Efficient coding in speech sounds: Cultural evolution and the emergence of structure in artificial languages
Verhoef, T.

Citation for published version (APA):
Verhoef, T. (2013). Efficient coding in speech sounds: Cultural evolution and the emergence of structure in artificial languages

General rights
It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations
If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: http://uba.uva.nl/en/contact, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.
The scribble to sound experiment as described in the previous chapter did not entirely result in the findings that were expected. Especially with respect to questions on the origins of combinatorial structure, it did not lead to the expected insights. Many issues that arose during the scribble to sound study were used as the basis for new ideas that were implemented in a follow-up study, presented in this chapter. The most crucial changes that were made involve the lack of referential meanings in the new experiment and an entirely different way of sound production, replacing the scribble interface.

Given the difficulty participants had in learning to use the scribble interface, it was necessary to replace it with the use of a more intuitive sound production interface. Natural speech was still ruled out, because the natural vocalisations of human participants would already have discrete and combinatorial structure. As a solution to this problem, slide whistles were used in the experiment described in this chapter (see figure 4.1). 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. Asking participants to whistle with their mouth seems less practical, since not everyone is able to do this and even for those who are able to whistle, doing it for an hour straight in an experimental setting most likely is not comfortable and would perhaps result in cheek muscle soreness.

![Figure 4.1: Plastic slide whistle from the brand Grover-Trophy](image)

This chapter contains parts that also appear in: Verhoef, T., Kirby, S. & de Boer, B.G. (under review). Emergence of combinatorial structure and economy through iterated learning. *Journal of Phonetics*
4. Whistles

4.1 Experimental iterated learning with whistles

The experiment described in this chapter shows that it is possible to apply the experimental iterated learning paradigm to acoustic, continuous signals and that this can provide new insights into how combinatorial structure emerged. Concerning the different views on such emergence that have been reviewed in chapter 2, the results will be compared with predictions expected from either dispersion models or theories based on principles of economy. We will see that in this laboratory experiment, the emergence of combinatorial structure is not necessarily driven by pressures for distinctiveness in a growing vocabulary, as Hockett (1960) and others proposed, and that a simple dispersal model alone cannot account for the results. Instead, the results show that combinatorial structure can emerge as an adaptation to cultural transmission and this happens in a way that seems to conform to economy principles (Clements, 2003; Ohala, 1980).

4.2 Methods

The experiment involves the task of learning and reproducing an artificial whistled language, again with the crucial manipulation that each person is exposed to the language the previous participant produced (Kirby et al., 2008). This allows us to study the whistled languages closely while they are being passed on from person to person, simulating cultural transmission.

The languages in this experiment consist of continuous acoustic signals that are produced with a slide whistle (plastic version by Grover-Trophy, see figure 4.1). To reduce interference of pre-existing experience with speaking, slide whistles were used for sound production. The slide whistle has a plunger that can be used to adjust the pitch of the whistle sounds within a range of between about 450 Hz and 2500 Hz. Note that this range is different from the fundamental frequency range of human speech, which roughly ranges from about 85 Hz for a low male speaking voice to about 400 Hz for infants cries (Baken and Orlikoff, 2000).

The artificial languages contain some radical, but necessary, abstractions from natural human languages. In real languages words have meanings, while in this experiment the whistle sounds do not refer to anything. This allows us to control for influences of for instance compositionality, iconicity or vocabulary size, while closely investigating the emergence of phonological structure as a set of meaningless building blocks that are combined into larger signals. The level of structure that is the focus of this experiment is what Fitch (2010) calls “bare phonology” and this structural characteristic is also found in for instance music (Fitch, 2006, 2010). The process studied here may therefore be relevant for the evolution of music as well.
4.2. Methods

4.2.1 Procedure
During the experiment participants were asked to memorise and reproduce a set of twelve different whistle sounds. They completed four rounds of learning and recall. In the learning phase they were exposed to all twelve 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 in their own preferred order from memory. The input stimuli of one participant consisted of the output that the previous participant produced in the last recall round (or the initial input set). Transmission was continued in this manner until there were ten participants in each chain and 4 parallel chains were completed. The experiment took place in a sound proof recording studio and it lasted about 60 minutes in total per participant. After entering the studio, the participants were first informed about what was expected from them during the experiment both in written and spoken form. The written instructions can be found in appendix B.1. Then they were given the opportunity to ask questions and were asked to give informed consent and to fill out a background questionnaire. Most participants had never used a slide whistle before, so they were allowed to familiarise themselves with the instrument before starting the experiment. When the last recall phase was finished, the participants were asked to fill out a post-participation questionnaire in which they could inform us about the strategies they had used for learning and recall and to give feedback on how they felt about their performance. Appendix B.2 shows a screenshot of the user interface that was created for this study.

4.2.2 Initial input set
To construct the initial whistle language that was used as training input for the first participant in each chain, a whistle database was used. This database consisted of whistle sounds that were created by people who participated in an early exploratory pilot study and were asked to freely record a number of whistle sounds. The set was constructed so as not to exhibit any combinatorial structure. It was a collection of sounds that exhibited many different ‘techniques’ of whistling (such as staccato, glissando, siren-like, smooth or broken) with as little as possible reuse of basic elements. Figure 4.2 shows the complete set of twelve whistles from this set plotted as pitch tracks on a semitone scale using Praat (Boersma, 2001).

4.2.3 Reproduction constraint
During the recall phase of the experiment there is one constraint on the whistle reproductions. Participants have to produce twelve unique whistles and are not allowed to record the same signal (defined more precisely below) twice within a recall phase. Previous work by Kirby et al. (2008) on iterated learning in the laboratory has shown that without
4. Whistles

Figure 4.2: Whistles from the initial whistle set, plotted as pitch tracks on a semitone scale. Note the diversity and complex structure of the whistles.

preventive measures against homonymy, the transmitted language is likely to collapse and end up with only a few words covering lots of meanings. A simple filtering approach, which made sure that the next participant was never exposed to a language with homonymy, solved this issue (Kirby et al., 2008). Because it is likely that participants in the experiment forget which whistles they already recorded and because there is no natural communicative pressure to preserve expressivity, a constraint had to be introduced here as well. During recall, the experimental software automatically compared each newly recorded whistle sound with the other whistles that had already been recorded in the same recall round. If the whistle sound was too similar to one of these previously recorded ones, it was rejected and the participant was asked to record another one. Similarity between whistle sounds was determined using a whistle distance measure defined as follows:

$$0.5D_p + 0.2D_i + 0.2D_s + 0.05D_{sd} + 0.05D_{pv}$$

where $D_p$ is the Dynamic Time Warping (DTW) (Sakoe and Chiba, 1978) distance between the two pitch tracks with pitch in Hz and 500 samples per second, $D_i$ is the DTW distance between the two intensity tracks, as computed using Praat (Boersma, 2001), $D_s$ is the difference in the number of segments (where segments are defined as sounding parts separated by silent pauses), $D_{sd}$ is the difference in variation of segment duration, where the variation is measured as the difference between the duration of the longest and shortest segments in the signal, and $D_{pv}$ is the difference in variation of pitch. Data collected in a pilot study was used to create this measure and to determine the coefficients. Participants in this pilot were all asked to imitate the same set of 10 whistles and the dataset created from
these responses was used to find the set of coefficients that resulted in the highest whistle recognition score. The distance below which two whistles were considered the same was set at a relatively low value of 0.06. In this way, participants could still produce relatively similar whistles. A low value was chosen because it was not supposed to influence the outcome of the recall phase in any way other than to reject repetitions of the same signal.

4.2.4 Participants

Forty participants took part in the experiment. They were divided over four parallel chains, each containing ten generations of learning and recall. All participants were university students from either the University of California San Diego, or the University of Amsterdam, ranging in age from 18 to 32 (with a mean of 22). Twenty-six were female. Each chain contained either three or four male participants. They were paid 10 euros or 10 dollars in cash to compensate for their time.

4.2.5 Expectations

Based on the results of Kirby et al. (2008) on the emergence of compositional structure and the results of del Giudice et al. (2010) on combinatorial structure in systems of graphical signals, the expectation is that cultural transmission also causes the emergence of combinatorial structure in the systems of whistled signals and leads to increased recall performance towards the end of the chains. Constraints on memory and learning biases are expected to cause the transmitted systems to become more structured and when there is more structure, participants learn faster and perform better. The whistled systems are therefore expected to change to become optimised for learnability.

4.3 Qualitative results

In this section we first take a close look at specific examples from individual chains and analyse the results qualitatively, in order to get an idea of what the participants seem to be doing. In appendix B.3 the complete transmission chains that resulted from this experiment are shown.

Participants are asked to learn and reproduce the whistle sounds they are exposed to and they try their best to do this as well as they can, but the task is very difficult. Because of this, people make mistakes and they do not recall all whistles flawlessly. In their reproductions they tend to over-generalise some of the structure that they try to discover in the set. 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 consists of subsets of related elements, essentially exhibiting combinatorial reuse, which appears to be more easily remembered and results in increased recall on the whole set.

Figures 4.3 and 4.4 show specific examples of recall behaviours that eventually lead to a gradual increase of structure. The whistles are plotted as pitch tracks on a semitone scale using Praat (Boersma, 2001). Figure 4.3 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 could be considered to resemble co-articulation. In co-articulation a speech sound (or manual articulation in sign language) is influenced by a surrounding articulation. The effects of one articulation can for instance carry over to the next, which then becomes more like its predecessor. A syllable that ends with a rounded consonant may for instance cause the following syllable to also be produced in a more rounded way. The example with the two whistle sounds shows how the final pitch of the first part of the whistle influences the initial pitch of the second part. When the first part contains a falling pitch movement and ends low, the following segment
(with a repeated falling-rising pattern) starts low, but when the first part ends higher, the following falling-rising pattern starts high. Figure 4.4 shows how a combination of mirroring, repetition and borrowing results in a predictable system that is stable and persists after its innovation. In the productions of generation four there is no whistle yet that resembles the one with two falling slides shown here, but in generation five a mirrored version of this whistle appears. Then in generation six the falling one is borrowed and combined with a new final element into a new whistle. In generation seven, this final element may have been reanalysed as having meant to be a repetition of the falling slide element present in the original two, because suddenly a version with three falling slides appears. In the same generation, a mirrored version with three rising slides also appears thus filling a gap and making the system regular.

To take a closer look at the cumulative effect of the participant’s recall behaviours on the transmitted system of signals, figure 4.5 shows a fragment of the set of whistles produced by the tenth and last participant in a chain. In this set we can identify a clear combinatorial structure. There is a set of building blocks (falling-rising slides and short level notes) and these are reused and combined in different but systematic ways to create the whistles in the set. The whistles for instance differ from each other in the number of short level tones they start and end with and for each there is often a version mirrored in order as well. In addition, the set has become more constrained, for instance in the number of falling-rising movements per segment. In the initial set (see figure 4.2) there were whistles with several falling-rising movements in one segment, but this has reduced to a maximum of two movements in the last generation of this chain. Another constraint is the fact that in this set all segments with slides start with a falling tone and there is no longer any version that is mirrored in pitch. Note that this is specific for this particular chain; in other chains rising-initial patterns did occur. 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 realised in different ways. In fact, it appears that each chain results in a set of signals that has recognisable structure in a way that it should be possible to determine whether any given whistle belongs to a set or not.

To summarize the qualitative analysis: 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 and this strategy was indeed reported by participants in a post-test questionnaire.
Figure 4.4: An example of cumulative mirroring, repetition and borrowing. 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 generalised to fit the pattern of the original two, but repeated. This predictable system stays stable until the end of the chain. The whistles are plotted as pitch tracks on a semitone scale.
4.4 Quantitative results

To find quantitative confirmations for the observations that were made in the qualitative analysis, several measures were used to find out how structure and learnability develop in the transmitted whistled systems. Details on the implementation of the analysis and the signal preprocessing steps can be found in appendix B.4.

4.4.1 Recall error

To determine the recall error at each generation in the chains, the distance (defined more precisely below) between the input set and the output set for each participant in each chain was measured. The expectation is that the recall error is lower for participants that came later in the chains, assuming that the observed increase in the reuse of basic elements makes the set more learnable. The recall error is measured as the sum of distances between each whistle in the output and its corresponding whistle from the input. Each whistle from one set is paired with a unique whistle from the other set and this is repeated in all possible ways to find the pairing for which the sum of distances is minimal. To compute the distance between a pair of whistles, a whistle distance measure was used that is different from the one that has been described in section 4.2.3. After the data was collected and the results were analysed qualitatively, participant behaviour was found to be predicted better by the movements of the plunger than by the acoustic signals (on which the first distance measure was based). People seemed to remember and classify the sounds according to the plunger ‘gestures’ they made to produce them. A movement (representing a building block) would be performed with the same displacement when the plunger was at the bottom of the whistle (with low pitch) as when the plunger was at
the top (with high pitch). But in terms of pitch differences, this same motion results in a much bigger difference when it is produced at higher pitch than at lower pitch, because of the non-linear relation between the pitch change and plunger movement of the whistle. This means that if acoustical features are used, distances between building blocks tend to be overestimated in the high pitch range, while they are underestimated in the low pitch range, even when the semitone scale is used.

For the new ‘articulatory’ measure the pitch tracks are first transformed into sequences of plunger positions (from approximately 3 cm to 20 cm) following equation 4.1, where \( l \) is the length in cm between the mouthpiece and sliding stopper, \( c \) is the speed of sound at body temperature (35000 \( cm/s \)) and \( f \) is the measured frequency in Hz. These new tracks approximately represent the actual movements the participants made, and the distance between two whistles is the Derivative Dynamic Time Warping (Keogh and Pazzani, 2001) distance between two movement tracks. This measure therefore focuses on the similarities of whistle shapes and ignores absolute pitch.

\[
l = \frac{c}{4f}
\]  

Figure 4.6 shows the development of the recall error over the four chains, with increasing generations on the horizontal axis. A significant cumulative decrease in recall error was measured using Page’s (1963) trend test (\( L = 1317, m = 4, n = 10, p < 0.05 \)), implying an increase of learnability and reproducibility of the whistle sets over generations.

### 4.4.2 Structure

To define a measure of structure for the emerging whistle sets, an attempt was made to find a way to show that the sets in later generations were composed of a smaller set of basic building blocks that were increasingly reused and combined. This means that the sets would have become more compressible. This type of compressibility can be measured with the information-theoretic measure of entropy (Shannon, 1948). To compute entropy for a set of whistles, the whistles were divided into segments. The silences within a whistle were used as segment boundaries. Then, using all segments that occur in the set of twelve whistles, (average-linkage) agglomerative hierarchical clustering (Duda et al., 2001) was used to group together those segments that were so similar (according to the measure described in section 4.4.1) that they could be considered the same category or building block. Clustering continued until there was no pair of segments left with a distance smaller than 0.08. Equation 4.2 from Shannon (1948) was used to compute entropy, where \( p_i \) is the probability of occurrence of building block \( i \).

\[
H = - \sum p_i \log p_i
\]
Figure 4.6: Recall error on the whistle sets over generations for all four chains, demonstrating that the whistle systems evolve through cultural transmission and become more learnable.

Figure 4.7 shows the development of entropy for the four chains, with the generations again on the horizontal axis and 0 referring to the initial set. A significant decrease in entropy was measured using Page’s (1963) trend test ($L = 1427$, $m = 4$, $n = 10$, $p < 0.001$), excluding the artificially inserted initial set (because this set is not an output produced by a participant, but was constructed by the experimenter). This result implies an increase of structure and predictability as well as more efficient coding.

The measure of entropy described above captures the increase of reuse of basic building blocks and as such it is a good first measure of structure. However, to investigate the combinatorial rules and structure more closely, associative chunk strength (Knowlton and Squire, 1994) was measured in addition. This measure originates from the field of artificial grammar learning and has been adopted before for analysing experimental iterated learning results (Cornish et al., 2010). The associative chunk strength takes the order of appearance of the different building blocks into account and this measure would allow us to find out whether ‘phonotactic’ or sequence constraints can be detected. The structure that was described for instance in section 4.3 for the last generation of chain one, where short level notes surround falling-rising
slides in a systematic way, should result in a higher chunk strength. This measure was computed by using the building blocks that were found as described above for measuring entropy. All bigrams and trigrams (sequences of two or three building blocks) that occurred in the whistles were identified and their frequencies in the whistle sets were counted. The associative chunk strength of a whistle set is the average of the bigram and trigram frequencies.

Figure 4.8 shows the associative chunk strength for the four chains, with the generations again on the horizontal axis and 0 referring to the initial set. A significant increase was measured using Page’s (1963) trend test ($L = 1322, m = 4, n = 10, p < 0.05$), excluding the artificially inserted initial set. This implies that there is a trend towards sequential structure, although as can be observed in figure 4.8, this trend is clear in chain one and chain four, but seems to be absent in the other two chains.

### 4.4.3 Dispersion

As mentioned in chapter 2, it has been suggested that the emergence of combinatorial structure is driven by optimisation for articulatory ease and signal distinctiveness in line with dispersion theories (e.g. de Boer,
4.4. Quantitative results

Figure 4.8: Associative chunk strength of the whistle sets over generations for all four chains, showing an increase in reoccurrence of bigram and trigram sequences of basic whistle patterns.

2000; de Boer and Zuidema, 2010; Liljencrants and Lindblom, 1972). It is therefore interesting to measure whether the whistled signals in this experiment become more dispersed towards the end of the chains. In order to do this the measure of energy \( E \) was adopted from Liljencrants and Lindblom (1972). They used this measure to quantify the acoustic dispersion of vowels systems. The dispersion of whistles in the emerged languages was computed following equation 4.3 which is the same as Liljencrants and Lindblom’s equation (2). Here \( r_{ij} \) is the distance between whistles \( i \) and \( j \). The distance is calculated with the distance measure described in section 4.4.1. A lower value of energy means more dispersion in the whistle sets.

\[
E = \sum_{i=1}^{n-1} \sum_{j=0}^{i-1} \frac{1}{r_{ij}^3}
\]  

(4.3)

Figure 4.9 shows how the energy between whistles in the sets develops over generations. At a glance we already see that the energy level does not appear to decrease. Page’s trend test also reveals that there is no
significant decrease of energy \((L = 1138, m = 4, n = 10, p > 0.05)\), excluding the artificially inserted initial set. We can therefore conclude that the whistles in the sets do not become more dispersed over generations. Actually, in one of the four chains there appears to be a rather sharp increase of energy towards the end of the chain, as can be seen in figure 4.9.

Based on the idea that economy and maximal reuse of basic elements also play a role (Clements, 2003; Ohala, 1980), it is no surprise that dispersion did not increase. Given the qualitative analysis as described in section 4.3 and the outcome of measuring structure as described in this section, we would actually expect an increase in similarities between whistles. With the increasing rate of reuse of basic elements, one may expect that for most whistles in the set there is another one that is similar for some features. This can also be quantified, by measuring the average Nearest Neighbour distance for the whistles within a set. For all chains over all ten generations, the whistles within a single set were compared. For each of the twelve whistles in the set of a generation, the distance to their nearest neighbour was computed. The average of these values was
4.4. Quantitative results

Figure 4.10: Average nearest neighbour distance between the twelve whistles of the set in each generation. The whistles tend to become more similar towards the end of the chains.

then used to test whether whistles have a close neighbour in the set, with which they may share elements. Note that the energy measure defines a more global measure of dispersion and takes distances between all signals in a set into account, while the nearest neighbour distance only measures the distance to one nearest neighbour to see for each signal if there is another one in the set with similar features.

Figure 4.10 shows these average distance values for each chain with increasing generations on the horizontal axis (including the initial set at generation 0). It is clear that the whistles indeed increasingly have close neighbours in the set over generations. The whistles become gradually more similar to each other and this decrease in average nearest neighbour distance is significant according to Page’s trend test ($L = 1322, m = 4, n = 10, p < 0.05$), excluding the artificially inserted initial set. Although in general lower average distance is not necessarily the result of higher reuse, the combination with the qualitative results and other measures makes it likely that in this case it is related to the increased reuse and sharing of features.

The signals within the whistled languages thus seem to become closer to each other, but this does not immediately imply that dispersion plays
no role. When we closely inspect the building blocks that construct these signals, we are able to observe effects of dispersion. One can compare, for instance, the short level note with the falling-rising slide pattern, two building blocks that emerged in the set shown in figure 4.5. The first is very short and involves no plunger movement at all while the second is long and involves a plunger movement over a large part of the pitch range. Observations of this kind can also be quantified, by measuring the energy between building blocks within a set. For each generation in each chain, the building blocks that were found by the clustering procedure as described in section 6.2 were used. The energy between the building blocks within a set was measured with equation 4.3, where $r_{ij}$ is the distance between building blocks $i$ and $j$. The distance between building blocks is calculated in the same way as the distance between whistles, with the distance measure described in section 4.4.1.

Figure 4.11 shows the energy values measured between building blocks for each chain with increasing generations on the horizontal axis (including the initial set at generation 0). The building blocks seem to become significantly more dispersed towards the end of the chains according to Page’s trend test ($L = 1351$, $m = 4$, $n = 10$, $p < 0.01$), excluding the arti-
4.5 Discussion

The experiment presented in this chapter demonstrates that it is possible to study questions of evolutionary phonology in the laboratory using the method of experimental iterated learning (Kirby et al., 2008). The results suggest that cultural evolution can cause a system of whistled signals to become organised in such a way that it is reminiscent of how speech is organised: a small number of (dispersed) building blocks is combined into a larger number of utterances, while the elements and the ways in which they can be combined differ from one chain to the other, resulting in distinct ‘traditions’. The qualitative analysis showed different strategies that caused combinatorial structure to increase in the transmission chains. Towards the end of the chains, a clear discrete set of basic building blocks could be identified and these blocks were systematically reused and combined. A quantitative analysis revealed that the learnability and reproducibility of the whistled signals increased cumulatively over generations. This is in line with earlier findings within the iterated learning paradigm (Kirby et al., 2008; Kirby and Hurford, 2002; Kirby et al., 2004). In addition, the increase of combinatorial structure could be measured quantitatively and the results suggest that the whistled languages become more compressible and predictable with increasing repetitions of learning and recall.

According to Hockett (1960), the emergence of combinatorial structure could be explained by a gradual growth of the vocabulary. When the number of meanings that are expressed increases, the signals referring to those meanings have to be closer, filling up the signal space. Hockett suggests that the signal space is first maximally exploited holistically until the signals cannot be reliably discriminated anymore. Combinatorial structure then allows for an expansion of expressivity, while discriminability is maintained. The experiment discussed in this chapter shows a different route to combinatorial structure. The whistled languages have only twelve signals and the vocabulary does not grow during the experiment. Even with this tiny vocabulary, combinatorial structure emerges while the signal space is not used maximally. The use of the signal space actually reduces over generations. In the experiment presented in this chapter combinatorial structure therefore does not seem to follow from an interaction between vocabulary size and signal dispersion, but rather from the fact that a vocabulary of a certain size needs to be learned within a very limited time frame. Cognitive biases and pressures favouring a more learnable system seem to be driving the emergence of structure in this case. A system of signals that does not
4. Whistles

use combinatorial structure can be hard to learn, because it is entirely unpredictable: everything that can be produced can potentially be part of the system. In contrast, a discrete and combinatorial system limits possibilities, where only a few elements can be used and combined in restricted ways, and is therefore much more predictable. The signals that fit the structure are more likely to be learned and preserved over generations.

To further interpret the results, we return to the principles of dispersion and economy. As mentioned in chapter 2, theories about the emergence of structure in phonology and phonetics can roughly be divided into two groups. The first group focuses on the importance of optimisation for signal distinctiveness (e.g. de Boer, 2000; de Boer and Zuidema, 2010; Liljencrants and Lindblom, 1972; Oudeyer, 2006) and the second focuses on drives that optimise (feature or gesture) economy (e.g. Clements, 2003; Maddieson, 1995; Ohala, 1980). A theory favouring dispersion would predict that the signals in the whistled languages would become less similar and more dispersed in the signal space. A theory favouring economy would predict that a small set of distinct elements would come to be reused and combined maximally. At first sight, the results seem to favour economy, but the results are not entirely in contradiction with maximisation of distinctiveness either, as was demonstrated by measuring dispersion of the basic building blocks. However, the formation of building blocks and their role in the final signals does not resemble the simplest models favouring dispersion, but is more reminiscent of the models favouring economy. If a building block is present, it tends to get reused (possibly in mirrored form) before new ones appear.

The reuse of building blocks as observed in the experiment is not quite the same as the reuse of features in the theories of feature economy. Distinctive features in speech are related to, for instance, places or manners of articulation and these are realised simultaneously in speech sounds, while the objects of combination in the presented quantitative analysis are sound elements that are combined sequentially. A potential way of comparing the whistle structure with features could have been to define whistle features such as ‘pitch direction’, ‘amount of falling rising pitch movements’, ‘whistle duration’, ‘staccato or glissando style’ for example, but this seems too much like imposing feature theory on whistles. The way in which the whistles in the study described here are built up of building blocks is more comparable to the way morphemes are constructed from phonemes or syllables. Therefore, economy is not measured here at the same level as it is described in the theories of feature economy. This difference should not make the comparison less interesting however as it is useful for studying the general tendency towards efficient, combinatorial structure. It has been suggested previously that economy in phonology may be functioning at a general cognitive level: “Feature economy reflects a general predisposition to organize linguistic data into a small number of categories and to
generalise these categories maximally” (Clements, 2003). In addition, the role of compression in languages at other levels has been discussed at length (e.g. Ackerman et al., 2009; Brighton, 2002; Clark, 1994; Teal and Taylor, 2000). This experiment provides a demonstration of how such efficient coding, independent of the level of organisation, may emerge. More details are being studied in follow-up experiments. In one of these experiments the whistled signals cannot contain silences and the possibilities for combining elements sequentially is therefore more limited. A preliminary analysis of the data in this more limited signal space shows emerging systems with patterns that are reminiscent of categories found in tonal languages.

Most experimental iterated learning studies so far were based on discrete, symbolic signals (e.g. Kirby et al., 2008; Reali and Griffiths, 2009; Smith and Wonnacott, 2010). In contrast, the experiment presented in this chapter used continuous signals without pre-defined basic elements. Therefore, some challenges had to be faced. In previous experiments where for instance the signals were strings of existing characters (Kirby et al., 2008; Smith and Wonnacott, 2010), the cognitively salient building blocks corresponded more or less directly to the discrete symbols out of which the stimuli were constructed: letters or syllables. Therefore, in the analysis of these experiments, there was not much explicit thought given to how to find building blocks on which to base the structural analysis. In continuous signal spaces it turns out to be a much more difficult problem to identify what the basic elements are out of which the signal is constructed, what the boundaries between elements are, what within-category-variation is and what between-category-variation consists of. The decision to consider silences as boundaries between potential building blocks was based on the qualitative observation that participants reused and combined the pieces of sound surrounded by silences. Other ways of analysing the signals may have been possible and may have lead to slightly different results, since it may be the case that for some participants the building blocks were actually different from the ones that were analysed. Other ways of segmenting the signals could be for instance to consider local pitch maxima and minima or the pitch inflection points as segment boundaries.

The way in which the building blocks change over the experimental generations is, like in natural languages, (whistle-)language specific. This may explain why for some chains the measured increase in structure is clearer than for others. The difficulty of deciding how to segment the signals into basic elements is not unlike similar problems in natural language analysis, such as deciding whether pitch movements or pitch targets are the primary cognitive elements of intonational structure (see Arvaniti et al., 1998). It is probably true that even speakers of a language do not always use exactly the same analyses of what the building blocks are. It would be difficult to explain language change if this was not the case.
To be able to simulate language evolution in the lab, necessary abstractions from reality had to be made. One of these involved the lack of meaning conveyed by the whistled signals. However, note that the system is not entirely meaningless, because 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. Having to retrieve the whistles from memory also encourages participants to ‘label’ the whistles as for instance: ‘the one with many up and down movements’ or ‘the very first whistle I learned’. Moreover, once the whistles evolve towards sharing features, people tend to categorise them as subsets, such as ‘the ones that all start with one slide down’ or ‘the ones that only have slides up’. This adds meaning implicitly and makes learning and recall of the whole set of whistles easier because chunking of information in this way facilitates encoding more information in short-term memory Miller (1956). Given the results presented here that show how combinatorial structure can emerge independently from complex semantics, an interesting next step would be an experiment that includes meanings. Such an experiment is described in chapter 6.