frequencies of the first four formants of the vowel associated with this particular articulation. The basic data for the formants have been taken from Vallée [18, pp. 162–164].

In the experiments a certain amount of noise has been added to the formant frequencies that are produced by the agents. The adding of noise consist of multiplying the formant frequencies by:

1) $1 \pm U(a)$,

in which $U(a)$ is a random variable uniformly distributed over $[-a,a]$, where a varies for different experiments. The addition of noise makes the games more natural. Similarly, in human speech it cannot be expected that sounds will always be produced and perceived accurately. The noise also makes it impossible for the agents to copy each other's phonemes perfectly, thereby forcing them to create sound systems in which the phonemes are not too close together, as well as opening the possibility of change and language evolution.

For each phoneme an agent creates, it generates the formants of an ideal articulation of this sound. This ideal articulation is called the *prototype vector* and it is stored together with the articulatory description of the phoneme. Every time an agent hears a sound, it calculates the distance between the prototype vectors of all the phonemes it knows and the formants of the sound it just heard. The phoneme with the prototype vector that is closest to the sound that was heard is considered to be the recognised phoneme. This whole process could in principle be implemented using neural networks, thereby increasing the biological plausibility.

The distance measure that is used to compare phonemes is of crucial importance to the form of the vowel systems that will be generated by the agents. In order to get natural vowel systems, and in order to be able to compare the results of the experiments with those of at least one other group, a distance measure that has been adapted from Boë et al. [2] was used in a slightly modified form. The distance measure takes into account that the human auditory system distinguishes vowels by their formant frequencies, lower formants having a greater influence, that it does not distinguish well between formants that are very close together and that it works in an essentially logarithmic manner.

For the distance function two weights need to be calculated:

$$2)\ w_1 = \frac{c - (F_3 - F_2)}{c}$$

$$3)\ w_2 = \frac{(F_4 - F_3) - (F_3 - F_2)}{F_4 - F_2}$$

Where w_1 and w_2 are the weights, F_1–F_4 are the formants in Bark[§] and c is a critical distance, set to 3.5 Bark.

The weighted sum of F_2, F_3 and F_4 which we will call F_2' will now be calculated as follows:

$$4)\ F_2' = \begin{cases} F_2, & \text{if } F_3 - F_2 > c \\ \dfrac{(2-w_1)F_2 + w_1F_3}{2}, & \text{if } F_3 - F_2 \le c \wedge F_4\text{-}F_2 > c \\ \dfrac{w_2F_2 + (2-w_2)F_3}{2}\text{-1}, & \text{if } F_4 - F_2 \le c \wedge F_3 - F_2 < F_4 - F_3 \\ \dfrac{(2+w_2)F_3 - w_2F_4}{2}\text{-1}, & \text{if } F_4 - F_2 \le c \wedge F_3 - F_2 \ge F_4 - F_3 \end{cases}$$

The values of F_1 and F_2' for a number of vowels are shown in figure 2. We can see from this figure that the distribution of the vowels through the acoustic space is quite natural. However, as it is a two-dimensional projection of an essentially three-dimensional space, not all distances between all phonemes can be represented accurately. This is especially the case with the distinction rounded-unrounded. Unfortunately this is difficult to avoid in any system.

The distance between two vowels, a and b can now be calculated using a weighted Euclidean distance:

$$5)\ d = \sqrt{\left(F_1^a - F_1^b\right)^2 + \lambda\left(F_2^{a\,\prime} - F_2^{b\,\prime}\right)^2}$$

This again, in accordance with the work of Boë et al. [2]. The value of the parameter λ is chosen to be 0.5 for all experiments that will be described.

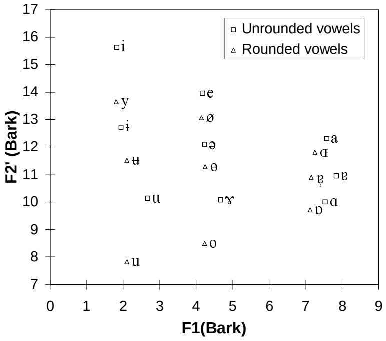

Figure 2: Vowels in F1-F2' space

[§] The Bark scale is a logarithmic frequency scale, which is based on human perception. If the distance in Bark between different frequencies is equal, this distance is perceived as equal by the human ear.

With the articulator function and the perception function that have been described in this section, the agents can produce and perceive speech sounds in a way that is sufficiently human. This means that the results that are generated with these systems can at least to some extent be compared to the results of research into human sound systems.

3. The Imitation Game

The experiments presented in this work are concerned with the emergence of a coherent and useful phonology in a population of initially empty agents. In order to investigate how this can happen, the agents engage in exchanges of sounds, so-called imitation games[**], the goal of which is to learn each other's speech sounds. If necessary, speech sounds are invented, in order to get the imitation games started, and also in order to introduce more possible sounds in the population.

The structure of the imitation games is based on Steels' ideas about the origins of language [16]. He considers language a cultural phenomenon that maintains coherence through self-organisation. Language is learnt by actively making hypotheses about the form of the language and by testing these in linguistic interactions, which he calls *language games*. Complexity arises through (cultural) evolution and co-evolution of linguistic structures. In his view, there is no need for innate mechanisms (a Language Acquisition Device) to explain the origin and the acquisition of language. According to Steels, the above mentioned mechanisms are able to explain both the historical origin as well as the acquisition of language.

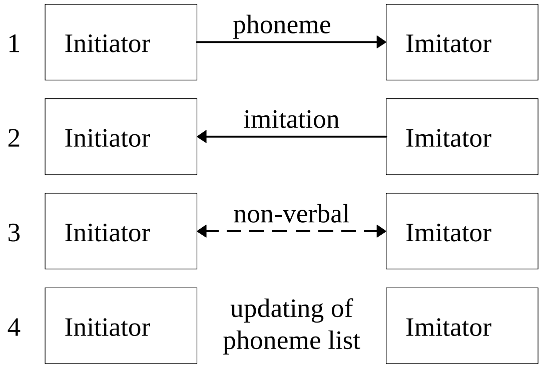

Figure 3: The imitation game

The basic rules of the imitation game that is played by two agents are very simple. Two agents are randomly selected from the population of agents. One of the agents, which we will call the *initiator*, selects one of its phonemes and says this to the other agent. The other agent, which we will call the *imitator*, interprets this sound in terms of its phonemes, and then produces the phoneme it thinks it has recognised. The other agent listens to this imitation, and also interprets it in terms of its phonemes. If the phoneme it recognises is the same as the one it just said, the imitation game is considered to be successful. If it is not equal, the game is unsuccessful. There follows a non-verbal communication, in which the imitator gets to know if its imitation was correct or not. The whole process is illustrated in figure 3.

For each phoneme in the phoneme list of both the initiator and the imitator, the number of times it is used and the number of times it was successful are kept. Every time a phoneme is uttered in a language game, its use count is increased. Every time it was successfully imitated, its success score is increased. If it was not successfully imitated, nothing happens to the success score. The quality of a phoneme is this success score divided by the number of times it was used.

Depending on the course of the language game, the initiator and imitator can change their repertoire of phonemes. The phoneme lists of the agents are initially empty, so at first the initiator has to choose a random articulator position, and use this as its first phoneme. If the phoneme list of the initiator is also empty, it tries to make an imitation of the sound it just heard, by saying sounds to itself, and using a hill-climbing heuristic in order to approach the sound it just heard. It then adds this imitation to its phoneme list.

If the initiator already has a list of phonemes, it picks one of these at random and utters it, or creates a new phoneme with a very small probability. If the imitator already has a list of phonemes, it picks the closest match (as described above) and uses this as imitation. If the imitation was successful, the imitator tries to shift the phoneme it said a bit closer to the sound it just heard, again using a hill-climbing heuristic. This in order to make the phoneme even better. If the imitation was not successful, and if the quality of the phoneme was low, the phoneme is also shifted, in order to try to improve the imitation. However, if the quality of the phoneme was high, the phoneme is not shifted, because its high score indicates that it is probably a good imitation of another phoneme. Therefore, we create a new phoneme (using again a hill-climbing heuristic) that sounds similar to the sound that had to be imitated.

Two other processes are going on. Firstly, phonemes that have low quality for a long time are removed from the phoneme list. With a certain probability, the initiator's phonemes that have a quality score that is below a certain threshold are removed. Secondly, phonemes that are too close together, are merged. Phonemes are considered too close together if they are so close together that they can be confused through the noise that is added to the formant frequencies. The phonemes are fused by taking the articulator position of the phoneme with the highest score as the new articulator position. The success and use counts of the new phoneme are calculated by adding the success and use counts of the old phonemes.

All the steps of the language game as have been described above, are both necessary within the system and could in principle be performed by humans. Without some of the steps outlined above the system does not function as well. If phonemes are not shifted closer together, they stay too far

---

[**] Not to be confused with Suzuki and Kaneko's imitation games [17], which are completely different.

apart, they get confused and the number of phonemes can not be increased. If bad phonemes are not removed, they degrade the per performance. If similar phonemes are not merged, they tend to get confused and degrade the performance as well.

4. The Experiments

In this section we will present a number of experiments that have been conducted with the language games and the agents described above. The goal of the experiments was to investigate whether it was possible to develop a successful sound system in a population of initially empty (*tabula rasa*) agents, and what form this sound system would take under different conditions of noise, and for different population sizes. The experiments that were conducted consisted of a predetermined number of iterated imitation games in a homogeneous population of agents.

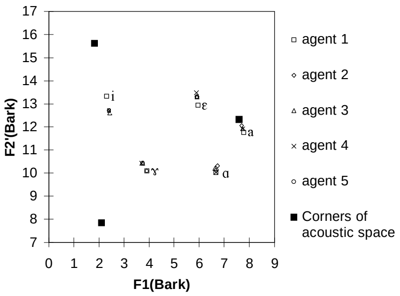

Figure 4: Sound systems of five agents

The results of the first experiment are presented in figure 4. It shows the sound systems that were developed in a population of five agents after 1000 imitation games were played. The acoustic realisation of the phonemes was subject to 10% noise (a=0.1 in equation 1). It is clear from the clusters in the figure that the five agents share the same phonemes. The corresponding phonemes for the different agents are close together, while the phonemes within one agent are far apart. This is optimal for a sound system that is meant for communicating different sounds between agents. It can also be observed that the phonemes are spread through the available acoustic space in a way that is reminiscent of the way vowels of human languages are spread through acoustic space, even though the vowel system that was arrived at: [i,ɛ,a,ɑ,ɤ] probably does not appear in any human language.

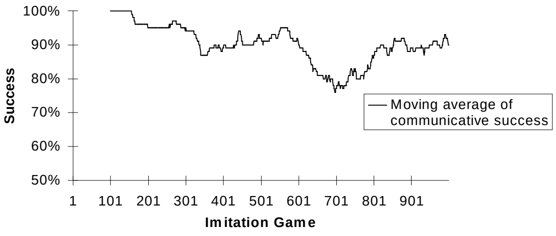

Figure 5: Success of agents with 10% noise

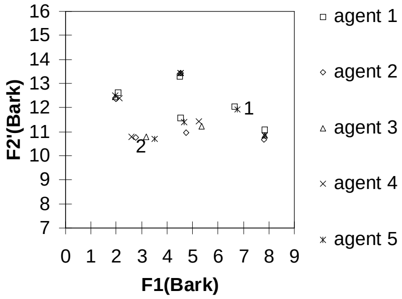

Figure 6: Sound systems in transition

The imitation success of the agents, as illustrated in figure 5 is constantly between 70% and 100%. The success starts at the 100% level in the beginning of the experiment, because at that time the agents only have one phoneme each and confusion is not possible. As soon as the agents start creating new phonemes, however, the success score drops, because phonemes are being confused. After a while, the agents succeed in making copies of the phonemes, and the success score returns to near 90%. The results shown are of the same run that resulted in the sound systems of figure 4, and are representative for the runs that are normally generated by the simulation.

If the amount of noise in the formant frequencies is increased, the area over which phonemes are “smeared” in acoustical space will also increase, and the number of phonemes that can coexist without confusion in the agents’ vowel systems will decrease. We therefore expect smaller vowel systems and more variation within the realisation of individual vowels. This is illustrated in figure 7, which represents a typical sound system[††] of the agents after 5000 imitation games. This number is higher than in the previous experiment, as the agents apparently take longer to develop multiple phonemes if there is more noise. This is logical because newly generated phonemes have a higher chance of interfering with existing phonemes. Note that the realisation of a formant can be shifted as much as 2 Bark down or 1.5 Bark up by 30% noise, so any phoneme can be realised in a significant part of the acoustic space.

When one agent starts using a new phoneme, this can be

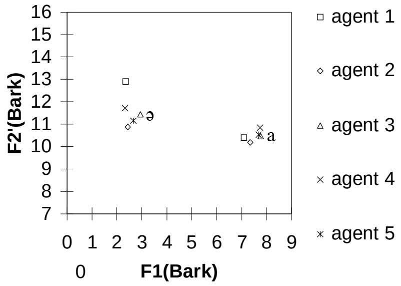

Figure 7: Sound systems with 30% noise

[††] The vowel system, consisting of /a/ and /ə/, coincidentally is similar to the vowel system of Oubykh, a West-Caucasian language.

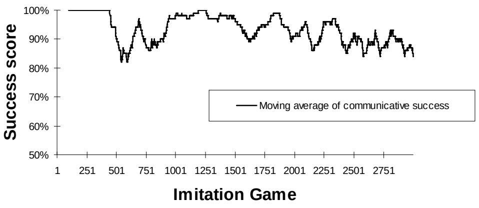

Figure 8: score of twelve agent experiment

adopted by the other agents in the population. First one agent invents a new phoneme at random. When it uses this phoneme in an imitation game, the imitation game is bound to fail. However, if a language game fails in an agent whose phonemes otherwise have a good quality score, a new phoneme will be generated that is like the phoneme that was just heard, as has been described in section 3. If this new phoneme does not interfere with the phonemes that are already present, it will be accepted by the population of agents, and will become successful as well. This process can be observed in figure 6. Here one of the agents, agent 1, seems not to have the phoneme marked with 2, that is otherwise shared by all other agents, but it does seem to have an extra phoneme, marked with 1, which it shares with one other agent, agent 5. Actually these two facts are unrelated. The phoneme marked with 2 is a phoneme that has been created by another agent than agent 1, some time before the moment at which figure 6 was made. Agent 1 has not yet had the opportunity to make a successful copy of this phoneme. The phoneme marked with 1, however, has been recently created by agent 1. The only agent that has had the opportunity to make a successful imitation of this phoneme is agent 5. It can be observed that new phonemes are created in gaps between existing phonemes in the acoustical space. Phonemes that are created outside such gaps will quickly be merged with the existing phonemes, or will interfere with existing phonemes, and be removed from the sound systems, because their quality scores will remain too low.

A last observation that will be made is what happens when the agent population is made larger. For this, experiments with 12[‡‡] agent have been conducted. The experiments were run for 3000 cycles and had 10% noise on the acoustic space. The success score of a typical experiment is shown in figure 8. We can see that the score stays above 80%, although it does seem to be decreasing a bit over time. This is undoubtedly due to the increasing number of phonemes in the population of agents. But there does not seem to be a big differences between figure 8 figure 5, which showed the success score of a population of five agents.

The phonemes of the twelve agents of this experiment after 3000 imitation games are shown figure 9. We can observe that there are four to six clusters of phonemes. Three of these are compact and unambiguous. Another cluster, which can be found between 4 and 6 Bark on the F1 axis and around 11 Bark on the F2' scale is also unambiguous, but much more dispersed. This cluster is quite close to another diffuse cluster, which can be found between 2 and 3 Bark on the F1 scale and 9 and 12 Bark on the F2' scale. This cluster could also be considered as two separate clusters, as some agents (for example agents 6, 10 and 11) have two phonemes near the densest points in this cluster, whereas other agents (3 and 7) have only one phoneme in the centre of this cluster. This could indicate that the cluster represents a phoneme in the process of splitting. More research is needed, however, in order to make this clear.

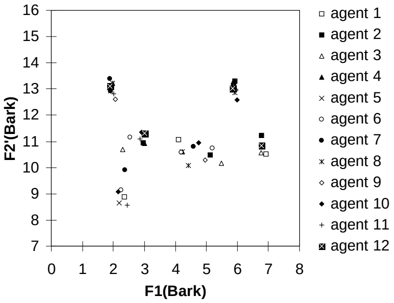

Figure 9: Phonemes of twelve agent experiment

In any case, it does not seem that the increase in the number of agents influences the success of the imitation very much. Of course, there is bound to be some influence, as an agent will play games with more other agents, so that its phonemes get shifted in more different directions and therefore converge less quickly to a common point. The fact that the number of agents does not greatly influence the success of imitation is promising, as for realistic experiments the number of agents has to be much larger than the five or twelve used in the present experiments. Fortunately, the simulations are not computationally intensive, so it should be possible to increase the number of agents to about a hundred times the number of agents that were used in the experiments presented here.

5. Conclusions and Future Work

The results of the experiments show that it is possible to generate realistic vowel systems in a distributed population of agents that try to imitate each other under constraints. No innate features that determine the form of the vowel systems were needed, nor does it appear to be necessary for the agents to inspect each other's internal state.

The experiments have also shown that the generated vowel systems are not static. They are constantly changing as a result of the invention of new phonemes, the shifting of existing phonemes due to noisy production, and the deletion and merging of phonemes. This is a phenomenon that is also found in natural language, albeit in a less extreme way than in our system. The agents in our system are probably not conservative enough. However, the observed changes seem to indicate that sound change in human language can be explained by the mechanisms that have been proposed in this paper. Previous attempts to explain vowel systems on functional grounds [1,2,3,5,6,10,13,18] have always resulted

---

[‡‡] The number of 12 agents was chosen because this is the number Berrah et al.[1] use in their experiments.

in static systems and could therefore not account for language change. Especially the fact that the agents actively imitate each other seems to be important, as this makes it possible for newly invented phonemes to become successful. In Berrah et al,'s system [1,5,6], for example, introducing a new phoneme would lead to its immediate rejection, as no matching phonemes would be found in the other agents. This would lead to a lower fitness of the agent that invented the phoneme, and thus both the agent as well as the phoneme would eventually disappear from the population. Only simultaneous invention of a new phoneme in multiple agents would make it possible for a new phoneme to be accepted.

Apparently the shape of vowel systems can be explained by considering them as the result of a self-organising process consisting of interactions (imitation games) in a population of independent agents that each change their local phonological knowledge according to the outcome of these interactions. Of course, this situation is a gross simplification of the way humans learn the sounds of their language. However, it does give an indication that we do not have to resort directly to innate mechanisms for explaining phonological phenomena.

These observations agree quite well with the observations that Steels [15,16] has made in trying to apply the ideas of self-organisation to other parts of language, notably lexicon formation. It appears that for more parts of language it is not necessary to invoke innate mechanisms, but that they can be explained by self-organising processes.

In the immediate future the system can be extended in several ways. Because of its dynamic nature, it can easily be extended to accommodate a changing population of agents. One could, for example, add and remove agents from the population, and see how this influences the dynamics. These agents can be made to differ in "age" so that older agents are more conservative than younger ones. An interesting question would then be whether this conservatism would stabilise the population. One could also investigate how the influx of new, empty agents would influence the stability of the population.

Another modification of the system would be to investigate more complex sounds. Investigating only vowels is easy, but also quite unrealistic if one wants to learn things about human language. One possible extension would be to investigate consonant-vowel syllables, as have already been investigated in a static system by Lindblom et al. [11]. For this, one would have to add constraints on articulation, as well as constraints on perception.

Considering (the phonology of) language as an emergent phenomenon of the interaction of language users allows us to use the tools of the study of artificial life and dynamic systems for the study of language. It opens up a new perspective that can make it easier to explain a number of phenomena that can nowadays only be explained by postulating innate mechanisms.

6. Acknowledgements

The work presented here has been done in part at the AI-lab of the Vrije Universiteit Brussel in Brussels, Belgium and in part at the Sony Computer science laboratory in Paris, France. It forms part of ongoing research project into the origins of language. It was financed by the Belgian federal government FKFO project on emergent functionality (FKFO contract no. G.0014.95) and the IUAP 'Construct' project (no. 20). I thank Luc Steels, Edwin de Jong and Dolores Cañamero for valuable suggestions on– and discussion of the ideas that are fundamental to the work and Hervé Glotin for personally explaining his Alife-work on phonology.

7. References

1. Berrah, Ahmed-Reda, Hervé Glotin, Rafael Laboissière, Pierre Bessière and Louis-Jean Boë,(1996) From Form to Formation of Phonetic Structures: An evolutionary computing perspective, in: Terry Fogarty and Gilles Venturini, eds. *ICML '96 workshop on Evolutionary Computing and Machine Learning*, Bari, pp. 23–29
2. Boë, Louis-Jean, Jean-Luc Schwartz and Nathalie Vallée(1995), The Prediction of Vowel Systems: perceptual Contrast and Stability, in: Eric Keller (ed.), *Fundamentals of Speech Synthesis and Speech Recognition*, John Wiley, pp. 185–213
3. Carré, René and Mohamad Mrayati, Vowel transitions, vowel systems, and the Distinctive Region Model, in: C. Sorin et al. (eds.) *Levels in Speech Communication: Relations and Interactions*, Elevier pp. 73–89
4. Chomsky, Noam and Morris Halle (1968) *The sound pattern of English*, MIT Press, Cambridge, Mass.
5. Glotin, Hervé, (1995) *La Vie Artificielle d'une société de robots parlants: émergence et changement du code phonétique*. DEA sciences cognitives-Institut National Polytechnique de Grenoble
6. Glotin, Hervé, Rafael Laboissière(1996) Emergence du code phonétique dans une societe de robots parlants. *Actes de la Conférence de Rochebrune 1996 : du Collectif au social*, Ecole Nationale Supérieure des Télécommunications - Paris
7. Grimes, Barbara F. (ed.)(1996) *Ethnologue: Languages of the World, 13th edition*, SIL.
8. Jakobson, Roman and Morris Halle (1956) *Fundamentals of Language*, the Hague: Mouton & Co.
9. Ladefoged, Peter and Ian Maddieson (1996) *The Sounds of the World's Languages*, Blackwell.
10. Liljencrants, L. and Björn Lindblom (1972) Numerical simulations of vowel quality systems: The role of perceptual contrast, *Language* **48** pp. 839–862.

11. Lindblom, Björn, Peter MacNeilage and Michael Studdert-Kennedy(1984), Self-organizing processes and the explanation of language universals, in: Brian Butterworth, Bernard Comrie and Östen Dahl (eds.) *Explanations for language universals*, Walter de Gruyter & Co. pp. 181–203
12. Lindblom, Björn and Ian Maddieson (1988), Phonetic Universals in Consonant Systems, in: Hyman, Larry M. and Charles N. Li (eds.) *Language, Speech and Mind*, pp. 62–78.
13. Lindblom, Björn(1996), Systemic constraints and adaptive change in the formation of sound structure, in: James R. Hurford, Michael Studdert-Kennedy and Chris Knight, *Evolution of Language: Social and Cognitive Bases for the Emergence of Phonology and Syntax*
14. Maddieson, Ian,(1984) *Patterns of sounds*, Cambridge University Press.
15. Steels, Luc (1996) The Spontaneous Self-organization of an Adaptive Language, in: S. Muggleton (ed.) *Machine Intelligence* **15**.
16. Steels, Luc (1997) Synthesising the origins of language and meaning using co-evolution, self-organisation and level formation, in: J. Hurford (ed.) *Evolution of Human Language*, Edinburgh: Edinburgh University Press.
17. Suzuki, Junji, Kunihiko Kaneko,(1994) Imitation Games, in: *Physica D* 75 pp. 328–342
18. Vallée, Nathalie, (1994) *Systèmes vocaliques: de la typologie aux prédictions*, Thèse préparée au sein de l'Institut de la Communication Parlée (Grenoble-URA C.N.R.S. no 368)