4 From holistic to discrete speech sounds: the blind snowflake-maker hypothesis

Pierre-Yves Oudeyer

4.1 The speech code

Sound is a medium used by humans to carry information when they speak to each other. The existence of this kind of medium is a prerequisite for language. It is organized into a code, called speech, which provides a repertoire of forms that is shared in each language community and that allows its users to encode content information. This code is mainly conventional, and thus intuitively requires coordinated interaction and communication to be established. How, then, might a speech code be formed prior to the existence of communication and of language-like interaction patterns?

Moreover, the human speech code is characterized by several properties which we have to explain. Here are some of them:

Property 1: Discreteness and systematic re-use. Speech sounds are phonemically coded as opposed to holistically coded. This implies two aspects: (i) in each language, the continuum of possible sounds is broken into discrete units; (ii) these units are systematically re-used to build higher-level structures of sounds, like syllables.

For example, in articulatory phonology (see Studdert-Kennedy and Goldstein 2003; Studdert-Kennedy, Chapter 3), a vocalization is viewed as multiple tracks in which gestures are performed in parallel (the set of tracks is called the gestural score). A gesture is the combination of several articulators (e.g. the jaw, the tongue) operating to execute a constriction somewhere in the mouth. The constriction is defined by the place of obstruction of the air as well as the manner. Given a subset of organs, the space of possible places of constrictions is a continuum (for example, the vowel continua from low to high, executed by the tongue body),

though each language uses only a few places to perform gestures. This is what we call discreteness.[1] Furthermore, gestures and their combinations, which may be called 'phonemes', are systematically re-used in the gestural scores that specify the syllables of each language. Some researchers call this 'phonemic coding'.

Property 2: Universal tendencies. Reoccurring units of vocalization systems are characterized by universal tendencies. For example, our vocal tract makes it possible to produce hundreds of different vowels. However, each particular vowel system typically uses only three, four, five, or six vowels, and extremely rarely more than twelve (Schwartz et al. 1997a). Moreover, some vowels appear much more often than others. For example, most languages contain the vowels [a], [i] and [u] (87 per cent of languages) while other vowels are very rare, such as [y], [œ] and [ɯ] (5 per cent of languages). Also, there are structural regularities: for example, if a language contains a front unrounded vowel at a certain height, for example the /ɛ/ in *bet*, it will also usually contain the back rounded vowel at the same height, which would be the /ɔ/ in *hawk* in this case.

Property 3: Sharing. The speakers of a particular language use the same phonemes and they categorize speech sounds in the same manner. However, they do not necessarily pronounce each of them exactly the same way.

Property 4: Diversity. At the same time, each language categorizes speech sounds in its own way, and sometimes does this very differently from other languages. For example, Japanese speakers categorize the <l> of *lead* and the <r> of *read* as identical.

This chapter addresses the question of how a speech code with these properties might have formed from non-speech prior to the ability to have linguistic interactions. The mechanism I present is based on a low-level model of sensory-motor interactions. I show that the integration of certain very simple and non-language-specific neural devices allows a population of agents to build a speech code that has the properties

[1] The fact that the audible speech stream is continuous and produced by a mixture of articulatory movements is not incompatible with 'discreteness': 'discreteness' applies to the command level, which specifies articulatory targets in time, which are then sequentially and continuously reached by the articulators under the control of a low-level motor controller.

outlined above. The original aspect is that this presupposes neither a functional pressure for communication, nor the ability to have coordinated social interactions (agents do not play language or imitation games). It relies on the self-organizing properties of a generic coupling between perception and production both within agents and in the interactions between agents.

4.2 Existing approaches

4.2.1 *The reductionist approach*

One approach is 'reductionist': it tries to reduce properties of the speech system to properties of some of its parts. In other words, this approach hopes to find a physiological or neural structure, the characteristics of which are sufficient to deduce the properties of speech.

For example, cognitive innatism (Chomsky and Halle 1968; Pinker and Bloom 1990) defends the idea that the brain features a neural device specific to language (the Language Acquisition Device) which 'knows' at birth the properties of speech sounds. This 'knowledge' is supposed to be pre-programmed in the genome. A limit of this approach is that its defenders have remained rather imprecise on what it means for a brain to know innately the properties of language. In other words, this hypothesis is not naturalized. Also, no precise account of the origins of these innate devices has ever been provided.

Other researchers focus on the vocal tract physics as well as on the cochlea electro-mechanics. For example, they claim that the categories that appear in speech systems reflect the non-linearities of the mapping from motor commands to percepts. Phonemes would correspond to articulatory configurations for which small changes lead to small changes in the produced sound. Stevens (1972) defends this idea. There is no doubt that the morphoperceptual apparatus influences the shape of speech sounds. However, this reductionist approach has straightforward weaknesses. For example, it does not explain the large diversity of speech systems in the world's languages (Maddieson 1984). Also, there are many experiments which show that the zones of non-linearity of perception in some languages are not compatible with those of certain other languages (e.g. Japanese <l> and <r>, as noted above).

Another example of this approach is that of Studdert-Kennedy and Goldstein (2003); see also Studdert-Kennedy (Chapter 3) for the origins of discreteness, or 'particulate speech' in his terms. Studdert-Kennedy and Goldstein remark that the vocal apparatus is physiologically composed of discrete independent articulators, such as the jaw, the tongue, the lips, the velum, etc. This implies that there is some discrete re-use in complex utterances due to the independent articulators that move. I completely agree with this remark. However, other aspects of discreteness are not accounted for. Indeed, as Studdert-Kennedy and Goldstein (2003) note, once you have chosen to use a given set of articulators, there remains the problems of how the continuous space of possible constrictions or timings between gestures is discretized. Goldstein (2003) proposes a solution to this question that I will review later in the chapter (since it is not reductionist but is a mixture of self-organization and functionalism).

One has to note that this 'reductionist' approach proposes answers concerning the presence of properties (1) and (2) of the speech code, but addresses neither the diversity of speech sounds nor the fact that they are shared across communities of agents. This approach also does not provide answers to the chicken-and-egg problem of the formation of a code, although this was, of course, not its goal.

4.2.2 *The functionalist approach*

The functionalist approach attempts to explain the properties of speech sounds by relating them to their function. Basically, it answers the 'why' question by saying 'the system has property N because it helps to achieve function F'. It answers the 'how' question by saying 'systems with property N were formed through Darwinian evolution (genetic or cultural) under the pressure to achieve function F'. This approach could also be called 'adaptationist':[2] systems with property N were designed for ('ad') their current utility ('apt'). Note that typically, functionalist explanations take into account constraints due to brain structure, perceptual, and vocal systems.

Typically, in the case of the four properties of speech sounds we are interested in, this function is 'communication'. This means that the sounds

[2] I use the term adaptationism in its general form: the adaptation may be achieved through genetic or cultural evolution.

of a speech code should be perceptually distinct enough so that they are not confused and communication can take place. The constraints which are involved typically include a cost of production, which evaluates how much energy is to be spent to produce the sounds. So, under this view, speech sounds are a reservoir of forms which are quasi-optimal in terms of perceptual distinctiveness and economy of energy.

For example, Lindblom (1992) shows that if we search for vowel systems which are a good compromise between perceptual distinctiveness and energy cost of articulation, then we find the most frequent vowel systems in human languages. Lindblom also showed similar results concerning the re-use of units to form syllables.

Operational scenarios describing how Darwinian cultural evolution formed these systems have also been described. For example, de Boer (2001a) builds a computer simulation showing how cultural evolution might have worked, through processes of imitation among agents. In this simulation, the same mechanism explains both the acquisition of vowels and its formation; this mechanism is imitation. As a consequence, he also proposes an answer to the question: 'How are vowel systems acquired by speakers?'

Note that de Boer's model does not deal with questions concerning discreteness (which is built in) and systematic re-use (indeed, his agents produce only simple static vowel sounds, and systematic re-use is a property of complex dynamic sounds). However, this model is very interesting since it shows a process of formation of a convention, i.e. a vowel system, within a population of agents. This really adds value to the work of Lindblom, for example, since it provides a mechanism for (implicit) optimization which Lindblom merely assumed.

However, the imitation game that agents play is quite complex and requires a lot of assumptions about the capabilities of agents. Each of the agents maintains a repertoire of prototypes, associations between a motor program and its acoustic image. In a round of the game, one agent, called the speaker, chooses an item from its repertoire, and utters it to another agent, called the hearer. Then the hearer searches its repertoire for the closest prototype to the speaker's sound, and produces it (he imitates). Then the speaker categorizes the utterance of the hearer and checks if the closest prototype in its repertoire is the one he used to produce the initial sound. Then he tells the hearer whether it was 'good' or 'bad'. Each item in the repertoires has a score, used to promote items which lead to successful

imitations and prune the others. In the case of bad imitations, depending on the scores of the prototype used by the hearer, either this prototype is modified so as to better match the sound of the speaker, or a new prototype is created, as close as possible to the sound of the speaker.

So to perform this kind of imitation game, a lot of computational/cognitive power is needed. First, agents need to be able to play a game, involving successive turn-taking and asymmetric role-changing. Second, they must voluntarily attempt to copy the sound production of others, and evaluate this copy. Finally, when they are speakers, they need to recognize that they are being imitated intentionally, and give feedback/reinforcement to the hearer about the (lack of) success. The hearer must understand the feedback, which says that from the point of view of the other, he did or did not manage to imitate successfully.

The level of complexity needed to form speech-sound systems in this model is characteristic of a society of agents which already possesses some complex ways of interacting socially, including a system of communication (which allows them for example to know who is the speaker and who is the hearer, and which signal means 'good' and which signal means 'bad'). The imitation game is itself a system of conventions (the rules of the game), and agents communicate while playing it. It requires the transfer of information from one agent to another, and so requires that this information be carried by shared 'forms'. So it presupposes that there is already a shared system of forms. The vowel systems that appear do not really appear 'from scratch'. This does not mean that de Boer's model is flawed, but rather that it deals with the subsequent evolution of language (or more precisely, with the evolution of speech sounds) rather than with language origins (in other words it deals with the formation of *languages—les langues* in French—rather than with the formation of *language—le langage*). Indeed, de Boer presents interesting results about sound change, provoked by stochasticity and learning by successive generations of agents. But the model does not address the bootstrapping question: how did the first shared repertoire of forms appear, in a society with no communication and language-like interaction patterns? In particular, the question of why agents imitate each other in the context of de Boer's model (this is programmed in) remains open.

Another model in the same spirit was proposed by Browman and Goldstein (2000) and Goldstein (2003). This model is very interesting since it is the only one I know of, apart from the work presented in the

present chapter, which tries to approach the question of the origins of the discretization of the continuum of gestures (they call this 'emergence of discrete gestures').[3] It involves a simulation in which two agents could produce two gestures, each parametrized by a constriction parameter taken in a continuous one-dimensional space (this space is typically the space of possible places of constrictions, or the continuous temporal interval between two gestures). Agents interacted following the rules of the 'attunement game'. In one round of the game, both agents produced their two gestures, using for each of them a parameter taken in the continuum with a certain probability. This probability was uniform for both gestures at the beginning of the simulation: this meant that a whole continuum of parameters was used. Next, agents recovered the parameter of the other agent's first gesture, and compared it to the parameter they used themselves. If this matched, then two things occurred: the probability of using this parameter for the first gesture was increased, and the probability of using the same value for the second gesture is decreased. This simulated the idea that agents are attempting to produce both of their gestures differently (so that they are contrasted and can be differentiated), and the idea that they try to produce each gesture in a similar fashion to the corresponding gesture of the other agent (so that a convention is established). At the end of the simulations, agents converged to a state in which they used only one value for each gesture, so the space was discretized, and these pairs of values were the same for the two agents in the same simulation and different in different simulations. Goldstein utilized simulations both using and not using non-linearities of the articulatory to acoustic mapping. Not employing it led to the uniform use of all parameters across all simulations, while employing it led to statistical preference for parameters falling in the stable zones of the mapping.

Like de Boer's simulation, in this model agents have coordinated interactions: they follow the rules of a game. Indeed, they both need to produce their gestures together in one round of the game. Secondly, as in the 'imitation game', a pressure for differentiating sounds is programmed in,

[3] There is also the work of Studdert-Kennedy (Chapter 3), but as explained earlier, this focuses on another kind of discreteness in speech, i.e. that related to the independent and parallel use of different sets of organs to perform gestures.

as well as a pressure to copy the parameters of the other agent. This means that it is again presupposed that agents already live in a community in which complex communication exists. However, this was certainly not a concern in that simulation, in the context of research in phonology, whilst the primary concern in the present chapter is the bootstrapping of language. Thus, it remains to be seen how discrete speech, which has been argued to be crucial for the rise of language (Studdert-Kennedy and Goldstein 2003), might have come to exist without presupposing that complex communication had already arisen. More precisely, how might discrete speech appear without a pressure to contrast sounds? This is one of the issues we propose to solve later in the present chapter.

Furthermore, in Goldstein's model, one assumption is that agents directly exchange the targets that they used to produce gestures (there is noise, but they are still given targets). However, human vocalizations are continuous trajectories, first in the acoustic space, and then in the organ relation space. So what a human gets from another's gesture is not the target, but the realization of this target which is a continuous trajectory from the start position to the target. And because targets are sequenced, vocalizations do not stop at targets, but continue their 'road' towards the next target. The task of recovering the targets from the continuous trajectory is very difficult, and has not been solved by human speech engineers. Maybe the human brain is equipped with an innate ability to detect events corresponding to targets in the stream, but this is a strong speculation and so incorporating it in a model is a strong (yet interesting) assumption. In the present chapter, I do not make this assumption: agents will produce complex continuous vocalizations specified by sequences of targets, but initially will be unable to retrieve any kind of 'event' that may help them find out where the targets were. Instead, they use a time resolution filter which ensures that each of the points on the continuous trajectory is considered as a target (while only very few of them actually are targets). This introduces a huge amount of noise (not white noise, but noise with a particular structure). However, I show that our society of agents converges to a state in which agents have broken the continuum of possible targets into a discrete repertoire which is shared by the population. Using the structure of the activation of the neural maps of agents, at the end it is possible to retrieve where the targets were (but this will be a result rather than an assumption).

4.3 The 'blind snowflake-maker' approach

Functionalist models have their strengths and weaknesses, which we are not going to detail in this chapter (for a discussion, see Oudeyer 2003). Instead, I propose another line of research, which I believe is almost unexplored in the field of the origins of language. This is what we might call the blind snowflake maker approach (by analogy with the 'blind watchmaker' of Dawkins 1986, which illustrates the functionalist approach).

There are, indeed, mechanisms in nature which shape the world, such as that governing the formation of snowflakes, which are quite different from Darwinism (Ball 2001). They are characterized by the property of self-organization, like Darwinism, but do not include any concept of fitness or adaptation. Self-organization is here defined as the following property of a system: the local properties which characterize the system are qualitatively different from the global properties of the system.[4]

The formation of snow crystals is illustrated in Figure 4.1. The local mechanism at play involves the physical and chemical interactions between water molecules. If one looks at these physical and chemical prop-

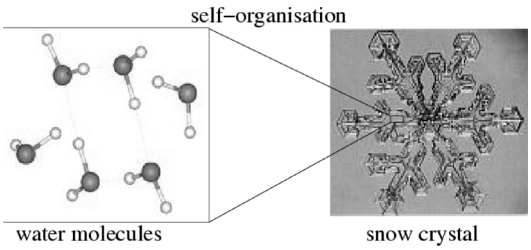

Fig. 4.1 The properties of water molecules and the way they interact are qualitatively different from the symmetrical large-scale structure of snow crystals: this is self-organization.

[4] Note that this definition of self-organization is 'non-magical' and differs from a definition stating that this is a property of systems in which the operations of the higher level cannot be accounted for solely by the laws governing the lower-order level, i.e. cannot be predicted from, nor is reducible to, its constituents. I do not include any dimension of surprise in the concept of self-organization; when I say that the system described in the chapter self-organizes, this does not mean that its behaviour is surprising or unpredictable from its components, but that its global behaviour has qualitative properties different from the properties of its constituents.

erties, one never finds anything that looks like the structure of snow crystals. However, if one lets these molecules interact at the right temperature and pressure, marvellous symmetrical structures, with a great diversity in exact shapes, form (Kobayashi and Kuroda 1987). This is an example of a mechanism that shows self-organization, and builds very complex shapes which are not adaptive or functional (it would be hard to claim that it helps the water to survive). There is no reason why this kind of 'free' formation of structures would not appear in the biological world too. This idea has been defended by Thompson (1932), Gould and Vrba (1982), and Kauffman (1995).

Thompson (1932) gives the example of the formation of the hexagonal honeycomb of the honeybee. Honeybees build walls of wax made by regular hexagonal cells which tile the whole plane. This is remarkable because (a) there are only three ways to tile the plane with regular shapes (squares, triangles and hexagons), and (b) hexagons are optimal since it takes less material to cover the same area with hexagons than with triangles or squares. There are two possible ways to account for this. First, one might think that honeycomb was designed as an adaptation by the honeybees to minimize their metabolic cost: this is the Darwinist functionalist approach. The honeybees would have tried out many possible shapes until they stumbled on hexagons, which they would have found to be less energy-consuming. This would imply that honeybees would have acquired sophisticated instincts that allow them to build perfect hexagons without compasses and set-squares. This explanation is plausible but elaborate, and requires a time-consuming search in the space of forms by the honeybees.

A second explanation, proposed by Thompson, is much more straightforward for the honeybees. Hexagonal forms are the consequence of purely physical forces: if the wax of the comb is made soft enough by the body heat of the bees, then it is reasonable to think of the compartments as bubbles surrounded by a sluggish fluid. And physics makes the bubbles pack together in just the hexagonal arrangement of the honeycomb, provided that initially the wax cells are roughly circular and roughly of the same size; see Figure 4.2. So it might be that initially, honeybees would just build wax cells which were roughly circular and of roughly the same size, and by heating them they automatically obtained hexagonal cells, owing to the self-organizing properties of the physics of packed cells. Note, this does *not* entail that modern honeybees lack an innate,

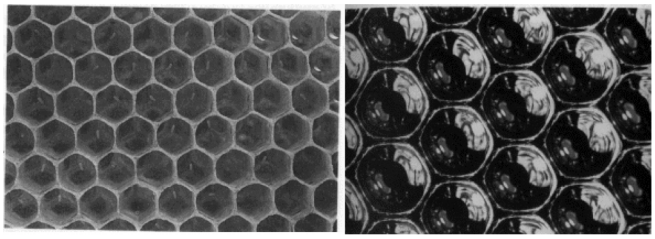

FIG. 4.2 (a) shows the regular hexagonal tiling of the honeycomb; (b) shows the same pattern taken by a raft of water bubbles.

hard-wired neural structure which allows them to build precisely hexagonal shapes; a proposal for such an innate capacity was suggested by von Frisch (1974). Rather, Thompson's proposal is that long ago in evolutionary history, the honeybees might simply have relied on the self-organization of heated packed wax cells, which would have led them to find the hexagon, but later on in their evolutionary history, they might have incorporated into their genome schemata for building those hexagons directly, in a process similar to the Baldwin effect (Baldwin 1896; see Briscoe, Chapter 14).

The goal of this chapter is to present an approach to the formation of speech codes which is very similar in spirit to the approach of D'Arcy Thompson to the formation of honeycomb. We will propose that the formation of sound systems with the properties of discreteness, systematic re-use, universal tendencies, diversity and sharing, may be a result of self-organization occurring in the interactions of modules which were not necessarily selected for communication. The mechanism is not based on the manipulation of the genetic material, but results from the interaction of agents and from a number of generic neural and physical modules (which may have a function on their own, not related to speech communication) during the lifetime of agents. Note that the scenario I propose explains how sound systems with the four properties above could have formed before being related to communication, but says nothing about how it could have been recruited later to be used as an information carrier in communication. If we take the example of the origins of bird feathers used by Gould and Vrba (1982), it is like explaining how the feathers came up with the thermoregulation pressure, but not saying how the feathers were recruited to fly.

4.4 The mechanism

The model is a generalization of that of Oudeyer (2001a), which was used to model a particular phenomenon of acoustic illusion, called the perceptual magnet effect. This model was itself a generalization and unification of the earlier models of Damper and Harnad (2000) and Guenther and Gjaja (1996).

It is based on the building of an artificial system, composed of robots/agents endowed with working models of the vocal tract, the cochlea, and some parts of the brain. The complexity and degree of reality of these models can be varied to investigate which aspects of the results are due to which aspects of the model. I stress that while some parts of the model are inspired by knowledge from neuroscience, we are not trying to reproduce faithfully what is in the human brain. Rather, I attempt to build an artificial world in which we can study the phenomenon described at the beginning of the chapter (i.e. the speech code). Because we know exactly what is happening in this artificial world, in particular what the assumptions are, I hope this will enhance our understanding of speech. The model does this by allowing us to give sufficient conditions for the appearance of a speech code, and it can also tell us what is not necessary (e.g. we will show that imitation or feedback are not necessary). Because the mechanisms that formed speech involve the interaction of many components and complex dynamics, artificial systems are a crucial tool for studying them, and it helps to obtain intuitive understanding of them. Our artificial system aims at proving the self-coherence and logical plausibility of the concept of the 'blind snowflake maker', applied to the origins of discrete speech sounds. For more details on this methodology of the artificial, see Steels (2001) and Oudeyer (2003).

4.4.1 *The architecture of the artificial system*

Here, I summarize the architecture of the system, and in particular the architecture of agents. Technical details can be found in Appendix 4.1 at the end of this chapter. Each agent has one ear which takes measurements of the vocalizations that it perceives, which are then sent to its brain. It also has a vocal tract, the shape of which is controllable and which allows it to produce sounds. The ear and the vocal tract are connected to a brain,

which is basically a set of interconnected neurons. There are two sets of neurons. One is called the 'perceptual map', which gets input from the measurements taken by the ear. Then the neurons of the perceptual map send their output to the second set of neurons, the 'motor map' (this could also be called an 'articulatory map'). These motor neurons send signals to a controller which drives the vocal tract. These signals should be viewed as commands specifying articulatory targets to be reached in time. The articulatory targets are typically relations between the organs of the vocal tract (like the distance between the lips or the place of constriction). They correspond to what is called a 'gesture' in the articulatory phonology literature (Browman and Goldstein 2000; Studdert-Kennedy, Chapter 3). Figure 4.3 gives an overview of this architecture. In this chapter, the space of organ relations will be two-dimensional (place and manner of constriction) or three-dimensional (place, manner of articulation, and rounding).

What we here call a neuron is a box which receives several inputs/measurements, and integrates them to compute its activation, which is propagated through output connections. Typically, the integration is made by first weighting each input measurement (i.e. multiplying the measurement by a weight), then summing these numbers, and applying to

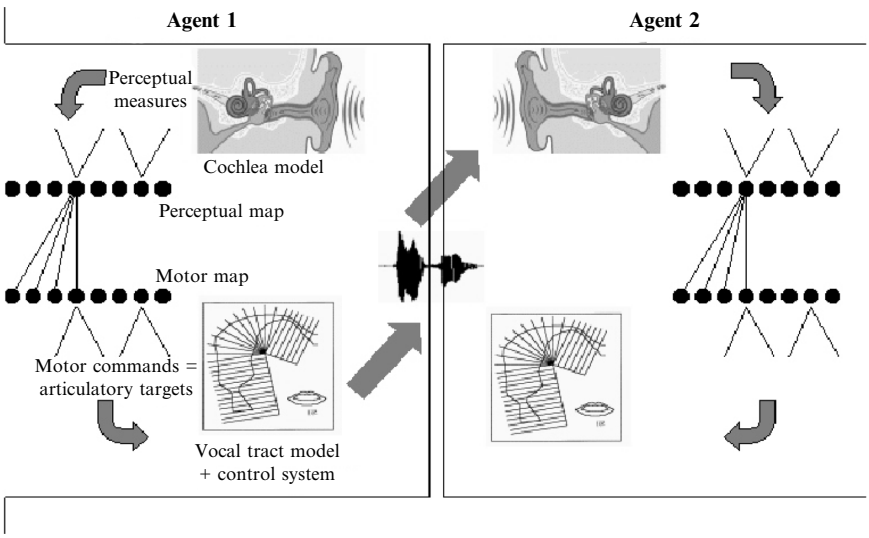

Fig. 4.3 Overview of the architecture of agents in the artificial world

the sum a function called the 'tuning function'. The tuning function is in this case a Gaussian curve, whose width is a parameter of the simulation. A weight is attached to every connection between neurons. In the model, all weights are initially random.

Also, the neurons in each neural map are all interconnected. This means that they receive inputs from all other neurons in the map. When a stimulus is perceived, this prompts an initial activation of all the neurons in the two maps. Next, the activation of each neuron, after the update of all the weights, is updated according to the new activation of the neurons to which it is connected. This is repeated until the activations stabilize. This is what is called an attractor in dynamical systems language. This attractor, i.e. a set of neuron activations which is stabilized, is the same for a number of different stimuli, called its basin of attraction. This models categorization behaviour. There are as many categories as there are attractors.

The production of a vocalization consists of choosing a set of articulatory targets. To choose these targets, one activates neurons in the motor map of agents sequentially and randomly. This activation is a command which we take as the definition of a gesture in this chapter. A target is specified by the weights of the output connections of the activated motor neurons. Then there is a control system which executes these commands by pulling the organs towards the targets continuously and sequentially.[5] Here, the control system simply amounts to generating a continuous trajectory in the organ relation space which passes through the targets. This is achieved through simple spline interpolation, which is basically a polynomial interpolation. Because initially the weights of the connections

[5] It is important to note that this way of producing complex articulation already contains some discreteness. I assume that syllables are specified as a sequence of targets. This is in fact in line with the literature on motor control in mammals (Kandel et al. 2001), which describes it as being organized in two levels: a level of high-level discrete commands (our targets), and a low level which takes care of the execution of these motor commands. So this level of discreteness at the level of commands may not be a feature to be explained in the context of research on the origins of language, since it is already present in the motor-control architecture of mammals. However, I do not assume that initially these targets are organized: the set of commands used to define targets is taken as a continuum of possible commands and there is no re-use of targets from one syllable to another; discreteness and systematic re-use are a result of the simulations. Also, I do not assume that there is discreteness at the perceptual level: agents are not able to detect 'events' in the acoustic stream. (However, at the end they are able to identify the categories of targets which were used to produce the sound.)

are random, agents produce vocalizations with articulatory targets that are spread uniformly across the space of possible targets. This implies that their vocalizations are initially holistic as far as the commands are concerned (the whole continuum of physically possible commands is used). They are not phonemically coded.

Agents produce vocalizations not by a static configuration of the vocal tract, but rather by continuous movement of the vocal tract. This implies that agents receive a continuous trajectory (in the acoustic space) from the vocalizations of other agents. Next I explain how this trajectory is processed, and how it is used to change the weights of the connections between the neurons.

First of all, agents are not able to detect high-level events in the continuous trajectory, which would allow them, for example, to figure out which points were the targets that the other agents used to produce that trajectory. Instead, they segment the trajectory into very small parts, corresponding to the time resolution of perception (this models the time resolution of the cochlea). Then all these small parts are integrated, giving a value in the acoustic space, which is sent to the perceptual neurons. Each perceptual neuron is then activated.

The weights change each time the neurons to which they are connected are activated. The input connections of the perceptual neurons are changed so that the neurons become more sensitive to the stimuli that activated them, and the change is larger for neurons with a high activation than for neurons with a low activation (this is sensitization of neurons). Then the activation of the perceptual neurons is propagated to the motor neurons. Two possibilities ensue: (i) the motor neurons are already activated because the vocalization was produced by the agent itself, and the weights of the connections between the perceptual and the motor neurons are reinforced if they correspond to a link between two neurons whose activation is correlated, and weakened if they correspond to a link between neurons whose activation is not correlated (this is Hebbian learning). This learning rule allows the agent to learn the mapping between percepts and motor commands during babbling. (ii) If the motor neurons were not already activated (the sound comes from the vocalization of another agent), then the weights of the connections between the two maps are not changed, but the weights of the connections between the motor neurons and the control system are changed. The neuron with the highest activation in the neural map is selected, and its output weights, which

specify an organ relation, are used as a reference to update the other weights: they are changed so that the organ relation they specify looks a little more like that of the reference neuron, and this change is weighted by the current activation of each motor neuron.

A crucial point is the coupling between the production process and the perception process. Let us term the weights of the input connections of the perceptual neurons the preferred vectors of these neurons. This term comes from the fact that the set of weights of a neuron forms a vector, and the stimulus that has the same values as the weights will activate the neuron maximally. We also call the output weights of the motor neurons their preferred vector. The set-up and the dynamics of the two neural maps ensure that the distribution of preferred vectors in the motor map corresponds to the distribution of preferred vectors in the perceptual map: if one activates all the neurons in the motor map randomly, many times, to produce sounds, this then gives a distribution of sounds that is the same as the one coded by the neurons of the perceptual map. The distribution of the preferred vectors of the neurons in the perceptual map changes when sounds are perceived, which in turn changes the distribution of preferred vectors in the motor map, which then implies that if an agent hears certain sounds more often than others, he will also tend to produce them more often (here, a 'sound' refers to one small subpart of a vocalization, generated by the time-resolution filter described earlier). It is important to see that this process of attunement is not realized through imitation, but is a side effect of an increase in sensitivity of neurons, which is a very generic, local, low-level neural mechanism (Kandel et al. 2001).

Agents are put together in a world in which they will wander randomly. At random times, they produce a vocalization, and agents next to them hear the sound and adapt their neural maps. Each agent also hears its own sounds, using it to learn the mapping from perception to motor commands.

At the start, every agent produces sounds with targets that are randomly spread across the continuum: this means that this continuum is not discretized and there is no systematic re-use of targets. In other words, agents' vocalizations are holistic. I will show that their neural maps self-organize and synchronize so that after a while they produce complex sounds with targets belonging to a small number of well-defined clusters: the continuum is then discretized. Moreover, the number of clusters is small compared to the number of vocalizations they produce during their lifetime, which implies a systematic re-use of targets across vocalizations.

Finally, these clusters are the same for all agents: the code is shared and specific to each agent community, because in each simulation run, the set of clusters that appears is different (so there is diversity).

I use two kinds of model for mapping from motor configurations to sounds and then perception. The first kind is abstract and trivial: this is a random linear mapping from one space to the other. This allows us to see what we can get without any special mapping properties, in particular without non-linearities. In fact, I show that we get quite far without these, obtaining discreteness, systematic re-use, sharing, and diversity. The second kind is a more realistic model of the mapping, using three motor parameters: tongue height, tongue front–back position, and lip rounding. The formants corresponding to any configurations are then calculated using de Boer's (2001a) model, which is based on human data. This model allows us to predict the vowel systems that appear in human languages, thus allowing us to account for some universal tendencies in human vowel systems.

4.4.2 *Non-assumptions*

Agents do not play a language game in the sense used in the literature (Hurford et al. 1998), and in particular do not play the 'imitation game' which is, for example, used in de Boer (2001a). Their interactions are not structured, there are no roles and no coordination. In fact, they have no social skills at all. They do not distinguish between their own vocalizations and those of others. They do not communicate. Here, 'communication' refers to the emission of a signal by an individual with the intention of conveying information which will modify the state of at least one other agent, which does not happen here. Indeed, agents do not even know that there are other agents around them, so it would be difficult to say that they communicate.

4.5 The dynamics

4.5.1 *Using the abstract linear articulatory/perceptual mapping*

This experiment used a population of twenty agents. I describe first what was obtained when agents use the linear articulatory synthesizer. In the

simulations, 500 neurons were used per neural map, and $\sigma = 0.05$ (width of their tuning function). The acoustic space and the articulatory space are both two-dimensional, with values in each dimension between zero and one. These two dimensions can be thought of as the place and the manner of articulation.

Initially, as the preferred vectors of neurons are randomly and uniformly distributed across the space, the different targets that specify the productions of the agents are also randomly and uniformly distributed. Figure 4.4 shows the preferred vectors of the neurons in the perceptual map of two agents. We see that these cover the whole space uniformly, and are not organized. Figure 4.5 shows the dynamic process of relaxation associated with these neural maps, and due to their recurrent connections. This is a representation of their categorizing behaviour. Each small arrow represents the overall change of activation pattern after one iteration of the relaxation (see the Appendix to this chapter). The beginning of an arrow represents a pattern of activations at time *t* (generated by presenting a stimulus whose coordinates correspond to the coordinates of this point;

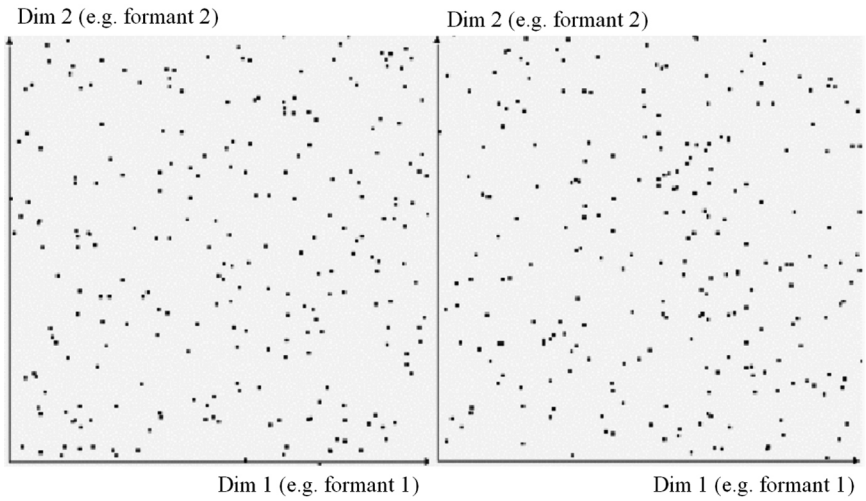

Fig. 4.4 Acoustic neural maps in the beginning. As with all other figures, the horizontal axis represents the first formant (F1), and the vertical axis represents the effective second formant (F2′). The unit is the Bark, and they are oriented from low values to high values. (The Bank is the standard unit corresponding to one critical band width of human hearing.)

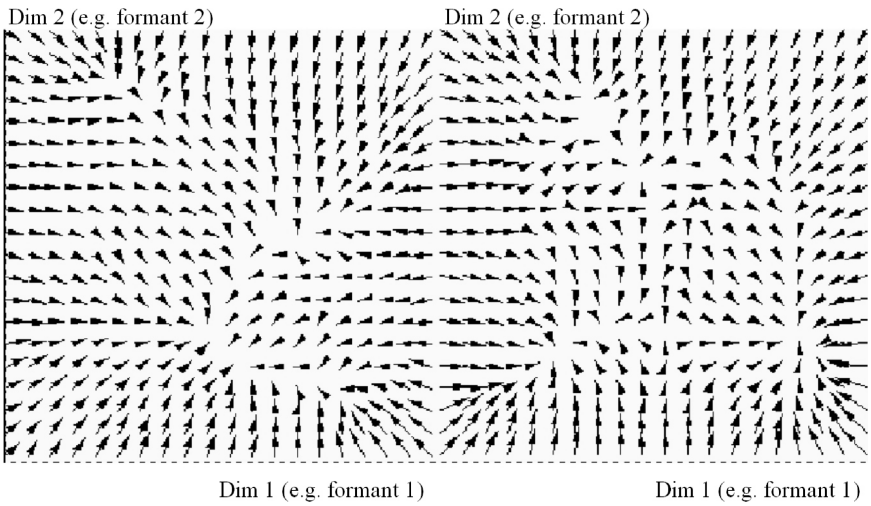

FIG. 4.5 Representation of the two agents' attractor field initially

this is possible because the population vector is also a decoding scheme which computes the stimulus which activated a neural map). The end of the arrow represents the pattern of activations of the neural map after one iteration of the relaxation. The set of all arrows allows one to visualize several iterations: start somewhere on the figure, and follow the arrows. At some point, for every initial point, you get to a fixed point. This corresponds to one attractor of the network dynamic, and the fixed point to the category of the stimulus that gave rise to the initial activation. The zones defining stimuli that fall into the same category are visible on the figure, and are called basins of attractions. With initial preferred vectors uniformly spread across the space, the number of attractors as well as the boundaries of their basins of attractions are random.

The learning rule of the acoustic map is such that it evolves so as to approximate the distribution of sounds in the environment (though this is not due to imitation). All agents produce initially complex sounds composed of uniformly distributed targets. Hence, this situation is in equilibrium. However, this equilibrium is unstable, and fluctuations ensure that at some point, symmetry breaks: from time to time, some sounds get produced a little more often than others, and these random fluctuations may be amplified through positive feedback loops. This leads to a multi-peaked distribution: agents get into the kind of situation in Figure 4.6 (for

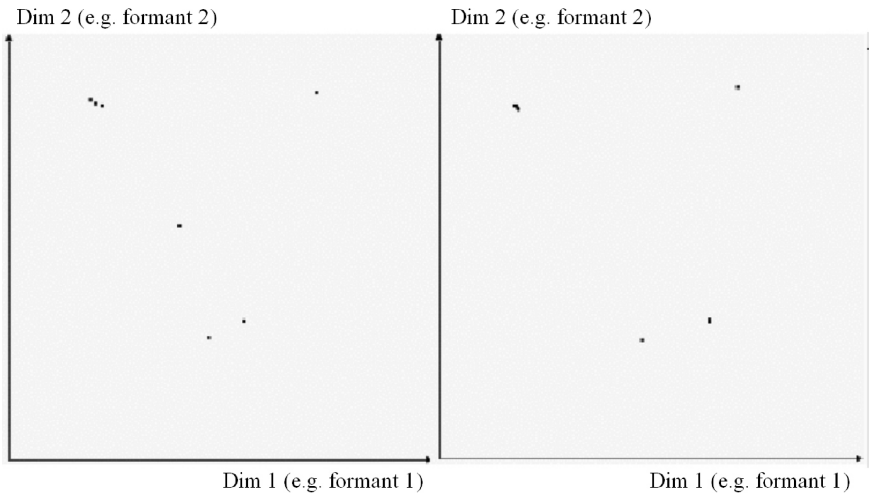

Fig. 4.6 Neural maps after 2000 interactions, corresponding to the initial state of Fig. 4.4. The number of points that one can see is fewer than the number of neurons, since clusters of neurons have the same preferred vectors and this is represented by only one point.

the unbiased case) which corresponds to Figure 4.4 after 2000 interactions in a population of twenty agents. Figure 4.6 shows that the distribution of preferred vectors is no longer uniform but clustered. However, it is not so easy to visualize the clusters with the representation in Figure 4.6, since there are a few neurons which have preferred vectors not belonging to these clusters. They are not statistically significant, but introduce noise into the representation. Furthermore, in the clusters, basically all points have the same value, so that they appear as one point. Figure 4.7 allows us to visualize the clusters better, by showing the attractor landscape that is associated with them. We see that there are now three well-defined attractors or categories, and that these are the same in the two agents represented (they are also the same in the eighteen other agents in the simulation). This means that the targets the agents use now belong to one of several well-defined clusters, and moreover can be classified automatically as such by the relaxation of the network. The continuum of possible targets has been broken; sound production is now discrete. Moreover, the number of clusters that appear is low, which automatically brings it about that targets are systematically re-used to build the complex sounds that

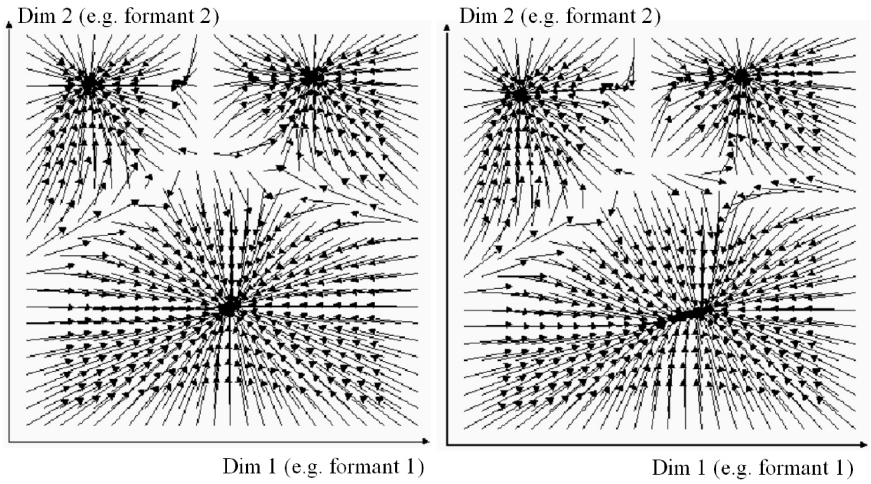

FIG. 4.7 Representation of the attractor fields of two agents after 2000 interactions. The number of attractors is fewer than the number of points in FIG. 4.6. This is because in the previous figures, some points correspond to clusters and other to single points. The broad width of the tuning function ensures that the landscape is smoothed out and individual points which are not too far from clusters do not manage to form their own basin of attraction.

agents produce. All the agents share the same speech code in any one simulation. However, in each simulation, the exact set of modes at the end is different. The number of modes also varies with exactly the same set of parameters. This is due to the inherent stochasticity of the process. I illustrate this later in the chapter.

It is very important to note that this result of crystallization holds for any number of agents (experimentally), and in particular with only one agent, which adapts to its own vocalizations. This means that interaction with other agents—i.e. the social component—is not necessary for discreteness and systematic re-use to arise. But what is interesting is that when agents do interact, then they crystallize in the same state, with the same categories. To summarize, there are, so far, two results: first, discreteness and systematic re-use arise because of the coupling between perception and production within agents; second, shared systems of phonemic categories arise because of the coupling between perception and production across agents.

We also observe that the attractors that appear are relatively well spread across the space. The prototypes that their centres define are thus percep-

tually quite distinct. In terms of Lindblom's framework, the energy of these systems is high. However, there was no functional pressure to avoid close prototypes. They are distributed in that way because of the intrinsic dynamics of the recurrent networks and their rather large tuning functions: indeed, if two neuron clusters get too close, then the summation of tuning functions in the iterative process of relaxation smooths their distribution locally and only one attractor appears.

4.5.2 *Using the realistic articulatory/acoustic mapping*

In the previous subsection, we assumed that the mapping from articulations to perceptions was linear. In other words, constraints from the vocal apparatus due to non-linearities were not taken into account. This is interesting because it shows that no initial asymmetry in the system was necessary to get discreteness (which is very asymmetrical). So there is no need to have sharp natural discontinuities in the mapping from the articulations to the acoustic signals and to the perceptions in order to explain the existence of discreteness in speech sounds (I am not saying that non-linearities of the mapping do not help, just that they are not necessary).

However, this mapping has a particular shape that introduces a bias into the pattern of speech sounds. Indeed, with the human vocal tract, there are articulatory configurations for which a small change effects a small change in the produced sound, but there are also articulatory configurations for which a small change effects a large change in the produced sound. While the neurons in the motor map have initially random preferred vectors with a uniform distribution, this distribution will soon become biased: the consequence of non-linearities will be that the learning rule will have different consequences in different parts of the space. For some stimuli, many motor neurons will have their preferred vectors shifted a lot, and for others, very few neurons will have their preferred vectors shifted. This will very quickly lead to non-uniformities in the distribution of preferred vectors in the motor map, with more neurons in the parts of the space for which small changes result in small differences in the produced sounds, and with fewer neurons in the parts of the space for which small changes result in large differences in the produced sounds. As a consequence, the distribution of the targets that compose vocalizations will be biased, and the learning of the neurons in

the perceptual maps will ensure that the distributions of the preferred vectors of these neurons will also be biased.

The articulatory synthesizer used is from de Boer (2001a). This models only the production of vowels. The fact that agents produce only vocalizations composed of vowel sounds does not imply that the model does not hold for consonants. I chose this articulatory synthesizer because it is the only one both fast enough and realistic enough for my computer simulations. The articulatory space (or organ relation space) is three-dimensional here: tongue height (i.e. manner of articulation), tongue front–back position (i.e. place of articulation), and lip rounding. Each set of values of these variables is then transformed into the first four formants, which are the poles of the vocal tract shaped by the position of the articulators. Then the effective second formant is computed, which is a non-linear combination of the second, third, and fourth formants. The first and effective second formants are known to be good models of our perception of vowels (de Boer 2001a). To get an idea of this, Figure 4.8 shows the state of the acoustic neural maps of one agent after a few interactions between the agents (200 interactions). This represents the bias in the distribution of preferred vectors due to the non-linearities.

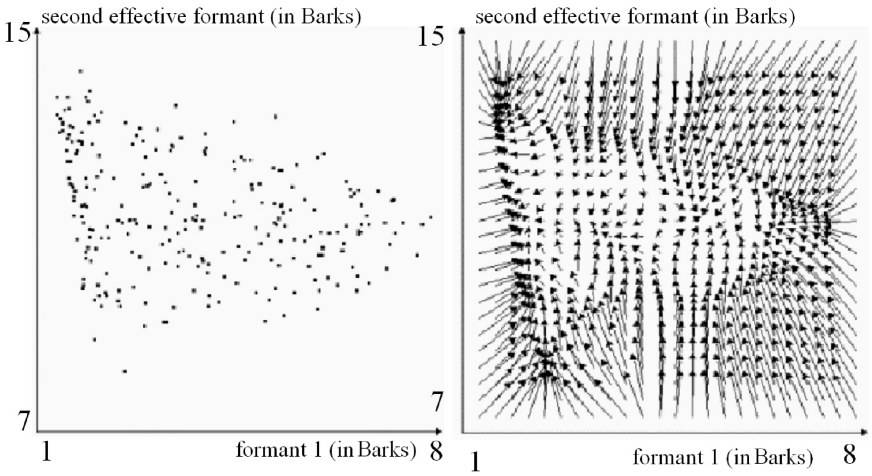

Fig. 4.8 Initial neural map and attractor field of one agent within a population of twenty agents. Here the realistic articulatory synthesizer is used.

A series of 500 simulations was run with the same set of parameters, and each time the number of vowels as well as the structure of the system was checked. Each vowel system was classified according to the relative position of the vowels, as opposed to looking at the precise location of each of them. This is inspired by the work of Crothers (1978) on universals in vowel systems, and is identical to the type of classification in de Boer (2001a). The first result shows that the distribution of vowel inventory sizes is very similar to that of human vowel systems (Ladefoged and Maddieson 1996): Figure 4.10 shows the two distributions (the plain line is the distribution corresponding to the emergent systems of the experiment; the dotted line is the distribution in human languages), and in particular the fact that there is a peak at five vowels, which is remarkable since five is neither the maximum nor the minimum number of vowels found in human languages. The prediction made by the model is even more accurate than that of de Boer (2001a), since his model predicted a peak at four vowels. Then the structure of the emergent vowel systems was compared to those in human languages as reported in Schwartz et al. (1997a). More precisely, the distributions of structures in the 500 emergent systems was compared to the distribution of structure in the 451 languages of the UPSID database (Maddieson 1984). The results are shown in Figure 4.11. We see that the predictions are fairly accurate, especially in the prediction of the most frequent system for each size of vowel system (less than eight). Figure 4.9 shows an instance of the most frequent system in both emergent and human vowel systems. In spite of the predictions of one four-vowel system and one five-vowel system which appear frequently (9.1 and 6 per cent of systems) in the simulations and never appear in UPSID languages, these results compare favourably to those obtained by de Boer (2001a). In particular, we obtain all this diversity of systems with the appropriate distributions with the same parameters, whereas de Boer had to modify the level of noise to increase the sizes of vowel systems. However, like de Boer, we are not able to predict systems with many vowels (which are admittedly rare in human languages, but do exist). This is certainly a limitation of our non-functional model. Functional pressure to develop efficient communication systems might be necessary here. In conclusion, one can say that the model supports the idea that the particular phonemes which appear in human languages are under the influence of the articulatory/perceptual mapping, but that their existence, which means the phenomenon of phonemic

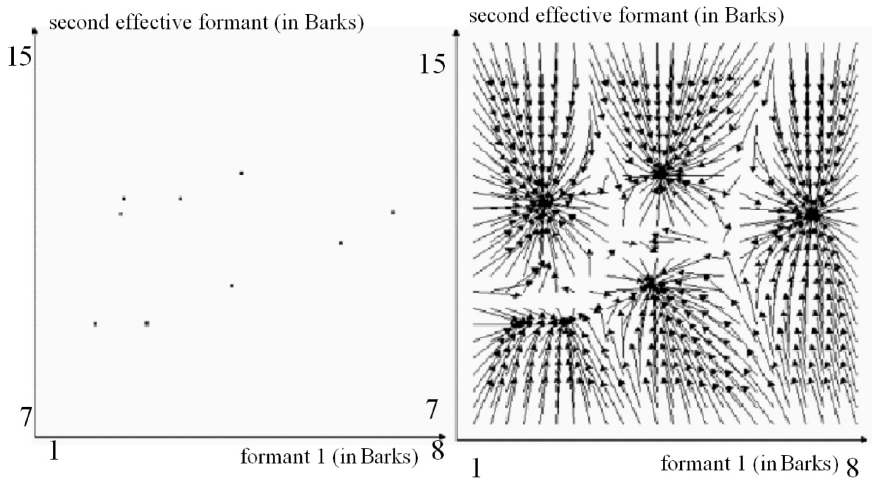

Fig. 4.9 Neural map and attractor field of the agent from Fig. 4.8 after 2000 interactions with the other twenty agents. The corresponding figures for other agents are nearly identical, as in Fig. 4.6 and 4.7. The vowel system produced corresponds to the most frequent five-vowel system in human languages.

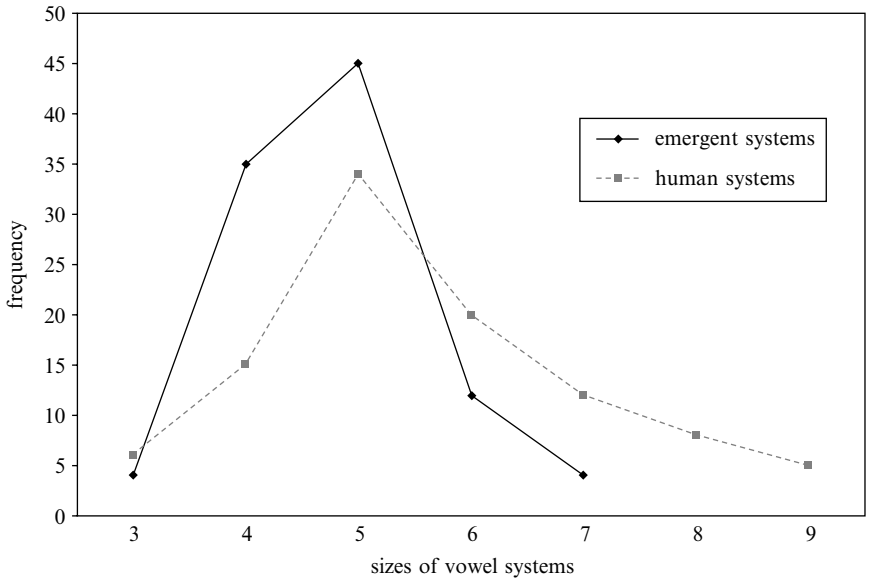

Fig. 4.10 Distribution of vowel inventory sizes in emergent and UPSID human vowel systems

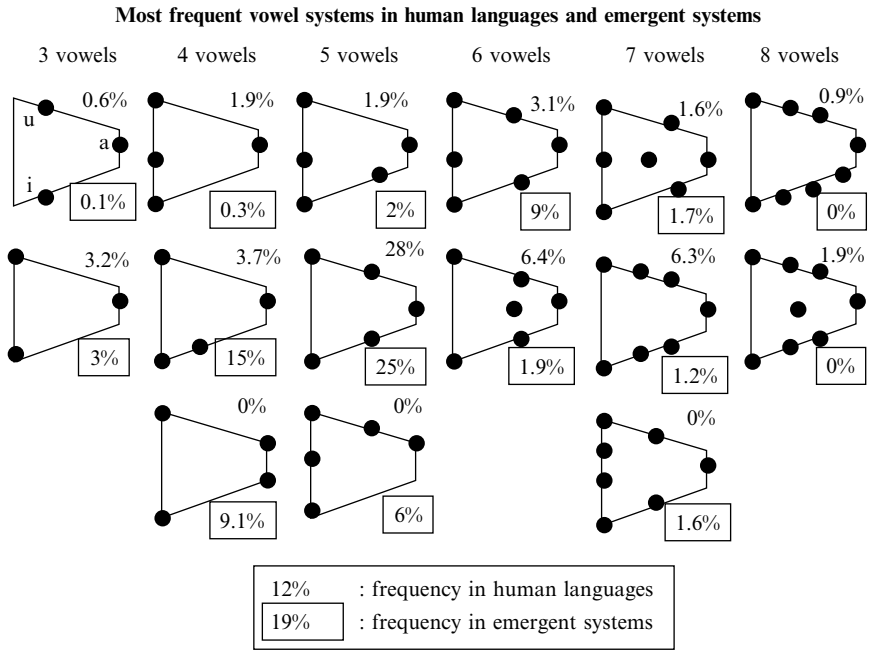

FIG. 4.11 Distribution of vowel inventory structures in emergent and UPSID human vowel systems. This diagram uses the same notations as Schwartz et al. (1997). Note that here, the vertical axis is also F2, but oriented from high values to low values.

coding, is not due to this mapping but to the sensory-motor coupling dynamics.

4.6 Discussion

A crucial assumption of the artificial system presented in this chapter is the fact that there are connections between the motor vocal neural map and the perceptual acoustic map which allow the agents to learn the mapping between the two spaces. How might these connections have appeared?

First, it is possible that they appeared through Darwinian genetic evolution under a pressure for language. But the simplicity and non-specificity of the neural architecture allows other ways of explaining

their origins which do not necessitate a pressure for language. These scenarios truly illustrate the 'blind snowflake maker' approach.

One alternative scenario is that these connections evolved for imitation. Imitation may have appeared for purposes very different from language: for example, it might have evolved to maintain social cohesion. Copying of types of behaviour might have been used to mark friendship, for example, as in some species of birds. Interestingly, this kind of imitation does not require a system of sounds made of discrete units that can be re-used to produce infinite combinations. Also, in this kind of imitation, agents do try to copy behaviours or sounds, but do not try to discriminate sounds. This means that there is no pressure to develop a system of sounds that are different from each other and categorized as such. There is no need to have a system of categories as a whole if all that is useful is just evaluating the similarity of the sound you yourself produce to one produced by another individual at a given moment. But discreteness, re-use, and a system of differences and categories are necessary for speech. The artificial system of this chapter shows precisely that with just a simple neural system which may very well have evolved for 'imitation for social cohesion' (it would be difficult to make a simpler system), we obtain freely, through self-organization, a system of sounds that is shared by a population and that is discrete, with systematically re-used units and a system of categorization. In other words, we obtain exactly what speech needs without the need for speech.

So, even if the neural devices that are assumed in this chapter evolved for imitation, they produce speech-sound systems without a functional pressure for speech (as used in the context of language). The simulations of de Boer and of Browman and Goldstein do assume this pressure for speech, since their agents do try to produce the sounds of their repertoire differently (and in the case of de Boer, they try to make the repertoire as big as possible). But the agents here do not try to distinguish sounds. The fact that they reach a system of categories that allows them to distinguish sounds is a self-organized result. Furthermore, the notion of repertoire is not pre-programmed, but appears as a result.

A second alternative scenario for the origins of the neural structure which allows the learning of the mapping between sounds and articulatory configurations is possible. The system just needs initially random neurons that are sensitive to sounds, random neurons that are sensitive to motor commands, and random connections between these two sets of neurons.

Then it needs the random connections between these two sets of neurons to adapt by following a very general dynamic: Hebbian learning. Next, activating the motor neurons randomly and uniformly leads to a movement of the vocal tract which produces sounds, which in turn activates the perceptual neurons, and then the connections between the two maps self-organize so that after a while the mapping is effectively learnt. This is what we call babbling.

Crucially, this architecture does not require precise pre-wiring during ontogeny which is pre-programmed by the genes. The neurons in the perceptual map and in the motor map certainly existed well before speech. In fact they have existed since ears and mouths have existed. So the question, of course, is how did they come to be connected? It is quite possible that these connections are a side-effect of general architectural design constraints of the brain; Gould and Vrba (1982) give many examples of other features of the bodies and brains of animals which appeared in a similar manner. Indeed, it is obvious that the connections between certain modalities (for example, vision and the motor control of arms) are necessary and thus existed very early in mammalian evolution. It might very well be that the most efficient strategy to produce these connections is to connect all modalities rather than just to connect particular modalities with other particular modalities. This might be more efficient because it requires fewer specifications for the growth process, and thus might be more robust, and the advantage of this robustness might be superior to the cost of having *a priori* unnecessary connections. In fact, this method of shaping the brain by initial generation of many random neural structures, followed by a pruning phase, is accepted by a large part of the neuroscience community (see Changeux 1983). But then all mammals should have these connections between neurons that perceive sounds and neurons that control the movements of the mouth. So, why are we the only mammal to have such a system of speech sounds? And in particular, why do monkeys or chimps not have speech sounds like ours? It is probable that they do have at birth the connections between the neurons that perceive sounds and those that control the mouth, but that they lose them because the key is somewhere else. The key might lie in *babbling*.

In fact, precisely one of the assumptions that I make (and which monkeys or chimps do not seem to implement) is that the agents activate spontaneously, often, and randomly, the neurons of their motor map. This

means that they spontaneously try out many articulatory configurations and repeat these trials. In other words, they practise. Monkeys or chimps do practise certain specific motor activities when they play, but these are very limited and they do not try to practise all the motor activities that their bodies allow, while human children do. For example, once monkeys have thrown a stone towards an objective, they will not try to do it again repeatedly. And it seems that a major evolutionary event which gave rise to primitive humans with increased skills as compared to their ancestors is the ability to practise any new motor activity that they encounter, in particular vocal babbling. Indeed, a general drive to explore all motor activities available to the body may have been very beneficial for the learning of many skills useful for primitive humans, who lived in a dynamic, quickly changing environment (for example, coping with changes in habitat, or living in complex dynamic social structures). This may have pushed them, in particular, to use their vocal apparatus for babbling. And then we come to the beginning of the simulation presented in this chapter, which shows that self-organization takes place and generates 'for free' a system of sounds shared by the agents who live in the same area and which is phonemically coded/discrete. Monkeys or chimps may have the connections, but because they do not practise, the neural structures connecting the two modalities certainly die (through the pruning process of activity-dependent neural epigenesis; Changeux 1983). But humans do practise, which not only allows the neural system to be kept alive, but also allows the generation of a shared speech code.

4.7 Conclusion

This chapter presents a mechanism providing a possible explanation for how a discrete speech code may form in a society of agents which does not already possess the means to communicate and coordinate in a language-like manner. Contrary to other computational models of the origins of language (see Cangelosi and Parisi 2002), the agents do not play language games. They have, in fact, no social skills at all. I believe the mechanism presented may be the kind of mechanism that could solve the language bootstrapping problem. I have shown how one crucial prerequisite, i.e. the existence of an organized medium that can carry information in a

conventional code shared by a population, may appear without linguistic features being already there.

Furthermore, this same mechanism allows us to account for properties of the speech code like discreteness, systematic re-use, universal tendencies, sharing, and diversity. I believe that this account is original because (a) only one mechanism is used to account for all these properties and (b) we need neither a pressure for efficient communication nor innate neural devices specific to speech (the same neural devices used in the chapter can be used to learn hand-eye coordination, for example).

Models like that of de Boer (2001a) are to be seen as describing phenomena occurring later in the evolutionary history of language. More precisely, de Boer's model, as well as, for example, that of Oudeyer (2001b) for the formation of syllable systems, deals with the recruitment of speech codes like those that appear in this chapter, and studies how they are further shaped and developed under functional pressure for communication. Indeed, whilst we have shown here that one can go a long way *without* such pressure, some properties of speech can only be accounted for *with* it. An example is the phenomenon of chain shifts, in which the prototypes of sounds of a language are all moved around the space.

However, in de Boer (2001a) and Oudeyer (2001b), the recruitment of the speech code is pre-programmed. How this could have happened in the origins of language is a problem which remains to be solved. A particular instantiation of the problem is: how do agents come to have the idea of using a speech code to name objects? In fact, the problem of the recruitment of features not initially designed for a certain linguistic function is present at all levels of language, ranging from sounds to grammar. The question of how recruitment comes about is a major challenge for research on the origins of language. In this chapter, we have shown one example of recruitment: individual discrete sounds were systematically re-used in the building of complex vocalizations, and this was not pre-programmed.

ACKNOWLEDGEMENTS

I would like to thank Michael Studdert-Kennedy for his very helpful comments which helped to improve this chapter greatly. I would also like to thank Luc Steels for supporting the research presented in the chapter.

FURTHER READINGS

For a comprehensive overview of research on the origins of speech sounds, Lindblom (1992), de Boer (2001), and Studdert-Kennedy and Goldstein (2003) are key references.

For the role of self-organization in the origins of patterns in the biological world, and in particular for the relationship between self-organization and neo-Darwinian natural selection, Kauffman (1995) is a good start, presenting the general arguments in an accessible way, and could be followed up by Ball (2001) and Thompson (1932), which present a great many biological examples, with more technical and empirical detail.

Appendix 4.1: Technical Details of the Mechanism

The neurons have a Gaussian tuning function. If we note $tune_{i,t}$ the tuning function of n_i at time t, s one stimulus vector, v_i the preferred vector (the weights) of n_i, then the form of the function is:

$$tune_{i,t}(s) = \frac{1}{\sqrt{2\pi\sigma}} * e^{-\frac{1}{2}v_i * s^2/\sigma^2}$$

The notation $v_1 * v_2$ denotes the scalar product between vector v_1 and vector v_2. The parameter σ determines the width of the Gaussian, and so if it is large the neurons are broadly tuned (a value of 0.05 means that a neuron responds substantially to10 per cent of the input space).

When a neuron in the perceptual map is activated because of an input s, then its preferred vector is changed. The mathematical formula of the new tuning function is:

$$tune_{i,t+1}(s) = \frac{1}{\sqrt{2\pi\sigma}} * e^{v_{i,t+1} * s2/\sigma 2}$$

where s is the input, and $v_{i,t+1}$ the preferred vector of n_i after the processing of s:

$$vi,t+1 = \ v_{i,t} + 0.001 * tune_{i,t}(s) * (s - v_{i,t})$$

Also, when a sound is perceived and through propagation activates the motor neurons, the weights of the output connections of these neurons also change. The preferred vector of the most active neuron is taken as a reference: the other preferred vectors are changed so that they get closer to this preferred vector. The change is made with exactly the same formula as for the neurons in the perceptual map, except that s is the preferred vector of the most active neuron. When an agent hears a vocalization produced by itself, the motor neurons are already activated when the perceived sound activates the neurons in the perceptual map. Then, the weights of the connections between the two neural maps

change. A Hebbian learning rule is used. If i is a neuron of the perceptual map connected to a neuron j of the motor neural map, then the weight $w_{i,j}$ changes:

$$\delta w_{i,j} = c_2 * (tune_{i,s_i} - < tune_{i,s_i} >)(tune_{j,\ s_j} - <tune_{j,s_j}>)(\text{correlation rule})$$

where s_i and s_j are the input of neurons i and j, $<tune_{i,s_i}>$ the mean activation of neuron i over a certain time interval, and c_2 a small constant. All neurons between the two maps are connected.

Both the perceptual and the motor neural map are recurrent. Their neurons are also connected to each other. The weights are symmetric. This gives them the status of a dynamical system: they have a Hopfield-like dynamics with point attractors, which are used to model the behaviour of categorization. The weights are supposed to represent the correlation of activity between neurons, and are learnt with the same Hebbian learning rule:

$$\delta w_{i,j} = c_2(tune_{i,s_i} - <tune_{i,s_i}>)(tune_{j,\ s_j} - <tune_{j,s_j}>)(\text{correlation rule})$$

These connections are used to relax each neural map after the activations have been propagated and used to change the connections weights. The relaxation is an update of each neuron's activation according to the formula:

$$act(i,\ t+1) = \frac{\sum_j act(i,t) * w_{i,j}}{\sum_i act(i,t)}$$

where $act(i)$ is the activation of neuron i. This is the mechanism of competitive distribution, together with its associated dynamical properties.

To visualize the evolution of the activations of all neurons during relaxation, we use the 'population vector'. The activation of all the neurons in a neural map can be summarized by the 'population vector' (see Georgopoulos et al. 1988): it is the sum of all preferred vectors of the neurons weighted by their activity (normalized as here we are interested in both direction and amplitude of the stimulus vector):

$$pop(v) = \frac{\sum_i act(n_i) * v_i}{\sum_i act(n_i)}$$

The normalizing term is necessary here since we are not only interested in the direction of vectors.