Cultural emergence of combinatorial structure in an artificial whistled language

Tessa Verhoef (t.verhoef@uva.nl)
ACLC, University of Amsterdam
1012 VT, Amsterdam, Netherlands

Simon Kirby (simon@ling.ed.ac.uk)
LEC, University of Edinburgh
Edinburgh EH8 9AD, UK

Carol Padden (cpadden@ucsd.edu)
CRL, University of California San Diego
La Jolla, CA 92093-0503, USA

Abstract

Speech sounds within a linguistic system are both categorical and combinatorial and there are constraints on how elements can be recombined. To investigate the origins of this combinatorial structure, we conducted an iterated learning experiment with human participants, studying the transmission of an artificial system of sounds. In this study, participants learn and recall a system of sounds that are produced with a slide whistle, an instrument that is both intuitive and non-linguistic. The system they are exposed to is the recall output of the previous participant. Transmission from participant to participant causes the system to change and become cumulatively more learnable and more structured. This shows that combinatorial structure can culturally emerge in an artificial sound system through iterated learning.

Keywords: iterated learning; duality of patterning; combinatorial structure; phonology; cultural evolution; emergence; learnability

Introduction

In human languages, a finite set of basic speech sounds (phonemes) is combined into a (potentially) unlimited set of well-formed morphemes (words, clitics, grammatical markers etc.). Hockett (1960) placed this phenomenon under the umbrella of 'duality of patterning' and listed it as one of the basic design features of human language. Such combinatorial structure requires a continuous signal space to be organized into a discrete set of basic building blocks that are reused in a systematic way (Oudeyer, 2006). In addition, there are constraints (e.g. phonotactic) on the ways the basic elements can be recombined. The specific building blocks and the rules for their recombination differ from one language to the other, but are shared among all members of a speech community. This paper presents an experimental investigation into how combinatorial organization of (acoustic) signals may have emerged.

Hockett (1960) suggested a possible advantage for a language that uses combinatorial structure: If there is a limit on how accurately signals can be produced and perceived, there is a practical limit to the number of distinct signals that can be discriminated. When a larger number of meanings needs to be expressed, structured (sequential) recombination of elements is needed to maintain clear communication. Not much data was available to Hockett regarding the origins of combinatorial structure that could be used as evidence in favor or against such a hypothesis. However, recently research on emerging sign languages, studies of computer simulations and experiments with human subjects have provided evidence that can shed light on this hypothesis.

Research on a newly emerging sign language, Al-Sayyid Bedouin Sign Language (ABSL), provides a potential challenge to Hockett's hypothesis of the emergence of phonological structure (Israel & Sandler, 2009; Sandler, Aronoff, Meir, & Padden, 2011). All established sign languages that have been analyzed have been shown to exhibit phonological structure with features of discreteness and recombination as in spoken systems. There is a discrete set of location, hand shape and movement features that are recombined into meaningful words and there are constraints on the ways in which features can be combined. ABSL is a young but already fully functional sign language in which the phonological structure does not appear to be fixed yet. This case questions whether the emergence of combinatorial structure is necessarily driven by a growing meaning space alone, because the language is fully functional without such structure.

The case of ABSL is rare and the language may change towards using combinatorial structure in future generations. This expectation is supported by several computer models which have shown that sound systems tend to develop discrete building blocks or combinatorial structure through self-organization in a population of interacting agents (de Boer, 2000; Oudeyer, 2002, 2005). De Boer and Zuidema (2010) simulated the transition from a system of holistic signals to a combinatorial system in this way. They introduced an important distinction between two types of combinatorial structure: 'superficial combinatorial structure', which is structure that can be identified by an outsider observing the system, but of which the users of the system are not aware; and 'productive combinatorial structure', which is structure that is actively used in production, perception, learning or storage of signals (de Boer & Zuidema, 2010). Only superficial combinatorial structure can be observed in this simulation, since it is hard to tell whether the agents are aware of it. Therefore, to be able to study active use of the system, we need to observe human behavior.

Kirby, Cornish, and Smith (2008) developed a novel exper-

imental paradigm: iterated learning with humans. This approach makes it possible to investigate the effects of cultural evolution on a transmitted (artificial) language in a controlled laboratory setting. This has the advantage over computational simulations that it does not abstract away from the complexities of human learning, though it still allows for control over environmental factors and the parameters of the artificial language. Such control is not possible when studying a language in the field. Iterated learning refers to (repeated) acquisition of a behavior by a person through observation of similar behavior by another person who acquired it in the same way (Kirby et al., 2008). In the experimental paradigm a chain of learners is created in which the outcome of the learning process of one participant is used as the input for the next person (Kirby et al., 2008). Passing through the minds of learners, the system that is being transmitted is expected to adapt to the learning biases, expectations and constraints of the learners (Deacon, 1997; Kirby & Hurford, 2002; Christiansen & Chater, 2008; Griffiths, Kalish, & Lewandowsky, 2008).

This idea is not new, having been successfully used in numerous computer simulations and, more recently, in behavioral experiments. In experiments it has been applied successfully to show the emergence of compositional structure (Kirby et al., 2008), combinatorial structure in visual symbols (del Giudice, Kirby, & Padden, 2010), predictability in plural marking (Smith & Wonnacott, 2010) and in other category or function learning tasks (Griffiths et al., 2008). We applied the same paradigm to study the emergence of combinatorial structure in an artificial whistled language. If structure emerges, it allows us to see whether people make productive use of it by observing their learning and production. We expected that an increase in productive combinatorial structure would result in a cumulative increase in learnability as well as a cumulative increase of superficial combinatorial structure.

Methods

The experiment described here involves learning twelve different signals using a slide whistle (see figure 1). We use slide whistles for sound production, because participants can easily use them to produce a rich repertoire of acoustic signals, while only very little interference from pre-existing linguistic knowledge is expected.

Figure 1: A slide whistle

Procedure

The participants completed four rounds of learning and recall. In the learning phase they were exposed to all twelve whistle sounds one by one, and were asked to imitate each sound with the slide whistle immediately after hearing it. After each learning phase, a recall phase followed in which they reproduced all twelve whistles without being able to listen to them again. After the fourth recall phase, the output of this last recall round became the input set for the next participant. For the recall procedure, we set in place a strategy to prevent participants from reproducing the same whistle twice. The user interface of the experiment automatically compares each new whistle produced to all other whistles already accepted and stored during that phase and it rejects that whistle if it was too similar to any other. To determine how close two whistles are, we defined a whistle distance measure. This measure is a weighted combination of several separate measures: the Dynamic Time Warping (DTW) (Sakoe & Chiba, 2003) distance between the two pitch tracks (D_p), the DTW distance between the two intensity tracks (D_i), the difference in the number of segments (separated by silent pauses) (D_s), the difference in variation of segment duration (D_{s_d}) and the difference in variation of pitch (D_{p_v}). These separate measures were scaled to have approximately a maximum value of one and then combined following: $0.5D_p + 0.2D_i + 0.2D_s + 0.05D_{s_d} + 0.05D_{p_v}$. Data collected in a pilot study was used to create this measure and to find the set of weights that resulted in the highest whistle recognition score. The rejection threshold was set at 0.06, a low value because it was not supposed to influence the outcome of the recall phase in any way other than to reject doubles.

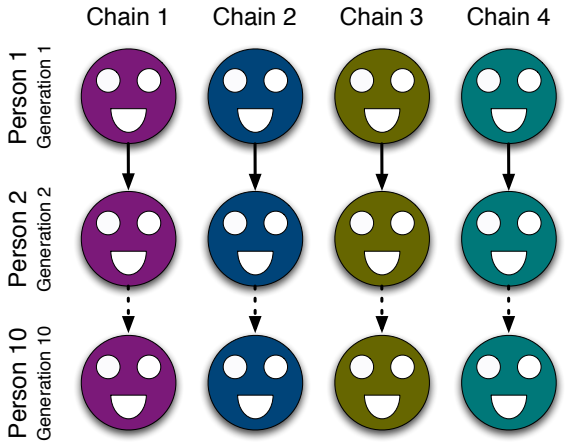

Figure 2: Four parallel chains of participants in an iterated learning experiment in which each person learns from the output of the previous participant.

Participants

Forty participants took part in four parallel chains of ten generations of learning and reproduction. This is schematically shown in figure 2. All participants were university students from either University of California San Diego, or University of Amsterdam, ranging in age from 18 to 32 (mean age of 22).

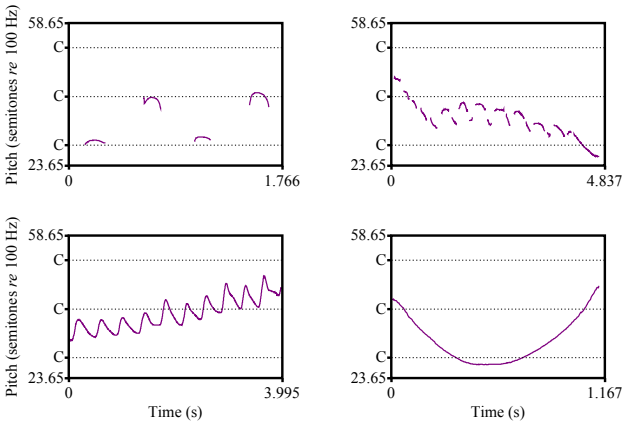

Figure 3: A few examples of whistles that appeared in the initial set of twelve, plotted as pitch tracks on a semitone scale.

Initial input whistles

All four chains started with the same initial set of twelve whistles, which were selected from a large database of whistles that were collected from nine participants in a pilot study. These participants were asked to freely produce a number of whistles. We selected a set of signals that uses a wide range of whistle techniques (e.g. slides, siren-like, staccato) such that the total set of whistles did not exhibit (superficial) combinatorial structure. A few examples are shown in figure 3, plotted as pitch tracks on a semitone scale using Praat (Boersma, 2001).

Results

The experiment resulted in four chains of ten participants in which each person produced an output set of twelve whistles as the result of learning from the productions of the previous person (or from the initial input set).

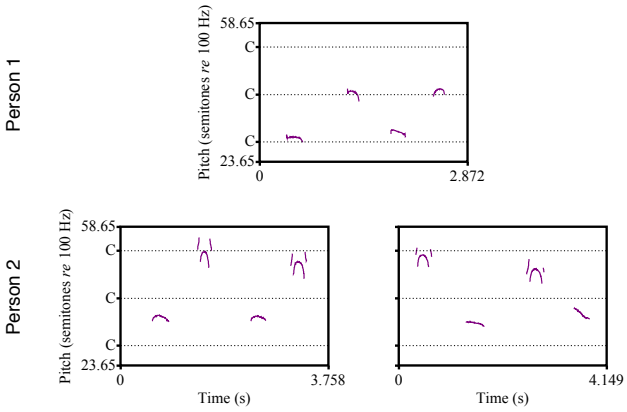

Figure 4: An example of mirroring in chain one. In the output of person one there was no other that was similar to the whistle shown, but in the next generation a mirrored version appears.

Qualitative results

If we take a close look at what happens over the course of the chains, there are a few recall behaviors that seem to lead to an increase of structure. Remembering twelve whistles after only four exposures is difficult, so participants generally don't recall all of them flawlessly. They seem to overgeneralize some of the (superficial) combinatorial structure that appears to be present. This results in an introduction of whistles which are related in form to other whistles learned: some of these whistles are inverted versions of learned whistles and others combine or repeat elements that are borrowed from existing whistles. As a result of this, whistles begin to share properties with one another but retain distinctive elements. This results in a set of whistles that consist of subsets of related elements, which appears to be more easily remembered and results in increased recall on the whole set.

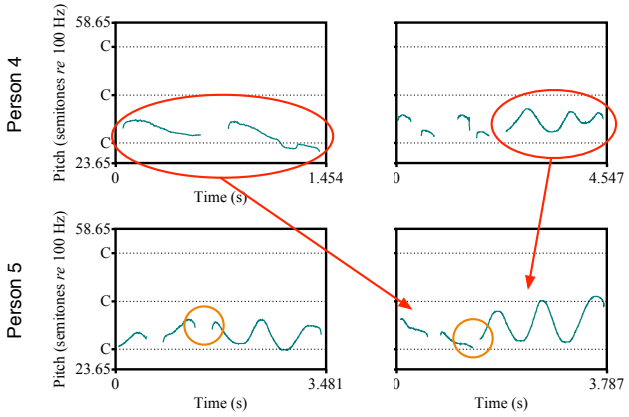

Figure 5: An example of recombination in chain four. One whistle and the second part of another whistle are combined in the next generation and the first part of this new whistle is mirrored in a second version. The orange circles indicate an effect that can be considered coarticulation.

Figures 4, 5, 6 and 7 show examples from several chains. Figure 4 shows an example of the creation of a new whistle through mirroring in chain one. Figure 5 shows an example of recombination in chain four in which one whistle from the previous generation is combined with the second part of another whistle to create a new whistle. In addition, the first part of this new whistle is mirrored in a second new whistle. Interestingly, these two whistles show an effect that can be considered coarticulation, because the initial pitch of the second part is influenced by the final pitch of the first part. Figure 6 shows an example of repetition (analogous to reduplication) in chain two in which a new whistle is formed by doubling an existing whistle learned from the previous generation. Figure 7 shows combined mirroring, repetition and borrowing from chain four, which results in a predictable system that is stable and persists after its innovation.

The set of whistles produced by the tenth participant in a chain is the end result of a process of repeated learning and

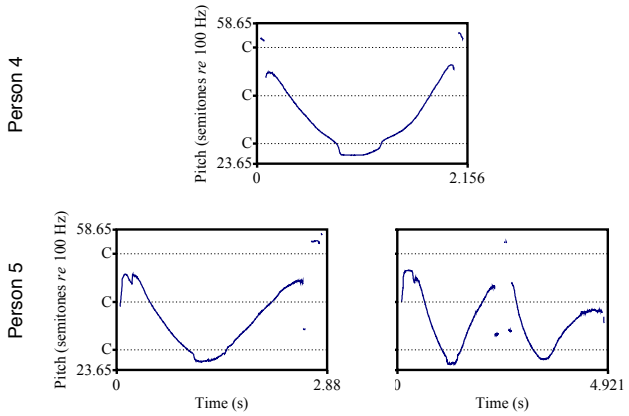

Figure 6: An example of repetition in chain two. The whole whistle is repeated to form a new whistle in the next generation.

imperfect recall and shows the cumulative effect of the mirroring, borrowing and repetition behaviors. Figure 8 shows part of the tenth set of chain one. In this set we can clearly identify a set of building blocks (slides up and down or single notes) and these are reused and combined in different ways in many whistles. Moreover, there appears to be 'language' specific constraints on the ways the elements are reused. In

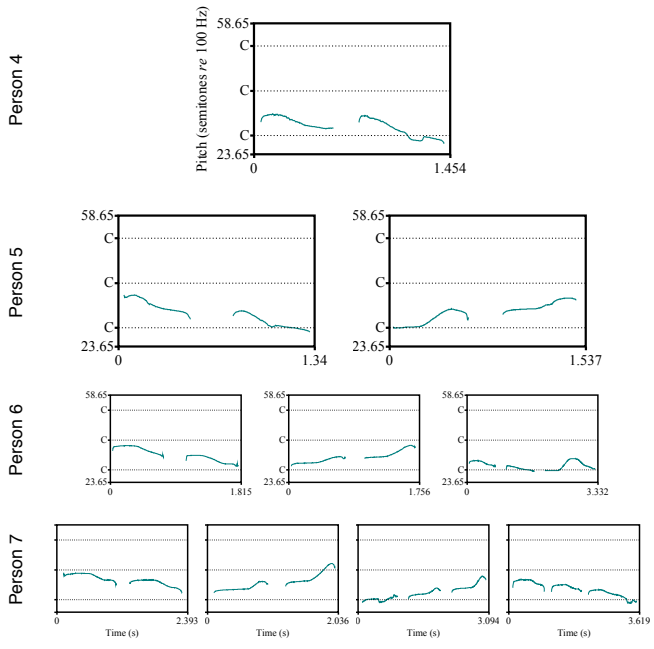

Figure 7: An example of cumulative mirroring, repetition and borrowing in chain four. Person 5 mirrors the whistle from the previous set, then person six borrows one of the two in a new whistle and finally this new whistle becomes generalized to fit the pattern of the original two, but repeated. This predictable system is stable towards the end of the chain.

this first chain for instance, single notes always follow each other on the same pitch and slides always go down first, never up-down. Such constraints exist in the other chains as well, but are not the same. Another interesting observation is that in the case of a silent pause, the signal mostly continues at the same pitch, which could be considered a coarticulatory effect.

In summary: qualitatively we can see an increase in the reuse of basic whistle elements in the sets. Once whistles that are composed of these elements appear in the set, they are more likely to be learned and recalled by later generations of participants who use the similarities across whistles to group them as subsets thus aiding their recall. This in turn makes it less difficult to remember the whole set.

Quantitative results

To attempt to confirm our qualitative observations and intuitions about the data, we present a quantitative analysis. First we will look at the development of learnability over the course of the chains and then we will investigate how the reuse of basic elements develops.

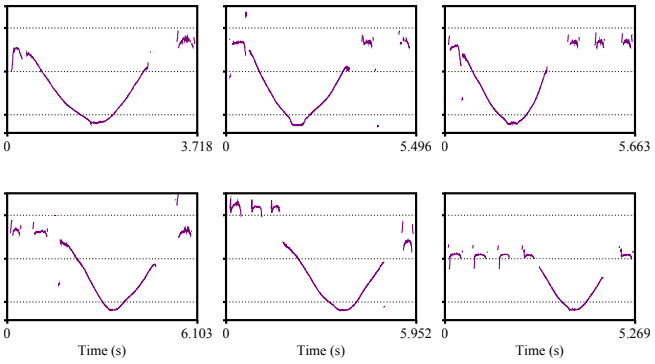

Figure 8: A fragment of the whistles in the tenth set of chain one. Clearly, a set of basic element can be identified that are reused and systematically combined.

To measure whether the sets of whistles really become easier to learn, we computed the distance between the input set and the output set for each participant in each chain to see if the recall error is cumulatively lower for participants that appear later in the chains. Recall error here is the sum of distances between each whistle in the output and its corresponding whistle from the input. To compute the distance between a pair of whistles, we used the measure that was described in the methods section. To compute the distance between two sets of twelve whistles (two consecutive generations' outputs), we used the following procedure: Each whistle in a set was paired with a unique whistle from the second resulting in a single set of 12 distinct pairs. Distance measures are computed for each pair and summed. This process was repeated for all possible permutations of pairs. The lowest sum of distances (representing the best set of pairs across the two sets of whistles) was chosen as the distance value for the two sets.

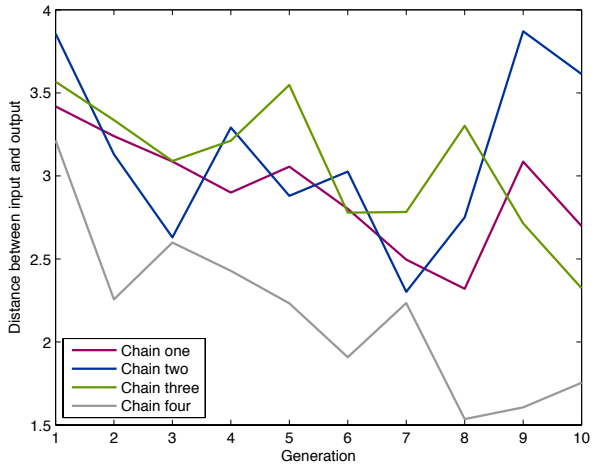

Figure 9: Recall errors for each participant in the four chains. A significant cumulative decrease ($p < 0.001$) is measured with Page's trend test.

Figure 9 shows the measured recall errors for all participants in all four chains. The recall error decreases towards the end of the chain for most chains which means that the learnability of the sets increases. Using Page's trend test (Page, 1963) we can see that there is indeed a significant cumulative decrease of recall error ($L = 1403, m = 4, n = 10, p < 0.001$).

To test whether the reuse of basic elements increases and the signals within a set increasingly share more features and become similar in certain ways, we also compared whistles within a set. In each generation, for all twelve whistles in the set the distance to their nearest neighbor in the same set was computed. The average of these distances was used as the value that would indicate a higher reuse and sharing of features in the case of a lower distance. We are aware of the fact that a lower diversity of signals is not necessarily the result of a higher reuse, but combined with the qualitative analysis this seems very likely in our data and therefore we use it as a suggestive measure.

Figure 10 shows a graph with these values for each chain including the initial set. The graph shows indeed that the signals become more similar to each other and that, increasingly, for most whistles in the set there is another one that is similar in some way. Using Page's trend test (Page, 1963) we can show that there is a significant cumulative decrease in variation among the whistles ($L = 1355, m = 4, n = 10, p < 0.01$), excluding the initial set.

Discussion

The work presented in this paper shows that effects of iterated learning on an artificial sound system in the laboratory can cause this system to become organized in a way that is reminiscent of how speech sounds and signs in sign languages are organized. Superficial combinatorial structure has been observed in the sound systems that emerge by analyzing the results qualitatively. From the increase in learnability, from the way in which participants invent new whistles and from the strategies the participants report to use themselves, we can conclude that it is productively used as well.

Hockett (1960) proposed that a growth in meaning space could be what makes combinatorial structure necessary. However, the high functionality but lack of combinatorial structure evident in ABSL shows that a language can have a large meaning space (histories, dreams, legends and gossips for instance are discussed effortlessly (Sandler et al., 2011)) without needing combinatorial signals. This suggests that combinatorial structure is not necessarily the result of a growing meaning space.

The data we presented here also shows that combinatorial structure can emerge without a growing meaning space. Hockett suggests that the signal space is first fully exploited holistically until signals become too easily confused. Recombination is then needed to expand possibilities while maintaining discriminability. In our experiments combinatorial structure emerges long before the signal space is fully exploited. Even in a system with a very small vocabulary of only twelve signals structure emerges. Similar results have been found in a related study using the visual modality (del Giudice et al., 2010). This finding is therefore not slide whistle, or even modality, specific. What seems to be driving the emergence of structure here relates to learnability. By developing from a holistic system (in which virtually everything is possible within the limits of the modality) towards a discrete and combinatorial system (in which only a few elements can be used and these elements can only be combined in restricted ways) the system becomes more predictable. Without structure, there are no constraints on which whistles are well-formed. This makes the signal space theoretically unrestricted and unpredictable. On the other hand, combinatorial structure limits possibilities and allows learners to focus

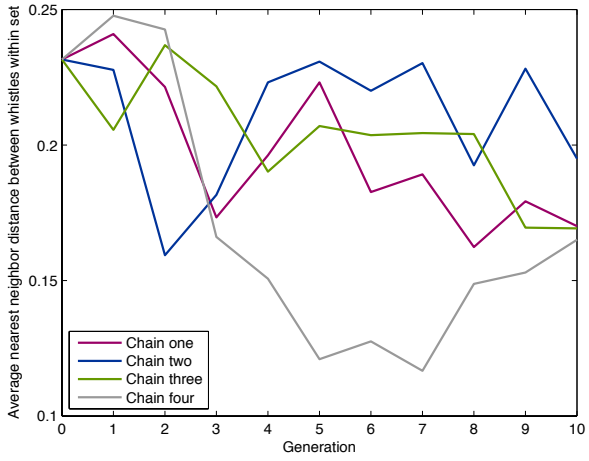

Figure 10: Average nearest neighbor distance between whistles within the set for each generation, including the initial set. A significant cumulative decrease ($p < 0.01$) is measured with Page's trend test.

only on the variations that are linguistically relevant. It has been argued that languages are generally organized to be predictable and Smith and Wonnacott (2010) have shown that the process of iterated learning can cause a linguistic system to lose unpredictable variation. Moreover, the principle of measuring predictability in a linguistic system has also been applied to explain the existence and learnability of complex morphological systems in real languages (Ackerman, Blevins, & Malouf, 2009).

In concert with an increase in predictability, the whistles evolve towards sharing features. This makes it possible to remember them as subsets, which makes learning and recall easier. The idea that chunking of information in this way facilitates encoding more information in short-term memory is well established (Miller, 1956). This strategy was often reported by our participants in a post-test questionnaire.

The whistles that fit the structure and conform to people's cognitive biases are more likely to be preserved from generation to generation in cultural evolution. Combinatorial structure therefore potentially emerges within a gradual (cultural) evolutionary process. This provides an alternative explanation for the origins of combinatorial structure in addition to the existing theory involving an increase in the number of meanings to be expressed.

We hope to have demonstrated a fruitful new approach for studying the influence of iterated learning and cognitive biases on the emergence of structure in speech. One simplification we made in the current design is that the whistles do not convey meanings. The requirement of reproducing twelve unique whistles provides an artificial pressure for expressivity, which would normally result naturally from the need to express distinct meanings. In Kirby et al. (2008) an expressivity constraint was used effectively as well, although in that case there was no absence of semantics. Our results involve a first investigation of combinatorial structure while controlling for effects of semantics (e.g. iconicity, compositionality). Our current work aims to build on the foundations laid by this study by attaching meanings to the whistles.

Acknowledgments

We thank Alex del Giudice, Bart de Boer, Wendy Sandler, Irit Meir and Mark Aronoff for helpful discussions, suggestions and improvements. This research was funded in part by NIH grant RO1 DC6473 to Carol Padden and NWO vidi project 016.074.324 'Modeling the evolution of speech'.

References

Ackerman, F., Blevins, J. P., & Malouf, R. (2009). Parts and wholes: Patterns of relatedness in complex morphological systems and why they matter. In J. P. Blevins & J. Blevins (Eds.), *Analogy in grammar: Form and acquisition* (pp. 54–82). Oxford University Press.

Boersma, P. (2001). Praat, a system for doing phonetics by computer. *Glot international*, *5*(9/10), 341–345.

Christiansen, M. H., & Chater, N. (2008). Language as shaped by the brain. *BBS*, *31*(5), 489–509.

de Boer, B. (2000). Self-organization in vowel systems. *Journal of Phonetics*, *28*(4), 441–465.

de Boer, B., & Zuidema, W. (2010). Multi-agent simulations of the evolution of combinatorial phonology. *Adaptive Behavior*, *18*(2), 141–154.

Deacon, T. W. (1997). *The symbolic species: The co-evolution of language and the brain*. WW Norton & Co Inc.

del Giudice, A., Kirby, S., & Padden, C. (2010). Recreating duality of patterning in the laboratory: a new experimental paradigm for studying emergence of sublexical structure. In A. D. M. Smith, M. Schouwstra, B. de Boer, & K. Smith (Eds.), *Evolang8* (pp. 399–400). World Scientific Press.

Griffiths, T., Kalish, M., & Lewandowsky, S. (2008). Theoretical and empirical evidence for the impact of inductive biases on cultural evolution. *Proc. Trans. R. Soc. B*, *363*(1509), 3503–3514.

Hockett, C. (1960). The origin of speech. *Scientific American*, *203*, 88–96.

Israel, A., & Sandler, W. (2009). Phonological category resolution: A study of handshapes in younger and older sign languages. In A. Castro Caldas & A. Mineiro (Eds.), *Cadernos de saude, vol 2, special issue linguas gestuais* (pp. 13–28). UCP: Lisbon.

Kirby, S., Cornish, H., & Smith, K. (2008). Cumulative cultural evolution in the laboratory: An experimental approach to the origins of structure in human language. *PNAS*, *105*(31), 10681–10686.

Kirby, S., & Hurford, J. (2002). 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). Springer Verlag New York.

Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review*, *63*(2), 343–355.

Oudeyer, P. Y. (2002). Origins and learnability of syllable systems, a cultural evolutionary model. *LNCS*, *2310*, 143–155.

Oudeyer, P. Y. (2005). How phonological structures can be culturally selected for learnability. *Adaptive Behavior*, *13*(4), 269–280.

Oudeyer, P. Y. (2006). *Self-organization in the evolution of speech*. Oxford University Press, USA.

Page, E. B. (1963). Ordered hypotheses for multiple treatments: A significance test for linear ranks. *Journal of the American Statistical Association*, *58*(301), 216–230.

Sakoe, H., & Chiba, S. (2003). Dynamic programming algorithm optimization for spoken word recognition. *IEEE Trans. Acoust., Speech, Signal Process.*, *26*(1), 43–49.

Sandler, W., Aronoff, M., Meir, I., & Padden, C. (2011). The gradual emergence of phonological form in a new language. *NLLT*. (To appear)

Smith, K., & Wonnacott, E. (2010). Eliminating unpredictable variation through iterated learning. *Cognition*, *116*, 444–449.