Adaptive Behavior

Multi-Agent Simulations of the Evolution of Combinatorial Phonology

Bart de Boer[1], Willem Zuidema[2]

[1] *Amsterdam Center for Language and Communication, University of Amsterdam, The Netherlands*

[2] *Institute for Logic, Language and Computation, University of Amsterdam, The Netherlands*

A fundamental characteristic of human speech is that it uses a limited set of basic building blocks (phonemes, syllables), that are put to use in many different combinations to mark differences in meaning. This article investigates the evolution of such *combinatorial phonology* with a simulated population of agents. We first argue that it is a challenge to explain the transition from holistic to combinatorial phonology, as the first agent that has a mutation for using combinatorial speech does not benefit from this in a population of agents that use a holistic signaling system. We then present a solution for this evolutionary deadlock. We present experiments that show that when a repertoire of holistic signals is optimized for distinctiveness in a population of agents, it converges to a situation in which the signals can be analyzed as combinatorial, even though the agents are not aware of this structure. We argue that in this situation adaptations for productive combinatorial phonology can spread.

Keywords phonemic coding · combinatorial structure · language evolution · language games

1 Introduction

Human speech is combinatorial. This means that it combines a limited number of basic sounds into a potentially infinite set of complex utterances that all differ in meaning. Languages can be extremely complex and diverse in the repertoire of speech sounds they use and in the way they are combined. For example, the Khoisan language !Xóõ is analyzed to use 186 different speech sounds (Traill, 1985) while the Caucasian language Georgian allows concatenation of many consonants such as in the word *prtskvna* "to peel" (Catford, 1977). Such complex learned systems are an essential component of the discrete infinity of meanings that languages must express.

Despite this apparent complexity and the wide scope of possibilities, children acquire the system of speech sounds of their native language remarkably quickly. At six months of age infants already learn which speech sounds are important in their native tongue, even before they can properly produce speech or understand the complete meaning of utterances (Kuhl, Williams, Lacerda, Stevens, & Lindblom, 1992).

Compared with our closest relatives, these are impressive feats. Although some primates use utterances that are built up from smaller units (Arcadi, 1996; Mitani & Marler, 1989) changing the order of the units does not substantially alter the meaning of the utterances. And although there are indications that chimpanzee vocalizations are partly learned (Crockford, Herbinger, Vigilant, & Boesch, 2004), chimpanzees only have very limited abilities for vocal imitation (Fitch, 2000).

These facts suggest that the last common ancestor of humans and other great apes did not use combinato-

Correspondence to: Bart de Boer, Amsterdam Center for Language and Communication, University of Amsterdam, Spuistraat 210, 1012 VT, Amsterdam, The Netherlands. *E-mail*: b.g.deboer@uva.nl.
Tel.: +31 20 525 2182, *Fax*: +31 20 525 3021.

(2010), Vol 18(2): 141–154.
DOI: 10.1177/1059712309345789
Figures 5, 6 appear in color online: http://adb.sagepub.com

rial phonology in semantic communication. Hence, the ability for learning and using combinatorial systems of speech sounds must have evolved in the hominid lineage since. We should thus ask the question how the transition from holistic to combinatorial repertoires of speech sounds could have taken place.

Although combinatorial systems are in general more robust against noise and therefore preferable from an information theoretic point of view (Nowak & Krakauer, 1999), this does not in itself constitute an evolutionary explanation. Crucially, evolutionary explanations must provide a path of ever-increasing fitness, where each new variant can invade in a population where it is initially infrequent. One can imagine that once a holistic repertoire is established in the population, adaptations for combinatorial speech never have a chance to spread. After all, if all the other agents in the population are using holistic systems, a mutant agent will not benefit from being able to produce, perceive, or learn more combinatorial utterances. A recurring problem in language evolution research is that the usefulness of an innovation depends on how many agents in the population can process it: fitness in language evolution is typically “frequency-dependent” (Cavalli-Sforza & Feldman, 1983).

Explanations that propose that the combinatorial nature is an exaptation of existing behavior, such as repetitive motion of the jaw in chewing and breathing (MacNeilage & Davis, 2000) are only partly satisfactory. They first of all only explain the syllable structure of speech, and not the internal combinatorial nature of syllables. But more importantly, they do not provide a formal account of how the proposed adaptations of preexisting behaviors were initially selected for and managed to spread in a population.

Such scenarios would be much more convincing if one could show that even when individuals do not make use of a combinatorial system of internal representations for speech, systems of utterances evolve—culturally or genetically—toward showing aspects of combinatorial systems. Such systems can be said to be *superficially combinatorial*. We define superficial combinatorial structure as combinatorial structure that can be observed by an outside observer in a system of signals, but that is not actively used by the agents using the signals. Whenever combinatorial structure is used by agents in producing, perceiving, learning, or storing signals, we prefer to refer to it as a *productive combinatorial* system. We propose a mechanism by which superficial combinatorial structure can emerge in a population. Subsequently, in a population that uses superficially combinatorial speech sounds, a mutation for using this combinatorial structure would have a chance of spreading, thus providing a pathway of continuously increasing fitness from holistic to productively combinatorial utterances. Adaptations for learning combinatorial structure could evolve in this way.

We propose that optimizing for acoustic distinctiveness a system of speech sounds that are extended in time results in a superficially combinatorial structure. We investigate this using a simulated population of agents that try to imitate each other as well as possible. The results from this study support the hypothesis, and hence fill an important gap in existing explanations for the origins of combinatorial speech.

The model differs in two important ways from the few existing formal models of the evolution of combinatoriality. First, combinatorial and holistic signals are represented in the same way as trajectories in an acoustic space, and the agents in our simulations produce and perceive signals purely holistically. This makes them different from the agents used in Lindblom, MacNeilage, and Studdert-Kennedy (1984) and Oudeyer (2002) where agents make use of the combinatorial structure of the utterances in their repertoire, and the combinatorial structure is built in. These models did not so much investigate the emergence of combinatorial structure itself, but the emergence of discrete categories if some degree of combinatorial structure was already present.

By avoiding an a priori distinction between holistic and combinatorial trajectories, our model is also different from the mathematical model in Nowak and Krakauer (1999), where individuals are assumed to have two independent repertoires in parallel, one holistic and one combinatorial. Moreover, our model also assumes that a new mutant repertoire can only invade in a population if it represents an increase in communicative success *even when talking mainly to speakers of the resident repertoire*. This is a much stricter criterion than used by Lindblom et al. (1984), where a new mutant repertoire is already assumed to invade if a whole mutant population does better than a resident population. It is also much stricter than Nowak and Krakauer (1999), where the holistic and combinatorial repertoires are assumed not to interact at all. In Oudeyer (2002), finally, communicative success is not monitored and does not play a role in the dynamics.

2 The Model

The model that was used in this research was an individual- or agent-based computer simulation. In individual-based models of language and its evolution, a population of language users is modeled with a set of simplified models of the individuals (called agents) in the population. Each agent is able to engage in certain language-like interactions with other individuals of the population. The agent-based paradigm has been used by other researchers of language evolution and examples include "language games" (Oliphant & Batali, 1996; Steels, 1997, 1998) and "iterated learning" (Kirby & Hurford, 2001; Smith, Kirby, & Brighton, 2003; Zuidema, 2003). The aim of the agents is to develop a repertoire of signals with which they can communicate as well as possible with the other agents in the population. For this it is required that the agents in the population agree on a repertoire and that the signals in the repertoire are as different from each other as possible.

Each agent has a repertoire of trajectories and is able to produce and perceive these as signals in an abstract "acoustic" space. In the experiments presented here, this space was chosen to be two-dimensional, but it could in principle have any number of dimensions. The dimensions can be imagined as features of the acoustic signal. For human perception, such features could be the pitch, the loudness, or the formants (peaks in the frequency spectrum) of the signal. In the model presented here the space is abstract. The dimensions do not correspond to any real feature of the signal. We decided to use an abstract space, because our aim was to investigate a general property of trajectories used for signaling, independent of the actual properties of perception and production. Using more realistic features would result only in an alteration of the shape of the acoustic space. Our acoustic space is a square with sides of length 10 in all simulations presented here.

Trajectories in this space consist of a fixed number (N) of points, representing acoustic signals with fixed duration. Points on a trajectory can be considered as samples of that trajectory, taken at fixed time intervals. Points on a trajectory can have any distance between 0 and R to their predecessor and their successor. The values of N and R were taken to be 20 and 1, respectively in the simulations presented here. The acoustic space and a trajectory are illustrated in Figure 1.

Distances between two trajectories T_1 and T_2 are calculated as the sum over all distances between corresponding points of the trajectories, corresponding to the following equation:

$$d = \sum_{i=1}^{N} \|t_{1,i} - t_{2,i}\| \tag{1}$$

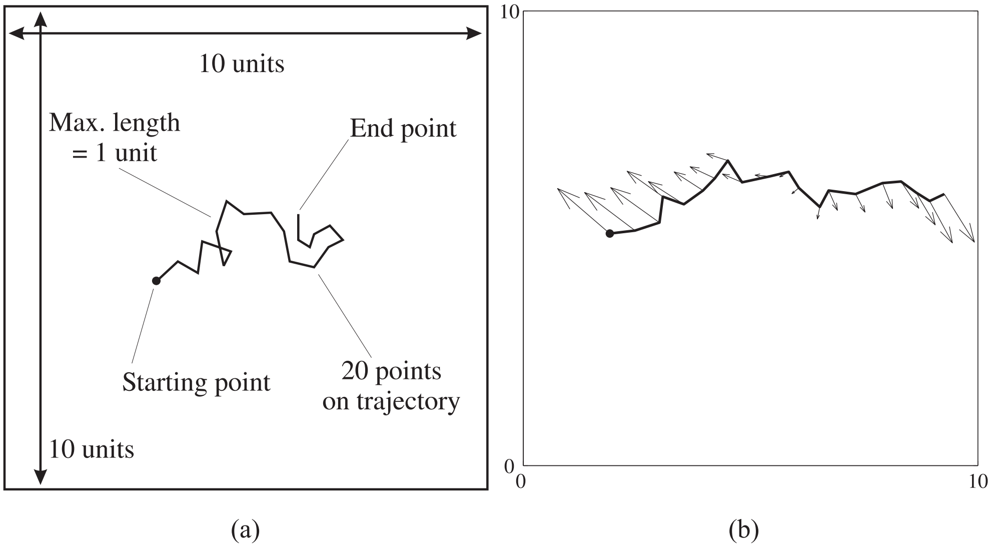


Figure 1 (a) Illustration of an abstract acoustic space and a trajectory in it. (b) Example of shape-preserving noise. Arrows indicate shift by noise. Note correlation between neighboring shifts.

where $t_{1,\,i}$ and $t_{2,\,i}$ are points from trajectories T_1 and T_2, respectively. The double bars give the Euclidean vector distance (the points $t_{1,\,i}$ and $t_{2,\,i}$ are of the same dimensionality as the assumed acoustic space). This distance measure is convenient and easy to calculate for trajectories with a fixed number of points. It can be argued that for a variable number of points, a distance measure such as dynamic time warping would be better and that this would probably correspond better to the way humans perceive acoustic signals (Sakoe & Chiba, 1978). Some preliminary experiments with dynamic time warping were performed but no qualitative difference in performance between the two distance measures was found.

When an agent produces an utterance, noise is added. Noise is added in a way that preserves the general shape of trajectories. We found that this leads to faster convergence than adding noise that is independent for each point on the trajectory. Adding noise that is correlated from point to point is more realistic, as it implies the existence of disturbances of longer duration (the behavior of systems with uncorrelated noise was qualitatively similar). The method adds independent shifts to the first and last points of the trajectory, and interpolates for the points in between. A smaller independent shift is also added to all points. In equation form:

$$t'_{x,\,i} \leftarrow t_{x,\,i} + \alpha_i \mathbf{s}_{x,\,1} + (1 - \alpha_i)\mathbf{s}_{x,\,N} + \mathbf{n}_i, \qquad (2)$$

where $\mathbf{s}_{x,\,1}$ and $\mathbf{s}_{x,\,N}$ are vectors that are the same for all points i on any one trajectory x, but different for different trajectories (their components are taken from the normal random distribution with mean 0 and standard deviation σ_{noise}). The vector $\mathbf{n}_i$ is different for each point on the trajectory (each component is taken from the normal random distribution with mean 0 and standard deviation $\frac{\sigma_{noise}}{N}$). The α is a weight factor that varies from 0 on one end of each trajectory, to 1 on the other end: $\alpha_i = \frac{i-1}{N-1}$. As this kind of noise preserves the overall shape of a trajectory, it is called shape-preserving noise. An example of shape preserving noise is presented in Figure 1.

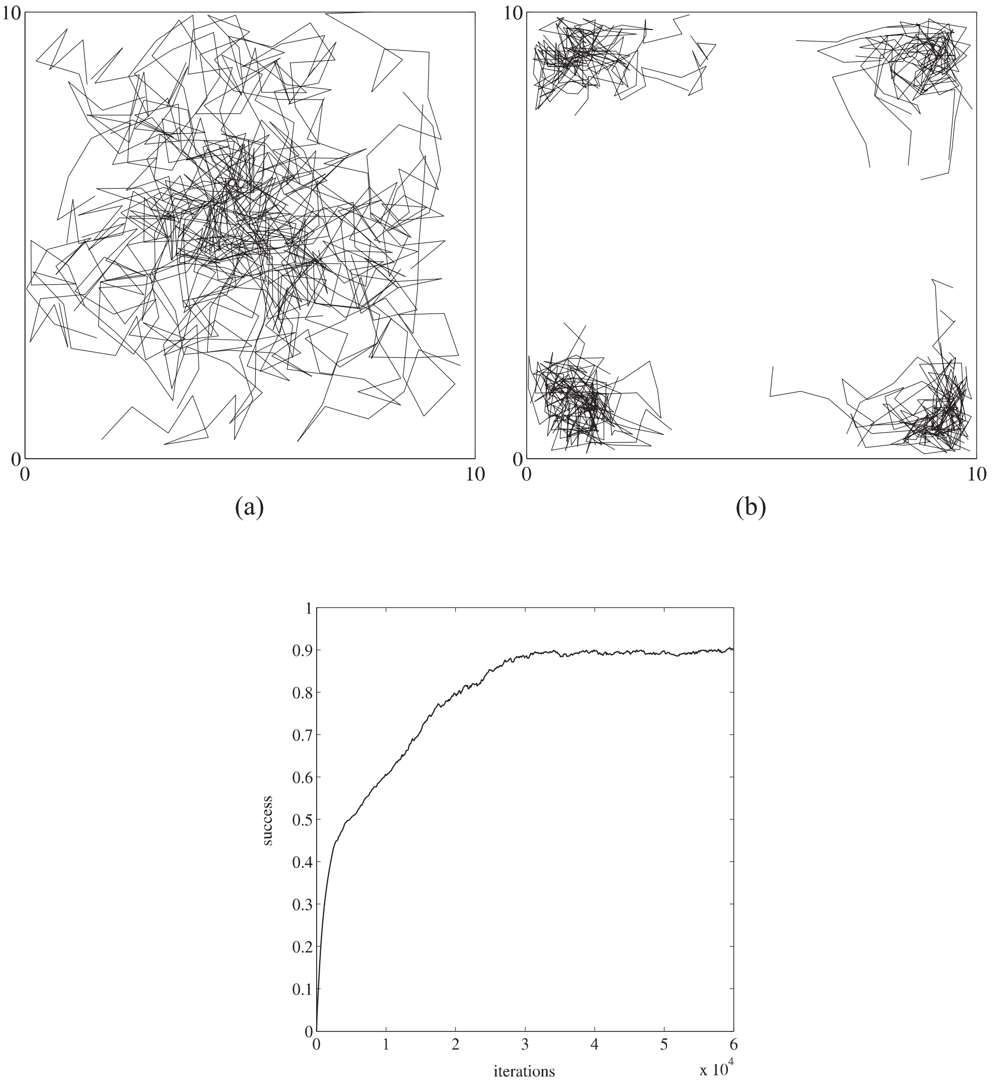
The interaction between agents is modeled by a variant of the "imitation game" (de Boer, 2000, 2001). An imitation game is an interaction between two agents that has been designed to create pressures for signals that are as distinctive as possible, but that does not require the signals to have meaning. This is achieved by letting the agents imitate each other using categorical perception and noisy transmission. Details of the procedure are given below.

The number of trajectories in every agent is set to a fixed number, K, beforehand. Trajectories are initialized randomly at the start of the simulations. The first point is randomly taken from the uniform distribution over a square of size 1 × 1 (1/100th of the available acoustic space) at the center of the acoustic space. Subsequent points on the trajectory are generated at distance R from the previous points and with uniformly distributed angle with respect to the previous point. An example of such a randomly initialized system can be found in Figure 2a.

A success score is associated with each trajectory of an agent. This score measures how successful a trajectory has been in previous interactions. This score is initially set to 0 for each trajectory.

For each iteration of a simulation, one agent (the *initiator*) is selected randomly from the population of agents. This agent selects one trajectory from its repertoire and makes a slight modification to it. The modification is made by selecting a random point and adding a vector to it. The vector has components that are taken from the normal random distribution with mean 0 and standard deviation σ_{shift}. The value of σ_{shift} was 1 for all simulations presented here. Additionally, we enforce two constraints on the trajectories. First, they must always remain within the bounds of the defined acoustic space. If the modification operation moves points outside of these bounds, they are therefore moved back to the nearest location inside the boundaries. Second, points on the trajectory may never lie more than a distance R apart. If the modification operator moves a point further than this distance from its neighbors, these neighbors are shifted toward the modified point such that their distance becomes equal to R again. This procedure is repeated for all points on the trajectory.

With the modified trajectory, the agent plays repeated imitation games with all other agents in the population (called *imitators*). In an imitation game, the initiator produces the modified trajectory with noise added to it. The imitator then selects the trajectory in its repertoire that is closest to this, and in turn produces it while adding noise. The initiator then checks whether the trajectory in its repertoire that is closest to this is



(a)

(b)

(c)

Figure 2 System of trajectories (a) before and (b) after 60,000 generations. Each cluster in (b) contains a trajectory for each agent in the population. Note how trajectories become bunched up in the corners. (c) The success over the iterations of a population of 10 agents using four trajectories each.

the same as the trajectory that was originally selected to start the game. If this is the case, the imitation game is successful, else it is a failure. In this way, 50 imitation games are played with each other agent in the population. Finally, the number of successful games is divided by the total number of games played. This number ($s_{modified}$) is compared with the success score ($s_{original}$) of the selected trajectory. If it is lower, the modified trajectory is discarded, and the original trajectory is kept. If the score is higher, the original trajectory is averaged with the modified trajectory and this is stored in the agent's repertoire. The calculation is as follows:

$$t_{new,\,i} \leftarrow \beta t_{original,\,i} + (1 - \beta) t_{modified,\,i}, \qquad (3)$$

where β is a weighting constant, whose value was set to 0.5 in the simulations presented here. In the case of both success and failure, the success score of a trajectory is updated similarly:

$$s_{new} \leftarrow \beta s_{original} + (1 - \beta) t_{modified}, \tag{4}$$

where $s_{modified}$ is the success of the modified trajectory in the imitation games. This procedure is repeated for a predetermined (but large) number of iterations. Trajectories eventually converge to a local optimum. For ease of reimplementation all the important algorithms are described in pseudocode in the Appendix. The notation is adopted from Cormen, Leiserson, and Rivest (1993).

3 Results

Running the system results in increasing average success of imitation as well as increasing structure in the repertoire of trajectories that is used in the population. All simulations were run with a population of 10 agents and with a noise standard deviation (σ_{noise} in the equations) of 2. The first thing to note is that success increases over the iterations and converges to an asymptotic value, as illustrated in Figure 2. This figure shows the running average of success (calculated as $\bar{x}_t \leftarrow 0.999\overline{x_{t-1}} + 0.001x_t$, where x_t is the success at time t and $\bar{x}_t$ is the running average at time t) for a typical run in which there were four trajectories per agent. Between 3,000 and 30,000 iterations, the success rises almost linearly, after which a plateau is reached and maintained. For larger numbers of trajectories, success rises slower and a somewhat lower final value is reached. This is to be expected, as only one trajectory gets adapted per iteration, so adapting more trajectories takes more iterations. Also, given a limited acoustic space and a fixed noise level, confusion probability is expected to be greater for larger numbers of trajectories, and success correspondingly lower.

More interesting for the purpose of this article is what happens to the configuration of the trajectories. In Figure 2, the initial (random) trajectories of all agents are shown in the acoustic space, as well as the trajectories after 60,000 generations. The trajectories form clusters with different shapes and positions. Each agent in the population has a trajectory in each cluster. It is found that the four trajectories of each agent bunch up in the four corners of the acoustic space. This seems understandable, as in this way, trajectories within a cluster are close together, while distance between clusters is maximized. Both properties contribute to higher success in imitation.

The question then arises what would happen if a fifth trajectory were added to the repertoires. It is conceivable that a structure would emerge in which the fifth trajectory is bunched up in the center of the acoustic space, giving a structure similar to the five dots on the face of a die. This would be analogous to the structures found when optimizing vowel systems, where the acoustic distances between vowels are maximized (for instance: de Boer, 2000, 2001; Liljencrants & Lindblom, 1972; Oudeyer, 2001, 2005). However, this is not what is found in our experiments.

When adding a fifth trajectory to the repertoire, it generally becomes stretched out from one corner of the acoustic space to another. This is shown in Figure 3. It can be observed that four trajectories are still bunched up in the corners of the acoustic space, but that the fifth trajectory is now on the diagonal. For clarity the repertoire of a single agent is shown in Figure 3b. This shows the identical structure.

In fact, with hindsight it is understandable that trajectories would stretch out over the diagonal, as this results in a larger distance between the points on the stretched out trajectory and the trajectories in the corners than for a trajectory bunched up in the center. The average distance of points on the trajectory to the two corners it visits remains equal, while the average distance to the two corners it does not visit increases (Zuidema, 2005).

The situation is less clear for larger numbers of trajectories. Therefore the simulation was run for other numbers of trajectories as well. As the complexity of the imitation games increases proportionally to the square of the number of trajectories, and the number of games it takes to converge increases at least linearly with the number of trajectories per agent, the total time for the simulations to run increases with at least the cube of the number of trajectories. Running these simulations is therefore very time consuming, and only six and 10 trajectories were tried. These are shown in Figure 4. The system with six trajectories uses two diagonal trajectories in addition to the four trajectories in the corners. The system with 10 trajectories is more complicated, and therefore an individual agent is shown in the rightmost frame of the figure. It can be observed that all trajectories make

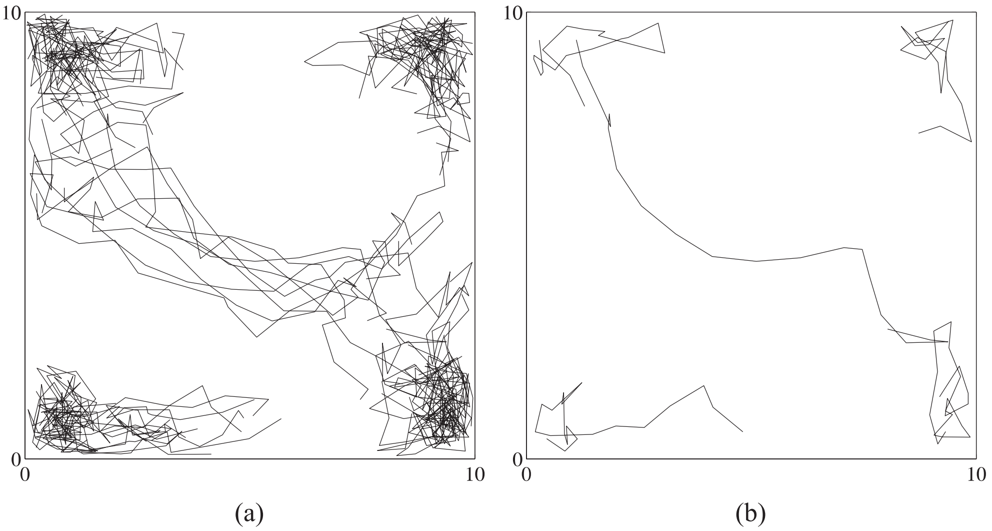


Figure 3 Population of 10 agents with five trajectories each after 60,000 generations. (a) Shows the whole generation, (b) shows the repertoire of one agent from the population.

use of the corners of the acoustic space, and that trajectories that make use of the same two corners, have different directions. It is also noteworthy that there is no trajectory bunched up in the upper right corner. Apparently stretched out trajectories are sufficiently distinctive.

In order to get more than an impressionistic view of how combinatorial structure emerges, a quantitative measure is needed. However, quantitative measures of the degree of combinatoriality appear to be non-existent in both the linguistic and animal communications literature. This is understandable, as all human languages have combinatorial structure, and therefore measuring the degree of recombination is not interesting linguistically. Animal communication systems, on the other hand do not tend to have combinatorial structure, so measuring it is also not usually done by biologists. The few efforts that have been made (Mitani & Marler, 1989) depended on measuring similarities in the (spectrograms of) observed signals by hand. This is not satisfactory for the present purpose.

A quantitative measure of combinatorial structure would ideally be based on the ratio between the number of building blocks and the number of utterances in the system. This reduces the question to finding the building blocks of a repertoire of signals. Unfortunately, there are no readily usable techniques for this, either. In linguistics, building blocks for speech are defined by using a combination of knowledge of articulatory gestures and *minimal pairs*. Minimal pairs are pairs of words that differ only in one sound, but have different meanings. As this procedure depends both on the knowledge of possible articulatory gestures and on meaning, it cannot be directly applied to our system.

However, the notion of trajectories moving in a predictable way from building block to building block can be used to define a crude measure of the amount of combinatorial structure that exists in the system. If it is assumed that trajectories move in a (more or less) straight line from building block to building block, then in a combinatorial system start- and endpoints of trajectories are expected to lie closer together on average than intermediate points on the trajectories.

Here a measure of phonemicity that is based on these considerations is proposed. The average weighted distance between start- and endpoints is calculated, as well as the average weighted distance between other points on a trajectory. A non-linear weighting function is used to give disproportionally higher weight to points that are close together. Another weighting function is used to give higher weight to points on a trajectory that are further away from the start- and endpoints.

The average weighted distance between start- and endpoints, E, is calculated as follows:

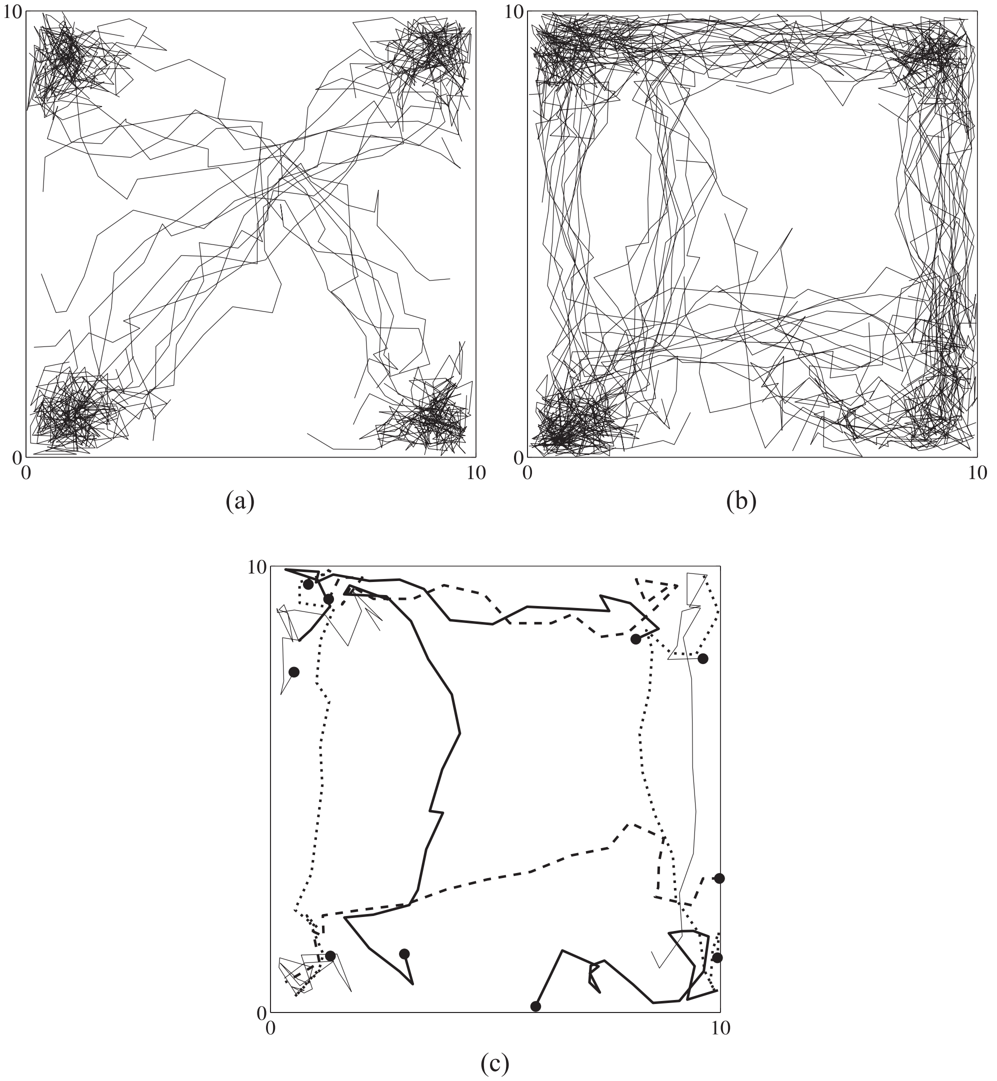


Figure 4 Results for larger number of trajectories. Shown are a system with six trajectories (a), a system with 10 trajectories (b) and the repertoire of an agent from the population with 10 trajectories. Trajectories in (c) have been given different styles to make them easier to follow. Note that trajectories that appear to follow a similar path have starting points (indicated with a black dot) in different corners.

$$E = \sum_{i=1}^{M}\sum_{j=i+1}^{M} W(\|t_{i,1}-t_{j,1}\|) + W(\|t_{i,1}-t_{j,N}\|) + W(\|t_{i,N}-t_{j,1}\|) + W(\|t_{i,N}-t_{j,N}\|)$$

where M is the number of trajectories, N is the number of points on a trajectory, and $t_{i,j}$ is point j on trajectory i.

The average weighted distance between intermediate points is calculated as follows:

$$D = \sum_{i=1}^{M}\sum_{j=i+1}^{M}\sum_{k=1}^{N} V(k)\Big(W(\|t_{i,k}-t_{j,N-k+1}\|) + W(\|t_{i,N-k+1}-t_{j,1}\|)\Big)$$

The weighting functions are as follows:

$$W(d) = 1 - erf(2d)$$

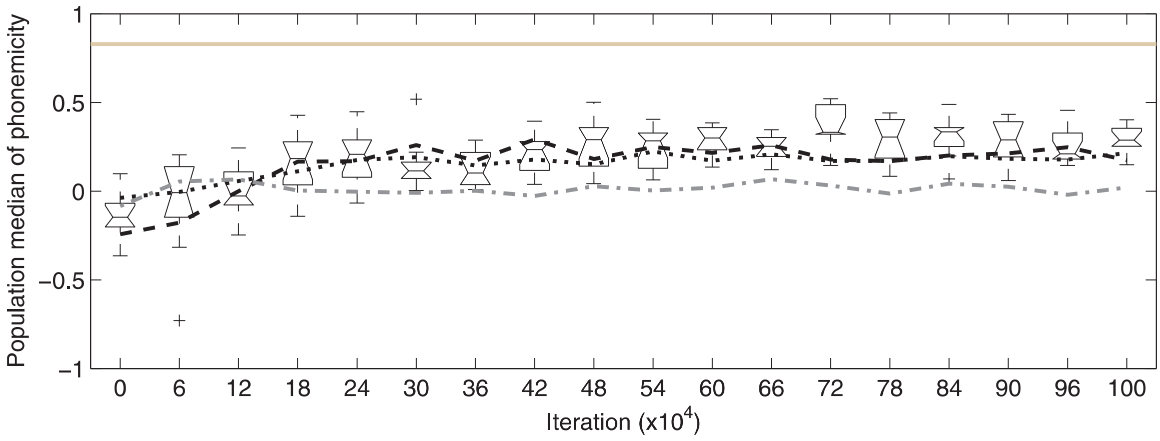


Figure 5 Boxplots of phonemicity measured for 10 runs of one million imitation games for populations of 10 agents with nine trajectories each. Notches in the box plots indicate the intervals for 5% significance. For reference, the phonemicity of an "ideal" system of nine trajectories (described in the text) is indicated as a gray line. Also shown are the medians of phonemicity for systems of six (dashed line) and 16 (dotted line) trajectories, as well as the phonemicity of nine trajectories in a circular space (gray dash-dotted line).

for the weighting of distances d, where erf(x) is the *error function*, the cumulative probability density function corresponding to the normal distribution. This function approximately models the probability of confusion of two trajectories when normally distributed noise is added.

The weighting of intermediate points is weighted as follows:

$$V(k) = \frac{(k-1)(N-k)}{\sum_{i=1}^{N}(i-1)(N-i)}$$

Finally, phonemicity is calculated as the logarithm of the ratio between E and D: $P = \ln\frac{E}{D}$.

This measure was calculated over 10 runs of a system with 10 agents that each had nine trajectories. For each run, one million imitation games were run. The results are presented in Figure 5.

In this figure it can be seen that phonemicity increases over time, and levels off after about 500,000 imitation games. Phonemicity levels off below the value for an ideally holistic system of nine trajectories (defined as a system that uses three corners in all their nine combinations) but higher than the phonemicity value for the original random systems. The increase in phonemicity appears to be independent of the number of trajectories, but the shape of the space in which the trajectories exist appears to be crucial. For square spaces, phonemicity increases, but for the circular space, no significant increase is observed. Some structure does appear to emerge (as illustrated in Figure 6) but according to our measure, this is not more phonemic than random structure. Whether this illustrates a weakness in the measure, or a real lack of structure in a space without corners, remains to be investigated. In any case, the phonemicity measure is still quite crude, as there is a rather large amount of fluctuation, and values still depend on the length of the trajectories and the size of the space. However, it is clear from these results that combinatorial structure increases because of the optimization for imitative success that occurs over time.

4 Conclusion

All experiments with more than four trajectories resulted in repertoires with trajectories that were stretched out through the available acoustic space. Only the corners were occupied by trajectories that were bunched up into point-like signals. Moreover, the trajectories in the repertoires reused the same start- and endpoints.

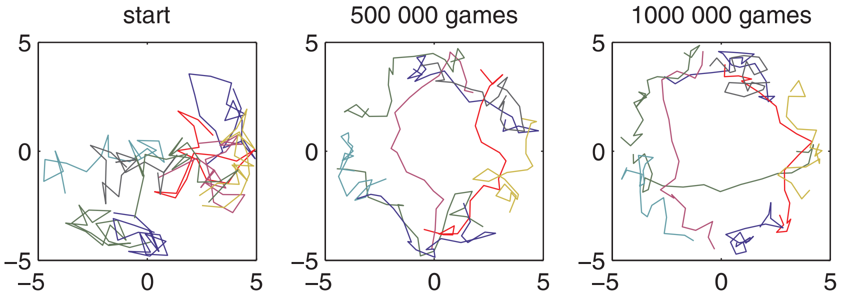


Figure 6 Development of one agent's trajectories in a circular space. Note that there appears to be some structure in the system in that start- and endpoints of trajectories tend to occur close together. Also note that the system remains quite stable between games 500,000 and 1,000,000.

They can therefore be considered to be constructed from a limited number of building blocks. In this sense, they are combinatorially coded. Using a quantitative measure of the degree of combinatorial structure, it was shown that in systems consisting of nine trajectories, the amount of combinatorial structure increases over time, and approaches the value of the optimally combinatorial system.

The emerged repertoire with 10 trajectories that is shown in Figure 4, can be analyzed as having four basic building blocks: the four corners of the acoustic space. The 10 trajectories can be coded by their starting corner and their ending corner. The trajectories then become: tl-tl, tl-tr, tl-bl, tr-tl, tr-br, bl-bl, bl-tl, br-br, br-tr, and br-tl (tl is top left, tr is top right, bl is bottom left and br is bottom right). Although some information about the actual path of the trajectory is lost in this way, these descriptions are sufficient for playing successful imitation games. The system of 10 trajectories can therefore be considered to be built up from four phonemes.

It must be stressed that the agents that use these repertoires of trajectories are not aware of this. They store, perceive, and produce the trajectories in a completely holistic way. The repertoire is therefore only *superficially* combinatorial.

As the structure is present in the repertoire of trajectories, however, it becomes advantageous for agents to exploit it. Agents can evolve learning mechanisms that learn utterances in a combinatorial way instead of holistically. In this way, agents can evolve toward using combinatorial structure productively, as in human speech.

In this way, the (biological or cultural) evolution of sound systems toward superficial combinatorial structure through the pressure of increasing acoustic distinctiveness allows biological evolution of the learning mechanisms to take place. Even though an agent might be the only one in the population to actively use combinatorial structure, it will still have an advantage over the other agents, as the combinatorial structure is already superficially present. This is completely different from the situation where an agent mutates toward the ability to use combinatorial structure in a population where no speech with combinatorial structure is used.

One important point that could be improved is the acoustic space that was used. In the experiments it is a square without internal structure. It would be interesting to investigate spaces that are more plausible, both from the perspective of production and perception. This would alter the shape of the allowed space. As the emerged trajectories in the experiments presented here appear to exploit the shape of the available space (trajectories tend to use corners) it is likely that trajectories in a more complex space would exploit its features to maximize distinctiveness. These features could conceivably be used to explain universal properties of human sound systems.

We have explored such more complex spaces using direct optimization of systems of trajectories (Zuidema & de Boer, 2009). No interactions between language users were modeled in that article, but similar systems of superficially combinatorial trajectories were obtained by optimizing systems of trajectories for acoustic distinctiveness directly. This was computationally less demanding and many more different settings could therefore be explored. Most notably, systems that existed in a three-dimensional space also showed the emergence of combinatorial structure (Figure 9b in Zuidema & de Boer, 2009). As the results from optimization appear to follow the same pattern as the results from the agent-based models, it is probably safe to conclude that agent-based models would also result in combinatorial systems under all the different conditions explored by optimization. However, pursuing this might be an avenue of future research.

An interesting but difficult question concerns the details of the effective acoustic space relevant for human evolution: that is the space that would realistically model the range of sounds early hominins could produce and the similarities between those sounds as they would perceive them. Would formants form important dimensions in that space, as they do in modern speech perception? Or would pitch or other functions of spectral energy be more salient? Answers to such questions might come from detailed modeling of hominin vocal tracts (e.g., de Boer, 2008, 2009) combined with data about primate auditory perception. However, in this article we avoid such questions by focusing on an abstract acoustic space and on results that are robust to the particular shape of that space.

In the most basic case investigated in this article, it has been shown that trajectories that are maximized for distinctiveness develop a structure that can be interpreted as combinatorial and that can be exploited by agents that learn speech combinatorially. This creates an interaction between cultural evolution of a repertoire and biological evolution that allows combinatorial learning of speech to evolve.

Appendix: Pseudocode

This appendix contains the pseudocode representation of the most important algorithms that were used in this study and that have been described in the text. Algorithms that were considered too obvious (such as finding the closest trajectory to another trajectory) have been omitted. The basic loop of the algorithm is called DISTRIBUTED OPTIMIZATION and all other functions are called from there. Notational conventions follow those of mathematics and computer science. Function names are written in small caps, variable names in italics, and control words in bold. Assignments are indicated with a left arrow, while comparisons are indicated with an equals sign. Properties (or members in object-oriented programming parlance) of variables are indicated with *variable.property* notation. Comments are preceded with a % sign.

```
DISTRIBUTED OPTIMIZATION($A$)
% A is the population of agents
repeat $N_{games}$ times
    initiator ← random agent from $A$
    $T_{original}$ ← random trajectory from initiator
    $T_{shifted}$ ← SHIFT TRAJECTORY($T_{original}$, $\sigma_{shift}$)
    $T_{shifted}.success$ ← 0
    replace $T_{original}$ with $T_{shifted}$ in initiator
        repeat $N_{test}$ times
            imitator ← random agent from $A$, different from initiator
        $T_{shifted}.success$ ← $T_{shifted}.success$ + PLAY GAME(initiator, $T_{shifted}$, imitator)
        end repeat
        if $T_{shifted}.success/N_{test} < T_{original}.success$
            restore $T_{original}$ in initiator
        else
            $T_{new}$ ← MIX TRAJECTORIES($T_{original}$, $T_{shifted}$, $\alpha$)
                $T_{new}.success$ ← $\beta \cdot T_{original}.success + (1 - \beta)T_{shifted}.success/N_{test}$
                    replace $T_{shifted}$ with $T_{new}$ in initiator
        end if
end repeat
```

```
$T_{out} \leftarrow$ SHIFT TRAJECTORY($T_{in}$, $\sigma_{shift}$)
% $T_{in}$ is the trajectory to be shifted, $\sigma_{shift}$ is the standard deviation of shift noise
% G($\mu$, $\sigma$) is 2-dimensional Gaussian noise with mean $\mu$ and standard deviation $\sigma$
$i \leftarrow$ random point on $T_{in}$
$T_{out}.i \leftarrow T_{in}.i + \mathrm{G}(0, \sigma_{shift})$
for $j \leftarrow i$ to $N_{points} - 1$
    if $|T_{out}.j - T_{in}.[j + 1]| > maxdist$
        $T_{out}.[j + 1] \leftarrow maxdist/|T_{out}.j - T_{in}.[j + 1]|\ (T_{in}.[j + 1] - T_{out}.j)$
    else
        $T_{out}.[j + 1] \leftarrow T_{in}.[j + 1]$
    end if
end for
for $j \leftarrow i$ down to 1
    if $|T_{out}.j - T_{in}.[j - 1]| > maxdist$
        $T_{out}.[j - 1] \leftarrow maxdist/|T_{out}.j - T_{in}.[j - 1]|\ (T_{in}.[j - 1] - T_{out}.j)$
    else
        $T_{out}.[j - 1] \leftarrow T_{in}.[j - 1]$
    end if
end for
```

```
$T_{out} \leftarrow$ MIX TRAJECTORIES($T_1$, $T_2$, $\beta$)
% Trajectories $T_1$ and $T_2$ are mixed with weighing parameter $\beta$
for all points $i$ on $T_1$
    $T_{out}.i \leftarrow \beta \cdot T_1.i + (1 - \beta)T_2.i$
end for
```

```
$success \leftarrow$ PLAY GAME($initiator$, $T_{init}$, $imitator$)
% $initiator$ and $imitator$ are agents that play the game
% $T_{init}$ is a trajectory from $initiator$'s repertoire with
% which the game is played
$T_{said} \leftarrow$ ADD NOISE($T_{init}$, $\sigma_{noise}$)
$T_{imit} \leftarrow$ $imitator$'s closest trajectory to $T_{said}$
$T_{resp} \leftarrow$ ADD NOISE($T_{imit}$, $\sigma_{noise}$)
$T_{succ} \leftarrow$ $initiator$'s closest trajectory to $T_{resp}$
if $T_{succ} = T_{init}$
    $success \leftarrow 1$
else
    $success \leftarrow 0$
end if
```

```
$T_{noise} \leftarrow$ ADD NOISE($T_{pure}$, $\sigma_{noise}$)
% Add shape preserving noise to trajectory $T_{pure}$
% G($\mu$, $\sigma$) is 2-dimensional Gaussian noise with mean $\mu$ and standard deviation $\sigma$
$S_{start} \leftarrow \mathrm{G}(0, ?\sigma_{noise})$
$S_{end} \leftarrow \mathrm{G}(0, ?\sigma_{noise})$
for $i \leftarrow 1$ to $N_{points}$
    $T_{noise}, i \leftarrow T_{pure}, i$
                $+ S_{start}(N_{points} - i)/(N_{points} - 1)$
                $+ S_{end}(i - 1)/(N_{points} - 1)$
                $+ \mathrm{G}(0, \sigma_{noise}/N_{points})$
end for
```

Acknowledgments

WZ is funded through a Veni-fellowship (639.021.612, Exacte Wetenschappen) and BdB through a Vidi-fellowship (016.074.324, Geesteswetenschappen) of the Netherlands Organisation for Scientific Research. We thank our anonymous reviewers for their comments.

References

Arcadi, A. (1996). Phrase structure of wild chimpanzee pant hoots: patterns of production and interpopulation variability. *American Journal of Primatology*, *39*, 159–178.

Catford, J. C. (1977). Mountain of tongues: the languages of the Caucasus. *Annual Review of Anthropology*, *6*, 283–314.

Cavalli-Sforza, L. L., & Feldman, M. W. (1983). Paradox of the evolution of communication and of social interactivity. *Proceedings of the National Academy of Sciences of the USA*, *80*, 2017–2021.

Cormen, T. H., Leiserson, C. E., & Rivest, R. L. (1993). *Introduction to algorithms*. Cambridge, MA: MIT Press.

Crockford, C., Herbinger, I., Vigilant, L., & Boesch, C. (2004). Wild chimpanzees produce group-specific calls: a case for vocal learning? *Ethology*, *110*, 221–243.

de Boer, B. (2000). Self organization in vowel systems. *Journal of Phonetics*, *28*, 441–465.

de Boer, B. (2001). *The origins of vowel systems*. Oxford: Oxford University Press.

de Boer, B. (2008). The joy of sacs. In A. D. M. Smith, K. Smith, & R. Ferrer i Cancho (Eds.), *The evolution of language* (pp. 415–416). New Jersey: World Scientific.

de Boer, B. (2009). Why women speak better than men (and its significance for evolution). In R. Botha & C. Knight (Eds.), *The prehistory of language* (pp. 255–265). Oxford: Oxford University Press.

Fitch, W. T. (2000). The evolution of speech: a comparative review. *Trends in Cognitive Science*, *4*, 258–267.

Kirby, S., & Hurford, J. R. (2001). The emergence of linguistic structure: An overview of the iterated learning model. In A. Cangelosi & D. Parisi (Eds.), *Simulating the evolution of language* (pp. 121–148). London: Springer Verlag.

Kuhl, P. K., Williams, K. A., Lacerda, F., Stevens, K. N., & Lindblom, B. (1992). Linguistic experience alters phonetic perception in infants by 6 months of age. *Science*, *255*, 606–608.

Liljencrants, J., & Lindblom, B. (1972). Numerical simulations of vowel quality systems. *Language*, *48*, 839–862.

Lindblom, B., MacNeilage, P., & Studdert-Kennedy, M. (1984). Self-organizing processes and the explanation of language universals. In M. Butterworth, B. Comrie, & O. Dahl (Eds.), *Explanations for language universals* (pp. 181–203). Berlin: Walter de Gruyter & Co.

MacNeilage, P. F., & Davis, B. L. (2000). On the origin of internal structure of word forms. *Science*, *288*, 527–531.

Mitani, J. C., & Marler, P. (1989). A phonological analysis of male gibbon singing behavior. *Behaviour*, *109*, 20–45.

Nowak, M. A., & Krakauer, D. (1999). The evolution of language. *Proceedings of the National Academy of Sciences of the USA*, *96*, 8028–8033.

Oliphant, M., & Batali, J. (1996). Learning and the emergence of coordinated communication. *Center for Research on Language Newsletter*, *11*(1), 1–46.

Oudeyer, P.-Y. (2001). Coupled neural maps for the origins of vowel systems. In G. Dorffner & K. H. Bischof (Eds.), *Proceedings of the International Conference on Artificial Neural Networks*, Lecture notes in computer science (Vol. 2130, pp. 1171–1176). Berlin: Springer Verlag.

Oudeyer, P.-Y. (2002). Phonemic coding might be a result of sensory-motor coupling dynamics. In J. Hallam (Ed.), *Proceedings of the International Conference on the Simulation of Adaptive Behavior (SAB)* (pp. 406–416). Cambridge, MA: MIT Press.

Oudeyer, P.-Y. (2005). The self-organization of speech sounds. *Journal of Theoretical Biology*, *233*, 435–449.

Sakoe, H., & Chiba, S. (1978). Dynamic programming optimization for spoken word recognition. *IEEE Transactions on Acoustics, Speech and Signal Processing*, *26*, 43–49.

Smith, K., Kirby, S., & Brighton, H. (2003). Iterated learning: a framework for the emergence of language. *Artificial Life*, *9*, 371–386.

Steels, L. (1997). The synthetic modelling of language origins. *Evolution of Communication*, *1*, 1–34.

Steels, L. (1998). Synthesising the origins of language and meaning using co-evolution, self-organisation and level formation. In J. R. Hurford, M. Studdert-Kennedy, & C. Knight (Eds.), *Approaches to the evolution of language* (pp. 384–404). Cambridge: Cambridge University Press.

Traill, A. (1985). *Phonetic and phonological studies of !Xóõ bushman*. Hamburg: Helmut Buske Verlag.

Zuidema, W. (2003). How the poverty of the stimulus solves the poverty of the stimulus. In S. Becker, S. Thrun, & K. Obermayer (Eds.), *Advances in neural information processing systems 15* (pp. 51–58). Cambridge, MA: MIT Press.

Zuidema, W. (2005). *The major transitions in the evolution of language*. Unpublished doctoral dissertation, University of Edinburgh.

Zuidema, W., & de Boer, B. (2009). The evolution of combinatorial phonology. *Journal of Phonetics*, *37*(2), 125–144. (0095-4470 doi: DOI: 10.1016/j.wocn.2008.10.003)

About the Authors

Bart de Boer studied computer science at the Rijksuniversiteit Leiden and did a PhD on modeling self-organization and the evolution of vowel systems at the Vrije Universiteit Brussel. He now works as a postdoctoral researcher at the Amsterdam Center for Language and Communication of the University of Amsterdam, using computer models, biological data, and human subject experiments to study the evolution of speech.

Willem Zuidema received his PhD in linguistics from the University of Edinburgh (2005). Since 2004 he has worked as a postdoctoral researcher at the Institute for Logic, Language and Computation at the University of Amsterdam, primarily on models of language learning and evolution. From 2007 until 2009 he also worked at the department of Behavioural Biology, Leiden University. He is a member of the Cognitive Science Center Amsterdam and teaches the MSc course Brain & Cognitive Science. In 2007 he received a 3-year NWO-Veni fellowship for the project Discovering Grammar to study the cognitive mechanisms underlying sequence learning in humans and other species.