Self Organisation in Vowel Systems through Imitation

Bart de Boer
Artificial Intelligence Laboratory
Vrije Universiteit Brussel
Pleinlaan 2
1050 Brussels, Belgium
bartb@arti.vub.ac.be

Abstract

In this paper an artificial life approach to the explanation of the shape of vowel systems is presented. A population of artificial agents (small independent computer programs) that are each able to produce and perceive vowels in a human-like way, engages in imitation games. In these imitation games one agent makes a sound and another agent tries to imitate it. Both agents use their list of phonemes for analysing and producing the sounds. Depending on the outcome of the language game, the agents update their phoneme lists, using only local information. It is demonstrated that in this way vowel systems that look remarkably like human vowel systems emerge. The process is insensitive to factors such as noise level, initial conditions and number of agents. It is argued that this could be a useful way of explaining the universal characteristics of human vowel systems.

1 Introduction

The world's languages contain a surprising number of different sounds. In the most recent version (1996) of the UCLA Phonological Segment Inventory Database (Maddieson, 1984) 921 different segments are recognised: 652 consonants and 269 vowels and diphthongs. However, in any particular language, only a limited number of these sounds are used. According to Maddieson, most languages have between 20 and 37 phonemes. The minimum number is 11 for the East-Papuan language Rotokas and the South-American language Múra-Pirahã, and the maximum number is 141 for the Khoisan language !Xũ (Grimes, 1996; Maddieson, 1984). The typical number of phonemes, according to Maddieson, lies between 20 and 37.

Also certain regularities are found in the sound systems of the languages of the world. If we concentrate on vowels, we find that certain vowels, such as [u] [a] and [i] appear almost universally, while other vowels, such as [ɤ], [ɵ] and [œ] are much rarer. The structure of vowel systems also shows great regularities. They tend to be symmetric. If a language has a front vowel of a certain height, for example [e], it is likely to have the back vowel of the same height [o] as well, although the two vowels will usually differ in rounding. Languages tend to prefer vowel systems in which the acoustic difference between the vowels is as big as possible. For this reason, vowel systems with just [i], [a] and [u] appear more frequently than systems with just [e], [a] and [o], or with [i], [a] and [ɯ].

Traditionally, these phenomena have been explained through distinctive feature theory (Chomsky and Halle, 1968; Jakobson and Halle, 1956). The preferred shapes of vowel systems are explained by innate distinctive features, and by their markedness. Features split up the continuous articulatory space. As there are only a limited number of features, vowel systems will only contain a limited number of vowels, and because some features are more marked than

others, some vowels and some combinations of vowels will appear less often than others.

Unfortunately, this theory does not address the question where the distinctive features come from, nor how a discrete set of phonemes came to be used for communication in the first place. It still remains to be explained what the reason for the presence of discrete phonemes and distinctive features is. As (Lindblom et al, 1984, page 187) wrote: "...postulating segments and features as primitive universal categories of linguistic theory should be rejected...". Also, distinctive feature theory does not explain why certain minor differences in pronunciation are replicated so closely by speakers of a certain dialect. For example, there is a difference between English *do*, French *doux* (soft), German *du* (thou) and Dutch *doe* (do) that is perceived and recognised by speakers of these languages, even though all these words are described as an anterior and coronal voiced consonant followed by a high, back, rounded vowel.

In order to explain the shapes of sound systems of the world's language without having to resort to innate features, a number of functional explanations have been put forward. For vowel systems different researchers (Liljencrants and Lindblom, 1972; Boë et al, 1995) (among others) have given elaborate computational models. These models predict the qualities of the vowels in vowel systems with a given number of vowels by calculating a maximum for the acoustic distances between the vowels. Carré and Mrayati (Carré and Mrayati, 1995) have also used computer models for predicting vowel systems, based on articulatory as well as acoustic constraints. Furthermore Stevens (Stevens, 1989) has developed a theory that explains the shape of sound systems through non-linear characteristics of the human vocal tract and auditory system.

All these theories, although to some extent controversial, provide good explanations of *why* vowel systems are the way they are. They maximise acoustic contrast and minimise articulatory effort. However, the theories do not provide a mechanism to explain *how* these characteristics obtain in a population of language users. They all consider language as an independent system that somehow optimises a number of constraints. They do not take into account that languages are used by individual speakers that are each quite capable of learning and using any vowel system. Somehow, the interactions between these speakers cause the functional constraints as mentioned above to *emerge*.

The emergence of constraints on vowel systems through the interactions of individual agents has already been studied by Hervé Glotin and others (Berrah et al, 1996; Glotin, 1995; Glotin and Laboissière, 1996). Unfortunately, their work contains a number of unrealistic assumptions about the way in which sound systems are transferred from generation to generation. In their system agents use vowels to make sounds to each other. Vowels are shifted to make them more similar to the ones from the other agents. After a while the agents that have least shifted their vowels create offspring that replace agents that have much shifted their vowels. The initial position of the new agents' vowels is determined from the initial position of the vowels of the agents' parents. The number of vowels in each agent is fixed, which makes it less realistic. Another disadvantage of this system is that it does not model the way in which new agents acquire their phonemes (they already have a set from birth). Also the pseudo-genetic component obfuscates the actual processes (the language-like interactions between the agents) that shape the vowel systems.

The work that will be presented here is based on the theory of Steels (Steels, 1997b). Steels considers language to be a phenomenon that is the result of self-organisation and cultural evolution in a population of language users. Knowledge of the language is transferred through linguistic interactions that Steels calls *language games*. Individuals actively form and test hypotheses about the language in these games. Innovation is introduced by random variations and errors in imitation. Selection pressure for more efficient and effective communication causes certain variations to be preferred over others. Self-organisation ensures that coherence

is maintained. According to Steels, this mechanism can both explain the origin of language, as well as the acquisition of language by a single individual. Steels has mainly tested his theory in the area of lexicon formation (Steels, 1995) and semantics (Steels, 1997a). In the present paper the theory is applied to the field of phonology.

In the next section some more background on self-organisation is given. In section 3 the simulation that was used for investigating the theory is described. In section 4 the experiments that have been done are described, and in section 5 some conclusions are drawn.

2 Self-Organisation

Quite often spontaneous order can emerge in systems that are not controlled centrally. An example of this is the construction of a honeycomb. No single bee (not even the queen) has control over the building behaviour of the whole swarm. Still, a very regular pattern of hexagons emerges. This happens because bees start to build cells at a certain distance from other bees that build cells. After a while they will encounter the neighbouring cells. Thus a pattern of hexagons emerges. Other examples of the outcomes of self-organising processes are termites' nests, sand dunes and the formation of paths.

All self-organising systems have a large number of constituent parts that interact on a small scale. Order emerges on a large scale. This order is obtained from initial random behaviour of the constituent parts through positive feedback processes. These feedback processes cause the constituent parts to settle collectively in a certain state, once an accidental majority of them happens to be in that state. The field of "artificial life" is concerned with the investigation of self-organising processes that are inspired by living systems through computer simulations.

The approach that is followed in this paper and that was introduced by Steels (Steels, 1997b) is an artificial life approach. Language is a self-organising process. It exists in a community of speakers, and persists through the interactions of the speakers. No individual has central control over the language and no individual speaker is necessary for the persistence of the language. They are born and they die and still the language remains more or less continuous over time and throughout a population.

The computer simulations that are presented here model linguistic interactions in an artificial life way. This means that the emergence of order in a population of *agents* (small computer programs that can operate autonomously) is studied. The agents are able to produce and perceive speech sounds in an approximately human way, they have only local knowledge (i.e. about their own speech sounds) and engage in local interactions with only one other agent at a time. It will be shown that phenomena that are also found in human vowel systems emerge.

3 The System

The agents in the simulation are equipped with a speech synthesiser, a speech perception system and a list of phonemes. It should be stressed that the agents are not restricted to any particular natural language. The speech synthesiser is capable of generating all simple vowels. It takes as input the three major vowel features: tongue position, tongue height and lip rounding. Its output consists of the first four formant frequencies of the vowel that would be generated by the specified articulator positions. The production model is based on an interpolation of artificially generated formant patterns of 18 different vowels taken from (Vallée, 1994, page 162–164). A certain amount of noise is added to the formant frequencies: they are shifted up or down a random percentage. The speech perception system is based on a model developed by (Boë et al, 1995) who based their system on a substantial amount of observations of human perception of speech. In this model low frequency formants are considered to be more salient than high frequency formants and if two formants are close together, they are perceived approximately as one formant with an intermediate frequency. These characteristics ensure that the agents perceive formant patterns as similar if humans would also perceive them as

similar. Both the speech synthesiser and the speech perception system are described in more detail in (de Boer, 1997)

The agents start with an empty phoneme list: they know no phonemes at all. They learn their phonemes through interactions with each other. The shape of the resulting vowel system will be determined for a small part by coincidence and for the largest part by self-organisation under acoustical and articulatory constraints.

The interactions between the robots are called *imitation games.* For each imitation game, two agents are chosen randomly from the population. One agent will initiate the game and is called the *initiator*, the other one is called the *imitator.* The initiator randomly chooses a phoneme from its phoneme list, or creates a new phoneme randomly if its phoneme list is empty. It then generates the corresponding sound (the formant pattern). The imitator listens to this sound, and analyses it in terms of its own phonemes. It tries to find among its own phonemes the phoneme whose formant pattern most closely resembles the sound it just heard. If its phoneme list is empty, it generates a new phoneme. The imitator then generates the sound that corresponds to its best matching phoneme. The initiator listens to this sound and also analyses it in terms of its own phonemes. It then checks whether the phoneme that most closely matches the sound it just heard is the same as the phoneme it originally said. If they are the same, the imitation game is successful. If they are not the same, the game is unsuccessful.

Depending on the outcome of the language game, the imitator undertakes a number of actions. If the language game was successful, it shifts the phoneme it said in such a way that it will sound more like the sound it just heard. This is done by making slight changes to the phoneme and by checking whether these increase the resemblance. The change that most increases the resemblance is kept. This procedure is called *hill climbing* in artificial intelligence, and it is comparable to making sounds to oneself in order to learn how to pronounce a given sound.

If the imitation game was unsuccessful, the agent can either create a new phoneme or shift the old phoneme, depending on whether the phoneme it used for imitating the sound had previously been successful or not. The success of a phoneme is calculated by keeping track of the number of times a phoneme was used in an imitation game (both by initiator and by imitator) and the number of times the imitation game in which the phoneme was used was successful. The ratio between these numbers is used as a measure of success of the phoneme.

If the phoneme has been unsuccessful, it is shifted to resemble more closely the sound that was heard. If it has been successful, however, it is assumed that the failure of the imitation game was caused by the fact that two phonemes are confused. The initiator has two phonemes that are matched by only one phoneme in the imitator. Hence the imitator creates a new phoneme that closely resembles the sound that was heard. This usually resolves the confusion.

Two more processes are taking place in the agents. First of all, an agent's phonemes that resemble each other too closely are merged. Two phonemes are merged by keeping the most successful one and by throwing away the least successful one. The successfulness of the new phoneme is calculated by adding the use- and success counts of the original phonemes. Secondly, phonemes that have a use/success ratio that is too low, are discarded. This causes bad phonemes to disappear eventually from the phoneme repertoire of the agents.

4 The Experiments

A large number of experiments have been done with the system described above. Experiments have been performed with varying numbers of agents and under various conditions of noise. The system consistently produced populations of agents that were able to imitate each other successfully with the vowel systems that emerged. These vowel systems showed remarkable similarities with vowel systems found in human languages. A typical example of the vowel systems of a population of 20 agents, with

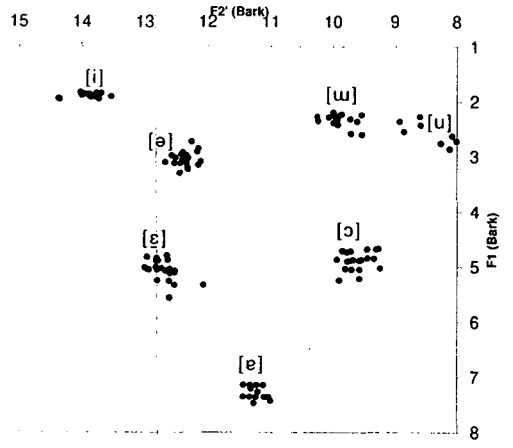

Figure 1: Acoustic representation of the vowel systems of a population of twenty agents after 2000 imitation games.

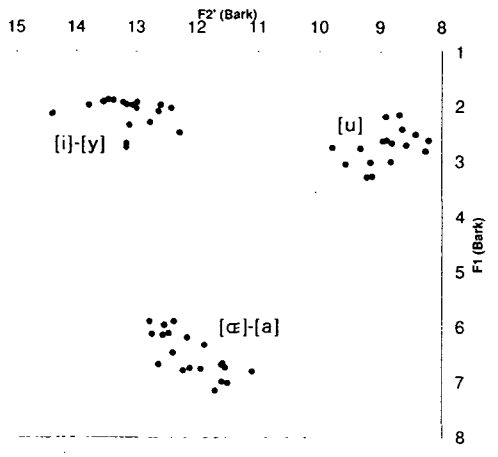

Figure 2: Vowel system of a population of twenty agents that communicate with 20% noise. The three vowel system has been stable for over 1000 language games.

a maximum of 10% noise on the formant frequencies, is given in figure 1. This figure is an acoustic representation of all the phonemes of all the agents in the population. In this figure a number of clear clusters can be discerned. Almost all the phonemes of the agents tend to appear in one of the seven clusters. In addition, almost all agents have a phoneme in the six largest clusters. Only in the small cluster in the lower left corner, representing the [u], few agents have a phoneme. This is probably because this phoneme has recently been created, and not all agents have been able to make an imitation, yet.

The vowel systems that emerge from the imitation games are not static. They are constantly changing as new phonemes are formed and old phonemes shift through the available acoustic space. This process is illustrated in figure 3, the result from a different simulation with the same starting conditions (twenty agents and 10% noise) but with slightly different random influences. In this figure we see two vowel systems that are snapshots of one population of agents, taken 1000 language games apart. We see that clusters move through the acoustic space and that clusters tend to compact. However, a certain distance appears to be kept between the clusters. Also the clusters seem to remain spread over a certain area; they do not reduce to points completely.

Under various conditions of noise, systems with different numbers of clusters emerge. If the amount of noise is increased, systems with fewer clusters are generated (an example is given in figure 2). However, the success of the imitation games stays approximately the same. Also the number of agents does not seem to matter much. Experiments with five to forty agents have all resulted in stable systems. Furthermore, the systems seem to be resistant to population change. If old agents are removed at random, and new empty agents are added at random, the vowel systems remain stable. The empty agents will rapidly learn the existing phonemes by imitating more experienced agents. If the inflow of new agents becomes too large, however, instability arises.

5 Conclusions and Discussion

The first conclusion that can be drawn from the work presented above is that stable sound systems do emerge in a population of artificial agents that play imitation games. Moreover, these systems have discrete clusters in a continuous acoustic space that could be described by (discrete) distinctive features, even though there was no predetermined partition of

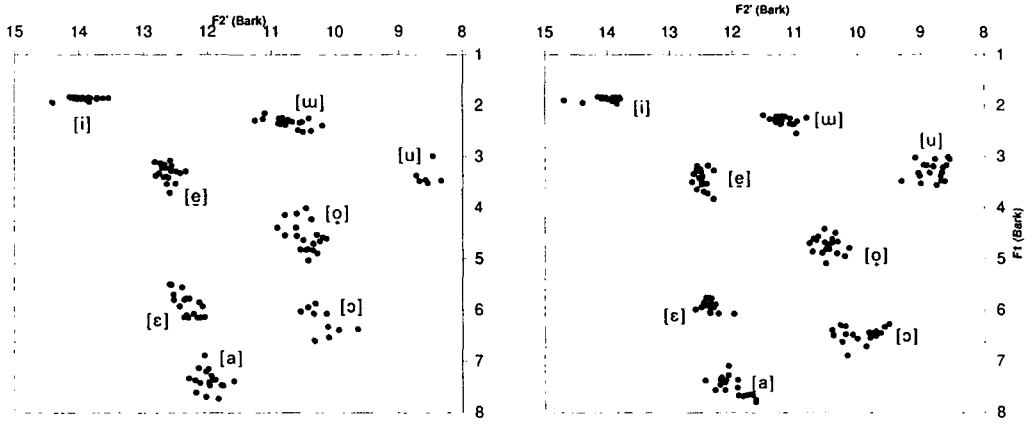

Figure 3: Dynamics of artificial vowel systems: the system at the right obtained from the one at the left after 1000 imitation games.

the acoustic space. In addition, it can be concluded that the shapes of these emergent systems show remarkable similarities to the shapes of the most frequent vowel systems found in human languages.

It remains to be seen to what extent these results are applicable to human language. It must be admitted that the language capabilities of the simulated agents are a gross simplification of the language capabilities of humans. However, the agents are entirely biologically plausible. This means that they can do nothing that humans could not do *in principle*. Also, they provide a possible mechanism by which functional constraints on vowel system that were first researched with computers by Liljencrants and Lindblom (Liljencrants and Lindblom, 1972) can emerge from interacting language users.

The system described here provides a model for predicting certain universals of vowel systems. It does not have to postulate innate distinctive features or innate mechanisms other than the fact that agents communicate with a limited set of sounds. Also the system shows individual variation and language change that do not decrease the agents' ability to analyse each other's sounds. A remarkable property of the simulations that have been presented is that both the learning of speech sounds as well as sound change can be generated by the same mechanism.

The author thinks that these results justify considering phonological processes in language as self-organising processes. By taking this point of view it also becomes possible to bridge the gap between language as behaviour of individuals and language as a system by using computational models.

6 Acknowledgements

The work was done at the AI-laboratory of the Vrije Universiteit Brussel in Brussels, Belgium and at the Sony Computer Science Laboratory in Paris, France. It is part of ongoing research into the origins of language. It was financed by the Belgian federal government FKFO emergent functionality project (FKFO contract no. G.0014.95) and the IUAP 'Construct' project (no. 20). I thank Luc Steels for valuable suggestions on- and discussion of the fundamental ideas of the work.

References

Ahmed-Reda Berrah, Hervé Glotin, Rafael Laboissière and Louis-Jean Boë 1996. From Form to Formation of Phonetic Structures: An evolutionary computing perspective In: Terry Fogarty and Gilles Venturini, editors *Proceedings of the ICML '96 workshop on Evolutionary Computing and Machine Learning* pages 23–29

Louis-Jean Boë, Jean-Luc Schwartz and Nathalie Vallée 1995. The Prediction of Vowel Systems: perceptual Contrast and Stability In: Eric Keller, editor, *Fundamentals of Speech Synthesis and Speech Recognition*, John Wiley, pp. 185–213

René Carré and Mohammed Mrayati 1995. Vowel transitions, vowel systems and the Distinctive Region Model In: C. Sorin et al. (editors) *Levels in Speech Communication: Relations and Interactions*, Elsevier, pages 73–89

Noam Chomsky and Morris Halle 1968. *The sound pattern of English*, Cambridge, MS: MIT Press.

Bart de Boer 1997. *A second report on emergent phonology*, AI-memo 97-04, AI-lab, Vrije Universiteit Brussel

Hervé Glotin 1995. *La Vie Artificielle d'une société de robots parlants: émergence et changement du code phonétique*, DEA sciences cognitives-Institut National Polytechnique de Grenoble

Hervé Glotin and Rafael Laboissiére 1996. Emergence du code phonétique dans une société de robots parlants. *Actes de la conférence de Rochebrune 1996: du Colectif au social*, Ecole Nationale Supérieure des Télécommunications-Paris.

Barbara F. Grimes, editor 1996. *Ethnologue: Languages of the World, 13^{th} edition*, Summer Institute of Linguistics

Roman Jakobson and Morris Halle 1956. *Fundamentals of Language*, the Hague: Mouton & Co

L. Liljencrants and Björn Lindblom 1972. Numerical simulations of vowel quality systems: The role of perceptual contrast, *Language* **48**, pages 839–862

Björn Lindblom, Peter MacNeilage and Michael Studdert-Kennedy 1984. Self-organizing processes and the explanation of language universals In Brian Butterworth, Bernard Comrie, Östen Dahl, editors *Explanations for language universals*, Berlin, Walter de Gruyter & Co

Ian Maddieson, 1984. *Patterns of Sounds*, Cambridge: Cambridge University Press

Luc Steels 1995. A Self-Organizing Spatial Vocabulary, *Artificial Life* **2**, Cambridge (MS): MIT Press, pages 319–332

Luc Steels 1997a. Constructing and Sharing Perceptual Distinctions. In: Maarten van Someren and G. Widmer, editors: *Proceedings of the ECML*, Berlin: Springer Verlag. To appear.

Luc Steels 1997b. Synthesising the origins of language using co-evolution, self-organisation and level formation. In J. Hurford et al. editor, *Evolution of Human Language*, Edinburgh: Edinburgh University Press. To appear.

Kenneth N. Stevens 1989. On the quantal nature of speech. *Journal of Phonetics* **17**, 1, pages 3–45

Nathalie Vallée 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. n^o 368)