CULTURAL EMERGENCE OF FEATURE ECONOMY IN AN ARTIFICIAL WHISTLED LANGUAGE

Tessa Verhoef & Bart de Boer

University of Amsterdam, the Netherlands
t.verhoef@uva.nl; b.g.deboer@uva.nl

ABSTRACT

To investigate the origins of combinatorial structure in speech, we conducted an iterated learning experiment with human participants, studying the transmission of a system of whistled signals. Participants learn and reproduce a system of sounds with a slide whistle and their recall output is the input for the next participant. Vertical transmission causes the system to change and become cumulatively more learnable and more structured, yielding increasing combinatorial structure.

Keywords: iterated learning, combinatorial structure, phonology, emergence, learnability

1. INTRODUCTION

Human speech is unique compared to other primate vocalization systems in that it uses a finite number of building blocks to produce a (potentially) unlimited number of utterances [7]. Such *combinatorial structure* occurs at multiple levels: individual speech sounds (phonemes or syllables for example) are combined into words (or morphemes) and words are combined into phrases and sentences. This paper seeks to contribute to the study of how combinatorial structure emerged, and focuses on the level of individual sounds.

Hockett suggested that languages use combinatorial structure because there is a practical limit to the number of holistic signals that can be reliably produced and distinguished [7]. By combining basic signals, an unlimited set of utterances can be produced. In addition, a system of combinatorial signals may be easier to learn than a system without such structure. For combinatorial systems to evolve two steps are needed: the available acoustic/articulatory space needs to be split up into a set of discrete building blocks and these building blocks need to be combined into utterances. The first step of this process has been investigated with computer simulations [1, 13] and is relatively well understood. The second step is harder to comprehend, as it is less well understood how humans split up continuous signals into their basic building blocks. Some insight can be gained from how newly emergent sign languages gain phonological structure [15] but such natural data is very rare.

Zuidema and de Boer [17] have proposed that combinatorial structure can be superficial or productive. Superficial combinatorial structure is present when a system of signals can be analyzed in terms of combinations of building blocks, but the users of the signals are not aware of this structure. Many animal vocalizations have superficial combinatorial structure. Productive combinatorial structure requires that the structure is also used to learn and generalize utterances. Within the order of primates, productive combinatorial structure may be unique to humans [17].

We use the paradigm of experimental iterated learning [5, 6, 9] to study the emergence of productive combinatorial structure. In this paradigm, participants in the laboratory learn a set of utterances from the output of the previous participant, and their output is used as input for the next 'generation'. This models transfer of language from one generation to the next and thus can be used to study (cultural) language evolution experimentally. Moreover, iterated learning tends to amplify learning biases [4, 6, 10] and can give a detailed picture of how humans learn and how learning shapes the system that is being transmitted.

Using this paradigm we seek to investigate whether productive combinatorial structure emerges when humans learn and reproduce small sets of acoustic signals. We expect productive use of combinatorial structure to yield a cumulative increase in learnability and superficial combinatorial structure. We also aim to investigate what strategies humans employ to productively use combinatorial structure and whether these strategies are similar to those used in real human languages. Because we want to avoid linguistic biases, we do not use sets of speech sounds; instead, we use sets of whistled sounds that participants produce with slide whistles.

2. METHODS

The experiment described here involves learning twelve different signals using a slide whistle (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.

2.1. Procedure

The participants completed four rounds of learning and recall. In the learning phase they were exposed to all signals one by one, and asked to imitate each sound with the slide whistle immediately. After this, a recall phase followed in which they reproduced all twelve whistles from memory. The input stimuli consisted of the output the previous participant produced in the last recall round (or the initial input set which exhibited no combinatorial structure). More detail is reported in [16].

2.2. Participants

Forty participants took part in four parallel ten-generation chains of learning and recall. All participants were university students from either University of California San Diego, or University of Amsterdam, ranging in age from 18 to 32 (μ=22).

3. RESULTS

3.1. Qualitative results

Behaviors that are observed in the recall phase eventually lead to an increase of structure. Remembering twelve whistles after only four exposures is difficult, so participants generally do not recall all of them flawlessly. They appear to over-generalize some of the (superficial) combinatorial structure that they perceive. This results in the introduction of whistles that are related in form to other learned whistles: 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 an inventory 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.

Figures 2 and 3 show whistles plotted as pitch tracks on a semitone scale using Praat [2]. Figure 2 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 co-articulation, because the final pitch of the first part influences the initial pitch of the second part. Figure 3 shows combined mirroring, repetition and borrowing from chain four, which results in a predictable system that is stable and persists after its innovation.

Figure 2: Example of recombination and co-articulation (see section 3.1).

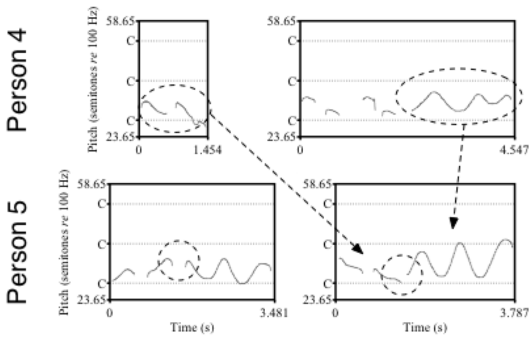
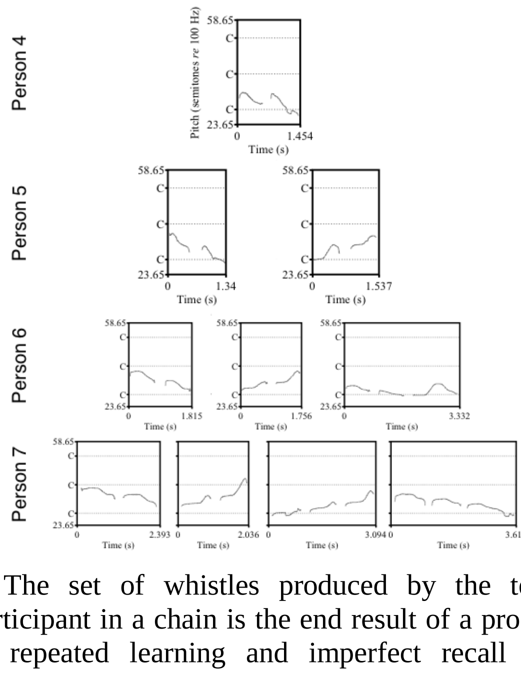
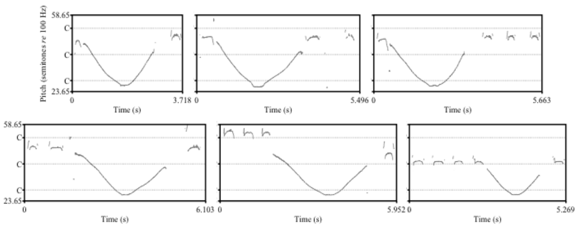


Figure 3: An example of cumulative mirroring, repetition and borrowing (see section 3.1).



The set of whistles produced by the tenth participant in a chain is the end result of a process of repeated learning and imperfect recall and shows the cumulative effect of the mirroring, borrowing and repetition behaviors. Figure 4 shows part of the tenth set of chain one. In this set we can 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.

Figure 4: Example of reuse of basic elements in the last set of chain one (see section 3.1).



When comparing outcomes of different chains, there appear to be ‘language’-specific constraints on the ways the elements are reused. In this first chain, for instance, short single notes always follow each other on the same pitch and slides always go down first, never up-down. Constraints also exist in the other chains, but these are different. Overall, similar patterns of borrowing, mirroring and reuse are found in all four chains, resulting in systems that exhibit similar degrees of combinatorial structure, which is realized in different ways.

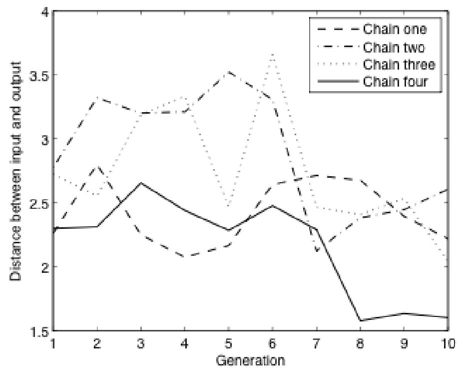
In summary: qualitatively we can see an increase in the reuse of basic whistle elements in the sets. Once whistles that are composed of these elements appear in the set, they are more likely to be learned and recalled by later generations who use the similarities across whistles to group them as subsets,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

3.2. Quantitative results

Combinatorial systems are supposedly easier to learn than systems without such structure. In order to check this, we measured the distance between the input set and the output set for each participant in each chain. We would expect the recall error to be 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 a measure including Derivative Dynamic Time Warping [8] of pitch tracks and intensity tracks (for more details, see [16]). The distance between two sets of twelve whistles is then the sum of the distances in the set of 12 distinct pairs (where each whistle is paired with a unique whistle from the other set) for which this distance is minimal.

Figure 5 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. Page’s trend test [14] shows that there is a significant cumulative decrease of recall error ($L = 1362$, $m = 4$, $n = 10$, $p < 0.01$).

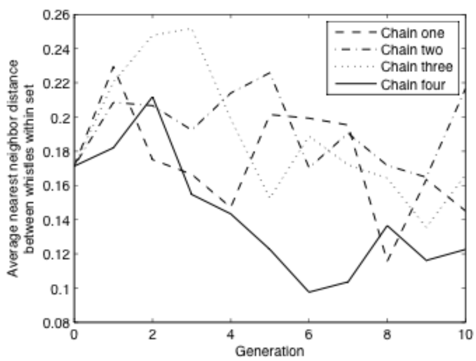
Figure 5: Recall errors for participants in all chains (see section 3.2).



To test whether the reuse of basic elements increases and the signals within a set increasingly share more features, we also compared whistles within a set. In each generation, for all twelve whistles in the set the distance to their nearest neighbor was computed. The qualitative analysis suggests that in our experiments lower average distance indicates more reuse and sharing of features (although in general lower average distance is not always the result of higher reuse).

Figure 6 shows the average distance values for each chain including for the initial set. The signals become more similar to each other and, increasingly, for most whistles in the set there is another one that is similar for some features. The decrease in variation among the whistles, excluding the initial set, is significant according to Page’s trend test ($L = 1407$, $m = 4$, $n = 10$, $p < 0.001$).

Figure 6: Average nearest neighbor distance between whistles (see section 3.2).



4. DISCUSSION

The work presented in this paper shows that experimental iterated learning can cause an artificial whistled language to become organized in a way that is reminiscent of how speech sounds are organized. Qualitative analysis of the systems of signals shows that superficial combinatorial structure emerges. From the increase in learnability, from the way in which participants invent new whistles and from the strategies the participants report using, we can conclude that it is productively used as well. The way participants use combinatorial structure productively seems similar to how it is used in language (combination of building blocks and co-articulatory effects).

Hockett [7] proposed that a growth in meaning space could be what makes combinatorial structure necessary. He suggests that the signal space is first fully exploited holistically until signals become too easily confused. Recombination is then needed to expand expressivity 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 [5]. This finding is therefore not specific to slide whistles or modality. What seems to be driving the emergence of structure is 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 combined in restricted ways) the system becomes more predictable. The whistles that fit the structure and conform to people's expectations are more likely to be learned and preserved over generations.

Although the *utterances* appear to become more similar in the process of cultural evolution, the *building blocks* tend to become more different. In that respect the experimental results are not in contradiction with the importance of distinctiveness in the formation of combinatorial repertoires of signals. However, the formation of building blocks here does not resemble the simplest dispersal models [1, 11], but is more reminiscent of the 'Maximal Utilisation of Available Distinctive Features' principle proposed by Ohala [12] or 'feature economy' (see [3]). If a building block is present, it tends to get mirrored and reused before new ones appear.

The experiment presented here has only scratched the surface of the question of emergence of combinatorial structure. However, we hope to have shown that the experimental paradigm of iterated learning provides a useful new instrument to investigate this question.

5. AKNOWLEDGEMENTS

We thank Simon Kirby, Carol Padden and Alex del Giudice for helpful discussions. This research was funded in part by NIH grant RO1 DC6473 and NWO vidi project 016.074.324.

6. REFERENCES

[1] de Boer, B. 2000. Self-organization in vowel systems. *J. Phon.* 28(4), 441-465.

[2] Boersma, P. 2001. Praat, a system for doing phonetics by computer. *Glot International* 5(9/10), 341-345.

[3] Clements, G.N. 2003 Feature economy in sound systems, *Phonology* 20, 287-333.

[4] Deacon, T.W. 1997. *The Symbolic Species: The Co-evolution of Language and the Brain*. WW Norton & Co.

[5] del Giudice, A., Kirby, S., Padden, C. 2010. Recreating duality of patterning in the laboratory. In Smith, A.D.M., Schouwstra, M., de Boer, B., Smith, K. (eds.), *EVOLANG* 8, 399-400.

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

[7] Hockett, C. 1960. The origin of speech. *Scientific American* 203, 88-96.

[8] Keogh, E.J., Pazzani, M.J. 2001. Derivative dynamic time warping. *Proc. SDM'2001* Chicago.

[9] 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.

[10] Kirby, S., Hurford, J. 2002. The emergence of linguistic structure: an overview of the iterated learning model. In Cangelosi, A., Parisi, D. (eds.), *Simulating the Evolution of Language*. Springer Verlag NY, 121-148.

[11] Liljencrants, J., Lindblom, B. 1972. Numerical simulations of vowel quality systems. *Language* 48, 839-862.

[12] Ohala, J.J. 1980. Moderator's introduction to symposium on phonetic universals in phonological systems and their explanation. *Proc. ICPhS IX* Copenhagen, 3, 181-185.

[13] Oudeyer, P.Y. 2006. *Self-organization in the Evolution of Speech*. Oxford University Press, USA.

[14] Page, E.B. 1963. Ordered hypotheses for multiple treatments: A significance test for linear ranks. *J. Am. Stat. Assoc.* 58(301), 216-230.

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

[16] Verhoef, T., Kirby, S., Padden, C. 2011 Cultural emergence of combinatorial structure in an artificial whistled language. In Carlson, L., Hölscher, C., Shipley, T. (eds.), *Proc. CogSci 33*, Austin, TX (to appear).

[17] Zuidema, W., de Boer, B. 2009. The evolution of combinatorial phonology. *J. Phon.* 37(2), 125-144.