

Simulated Evolution in a Linguistic Model

CARSTEN KNUDSEN^{a,*} and LYNNE CAMERON^b

^a*Department of Physics, Technical University of Denmark, DK-2800 Kongens Lyngby, Denmark;*

^b*School of Education, University of Leeds, Leeds LS2 9JT, England*

(Received 15 March 2000)

In this paper we present a simple evolutionary model of childrens' language development, whose central nonlinearity is represented by noninvertible discrete dynamical systems. The underlying assumption of the model is that children learn from other children through their interactions. The concrete learning mechanism used is based on imitation, where childrens' languages evolve through attempting to imitate other childrens' utterances. The use of imitation in evolutionary models has been used, for instance, in evolution of bird song by Kaneko and Suzuki. The model to be presented here is similar to Kaneko and Suzuki's model, the primary difference being the continuous nature of bird song, in contrast to the discrete nature of childrens' utterances.

Keywords: Learning; Speech generation; Imitation; Evolution; Noninvertible mappings; Chaos

1. INTRODUCTION

Recent work in a complex systems approach is highlighting the co-evolution of vocabulary and grammar in child language [1], and the interdependence of vocabulary and syntax in language learning modelled by simple recurrent networks in a connectionist approach [5].

Before proceeding we shall introduce some notation, by the following definitions: A vocabulary is a fixed collection of words; an utterance is a sequence of words generated by a child; the language of a child is the set of utterances that can be generated by the child; and finally, an interac-

tion is when two children talk together. It should be noted that in our simple model neither words nor utterances have any particular meaning associated with them (apart from an inherent ordering; see later). Improvements on the model would obviously have to address the issue of including some meaning, for example, through the inclusion of a grammar.

In order to keep the model as simple as possible we will make the following two basic assumptions about childrens' development of simple language. We stress that the assumptions are meant as suggestive and pedagogical rather than as a full scale model of the complex phenomena involved in

*Corresponding author.

real life language development. Firstly, it is assumed that children learn through interaction with other children, that is, the learning is an instance of unsupervised learning, contrasted by neural networks which normally learn under supervision [17]. Secondly, the primary learning is through imitation of other children. Our model is similar to the bird song model by Kaneko and Suzuki [10].

The model is reduced from the complications of real life language use and learning to a few important qualitative features: basically, speech generation and learning through reproducing the speech of others heard in interaction. Current views of both first and second language learning emphasise the importance of social interaction in the development [13, 14]. The model incorporates key aspects of second language learning in which both input to the learner and output produced by the learner are necessary [15]. The input to the child in the model arises through interaction, albeit of a very simplified form that downplays the importance of comprehensibility or understanding to learning [12, 13].

The co-adaptive language development being modelled might be seen as either (a) first language development after about three years of age, such as might occur in nursery or pre-school classes, or (b) classroom foreign language learning through pair or group work. The major difference between these two contexts of language development, as far as the constituents of the model are concerned, lies in the size of the vocabularies brought to the learning task. In the first language context, the vocabularies of pre-school children are estimated at 3000 words or beyond [14]; in a typical foreign language context, this level of vocabulary would take at least four or five years of tuition to reach. The vocabulary used in the present model is very small relative to the first language figures, but might be considered to represent a subset of vocabulary for a particular group of children, perhaps referring to a relatively unfamiliar semantic field that is met only through formal schooling, such as weight and mass.

The modelling process will involve defining the development of the following three elements: a model of a child; a model of interaction between two children; and a model of the learning process. Having defined these elements we can continue with exposing the model to simulated evolution. This will consist of letting a group of children interact, and letting (some of) the children learn from these interactions. After this we can then evaluate the results qualitatively as well as quantitatively. First, we shall simply observe characteristic features of the evolutionary processes, such as determining what groups of children perform well, and describing what stable (and unstable) evolutionary structures are present in the system; then we will evaluate the complexity of the languages involved using finite state machines as models of computation.

In the model, a child will be capable of listening to other children (though only one at a time), to think about what it has heard, or rather is currently hearing, as well as generate speech. Thus, the child is merely a specific instance of a much more general situation: An entity capable of receiving information, processing information, and communicating information, *i.e.*, the components characterising a computer.

The organisation of this paper is as follows. In Section 2 the mathematical model is derived and formulated. Section 3 introduces the fitness landscape and examples are given. Then in Section 4 we present the results of simulated evolution in the model by describing the evolved structures. In Section 5 we evaluate the complexity of the languages developed and interpret the results in this context. Finally, Section 6 contains a discussion of the presented results.

2. THE MODEL

Let us briefly go through the main model components, listening and speaking. The first component is the listening device. This determines how much we listen to ourselves and others. In one

end we mainly ignore others and essentially we have autonomous speech. This is referred to as low coupling. At the other end we mainly listen to others and ignore our own internal state; this is called high coupling. For the speech generation we need to know what we say in the absence of input, *i.e.*, autonomous generation of speech. We shall pick the speech generation process as the model's primary nonlinearity, and most interpretations will be in terms of this parameter. The model of a single child will be characterised by three main parameters, one determining the speech process: a parameter $p \in [0, 1]$ (which will be referred to as the parameter); one determining how much the child listens to others during imitation: a coupling $\varepsilon \in [0, 1]$. Finally, it is also necessary to specify how the child starts to speak autonomously through a state $x_0 \in [0, 1]$. In the present model this is always chosen randomly (*i.e.*, x_0 in Eq. (2) below is always chosen by a pseudo random number generator).

To simplify matters we will let each child have access to the exact same vocabulary consisting of, say, 20 words. As an example the vocabulary could consist of the following words:

and, dog, not, cat, why, pat, sit, bat, let, say, has, him, her, dot, wet, hot, hat, run, all, win.

The only significance of the words is that they are ordered as we have just listed them. That means that we consider the word pairs (and, dog), (sit, bat), and (hot, hat) as being close, and similarly the word pair (and, win) is considered distant. The number of words in the vocabulary is a parameter in the model and the effect of changing it will be considered. Although all children have the same basic vocabulary they may not all be able to use all of the words in the sentence, especially the words at the end of the list are not used as much as those in the middle (the explanation for this lies in the used speech generation process outlined below).

Let us now explain the autonomous discrete speech generation (as in the case of a child being asked to say "something"). The process must

generate a sequence of discrete elements, *i.e.*, words taken from the vocabulary. Discrete state systems are then a natural choice, *e.g.*, finite state machines, Turing machines and cellular automata. However, we have chosen to use a continuous variable, namely by the use of the logistic map, and then apply coarse graining (or symbolic dynamics) to generate utterances, as we describe in the following. We start by generating a sequence of fixed length, say, 20, $\{x_0, x_1, \dots, x_{19}\}$ by the logistic equation

$$x_{n+1} = ax_n(1 - x_n) \quad (1)$$

where a is determined by the parameter p through the relation $a = a_0 + p/2$ where $a_0 = 3.5$ and the initial condition x_0 . This sequence is converted to an integer sequence $\{I_0, I_1, \dots, I_{19}\}$ by transforming each x_i to an integer by $I_i = \text{integer part of } (20x_i)$, *i.e.*, by a coarse graining of the orbit. Finally, the integer I_i is transformed into a word by the vocabulary, *i.e.*, $I_i = 0$ gives and, and $I_i = 6$ gives sit. As an example of the creation of an utterance consider the orbit $\{0.01, 0.16, 0.49, 0.71, 0.82\}$. This is turned into the integer sequence $\{0, 3, 9, 14, 16\}$, which in turn yields the utterance $\{\text{and cat say wet hat}\}$. Let us in passing note that the choice of the logistic map as the underlying dynamical system is rather arbitrary; any system capable of generating complex dynamics could presumably be used.

Having defined the autonomous speech generation we can proceed to define the interaction between two children. The interaction between is an imitation process. A child evaluates its language by attempting to imitate an utterance generated by another child and *vice versa*. This is carried out as follows. The imitated child talks autonomously (using Eq. (1)). The imitating child listens to the other child while simultaneously "talking" but not out loud; this is a transient phase. After a while (an utterance of fixed length) the imitating child stops listening and attempts to imitate the other child by autonomous speech and initial condition as generated by the transient

phase. The roles are reversed and the imitation process repeated. We then assign a score depending on their ability to imitate each other.

The mathematical details of the interaction are as follows. The child that is being imitated generates a sequence autonomously by the recurrence relation

$$y_{n+1} = ay_n(1 - y_n), \quad n = 0, 1, \dots, 18 \quad (2)$$

whereas the child attempting to imitate is being coupled to the child being imitated. The non-autonomous recurrence relation used here is

$$\begin{aligned} x_{n+1} = a((1 - \varepsilon)x_n + \varepsilon y_n) \\ (1 - ((1 - \varepsilon)x_n + \varepsilon y_n)), \quad (3) \\ n = 0, 1, \dots, 8 \end{aligned}$$

For small values of the coupling parameter ε the result very similar to that generated by the autonomous process described in Eq. (2). After the non-autonomous speech the imitating child uses

$$x_{n+1} = ax_n(1 - x_n), \quad n = 9, 10, 11, \dots, 18 \quad (4)$$

where $a = a_0 + p/2$ and ε is the coupling parameter with different parameter and coupling values for the two children.

For future reference, Figure 1 shows the bifurcation diagram for the nonlinear map in Eq. (2). Note that this is a truncated version of the usual logistic map for the chosen parameter range $[0, 1]$. Also note the periodic windows, *e.g.*, the period three window near $p = 0.68$. More information on the behaviour of the logistic map can be found in, *e.g.*, Guckenheimer and Holmes [9] and Collet and Eckmann [2].

The learning process is clearly an important part of the model. If a child performs poorly relative to other children it will learn; if it performs well it doesn't learn from the interactions. We have chosen to let the speech generation process be the only dynamic quantity during evolution, meaning that the coupling remains fixed over time. This is done in order to make it easier to interpretate the results. To make sure that learning is gradual, the learning involves adding or subtracting a random number between 0 and 0.01 to the current value of the nonlinearity parameter. The only exception is when p becomes

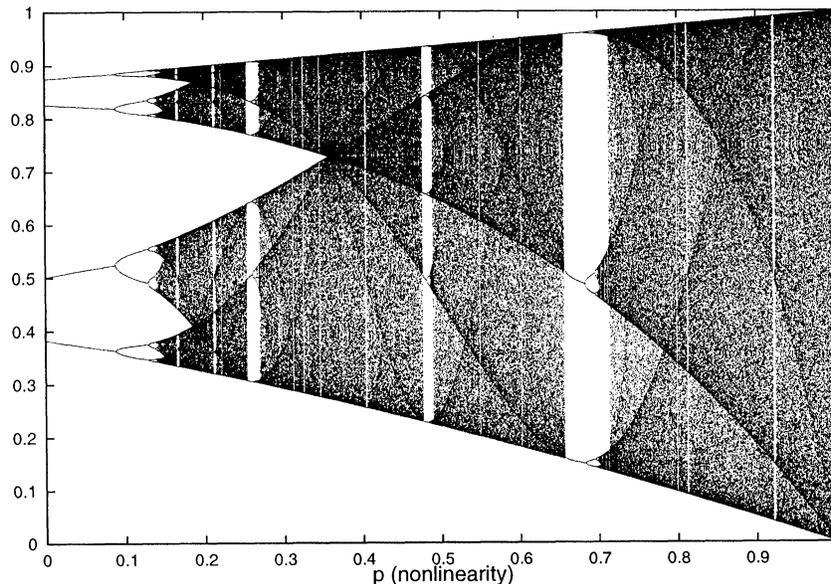


FIGURE 1 Bifurcation diagram for the reparameterized logistic map.

larger than one we subtract one, and similarly when the parameter becomes negative we add one. This effectively places the nonlinearity parameter on the circle, *i.e.*, $p \in S^1$.

As mentioned above we assign a score to the performance of the children in the imitation game. This is done by measuring the quality of the imitation. To do this we define the difference between two utterances by the formula:

$$\sum_{\text{words in utterance}} (\text{distance between words})^2 \quad (5)$$

or more precisely

$$\sum_{k=10}^{19} (\text{integer part of } (20|x_k - y_k|))^2 \quad (6)$$

The better speaking child receives 10 points, the poorer 1 point; in case of a tie both receive 5.5 points. As described in Section 4, in the round robin tournament, dealing with simulated evolution, the scores are normalized such that the best overall gets a unit score.

3. THE FITNESS LANDSCAPE

To represent the performance of the children (or the individual speech generation processes) we can produce a so-called fitness landscape. The fitness landscape represents the performance or fitness of all individuals in a population relative to all other individuals. To determine the landscape we take a population of children and let all children interact with all other children in a round robin tournament. The total score of each child is then normalized with respect to the best score. The fitness landscape is plotted by showing the fitness *versus* the nonlinearity parameter. A general feature of the fitness landscape is that it is in general rugged, see, *e.g.*, Kauffman [11]. We note that the fitness landscape depends on the population, and hence when evolution takes place (as we will include later on), the fitness landscape changes over time. In other words it may be considered the goal of each individual to move around in the changing fitness landscape attempting to optimize its performance.

To test the importance of the size of the vocabulary we have computed the fitness

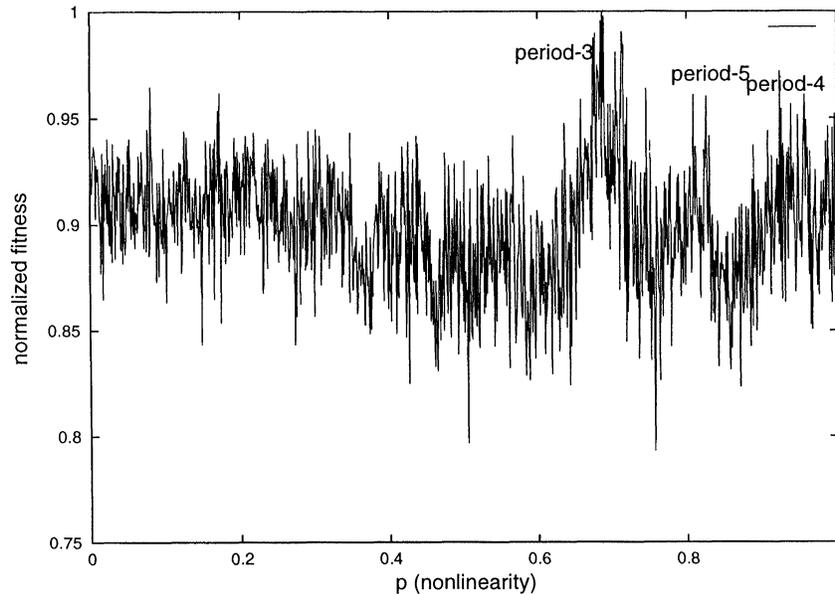


FIGURE 2 Fitness landscape for a vocabulary with two words.

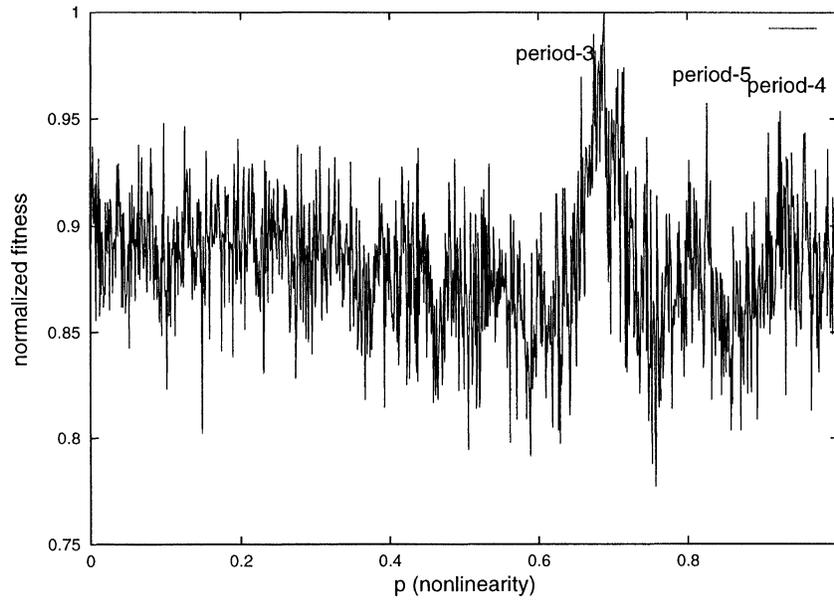


FIGURE 3 Fitness landscape for a vocabulary with twenty words.

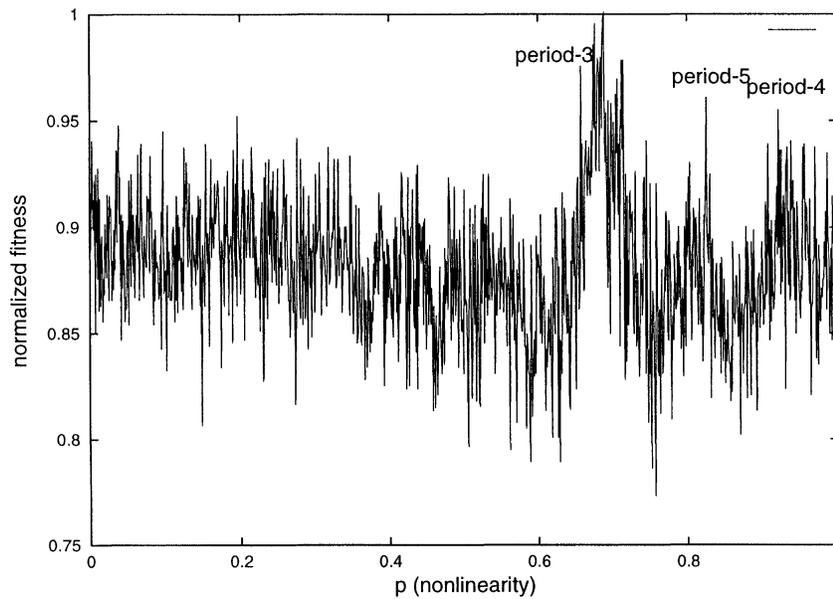


FIGURE 4 Fitness landscape for a vocabulary with one thousand words.

landscape for different vocabulary sizes, and a population containing 1000 individuals (small variations occur depending on the number of

individuals, but the qualitative appearance is the same). Each child is assigned a randomized coupling between 0 and 0.1 which remains fixed.

In Figure 2 we have shown a fitness landscape for a vocabulary with only two words. Note the ruggedness of the landscape. Figure 3 shows a fitness landscape when the vocabulary contains twenty words. There are obviously many differences between the fitness landscapes shown in Figures 2 and 3, but also similarities are present, such as the ruggedness. More important similarities are the peaks in the landscape, for example, the largest peak is close to 0.7, located near the period-3 window (as is also indicated in the figures). Also note the smaller peaks near some of the period-4 and period-5 windows. These peaks are all clearly present in both fitness landscapes. Dramatically increasing the size of the vocabulary to one thousand words yields a fitness landscape as shown in Figure 4. The qualitative similarity with the previous landscapes is clear. Hence we conclude that the system dynamics does not critically depend on the size of the vocabulary. For simplicity we shall therefore fix the size of the vocabulary at twenty words until Section 5 where it will be reduced to two words.

4. SIMULATED EVOLUTION

In this section we will present the results of carrying out simulated evolution in the model. The recipe is very straightforward. We take a uniformly randomized population of children, *i.e.*, they have random nonlinearity and coupling parameter. The children then interact with all other children in a round robin-like tournament. Through each interaction the children accumulate a score determining their fitness in the current population. A ranking is performed and the 10% poorest performers then learn, according to the simple ruled described above. These steps are then repeated over and over. We shall now show some examples of simulated evolution in the model with a population of 100 children and discuss the results.

As we are only including evolution on the nonlinearity parameter, we can plot this parameter against time, *i.e.*, for each, say ten time units (or longer depending on the time scale) we show the parameter for each individual against time. In Figure 5 we have shown the evolution taking place

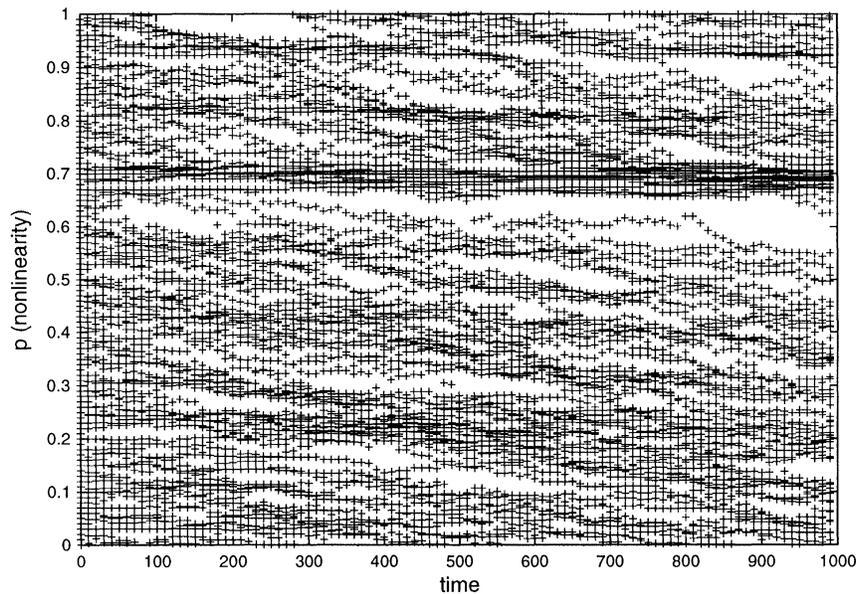


FIGURE 5 Simulated evolution in a population of 100 children.

on a time scale from 0 to 1000. The initial population is uniformly distributed as can be seen on the ordinate axis. Around time 50 we note that there is a depletion of individuals around the parameter 0.6 which lasts until time 200. The explanation for this can be seen in the fitness landscape shown in Figure 3. The average fitness of individuals near this parameter value is rather poor, and hence the individuals located here initially have learned and moved away from this region. Similar depletions are seen many places in the figure, but is most notable in this region where it occurs repeatedly. Another feature we can note is the clustering of individuals, particularly near 0.7. Figure 3 reveals that this is where the main peak is located. Another concentration is found around 0.8 at another peak. This structure also seems to persist on the time scale shown. A smaller clustering can be found near 0.95, but this is less stable, *i.e.*, it can be seen how the individuals here disappear around time 600 only to return later. The reason for the latter being unstable is found in Figure 3 where we can see that there are individuals with poor performance in the region

around 0.95, and hence the noise level (in the evolution) makes this evolutionary structure unstable. Increasing the time horizon to 10000 we obtain the result shown in Figure 6. Here the structure near 0.7 becomes even more pronounced and its stability emphasized. The two other previously mentioned regions are seen to reappear occasionally but are never stable for long. An interesting feature present in this figure but not in the previous one, is the clear trend of individuals below 0.7 moving towards lower values of the nonlinearity parameter. Again Figure 3 provides a clue. It can be seen that for decreasing values of the nonlinearity the average fitness increases slightly, but of course with many fluctuations. Increasing the time horizon once more to 100000 yields Figure 7. Here the structure around 0.7 is still stable, and we conclude that it is stable for all times. The two other regions of clustering of children can be seen to exist once in a while but not for long judged on this time scale. The downwards moving trend is now barely visible which is due to the fact that it now moves downwards with a slope ten times larger, *i.e.*, almost vertically.

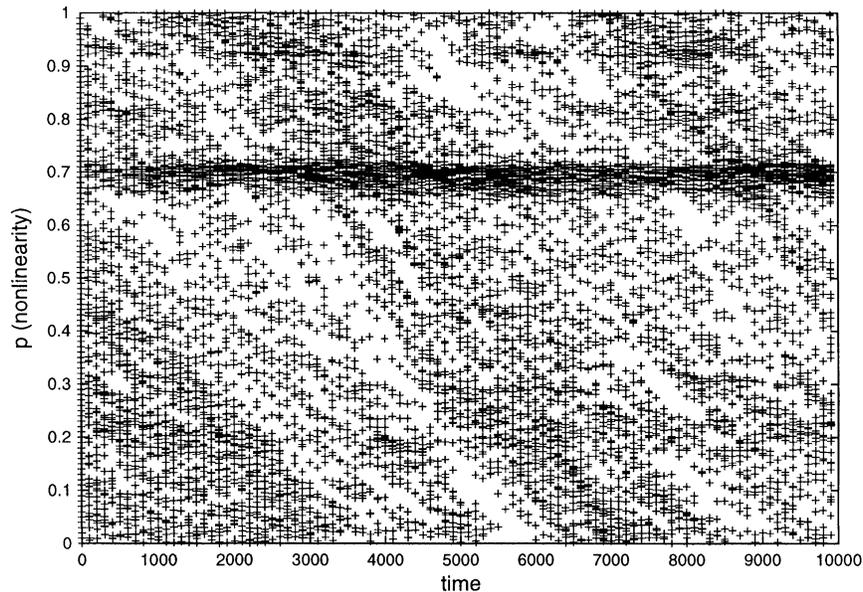


FIGURE 6 Simulated evolution in a population of 100 children.

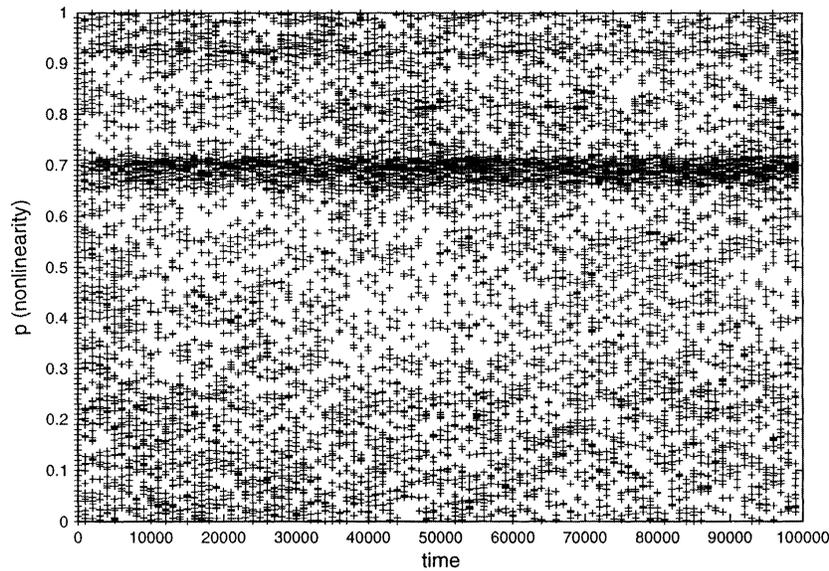


FIGURE 7 Simulated evolution in a population of 100 children.

5. COMPLEXITY OF LANGUAGES

In the previous section we saw that simulated evolution leads to clustering of children, with some clusterings being stable, some unstable. The clusterings were located around specific parameter ranges, which corresponded to the peaks in the uniformly randomized populations fitness landscape (for example, as shown in Fig. 3). These high performance peaks were located near large periodic windows for the underlying logistic map (see Fig. 1). Thus, simulated evolution does tend to produce a clustering around parameter values for the speech generation process that performs well in the imitation process. However, it does not answer any questions about what kind of languages are doing well. To investigate this we shall attempt to evaluate the languages used by the children in the model.

To evaluate the language of a child we need to decide what is meant by the complexity of a language. This is obviously not an easy question to answer satisfactorily, so we shall settle for an operational definition. We will define the language in terms of how difficult it is to recognize the set of

utterances generated by a particular child. In other words we wish to construct a machine that, given an utterance, is capable of determining whether it could have been uttered by that child.

The language of a child can be difficult to describe, and we will approach the problem with a straightforward operational definition. Many children in the model are capable of chaotic word composition and hence can generate infinitely many utterances. Thus to simplify we only consider utterances of fixed length, more specifically we will only consider utterances of length ten. To determine the language we first have the child generate ten utterances, each of length one thousand. These utterances are then searched with a template of length ten for all possible occurrences of strings of length ten (out of the total possible number which is $2^{10} = 1024$). The set of different utterances found is then defined as the language.

As we recall that the fitness landscape does not depend sensitively on the number of words in the vocabulary, we shall for simplicity only consider the case of a binary language.

A language recognizer is a special case of a finite state machine, see, *e.g.*, Grimaldi [8] (more

advanced machine models of the computation of the logistic map are described in Crutchfield and Young [4]). It can formally be described as a triple $\langle S, I, f \rangle$ where S is the set of states of the finite state machine, I is the input alphabet, in this case is the binary alphabet $I = \{0, 1\}$, and $f: S \times I \rightarrow S$ is the function determining the next state as a function of the present state and the input.

Figure 8 shows an example of a simple finite state machine that recognizes the language defined by the binary strings $\{000, 001, 100, 110\}$. In the figure it is clear that the two states 4 and 5 are identical since they give rise to the same next state on the same input, *i.e.*, $f(4,0) = f(5,0)$ and $f(4,1) = f(5,1)$. Collapsing these two states to one single state leads to a smaller finite state machine equivalent to the original. Continuing to collapse identical states is a minimization process that leads to the smallest finite state machine equivalent to the original one, which is easily constructed. We then define the complexity of a language as the number of states in this minimal finite state machine.

To illustrate the minimization procedure we shall now minimize the finite state machine shown in Figure 8. As mentioned above states 4 and 5 are equivalent and hence can be collapsed. The same is true for state 3 and the acceptance state (note that we haven't included the arrows on the error and

acceptance states: the arrows should point to the states themselves). Reducing the recognizer we obtain the finite state machine shown in Figure 9. Considering states 1 and 4 we can see that they map to the same states under the same input, and hence they can be collapsed into one single state. This yields the finite state machine depicted in Figure 10. We observe that this machine cannot be reduced any further, and hence it represents the minimal finite state machine capable of recognizing the language $\{000, 001, 100, 110\}$.

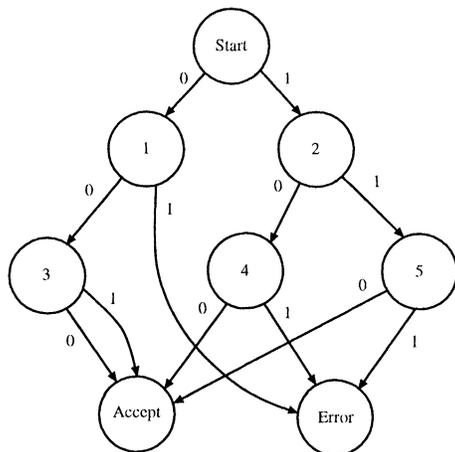


FIGURE 8 Sample finite state machine recognizing the language $\{000, 001, 100, 110\}$.

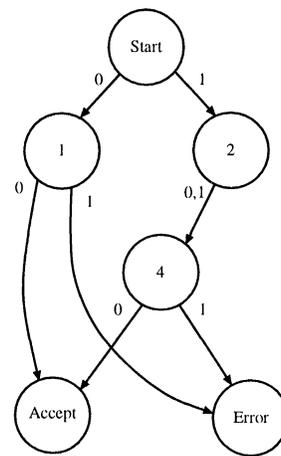


FIGURE 9 Partially reduced finite state machine recognizing the language $\{000, 001, 100, 110\}$.

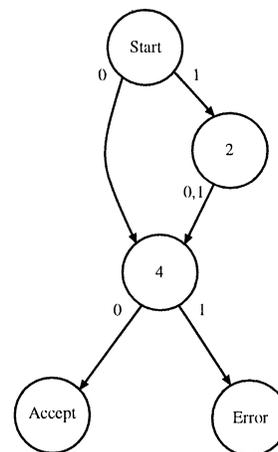


FIGURE 10 Minimal finite state machine recognizing the language $\{000, 001, 100, 110\}$.

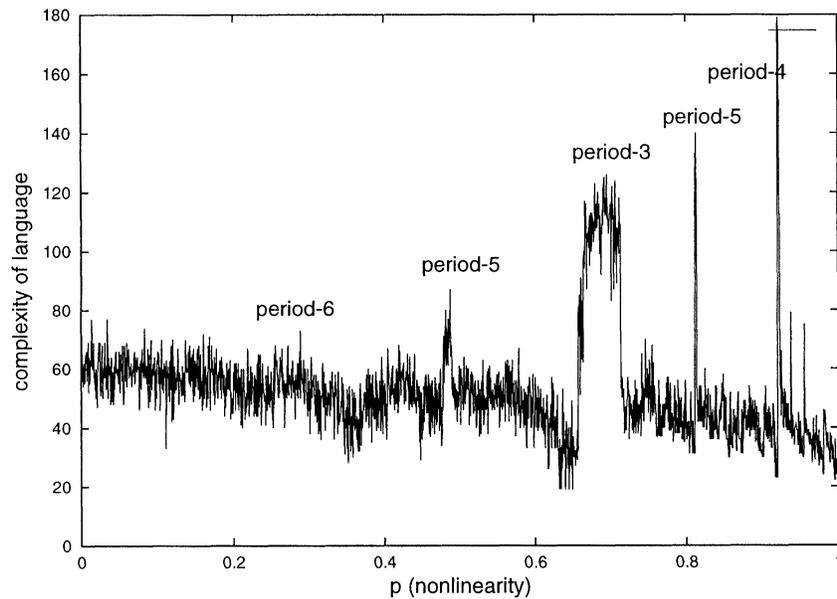


FIGURE 11 Complexity of the language of children against the nonlinearity parameter.

In Figure 11 we have shown the computed complexity of the languages against the nonlinearity parameter for a total of 2049 children, uniformly distributed along the nonlinearity parameter axis. The most notable feature of the figure is the peaks, which are much more pronounced than the performance peaks in the fitness landscape. We recognize the largest peak as the period-3 window in the logistic map where the most dominant clustering took place during simulated evolution. Also we note the period-4 and period-5 windows that had visible clustering during evolution. In addition there are a few peaks to be seen in the left hand side of the figure, most prominently a significant peak associated with another period-5 window. The figure shows that the most complex languages are not created in the period-3 window but rather in the period-4 and period-5 windows. However, as we saw these are not as important during simulated evolution. This is due to the narrow structure of the peaks. Remember that there are a number of inherent noise factors. First, initial conditions before imitation are always chosen at random, and secondly the learning

process adds a random number to the individuals nonlinearity parameter when they learn. Especially the latter has the effect of destabilizing very narrow peaks. The most interesting information we gain from Figure 11 is that the system dynamics clearly selects individuals with languages that are complex in terms of recognition, although there is nothing in the system dynamics that refers to the complexity of the language (only the ability to imitate matters).

6. DISCUSSION

We have described a model of childrens' language development whose central element was the speech generation process. This process was based upon discretization of the orbits of noninvertible maps, here logistic maps. The model of language development was based on imitation, *i.e.*, the childrens' skills were measured by their ability to imitate utterances of other children. The relative abilities were presented in terms of the speech generation parameter by computing the fitness landscape. The

peaks in this landscape, which correspond to good language skills with respect to imitation, were seen to be located mainly in periodic windows for the logistic map, with only the largest ones being visible, the largest by far being the period-3 window.

The model was extended with a learning mechanism by which the poor performing children could learn from the better. This was used in simulating evolution in a model with many individuals interacting through imitation. The dynamic interactions and hence learning lead to clustering of children near parameter values where periodic windows are located.

We introduced the complexity of the language of a child through the number of states of a minimal finite state machine capable of recognizing the child's language. It was demonstrated that the imitation mechanism during evolution selected exactly those children with complex languages, in the above sense. This was despite the fact that the imitation process contains no information about the complexity of the language. We saw that even more complex languages existed but that they were destabilized these in evolutionary terms by noise.

Evolutionary models such as the present represent a qualitative approach to analysing complex dynamical systems, in that they are constructed as analogous to the real life system, not as a direct representation of it [7]. The model of language learning that we have explored contains only a few features characteristic of real life language use and development. In evaluating the usefulness of computational modelling in the investigation of extremely complicated real life systems, such as language development, the extent to which the model mimics, or not, the behaviour of real life systems will give some indication of usefulness. If a simple model performs appropriately, it might be gradually developed by introducing further features.

The features that were selected to be included in the model—speech generation and learning through reproducing the speech of others—are

central to language development in both first and foreign language contexts. Evolving the model led to the clustering around certain values, and to a preference for the children who succeed in learning to be also those who develop more complex languages, but not the most complex languages which was due to the noise level as discussed above. The model produces this as an emergent outcome; it could not have been predicted from the components of the model.

How does the emergent behaviour of the model compare with real life language development? In some key ways, the model reflects what applied linguists and teachers would recognise as intuitively likely: reproducing the speech of others can lead to the development of a more complex language, as it has been defined here, which contains some sense of being original or unexpected. The finding that this process is more successful for the good speaker (those around the period three window) than for really excellent speakers, or those who are not so good, is initially surprising, but might reflect the need for alternative ways of learning for children of different proficiency levels. On the other hand, the less positive learning experience of those children with just slightly less complex languages than the good learners is a counter-intuitive outcome that needs further study.

The model we have presented is clearly extremely simplified and, of course, many aspects have been left out. There are obviously many ways in which one could develop the model. If we start with the speech generation mechanism it is important that the underlying model is capable of generating complex utterances, which was the main reason for our choice of the logistic map. It is possible to use other maps, for instance, higher dimensional maps that include more parameters, all or some of which could then be included in the learning process. Since the utterances we require are composed from a discrete set one could also consider the use of finite state machines or Turing machines, both of which could immediately deliver utterances without coarse graining. The richness of

the dynamics of the logistic map would require large finite state machines and Turing machines if we were to mimic these.

Other models of interaction could be tried. For example, it would be better to include a dynamic coupling rather than the static being used, in order to reflect how speakers accommodate to each other in the process of word composition [3].

It is possible to maintain reproduction as the main mechanism of learning but rather than the current use of unsupervised learning one could introduce supervised learning with a teacher or other helpful adult who could present more learnable utterances for reproduction. Likewise the model of learning, which is extremely simple in the present model could be replaced by something more sophisticated, for instance, a learning mechanism that takes some account of the child's zone of proximal development [16], *i.e.*, that at any given time, certain things are easier to learn than others [6]. It would also be interesting to include additional effects into the model, such as a grammar which would constrain the order of words in an utterance.

References

- [1] Bates, E. and Goodman, J., On the emergence of grammar from the lexicon, *The emergence of language* (Mahwah, New Jersey) (MacWhinney, B. Ed.), Lawrence Erlbaum Associates, 1999.
- [2] Collet, P. and Eckmann, J.-P., *Iterated maps on the interval as dynamical systems*, Birkhäuser, Basel, 1980.
- [3] Coupland, N. and Giles, H. (1988). Communicative accommodation: Recent developments, *Language and Communication*, **8**, 175–182.
- [4] Crutchfield, J. P. and Young, K., Computation at the onset of chaos, *Complexity, entropy and the physics of information* (Redwood City) (Zurek, W. H. Ed.), Addison-Wesley, 1990.
- [5] Elman, J., Language as a dynamical system, *Mind as motion: Explorations in the dynamics of cognition* (Cambridge, Massachusetts) (Port, R. and van Gelder, T. Eds.), MIT Press, 1995, pp. 195–223.
- [6] van Geert, P. (1998). A dynamic systems model of basic developmental mechanisms: Piaget, Vygotsky, and beyond, *Psychological Review*, **105**, 634–677.
- [7] van Gelder, T. and Port, R., It's about time: An overview of the dynamical approach to cognition, *Mind as motion: Explorations in the dynamics of cognition* (Cambridge, Massachusetts) (Port, R. and van Gelder, T. Eds.), MIT Press, 1995, pp. 1–44.
- [8] Grimaldi, R. P., *Discrete and combinatorial mathematics*, Addison-Wesley, Redwood City, 1999.
- [9] Guckenheimer, J. and Holmes, P., *Nonlinear oscillations, dynamical systems and bifurcations of vector fields*, Springer-Verlag, New York, 1983.
- [10] Kaneko, K. and Suzuki, J., Evolution to the edge of chaos in an imitation game, *Artificial life III* (Redwood City) (Langton, C. G. Ed.), Addison-Wesley, 1994, pp. 43–54.
- [11] Kauffman, S. A. (1990). Requirements for evolvability in complex systems: orderly dynamics and frozen components, *Physica D*, **42**, 135–152.
- [12] Krashen, S., *Principles and practice in second language acquisition*, Pergamon, Oxford, 1982.
- [13] Long, M. H. (1983). Native speaker/non-native speaker conversation and the negotiation of comprehensible input, *Applied Linguistics*, **4**, 126–141.
- [14] Snow, C., Social perspectives on the emergence of language, *The emergence of language* (Mahwah, New Jersey) (MacWhinney, B. Ed.), Lawrence Erlbaum Associates, 1999.
- [15] Swain, M., Three functions of output in second language learning, *Principles and practices in applied linguistics* (Oxford) (Cook, G. and Seidlhofer, B. Eds.), Oxford University Press, 1996, pp. 245–256.
- [16] Vygotsky, L., *Thought and language*, Wiley, New York, 1962.
- [17] Wasserman, P. D., *Neural computing: theory and practice*, Van Nostrand Reinhold, New York, 1989.