Phonemic Coding: Optimal Communication Under Noise?

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

Willem Zuidema
Language Evolution and Computation Research Unit
School of Philosophy, Psychology and Language Sciences
and Institute of Animal, Cell and Population Biology
University of Edinburgh
40, George Square
Edinburgh EH8 9LL, United Kingdom
jelle@ling.ed.ac.uk

http://arti.vub.ac.be/ ∼ bartb
http://www.ling.ac.uk/ ∼ jelle

Abstract

Human languages are universally phonemically coded, whereas many animal signal systems are not. A number of theories and models have been developed to explain this evolutionary transition, but some major problems remain. We present a simulation to investigate the hypothesis that phonemic coding is an side effect of optimizing signal systems for success in imitation. Crucially, signals in our model are trajectories in an (abstract) acoustic space. Hence, both holistic and phonemically coded signals have a temporal structure. Using both qualitative inspection of emerged systems of trajectories and a statistical analysis of a measure of phonemicity, we find that phonemically coded systems are indeed preferred. The model thus provides a new explanations for the evolutionary pathway to the emergence of phonemic coding.

1 Introduction

One of the universal properties of human language is the fact that it is phonemically coded. Linguistic utterances can be split into units that can be recombined into new linguistic utterances. For instance, the words “we”, “me”, “why” and “my” as pronounced in standard British English are built-up from the units “w”, “m”, “e” and “y”, which can all be used in many different combinations.

There is some controversy about the exact level at which combination takes place. In the traditional view the atomic units are phonemes: minimal speech sounds that can make a distinction in meaning. An increasingly popular alternative view is that the atoms are syllables, or the possible onsets, codas and nuclei of syllables. Nevertheless, there is general agreement that in natural languages, atomic units are combined into larger wholes. For the purposes of this paper, we do not need to take sides in the debate about the exact nature of the combinatorial elements of human language. Instead, we study signals that occur in an abstract acoustic space, and address the question of why and how phonemically coded sets of signals have emerged.

The combinatorial nature of human speech is in contrast with many animal calls and non-linguistic human utterances, which generally cannot be split into smaller units. The songs of some songbirds and whales, however, do seem to have combinatorial structure. The fact that in evolutionary unrelated lineages combinatorial systems have emerged indicates that such systems can be considered as evolutionary attractors. Recombination apparently has major evolutionary advantages. Two views on the advantages that recombination offers are:

1. It makes it possible to transmit an infinite number of messages over a noisy channel (the "noisy coding argument", an argument from information theory, e.g. Nowak & Krakauer 1999).

2. It makes it possible to create an infinitely extensible set of signals with a limited number of building blocks. Such productivity provides a solution for memory limitations, because signals can be encoded more efficiently, and for generalization, because new signals can be created by combining existing building blocks (the "productivity argument", a point often made in the generative syntax tradition, e.g. Jackendoff 2002);

These advantages are a good starting point for answering the questions of *why* combinatorial coding would emerge, and *how* initially holistic systems (which seem to be the default for smaller repertoires of calls) can change into phonemically coded systems. In this paper we will address both questions. In the following we will discuss some existing formal models of phonemic coding, discuss which open problems remain and then develop a model of our own that addresses some of these problems.

2 Previous work

2.1 Natural selection for combinatorial phonology

Several mathematical and computational models have shown that under noisy transmission, digital, combinatorial coding is more efficient than continuous coding. Nowak & Krakauer (1999) apply this insight in the context of the evolution of language, and derive an expression for the "fitness of a language". Imagine a population of individuals that all agree on which signals to use for which objects or events. The fitness of a language is now given by the expected success of a random individual to communicate about a random object or event with a random other individual. Nowak et al. show that when communication is noisy and when just a single sound is used for every meaning, the fitness is limited by an "error limit": only a limited number of sounds can be used — and thus a limited of meanings be expressed — because by using more sounds the successful recognition of the current signals would be impeded. Nowak et al. further show that in such noisy conditions, fitness is higher when (meaningless) sounds are combined into longer words. When the environment is combinatorial (i.e. objects and actions occur in many combinations) the fitness is highest when meaningful words are combined into longer sentences.

These results are essentially particular instantiations of Shannon's more general results on "noisy coding" (Shannon, 1948), as is explored in a later paper by the same group (Plotkin & Nowak, 2000). More interesting is the question how natural selection could favor a linguistic innovation in a population where that innovation is still very rare. Nowak & Krakauer (1999) do a game theoretic analysis of "compositionality". They consider all mixed strategies where both holistic and compositional signals are used, and show that strategies that use more compositionality can invade strategies that use less. This means that the adaptive dynamics of languages under natural selection should lead to compositionality. For combinatorial phonology a similar analysis can be given.

Although this model is a useful formalization of the problem and gives some important insights, as an explanation for the evolution of phonemic coding and compositionality it is still insufficient. The main problem is that the model only considers the advantages of combinatorial strategies, and ignores two obvious disadvantages: (1) by having a "mixed strategy" individuals have essentially two languages in parallel, which one should expect to be costly because of memory and learning demands and additional confusion; (2) combinatorial signals that consist of two or more sounds take longer to utter and are thus more costly. A fairer comparison would be between holistic signals of a certain duration (where repetition of the same sound decreases the effect of noise) and combinatorial signals of the same duration (where the digital coding decreases the effect of noise). This is the approach we take in this paper.

2.2 Crystallization in the perception–imitation cycle

A completely different approach to phonemic coding is based on "categorical perception". Categorical perception (Harnad, 1987) is the phenomenon that categorization influences the perception of stimuli in such a way that differences between categories are perceived as larger and differences within categories as smaller that they really are (according to an "objective" similarity metric). For instance, infants already perceive phonemes as closer to the closest prototype phoneme from their native language than it is according to an "objective" (cross-linguistic) acoustical metric (Kuhl *et al.*, 1992). Hence, when presented vowels as stimuli ranging from /o/ to /a/ in fixed increments, British subjects will hear the first stimuli as o's or almost o's, and the last as a's or almost a's. Apparently, the frequency and position of acoustic stimuli gives rise to particular phoneme prototypes, and the prototypes in turn distort the perception.

Oudeyer (2002) studies a model that yields such a perceptual distortion. In this model, signals are modeled as points in an acoustic space, and are thus instantaneous. Oudeyer considers that signals survive from generation to generation because they are perceived and imitated. Oudeyer shows that categorical perception *shapes* a signal repertoire such that it conforms more and more to the prototype phonemes. Thus, emitted signals shape perception, and distorted perception shapes the repertoire of signals in the cycle from emission to perception to emission (the perception–imitation cycle; see also Westermann 2001 for a model of sensori-motor integration and its relevance for imitation and categorical perception). Oudeyer calls the collapse of signals in a small number of clusters "crystallization".

Oudeyer's model is fascinating, because it gives a completely non-adaptive mechanism for the emergence of phonemic coding. However, it is not clear how well it would work if signals have a time structure rather than being instantaneous[1]. Moreover, even if the mechanism works also in these conditions, it remains an important question whether phonemic coding increases the functionality of the language, and thus the fitness of the individual that uses it. If not, one would expect selection to work against it. In particular, in Oudeyer's model, where signals are instantaneous, a large repertoire of signals is collapsed into a small number of clusters. A functional pressure to maintain the number of distinct signals would thus have to either reverse the crystallization, or combine signals from different clusters. This aspect, which seems the core issue in understanding the origins of phonemic coding, is

[1]Oudeyer has also tested the model for sequences of sounds (Oudeyer, p.c.), but, as far as we know, not for continuous trajectories. It seems that in this version of the model the "combinatorial" aspects of phonemic coding is imposed and only the "categorical" (see section 2.3) aspect is emergent, such that our criticism still holds.

not modeled by Oudeyer. In our model, we ensure that the number of distinct signals remains at least at the same level; i.e. the functionality increases rather than decreases.

2.3 Aspects of phonemic coding

Other models of phonemic coding assume the sequencing of phonetic atoms into longer strings as given. They concentrate rather on the structure of the emerged systems (Lindblom *et al.*, 1984; de Boer, 2001; Redford *et al.*, 2001) or on how conventions on specific combinatorial signal systems can become established in a population through cultural transmission (Steels & Oudeyer, 2000). These models are interesting, and, importantly, bridge the gap with empirical evidence on how phonemic coding is implemented in the languages of the world.

It appears from this discussion that there are 4 related, but distinct aspects to phonemic coding:

1. Phonemically coded systems are *categorical*, in that they allow only a small number of basic sounds and not all feasible sounds in between;

2. they are also *superficially combinatorial*, in that all parts of each signal overlap with parts of other signals;

3. they are also *productively combinatorial*, in that the cognitive mechanism that produces and interprets signals uses the common parts of signals as building blocks that can be combined in all sorts of combinations;

4. the possible sets of categories and combinatorial rules show particular (cross-linguistic) constraints.

These aspects form a hierarchy, where the aspects further down the list imply the aspect above it. Oudeyer (2002) shows a non-adaptive mechanism that can yield aspect 1 (and gives a starting point for 4), but does not explain how the other aspects come about and how the functionality of the signal system is preserved. Nowak & Krakauer (1999) show how natural selection could favor 2, but ignore the temporal aspects of holistic signals. Zuidema & Hogeweg (2000) and Zuidema (2003) can be viewed as assuming aspects 1 and 2, and addressing the emergence of aspect 3 under natural selection and cultural evolution respectively (but the models are not discussed in these terms). Lindblom *et al.* (1984); de Boer (2001); Redford *et al.* (2001); Steels & Oudeyer (2000) all address aspect 4.

The question thus remains open as to under what circumstances a system of holistically coded signals with finite duration would change into a phonemically coded system of signals. In the paper we study a single mechanism that can yield aspects 1 and 2.

3 The model

In our model, we do not assume combinatorial structure, but rather study the gradual emergence of phonemic coding from initially holistic signals. We do take into account the temporal structure of both holistic and phonemically coded signals. We view signals as continuous movements (“gestures”, “trajectories”) through an abstract acoustic space. We assume that signals can be confused, and that the probability of confusion is higher if signals are more similar, i.e. closer to each other in the acoustic space according to some distance metric. We further assume that a functional pressure on distinctiveness maximizes the distance between trajectories.

3.1 Representing trajectories

The model is based on part-wise linear trajectories in a bounded 2-D continuous space (of size 15.0×15.0 in all simulations reported here). Trajectories are sequences of points with fixed length (here: 20). Each point has a fixed distance of 1.0 to the immediately preceding and following points in the sequence. The following and preceding points to a point can lay anywhere on a circle of radius one with that point at the center. Trajectories always stay within the bounds of the defined acoustic space.

Signals in the real world are continuous trajectories, but in the model we need to discretize the trajectories. However, to ensure that we do not impose the phonemic structure we are interested in, we discretize at a much finer scale than the phonemic patterns that will emerge. Hence, the points on a trajectory are not meant to model atomic units in a complex utterance.

3.2 Measuring distances

The distance between two trajectories t and r is defined as the sum of the distances between all corresponding points in the best possible alignment of the two trajectories. In finding the best possible alignment, one point from t can be mapped on several neighboring points in r and vice versa. In this way trajectories that resemble each other in shape, but that do not align perfectly still are considered close. This models the way humans perceive signals. The distances are calculated using "dynamic time warping", an efficient method that before the advent of statistical models, has been used with reasonable success in computer speech recognition (e.g. Sakoe & Chiba, 1978).

3.3 Maximizing the total mutual distance

In the first set-up of the model, we consider an idealized single repertoire of trajectories that, in a sense, repel each other. That is, the total distance between trajectories is optimized using a simple hill-climbing algorithm. The model goes through a large number of iterations. At every iteration, the sum of all mutual distances is calculated. Then a random change is applied to a random trajectory t, and the total distance is measured again. If this second measurement is larger than the first, the change is kept. If not, the change is reverted.

Random changes always respect the constraints on well-formed trajectories. Hence, a random point, t_x, is moved to a new random position (from a Gaussian distribution around the old position, provided it falls within the boundaries of the acoustic space). The two points on both sides of the moved point, t_{x+1} and t_{x-1}, are moved closer or further away such that the distance to t_x is again 1. The direction from t_x to t_{x+1} or t_{x-1} remains the same, unless the point would cross the boundary of the space, in which case it is rotated to the closest point within the boundary at distance 1 from t_x. The same procedure is applied recursively to the neighbors of t_{x+1} and t_{x-1} until the ends of the trajectory are reached.

In the second set-up of the model, we investigate what kind of repertoires of trajectories emerge in a *population* of agents that try to imitate each other in noisy conditions. The model is very similar, but now each agent in the population has its own repertoire, and it tries to optimize its own success in imitating and being imitated by other agents of the population.

This version of the model is like the imitation games of de Boer (2000). These only modeled holistic signals (vowels) and did not investigate phonemic coding. The game implemented here is a slight simplification of the original imitation game. First, all agents in the population are initialized with a random set of a fixed number of trajectories. Then for each game, a speaker is randomly selected from the population. This speaker selects a trajectory, and makes a random modification to it. Then it plays a number of imitation games (50 in all simulations reported here) with all other agents in the population. In these games, the *initiator* utters the modified trajectory with additional noise. The *imitator* finds the closest trajectory in its repertoire and utters it with noise. Games are successful if the imitator's signals is closest to the modified trajectory in the initiator's repertoire. If it turns out that the modified trajectory has better imitation success than the original trajectory, the modified trajectory is kept, otherwise the original one is restored.

4 Results

4.1 Optimizing a single repertoire

We ran the model under the single repertoire condition with a number of different parameters. In all simulations the initial trajectories are random sequences of positions, where the only constraints are that neighboring points are at distance 1 from each other and that all points are within the permitted space.

In simulations with few trajectories (up to 4), we find that the trajectories "bunch-up" and remain within a very small area at maximum distance from the areas used by the other trajectories. Each of these signals is thus a holistic signal, but the signals are "categorical".

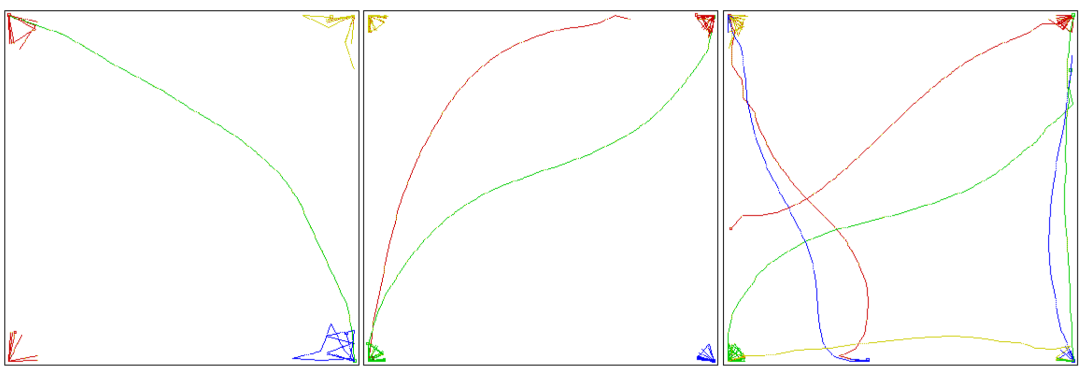

Figure 1: Comparison of optimized systems of 5, 6 and 10 trajectories. Note the reuse of start- and endpoints (squares indicate start points).

In simulations with 5 trajectories, 4 occupy the corners areas of the acoustic space in the same way as in simulations with 4 trajectories. However, the fifth trajectory stretches from one corner to another, and thus shares the areas for its begin and end points with two different other trajectories (see fig. 1, leftmost panel). This can be interpreted as a rudimentary phonemic code.

With more trajectories, the reuse of beginning and end points becomes more pronounced. In the simulation with 6 trajectories, the first 5 are similarly organized, but the sixth is essentially the inverse of the fifth. In the simulation with 10 trajectories, 3 trajectories are still bunched up in a small area of the acoustic space, but the other 7 are stretched out, sharing begin and end points with one another. Frequently one can find trajectories that are more or less the inverse of another trajectory.

In order to perform a statistical analysis, a numerical measure of the extent to which emerged systems were phonemically coded had to be defined. This measure, called the phonemicity $\mathcal{P}$, is defined as the ratio between the average distance between the start and end points of all trajectories and the average distance between all other corresponding points of all trajectories. Corresponding points are defined as points that are an equal number of steps away from either a start or an end point (i.e., two points that are at position 3 are corresponding points, but so are a point at position 3 and position L-2). The details for this measure are in the appendix.

In a phonemically coded system of trajectories, start and end points are expected to be closer together than the other points on a trajectory, while in a holistically coded system of trajectories, the average distance is expected to be approximately equal. Therefore, the measure should give lower values for phonemically coded systems. It is quite likely that better measures of phonemicity can be defined, but this measure does make a distinction between holistically and phonemically coded systems, and was therefore adopted for the analysis.

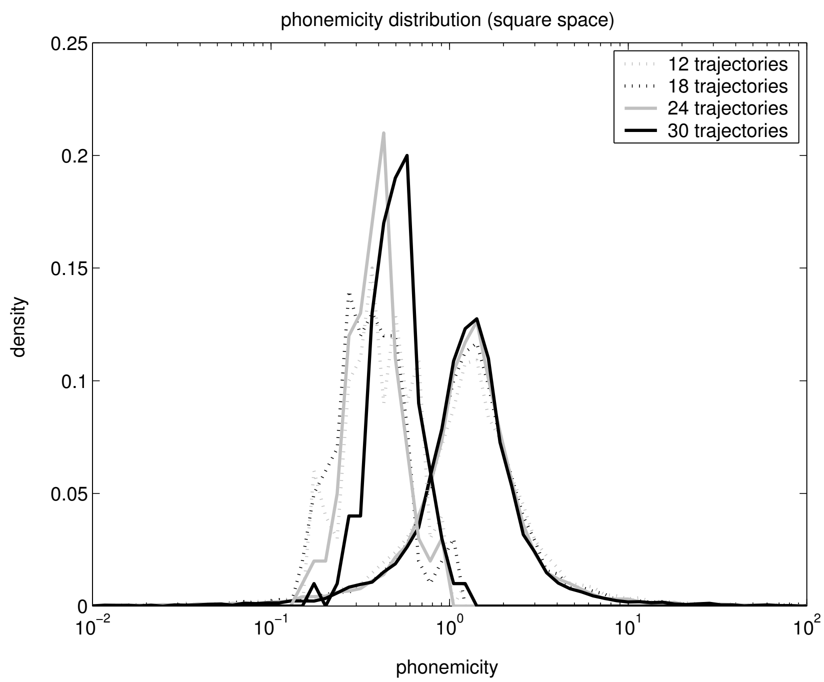

Figure 2: Distributions of phonemicity of random systems (right peak) and optimized systems (left peak). Note that the phonemicity measure for optimized systems is lower, indicating that the optimized systems are more phonemically coded than random systems.

Two conditions were compared. In both conditions the systems of trajectories were initialized randomly; only in the second condition were systems of trajectories optimized for distance first using 30,000 optimization steps. The results were measured for systems of many different sizes, but are presented for systems of 12, 18, 24 and 30 trajectories in figure 2. 10,000 random systems were evaluated, but for computational reasons only 100 optimized systems, as the amount of computation needed for optimization precluded larger numbers of systems to be evaluated. Note that the horizontal axis (showing the phonemicity) is logarithmic. This has the advantage of both making the peaks more distinct and making the distributions more similar to the normal distribution. When using the t-test, on both the phonemicity and its logarithm, it turns out the difference between the distribution of the random systems and the optimized systems is significant with $p < 0.05$ (the t-test is less appropriate for the non-log measure, because of the highly skewed distribution).

This result indicates that optimization for acoustic time-warped distance between trajectories results in more phonemically coded systems.

4.2 Optimizing repertoires in a population

For vowel systems, it has been shown that optimizing a single repertoire leads to similar systems as a population-optimization system (compare de Boer, 2000; Liljencrants & Lindblom, 1972). It can be shown that for trajectories the same is true, under the condition that noisy distortions of trajectories do not distort the shape of these trajectories too much. This is illustrated in figure 3.

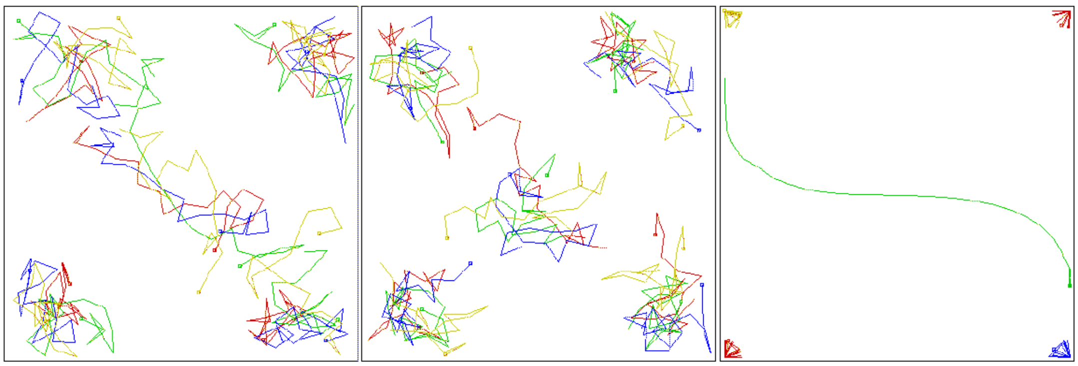

Figure 3: From left to right: emerged system with five trajectories in a population of ten agents (four agents shown), emerged system with five trajectories and uncorrelated noise, and optimized system of five trajectories. Small squares indicate the starting point of trajectories.

In this figure the left frame shows the system of five trajectories that resulted from playing imitation games in a population, using form-preserving noise. The right frame shows a system of five trajectories that resulted from optimizing distance. It can be observed that in both cases, the corners are populated by four trajectories, which are bunched up. The fifth trajectory, in contrast, follows the diagonal. As before, an analysis in terms of phonemes suggests itself: the four corners are basic phonemes, while the fifth trajectory uses one as the corners as a starting phoneme and the opposite corner as the ending phoneme. Both models result in similar systems of trajectories.

The middle frame, on the other hand, shows that when noise does not preserve shape of trajectories, a system results in which all trajectories are bunched up and an analysis in terms of phonemes is therefore not possible. As noise in real signals is band limited, it follows that shape will always be preserved to some extent. Therefore the shape-preserving model is indeed the correct model. Instead of investigating computationally extremely costly population models, it is therefore possible to investigate emergence of phonemic coding using the optimization model. For computational reasons, we have not performed simulations in the population condition with more than 5 trajectories.

5 Conclusion

We have investigated whether systems of trajectories that are used for imitation in a population would tend toward phonemic coding when agents tried to maximize their imitation success. It was found that running simulations of populations directly was too time-consuming. However, it was also found that direct optimization of time-warped distance between trajectories resulted in systems of trajectories that were similar to those found in preliminary experiments with imitation in a population. For this to be true, it was necessary to assume that in the population case, shape of trajectories was preserved under noise. This is a realistic assumption, as it turns out to be true for all noise that is band-limited, i.e. for which the energy of higher frequencies tends to zero. This is the case for all real-world noise.

When systems of trajectories were optimized for time-warped distance, it turned out that start- and endpoints were reused and that there were no trajectories (at least for limited numbers of trajectories) that had the same start- and endpoint and that only differed in the shape of the trajectory in between. This is indicative of phonemic coding. A measure of phonemicity was defined and it was found that optimized systems had significant lower values for this measure than random systems, indicating that they were more phonemically coded.

The conclusion to be drawn from this is that systems of complex articulations (trajectories) that have maximum distance to each other tend to show aspects of phonemic coding. Systems that have trajectories that are maximally distant from each other are most robust to noise. This means that optimizing systems of large numbers of complex articulations for robustness to noise, which is likely to happen when they are used for communication in a population, would result in systems of trajectories that can be analyzed in terms of phonemes.

The relevance for the evolution of speech is clear. When populations of agents start to communicate using small numbers of signals, it is unlikely that they would use phonemic coding, or be able to use it if it occurred. However, when extending the number of signals, the most robust systems would be the ones that can be analyzed as phonemically coded. Agents that have adaptations to detect and use this property would have an evolutionary advantage, as they would be able to learn faster, and probably to communicate more accurately as well. This provides a cultural beginning of a possible biological adaptation for using phonemically coded signals. This adaptation in the area of speech could later be exapted for use in combining words, in other words, for syntax.

6 Future work

The results described in this paper are preliminary, and need to be extended in several ways. Firstly, the model, especially in the population condition, should be tested with larger number of trajectories, and with trajectories of longer length. Presumably, the “phonemic coding” would then not just apply to the start and end points of the trajectories. Consequently, another measure of phonemicity needs to be defined.

Further, the model can be altered such that it allows trajectories of varying length in a single repertoire, and perhaps varying distances between the points of a trajectory.

Finally, and most ambitiously, the model should be extended to incorporate the aspects of phonemic coding that are currently not addressed: productive combinatorics and realistic constraints on the categories and rules of combination.

References

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, UK: Oxford University Press.

HARNAD, S. (1987). *Categorical Perception*. Cambridge, UK: Cambridge University Press.

JACKENDOFF, R. (2002). *Foundations of Language*. Oxford, UK: Oxford University Press.

KUHL, P., WILLIAMS, K., LACERDA, F., STEVENS, K. & LINDBLOM, B. (1992). Linguistic experience alters phonetic perception in infants by 6 month of age. *Science* **255**, 606–608.

LILJENCRANTS, J. & LINDBLOM, B. (1972). Numerical simulations of vowel quality systems: the role of perceptual contrast. *Language* **48**, 839–862.

LINDBLOM, B., MACNEILAGE, P. & STUDDERT-KENNEDY, M. (1984). Self-organizing processes and the explanation of language universals. In: *Explanations for language universals* (Butterworth, M., Comrie, B. & Dahl, ., eds.), pp. 181–203. Berlin: Walter de Gruyter & Co.

NOWAK, M. A. & KRAKAUER, D. C. (1999). The evolution of language. *Proc. Nat. Acad. Sci. USA* **96**, 8028–8033.

OUDEYER, P.-Y. (2002). Phonemic coding might be a result of sensory-motor coupling dynamics. In: *Proceedings of the 7th International Conference on the Simulation of Adaptive Behavior* (Hallam, B., Floreano, D., Hallam, J., Hayes, G. & Meyer, J.-A., eds.), pp. 406–416. Cambridge, MA: MIT Press.

PLOTKIN, J. B. & NOWAK, M. A. (2000). Language evolution and information theory. *Journal of Theoretical Biology* pp. 147–159.

REDFORD, M. A., CHEN, C. C. & MIIKKULAINEN, R. (2001). Constrained emergence of universals and variation in syllable systems. *Language and Speech* **44**, 27–56.

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

SHANNON, C. E. (1948). A mathematical theory of communication. *The Bell Systems Technical Journal* **27**, 379–423 and 623–656.

STEELS, L. & OUDEYER, P.-Y. (2000). The cultural evolution of syntactic constraints in phonology. In: *Proceedings of the VIIth Artificial life conference (Alife 7)* (Bedau, M. A., McCaskill, J. S., Packard, N. H. & Rasmussen, S., eds.). Cambridge (MA): MIT Press.

WESTERMANN, G. (2001). A model of perceptual change by domain integration. In: *Proceedings of the 23d Annual Conference of the Cognitive Science Society*. Mahwah, NJ: Lawrence Erlbaum.

ZUIDEMA, W. (2003). How the poverty of the stimulus solves the poverty of the stimulus. In: *Advances in Neural Information Processing Systems 15 (Proceedings of NIPS'02)* (Becker, S., Thrun, S. & Obermayer, K., eds.). Cambridge, MA: MIT Press. (forthcoming).

ZUIDEMA, W. & HOGEWEG, P. (2000). Selective advantages of syntactic language: a model study. In: *Proceedings of the 22nd Annual Meeting of the Cognitive Science Society* (Gleitman & Joshi, eds.), pp. 577–582. Mahwah, NJ: Lawrence Erlbaum Associates.

Appendix: Measuring phonemicity

The average distance $\mathcal{E}$ between the extreme points (start and end points) is given by:

$$\mathcal{E} = \frac{1}{2N(N-1)} \sum_{i=0}^{N} \sum_{j=i+1}^{N} \left(D\left(i_1, j_1\right) + D\left(i_1, j_L\right) + D\left(i_L, j_1\right) + D\left(i_L, j_L\right) \right) \tag{1}$$

where N is the number of trajectories and L is the length of each trajectory. The function D is a measure of distance between points and will be explained below. The average distance between all other corresponding points is given by:

$$\mathcal{C} = \frac{1}{N(N-1)(L-2)} \sum_{i=0}^{N} \sum_{j=i+1}^{N} \sum_{k=2}^{L-1} \left(D\left(i_k, j_k\right) + D\left(i_{L-k+1}, j_k\right) \right) \tag{2}$$

The phonemicity $\mathcal{P}$ is then simply:

$$\mathcal{P} = \frac{\mathcal{E}}{\mathcal{C}} \tag{3}$$

The distance function $D(i_a, j_b)$ is the inverse, squared Euclidean distance between point a of trajectory i and point b of trajectory j:

$$D\left(i_a, j_b\right) = \frac{1}{\epsilon + |p_a(i) - p_b(j)|^2} \tag{4}$$

where $p_a(i)$ is the position of point a of trajectory i. The term ϵ ($\epsilon = 0.01$ throughout this paper) is added to avoid division by zero. Note that this is a different distance function than was used in the optimization of the distances between trajectories.