# C E N T E R F O R R E S E A R C H I N L A N G U A G E

**March 2011** **Vol. 23, No. 1**

CRL Technical Reports, University of California, San Diego, La Jolla CA 92093-0526
Tel: (858) 534-2536 • E-mail: editor@crl.ucsd.edu • WWW: http://crl.ucsd.edu/


**TECHNICAL REPORT**

## Cultural evolution of combinatorial structure in ongoing artificial speech learning experiments


**Tessa Verhoef[1], Bart de Boer[1], Alex del Giudice[2], Carol Padden[2] & Simon Kirby[3]**

[1] University of Amsterdam
[2] University of California San Diego
[3] University of Edinburgh

*Address for correspondence:*

Tessa Verhoef, t.verhoef@uva.nl


## EDITOR'S NOTE

This newsletter is produced and distributed by the **CENTER FOR RESEARCH IN LANGUAGE**, a research center at the University of California, San Diego that unites the efforts of fields such as Cognitive Science, Linguistics, Psychology, Computer Science, Sociology, and Philosophy, all who share an interest in language. We feature papers related to language and cognition (distributed via the World Wide Web) and welcome response from friends and colleagues at UCSD as well as other institutions. Please visit our web site at `http://crl.ucsd.edu`.

## SUBSCRIPTION INFORMATION

If you know of others who would be interested in receiving the Newsletter and the Technical Reports, you may add them to our email subscription list by sending an email to `majordomo@crl.ucsd.edu` with the line "`subscribe newsletter <email-address>`" in the body of the message (e.g., `subscribe newsletter jdoe@ucsd.edu`). Please forward correspondence to:


Jamie Alexandre, Editor
Center for Research in Language, 0526
9500 Gilman Drive, University of California, San Diego 92093-0526
Telephone: (858) 534-2536 • E-mail: editor@crl.ucsd.edu

Back issues of the the CRL Newsletter are available on our website. Papers featured in recent issues include:

*Effects of Broca's aphasia and LIPC damage on the use of contextual information in sentence comprehension*
**Eileen R. Cardillo**
CRL & Institute for Neural Computation, UCSD
**Kim Plunkett**
Experimental Psychology, University of Oxford
**Jennifer Aydelott**
Psychology, Birbeck College, University of London)
Vol. 18, No. 1, June 2006

*Avoid ambiguity! (If you can)*
**Victor S. Ferreira**
Department of Psychology, UCSD
Vol. 18, No. 2, December 2006

*Arab Sign Languages: A Lexical Comparison*
**Kinda Al-Fityani**
Department of Communication, UCSD
Vol. 19, No. 1, March 2007

*The Coordinated Interplay Account of Utterance Comprehension, Attention, and the Use of Scene Information*
**Pia Knoeferle**
Department of Cognitive Science, UCSD
Vol. 19. No. 2, December 2007

*Doing time: Speech, gesture, and the conceptualization of time*
**Kensy Cooperrider, Rafael Núñez**
Depatment of Cognitive Science, UCSD
Vol. 19. No. 3, December 2007

*Auditory perception in atypical development: From basic building blocks to higher-level perceptual organization*
**Mayada Elsabbagh**
Center for Brain and Cognitive Development, Birkbeck College, University of London
**Henri Cohen**
Cognitive Neuroscience Center, University of Quebec
**Annette Karmiloff-Smith**
Center for Brain and Cognitive Development, Birkbeck College, University of London
Vol. 20. No. 1, March 2008

*The Role of Orthographic Gender in Cognition*
**Tim Beyer, Carla L. Hudson Kam**
Center for Research in Language, UCSD
Vol. 20. No. 2, June 2008

*Negation Processing in Context Is Not (Always) Delayed*
**Jenny Staab**
Joint Doctoral Program in Language and Communicative Disorders, and CRL
**Thomas P. Urbach**
Department of Cognitive Science, UCSD
**Marta Kutas**
Department of Cognitive Science, UCSD, and CRL
Vol. 20. No. 3, December 2008

*The quick brown fox run over one lazy geese: Phonological and morphological processing of plurals in English*
**Katie J. Alcock**
Lancaster University, UK
Vol. 21. No. 1, March 2009

*Voxel-based Lesion Analysis of Category-Specific Naming on the Boston Naming Test*
**Juliana V. Baldo, Analía Arévalo, David P. Wilkins**
Center for Aphasia and Related Disorders, VANCHCS
**Nina F. Dronkers**
Center for Aphasia and Related Disorders, VANCHCS
Department of Neurology, UC Davis
Center for Research in Language, UC San Diego
Vol. 21. No. 2, June 2009

*Phonological Deficits in Children with Perinatal Stroke: Evidence from Spelling*
**Darin Woolpert**
San Diego State University
University of California, San Diego
**Judy S. Reilly**
San Diego State University
University of Poiters
Vol. 21. No. 3, December 2009

*Dynamic construals, static formalisms: Evidence from co-speech gesture during mathematical proving*
**Tyler Marghetis, Rafael Núñez**
Department of Cognitive Science, UCSD
Vol. 22. No. 1, March 2010

*Re-mapping topographic terms indoors: A study of everyday spatial construals in the mountains of Papua New Guinea*
**Kensy Cooperrider, Rafael Núñez**
Department of Cognitive Science, UCSD
Vol. 22. No. 2, June 2010

# CULTURAL EVOLUTION OF COMBINATORIAL STRUCTURE IN ONGOING ARTIFICIAL SPEECH LEARNING EXPERIMENTS

**Tessa Verhoef**[1], **Bart de Boer**[1], **Alex del Giudice**[2], **Carol Padden**[2] **& Simon Kirby**[3]

[1] University of Amsterdam
[2] University of California San Diego
[3] University of Edinburgh

## Abstract

Speech sounds are organized: they are both categorical and combinatorial and there are constraints on how elements can be recombined. To investigate the origins of this structure, we conducted an iterated learning experiment with humans, studying the transmission of artificial systems of sounds. In this study, participants learn a system of sounds that are produced with an interface in which they draw trajectories on a computer screen in a continuous two-dimensional space. These trajectories are transformed into sounds. Through transmission from participant to participant, some structure emerged, but it turned out not to be stable, most probably because the learning task was too difficult. Even though the results were not entirely as expected, they were promising and led to the ideas for a follow-up, ongoing study involving transmission of a whistled sound system. A preview will be given into the first results of this second study, which shows that experimental iterated learning of an artificial sound system can cause a system of signals to gain combinatorial structure.

Keywords: *Experimental iterated learning, combinatorial structure, evolution of speech, emergence.*

**Tessa Verhoef**[1], **Bart de Boer**[1], **Alex del Giudice**[2], **Carol Padden**[2] **& Simon Kirby**[3]

[1] University of Amsterdam
1012 VT Amsterdam, The Netherlands
[2] University of California San Diego
La Jolla, CA 92093-0503, USA
[3] University of Edinburgh
Edinburgh EH8 9AD, UK

## Introduction

Sounds of human language are structured. They are both discrete and combinatorial (Oudeyer, 2005b; Fitch, 2010). Discreteness means that the continuous acoustic space of sounds that can be produced is organized into a finite set of basic building blocks. The combinatorial nature of speech means that these elements are reused and recombined in a systematic way. There are constraints on how the elements can be recombined, and the specifics of these constraints and the basic elements differ from one language to the other, but are shared among all members of a speech community. How did speech become organized in this way? This is the question we will address in this report.

Hockett (1960) identified the discrete and combinatorial organization of speech as one of thirteen basic design features of language. He called it 'duality of patterning' and it refers, in part, to how meaningless phonemes are recombined into grammatical morphemes. Hockett (1960) already had an idea about why it would be an advantage if such a structure were present in language. When the meaning space grows and each signal refers to its meaning as a whole, the signal space for creating these holistic signals will fill up and the individual signals will become closer to each other. 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. Therefore structural recombination of elements is needed to maintain clear communication with a growing meaning space. When Hockett wrote his paper (1960) there was no data available about the origins of combinatorial structure that could be used as evidence in favor or against such a hypothesis. Now, data is accumulating, for instance from the study of emerging sign languages, from the use of computer simulations and from experiments in the laboratory.

A newly emerging sign language, Al-Sayyid Bedouin Sign Language (ABSL), is currently being studied and shows the emergence of phonological structure (Israel and Sandler, 2009; Sandler et al., 2011). Sign languages usually

have phonological structure with the same features of discreteness and recombination as speech. There is a discrete set of location features, handshape features and movement features that are recombined into meaningful words and there are constraints on the ways in which they can be combined. ABSL is a young but fully functional sign language in which the phonological structure is not yet fixed (Israel and Sandler, 2009). This research provides important evidence about the emergence of phonological structure since it throws light on how such structure can develop, for instance through conventionalization processes within families of signers (Sandler et al., 2011).

In addition to observations that can be made of real language data, computational models provide insight into the emergence of phonological structure. For instance, such simulations have shown how a discrete set of vowel categories can emerge (de Boer, 2000; Oudeyer, 2005b), how the organization of syllable systems can be established (Oudeyer, 2005a) and how it is possible for a system of holistic signals to turn into a system with combinatorial structure (de Boer and Zuidema, 2010). Typically, in these simulations, there is a population of interacting computer agents and cultural evolution is studied by simulating conventionalization through social coordination and/or transmission through iterated learning. Social coordination involves establishment of a shared communication system through interactions between the agents in the population, and iterated learning involves (repeated) acquisition of a behavior by an agent through observation of the same behavior by another agent that acquired it in the same way (Kirby et al., 2008)

Unfortunately newly emerging (sign) languages are extremely rare and computer models generally abstract away from the full complexity of the human brain. In order to verify computer modeling findings experimentally, Kirby et al. (2008) introduced iterated learning experiments 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 of using real human learners, while still being able to control the environment, the parameters of the artificial language and the level of complexity, which is impossible to achieve in field studies of emerging languages. The idea is to create a chain of learners in which the outcome of the learning process of one participant is used as the input for the next person (Kirby et al., 2008). As each succeeding learner processes the previous learner's output, it is thought that the system itself will be shaped by the biases, expectations, and constraints of the learners (Deacon, 1997; Kirby and Hurford, 2002; Christiansen and Chater, 2008; Griffiths et al., 2008). This idea has been shown to work in numerous computer simulations and is now starting to be investigated experimentally. In experiments it has for instance been applied successfully to show the emergence of compositional structure (Kirby et al., 2008), combinatorial structure in visual symbols (del Giudice et al., 2010), color terms (Dowman et al., 2008), predictability in plural marking (Smith and Wonnacott, 2010) and in other category or function learning tasks (Griffiths et al., 2008). The current study employs this paradigm to examine the emergence of discrete combinatorial elements in speech by studying the transmission of artificial systems of sounds.

## Methods

The experiments described in this section were conducted within the experimental iterated learning paradigm (Kirby et al., 2008; Cornish, 2006). Participants had to learn an artificial system of sounds and the result of their learning was used as input for the next participant. Four parallel transmission chains were performed, with several successive learners in each chain.

### Participants

In total, 38 people participated in this study. Test subjects were recruited from the student population of the University of Amsterdam. 25 participants were female, 13 male and the mean age was 26.7. The participants were first asked to do a very short hearing test. All subjects had normal hearing. Participants were paid 10 euro in cash to compensate for their time.

### Stimuli

The sounds in the system of signals that was transmitted were produced by drawing continuous trajectories on a computer screen. The trajectories (and hence the signaling space) consisted of a single, continuous line in a two-dimensional space. These trajectories were transformed into sounds. Participants needed to learn to recognize and reproduce these sounds by drawing the right trajectories. In addition, these sounds (creating the signal space) were used as labels for different pictures (creating the meaning space) and the participants had to learn these sound-picture relationships.

**Signal space** Participants create sounds by scribbling trajectories. A trajectory is produced by placing the mouse pointer in the scribble area, pressing the mouse button, drawing (scribbling) the trajectory, and releasing the mouse button to indicate the trajectory is finished. The transformation of scribbles into sounds uses a mapping that resembles a vowel chart representation. Different locations in the scribble area sound like different vowel sounds. Vertical movements in the scribble space manipulate the first formant (increasing from 250 Hz to 1050 Hz when moving down) and horizontal movements manipulate the second formant (decreasing from 2900 Hz to 1100 Hz when moving from left to right). This creates a two-dimensional continuous space with differing vowel qualities. The participants were not told beforehand that they were going to create vowel trajectories, they had to discover this themselves.

Figure 1 shows a screenshot of the experiment with an explanation of the mapping in the scribble space. At the beginning of the experiment a random set of sounds was created by letting the computer draw random trajectories in the scribble space. This set of random sounds was used as input in the training set of the first person of each transmission chain. In order to measure the accuracy of an imitation of the sounds, a distance measure for comparing trajectories was needed. We used a normalized Dynamic

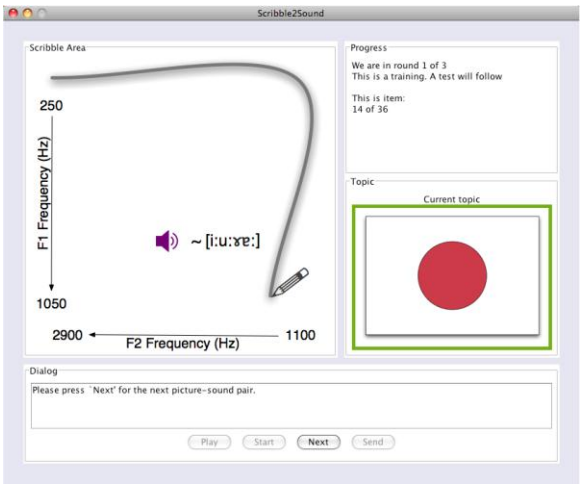



*Figure 1: Screenshot of the experiment, including a representation of the sound mapping. Note that participants did not get to see the axes or transcriptions, the Scribble Area was entirely empty.*

Time Warping Distance (Sakoe and Chiba, 2003) on the sequences of *x*, *y* coordinates in the scribble space to determine this distance.

**Meaning space** The meaning space consisted of nine pictures of different objects (squares, circles and rings) that had different colors (red, green or blue). Figure 2 shows these pictures. At the beginning of the experiment, each picture in the meaning space was randomly assigned to a unique sound in the set of random sounds in the signal space to create the initial set of sound-meaning pairs.

## Procedure

Before the experiment started the task was explained to the participants, both verbally by the experimenter and in written form on the screen. The participants were given a chance to ask questions before we started with the practice phase. In this phase the subjects were asked to familiarize themselves with the scribble area. They were given 30 trials in which they could explore the space by producing different scribbles and hearing the sounds they produced with these trajectories. After the practice phase, the real experiment started. The experiment consisted of three rounds of training and testing. Each round started with a training phase in which the participants were exposed to the training set six times, each time in a different random order. This means that they were shown the picture, heard the sound that labeled this picture and were given one chance to imitate the sound. Feedback on the imitation accuracy was provided by showing a colored border around the picture, which gradually changed from red to green when the imitation became more accurate. Then, in round one and two a short test of five items followed in which only the picture was shown and the participants had to reproduce the right sound from their memory. After the third training phase, a longer test followed which included all nine meanings. The signal productions in this last test were used as input for the next participant. After completing the final test, the participants were asked to provide feedback about their own performance and experience. The first two chains consisted of ten participants in each chain. Later chains were slightly shorter (as described below).

**Learning bottleneck** As has been shown with the use of computer models studying iterated learning and previous experimental iterated learning studies, the emergence of structure relies on the poverty of the stimulus (Smith et al., 2003). When a learner is not exposed to every possible expression during acquisition, there is a learning bottleneck (Smith et al., 2003). It has been shown that as a result of such a bottleneck in transmission, structure emerges both in computer simulations (Smith et al., 2003) and in experiments with humans (Kirby et al., 2008) for instance because expressions for new items are constructed by generalization on the learned signals. In the experiment

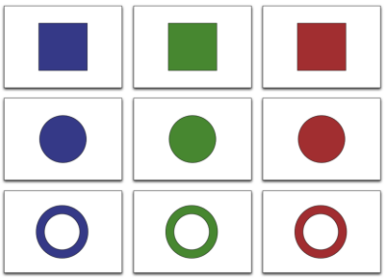


*Figure 2: Meaning space.*

described in this report the learning bottleneck was introduced by training the participants on only six out of the total of nine sound-meaning pairs in the training phase, but testing them in the final test on all nine pairs.

## Modifications

After the first two diffusion chains were completed we could make a few observations that led to two different adjustments in the third and fourth chain. The first involved the addition of another task in the testing phases and the second involved the introduction of adaptive learning in the training phases.

**Guessing task** We observed that some participants were paying very little attention to the sounds during the task. Once they thought they had discovered which trajectory would give them a reasonable feedback, they would remember this trajectory and its relation to the right picture. During post-test questioning, participants sometimes reported that they stopped listening to the sounds once they remembered what they thought were the right gestures. In order to make sure that the participants would not start to ignore the sounds, an additional task was included in the testing phase. This task was a guessing task in which a sound was played and four pictures were shown, one of which belonged to the sound. The participant was asked to choose the right picture. This modification was added in the third chain. This chain consisted of only 6 generations.

**Adaptive learning** Another observation we made was that participants had much difficulty learning to imitate sounds in the task. Their performance on most items stayed very poor throughout the course of the experiment and therefore an alternative learning structure was introduced, using adaptive learning. In this version, the participants would not be exposed to the complete training set at the beginning of the experiment, but the number of items they were trained on grew according to the imitation performance. At first, training would occur on only two different items. Then, when the participant was able to imitate those two closely enough, another example was added and so on. This modification was added in the fourth chain.

## Hypothesis

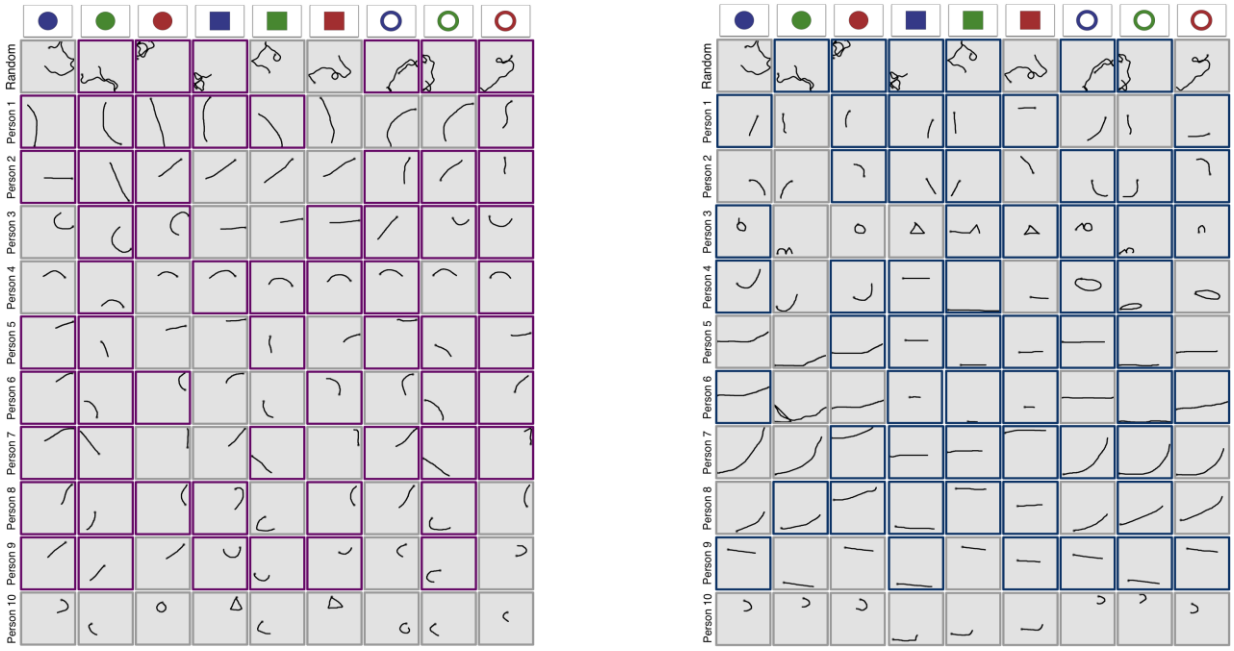
At the end of each transmission chain we expect to find an increase in the amount of structure in the systems of sounds that were transmitted. We would call this structure combinatorial if it consists of a systematic reuse of basic building blocks in the sounds. It has been shown before that the mechanism of (human) iterated learning can lead to the emergence of compositional structure (Kirby et al., 2008; Kirby and Hurford, 2002) and our hypothesis is that it will lead to structure on the sub-lexical, phonetic level as well (de Boer and Zuidema, 2010). In addition, we expect to find an increase in the learnability of the set of signals as the chain progresses, because the sound systems change to become optimized for learnability. When the system is more structured, and only the sounds that are remembered easily persist in the system, participants are expected to learn faster and perform better.

# Results

In this section we will first present the qualitative results, showing the development of the sound systems from generation to generation. This will give insight into the kinds of structure that did and did not occur. Second, we will present quantitative data, showing how the learning ability changed over the course of each chain.

## Qualitative results

In figure 3 the outputs in the first two chains are shown. The first row shows the trajectories for the random input sounds and then each row shows the output of a participant who got the previous row as input. The colored border around the picture means that this item was part of the training set for the next person. This person was not trained, but only tested, on the other three. Note that the participants never saw the actual scribbles. Only the sounds were transmitted, as was their relation to one of the pictures in the meaning space.




(a) Chain one

(b) Chain two

*Figure 3: The first row shows the trajectories for the random input sounds and then each row shows the output of a participant who got the previous row as input. The colored border around the picture means that this item was part of the training set for the next person.*

In both chains it can be observed that often the same signals are used for all objects with the same color or shape and right from the beginning there seems to be a tendency to search for patterns and apply generalizations. Often features such as the length of the sound, or the location of the trajectory in the space (influencing vowel quality) are linked to colors or shapes in the pictures. For instance in generation one of chain one, the trajectories that had to be created for the unseen pictures in the last test were often based on, or almost the same as the ones that were remembered for the seen pictures that had the color or shape in common. The red square, for instance, starts to be indicated by a trajectory going down, like the red circle and the blue square, while the green square gets a trajectory going up, like the green circle.

But in this first chain it is not until generation nine that more than one dimension in the picture (color and shape) is distinguishably indicated in the signals (see figure 4). For person nine, all circles are expressed as straight lines, squares as cup-shaped trajectories and rings as hooks. Green colored shapes are indicated by the use of the lower left corner, the others by the use of the upper right corner in which the trajectories for blue go in the opposite direction

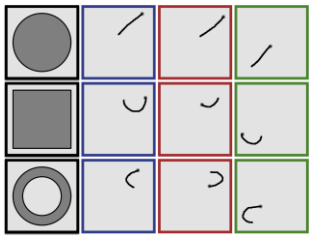


*Figure 4: Chain one, generation nine. Note that the shape of the trajectory appears to express the shape of the object, while the position of the trajectory expresses the color of the object.*

from those for red (except for the circle, but this happened only in this last output round, it was consistent in previous rounds). The type of structure that emerges in chain one does not persist in the chain, not even over one generation and the structure appears to be more visually oriented than auditory. We will come back to this observation in the discussion section.

In chain two, the first hints of structure appear in generation two (see figure 5). In this set, the location of the scribbles is clearly linked to the colors of the pictures in the meaning space. Red objects are always linked to scribbles in the upper half of the scribble space (corresponding to close/close-mid vowel sounds), green objects are linked to

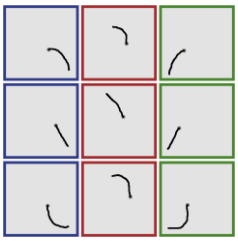


*Figure 5: Chain two, generation two. Note that the location of the trajectory indicates the color of the object in the meaning space.*

scribbles in the lower left corner (corresponding to open, front vowel sounds) and blue objects are linked to scribbles in the lower right corner (corresponding to open, back vowel sounds).

Then in generation four more structure emerges when the shape of the scribble is also used to make a meaningful distinction between different shapes in the meaning space (see figure 6). The structure that appeared in generation 4 was learned almost perfectly by the next person, except for the fact that the sounds for the ring shaped meanings did not stay the same. Only one (very clearly audible) feature that distinguished rings and squares in generation four was adopted by the next person, namely the longer duration of the sound. Following this, in generation six the structure is learned perfectly and even the sounds created for the unseen objects are correct.

In chain three we added the additional guessing task in response to the observation that participants did not pay much attention to the sounds during the experiment. The results in this chain were qualitatively the same to those in the first two chains and there was no noticeable difference in listening behavior. In the discussion section we will explain why we think this happened.

In chain four an adaptive learning regime determined the amount of training items that were presented at each time during the experiment, with a growing training set when the performance improved. While we thought this regime would help the participants to learn the sound-meaning pairs better, it actually revealed even more strikingly how difficult the learning task was. It turned out that about half of the participants did not progress beyond the initial stage in which there were only two training items in the set. Therefore the output data of most participants who did this version could not be used as input for the next person, because the learning bottleneck was simply too tight.

In summary, the qualitative results indicate that the structures that emerged did not persist throughout the chain until the end. We will explain why we think this happened and discuss these issues further in the discussion section.

## Quantitative results

In order to find out whether the sound-meaning systems were optimized to become more learnable by being transmitted through chains of human learners, we measured

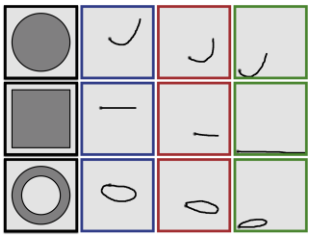


*Figure 6: Chain two, generation four. Note that the shape of the trajectory appears to express the shape of the object, while the position of the trajectory expresses the color of the object.*

the performance from generation to generation in each chain. For each participant the distance between the input set and the output they created for each meaning was measured, by using the distance measure as described above. Figure 7 shows these measures for the first three chains in three different situations: at the beginning of the experiment including only the training set, at the end of the experiment including only the training set, and at the end of the experiment including the complete set (also the three meanings they were never trained on). In the case that the average distance between input and output is approximately the same on the training and test set, it means that the participant performed just as well on the meanings they never saw as on the other six. This therefore probably means that this person generalized by using the structure to decide on the sounds for the unseen meanings. Figure 7 shows that this happens only a few times throughout the chains. It is clear that there is a relationship between the emergence of

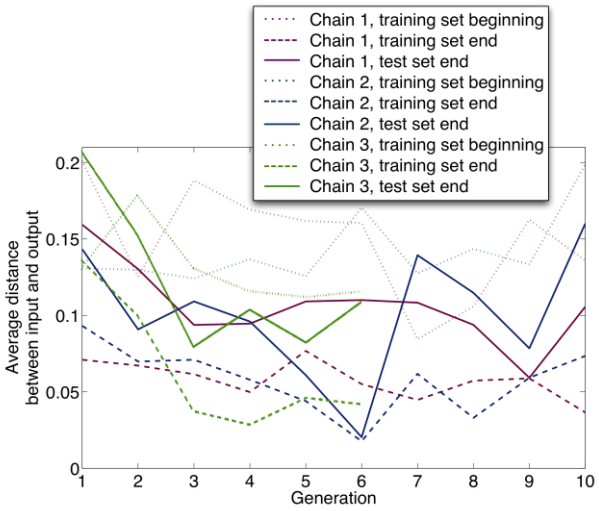



*Figure 7: Average distance between input and output for chain one, two and three in three different situations: at the beginning of the experiment including only the training set, at the end of the experiment including only the training set, and at the end of the experiment including the complete set (also the three meanings they were never trained on).*

structure and the increase of learnability (decrease of average distance). In chain one for instance, the performance on the complete set increases from generation seven to generation nine, where the performance is the same on the complete set and on the training set alone. This coincides with the appearance of structure in generation 7 and 8 where location in the scribble area is linked to color in the meaning space. Person nine uses this structure to create sounds for unseen meanings. In chain two we can see a similar development starting in generation four. With the emergence of the structure that was described in the qualitative results, the performance on the complete set increases over the next few generations. In generation six, the performance is again the same on the complete set and on the training set alone, indicating that this person could guess the right sounds for unseen meanings by using generalization.

Even though it happens a few times that learnability increases rapidly from generation to generation, it does not persist throughout the entire chain until the end. Just as the structure that sometimes emerges disappears again, the increased learnability disappears with it.

## Discussion and future direction

The series of experiments described in this report was intended as a first investigation of the emergence of combinatorial structure in speech-like signals. With this first attempt to study the cultural evolution of an artificial sound system in the laboratory, we expected to find an increase in learnability of the systems that were being transmitted, as well as an increase of the combinatorial structure within the systems. Although these improvements could be observed qualitatively as structure emerged from time to time and survived for a few generations, structure did not emerge as a permanent feature, nor was there a cumulative increase of learnability or of the degree to which combinatorial structure was present. The disappearance of structure was probably caused by the difficulty of the learning task, causing some participants to erase structure that emerged previously. The difficulty of using the scribble area interface caused a tight learning bottleneck in this experiment, which hindered transmission and emergence of structure. However, the results are promising, because there were a few participants who had less difficulty with the task and in these cases generalization and introduction of structure happened in the way we expected. These participants were mostly familiar with the vowel chart (for instance due to courses they followed in phonetics/phonology), which provided them with a mental map that made the task cognitively easier. The current findings are useful for considerations in future work, in which a more intuitive sound production interface and a less narrow bottleneck are needed.

One problem with the current study involves the analysis of the results and the relation to the original question of the emergence of combinatorial/sub-lexical structure. Structure does occur from time to time, but this structure cannot immediately be paralleled to combinatorial phonology. Actually it is more comparable to syntactic compositional structure, because the location and shapes in the scribble space are directly linked to colors or shapes in the meaning space. The building blocks are therefore meaningful and the structure compositional. There is no observable further recombination below this level. We are interested in the emergence of structure that is sub-lexical and more like ‘bare phonology’ (Fitch, 2010), but the use of a very structured meaning space in this study did not yield combinatorial structure of this kind.

Furthermore, the structure that emerges appears mainly in the visual modality. The use of location in the scribble area (manipulating vowel quality) creates audible distinctions, but it can also be observed that sometimes structure emerges that is visible when inspecting the scribbled trajectories, but involves barely audible distinctions in the auditory modality. An example is shown in figure 4. This figure shows the entire set of generation nine in the first chain. In this set the location in the scribble area is used to distinguish green colored objects from the others, while the shape of the trajectory scribbled indicates the shape of the object: a straight line for the circles, a cup-shaped trajectory for the squares and a hook-shaped trajectory for the rings. The use of location (and therefore the manipulation of vowel quality) is clearly audible, but the subtle differences between hook-shapes and cup-shapes for instance, are clearly visible, but barely audible. Since the learners in each chain are never exposed to the scribbled trajectories, but only to the sounds, a logical consequence is that this type of inaudible structure does not persist into following generations.

Why do participants focus so much on the visual modality and ignore the sounds? We think this is due to the feedback we give participants when they imitate the sounds. By providing feedback after imitation, a possibility is created for participants to solve the task without listening at all. They can directly focus on and remember the visual trajectory-meaning pairs that work well and result in positive feedback. This may be a more direct and easy memory task than having to remember sound-meaning pairs in addition to having to know how to produce these sounds in a multi-modal fashion. As mentioned before, we observed that some participants did not pay enough attention to the sounds, which confirms this concern.

The fact that part of the emerged structure was imperceptible is not the only factor in this experiment that hindered transmission and persistence of the structure in the sound sets. The learning task also appeared to be very difficult, especially because it was hard for participants to figure out how to reproduce the sounds by drawing trajectories. This may have been caused by the fact that the scribble area was a very unnatural interface for the production of sounds and on top of this it involved a multi-modal task with a difficult to interpret visual-auditory mapping (at least for people unfamiliar with the vowel space). The difficulty of the task became especially clear in chain four with the addition of active learning.

Even though there were issues about the experiment described above that did not turn out as expected, the results are interesting and informative as a first attempt to experimentally investigate the emergence of structure in speech sounds. Learning did take place and structure did emerge from time to time. These results definitely shed light on many important issues that need to be considered in future designs, such as the need for a more intuitive sound production interface to make sure the learning bottleneck will not be too narrow and the use of a less structured meaning space, or no meaning space at all. The lessons learned from this study gave rise to ideas for a follow-up experiment. A sneak peek of this work is presented in the next paragraph.

*Figure 8: Slide whistle*

## Future direction: a whistled sound system

Given the difficulty participants had in learning to use the

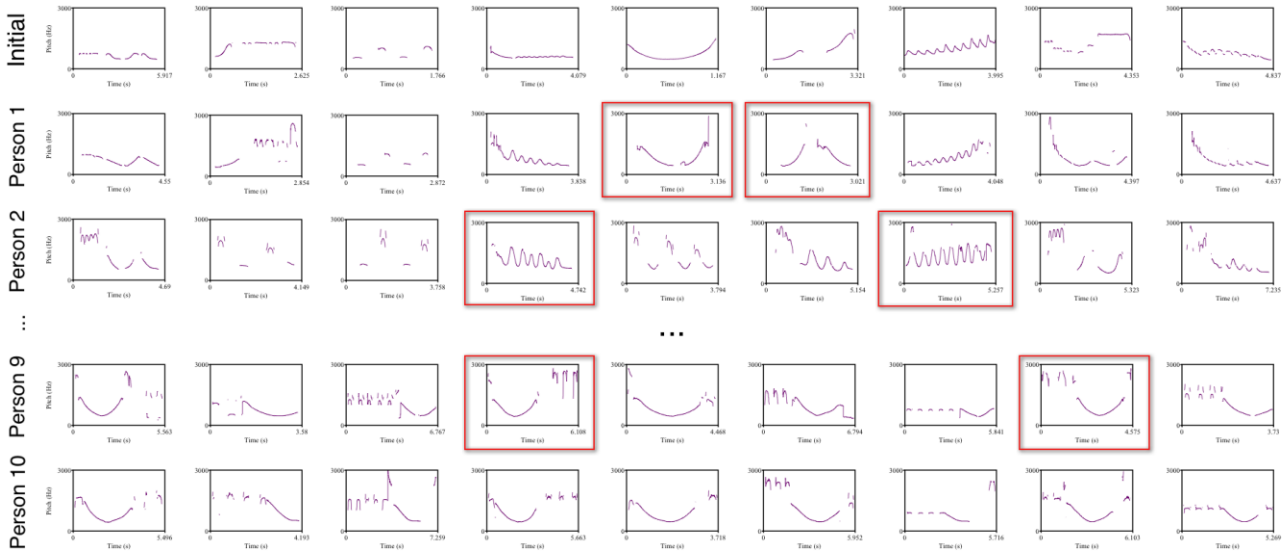



*Figure 9: Part of a chain in the ongoing experiment of iterated learning of a whistled system, showing pitch tracks. The first row shows the initial whistles and the other rows show the recall output whistles of several persons in the chain. The highlighted whistles are discussed in the text.*

scribble interface, we replaced it with the use of a slide whistle (see figure 8). Slide whistles are suitable because participants can easily use them to produce a rich repertoire of acoustic signals in an intuitive way, while only very little interference from pre-existing linguistic knowledge is expected. In this experiment combinatorial structure emerges readily. Participants in this study learn and reproduce twelve different short whistle sounds in a procedure that consists of four rounds of training and recall. During training they listen to the whistles one by one, and imitate them with the slide whistle and during recall they need to reproduce all twelve whistles. The output of the final recall phase is used as input for the next person in the chain. Testing is still in progress, but the results thus far indicate that through a process of incidental mirroring and borrowing of existing pieces in the recall phase combinatorial structure emerges. Figure 9 shows a part of one of the finished chains with the whistles represented as pitch tracks. Mirroring can for instance be observed here in whistle five and six of person one, four and seven of person two and whistle four and eight of person nine. It is clear that elements from whistle three and five from the initial set are often borrowed as a building block into new whistles. The evolved set of twelve whistles in the last generation (person 10) is clearly structured in a discrete and combinatorial way. There is a set of basic building blocks: single notes, slides that go down and then up and slides that go down. These elements are re-used and combined in a systematic way into twelve unique whistles and there appear to be constraints on the way they can be combined. Single notes for instance always follow each other on the same pitch and slides always go down first, never the other way around. Similar observations can be found in the other chains, but the specific elements, rules and constraints differ from one chain to the other. In short, these preliminary results show that experimental iterated learning of an artificial sound system can cause this system to become organized in a way similar to how speech sounds and signs in sign languages are organized.

## Acknowledgments

We thank Jelle Zuidema and Wendy Sandler for helpful discussions and suggestions. 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

Christiansen, M. H. and Chater, N. (2008). Language as shaped by the brain. *Behavioral and Brain Sciences*, 31(5): 489–509.

Cornish, H. (2006). Iterated learning with human subjects: an empirical framework for the emergence and cultural transmission of language. Master's thesis, University of Edinburgh.

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

de Boer, B. and 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.), *The evolution of language: Proceedings of the 8th international conference* (pp. 399–400). World Scientific Press.

Dowman, M., Xu, J., and Griffiths, T. L. (2008). A human model of color term evolution. In Smith, A. D. M., S. K. and i Cancho, R. A., (Eds.), *The Evolution of Language: Proceedings of the 7th International Conference*, pp. 421–422. World Scientific Press.

Fitch, W. (2010). *The evolution of language*. Cambridge University Press.

Griffiths, T., Kalish, M., and Lewandowsky, S. (2008). Theoretical and empirical evidence for the impact of inductive biases on cultural evolution. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 363(1509): 3503.

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

Israel, A. and Sandler, W. (2009). Phonological category resolution: A study of handshapes in younger and older sign languages. In Castro Caldas, A. and Mineiro, A., (Eds.), *Cadernos de Sade, Vol 2, Special Issue Linguas Gestuais*, pp. 13–28. UCP: Lisbon.

Kirby, S., Cornish, H., and Smith, K. (2008). Cumulative cultural evolution in the laboratory: An experimental approach to the origins of structure in human language. *Proceedings of the National Academy of Sciences*, 105(31): 10681.

Kirby, S. and Hurford, J. (2002). The emergence of linguistic structure: an overview of the iterated learning model. In Cangelosi, A. and Parisi, D., (Eds.), *Simulating the evolution of language*, pp. 121–148. Springer Verlag New York.

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

Oudeyer, P. (2005b). The self-organization of speech sounds. Journal of Theoretical Biology, 233(3): 435–449.

Sakoe, H. and Chiba, S. (2003). Dynamic programming algorithm optimization for spoken word recognition. Acoustics, *IEEE Transactions on Speech and Signal Processing*, 26(1): 43–49.

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

Smith, K., Kirby, S., and Brighton, H. (2003). Iterated learning: A framework for the emergence of language. *Artificial Life*, 9(4): 371–386.

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