Efficient coding in speech sounds

Cultural evolution and the emergence of structure in artificial languages

Tessa Verhoef

Efficient coding in speech sounds

—

Cultural evolution and the emergence of structure in artificial languages

Tessa Verhoef

ISBN: 978–90–8891–675–5
NUR: 616
Printed by: Proefschriftmaken.nl || Uitgeverij BOXPress

Layout: Typeset in LaTeX using a template created by Rolf de By
Cover design: Tessa Verhoef

EFFICIENT CODING IN SPEECH SOUNDS
—
CULTURAL EVOLUTION AND THE EMERGENCE OF STRUCTURE IN ARTIFICIAL LANGUAGES

A C A D E M I S C H P R O E F S C H R I F T

ter verkrijging van de graad van doctor
aan de Universiteit van Amsterdam
op gezag van de Rector Magnificus
prof. dr. D.C. van den Boom
ten overstaan van een door het college voor
promoties ingestelde commissie, in het openbaar
te verdedigen in de Agnietenkapel
op vrijdag 27 september 2013, te 12:00 uur

door

Tessa Verhoef

geboren te Maarssen

PROMOTIECOMMISSIE

Promotores:	prof. dr. S. Kirby
	prof. dr. P. Boersma
Co-promoter:	prof. dr. B. de Boer
Overige leden:	prof. dr. H. J. Honing
	prof. dr. W. Sandler
	dr. M. Tamariz
	dr. W. H. Zuidema

Faculteit der Geesteswetenschappen

Contents

Contents i

Acknowledgements v

1 Introduction 1
- 1.1 Evolution of language . . . 1
- 1.2 Sound organisation . . . 5
- 1.3 Overview . . . 6

2 Background 7
- 2.1 The protolanguage debate . . . 7
- 2.2 Emergence of combinatorial structure . . . 8
- 2.3 Language evolution experiments . . . 12
- 2.4 Compression and the brain . . . 16
- 2.5 The present work . . . 18

3 Scribbles 21
- 3.1 Experimental iterated learning with continuous signals . . . 22
- 3.2 Scribble to sound . . . 22
- 3.3 Methods . . . 23
  - 3.3.1 Participants . . . 23
  - 3.3.2 Stimuli . . . 23
  - 3.3.3 Procedure . . . 25
  - 3.3.4 Modifications . . . 26
  - 3.3.5 Expectations . . . 27
- 3.4 Results . . . 27
  - 3.4.1 Qualitative results . . . 27
  - 3.4.2 Quantitative results . . . 34
- 3.5 Discussion . . . 37

4 Whistles 41
- 4.1 Experimental iterated learning with whistles . . . 42
- 4.2 Methods . . . 42
  - 4.2.1 Procedure . . . 43
  - 4.2.2 Initial input set . . . 43
  - 4.2.3 Reproduction constraint . . . 43

4.2.4 Participants . . . 45
4.2.5 Expectations . . . 45
4.3 Qualitative results . . . 45
4.4 Quantitative results . . . 49
4.4.1 Recall error . . . 49
4.4.2 Structure . . . 50
4.4.3 Dispersion . . . 52
4.5 Discussion . . . 57

5 Games **61**
5.1 Perceptual category learning game . . . 62
5.1.1 Methods . . . 63
5.1.2 Results . . . 65
5.2 Follow-up experiment . . . 66
5.2.1 Methods . . . 66
5.2.2 Results . . . 67
5.3 Discussion . . . 68

6 Meanings **71**
6.1 Combinatorial structure versus iconicity . . . 71
6.2 Methods . . . 74
6.2.1 Procedure . . . 75
6.2.2 Initial input sets . . . 77
6.2.3 Reproduction constraint . . . 77
6.2.4 Participants . . . 79
6.3 Qualitative results . . . 79
6.3.1 Internal structure in whistle sets . . . 80
6.3.2 Segmenting whistles into building blocks . . . 82
6.3.3 Iconic whistle-object mappings . . . 83
6.4 Quantitative results . . . 85
6.4.1 Recall error . . . 86
6.4.2 Combinatorial structure . . . 88
6.4.3 Transparency . . . 90
6.5 Discussion . . . 91

7 Agents **97**
7.1 Sensitive periods . . . 97
7.2 Age sensitivity in language acquisition . . . 100
7.3 Age effects in language emergence, change and growth . . . 102
7.4 Vowel systems in a population of agents . . . 103
7.4.1 Memory . . . 104
7.4.2 Production . . . 104
7.4.3 Perception . . . 104
7.4.4 Interactions . . . 105
7.4.5 Memory update steps . . . 105
7.4.6 Population dynamics . . . 107
7.5 Simulations . . . 108
7.5.1 Experiments . . . 108

7.5.2 Measures . . . 109
7.5.3 Results . . . 110
7.6 Discussion . . . 112

8 Discussion **117**
8.1 Main findings . . . 118
8.2 The protolanguage debate . . . 120
8.3 Cultural transmission and efficient coding . . . 122
8.4 Possible concerns . . . 123
8.5 Plans for the future . . . 125
8.5.1 Experimental designs . . . 126
8.5.2 Neuroscience-inspired computer model . . . 127
8.5.3 Conclusion . . . 127

References **129**

Appendix A: Scribbles **149**
A.1 Instructions . . . 149
A.2 User interface . . . 150
A.3 Random scribble trajectory generation . . . 151

Appendix B: Whistles **153**
B.1 Instructions . . . 153
B.2 User interface . . . 154
B.3 Transmission chains . . . 155
B.4 Analysis details . . . 164
B.4.1 Pre-processing of whistle sound files . . . 164
B.4.2 Jump removal . . . 164
B.4.3 Segmenting whistle sounds . . . 166
B.4.4 Dynamic time warping . . . 167

Appendix C: Meanings **169**
C.1 Instructions . . . 169
C.2 User interface . . . 170
C.3 Transmission chains . . . 171

Summary **189**

Samenvatting **195**

Curriculum Vitae **201**

Acknowledgements

As part of the work I did for this thesis, I spent hours closely observing the evolution of miniature artificial languages in the lab. It almost looked like a miracle how spontaneously structure emerged in these languages. In contrast, the way this thesis took shape involved quite a bit more conscious effort and intentional design... But, voilà, here it is! And just as languages are shaped as the result of biases and influences from many individuals, the appearance of this thesis is neither a miracle, nor the result of my effort alone.

First and foremost, I would like to express my deepest gratitude to my supervisor Bart de Boer. I could not have wished for a more dedicated, supportive and reliable teacher. Your door was always open and I thank you for the innumerable times you gave me helpful advise, insightful ideas and constructive suggestions. You were there, as a robotics teacher in the Artificial Intelligence programme, when I set my first steps into the world of science as a first-year BSc student. After having been your student for almost ten years, it will be hard to get used to doing research without your guidance. Second, I am immensely grateful for the fact that my external supervisor, Simon Kirby, patiently gave me the chance to talk about my plans when we first met and generously gave me advice on how to proceed. Thank you so much Simon, for adopting me as your PhD student even though I was not even at the LEC. Your invaluable input has enormously influenced my work and you never stopped to inspire and motivate me. I have the greatest admiration for your research, enthusiasm and ideas. It was always a great pleasure to see you, wherever we got a chance to meet, in Edinburgh, San Diego and at various conferences or meetings. My third supervisor, Paul Boersma, also deserves my gratitude. He gave me a home in his phonetic sciences group and helped to improve my writing.

I also wish to sincerely thank the members of my reading committee: Wendy Sandler, Henkjan Honing, Jelle Zuidema and Mónica Tamariz. Thank you for making the time to read my thesis and for travelling to Amsterdam for my defense. My paranymphs, Jan-Willem van Leussen and Marieke Schouwstra, deserve my special thanks for their help in the preparations for the defense and support during the ceremony.

The ACLC has been a very pleasant and stimulating environment and I thank all my colleagues for their sociability and collaboration. I particularly thank my roommates, Joke and Mirjam, for the many lovely chats, occasional ice-cream breaks and for making the office a cozy place, and also Elly and Gerdien for their care and support as well as my fellow PhD students with whom I had the pleasure to share many meetings, lunchtimes, borrels, Sinterklaas celebrations and more. In addition I would like to acknowledge the ACLC for the financial support I was given for travelling abroad and to help me realise a science exhibition project in Nemo. Dirk-Jan Vet was always available to provide technical assistance and equipment and I am very grateful especially for his help with the exhibition. Furthermore, I sincerely thank Kees Hengeveld for supporting extracurricular activities like the exhibition and for giving me the chance to present my work to broader audiences.

Also outside the ACLC, I have received support from many different angles. Diana Issidorides and her colleagues at Science Center Nemo gave me the chance to collect data in the museum and provided me with all the assistance I needed. Jelle Zuidema was willing to collaborate with me on the museum project and he and Vanessa Ferdinand kindly allowed me to use and adapt their UFO game code. Rolf de By shared his Latex template with me so that I could use it to lay-out this thesis. Isabelle Boutriaux gave me a few minutes of fame by inviting me to be interviewed for Belgian television. I sincerely thank each of you.

During my PhD project I collaborated with Carol Padden and the researchers in her lab at UC San Diego and I am very grateful to have gotten the chance to become part of this unique, inspiring and motivating group. My visits have played a major role in the shaping of my research and I always had a wonderful time at UCSD. Carol, thank you so much for making me feel at home in San Diego and for providing such a stimulating environment, I cannot wait to be back in the fall. Many friends, colleagues and roommates in San Diego also made me want to come back time after time. Among others, I particularly wish to thank Deniz, Alex, So-One, Ryan, Sharon and Carson. My special thanks also goes to Margaret for her excellent help with the preparations for all my visits and her warm welcome upon each arrival.

In addition, I had the opportunity to visit the LEC in Edinburgh a few times and I am extremely grateful, not only for the meetings with my supervisor Simon, but also for the inspiring chats with Kenny Smith, Mónica Tamariz, Bob Ladd and James Kirby and for the enormously helpful feedback I always received from the audience at LEC talks. Furthermore, I wish to thank LEC PhD friends and colleagues, Hannah, Keelin, Vanessa, Sean, Justin, Andrea and others, for their great company and also for inviting me on their amazing LEC PhD retreats.

In the last few stages of finishing my dissertation I had the pleasure to be working at the AI lab at the Free University of Brussels and I thank my colleagues, Sabine, Hannah, Kerem, Heikki, Katrien, Ellen, Pieter and

Johan for the lovely chats and good times in Brussels.

The vast majority of the results presented in this thesis would never have been found if it wasn't for the $\pm$ 600 participants (and I am probably forgetting a few pilot studies) who were willing to be test subjects for my experiments, whether in Amsterdam, San Diego, in the museum or online. I am grateful to each and every one of them individually.

I had the opportunity to work with several students who were involved in my research when writing their thesis, as student-assistants or as interns. In particular I would like to thank Gisela Govaart, Doriene Diemer, Rob Grayson, Alessandro Lopopolo and Jared Markowitz for their enthusiasm and effort.

With the appearance of this thesis I close an important and so far the most exciting chapter of my life. During these four years many people became or continued to be part of my story and I am immensely grateful for all the support I have received from my dear friends and family. Unfortunately I cannot name everyone individually, but I would like to especially mention Carolina, José, Ana, Victor, René, Rafa, Linda, Rolf, Stephanie, Sandra, Tijs, Deborah, Erik, Madelon, Wilco, Wolf, Marieke, Winand, Annemarie.

My parents, Wout en Lia, I would never be able to thank enough for their unlimited trust in my abilities and their unconditional support in everything I undertake. Thank you both for always being there to help me, you are heroes.

Finally I convey my love and gratitude to Jaldert, who encouraged me throughout the entire process. Thank you for your support and for forgiving me for being absent and occupied so many times. I know I tested your patience and I warmly thank you for the fact that you are always there for me.

Introduction

1

No species on this planet other than mankind uses a system for communication as intricate as human language. How did we get from the chirps, howls and calls of monkeys and apes to the complex and sophisticated signal of human speech? What is the origin of this unique form of communication? This is a question that has fascinated researchers since long ago and the work presented in this thesis belongs to the scientific field in which it is studied. The specific area addressed here is cultural evolution and the emergence of structure in sound systems used for speech. This first chapter sketches the context for this thesis and provides an overview of what can be expected to be found in the following chapters.

1.1 Evolution of language

Language is one of the most important features that separate us humans from the rest of the animal kingdom. Yet, we do not have a clear picture of how it arose and what it is exactly that gives humans the ability to use it. Until relatively recently it was hard to approach questions on language evolution without resorting to speculation because there is not much tangible evidence to be found in this area (Müller, 1861). Speech is a rapidly fading signal and we do not have recordings of human's first utterances. Written language is a relatively recent phenomenon, so the history of writing systems will not help us to study the origins of spoken language. Fossil records may reveal data about the evolution of the human vocal tract and biological adaptations such as the descended larynx and the loss of air sacs can be shown to aid the production of speech (de Boer, 2012; Fitch, 2000), but there are other functions that could have driven the evolution of these adaptations as well (de Boer, 2009; Fitch, 2000). We can therefore not be sure they evolved especially for speech. So for a long time the data that could be used for developing theories about the evolution of language was limited. The results of early surmises received nicknames such as 'the bow-wow theory' for the idea that the first words were imitations of sounds such as animal vocalisations or other sounds from the environment and 'the pooh-pooh theory' for ideas assuming that the first words were the sounds people make when expressing emotions such as fear or joy (Müller, 1861).

It was at first assumed that there had to be a special innate language module unique to humans (Chomsky, 1976; Piattelli-Palmarini, 1989; Pinker and Bloom, 1990). How else could we explain why children acquire their language so easily and reliably, while other species did not seem to have these abilities? It was assumed that language "belongs more to the study of human biology than human culture" (Pinker and Bloom, 1990). Whether such a specialised language module evolved as a biological adaptation through natural selection (Pinker and Bloom, 1990) or by accident (Piattelli-Palmarini, 1989) is a matter of debate between proponents of this view, but they share the idea that humans are born with a special language faculty and that language should be studied as if it is a biological organ like any other in the human body (Chomsky, 1976). Until now, researchers have not been able to identify such an organ or module unique to humans that may account for our linguistic abilities. Studies involving the human brain (Deacon, 2009; Fisher and Marcus, 2006) as well as investigations into molecular genetics (Fisher and Marcus, 2006) suggest that language most likely arose in response to the reorganisation of many different systems that humans share with their ancestors and evidence for the existence of a single special module is therefore lacking.

In 1976, a conference was organised by Stevan Harnad and others (Harnad et al., 1976), in which researchers from many different disciplines were brought together to discuss issues on 'the origins and evolution of language and speech'. It was recognised that the speculative nature of research into this topic could only be overcome by taking a multi-disciplinary approach. This meeting involved sessions on a variety of topics including perception and cognition in humans as well as non-humans to explore the basis of language and intelligence; artificial intelligence to see to what extent machines can copy human (linguistic) abilities; comparative biological research to learn from communicative behaviour in animals; neuroscience to find out how the brain is involved and paleobiology to study what our ancestor's use of symbols and tool making can reveal. This could have been the start of a fruitful collaborative programme but it was not until 1996 before the field really took shape and the first EvoLang conference, an international and interdisciplinary conference on the evolution of language, was organised in Edinburgh. This became a series of biannual meetings with contributions from the different disciplines that were represented at the 1976 meeting, as well as the introduction of other modern and empirical methods. Geneticists for instance now search for unique genes that may explain human linguistic behaviour; computer modellers analyse and simulate evolutionary scenarios and interactions between individuals; linguists head into the field and study newly emerging (sign) languages; cognitive scientists and psychologists conduct experiments in which human participants learn or invent artificial languages and so on. In sum, there is now a wealth of data available and the development of suitable methods for studying language evolution is growing.

As modern data is accumulating, it becomes progressively clear that there are viable alternatives to the theory that assumes an innate language faculty. Computational simulations, laboratory experiments and other methods have yielded results (discussed in more detail in chapter 2) that are in line with the suggestion that language is shaped by the brain (Christiansen and Chater, 2008) and that not only biological evolution but also cultural evolution can explain the emergence of linguistic structure (Deacon, 1997; Kirby and Hurford, 2002; Kirby et al., 2004). As Deacon (1997) wrote in The Symbolic Species, "The structure of a language is under intense selection because in its reproduction from generation to generation, it must pass through a narrow bottleneck: children's minds" (Deacon, 1997, p.110). An idea that has become increasingly popular is that language is a system that culturally evolves in a way that can to some extent be compared to the process of natural selection in biological evolution. As Kirby and Hurford (2002), Kirby (2002), Zuidema (2003) and others demonstrated, transmission of a language from generation to generation can make the language more learnable and more structured. Each time the language is passed on it is filtered by the brains that are learning it. It is impossible for a learner to be exposed to every possible utterance in a language because languages are open-ended systems, so all learners have to form their own hypotheses about the structure of the system. Only those structures that can be inferred will be reproduced and therefore there is selection on learnable structures. The structures that are easily transmitted pass through the bottleneck and remain part of the language. In addition, typological data on many different languages revealed that languages around the world are much more diverse than originally thought, which makes the assumption of highly specialised biological adaptations even more implausible (Evans and Levinson, 2009).

Traditionally, it was assumed that the nature of language could be unravelled by studying individual language users (Pinker and Bloom, 1990) and by identifying the universal structures found in languages around the world as an indication of what is encoded innately. The newer ideas mentioned in the previous paragraph imply that language should be viewed as a complex adaptive dynamical system (Beckner et al., 2009; Brighton and Kirby, 2001; Kirby, 2002; Steels, 1997b). From this point of view, it follows that it would be naïve to study language as a system independent of culture and context. Language is the result of many systems that all influence each other in complex ways. The characteristics of the linguistic utterances produced by the individual is only a very small part of this system. Language is a complex system because it emerges as a result of interactions between multiple individuals. At the population (macro) level, language is more than a sum of all the utterances produced at the individual (micro) level.

As mentioned before, languages are transmitted over generations and are dynamic; they change over time and adapt to the selective pressures created by constraints on learning, interaction and population

structure. The role of the first has been explained in the previous paragraph. The second, interaction between multiple individuals, can cause the emergence of a conventionalised shared system when individuals align their behaviour. This has been demonstrated with the use of computer simulations (e.g. de Boer, 2000; Steels, 1997b; Zuidema and de Boer, 2009) and can also be observed in real languages when communities with no shared language start to share a living environment and a new language results from the interactions between members of the communities (Bakker, 1994). Third, population structure and social factors have been found to be related to language structure and complexity. Lupyan and Dale (2010) used statistical analysis techniques on a large sample of languages and found that factors such as population size and contact with other languages could predict certain characteristics of the language structure. Languages spoken by more people tend to be less complex. Wray and Grace (2007) similarly proposed that the pattern of language use may be of influence on the structure. In small, cohesive populations where everyone knows each other and the language is rarely used to talk with strangers, the content of what is talked about is expected to be predictable, for instance because roughly the same knowledge is shared by all members of the population. In contrast, in larger populations in which a greater proportion of conversations is held with strangers, the content cannot always be so easily predicted on the basis of a shared cultural background and context. Wray and Grace (2007) therefore argue that languages that are more often used for talking with strangers are more likely to develop towards having predictable structures and being transparent. Languages of isolated populations on the other hand are expected to be more opaque. In summary, different sources of data all indicate that influences of social and cultural factors should be taken into account in the study of language evolution.

The research presented in this thesis builds on the interdisciplinary work that views language as a complex adaptive dynamical system. Two of the relatively novel methods that have been developed in the field of language evolution, experiments with human participants and computer simulations, are central to this thesis. Computer simulations provide an excellent tool for investigating evolutionary processes and help shed light on the non-trivial relation between micro-level behaviours of individuals and macro-level structures in linguistic systems. The outcome of the complex interactions between these levels are hard to predict and simulations may lead to surprising new insights. However, assumptions and simplifications need to be made when creating computer models, which means that computer agent speakers do not necessarily resemble real speakers in every aspect, especially in terms of their cognitive power. Therefore, it is important to incorporate real human participants in research about language evolution as well. The method of experimental iterated learning (Kirby et al., 2008) has proven to be very suitable for this and

forms the main inspiration for the experimental work presented in this thesis. These experiments involve an exploratory investigation in which the experimental iterated learning paradigm was extended and developed further for the application to the study of the emergence of a specific property of language: the combinatorial organisation of sounds for speech. This property has received relatively little attention as compared to other aspects of language and has only very recently started to be addressed more widely (de Boer et al., 2012). Computer models and experiments together provide a good basis for testing existing theories and generating new ones.

1.2 Sound organisation

This thesis focuses on one particular characteristic of human language: the organisation of speech sounds. Speech sounds are part of a discrete repertoire of primitives that are organised in combinatorial structures. Where does this kind of structure come from? Compared to other species, humans are generally able to produce a larger range of different sounds and these sounds are organised and combined more elaborately (Hurford, 2011). In addition, humans are able to speak about an enormously rich set of meanings. Animal communication systems show very little semantics and complex, acquired meanings are rare. Some bird species use their song to convey fitness in the competition for mating and territory (Doupe and Kuhl, 1999), bottlenose dolphins refer to individuals within a group and maintain group cohesion by producing distinct signature whistles (Janik and Slater, 1998) and there are monkeys that associate different alarm calls with the threats of different predators (Zuberbühler, 2000), but none of these examples even remotely resemble the rich compositional semantics human language has (Hurford, 2011).

Unlike complex semantics, combinatorial structure is not something that is strictly unique to human language. At the level of (phonological) combinatorial structure, there are clear analogous structures in animal song systems. As Hurford (2011) shows with a detailed analysis of such systems: “Apart from the obvious lack of compositional, and referential, semantics, these songs are not qualitatively, but only quantitatively, different in their basic combinatorial structure.” (Hurford, 2011, p. 24). Examples are the structures found in the songs of birds, whales and non-human primates. Certain species of birds that typically acquire their song when growing up, such as the white-crowned sparrow or the zebra finch, produce songs that can be analysed into hierarchical structures in which basic building blocks (notes) are combined into syllables and syllables are organised into larger motifs (Doupe and Kuhl, 1999). A similar type of predictive and hierarchical pattern is found in the songs of humpback whales (Payne and Mcvay, 1971). Payne and Mcvay (1971) describe how the structure of the songs of these whales spans a much

longer duration than those of birds, but also consists of basic sound 'units' that are combined into larger constructs called themes, phrases, songs and song sessions. These whale songs have been analysed by Suzuki et al. (2006) with a computerised unit classifier and measures based on information theory to provide additional evidence for the presence of hierarchical combinatorial structure. Within the primate lineage, gibbons are known to produce complex songs as well (Clarke et al., 2006). They use a set of basic vocal units to form complex phrases and songs and individuals engage in 'duets' by taking turns in a systematic way. These examples suggest that perhaps very general cognitive structures are involved in processing and dealing with combinatorial structure of this type, and that no language-specific biological adaptations need to be assumed for explaining the emergence of such structure.

The evolution of complex sound systems for speech is investigated here within a framework that recognises the importance of cultural evolution. In this thesis I study how sound systems emerge, develop and are preserved when being transmitted over generations. One of the main aims is to investigate to what extent structures in sound systems for speech can be explained as the result of general cognitive biases and the process of cultural transmission. Several issues are addressed: the influence of cultural transmission on the emergence of phonological structure; the role of referentiality and semantics in such emergence and the way population structure affects the preservation of emerged systems.

1.3 Overview

The next chapter provides a background on a selection of areas in the field of language evolution that are relevant for the main subjects of this thesis. It provides a brief general overview of different views on the nature of human protolanguage, reviews current hypotheses and ideas that have been proposed to explain the emergence of combinatorial structure, summarises different experimental methods that have been used in the field and links these to ideas about efficient coding in the brain. Chapter 3 subsequently describes a first experiment in which the cultural emergence of combinatorial structure is studied. Chapter 4 then describes a more elaborate experimental study in which combinatorial structure emerges through cultural transmission in artificial whistled languages. Chapter 5 describes experiments disguised as online games that were conducted to further analyse the data from chapter 4. In chapter 6 results from a follow-up experiment with artificial whistled languages is described in which semantics is added. Chapter 7 is about a computational model that was used to study the preservation of emerged vowel systems in populations of interacting computer individuals. The thesis ends with a general overall discussion and conclusion.

Background

2

This chapter provides a compact review of select research areas in the field of language evolution relevant for the topic of this thesis. It sketches a framework for understanding the motivation behind the work presented and to help interpret the results. First, the debate on what a possible primitive ancestor of modern human language sounded or looked like is discussed. Whether or not there has been such a *protolanguage* and the details of the route from that stage to modern language is still a matter of debate. Then a section with background on the origins of combinatorial structure in language follows. Combinatorial structure is the main focus of this thesis and the studies presented in the following chapters investigate its emergence. Section 2.3 reviews an important experimental method, iterated learning with human participants, which plays a prominent role in almost all chapters of this thesis and the section thereafter links the findings from such experiments to ideas on efficient coding in the brain. The last section describes how the work in this thesis compares with earlier work that is related.

2.1 The protolanguage debate

Theories about a possible ancestral protolanguage have been the source of a longstanding debate and still form an unresolved issue in the field of language evolution. The ideas that have been proposed about what protolanguage looked or sounded like and how it developed into modern language, can roughly be categorised in two scenarios. One view, referred to as holistic protolanguage or the analytic route from proto- to modern language proposes that initially holistic utterances were segmented into smaller elements (Arbib, 2005; Wray, 1998). Examples of modern theories of this type may differ extensively on the details concerning the protolanguage modality. Arbib (2005) for instance describes a scenario in which *protosign*, a system of holistic manual

This chapter contains parts that also appear in the following articles:

Verhoef, T., Kirby, S. & de Boer, B.G. (under review). Emergence of combinatorial structure and economy through iterated learning. *Journal of Phonetics*

Verhoef, T. (2013) Cultural evolution, compression and the brain. *The Past, Present and Future of Language Evolution Research* (to appear).

utterances, first emerged as a result of combined pantomimic behaviour and conventionalised gestures. A system of vocal communication was assumed to emerge at a later stage. Fitch (2010) describes a modern version of Darwin's (and other's) musical protolanguage theory which he calls prosodic protolanguage. This scenario is more focused on the vocal-auditory modality and describes how phonology emerged first as a system independent of meaning out of a system of *protosong*. Possibly driven by mother-child bonding rituals and kin-selection, it is proposed that holistic meanings came to be attached to prosodic utterances and then these sounds became segmented through a process of regularisation and cultural transmission (Fitch, 2010). The other view is called the synthetic route from proto- to modern language. With this route it is assumed that simple words were combined into more complex structures (Bickerton, 1992; Tallerman, 2007). Bickerton (1992) for instance proposed that protolanguage first consisted only of lexical items that were strung together in an arbitrary order: just words without any syntactic structure. Syntax is assumed to have entered language later, although researchers differ in their belief on whether this happened gradually or abruptly (Schouwstra, 2012).

Many arguments have been proposed in favour of and against the different ideas on the nature of protolanguage. In chapter 8 this is addressed in more detail because it can be argued that the experimental results described in this thesis provide new evidence in this debate. However, the main research questions dealt with in this thesis do not involve hypotheses about protolanguage directly, therefore I refer the reader to Schouwstra's (2012) thesis for a more elaborate recent review of the debate and to a special issue dedicated to protolanguage edited by Arbib and Bickerton (2008). The next section reviews the area that is the main subject of this thesis: the emergence of combinatorial structure.

2.2 Emergence of combinatorial structure

One of the basic ways in which languages are organised is through their combinatorial structure: a small set of meaningless building blocks is combined into an unlimited set of words and at the same time, meaningful elements are combined into utterances and larger constructs (Hockett, 1960). This type of multi-level regularity is what Hockett (1960) called *duality of patterning* and he identified it as one of the basic design features of human language. The same phenomenon has also been termed *double articulation* by Martinet (1984) but as Ladd (2012) pointed out, there are subtle differences between the two definitions. Both however are consistent with the view that this phenomenon may reflect the "application of complex combinatoric principles at different levels in a hierarchical structure" (Ladd, 2012, p.271). In this thesis the focus is on one of the two proposed levels of organisation:

combinatorial structure at the sub-lexical level in speech. This refers to the combination of meaningless sounds into words. Hockett (1960) proposed a possible way in which such combinatorial structure of speech could have emerged. According to him, a growing vocabulary increased the need for combinatorial structure and drove its emergence. The signal space limits the number of holistic signals that can be distinguished. When the number of meaningful elements that need to be expressed increases, signals get closer to their neighbours in that space and discriminability decreases. This problem can be solved by combining a smaller number of elements into a larger repertoire of signals. Hockett's account therefore suggests that structure emerged out of pressures for expressivity and discriminability. Similar ideas have been proposed by drawing a parallel between duality of patterning in language and the structure that is found in chemical systems and genetics. It is argued that the emergence of structure in these domains as well as in language is attributable to more general properties of material nature that are necessary to maintain 'self-diversification' (Abler, 1989; Studdert-Kennedy and Goldstein, 2003).

The idea that optimisation for distinctiveness played a role in the emergence of combinatorial structure has been studied with the use of computer models. Liljencrants and Lindblom (1972) defined a measure to determine the overall discriminability of vowel systems (described in more detail in chapter 4 and also used in chapter 7). Their algorithm searched the space of possible vowel systems while optimising for discriminability and articulatory ease. These optimisations resulted in realistic (small) vowel systems, suggesting that these pressures may play a role in the emergence of a discrete set of vowel categories. This model can, however, not explain what drives this optimisation. To address this issue, agent-based simulations have been used to demonstrate that optimally dispersed discrete signal systems can emerge without explicit optimisation. The optimisation in these models is the result of self-organisation under pressures of good communication and learnability (de Boer, 2000; Oudeyer, 2006). de Boer (2000) modelled a population of interacting individuals. These *agents* (virtual robots) play imitation games and dynamically update their vowel repertoire in response to the success or failure of these interactions. With this model it was shown that optimisation for signal distinctiveness can be the result of self-organising principles arising from the interaction dynamics and realistic vowel systems emerged. This model is discussed in more detail in chapter 7 in which a study is described that uses a re-implementation of this simulation. Oudeyer (2006) also studied the emergence of vowel systems but his model did not involve a pre-defined interaction protocol. The agents did not engage in language games and there were no predefined rules for turn-taking in the speaking and listening behaviour. The brains of the agents in this model developed by dynamically adapting vectors of neurons of perceptual and articulatory networks in response to perceived sounds in the environment. As in

de Boer's model, realistic vowel systems emerged. This work was also extended to learn sequences of vowel targets to explain syllable structures (Oudeyer, 2006), but by pre-defining signals as sequences of sound primitives this work could not explain how or why combinatorial structure would emerge. Building on these earlier results, de Boer and Zuidema (2010) showed that combinatorial structure can also result from self-organisation in a population in which agents interact through imitation games with a pressure to keep signals distinct. Both holistic and combined signals were represented as continuous trajectories in this model and it could therefore be studied what was needed to cause the trajectories to 'stretch out' in the acoustic space, spanning more than one target in the space, which could be analysed as having combinatorial structure. In their model, systems with such combinatorial structure indeed emerged. In addition, Nowak et al. (1999) showed that, in the case that there is noise, there is a logical error limit to the number of signals that can be discriminated without loss of communicative success, which can be overcome by combinatorial structure. These results appear to conform to Hockett's (1960) proposal in which he explains the emergence of combinatorial structure on the basis of pressures from signal distinctiveness and vocabulary expansion.

However, it has been suggested that an explanation focusing on optimisation for distinctiveness alone may not be enough. Liljencrants and Lindblom (1972) already observed that, while smaller vowel system can be predicted quite accurately with their optimisation model, larger vowel systems are less well explained on the basis of dispersion. Looking at consonant inventories, Ohala (1980) suggested that the organisation in speech sounds instead seems to follow a principle of "Maximal use of available distinctive features". This was based on the observation that features used in the inventories are efficiently recombined and maximally reused, which does not always result in more dispersion. If consonant inventories were optimised for distinctiveness, we would assume that the members of one set would use as many different places and manners of articulation as possible. This is however not what is observed in real languages. If for instance a certain place of articulation is used to contrast one pair of phonemes in a system, it tends to be present for other pairs of phonemes with different manners of articulation as well. Berrah and Laboissière (1997) have shown, using a computer model that is similar to the imitation game model described in the previous paragraph (de Boer, 2000), that applying this idea to vowel systems leads to improved prediction of larger systems. Clements (2003), when referring to the theory of feature economy, expressed similar ideas about the importance of re-using features: "languages tend to maximize the combinatory possibilities of features across the inventory of speech sounds: features used once in a system tend to be used again" (Clements, 2003, p. 287). Both Ohala's and Clements' principles focus on the efficient reuse of distinct features to make up a system of sounds. A related proposal was made involving

speech gestures by Maddieson (1995), who described structure in speech in terms of articulatory gestures and efficient reuse of places of articulation.

These theories based on principles of economy may differ in the assumptions about whether the basic elements for reuse are abstract features or physical gestures, but what they have in common is that they all propose a rather different approach compared to the dispersal models mentioned before. A general tendency towards efficient representation of information appears to be assumed. This implies a more direct involvement of language learning and cognitive biases in explaining combinatorial structure. Perhaps an explanation based only on the need for distinctiveness under pressure of semantic complexification is therefore incomplete.

A possible source of evidence that can shed light on the question whether combinatorial structure was the result of pressures for discriminability when vocabularies expanded, is the study of a newly emerging sign language. Established sign languages have phonological structure that uses discreteness and recombination just as spoken languages do (Corina and Sandler, 1993). Al-Sayyid Bedouin Sign Language (ABSL) is a sign language that is only a few generations old and in which the emergence of phonological structure is currently being observed (Israel and Sandler, 2011; Sandler et al., 2011). Even though it is a fully functional and expressive sign language with a large vocabulary and a rich, open-ended meaning space, it appears that its combinatorial structure is less discrete than those of established sign languages (Sandler et al., 2011). This example shows that a growing vocabulary can be maintained without combinatorial structure.

A different source of evidence that weakens the assumption of dependence between combinatorial structure and complex semantics is the study of song systems of for instance birds and whales (Doupe and Kuhl, 1999; Payne and Mcvay, 1971). Here we find systems of predictable patterns similar to combinatorial structure in human language, with absence of complex semantics. This shows that combinatorial structure can exist without apparent pressure from a large repertoire of signals. Combined with the case of ABSL, the connection between combinatorial structure and growing repertoires of meanings is weakened in both directions: large repertoires of meanings exist without combinatorial structure and combinatorial structure exists without large repertoires of meanings.

In addition, the existence of pseudo words in human language suggests independence (Fitch, 2010). There are many more possible words that are well formed in a language than are actually used in the vocabulary, which is puzzling if one assumes that vocabulary drove the expansion of possible words.

In summary, many sources can be used to answer questions about the emergence of combinatorial structure, but the results so far are inconclusive. This thesis presents a collection of studies in which the

emergence of combinatorial structure is investigated experimentally. Chapter 3, chapter 4 and chapter 6 all describe studies in which the *experimental iterated learning* method is used. The next section provides a review of previous work in which experimental methods were used to study language evolution.

2.3 Language evolution experiments

In the field of language evolution, the acquisition of tangible evidence to support the many diverse theories is a difficult endeavour. As illustrated in chapter 1 important breakthroughs were made when new methods were discovered and interdisciplinary collaborations were made. The adoption of techniques from the field of Artificial Life for instance, such as computer simulations with interacting agents and robotic societies, provided an entirely novel way to test hypotheses and obtain new perspectives. One of the important new insights gained with this method is the fact that language should be viewed as a complex adaptive dynamical system (Brighton and Kirby, 2001; Kirby, 2002; Steels, 1997b), also explained in chapter 1.

Computer simulations have been used to show the importance of social and cultural processes in language evolution dynamics. When studying language as a dynamic cultural system in populations of language users, causes and effects quickly become hard to determine. Computer simulations help to investigate effects of certain variables in a controlled way. It was shown that self-organising principles could explain the emergence of certain linguistic structures without the need to assume that humans are born with language-specific innate cognitive biases. Models that simulated the interactions between artificial agents in populations demonstrated for instance how gradual conventionalisation and alignment could result in shared artificial languages (e.g. de Boer, 2000; de Boer and Zuidema, 2010; Steels, 1997b). Simulations of the cultural transmission of language from generation to generation as modelled within the *iterated learning framework* (Kirby, 2002; Kirby and Hurford, 2002) showed how languages become learnable and structured by being passed through a *transmission bottleneck*. Agents in these models learn their language by observing the productions of other individuals who also learned it in that way. As explained in chapter 1, a bottleneck is introduced because naïve individuals with no previous experience have to acquire the language but they are never exposed to every possible utterance in that language. Therefore these individuals have to generalise and make hypotheses about the structure of the language and they will produce utterances that are in line with those hypotheses. When this happens repeatedly, the language as a population-level system will adapt to the learning biases and constraints of the agents and become easier to learn (Brighton and Kirby, 2001; Kirby, 2002; Kirby and Hurford, 2002). Language therefore seems to be

shaped by its own transmission and the brains of its users (Christiansen and Chater, 2008; Deacon, 1997; Griffiths and Kalish, 2007; Kirby and Hurford, 2002).

Computer models generally abstract away from the full complexity of human communication. Because of this there is some resistance in the acceptance of findings from computer simulations on language evolution, which is illustrated clearly with the quote mentioned by Kirby et al. (2008) from Bickerton: “Powerful and potentially interesting although this approach is, its failure to incorporate more realistic conditions (perhaps because these would be more difficult to simulate) sharply reduces any contribution it might make toward unraveling language evolution. So far, it is a classic case of looking for your car-keys where the street-lamps are.” (Bickerton, 2007, p. 522).

One step towards more realism was taken when computer agents in simulations were given a body. These computer agents were embodied in the shape of robots and could therefore operate in the real world (Steels, 1997b). The Talking Heads experiment (Steels, 1997c) is one specific example of such a study, which had a set-up with two robotic heads, each with a camera and both observing a scene. Computer agents could ‘load’ themselves into the physical head and interact with the agent in the other head about the scene, where one of the two had to guess the topic the other was interacting about and shared lexicons emerged. This resulted in the study of a more realistic meaning space and provided insights into category formation and co-evolution between language and meaning. The brains of these agents however were still abstract and simplified.

Meanwhile, in the field of linguistics researchers were carrying out Artificial Language Learning (ALL) experiments in which human participants had to learn invented artificial languages. An example of this is an influential study by Saffran et al. (1996), intended to investigate how 8-month old infants segment words from fluent speech. They hypothesised that there was a potential source of statistical information infants could use, namely that the transitional probabilities of sound changes in the speech stream are higher within words than between words. The researchers wondered whether the infants could use this information. They exposed them to continuous streams of artificial speech in which the only cue to word boundaries was the statistical information. In a later test phase the infants were able to distinguish words from non-words, which indicated they indeed use statistical information.

It was recognised that ALL could potentially be a fruitful method to explore in the field of language evolution (Christiansen, 2000) and this technique (also known under the heading of the Artificial Grammar Learning paradigm) became widely used, as reviewed by Fitch and Friederici (2012), not only with humans as test subjects but also with animals. Eventually a seminal study was done by Galantucci

(2005) in which there was no initial invented language, but artificial communication systems emerged from scratch in the laboratory. In this experiment, two participants had to play a multiplayer video game in which they could only communicate with the use of a special graphical device. This device prevented the use of symbols or pictures because there was no direct mapping between the drawing action and what appeared on the screen. The success of solving the game depended on cooperation between the participants and towards the end of the experiment communication systems emerged quickly. At the end of the study the pairs had to play the game for 5 minutes without being able to use the communication device and they performed significantly worse in this case. Interestingly, the sign systems that emerged all were approximately equally effective as communication systems, but there was a wide variation in terms of the encoded messages and how these messages were coded.

Following this study of language emergence in the laboratory, a variety of other experimental designs were studied, as reviewed by Scott-Phillips and Kirby (2010). Other paradigms for strategy game experiments for instance were created, such as the embodied communication game, in which there is no pre-defined communication channel, but the actions players take to solve the game become communicative (Scott-Phillips et al., 2009). Another paradigm makes use of pictionary-style tasks (e.g. Garrod et al., 2007; 2010; Theisen et al., 2010) in which one person has to guess the topic the other is trying to communicate by drawing. A third paradigm studies iterated learning with human participants (e.g. Kirby et al., 2008) by simulating in the laboratory how behaviours such as language are culturally transmitted. In this thesis the focus is on this method, therefore it is explained in more detail below.

As mentioned above, iterated learning has been studied with computer models through agent-based simulations with a variety of learning mechanisms such as grammar induction (Kirby, 2000; Kirby, 2001), neural networks (Hare and Elman, 1995; Smith, 2002), minimum description length learning (Brighton and Kirby, 2001) and Bayesian inference (Griffiths and Kalish, 2007; Kirby et al., 2007). Kirby et al. (2008) introduced a method that allowed them to replicate the findings of these computer models in the laboratory by conducting experiments with human participants.

In iterated learning experiments participants have to learn and reproduce a set of signals. The set of signals is based on the output of a previous participant in the same experiment and the participants' own output is used as input for the next participant. In this way *chains* of transmission are created. The development of the set of signals that is being transmitted can be closely investigated and it reveals how individual (cognitive) biases and learning behaviour gradually influence this system (Christiansen and Chater, 2008; Deacon, 1997; Griffiths and Kalish, 2007; Kirby and Hurford, 2002). Kirby et al. (2008) demonstrated the emergence of compositional syntactic structure using

this experimental method. The utterances in these experiments were typed strings of characters referring to objects that differed in shape, colour and movement. Over experimental 'generations' of learning and reproduction, the compositional structure in these languages cumulatively increased and the languages became easier to learn. This happened without conscious invention of structures by individual participants and without an influence of communication. At the end almost all words of one of the languages were composed of three 'morphemes', where each morpheme consistently coded one of the three dimensions in the meaning space. This regularity made it possible for participants to predict the words for objects they had never been exposed to during training.

Not only the emergence of compositional syntax has been studied with this method, but several other features of linguistic systems have been investigated as well. Reali and Griffiths (2009) studied the development of an artificial language consisting of spoken sequences of syllables as words for objects, where each object was associated with one of two different words with a certain probability. Participants' knowledge of the learned language was tested by asking them to select one of the two words as the right one with a forced choice task. Based on the responses of one participant, the probabilities of the word-object pairings for the input-language for the next person were determined. After some iterations of this procedure it became clear that synonymy in the languages disappeared. The unpredictable variation in the word-object relations became regularised.

A similar loss of unpredictable variation was found by Smith and Wonnacott (2010) in artificial languages with morphological variability. Here, participants learned and reproduced sentences describing a scene involving either one cartoon animal or a pair of the same cartoon animals. Plurality was indicated with two different markers that were both used in combination with each of the nouns referring to the cartoon animals, but with different frequencies. This made the use of plural marking unpredictable and irregular. In the language that was passed on to the next participant, the produced sentences from the previous person were used. After repeated iterations of learning and production, the variability in plural marking did not disappear in all languages, but it did become more regular. The nouns ended up being used exclusively with one of the two markers, which made the system more predictable.

Combinatorial structure in visual signals was studied by del Giudice et al. (2012; 2010) in an iterated learning experiment in which participants had to learn and reproduce a set of graphical signals. These signals were produced with the use of a graphical device that was built following the design by Galantucci (2005) which was mentioned above. An initial set of random squiggles developed into a set with reuse of basic elements over generations. Combinatorial structure therefore increased as a result of repeated transmission.

In addition, the method has been applied to study the emergence of colour terms (Dowman et al., 2008) as well as non-linguistic category or function learning tasks (Griffiths et al., 2008a; Griffiths et al., 2008c).

The results of this body of experimental work confirm the idea that structure in language-like systems evolves culturally and comes to reflect human cognitive biases and constraints on learning, memory and production (Christiansen and Chater, 2008; Deacon, 1997; Griffiths and Kalish, 2007; Kirby and Hurford, 2002). After systems have been transmitted over a number of experimental generations, human over-generalisation causes them to become regularised. In the experimental results a tendency towards the emergence of compressible, predictable systems appears to be a recurring theme. Another example that clearly demonstrates this is an experiment where the iterated learning paradigm was used to study human inductive biases for learning different types of category structures (Griffiths et al., 2008b). Griffiths et al. (2008b) used a set of category structures for which it had previously been shown that the difficulty of learning these structures could be predicted by the incompressibility of the member concepts (Feldman, 2000). The concepts in these studies were 'amoebas' that contained a nucleus which differed according to three binary features: shape, size and colour. In the iterated learning study (Griffiths et al., 2008b), participants were presented with examples from categories of amoeba and were asked to select a hypothesis (choosing from a number of different completions of the set) that they thought best described the underlying category structure. New input data was generated following the distribution of the chosen types of category structures in a participants' responses. The results showed that those category structures that Feldman (2000) found to be more easily learned and for which the member concepts are more compressible, were increasingly chosen across generations of iterated learning. This reflects a bias towards these more compressible structures and shows that human learning and generalising from a few examples result in categories of amoebas that can be more efficiently coded.

2.4 Compression and the brain

The experiments discussed in the previous section indicate not only that linguistic structure may be the result of an evolutionary process in which languages gradually adapt to be learnable by their users (Kirby et al., 2008), but the results also seem to reflect a general tendency of the brain to compress information and make predictions. Some of the computer models about iterated learning have incorporated this idea, using models that implement inductive learning strategies such as minimal description length learning (Brighton, 2005; Brighton and Kirby, 2001; Teal and Taylor, 2000) or Bayesian prediction (Griffiths and Kalish, 2007; Reali and Griffiths, 2009; Smith, 2009). These models successfully simulate

behaviour of participants, but to gain a deeper understanding of what it is exactly about the human brain that leads to the observed iterated learning results and to learn more about the nature of relevant cognitive biases and where these biases may come from, it may be informative to look at some relevant results from the field of neuroscience.

The idea that brains encode information efficiently is not at all new. Barlow (1961) proposed that efficiency plays a role in the coding of sensory information and at present many brain theories and learning models exist that are based on this assumption (see for instance Chater and Vitányi (2003); Friston (2010); Olshausen and Field (2004); Schmidhuber (2009)). In the domain of cognitive processing, Chater and Vitányi (2003) present a review of studies that link cognitive tasks with efficient coding and discuss empirical evidence in line with their 'simplicity principle'. These studies encompass all kinds of cognitive and perceptive tasks, including linguistic processing. This principle has been applied to model language acquisition (Onnis et al., 2002) and ease of language acquisition has been linked to information theoretic principles before by Clark (1994). For decades neuroscientists have studied the hypothesis that compression and simplicity are important principles in neural processing with advanced computational techniques and precise measurements of neural responses of for instance cats, rats, monkeys and rabbits (as reviewed by Olshausen and Field (2004)). The studies that seem particularly interesting to the work described in this thesis are those in which it is demonstrated that brain processes are adapted to encode natural stimuli most efficiently.

A large body of work on this has been dedicated to the visual domain and more recently similar results have been found for auditory signals. Simoncelli and Olshausen (2001) review work in which the efficient coding principle is tested in visual systems. They give an overview of the regularities and statistical structure that can be found in natural images (such as mountains, rocks, trees) and present many examples of quantitative evidence in which these regularities are linked with structured neural responses. The main approach in this field is to create a model, and to adjust the parameters in such a way that the model optimally encodes the input data, for instance a set of images. Optimality is usually some measure on how well the input data can be reconstructed from the coded data. The resulting representations are then compared with real neural data. Olshausen and Field (1996,1997) for instance define a model in which images are encoded using linear combinations of basis functions. The set of functions is updated in the direction of an optimally efficient code. Properties of the basis functions that emerge as the final solution resemble those of single cell receptive fields in the early visual (V1) system (Olshausen and Field, 1996,1997), suggesting that these receptive fields encode natural stimuli efficiently.

In the auditory domain similar methods have been used (Lewicki, 2002; Smith and Lewicki, 2006). Smith and Lewicki (2006) used a model that

encoded sounds as a set of basis functions. These functions could have different shapes, lengths and onset times and they were optimised so that they encoded natural sounds (such as animal vocalisations, rain, cracking twigs) most efficiently. In parallel, response functions were computed for auditory nerve fibre measurements of a cat listening to the same set of sounds. The set of basis functions that emerged in the computational model was compared to the set from the actual brain measurements and these were found to be remarkably similar. This suggests that the cat brain encodes the structure present in natural sounds in an efficient way. Interestingly, Smith and Lewicki (2006) performed the same procedure with their model to find a set of functions optimised for the sounds of human speech. What they found was very similar to the results with natural sounds, namely that the basis functions that efficiently encode speech also closely resemble auditory response functions of a cat. This suggests that the sounds used for speech are likely adapted to the efficient auditory coding of the mammalian brain. Comparable results have recently been found with another efficient coding model for speech and comparisons with neural structures higher up the auditory pathway, as measured in cats and gerbils (Carlson et al., 2012). Since it is implausible that cat auditory processing has evolved to efficiently encode human speech, we may well assume that the sounds used in language are adapted to the (mammalian) auditory cortex. This therefore provides another convincing source of evidence supporting the view that linguistic organisation may have emerged in adaptation to the brain (Christiansen and Chater, 2008) and is not a reflection of innate biological adaptations. Although neuroscientific data does not play a role in the current thesis directly, it will be addressed again in chapter 8.

2.5 The present work

So far, there have not been many studies on experimental iterated learning of continuous signals and to the best of my knowledge this thesis provides the first investigation of the emergence of combinatorial structure in continuous systems of sounds. The previous work on experimental iterated learning has mostly focused on either aspects of language that can be represented in a discrete, symbolic way such as for instance morphology and compositional syntax (Kirby et al., 2008; Smith and Wonnacott, 2010) or on the emergence of graphical symbols. del Giudice et al. (2012; 2010), as mentioned before, used the device that was created by Galantucci (2005) in an iterated learning experiment about combinatorial structure in graphical signals. Galantucci et al. (2010) and Roberts and Galantucci (2012) also used Galantucci's (2005) device to study combinatorial structure in the laboratory but in communicative game settings, with no vertical transmission. Galantucci et al. (2010) showed that rapidity of fading influences the emergence of combinatorial structure. When signals fade

faster, more reuse of basic graphical elements was observed. In the study conducted by Roberts and Galantucci (2012), participants play naming games and communicate about animal silhouettes using the graphical device. They studied the influence of both conventionalisation and set size (vocabulary size) and found that the first could indeed cause combinatorial structure to emerge while the link to set size was less clear. Garrod et al. (2010) studied iterated learning chains in which participants did the pictionary task, but the focus was neither on combinatorial structure nor on sounds. Theisen et al. (2010) used the pictionary task in an experiment in which pairs of participants communicated about concepts from a structured meaning space. The graphical signals increasingly exhibited systematic reuse of arbitrary elements. Some previous studies have used sounds, but not in the same way as it is used in the studies presented here. Fay and Lim (2010) for instance asked participants to communicate using non-speech vocalisations, but no transmission chains were created and the signals themselves were not analysed, only the communicative success. Reali and Griffiths (2009) used non-existing spoken words, but there the aim was not to study the emergence of combinatorial structure and the participants did not produce the signals.

Most of the chapters in this thesis describe experiments in which experimental iterated learning was used to investigate the emergence of combinatorial structure in artificial languages that consist of continuous signals in the auditory modality. The experiment conducted by Kirby et al. (2008) formed the main example for this work. The reason for the use of experimental iterated learning as opposed to communicative or game strategic experimental paradigms is the fact that principles of economy (as described in section 2.2) had been proposed to play a role in the formation of combinatorial structure. As mentioned above, iterated learning seems to have a relation with efficiently coded, predictable systems. It could therefore be a strong method for demonstrating how combinatorial structure may emerge and potentially provide an explanation for the observed patterns in phonology. In order to make the method applicable to the study of combinatorial structure and continuous signals, the paradigm had to be adjusted in several ways. As will become clear especially in chapters 3, 4 and 6 this included an exploration of issues such as the nature of the signals, the production apparatus for creating these signals, the design of measures for quantifying combinatorial structure and the use of meanings, among others.

Scribbles

3

Speech sounds are organised: they are both categorical and combinatorial and there are constraints on how elements can be recombined. How did speech become organised in this way? As we have seen in chapter 2, different theories exist about the origins of combinatorial structure in language. Did it emerge because structural recombination of elements is needed to maintain clear communication with a growing meaning space, as Hockett (1960) suggested? Was the main pressure that drove the emergence signal dispersion? In chapter 2 several sources of evidence were highlighted that have been used to gain insight into these questions, from computer models via newly emerging sign languages to animal communication systems and more. There is a growing wealth of data, but together these findings still do not lead to a consistent answer. As reviewed in chapter 2, experimental methods have recently gained popularity in the field of language evolution. This chapter describes an experiment that was conducted using this method as a first attempt to simulate the emergence of combinatorial structure with human participants in the laboratory.

This chapter contains parts from the article:
Verhoef, T., de Boer, B.G., del Giudice, A., Padden, C. & Kirby, S. (2011). Cultural evolution of combinatorial structure in ongoing artificial speech learning experiments. *CRL Technical Report*, 23(1), 3-11.

3.1 Experimental iterated learning with continuous signals

The study by Kirby et al. (2008) was the most important example for the experiment described below. Kirby et al. (2008) exposed participants to an artificial language in which strings of characters, typed in using a keyboard, were words for objects that differed in colour, shape and style of movement. During training participants in their study only got to see about half of the objects, so there was a strong learning bottleneck. After a learning phase, participants were asked to (re)produce the strings for the objects, even those they had not been exposed to in training. The words that one participant reproduced were used to train the next person. After repeated transmissions, compositional structure emerged in the artificial languages (Kirby et al., 2008).

The strings that formed the signals in the experiments of Kirby et al. (2008) are composed of letters, so they are based on an already discretised set of primitives. However, in language there are (at least) two layers of combination (which Hockett (1960) called duality of patterning as discussed in chapter 2). Meaningless sounds (in the case of speech) are combined into meaningful words and phrases, but meaningful words and phrases are also combined to compose other meaningful expressions. The second layer represents compositional structure and this is what emerged in the experiment of Kirby et al. (2008). To be able to investigate the emergence of the type of organisation that is typical of the first layer, we need to use an artificial language with continuous signals. The experiment described in this chapter is designed as a first attempt to do this. The experiment is otherwise kept as similar as possible to the original study by Kirby et al. (2008), but with a simpler version of the meaning space and continuous signals.

3.2 Scribble to sound

Many experimental paradigms that have emerged in the field of language evolution are in one way or another based on or related to designs that were used in computer models studying the same phenomenon, as reviewed by Scott-Phillips and Kirby (2010). Studies involving iterated learning in the laboratory (Kirby et al., 2008; Smith and Wonnacott, 2010), for instance, followed findings that had been obtained with agent-based computer simulations (Kirby and Hurford, 2002). Experiments that investigate social coordination and the emergence of communication systems (e.g. Galantucci (2005); Scott-Phillips et al. (2009)) have commonalities with computer agent and robot experiments that involve language games (Steels, 1997b) or coordination tasks (Quinn, 2001). Phonological combinatorial structure has also been studied with the use of computer models. It has for

instance been investigated how discrete categories can emerge in acoustic communication systems.

As reviewed in chapter 2, a discrete set of vowel categories can emerge through self-organisation (de Boer, 2000; Oudeyer, 2006). In addition, de Boer and Zuidema (2010) have shown that self-organisation in a population of interacting agents can lead to combinatorial structure. In their model, the signals that are used for communication are continuous trajectories in a two-dimensional acoustic space. Both holistic and combinatorial signals are produced as signals that change over time and are therefore constructed in the same way. This system formed the inspiration for the type of artificial languages used in the experiment described below.

For the artificial languages in the current experiment, sounds produced with the voice had to be avoided because this study aims to investigate the emergence of discrete and combinatorial organisation, but humans already have such structure in their speech. An artificial articulatory apparatus was therefore designed and implemented. With this device, participants scribbled trajectories like the ones in de Boer and Zuidema (2010), in a two-dimensional square on a computer screen with the mouse. The software transformed these scribbles into sounds. The experiment described in this chapter therefore roughly combines the experimental set-up of Kirby et al. (2008) with the artificial linguistic signals design of de Boer and Zuidema (2010).

3.3 Methods

The experiment described in this chapter is a first attempt at investigating the emergence of combinatorial structure in sound systems through experimental iterated learning. 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.

3.3.1 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 euros in cash to compensate for their time.

3.3.2 Stimuli

The signals that were transmitted were produced by drawing continuous trajectories on a computer screen. The trajectories were composed of a

single, continuous curve in a two-dimensional space. These trajectories were transformed into sounds. Participants needed to learn to recognise 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 created 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.

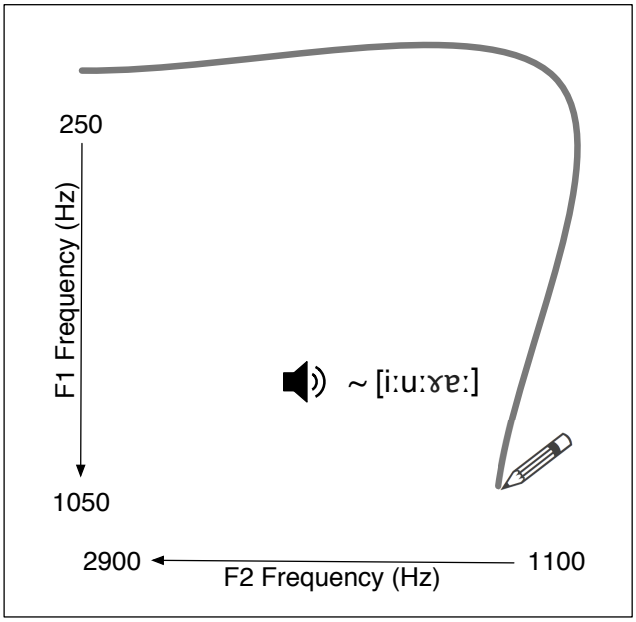


Figure 3.1: *Representation of the scribble to sound mapping. The trajectory that is shown in the figure would approximately sound like "iiiiiiuuuuuuaaaaa". Note that participants did not see the axes or transcriptions, the scribble area on the screen was empty.*

Figure 3.1 shows an explanation of the mapping in the scribble space. A screenshot of the user interface for the experiment is shown in appendix A.2. At the beginning of the experiment a random set of sounds was created by letting the computer draw random trajectories in the scribble space, with certain constraints (details can be found in appendix A.3). 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. The Dynamic Time Warping distance (Sakoe and Chiba, 1978) on the sequences of x, y coordinates in the scribble space was used to determine this distance.

Meaning space The meaning space consisted of nine pictures of different objects (squares, circles and rings) that had different colours (red, green or blue). Figure 3.2 shows these pictures. At the beginning of the experiment, each picture in the meaning space was randomly assigned to a unique sound, from the set of random sounds in the signal space, to create the initial set of sound-meaning pairs.

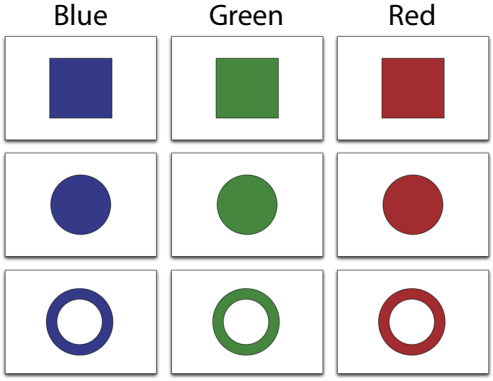


Figure 3.2: *Meaning space*

3.3.3 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 written instructions can be found in appendix A.1. The participants were given a chance to ask questions before they started with the practice phase. In this phase the subjects were asked to familiarise 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 with the scribble device. Feedback on the imitation accuracy was provided by showing a coloured border around the picture. This border could have any colour on the continuum from red to green, where red indicated a very low imitation accuracy and green indicated a high accuracy. 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 within this paradigm relies on a transmission bottleneck (Kirby et al., 2008; Smith et al., 2003; Zuidema, 2003). Learners are not exposed to every possible expression during acquisition. 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 generalising from learned items. In the experiment described in this report the transmission 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.

3.3.4 Modifications

After the first two diffusion chains were completed, a few observations could be made 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 It was 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 score as 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 6 generations.

Adaptive learning Another observation that was 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.

3.3.5 Expectations

The expectation was to find an increase in the amount of structure in the systems of sounds that were transmitted at the end of each transmission chain. This structure is combinatorial if it consists of a systematic reuse of basic building blocks in the sounds. It has been shown before that the mechanism of iterated learning can lead to the emergence of compositional structure (Kirby et al., 2008; Kirby and Hurford, 2002) and my hypothesis is that it leads to structure on the sub-lexical, phonetic level as well. In addition, an increase in the learnability of the set of signals was expected as the chain progresses, because the sound systems change to become optimised 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.

3.4 Results

In this section the qualitative results are presented first, showing the development of the sound systems from generation to generation. This gives insight into the kinds of structure that did and did not occur. Second, a quantitative analysis is shown, demonstrating how the learning ability changed over the course of each chain.

3.4.1 Qualitative results

In figures 3.3 and 3.4 the output in the first two chains is shown. The first row shows the trajectories for the random input sounds and each following row shows the output produced by a participant who received the data from the previous row as input.

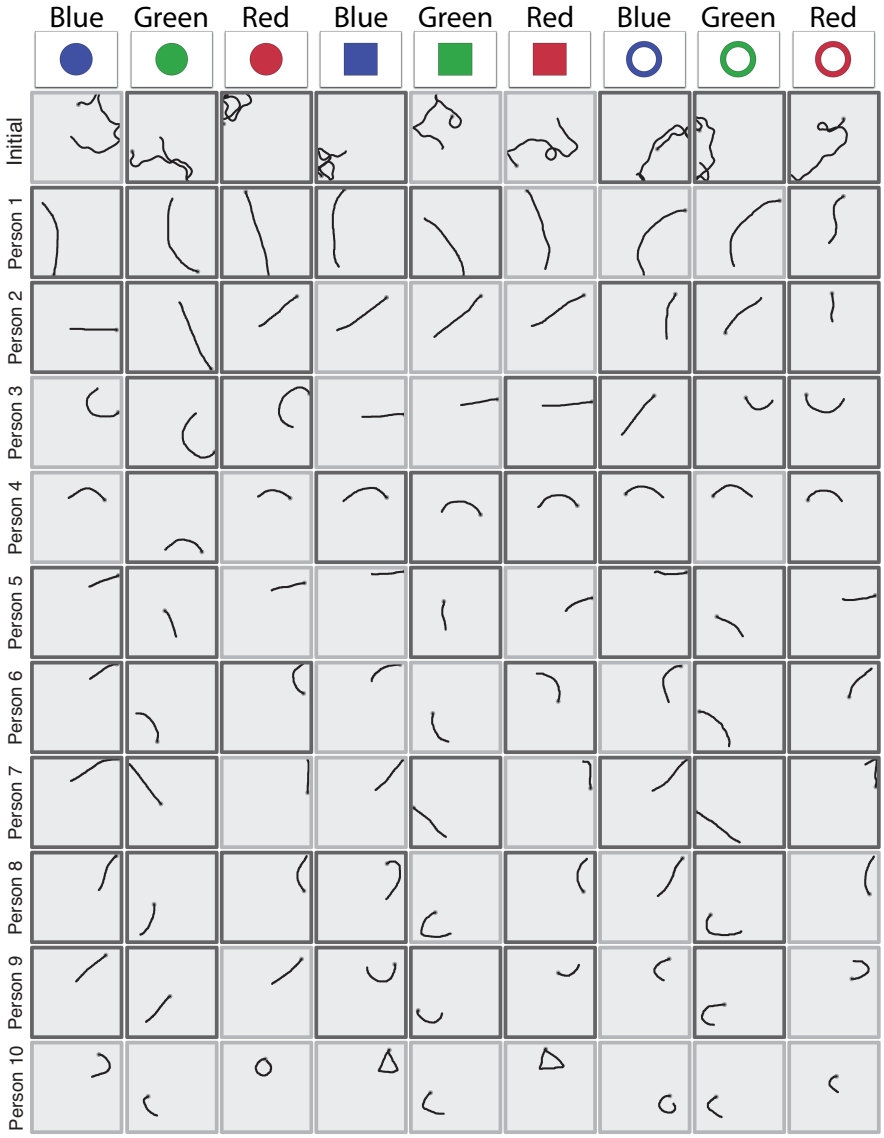


Figure 3.3: *Scribbles produced by participants during the final test in chain one. The first row shows the trajectories for the random input sounds and each following row shows the output of a participant who received the data from the previous row as input. The darker border around the picture means that this item was part of the training set for the next person. The grey dots indicate the starting point of the trajectories.*

Figure 3.4: *Scribbles produced by participants during the final test in chain two. The first row shows the trajectories for the random input sounds and each following row shows the output of a participant who received the data from the previous row as input. The darker border around the picture means that this item was part of the training set for the next person. The grey dots indicate the starting point of the trajectories.*

The darker 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. The starting point of each of the scribble trajectories is indicated with a grey dot. 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.

In both chains it can be observed that there is a tendency towards structure in which signals relate to parts of the meanings. Often the same signals are used for all objects with the same colour or shape. Right from the beginning participants seem to search for patterns and apply generalisations. Often features such as the length of the sound, or the location of the trajectory in the space (influencing vowel quality) are linked to colours 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 colour 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 (colour and shape) is distinguishably indicated in the signals (see figure 3.5). For person nine, all circles are expressed as straight lines, squares as cup-shaped trajectories and rings as hooks. Green coloured 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 from those for red (except for the circle, but the participant only made this mistake in the last output round, not 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. This observation will be discussed further in the discussion section.

In chain two, the first hints of structure appear in generation two (see figure 3.6). In this set, the location of the scribbles is clearly linked to the colours 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 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 3.7). 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

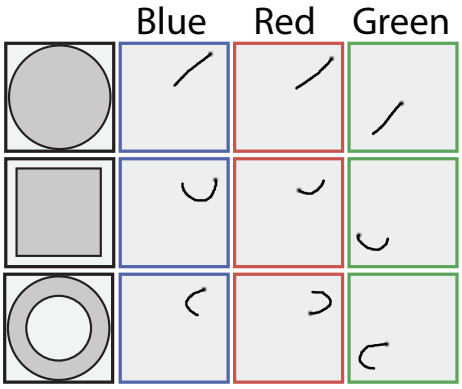


Figure 3.5: *Chain one, generation nine. Note that the shape of the trajectory appears to express the shape of the object, while the position of the trajectory expresses the colour of the object.*

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 both chain one and chain two it is clear that the range of different signals quickly becomes more constrained with increasing generations of transmission. In the initial set, every possible trajectory could be a part of the set, but towards the end of the chains the range of possible 'well-formed' scribbles is much more reduced. In the beginning the trajectories can start anywhere in the two-dimensional space and it can progress in any direction, with an undefined number of changes of direction. But in chain two for instance towards the end, each trajectory in the set starts on the left, moves to the right and has only a very limited number of changes of direction (mostly none). For the objects to which the participants are

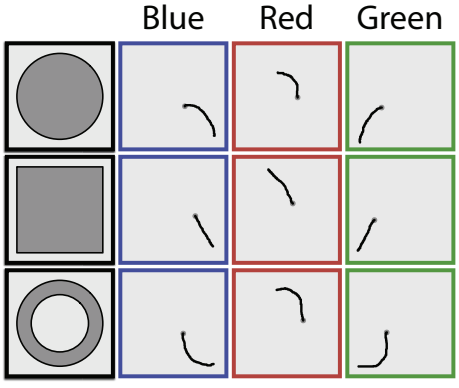


Figure 3.6: *Chain two, generation two. Note that the location of the trajectory indicates the colour of the object in the meaning space.*

not exposed in training, the produced trajectories appear to stay in line with these ‘rules’.

In chain three, the additional guessing task was added in response to the observation that participants did not pay much attention to the sounds during the experiment. Although this change was introduced to improve listening behaviour, such improvement could not be detected. The results in this chain were qualitatively the same as those in the first two chains without a noticeable difference in listening behaviour. In the discussion section a possible explanation for this will be proposed, but for now we will take a look at the qualitative results. Because of the fact that an improvement in the listening behaviour of our participants could not be observed, this chain was terminated after six generations, so as to start a new chain with another modification (as described below). In figure 3.8 the output produced by participants in chain three is shown. The first row again shows the trajectories for the random input sounds and each following row shows the output produced by a participant who received the data from the previous row as input. The darker border around the picture again means that this item was part of the training set for the next person and the grey dots indicate the starting points of the trajectories.

In this chain we can see, like in chains one and two, the emergence of a relation between location and colour. In generation two for instance, high scribbles are for blue objects, low scribbles are for green objects and scribbles at medium height are for red objects. However, per colour the signals for the different shapes are all the same, therefore the signals can no longer be used to distinguish the objects along this dimension. This issue is addressed again in the discussion section. The structure does not persist towards the end but whenever there is a slight (local) regularity in the signal to meaning mapping, it does tend to survive longer. This can be illustrated by looking at the example in figure 3.9. This example shows

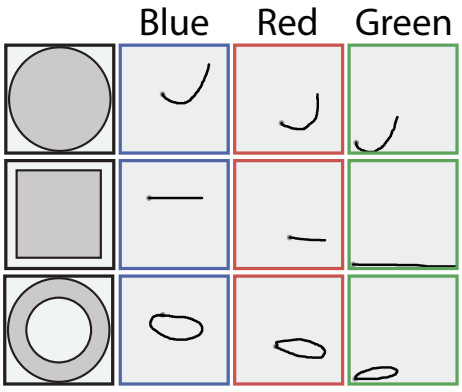


Figure 3.7: *Chain two, generation four. Note that the shape of the trajectory appears to express the shape of the object, while the position of the trajectory expresses the colour of the object.*

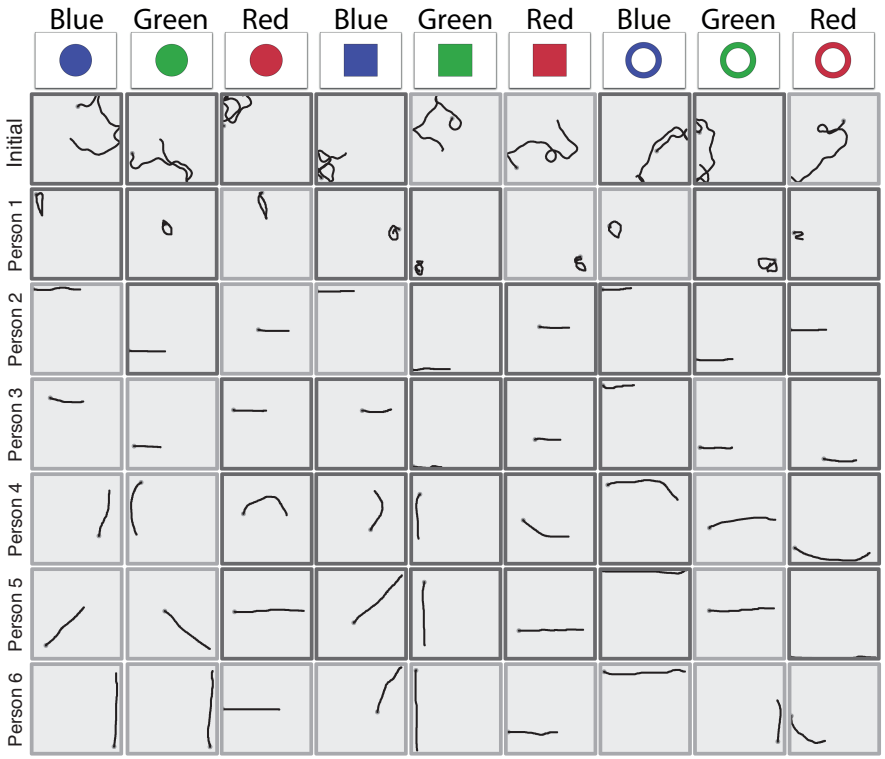


Figure 3.8: *Scribbles produced by participants during the final test in chain three. The first row shows the trajectories for the random input sounds and each following row shows the output of a participant who received the data from the previous row as input. The darker border around the picture means that this item was part of the training set for the next person. The grey dots indicate the starting point of the trajectories.*

the productions of three successive generations for the donut shaped objects. It can be observed that all three participants follow the 'rule' that connects colour to scribble height in the space, even though none of these participants were exposed to the green object. Apparently this is what makes sense to the participants (if blue is high and red is low, than green must be in the middle) and it is a mistake (note that the mapping in generation two is different) that consistently gets replicated. Like the first two chains, this chain also shows an increase in signal constraints towards the end. The variation in scribble length, direction and shape is strongly reduced.

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 this regime was introduced in the hope that it 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. This chain was therefore excluded in the further analysis.

In summary, the qualitative results indicate that some hints of structure did emerge from time to time in the chains, but it did not lead to the expected outcome. The structures that emerged mostly did not persist throughout the chain until the end and they were of a different type than the sort of regularities that were intended to be encountered. Possible explanations for these and other issues are presented in the discussion section.

3.4.2 Quantitative results

In order to find out whether the sound-meaning systems were optimised to become more learnable by being transmitted through chains of human learners, the performance from generation to generation in each chain was measured. For each participant the distance between the input set they received and the output they created for each meaning was measured, by using the distance measure as described above. Figures 3.10, 3.11 and 3.12 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 meaning-sound pairs they were never trained on). In the case that the average distance between input and

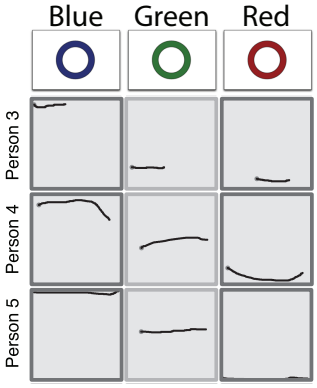


Figure 3.9: *Produced scribbles of three successive generations for the donut shaped objects. All three participants follow the 'rule' that blue is high, red is low and green is middle, even though none of these participants were exposed to the green object.*

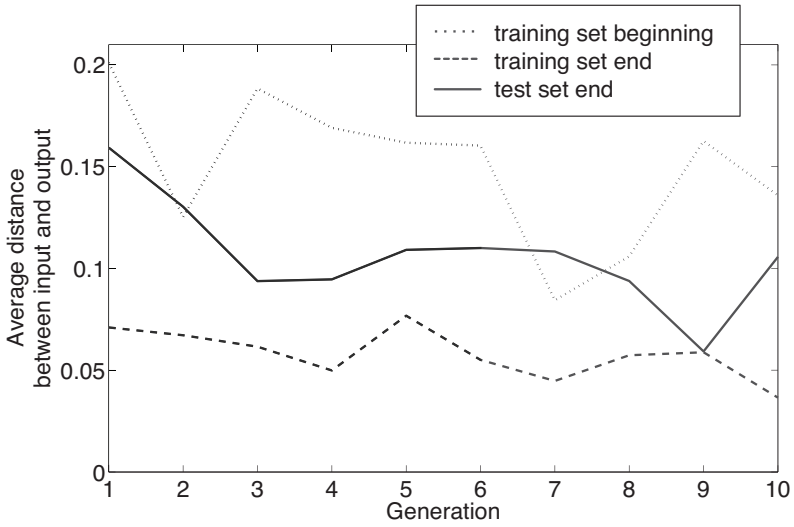


Figure 3.10: *Average distance between input and output for chain one in three different situations: at the beginning of the experiment including only the training set, at the end of the experiment including only the training set, and at the end of the experiment including the complete set (also the three meanings they were never trained on).*

output is approximately the same on the training and test set, it means that the participant performed just as well on the meaning-sound pairs they never saw as on the other six. This therefore probably means that this person generalised by using the structure to decide on the sounds for the unseen meanings. Figures 3.10, 3.11 and 3.12 show that this happens only a few times throughout the chains. It is clear that there is a relationship between the emergence of 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 colour in the meaning space (as illustrated in figure 3.3). 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 generalisation.

Even though it happens a few times that learnability increases rapidly from generation to generation, it does not persist throughout the entire

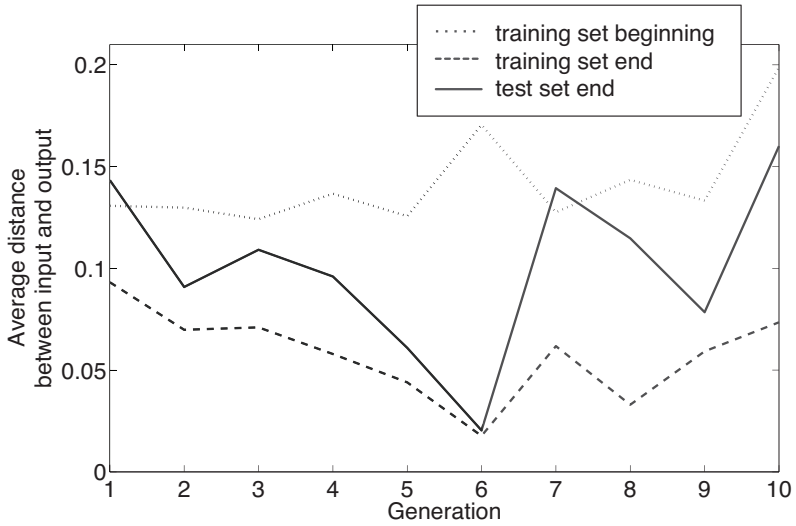


Figure 3.11: *Average distance between input and output for chain two in three different situations: at the beginning of the experiment including only the training set, at the end of the experiment including only the training set, and at the end of the experiment including the complete set (also the three meanings they were never trained on).*

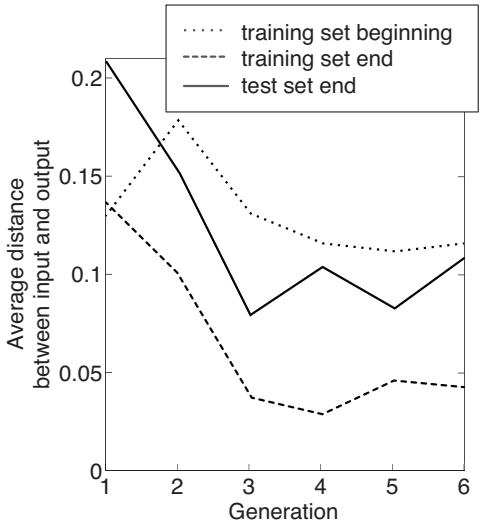


Figure 3.12: *Average distance between input and output for chain three in three different situations: at the beginning of the experiment including only the training set, at the end of the experiment including only the training set, and at the end of the experiment including the complete set (also the three meanings they were never trained on).*

chain until the end. Just as the structure that sometimes emerges disappears again, the increased learnability disappears with it.

3.5 Discussion

The experiment described in this chapter 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, an increase in learnability of the systems that were being transmitted, as well as an increase of the combinatorial structure within the systems was expected to be found. Although interesting changes 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. Because of this many participants failed to pick up on any potential structure that emerged previously and were therefore unable to transmit it. 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 generalisation and introduction of structure did happen. 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.

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 compared with combinatorial phonology, except perhaps in terms of the emerging constraints in the signal space. The observed structure is actually more comparable to syntactic compositional structure, because the location and shapes in the scribble space are directly linked to colours 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 sometimes structure emerges that is clearly visible when inspecting the scribbled trajectories directly, but involves barely audible distinctions in the auditory modality. An

example is shown in figure 3.5. 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 coloured 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? This may be due to the feedback that is given to 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. The feedback was meant to only help participants to learn the scribble to sound mapping, but it unintentionally also introduced a shortcut for solving the task. 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, it was 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.

In the reproductions produced by participants, it was not uncommon that the same signal would be repeated for different objects. This lead to underspecification and the loss of expressive power of the signals. In the experiments by Kirby et al. (2008) this also happened and in their study they prevented this by filtering the produced output of one participant for duplicates so that the next participant would never be exposed to homonymic examples. This successfully solved the problem of underspecification. With the design of the experiment described in this chapter I hoped and expected that underspecification would not play a role, because with continuous signals it is not easy to produce the

exact same signal twice, unlike in the case of typing strings of characters. Contrary to this expectation, underspecification did play a role, resulting in a system where different objects were mapped to signals that were very similar and only differed from each other by negligible variations. Perhaps this was due to the fact that small differences in the trajectories were barely audible and this is a point to keep in mind with future designs that involve continuous signals.

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. The results 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, the use of a less structured meaning space or no meaning space at all and the introduction of an intervention to prevent underspecification. The lessons learned from this study gave rise to ideas for a follow-up experiment. This experiment is discussed in the next chapter.

Whistles

4

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*

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*

4.1 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.

4.2.1 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

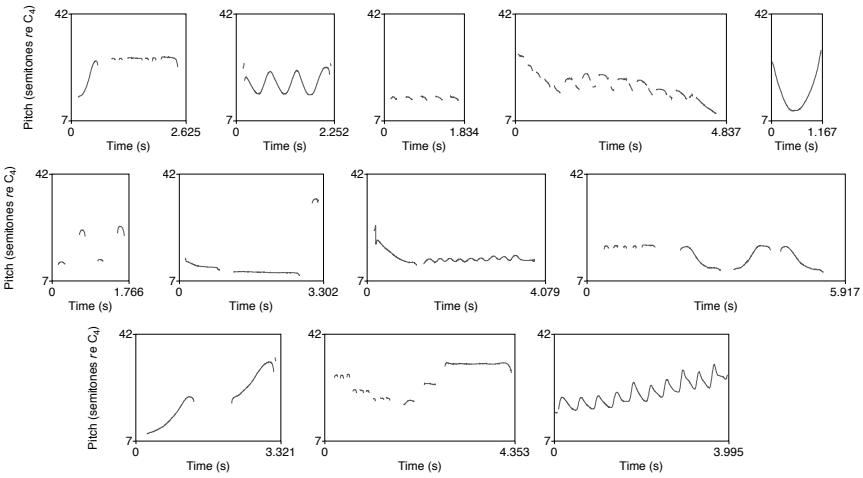


Figure 4.2: *Whistles from the initial whistle set, plotted as pitch tracks on a semitone scale. Note the diversity and complex structure of the whistles.*

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

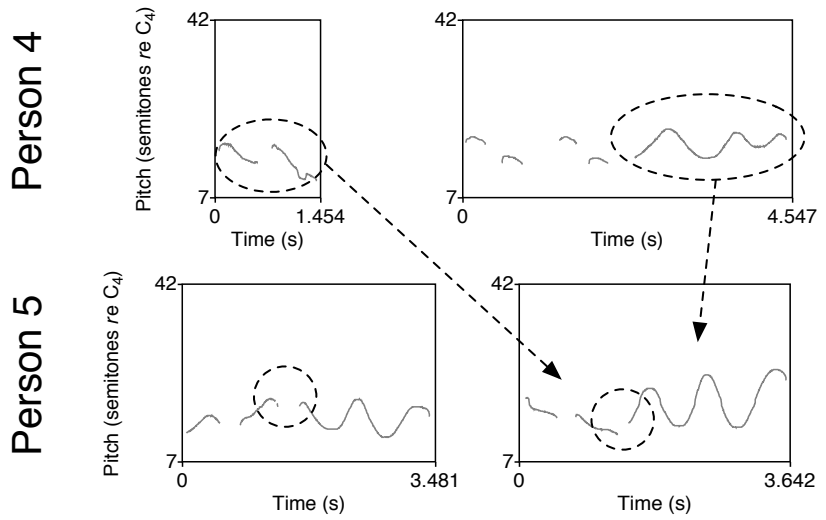


Figure 4.3: *An example of recombination in chain four: a whistle from the previous generation is combined with the second part of another whistle and a second version is added with a mirrored part. Note the co-articulation-like effect highlighted with circles: the final pitch of the first part influences the initial pitch of the second part.*

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.

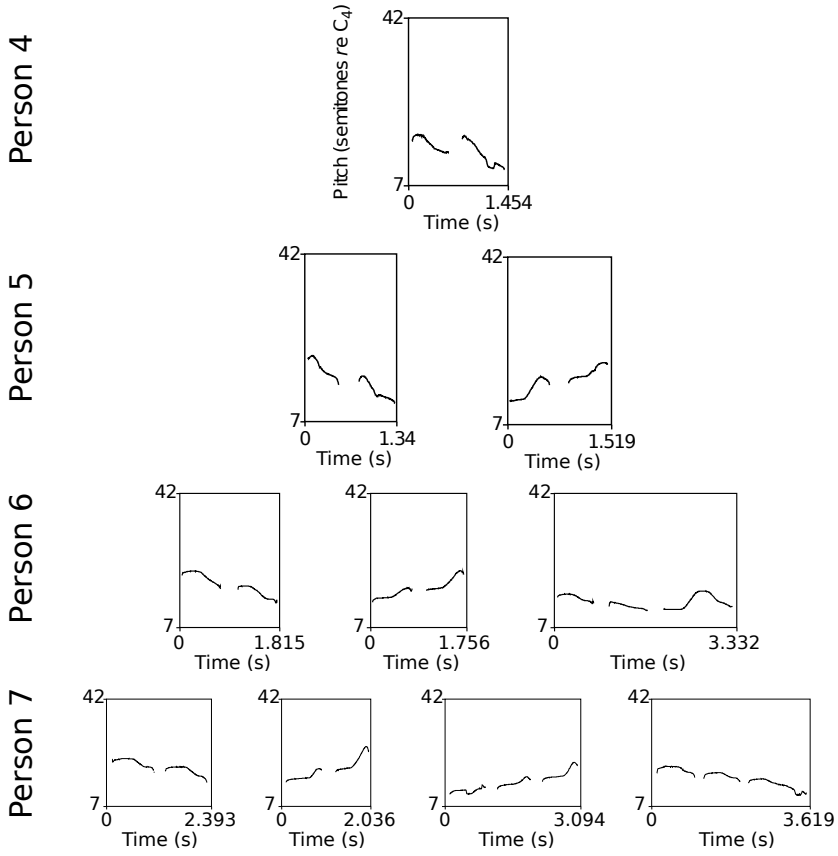


Figure 4.4: *An example of cumulative mirroring, repetition and borrowing. Person 5 mirrors the whistle from the previous set, then person six borrows one of the two in a new whistle and finally this new whistle becomes generalised to fit the pattern of the original two, but repeated. This predictable system stays stable until the end of the chain. The whistles are plotted as pitch tracks on a semitone scale.*

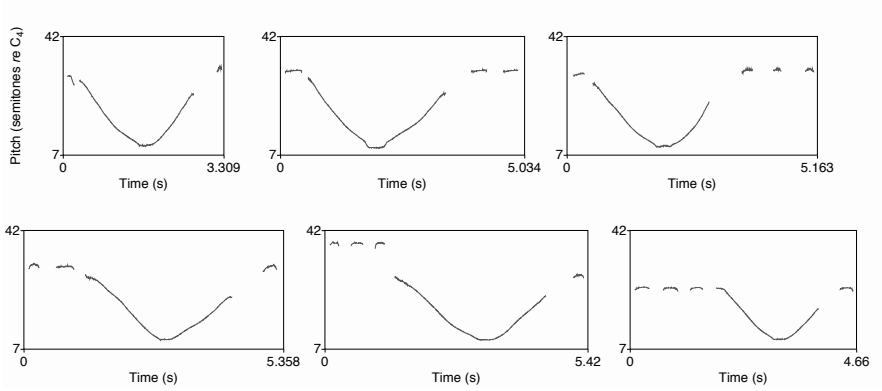


Figure 4.5: *Fragment of the whistles plotted as pitch tracks in the last set of a chain. Basic elements can be identified that are systematically recombined.*

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.

$$l = \frac{c}{4f} \tag{4.1}$$

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.

$$H = -\sum p_i \log p_i \tag{4.2}$$

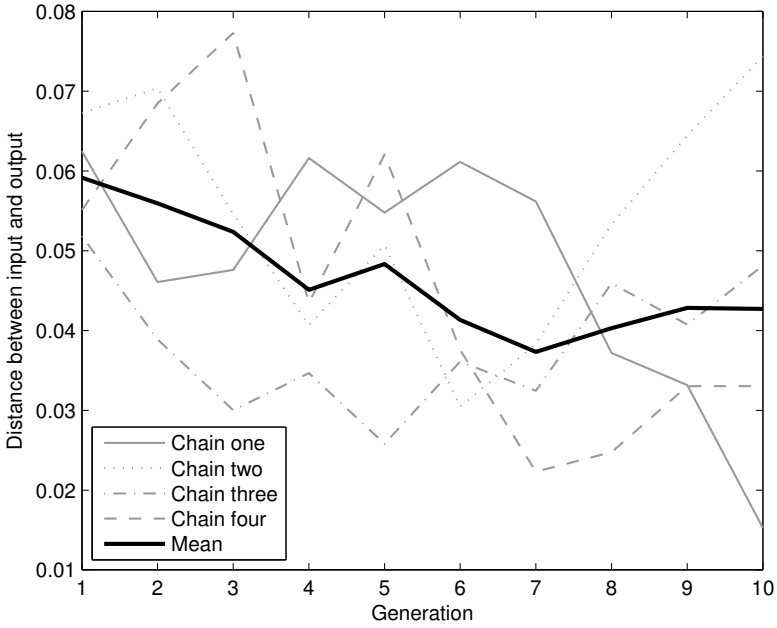


Figure 4.6: *Recall error on the whistle sets over generations for all four chains, demonstrating that the whistle systems evolve through cultural transmission and become more learnable.*

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

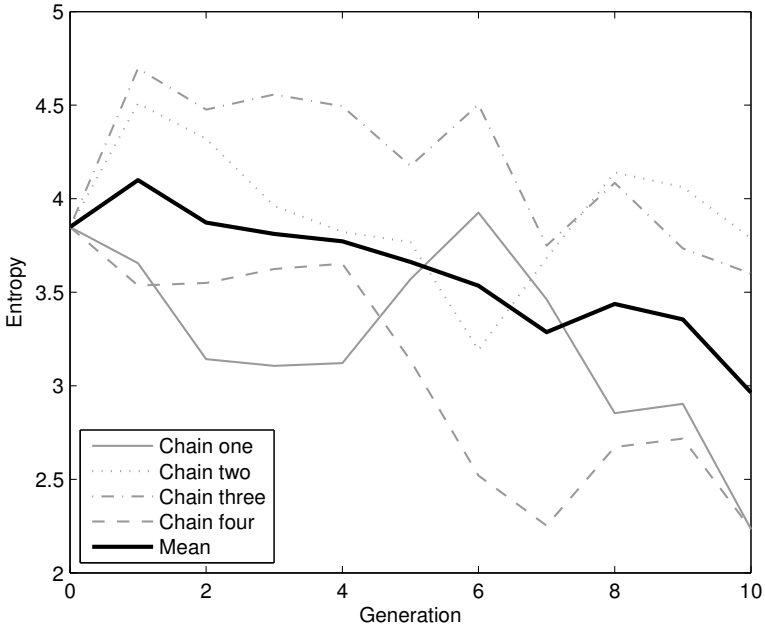


Figure 4.7: *Entropy of the whistle sets over generations for all four chains, demonstrating that the combinatorial structure increases.*

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,

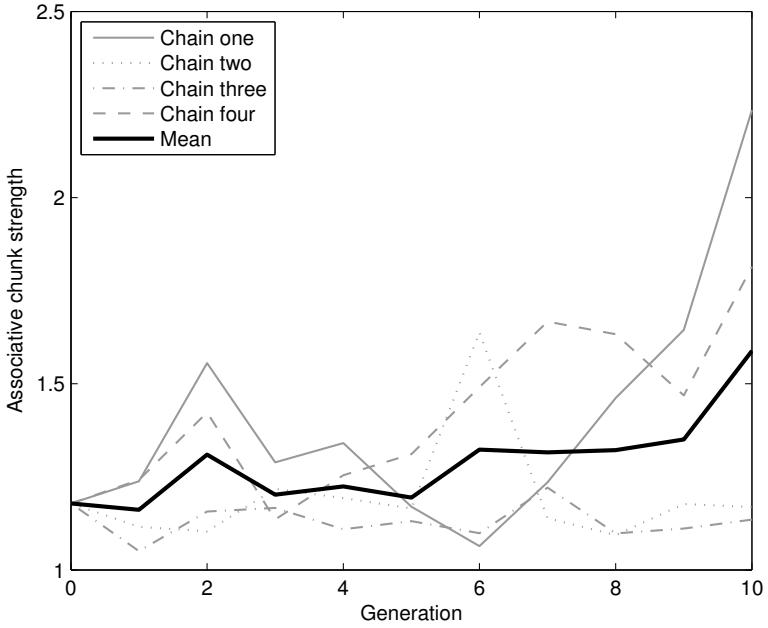


Figure 4.8: *Associative chunk strength of the whistle sets over generations for all four chains, showing an increase in reoccurrence of bigram and trigram sequences of basic whistle patterns.*

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.

$$E = \sum_{i=1}^{n-1} \sum_{j=0}^{i-1} \frac{1}{r_{ij}^2} \tag{4.3}$$

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

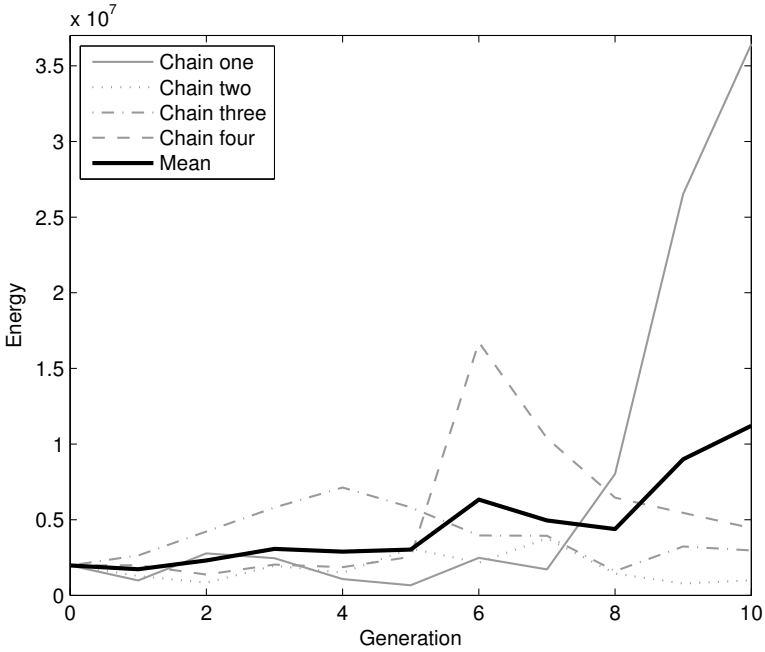


Figure 4.9: *Dispersion measured as energy between whistles in the set for each generation. The whistles do not tend to become more dispersed (no decrease in energy) towards the end of the chains. On the contrary, for at least one of the chains there appears to be an increase of energy.*

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

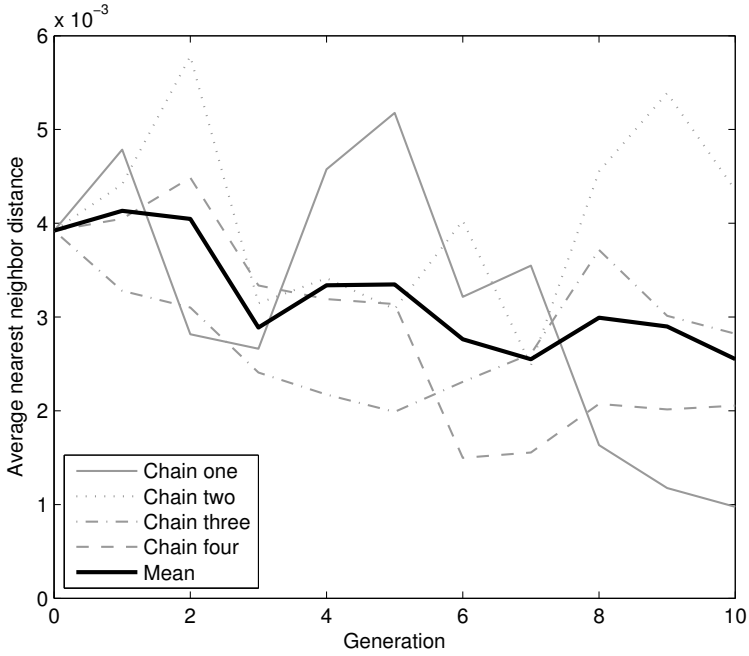


Figure 4.10: *Average nearest neighbour distance between the twelve whistles of the set in each generation. The whistles tend to become more similar towards the end of the chains.*

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

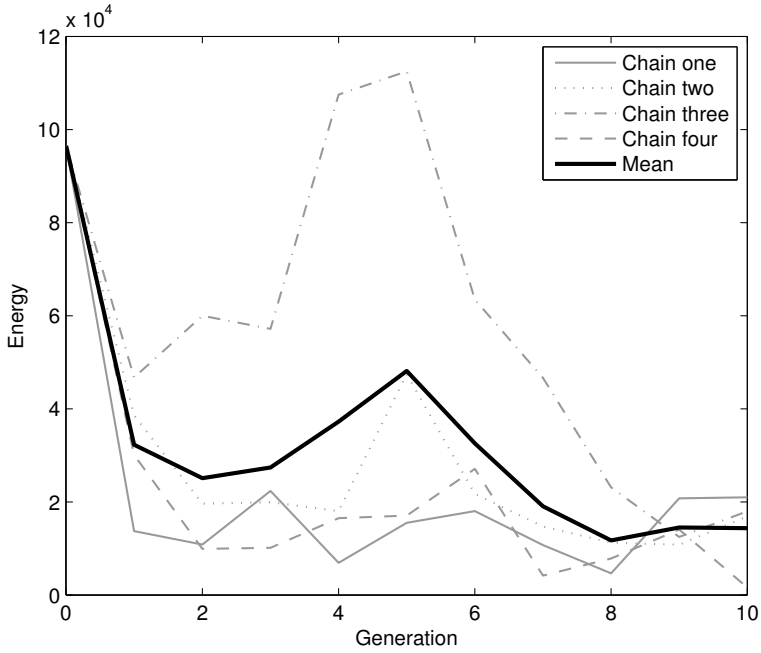


Figure 4.11: *Dispersion measured as energy between building blocks in each generation. The building blocks tend to become more dispersed (lower energy) towards the end of the chains.*

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-

ficially inserted initial set. These results imply that dispersion theories have a role to play in explaining empirical observations, but that they are not sufficient to capture the complexity in its entirety. Theories that take into account principles of economy also need to be considered.

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

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.

Games

5

In chapter 4, we could see how artificial whistled languages that are culturally transmitted in the laboratory gradually became easier to learn and more structured. The whistled languages were analysed with computational measures and it was shown that combinatorial structure increased over generations of learning and reproduction. To analyse this structure and its relation to learnability further, additional experiments were conducted and the results are described in this chapter.

Zuidema and de Boer (2009) introduced a distinction between two kinds of combinatorial structure that can be identified when studying systems of signals. The first kind is what they call *superficial combinatorial structure* and this refers to combinatorial structure that can be identified when a system is analysed by an outside observer, but the users of the system do not necessarily cognitively encode this structure. The second kind is called *productive combinatorial structure* and this refers to the structure that users of the system do encode and actively use in production, perception and learning. The results that were presented in the previous section show both qualitatively and quantitatively that a system of auditory signals gains (superficial) combinatorial structure and becomes more learnable when it is transmitted culturally. What has not been shown quantitatively yet is whether people who have to learn these emerged artificial languages, are able to actively use the combinatorial structure in a way that Zuidema and de Boer (2009) would call productive. Note that their definition does not require signal production before a system can be considered to have productive combinatorial structure. It involves the ability to make use of the structure in production as well as perception and learning. Given the combination of qualitative and quantitative results that were obtained in the previous chapter, the expectation is that the observed structure is not only observable by careful analysis, but to prove it can be used

The first experiment described in this chapter was previously described in:
Verhoef, T.(2012) The origins of duality of patterning in artificial whistled languages. *Language and Cognition*, 4(4), 357-380.

The second experiment was conducted as part of Science Live, the innovative research programme of Science Center NEMO that enables scientists to carry out real, publishable, peer-reviewed research using NEMO visitors as volunteers.

productively, this needs to be tested with new learners. In addition, the fact that an increase in learnability of the system was measured does not necessarily mean that it has become more learnable because of the increased structure and cognitive ease that comes with it. An alternative explanation may be that only the individual whistles have evolved to become easier to imitate and that therefore only articulatory constraints made the set more reproducible. The experiments described in this chapter were conducted to test the productive use of the emerged combinatorial structure in the whistled languages from chapter 4.

5.1 Perceptual category learning game

To test the possibilities for human productive use of the structure that seems to be present in the emerged whistle sets, and to identify whether cognitive constraints may indeed have been involved in shaping these sets, a separate experiment was conducted. In this experiment, the stimuli that were used came from the sets of signals from the last generation of chains one and four in the whistle experiment described in chapter 4. The aim of the current experiment is to test if human participants, who are exposed to a few examples of such an emergent whistle language, can decide for other examples if they belong to the set or not. For the design of this experiment I used a paradigm that was developed by Jelle Zuidema and Vanessa Ferdinand in which participants play a UFO game[1].

The task in this game can be compared with concept learning experiments from the field of cognitive psychology (Goodman et al., 2008). The methods for studying concept learning are popular as a means of unravelling the way generalisation and representation works in human cognition. Typically, participants have to learn to categorise or distinguish between several different concepts. They are first trained on a subset of examples, and then tested on a larger set to see whether they were able to learn the underlying category structure. In the experiments described in this chapter, the task is essentially the same, but it is presented in the context of a game in which participants need to learn to distinguish between two different types of aliens. The game environment makes the experiment more engaging and this was important since the participants were recruited on a voluntary basis both online and inside a science museum.

In the UFO game, two species of aliens exist: good aliens and bad aliens. The player's goal is to save the good aliens and kill the bad ones. The only way to distinguish a good alien from a bad one is to listen to their language. A screenshot of the game is shown in figure 5.1. First, there is a familiarisation phase. In this phase, UFO's keep flying by on the screen until the player catches one by clicking on it. When a UFO is

[1] The UFO game that was used in the experiments described in this chapter was created by Jelle Zuidema and Vanessa Ferdinand (http://www.webexperiment.nl/)

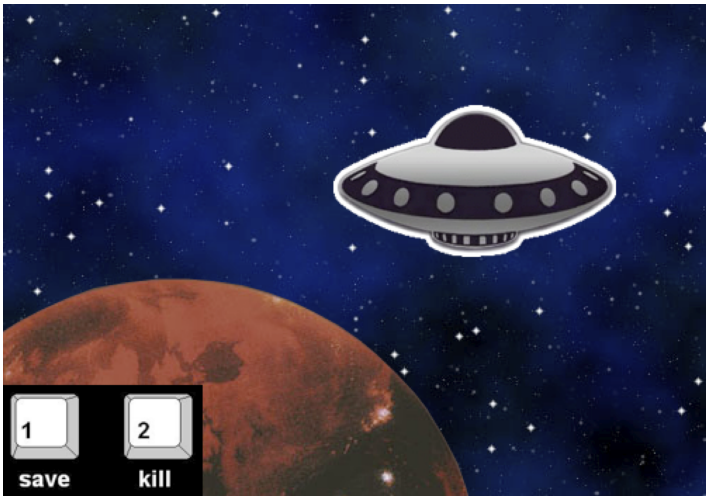


Figure 5.1: *Screenshot of the UFO game.*

caught, the alien inside makes a sound. In this phase participants are exposed to the language of the good aliens only and they practice to save the spaceships of these aliens. Participants are therefore asked to pay attention to the sounds these good aliens make and to press the 'save' button for all of them. In this familiarisation phase, half of the sounds from the language of the good aliens are played five times each. The next phase is the combat training, in which participants practice shooting UFO's. A few empty spaceships fly by and participants are asked to catch them and press the 'kill' button. This phase is only six items long, and no sounds are played. Finally, in the combat phase UFO's fly by again and when participants catch them, they have to listen to the sounds the aliens make, decide whether they are good or bad and kill or save them accordingly. This phase has 72 items, in which from both the good and bad alien languages, each whistle is played three times. Last, they see their final score.

5.1.1 Methods

Two conditions were created, differing in which individual whistle sounds from the two emergent languages were part of each alien species' language. In the 'intact' condition, each of the two alien species' languages consisted of a complete emergent whistle language. This means that one alien species had a vocabulary consisting of all twelve sounds produced by the last person in chain one (of the iterated learning experiment described in chapter 4) and the other alien species used those from the last person in chain four. In the 'mixed' condition, each alien species had six sounds in their language from the last person in chain one and six sounds from the last person in chain four, breaking up the emergent whistle languages from the iterated learning

experiment. This is illustrated schematically in figure 5.2. I selected the languages of chains one and four because, as can be seen in figure 4.7 of chapter 4, these were the two chains that resulted in emergent languages exhibiting the most combinatorial structure and their measured amount of structure was very similar. In the intact condition I alternated whether the good aliens used sounds from chain one or four. In the mixed condition, I used two different ways of breaking up the languages from the two chains. In both pairs of mixed languages, six sounds from each chain were randomly assigned to the language of the good aliens, and the other six of each to the language of the bad ones.

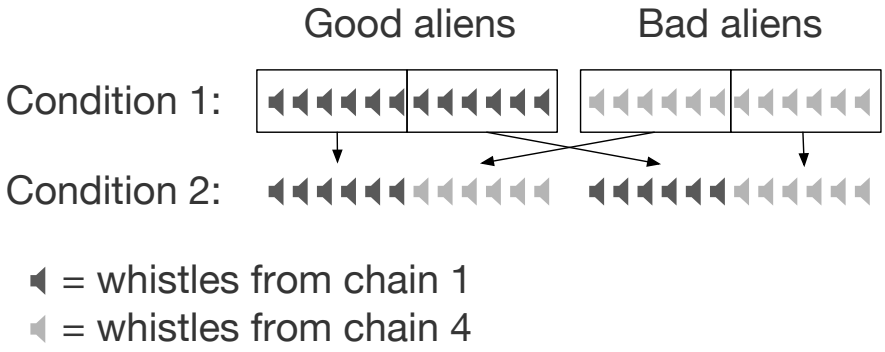


Figure 5.2: *Two experimental conditions: (1) the 'intact' condition, where each of the two alien species languages consisted of an intact emergent whistle set from the last generation of chain one and chain four of the experiment described in chapter 4. (2) the 'mixed' condition, where mixing sounds from both sets created the two languages.*

The aim of this design was to investigate whether participants generalise and use the combinatorial structure in the emerged whistle languages to classify new aliens as good or bad and save or kill them accordingly. In the familiarisation phase, participants are exposed to six out of the twelve sounds that the good aliens use. In the mixed condition, they are exposed to three sounds originating from each of the two chains. In the combat phase they are tested on all sounds of both species, including the ones of the good aliens they had never heard before. If the participants can learn the potential structure in the sounds and use it productively, they should perform better on the whistles they never heard before in the intact condition. The mixed condition, where the two emergent languages are broken up, should give participants much less evidence about potential rules, building blocks or constraints in the languages to generalise from. In the first condition, if structure is present in the emergent languages from the iterated learning experiment, participants should be able to generalise and classify the identity of UFO's with an accuracy above the baseline of random guessing.

This first experiment was conducted as an online game for which participants were recruited through Facebook. Ten participants completed the game in the mixed condition and eleven in the intact condition. Their ages ranged from 22 to 50 (mean age of 29). There were twelve male participants and six of them participated in each condition.

5.1.2 Results

To analyse the results, for each participant it was determined how well they could classify sounds that they had never heard in the familiarisation phase correctly as belonging to good or bad aliens. In total there were 54 new items in the combat phase (twelve sounds from the bad aliens and 6 from the good aliens that were never heard before, each appearing 3 times). As a measure of performance the discriminability index d' was used. This measure takes the individual response bias towards shooting or saving UFO's into account and is computed with the use of equation 5.1, where $z(H)$ and $z(F)$ are the z-transforms of the hit rate (H) and false alarm rate (F).

$$d' = z(H) - z(F) \tag{5.1}$$

The results are shown in table 5.1. In the intact condition, the median d' score was 2.585 and in the mixed condition it was -0.563. There is a significant difference between the distributions of the two groups (Mann-Whitney U = 110, n1 = 11, n2 = 10, $P < 0.001$). The expected baseline score measured as the number of items correctly classified in the case of random guessing would be 27 (54 x 0.5). In the intact condition, the median score of correct classification was 47, well above the baseline, and in the mixed condition it was 23.5, slightly below the baseline. There is a significant difference between the distributions of the scores in the two groups (Mann-Whitney U = 55, n1 = 11, n2 = 10, $P < 0.001$). These first results suggest that participants were able to learn the structure that was present in the emerged whistle sets. They could generalise from a few examples and make accurate predictions about group membership of sounds they had not been exposed to.

	Condition	
	Intact	Mixed
Median d'	2.585	-0.563
Median score	47	23.5

Table 5.1: *Results of the UFO experiment. There is a significant difference between the scores in the two conditions, measured as d' and as the median score of correct classification.*

5.2 Follow-up experiment

A potential problem with the experiment described above could be the fact that the two languages that participants have to distinguish are also produced by two different people. It could therefore be the case that the participants are picking up on differences in individual whistling style or characteristics. To make sure that this is not (only) playing a role in the results described above, another version of the experiment was conducted. In this version, the sounds were taken from the last two generations of each chain and they were first re-synthesised from extracted pitch tracks, so that the new alien words for each species contained whistles created by more than one person and the whistles retained only information about the plunger displacement and timing. In addition, the experiment was expanded by including all four chains. In the first pilot only the two chains that, according to our analysis, contained the most combinatorial structure were used. The aim of this follow-up experiment was to assess whether the initial results can be replicated with re-synthesised sounds for all four chains.

5.2.1 Methods

To prepare the sounds for the implementation of this follow-up experiment, all four chains that had emerged in the experiment described in chapter 4 were used. From each chain the set that was produced by the very last participant (generation ten) was taken, as in the first UFO game experiment, but this time half of the whistle recordings were replaced by the version of that same whistle that was produced by the preceding participant in the chain. By the last generation the sets were reproduced well enough for it to be straightforward to find a matching production for half of the whistles for each of the four chains. These new sets of twelve whistles were then preprocessed with Praat (Boersma, 2001) to re-synthesise the whistle recordings. This was done by extracting the pitch from the Sound object and then using the function ‘To Sound (sine)...’ to create a new Sound object. These re-synthesised sounds were used in the design of the UFO game.

The design of the UFO games were largely the same as for the first UFO game experiment. Six different versions were created in which the emerged whistled languages from chain one and four (from the experiment in chapter 4) were used in half of these. There was one version in which the good aliens spoke the language from chain one and the bad aliens the one from chain four, another version in which this was reversed and the third version was the mixed condition, in which whistles from both languages were used for both alien species. In the same way, the other three versions were constructed with emerged languages from chain two and three. The number of items in the familiarisation (30), practice (6) and combat (72) phases were the same as in the first game

design. There was one version with instructions in Dutch and one version in English. This was the case because the experiments were conducted inside a museum. It was part of a project, Science Live, that was carried out in collaboration with Science Center NEMO in Amsterdam. This project enables scientists to carry out real research using NEMO visitors as volunteers. These visitors are mostly Dutch, but many foreign tourists visit the museum as well. The experiment was again implemented as an online applet, but this time all participants completed it on a desktop computer, wearing Sennheiser HD202 headphones, in the designated Science Live space of the museum.

In total, 72 visitors completed the game in this experiment. Their ages ranged from 8 to 64 (mean age of 23) and 37.5 % were female. For five participants, recorded data had to be excluded from the analysis because the testing conditions were not always ideal in the museum. Sometimes it happened that other family members or friends would interrupt the participant during the game. Especially young children sometimes clearly got distracted by parents, brothers or sisters. In addition, some very young children wanted to participate only if they could 'do it together' with their parent, which was of course allowed in this setting, but then the data was excluded. These issues were all written down and linked to participant numbers on the testing days, so that they could easily be identified and excluded during the analysis.

5.2.2 Results

The analysis was carried out in the same way as for the first UFO game experiment. For each participant it was determined how well they were able to classify the new sounds as belonging to good or bad aliens correctly. In total there were again 54 new items in the combat phase. The results are shown in table 5.2. For chain one and four, the median d' score was 1.499 in the intact condition and -0.411 in the mixed condition. There is a significant difference between the distributions of d' in the two groups (Mann-Whitney $U = 132$, n1 = 12, n2 = 11, $P < 0.001$). For chain two and three, the median d' score was 0.443 in the intact condition and -0.443 in the mixed condition. There is also a significant difference between the distributions of d' in these two groups (Mann-Whitney $U = 45.5$, n1 = 25, n2 = 19, $P < 0.001$).

For chain one and four, the median score of correct classification was 41 in the intact condition and 24 in the mixed condition. There is a significant difference between the distributions of the scores in the two groups (Mann-Whitney $U = 132$, n1 = 12, n2 = 11, $P < 0.001$). For chain two and three, the median score of correct classification was 31 in the intact condition and 25 in the mixed condition. There is also a significant difference between the distributions of the scores in these two groups (Mann-Whitney $U = 58.5$, n1 = 25, n2 = 19, $P < 0.001$). These results suggest that also for the other two chains from the whistle experiment, participants were able to learn the structure that was present

in the emerged whistle sets. Moreover, this was not due to differences in individual whistle styles or characteristics.

	Condition			
	Chain 1&4		Chain 2&3	
	Intact	Mixed	Intact	Mixed
Median d'	1.499	-0.411	0.443	-0.443
Median score	41	24	31	25

Table 5.2: *Results of the follow-up UFO experiment. For both pairs of chains there are significant differences between the scores in the two conditions, measured as d' and as the median score of correct classification.*

5.3 Discussion

Two experiments have been presented in this chapter, both providing additional steps of analysis on the emerged whistle sets from the experiment described in chapter 4. In that chapter the presence of combinatorial structure was qualitatively determined by inspecting the whistles produced in the final generation as well as quantitatively by measuring a decrease of entropy over generations. These measures only captured structures as an outside observer, but did not take the productive use of the learner of such structures into account. The results of the experiments presented in this chapter demonstrate that the observed combinatorial structure can be helpful when participants are asked to identify whistles from different languages. In the first experiment, only the two emerged languages that seemed the most structured were used, and the whistles were presented to the UFO game players unaltered. In the second experiment, all emerged languages from chapter 4 were used and the whistle sets were altered in such a way as to make sure the whistles from one language were produced by different people and were re-synthesised to remove most individual characteristics.

Of course, the presence of combinatorial structure can not be directly inferred solely from the fact that participants were better able to distinguish the two languages in the intact conditions than in the mixed conditions. It is easy to imagine a situation in which the two languages both contain only one unique sound for each of the twelve words. In this case the task would be impossible in the mixed condition and very easy in the intact condition, without this being caused by any (interesting) combinatorial structure. However, combined with the analyses that were presented in chapter 4 and the fact that the whistles in the two languages were produced by participants in an experiment in which repeating the same whistle was prevented, the UFO game experiments provide strong evidence suggesting that ‘language specific’ constraints and regularities are present in the emerged whistled sets.

By means of this perceptual category learning game it is therefore shown that there is structural evidence available in the emergent whistle languages and learners use it to distinguish between distinct languages. Following the definitions proposed by Zuidema and de Boer (2009), the observed combinatorial structure could be concluded to be of the productive type. Human participants are are able to learn the regularities that emerged through experimental cultural transmission and they use it in perception and recognition.

Meanings

6

In chapter 4 we saw that efficient coding and combinatorial structure can emerge in a system of sounds that is culturally transmitted in the laboratory. Those results demonstrated a possible route towards the emergence of combinatorial structure in the sounds of speech. As discussed in chapter 4, the findings from this experiment challenge the hypothesis that Hockett (1960) introduced when he linked the emergence of combinatorial structure to vocabulary expansion and signal dispersal. Even in the case where only a small set of sounds is transmitted and the signal space does not become maximally used, combinatorial structure emerges in the experiment. The influence of semantics, compositional syntax or iconicity was controlled for, as the signals did not refer to any concrete meanings. In this manner, the emergence of combinatorial sound categories as an independent system could be studied. Obviously, in natural human languages meanings are important and the role of semantics in the evolution of linguistic structure should not be ignored (Schouwstra, 2012). Would the introduction of semantics influence the emergence of combinatorial structure at the level of phonology? In this chapter an experiment is presented in which, as in chapter 4, artificial whistled languages are culturally transmitted, but this time the whistled signals refer to meanings. As we will see, combinatorial structure emerges also in the case that semantic referentiality is present.

6.1 Combinatorial structure versus iconicity

Like 'duality of patterning', the design feature of language that is central to this thesis, another feature, 'arbitrariness', was listed by Hockett (1960) as essential to natural human language. This feature refers to the arbitrary/unmotivated mapping between words and their meanings. Hockett uses the words 'whale' and 'microorganism' as an example: 'whale' is a short word for a large animal, while 'microorganism' is the reverse. It has been argued that non-arbitrariness is rare in modern languages and that it is irrelevant for understanding linguistic structure (Newmeyer, 1992). More recently, however, researchers began to realise that non-arbitrary form-meaning mappings may be more widespread

than initially thought, both at the level of the word and at the level of the sentence structure (Perniss et al., 2010). When exploring beyond Indo-European languages, non-arbitrary form-meaning mappings seem to play a large role in many languages (Dingemanse, 2012; Imai et al., 2008; Perniss et al., 2010). This involves classes of words where for instance the shape, complexity, sound or some other characteristic of the meaning expressed is mimicked or iconically represented in the word. Examples have been identified as 'ideophones', 'mimetics' or 'expressives' and the phenomenon is often called sound-symbolism (Imai et al., 2008). Sound-symbolic mappings can take different forms. As Cuskley and Kirby (2013) describe, *conventional sound symbolism* refers to the statistical correspondences between certain clusters of similar forms and meaning classes, where sub-lexical elements are systematically used for a certain semantic domain. *Sensory sound symbolism* describes words that phonetically imitate the sound their referent makes, such as 'bang' or 'buzz' (which are called 'onomatopoeia'), or words that cross-modally imitate other characteristics of the referent, for instance based on vision, temporal structure, touch, taste, smell or other domains (Cuskley and Kirby, 2013; Dingemanse, 2011). Modern English may only have very little sensory sound symbolism but it is no longer considered to exclusively have arbitrary form-meaning mappings either, because conventional sound symbolism does occur often. Form-meaning pairings can be identified that reoccur with strikingly high frequencies, like words starting with *sn-*, that often refer to concepts that relate to the nose or mouth (snore, snack, snout, snarl, snort, sniff, sneeze, etc) (Bergen, 2004).

It has been shown that in the context of a lexical decision task non-arbitrary form-meaning pairs are processed faster than arbitrary form-meaning pairs (Bergen, 2004) and that sound-symbolic mappings help young children in acquiring new words (Imai et al., 2008). Moreover, it has been found that parents use sound-symbolic words in their infant-directed speech more often than in adult-to-adult conversations (Imai et al., 2008). These examples are among others that support the idea that there may be processing and acquisition benefits for iconic mappings in both spoken and signed languages (Perniss et al., 2010). Perhaps iconicity helps learners to ground linguistic expressions in sensory perception, although there are counterexamples as well. Some studies bring the presumed cognitive ease of iconic mappings into question, for instance by showing that very young children have more difficulty interpreting these (Tolar et al., 2008). Sound-symbolic mappings in language have been connected to cross-modal mappings in the human brain (Ramachandran and Hubbard, 2001; Simner et al., 2010). There appear to be many cognitive biases in cross-modal perception that are shared by humans. The bouba/kiki effect is one famous example that shows a strong preference to relate sharp shapes to the name 'kiki' (or 'takete') and round shapes to the name 'bouba' (or 'baluma') (Ramachandran and Hubbard, 2001). Many mappings have

been investigated and identified, especially in the visual-auditory domain (Hubbard, 1996; Ward et al., 2006), but also for instance relating taste to speech sounds (Simner et al., 2010). Such shared biases have been argued to play an important role in the evolution of language, by forming a starting point for the initial emergence of grounded speech (Ramachandran and Hubbard, 2001). Under the assumption that cultural transmission drives languages to become more learnable over time and with the presumed cognitive ease of processing iconic mappings, we may expect that iconicity would be preserved or even expanded in language evolution over time. This is, however, not what is usually reported. More often languages are assumed to develop towards more arbitrariness, where systematicity competes with iconicity (Goldin-Meadow et al., 1995; Theisen et al., 2010). Together, these issues illustrate the need for a more detailed investigation into the role of iconicity in language evolution.

Returning to the case of Al-Sayyid Bedouin Sign Language (ABSL), as discussed in chapter 2, this is an example of a fully functional, expressive sign language which lacks the clear discrete and combinatorial phonology that other languages have (Sandler et al., 2011). Could it be the case that this young sign language was able to survive up to now without duality of patterning because the manual modality allows for a large degree of iconicity and the language is learnable and transmissible even with limited phonological structure? When a system can support a large amount of transparent, holistic mappings, perhaps there is less need for combinatorial structure at the sub-lexical level (Sandler et al., 2011). On the other hand, it has been shown that there is actually an advantage for arbitrary mappings in acquiring word meanings in context (Monaghan et al., 2011). A secondary objective of the experiment described below is to investigate how iconic form-meaning mappings influence the emergence of combinatorial sub-lexical structure. Two conditions were studied: one in which the use of iconic form-meaning mappings is possible and one in which the use of iconic form-meaning mappings is experimentally made impossible. This is expected to provide insights into the possible role of iconicity in the emergence of duality of patterning since it may reveal whether a situation that allows for more iconicity, can 'survive' longer without the emergence of combinatorial structure. In the domain of iterated learning experiments with graphical systems, conflicting results have been found so far. del Giudice et al. (2010) studied systems in which graphical signals[1] were transmitted in iterated learning chains and they observed the emergence of combinatorial structure and a reduction of iconic forms over generations. On the other hand, Garrod et al. (2010) used graphical

[1]Most of the experiments conducted by del Giudice et al. (2012; 2010) made use of the graphical signalling device that was designed by Galantucci (2005) and included a transformation of the actual drawing, making iconic mappings less straightforward. However in this comparison I refer to a specific condition in which del Giudice did not use this device, but the actual drawings themselves were transmitted. This was therefore very similar to the conditions in the study by Garrod et al. (2010).

systems as well, but here the forms remained iconic and complex in the iterated learning chains.

In summary, the objective of this study is as follows. First and foremost, it is investigated whether the addition of meanings leads to a result that is similar to what was found in the whistle experiment without meanings, to see if combinatorial structure also emerges in the presence of semantics. Second, differences between the two conditions are investigated to see whether iconicity could cause a delay in the emergence of structure.

6.2 Methods

In this experiment participants are asked to learn and reproduce whistled signals with a slide whistle as labels for objects they see on a computer screen. As in the first whistle experiment, there were twelve whistled signals in the training set in total. The meanings in this study are part of a set of unusual objects that look like possible mechanical parts, but they are novel objects for which there are no conventional names in existing languages. The objects were selected as a subset of those created by Smith et al. (2011) and were slightly modified. To make sure that the meanings are not easy to categorise, all objects are in blue tone (transformed with a blue filter) and can therefore not be grouped by their colour. They also do not share shapes or parts and are not structured in any other obvious way[2]. Since this experiment attempts to investigate the emergence of sub-lexical combinatorial structure, the recombination of meaningless sounds into words, a meaning space with minimal structure is desirable. Any possible categorisations in the meaning space could cause semantics-related compositional structure to emerge, which would make our results harder to analyse. A few examples of objects that were used are shown in figure 6.1.

The last whistle sounds that a participant produced for each object were used as the words for those objects in the input given to the next participant. However, this is the point where the two conditions differ from each other. In one condition, the ‘intact’ transmission, the next participant is exposed to the output of the previous participant exactly as it was produced. The mapping from whistled signals to objects is kept intact. In the other condition, the ‘scrambled’ transmission, the output of the previous participant is altered before it is given to the next person. The produced form-meaning mappings are broken down by scrambling the mappings at each change of generation and by using a different set of objects between consecutive generations. In this way, if any iconic relations were to emerge in the sets, they would only be helpful for the participants in the first condition. For the second condition, any

[2]The meanings themselves have structure in the sense that they are complex objects with sometimes many different parts, but what is meant here is that there is no systematic structure between the items in the set, making it difficult to identify similarities or group items in the set into categories.

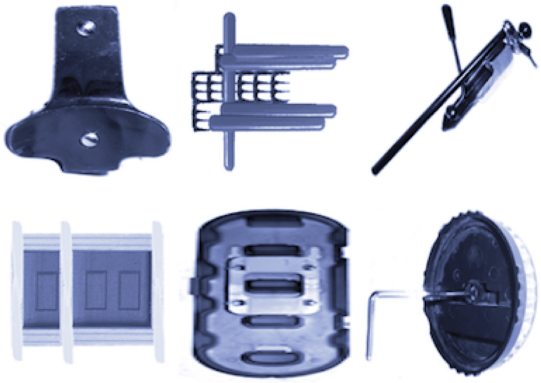
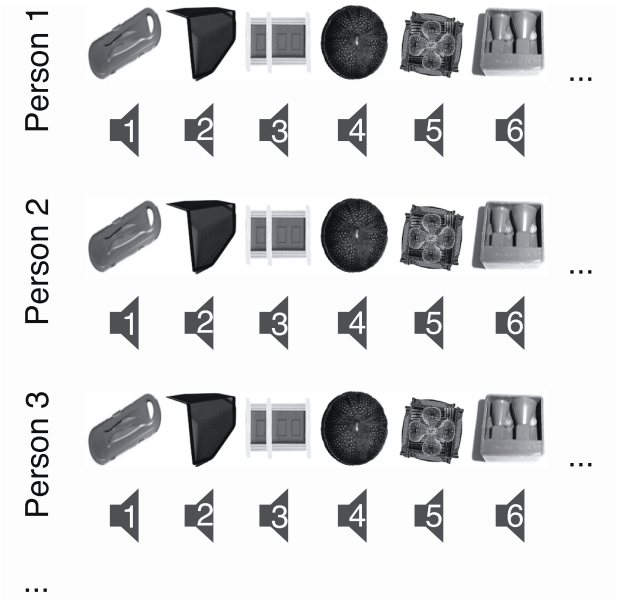


Figure 6.1: *Examples of novel objects used in the experiment. These objects were created by Smith et al. (2011) and were slightly modified. To reduce potential categorisation according to colours in the meaning space, all objects are in blue tone (transformed with a blue filter).*

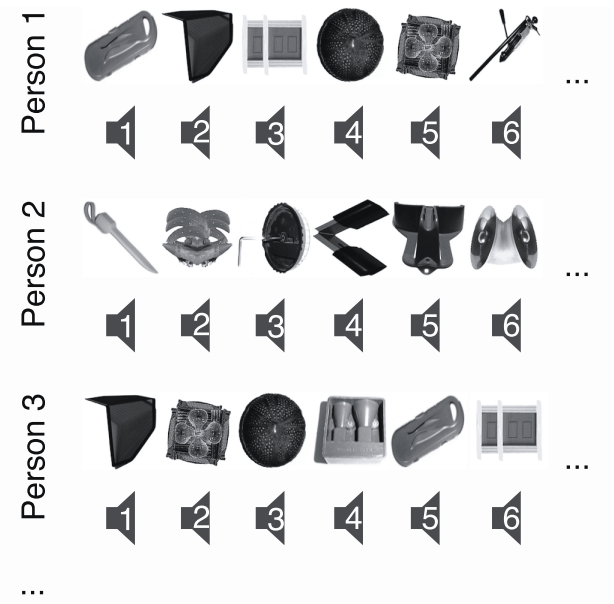
semantics-related structure is broken down in between the transmission steps. Only the signal sets themselves stay intact. Figure 6.2 shows a visual representation that explains the two conditions.

6.2.1 Procedure

Before the start of the experiment participants read a story to make the task more engaging. They were told that an alien space ship had crashed on earth and that the aliens need their help to repair their ship. To be able to help the friendly extraterrestrials, participants need to learn twelve words for alien space ship parts. The best way to imitate the sounds these aliens make is to use a slide whistle. Instructions on the task were given both in spoken and written form and there was time for participants to ask questions in case anything was not yet clear. The written instructions can be found in appendix C.1. Before the actual experiment started participants signed an informed consent form and completed a background questionnaire. After this, they were given some time to practice using the slide whistle. During the experiment they completed three rounds of learning and recall. The first two learning phases were followed by a ‘guessing game’ before the recall phase. In the learning phase the objects and their corresponding whistle were presented one by one in a random order, and participants recorded an imitation of the whistle. In the recall phase a panel was shown with a button for each object and the participant had to choose each of the objects once to record the right whistle for it from memory. The guessing phases were introduced in this version of the experiment to encourage people to keep paying attention to the mapping between whistle sounds to objects. In this guessing phase the whistles were played one by one in a random order and for each whistle the participant had to choose the



a The condition with intact form-meaning mappings



b The condition with disjoint object sets and scrambled form-meaning mappings

Figure 6.2: ***a****: The next person in a chain was exposed to the exact pairs of whistles and objects that the previous person created.* ***b****: The next person in a chain was exposed to the exact set of whistles that the previous person created but from one person to the other the set of objects was replaced and the whistles were randomly paired with the objects. Two sets of 12 objects were alternated and each was used every other generation so that the odd-numbered generations saw one set, and the even-numbered generations the other set.*

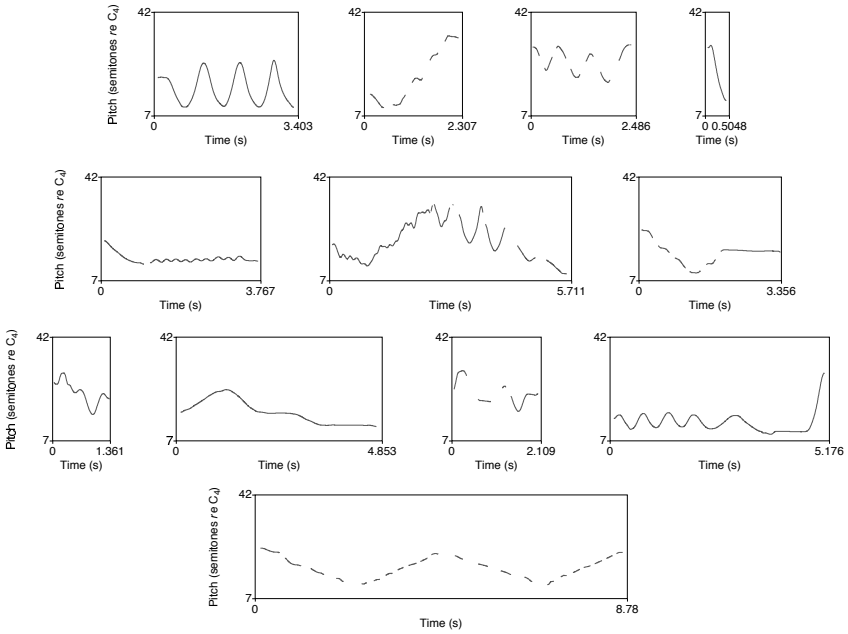
right object from a panel. This was done with half of the whistle-object pairs after the first learning phase and with the other half after the second. After the last recall phase participants were asked to complete a post-participation questionnaire and there was a debriefing. The whistles from the last recall phase were used as training input for the next participant, depending on the condition either with intact whistle-object mappings or scrambled and with other objects. Transmission was continued from person to person until there were eight generations in each chain and four chains per condition. The entire procedure took place inside a sound-proof booth and it took approximately 60 minutes in total. In Appendix C.2 a screenshot of the user interface that was created for this experiment is shown.

6.2.2 Initial input sets

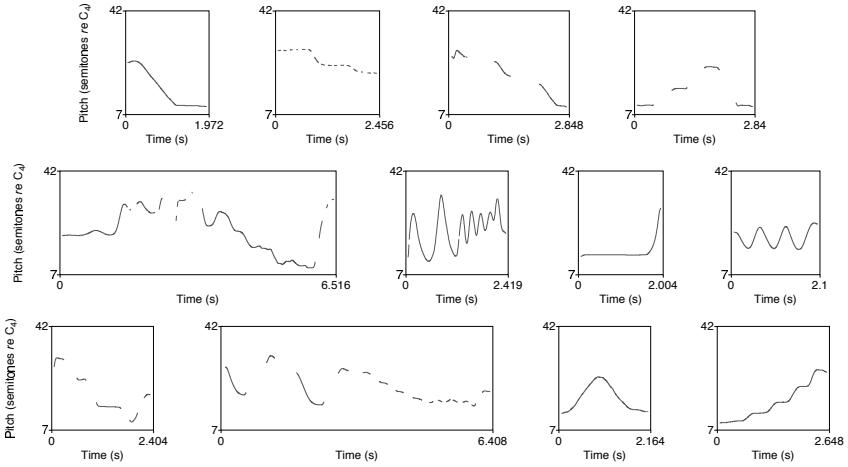
For this experiment two separate initial whistle sets were constructed. Each set was used as the starting point for half of the chains in each condition. The whistles were taken from the database of whistles that were collected during the pilot preceding the original whistle experiment described in chapter 4. During this pilot, whistle sounds were created by people who were asked to freely record a number of whistle sounds and a database was constructed from these recordings. The two initial sets were constructed so as not to exhibit combinatorial structure. To achieve this, the entropy measure for quantifying combinatorial structure from the original whistle experiment was used. Sets of twelve whistles were generated randomly from the database until two sets were found with no overlap, which had a comparable and relatively high measured entropy (4.18 and 4.28). Figure 6.3 shows the two sets of twelve whistles plotted as pitch tracks on a semitone scale using Praat (Boersma, 2001).

6.2.3 Reproduction constraint

As was described in chapters 3 and 4, experiments that involve iterated learning without a pressure for expressivity tend to result in systems of signals with under-specification. We have seen in chapter 3 that this problem is not resolved when the signals are continuous and less likely to be exactly the same as each other. Therefore, a reproduction constraint was used in this experiment as well. The constraint was very similar to the one that was used in the original whistle experiment. When a participant produced a whistle for an object that was too similar to another whistle that had already been produced for another object, the program told the participant that this whistle had already been produced and asked to redo the recording. Informed by the observation in chapter 4 that participants tend to remember whistles in terms of the movement they make with the whistle plunger, the whistles were compared using a distance measure that is different from the one that was used in the reproduction constraint in the first whistle experiment. The distance measure was a linear combination of different separate measures,



a Initial set one



b Initial set two

Figure 6.3: *The initial whistle sets used in the experiment*

combined as follows: $0.3D_m + 0.6D_{md} + 0.2D_i + 0.05D_d$ where D_m is the Dynamic Time Warping (DTW) (Sakoe and Chiba, 1978) distance between the two movement tracks which were computed from the pitch tracks in the same way as described in chapter 4, section 4.4.1, D_{md} is the Dynamic Time Warping distance between the derivatives (Keogh and Pazzani, 2001) of the movement tracks, D_i is the DTW distance between the two intensity tracks, D_d is the difference in duration, computed following equation 6.1, where d_1 and d_2 are the lengths of the sampled movement tracks (at 500 samples per second).

$$\frac{|\log(d_1/d_2)|}{\log(d_1 + d_2)} \tag{6.1}$$

Again, data collected in the pilot study was used to create this measure and to determine the coefficients. The participants in this pilot all imitated 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. As in the original whistle study, the distance below which two whistles were considered the same was set at a relatively low value (0.02). In this way, participants could still produce relatively similar whistles and it would not influence the outcome of the recall phase in any way other than to reject doubles. This was effective, since after all data was collected, we could measure that 70.3 percent of all participants were never asked to redo their recording and on average it happened only 0.6 times per participant within the entire duration of the experiment. This prevented the initial introduction of accidental repetitions well enough to prevent a collapse and variation was preserved much better than without the constraint. In pilots that were done with no constraint, the final whistle set often showed the reuse of the same whistle up to 5 times in the same set and most whistles were used at least twice. This was definitely not the case in the results presented below with the constraint in place.

6.2.4 Participants

In total 64 participants took part in the experiment. They were divided over eight transmission chains, four in each condition. Participants were recruited from the University of Amsterdam community through posters and e-mail invitations. All participants were between the ages of 19 and 41 years old, 43 were female and 21 male. In each chain either two or three men participated. They were compensated for their time with a cash payment of 10 euros.

6.3 Qualitative results

This section describes qualitative observations to give a first impression of the data. First, the internal structure of the whistle sets is investigated

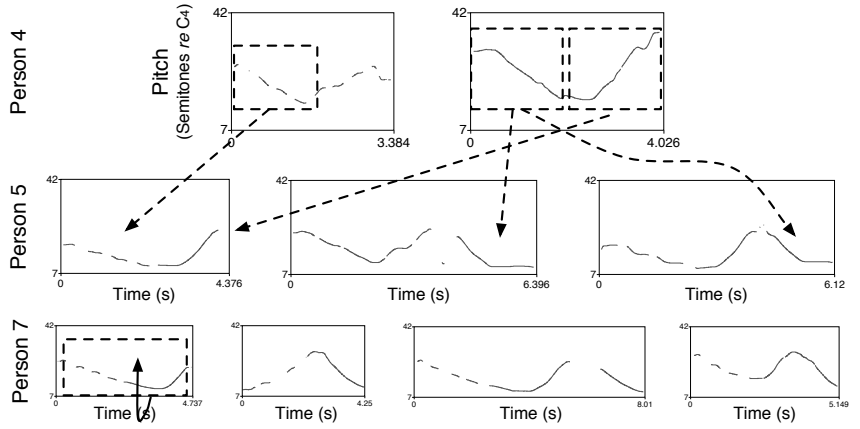


Figure 6.4: *Development of structure in a chain from the scrambled condition. Half of each of the whistles in the first row is borrowed and reused to form a new whistle. The left part of the smooth whistle is also reused and combined with existing whistles. These are then reproduced and all kinds of other variations on this appear.*

and compared to the structure that was found to emerge in the experiment without meanings. Second, the role of iconic form-meaning mappings is assessed. Appendix C.3 shows the complete transmission chains that resulted from this experiment.

6.3.1 Internal structure in whistle sets

On the level of the signals, independent of the objects they refer to, it can be observed that structure develops in a manner that is very similar to what could be observed in the experiment without meanings. Whistles were introduced that were clearly related in some way to the form of whistles that already existed in the set. For instance mirrored versions, combinations of existing whistles, repetitions of the same pattern within a whistle or whistles with similar shapes but different whistle manners appeared. Figure 6.4 shows an example of a development in one of the chains in the scrambled condition. Here, at generation four, two whistles are in the set that follow approximately the same shape in pitch contour (down and up), but are whistled in a different manner. One of them is whistled in a smooth and unbroken fashion and the other is more staccato-like and broken into pieces. In generation five, one half of each of these whistles is borrowed and reused to form a new whistle. The left part of the smooth whistle is also reused and combined with existing whistles. In later generations, these are reproduced and all kinds of other variations on this appear, such as ones that are mirrored again as a whole.

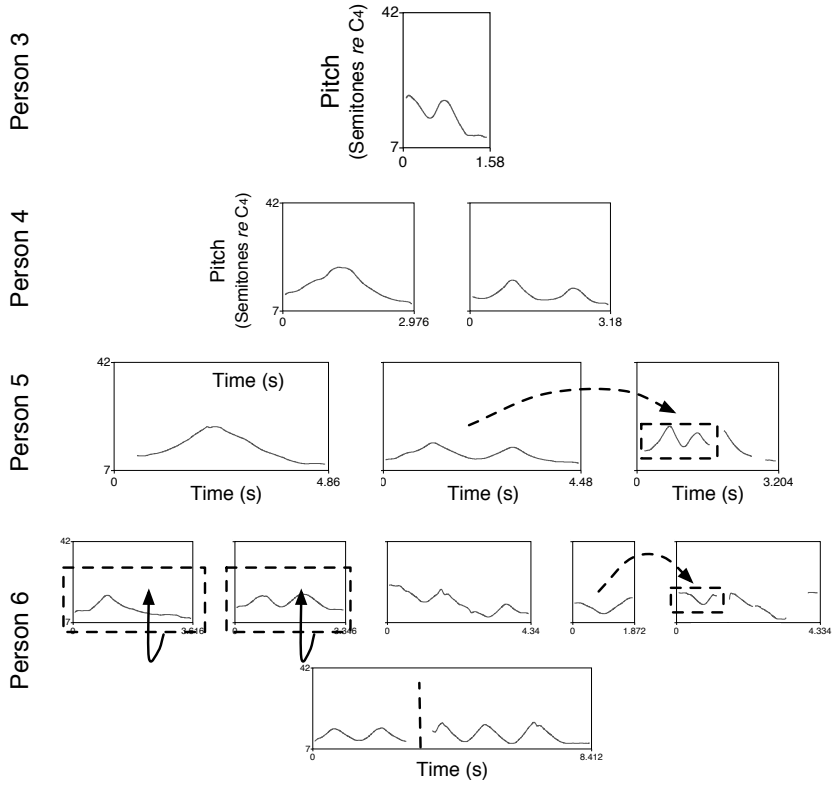


Figure 6.5: *Development of structure in a chain from the intact condition. The whistle on the first row seems to be an example for two new whistles in the next generation: one with one 'bump' and another with two. The 'two-bump' whistle is starting to be reused and combined with another pattern and in generation six both the one-bump and two-bump whistles are being reused, mirrored and recombined more widely.*

Figure 6.5 shows an example from one of the chains in the intact condition. In this example one whistle from generation three seems to be used as an example for two new whistles in the next generation: one with one 'bump' and another with two. In generation five the 'two-bump' whistle starts to be reused and combined with another pattern and in generation six both the one-bump and two-bump whistles are being reused, mirrored and recombined more widely. An existing whistle with several up and down movements is even segmented into two parts, where the first part is again the two-bump whistle.

To examine the final result of these gradual changes in the chains, we can look at the set of whistles produced by the eighth and last participant in a chain. Figure 6.6 shows a fragment of such a set

from the scrambled condition and here we can identify a clear combinatorial structure. There is a set of building blocks (short level notes, falling-rising slides, rising-falling slides and falling or rising slides) and these are reused and combined in a systematic way to create the whistles in the set. For some of the whistles, there is another version that is mirrored vertically and a pattern of short notes of alternating pitch height seems to be a recurring theme. The set has become very constrained as well, for instance in terms of the complexity of the falling-rising patterns and the overall variation in the type of building blocks that are left.

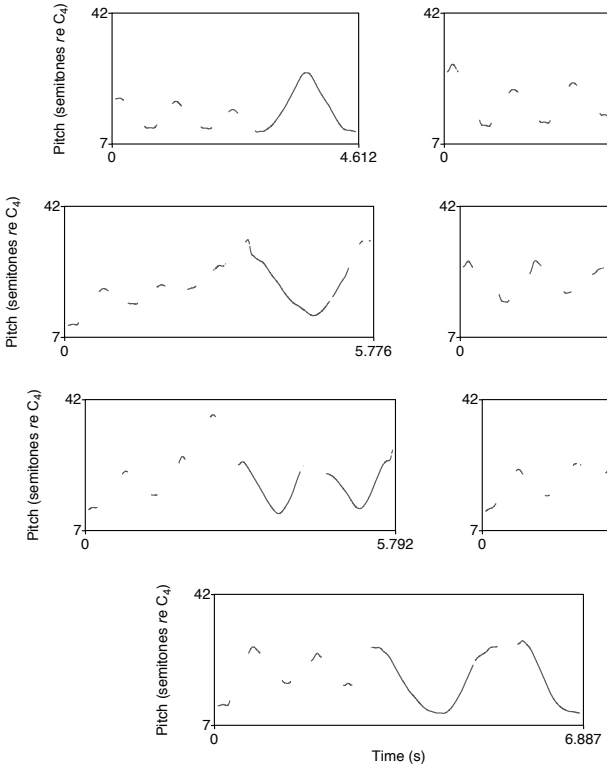


Figure 6.6: *Fragment from the whistle set produced by the last participant in a chain from the scrambled condition. Whistle sounds are plotted as pitch tracks on a semitone scale. Basic building blocks can be identified.*

6.3.2 Segmenting whistles into building blocks

As compared to the original whistle experiment, the emergence of a discrete set of basic elements seems to have happened in more varied ways in the current study. In the original experiment it was quite clear,

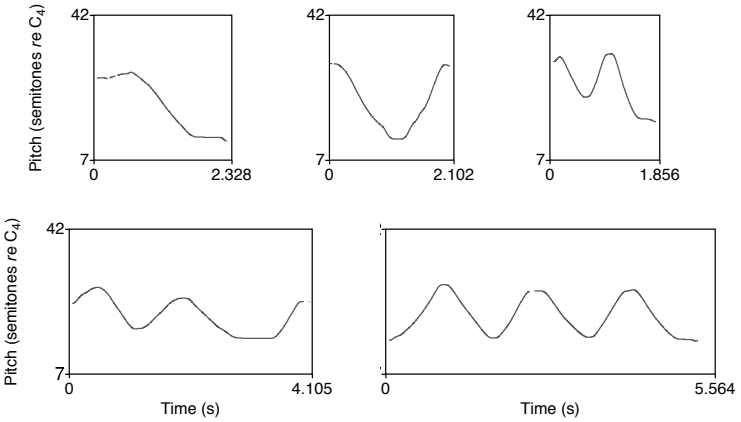


Figure 6.7: *A structure where silences do not determine segment boundaries.*

when observing the whistles qualitatively, that the silences (or pauses in the air stream) were solid indicators for where one segment ended and another one began. In the current study this appears not to be the only manner in which discretisation can be observed. Here, we also find structures that are combinatorial, but non-sequential or sequential without silences. Figure 6.7 for example shows a system in which all whistles are smooth, unbroken movements that differ from one another only in the number of falling and rising slides.

Figure 6.8 shows another example, where the same whistle shape, or movement, is reused several times, but each time with some parts realised in a different whistle manner (broken or smooth). This observation is taken into account in the quantitative analysis, described in section 6.4 and for which details can be found in appendix B.4.

6.3.3 Iconic whistle-object mappings

When talking about mappings between whistle sounds and alien objects one may wonder how a whistle sound can iconically depict such a visual object. In general, it is difficult to identify iconic relations as an outside observer, since iconicity is partly subjective and depends on experience and individual history. Whether a signal is iconic depends on how the receiver interprets it and this interpretation may be based on resembling associations for one person while they are purely symbolic for another (Keller, 1998). However, some examples could be found in the form-meaning pairs in the current data and iconicity could take several different forms in these examples. Most often, the shape of the whistle, or the pitch contour, would mimic certain features in the object. This could for instance be the overall shape of the object (round shape

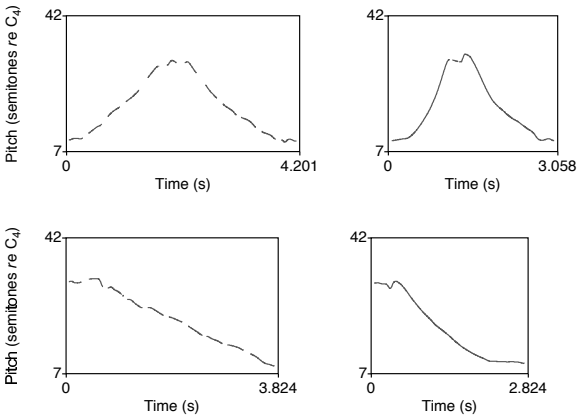


Figure 6.8: *A structure where recombination is not solely sequential.*

matched with curvy contour), the orientation of the object (long object placed on diagonal matched with one long falling contour) or the amount or direction of visually distinctive parts on the object (object with a certain number of distinctive parts on top of each other matched with whistle consisting of a comparable number of sounding parts with rising contour). It should be noted though that these are subjective observations and that it is not necessarily the case that the participants would agree with, or would be aware of the structural similarities between whistle and object as described. Judging from the observations, iconic mappings were not found to be widespread throughout the whole experiment. Figure 6.9 shows a few examples of clearly iconic form-meaning mappings that were encountered.

In some instances a clear shift could be observed in the data from iconic holistic signals towards non-iconic signals that became part of the combinatorial system. Figure 6.10 shows such an example. In this example a signal emerges that clearly mimics the shape of the object. This signal is copied by subsequent generations, although not perfectly. At some point a mirrored version of the signal is produced, which is equally iconic. Towards the end of the chain however, we see that the signal gets altered in such a way that it loses its iconic relation and starts to fit better with the rest of the system that emerged.

Participants filled out a post-participation questionnaire in which they were asked to describe their specific strategy (if any) for remembering the pairs and whether they thought the whistles and objects fit well together. Often participants reported strategies in line with the observations described in the previous paragraph. Other strategies that were reported involved: imagining how the object would sound and linking this with the whistle, imagining how the object would move and

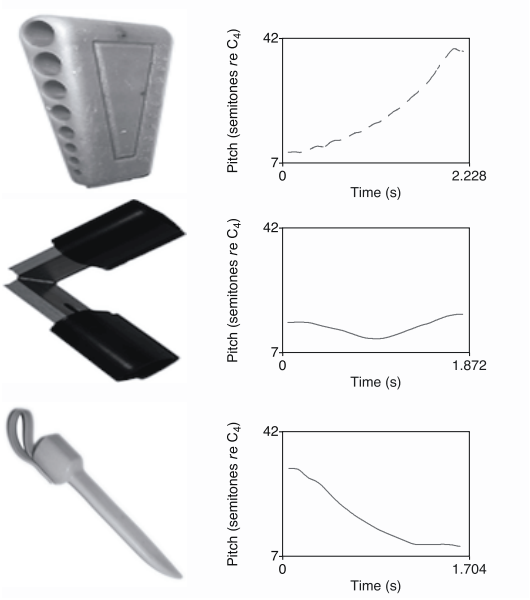


Figure 6.9: *Examples of iconic whistle-object pairs in the data. The first shows how the holes in the object that are arranged from the bottom to the top and become bigger are iconically depicted as a sequence of notes in a rising pattern. The second shows how the shape of the object is mimicked in the pitch contour. The third shows how the orientation of the object is imitated in the pitch contour.*

linking the pitch contour with that, or linking the object with some real object they know and linking the whistle with the sound that object would make. These reports further illustrate the subjectivity of form-meaning resemblance.

In summary, the structures that emerged in the sets of whistled signals resemble the discrete and combinatorial structure that emerged in the experiment without meanings, although there seems to be more variation in the way the signalling space is discretised: building blocks are not always straightforwardly segmented out by using silences as segment boundaries. Qualitatively, no difference could be observed between the structures in the two conditions. By observation, examples of iconic form-meaning mappings were not found to be abundant in the data, but participants did report often to make use of structural similarities between whistle and object as a strategy for remembering the pairs. However, these strategies were presumably very personal and subjective.

6.4 Quantitative results

This section describes a quantitative analysis that was used to assess whether the observed developments of structure are consistent across

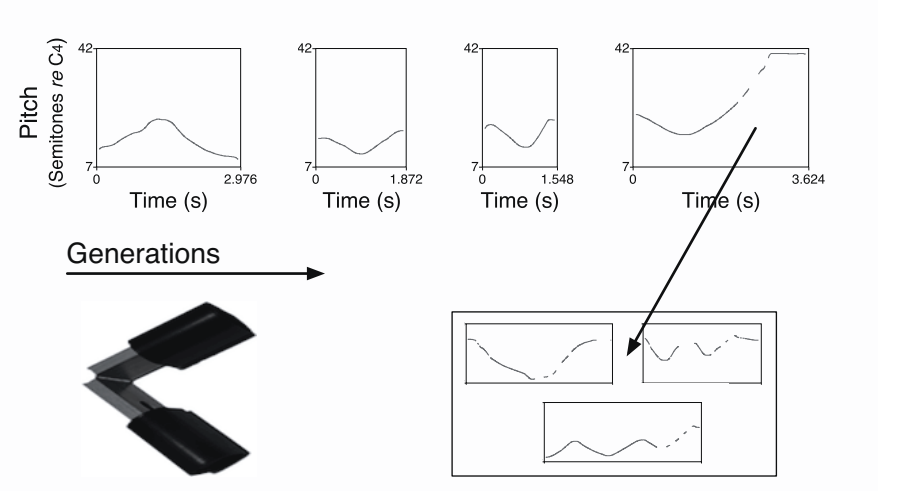


Figure 6.10: *An example of iconicity that is lost over generations.*

the data in all chains. First, the learnability is investigated by computing how well participants were able to recall the set of whistle-object pairs they had to remember. Then, the development of combinatorial structure is measured and compared over generations. Details on the implementation of the analysis and the signal preprocessing steps can be found in appendix B.4.

6.4.1 Recall error

To measure whether the sets of whistle-object pairs became easier to learn and reproduce, the recall error was measured by comparing for each participant the whistles that were produced with the whistles in the input. This was (first) computed in exactly the same way as in the analysis of the experiment without meanings. As described in chapter 4, section 4.4.1, whistles were matched by finding for each whistle a unique corresponding whistle in the other set in such a way that the sum of distances between the whistles is minimised. To determine the distance between two whistles, the same distance measure was used as the one described in chapter 4, section 4.4.1. This measure compares plunger movement tracks with the use of Derivative Dynamic Time Warping (Keogh and Pazzani, 2001).

Figure 6.11 shows the data for this measure of recall error for the four chains in both conditions, with increasing generations on the horizontal axis. The mean over the four chains for each condition is plotted with the standard errors. A significant decrease in recall error was measured using Page's (1963) trend test for the intact condition ($L = 724, m = 4, n = 8, p < 0.01$) as well as for the scrambled condition ($L = 732, m =$

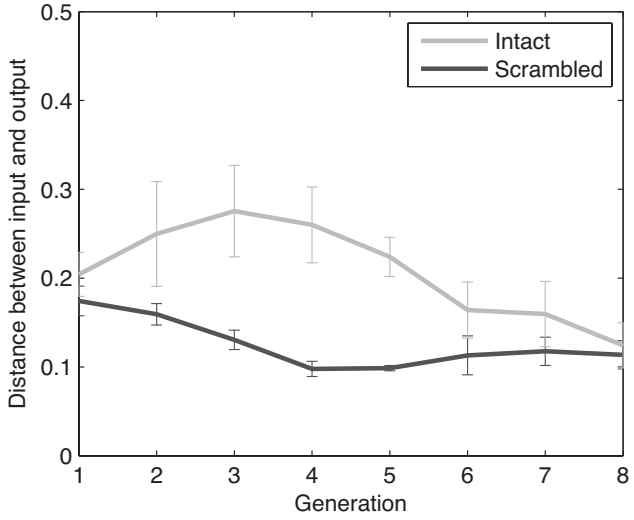


Figure 6.11: *Recall error over generations in both conditions, showing the mean and standard error. Recall error decreases significantly in both conditions.*

$4, n = 8, p < 0.01$). This means that there is an increase in the learnability and reproducibility of the form-meaning pairs over generations in both conditions.

In the measure of recall error described above, only the reproduction of the whistle sets independent of the meanings is assessed. Whether the participants were able to remember the right whistle for each of the objects is not taken into account. Although this measure allows for the most direct comparison with the results from the experiment without meaning, it makes more sense to include the correct form-meaning mapping in the analysis of the current data. To achieve this, the whistles that a participant produced for the objects were also directly compared with the whistles linked to those specific objects in the input.

Figure 6.12 shows the data for this measure of recall error on exact pairs for the four chains in both conditions, with increasing generations on the horizontal axis. Again, the mean over the four chains for each condition is plotted with the standard errors. A significant decrease in recall error was measured using Page's (1963) trend test for the intact condition ($L = 729, m = 4, n = 8, p < 0.01$) as well as for the scrambled condition ($L = 738, m = 4, n = 8, p < 0.01$), which means that there is also an increase in the learnability and reproducibility of the exact form-meaning pairs over generations in both conditions.

It should be noted that for most generations, there is a difference in the measured error between the two conditions. This issue is addressed in the next section.

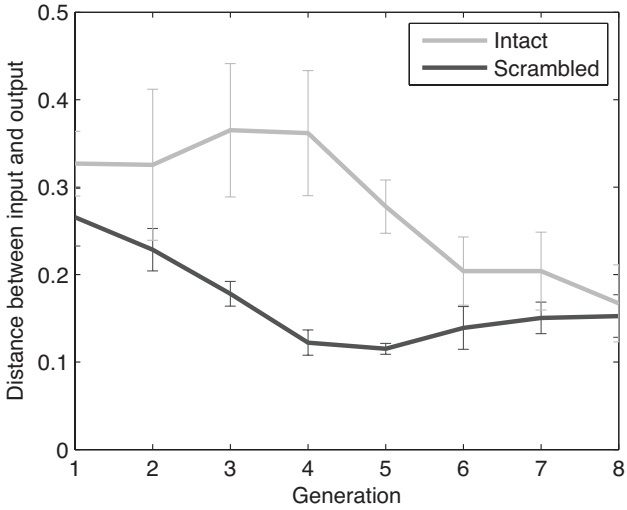


Figure 6.12: *Recall error on the exact whistle-object pairs over generations in both conditions, showing the mean and standard error. Recall error on the exact whistle-object pairs decreases significantly in both conditions.*

6.4.2 Combinatorial structure

To investigate whether the sets of whistles, like in the experiment without meanings, gradually become more structured after a number of transmissions, the entropy measure that was introduced in chapter 4 was applied to the current data. This measure makes use of the notion of entropy (Shannon, 1948) from information theory and is based on the idea that a set with more combinatorial structure is composed of fewer basic building blocks that are more widely reused and combined. One adjustment had to be made to the measure as it was described in chapter 4. Based on the qualitative observation that there was clearly no one 'right' segmentation that could be used to describe the discretisation of the signal space, three different types of segmentation were defined. The whistles were segmented in all three ways and the entropy was computed for each of the three sets of basic building blocks that resulted from the segmentations. The lowest entropy value that was measured was then considered to be the best minimal description length approximation and was used as the measure for (dis)order. The first type of segmentation was the original version, in which the silences were used as segment boundaries. The second type used the minima and maxima in the plunger movement track as segment boundaries and the third used the points of maximal velocity.

Figure 6.13 shows the development of entropy for the four chains in both conditions, where 0 refers to the initial whistle set. Again, the mean over the four chains for each condition is plotted with the standard error.

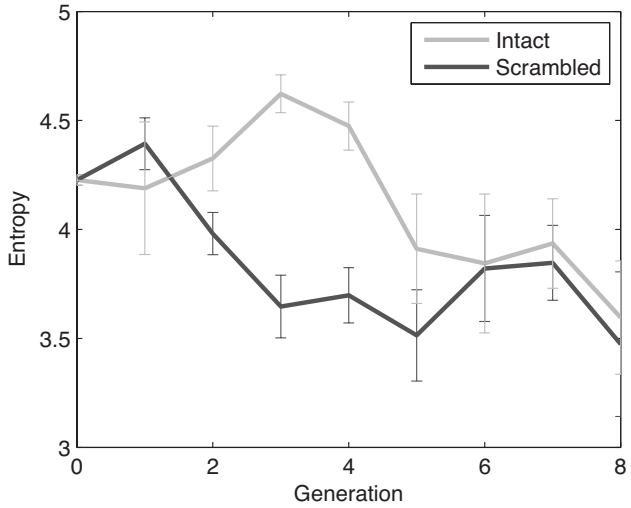


Figure 6.13: *Entropy of the whistle sets over generations in both conditions, showing the mean and standard error. Entropy decreases significantly in both conditions. This suggests that the combinatorial structure increased over generations.*

A significant decrease in entropy was measured using Page's (1963) trend test for the intact condition ($L = 728, m = 4, n = 8, p < 0.01$), excluding the artificially inserted initial set, as well as for the scrambled condition ($L = 712, m = 4, n = 8, p < 0.05$), excluding the artificially inserted initial set. These findings imply that the process of iterated learning in both conditions caused structure to emerge. Independent of the objects to which the whistles refer, there is an increase of structure and predictability and the whistles become internally more efficiently coded.

Thus, so far there does not seem to be a quantitative difference between the two conditions. Both the intact and the scrambled condition lead to a gradual increase of structure and more learnable systems toward the end of the chains. Is there any difference in the development of the structure in relation to the possibility for iconic mappings in the intact condition?

When looking at the development of entropy in the two conditions, we can see that the entropy in figure 6.13 in the intact condition tends to be higher than in the scrambled condition for almost all generations. A linear trend analysis of variance on the entropy with generation and condition as factors in a 2 x 9 mixed design ANOVA shows a main effect of condition, F(1,54) = 6.71 (p=0.012), as well as a main effect of generation, F(8,54) = 2.47 (p=0.023) (confirming the result of Page's trend test) and no interaction between generation and condition. This suggests that there is in fact a difference in the entropy between the two conditions. A post-hoc

Tukey's HSD test showed that entropy is significantly higher in the intact condition across generations as compared to the scrambled condition.

One perhaps surprising finding in the comparison between the two conditions may be that the average recall error seems to be higher at each generation for the intact condition. A linear trend analysis of variance on the recall error for exact pairs with generation and condition as factors in a 2 x 8 mixed design ANOVA reveals that there is indeed a main effect of condition, F(1,48) = 19.53 (p=5.63×10^{-5}), as well as a main effect of generation, F(7,48) = 2.35 (p=0.037) (confirming the result of Page's trend test) and no interaction between generation and condition. A post-hoc Tukey's HSD test showed that recall error is significantly higher in the intact condition across generations as compared to the scrambled condition. Given the expectation that iconicity would lead to more transparent and more learnable systems, we may have expected to see the reversed pattern. This issue will be addressed in more depth in the discussion section.

6.4.3 Transparency

Although it is difficult to assess the actual role iconicity played in the two conditions, the results of the guessing game phases could indirectly reveal a potential influence. If the mappings were more transparent in the intact condition, we would expect participants in that condition to score higher on the identification task after only very little exposure to the data. A linear mixed effects analysis was performed with *lme4* (Bates et al., 2013) in R to explore the effect of condition on the scores, with the fixed effect of round number (half of the items appeared in a guessing game round after the first exposure to the data and the other half after the second exposure) and intercepts for chain and generation as random effects. Likelihood ratio tests of this model against a null model excluding the effect of condition showed that condition does not affect performance in the guessing game ($\chi^2(1) = 0.210, p = 0.647$). This could suggest that the role of iconicity was minimal in both conditions, or at least did not play a large enough role in the intact condition to boost identification scores. However it needs mentioning that the participants had been exposed to the data before the guessing game phases, which is expected to have influenced the scores.

In order to deal with this issue, eight new participants were invited into the lab and asked to rate for each of the whistle-object pairs in all chains and for all generations in the intact condition how well they thought the sound fit with the object. This was expected to reveal whether a possible reduction of iconicity, measured as goodness-of-fit judgements, would coincide with the appearance of combinatorial structure in the intact condition. However, there did not seem to be any effect of generation on the degree of iconicity perceived by the participants on average, and when looking at each chain individually, only for one out of the four chains a significant decrease of rated goodness-of-fit could be measured over

generations with Page's (1963) trend test ($L = 2879, m = 12, n = 9, p < 0.01$). Perhaps more importantly, inter-rater consistency between the eight raters was very low, as measured with intraclass correlation (Shrout and Fleiss, 1979) (ICC(2,1) = 0.0406). This confirms the observation that iconicity is mostly subjective and is experienced differently from person to person.

6.5 Discussion

The experiment presented in this chapter shows that cultural evolution in the laboratory causes a system of whistled words for novel objects to become more learnable and more structured over time. This work expands a previous finding that showed the same result for whistled systems without meanings as discussed in chapter 4. For two different situations, one with transmission of intact form-meaning pairs and one with scrambled pairs, the transmitted whistled systems in the current experiments develop from sets of holistic signals towards having discrete and combinatorial structure. Sets of building blocks are efficiently reused and combined, similar to the structures of speech. In addition to the data presented in chapter 4, the current data forms another example to show that the emergence of combinatorial structure is not necessarily driven by vocabulary expansion and dispersal as was proposed by Hockett (1960). Even with vocabularies that are very small and the possibility for having solely holistic and iconic mappings without reaching the limits of the signalling space, structure emerges.

As a secondary objective the tension between combinatorial structure and possible iconic mappings was explored. It appeared that the potential for iconic mappings did not prevent the emergence of structure in this experiment. However, when looking at the development of entropy in the two conditions, we can see that there is a difference. Even though in both conditions structure emerged, entropy was higher in the intact condition than in the scrambled condition across generations. However, it is currently not possible to conclude that this difference is caused by a higher incidence of iconic form-meaning mappings in the beginning. The current data from the guessing game phase and the goodness-of-fit ratings does not suggest a large influence of iconicity. A more detailed analysis and probably also more data is needed. When we have a better picture of the development of iconicity, the development of structure can be linked to it more directly. However, finding an objective measure for quantifying the degree of iconicity in the data is not trivial. If it had been possible to identify a breakdown of structure in the meaning space, then it could be investigated whether the mapping between signals and meanings followed a pattern or seemed more or less arbitrary. However, the meaning space was chosen in such a way that there was no obvious structure present, which makes it impossible to group meanings according to criteria such as number of objects,

colours, spikiness or roundedness and so on. Quantitative measures such as the one designed by Tamariz and Smith (2008) could therefore not be used. On top of this, iconicity is partly subjective and whether it is present in a system may differ from one observer of that system to the other (Keller, 1998). As an initial attempt to define a measure for the transparency in the whistle-object mappings, the help of human participants was used. The results showed that participants did not agree on which mappings were iconic and which were not. This confirms the observation that iconicity is very subjective. A follow-up on this analysis would involve a guessing task, in which participants have to guess to which object a certain whistle belongs. In this manner it can be determined whether participants guess right more often on items from the earlier generations.

Under the assumption that iconic languages are more transparent and easier to understand and learn, one may expect that sound-symbolic mappings, if possible, should be exploited in linguistic systems, or even become more prevalent over time. Taken together, the findings from this study support a different expectation about iconicity, namely that it may not provide the perhaps expected reliable basis for building a linguistic system. From the current data we could conclude that iconicity may not be as easy as it seems at first. For the participants in the intact condition, where iconicity was possible, the recall error was significantly higher than for those in the scrambled condition. It is unclear why exactly this was the case, but it does show that the possibility for iconicity did not result in a set that was easier to remember. Assuming that the degree of iconicity was indeed higher in the intact condition, it could have been the case that the iconic signals were simply more complex and that they were therefore more difficult to reproduce precisely. Another reason for the higher error in the intact condition could be that the possible presence of a few iconic mappings created the false expectation that all meanings should have a resembling form and lead participants to reflect this in their reproductions. Since there is no proper way of determining the actual role of iconicity in the data yet, these are just guesses.

Even though there may be some strong cross-modal biases that are shared between individuals, many iconic links are based on individual history and experiences, as illustrated by the disagreement between goodness-of-fit raters. Moreover, there is often not only a single way in which a form can be iconic for a meaning. There are potentially many ways in which a form can be iconic, as well as many degrees of iconicity. For the objects in the experiment for instance, forms can mimic shape, orientation, number of parts, complexity, the sound it produces, the movement it makes, etc. In the domain of sensory sound symbolism, Fischer and Nänny (1999) describe different types of iconicity, such as ‘imagic’ iconicity, which refers to a direct resembling relation between form and meaning and ‘diagrammatic’ iconicity, which refers to consistencies in semantic or structural relations, or matching form-meaning topologies (de Boer and Verhoef, 2012). With so many

possibilities, a system with imagic, holistic iconicity may be transparent, but is very unpredictable. However, when a certain type of iconicity is used consistently for a semantic category or when iconic forms follow the constraints of combinatorial structure, it is part of a systematic and predictable system and may be more learnable than the type of mimicry/resembling iconicity alone. In the types of iconicity that have been found in speech, systematicity and regularities are indeed important. As Dingemanse pointed out: “It is the diagrammatic types of (...) iconicity that enable ideophones to move beyond the imitation of singular events toward cross-modal associations, perceptual analogies and generalisations of event structure” (Dingemanse, 2012, p.659). It is therefore possible that isolated iconic mappings are more likely to disappear than the type of iconicity that is part of a system, which may be the reason behind the relative rareness of imagic iconicity (Fischer and Nänny, 1999) as compared to diagrammatic iconicity. This second form may persist more easily if there is a good mapping between the topologies of the form and meaning spaces (de Boer and Verhoef, 2012). The emergence of topology-preserving form-meaning mappings has been investigated with the use of computer simulations (Zuidema and Westermann, 2003). The hypothesis that systems with unsystematic iconicity will develop towards having patterned iconicity (or arbitrary systematicity) through transmission gives the phenomenon of iconicity a role in theories on cultural evolution and the reduction of unpredictable variation in language evolution (Smith and Wonnacott, 2010). This leads to testable predictions and can be addressed in future experimental work.

Besides such a regularisation bias that may overrule possible cross-modal biases, there are also conformity biases that could influence the degree of remaining iconicity in a transmitted system. What I am referring to is what Tamariz (2011) calls ‘mindless imitation’ and it describes a widespread phenomenon in socially transmitted systems in which people simply copy an observed behaviour without knowing or understanding the function or origin of the behaviour. People tend not to focus on the meaningful or relevant parts when copying behaviours, but they focus on form and imitate this arbitrarily. Experiments with chimpanzees revealed that these animals can learn to solve problems with the use of tools by observing others (Tomasello et al., 1987). However, their tendency to precisely imitate each step of the observed behaviour seems to be different than that of human children (Horner and Whiten, 2005). Chimpanzees direct their attention to the environment and learn about affordances while observing others interacting with their surroundings (Tomasello et al., 1987). Human children on the other hand have been shown to have a much greater tendency towards imitating the precise pattern of behaviour of others, even if it is clear that part of that behaviour does not help towards solving problems (Horner and Whiten, 2005). The conformity bias can also be illustrated with the following story that Tamariz (2011) quoted from Gergely and Csibra (2006): “Sylvia always

cut the end of the ham when she cooked it because that is the way her mother did it; when the mother saw her do that, and asked why, Sylvia told her: “Because that is the way you always did it”. The mother explained that her pan was too small to hold a whole ham, and that was why she had to cut off the end.”. In linguistics it is therefore not unlikely that learners simply imitate utterances because many others are using them, without caring about, or even noticing iconic origins that perhaps are present. These iconic origins are subsequently likely to be regularised out of the system, a process which has been demonstrated to occur in iterated learning experiments by Caldwell and Smith (2012). In addition, such a process has been described for the early conventionalisation of forms in ABSL within families of signing individuals (Sandler et al., 2011). The sign for ‘egg’ for instance consisted of a compound of ‘chicken’, signed iconically with a handshape and movement resembling a pecking beak, and ‘oval shape’, signed with three fingers resembling an oval shape. In a later generation within a deaf family, this sign has changed into an assimilated form in which the oval handshape instead of the chicken beak handshape is used with the pecking movement, clouding the iconic origin of the movement.

Another example that illustrates the idea that perhaps people are not so much ‘helped’ by the iconic nature of form-meaning mappings when they try to remember such mappings, is a study on so called ‘tip-of-the-finger’ experiences. These experiences are similar to ‘tip-of-the-tongue’ experiences in speech, but for sign languages. It has been found that signers, when they can not retrieve the right sign from memory, often are able to remember a part of the phonological information (similar to not knowing a spoken word, but only knowing it starts with a particular sound). The signers would for instance know the handshape, but not the location or movement. Interestingly, while some signs had a movement, handshape or location that would iconically represent the meaning, the dimensions that were remembered first did not depend on whether these were the iconic dimensions. As an example, Thompson et al. (2005) describes how the sign for Switzerland has a movement that depicts the cross of the Swiss flag, but this part of the sign was not more likely to be remembered at first than other non-iconic dimensions. This phenomenon shows that iconic information is not always exploited even though it is there.

Concerning the design of the experiment discussed above, some changes could perhaps have prevented the problem that iconicity did not seem to be playing a large role in either condition. Perhaps if the chains would have been initialised with clearly iconic languages instead of random sets, we would have seen a more prominent difference between the two conditions. In addition, the use of novel unfamiliar objects and a novel peculiar ‘speech’ apparatus may have enhanced the fact that iconicity meant something different for each person in this experiment. A more natural and intuitive artificial language could perhaps be created with the use of gestures in the manual-visual modality. The manual-visual

modality probably also allows for a better mapping between form and meaning space, allowing for higher chance of encountering structured iconicity. Hearing participants without exposure to sign language will not be influenced by a pre-existing sign phonology, and it would be interesting to see if initially iconic artificial manual languages will become more like sign languages when they are experimentally transmitted.

To conclude, this chapter provides additional evidence to show that combinatorial structure in language can emerge through cultural evolution. The influence of iconicity in this evolutionary process still needs to be investigated in more depth, but with this study we provide an experimental platform that can be used to tackle this issue. In the future, additional experiments will be conducted in which gestures perhaps provide a more natural modality for iconic mappings and the design of more carefully controlled input languages allows for systematic investigations. These future studies are expected to provide more insights.

7 Agents

Most of the previous chapters in this thesis focus on how linguistic structure emerges and develops when it is transmitted over generations. With the experimental paradigm that was used in preceding chapters, we have seen that the structures get simplified, more constrained and they become easier to learn. In response to these results, the following question may arise: how are complexity and mutual intelligibility *preserved* over generations? This chapter presents a study in which the focus is on preservation of structure in artificial systems. Although the set-up is not directly comparable to the experimental results that were presented in previous chapters, it does provide an example of how experimental results may be complemented by computer simulation data. In addition, the simulations allow to model generations at a larger scale, with more than one individual per generation as well as the manipulation of age effects. Children for instance were not tested in the experiments, but their role can be modelled in computer simulations. The study described in this chapter involves an experiment in which a computer simulation is used to investigate the emergence and development of artificial vowel systems. The aim of this chapter is to show what the influence is on the preservation of complexity in vowel systems when children learn faster than adults. The computer model described in this chapter simulates populations of interacting agents in which the age structure varies.

7.1 Sensitive periods

Language learning is probably one of the few tasks that children are better at than adults. Adults, who have full-grown cognitive abilities and many years of experience in acquiring all sorts of knowledge and tasks, have the hardest time learning a new language. Children however, who are cognitively still underdeveloped, lack detailed motor control and are not deliberately trained, learn it perfectly. They pick up every aspect of

This chapter is a slightly adapted version of:
Verhoef, T. & de Boer, B.G. (2011). Language acquisition age effects and their role in the preservation and change of communication systems. *Linguistics in Amsterdam*, 4(1), 1-23.

the language without apparent effort. This phenomenon is often referred to as the ‘critical period’ for language acquisition (Lenneberg, 1969), or a period of ‘heightened sensitivity’ (Birdsong, 2005) which gradually declines. In this chapter it is investigated what the consequences are for a culturally evolving communication system if children learn faster than adults.

Time frames of heightened sensitivity to environmental input are not uncommon in nature. One well-known example is that of imprinting. Geese, chickens and ducks, soon after they leave the shell of their egg, will follow and keep following the first moving object or animal they perceive. The newborn gets imprinted with this object or animal and will continue to follow it as if it is following its parent (Hess, 1958; Lorenz, 1937; Spalding, 1873). There is an exception though. If the chicks and ducklings are not exposed to any moving object in the first 25 to 30 hours after their birth, this imprinting mechanism fails to work and the newborns miss out on their chance to imprint onto their mother (Hess, 1958). Likewise, rhesus monkeys that do not come in contact with other animals in their first year of life will be unable to develop normal social monkey behaviours (Harlow et al., 1965). Several bird species have only limited time to acquire their species specific song (Doupe and Kuhl, 1999) and visual systems of cats need to be exposed to external stimuli and developed within a certain time frame after birth (Hubel and Wiesel, 1970) because their window of opportunity closes.

These effects of age sensitivity are useful in the development of these animals. Sensitive periods have been identified in a more general context as being the result of evolution: “In many species and for different functional systems, strong selection pressures that favor great sensitivity to certain environmental stimuli and a maximum of learning during early stages of life may favor great stability of the results of such early experience, providing some kind or degree of protection against some later possible influences on the individual.” (Immelmann and Suomi, 1981). Therefore there might be advantages to a sudden or gradual change in learning abilities for humans as well, for instance in language. Just like it is crucial for the survival of a duck that it knows soon after birth who its mother is, language might be so important for humans that early learning of it is crucial to be able to function in our complex world of social structures.

Studying the exact causes and consequences of age sensitivity in language acquisition is not trivial. Language, and the mechanisms for acquiring it are shaped by processes on three different time scales (Kirby and Hurford, 2002): learning, cultural evolution and biological evolution. As Steels (1997b) has pointed out, extreme complexity arises in the interaction between these time scales. As has already been shown by Kirby and Hurford (2002), Steels (1997b) and others, computer models can deal with this complexity, and have already been applied successfully in the study of age sensitivity in language acquisition.

Simulations allow researchers to explore hypothetical situations that would be impossible to achieve in real language environments, as these cannot be manipulated experimentally. With computer models it is feasible to explore multiple potential scenarios, a method that can compensate to a certain extent for the information that is inevitably lost in historic processes. This is why computer models are excellent tools for studying scenarios for language evolution.

In explaining the cause of age sensitivity in language acquisition, evolutionary simulations have provided insights. Hurford (1991) created a model in which it is assumed that a selective advantage exists for linguistic abilities. The population goes through many generations of selection and mutation and a critical period effect evolves, caused by the selective pressures on language learning. In Hurford's view, the language acquisition capacity evolves as an adaptation, but the critical period arises as a side effect of "a lack of selective pressure to acquire (more) language (or to acquire it again) once it has been acquired" (Hurford, 1991). In a related model, Hurford and Kirby (1999) show that complex interactions between speed of language acquisition as a biological property and language size as a cultural property are important. The acquisition speed in this model evolves in such a way that full language acquisition is completed before puberty, the age at which agents start to reproduce. The speed will not evolve to become even greater once the complete language of the population can be acquired before puberty. When the language can be fully acquired before puberty, there is no pressure favouring faster acquisition than necessary to be 'ready' before it is needed for reproduction. When an 'innovative potential' is introduced, which gives the agents the ability to expand the language when they have acquired the complete existing language early, the language can grow and then again more speed is needed to acquire it on time, resulting in an arms race with ever increasing size and speed. Among many others, one more example concerns a more mathematical evolutionary model that was investigated by Komarova and Nowak (2001). In their model a critical period exists as an evolutionary stable strategy (Nash equilibrium). Their model assumes that there is a cost to language acquisition which has a negative effect on the reproductive success while at the same time a poor language performance reduces fitness as well. The interaction between these two forces yields an evolutionary stable optimal learning period. These examples all show that computer models can be very useful for the discovery of new ideas about language and age structure dynamics.

In this chapter, a simulation is introduced in which the focus is not, contrary to the contributions mentioned above, on what caused the age structure to biologically evolve, but on the consequences of such a structure on the scale of the population. The goal here is not to find evolutionary explanations for decreasing language learning abilities but it is assumed that this exists and investigated what the presence of such an age structure entails for a culturally evolving communication system.

The models of de Boer and Vogt (1999) and de Boer (2000) are used, which are very suitable for the investigation of the interaction between individual level behaviours and consequences at the cultural, population level. A conclusion that may be drawn from the results presented in this chapter, corresponding to what had been found by de Boer and Vogt (1999), is that an age structure improves stabilisation and preservation of complexity in the shared communication system of a changing population. These results are compared with related findings from the field of sociolinguistics in which the role of differences of learning ability in language change is adressed.

7.2 Age sensitivity in language acquisition

Lenneberg (1969) initiated modern research into age effects for language acquisition. He observed that traumatic brain lesions caused permanent aphasia in adults, while children with similar lesions initially had the same problems but were able to recover. Other important observations in this research are cases of language deprivation, such as the well-known case of Genie (Curtiss et al., 1974). When Genie started to acquire her first language, after being raised in almost complete social isolation, she was almost fourteen years old. With a lot of training researchers were able to teach Genie a small vocabulary but her language use stayed far from normal (Meadow, 1978). The case of Genie caused controversy because her situation was so extreme that it is actually hard to determine whether her abnormal linguistic development was due to linguistic stimulus deprivation or to the adverse conditions in which she grew up. Similar problems are connected to the observations that Lenneberg (1969) reported about aphasics. These are all abnormal conditions which makes it difficult to draw conclusions about the normal language acquisition pattern.

Cases of delayed sign language exposure form more natural examples (Mayberry, 2007). Newport (1988) describes differences between adult and infant learning of American Sign Language (ASL). The congenitally deaf subjects in this research were raised under normal circumstances. In contrast to the acquisition of spoken language, most ASL learners do not start to learn their language when they are born but most of the time they are exposed to the sign language for the first time when they start to go to school. However, in the cases where deaf children acquire their language from parents who are also deaf and communicate with ASL, they do get early exposure. Comparing the acquisition of ASL for early and late exposure and the level of ultimate fluency that is achieved, Newport (1988) found age effects. Mayberry (2007) investigated the influence of first (sign) language acquisition on the ability to learn a second language and also found an age of acquisition effect, showing that early exposure to a first language also benefits second language acquisition.

In addition to the observations of development in first language acquisition, age effects are being studied in second language learners. Second language learners and proficient bilinguals provide a source that teach us about the interaction between different languages (Flege et al., 1995), the influence of age of first language acquisition on second language learning (Mayberry, 1993; 2007), the ultimate proficiency that can be reached in the second language (Flege et al., 1995; Johnson and Newport, 1989; Oyama, 1976) and which aspects of the language are the hardest to acquire at an older age (Singleton, 2002; Weber Fox and Neville, 1996).

The exact mechanism behind language acquisition age effects is still unknown. Many different driving forces have been proposed. Some of them assume that the acquisition time frame is biologically determined, for instance coinciding with a decline of brain plasticity during development (Lenneberg, 1969). Brain plasticity has been found to relate to hormonal influences and the changes that occur around the age of puberty (Yun et al., 2004). The link between hormones and acquisition is also found in research done with song birds. As cited by Doupe and Kuhl (1999), experiments have been conducted in which birds were castrated before they had learned their song. The hormonal changes that accompanied this procedure influence their singing behaviour and by the time they would normally learn their song, these birds did not. However, when they would at a later moment receive certain hormones, they were still able to learn their song. This shows that in some cases the special learning ability seems to be extendable.

According to the ‘less is more’ hypothesis (Newport, 1988), children perform better because they are cognitively more limited when the acquisition takes place. The adult learners are able to store more whole forms and meanings, and may therefore face a more difficult analytic task. So, children have an analytical advantage because they have to make more generalised hypotheses about their linguistic input, making them more inclined to find patterns and regularities, even though there may be counterexamples. Adults on the other hand are able to handle more complex hypotheses and tend to have difficulty extracting more general patterns, especially when the linguistic input contains inconsistencies.

Other theories have been proposed that do not assume a direct link with developmental patterns. These theories assume that what we observe as age effects in language acquisition, actually has more to do with the quality and quantity of linguistic input. According to Kuhl (2000), the mechanism involves ‘neural commitment’, which means that a mental map is being created while learning, adjusting neural structures to the sounds of the native language. Perception is guided by a ‘native language magnet’. Newborns are able to distinguish sounds of many different languages, but when they grow up, their perceptual behaviour changes so that they are better able to distinguish the sounds of their native language while other, for them irrelevant, contrasts are no longer

perceived. Once their neural structure has committed to a language, other languages become harder to learn. Flege (1999) poses a similar hypothesis which also emphasises how a first and second language can influence each other. With his 'interaction hypothesis', he proposes a negative correlation between first language and second language proficiency.

7.3 Age effects in language emergence, change and growth

Several lines of research have shown that children and adults, with their differences in language learning ability, might be involved in language emergence, change and growth in different ways. When people come into contact who do not speak a mutually intelligible language and there is a need for communication, a pidgin language may emerge (Hall, 1962; Sankoff and Laberge, 1974), which contains features of both original languages, but does not contain a lot of structure. When the contact situation is over, the pidgin may disappear again (Hall, 1962) or else it may be transmitted to future generations and children will learn it. In this last situation, it has been argued that native learning might stimulate the emergence of structure (Sankoff and Laberge, 1974). A newly emerging sign language in Nicaragua, that did not originate from language contact but spontaneously emerged when a new school brought together deaf people, showed a comparable pattern in which children played an important role in the formation of the language. (Senghas et al., 2004).

A recent finding by Labov (2007) indicates that the differences in learning ability between adults and children could explain different observations in language change and stability. Two models for explaining linguistic change have dominated for a long time: the 'family tree model' and the 'wave model' (Labov, 2007). The family tree model describes how languages in the world are related and how protolanguages branched into new groups of related languages. One limitation of this model has been said to be that it assumes separate branches are independent. The wave model was introduced as an alternative view in 1872 (as described by Fox (1995)(p.129)) and accounts for the spreading of changes in the case of language contact, across language boundaries. Labov (2007) unifies these two models in one theory in which he accounts for both faithful 'transmission' of changes from generation to generation, resulting in family tree like pattern and the 'diffusion' of changes through language contact, resulting in a wave like pattern. The difference in learning abilities of young and old language users plays a major role in this theory. Labov (2007) presents a detailed description of two comparisons of sound system development: the first compares the stability of the short-*a* system within New York City with the spreading of this system to four other cities and the second compares the Northern Cities Shift in the areas of the Inland North with the spreading of this

system to the Midland cities. The spreading of the complex short-*a* system of New York City to other cities had resulted in a loss of structure and regularity, while it had been and remained stable within New York City. According to Labov, this difference is due to the fact that in New York City, there was an unbroken chain of faithful transmission from adults to children, while it was mainly individual adults who diffused the system to other cities. Likewise, the Northern cities shift resulted in a uniform and stable transmission of the system in the Inland North, probably because this case involved the migration of entire communities consisting of families with both adults and children. The changes were therefore steadily transmitted from generation to generation. At the same time, in the Midland cities the result was less stable and involved again a contact situation with single adults traveling around. Labov (2007) proposes therefore that the difference in adult and child learning causes the difference between transmission and diffusion.

The results of the simulations described below support this finding by Labov (2007). Here, it is also investigated how age effects in learning influence the cultural evolution of a communication system and the results show that an age structure in the population helps stabilisation and preservation of structure in emerged vowel systems. This chapter therefore provides additional support for the proposal that it is important to consider the consequences of language acquisition age effects in the study of language preservation and change.

7.4 Vowel systems in a population of agents

The starting point of this work is the agent-based imitation game model as described by de Boer (1997; 2000). This model follows the language game paradigm, which focuses on communication between individual agents in a population. Important in this work is the idea of viewing language as a population level phenomenon in which self-organisation causes optimisation and coherence. In de Boer and Vogt (1999), a version of this model is used which integrates transmission across multiple generations of interacting agents. This allowed them to introduce the concept of an age structure, which has also been applied in de Boer and Vogt (1999). Both models are re-implemented and used to investigate the influence of the age structure in detail.

The model consists of a population of individuals that interact with each other by means of imitation games (de Boer, 2000). The agents are equipped with a memory in which they can store articulatory (and acoustic) information about the vowels they have learned. With the use of a realistic articulatory synthesiser they can produce these vowel sounds and they are able to categorise the sounds that others produce with a model of vowel perception. In response to their interactions with other agents they can update their memory and learn new sounds.

7.4.1 Memory

Following de Boer (2000), the agents have a vowel memory in which they store prototypes of the vowels they have learned. Prototypes are the centres of the vowel categories that the agent has learned. A vowel prototype is described by three properties: tongue position (p), tongue height (h) and lip rounding (r). They have continuous values that can vary between 0 and 1. For each prototype the agent also keeps track of the number of times it has been used and the number of times it has been used successfully.

7.4.2 Production

Values for position, height and rounding form the input for the articulatory synthesiser. The output of the synthesiser consists of the frequencies of the first four formants of the vowel in Hertz. They are computed with a set of equations that are described by de Boer (2000). Thus the agents can produce a realistic range of sounds. Some noise is added to these formant frequencies to account for the fact that there is variation in the speaker's production of the same vowel. For every ith formant F_i, the frequency $F_i^{'}$ with added acoustic noise is computed by $F_i^{'} = F_i(1 + v_i)$, where v_i is randomly chosen from a uniform distribution such that $-\psi/2 \leq v_i < \psi/2$, where ψ is the maximal noise allowed.

7.4.3 Perception

The agents perceive a sound as the nearest vowel prototype in their memory. This implements categorical perception. To do this, they need to determine the distance between two vowels in acoustic space. This is done in a two-dimensional space: the first dimension is the first formant and the second is the effective second formant (both on a Bark frequency scale). The effective second formant is computed from the three higher formants (Schroeder et al., 1979). A graphical representation of this acoustic space is shown in figure 7.1.

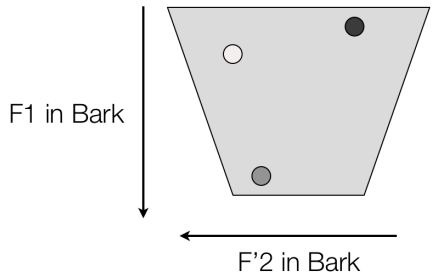


Figure 7.1: *Graphical representation of the acoustic space (the trapezium), with a repertoire of vowel prototypes (the dots)*

To compute the effective second formant $F_2^{'}$, the formant values are first converted from the Hertz to the Bark scale using equation (7.1) (Schwartz et al., 1997)

$$F_{Bark} = 7 \sinh^{-1}(F/650) \tag{7.1}$$

Then $F_2^{'}$ is computed for the two vowels (vowel *a* and vowel *b*) that are compared, following the equations in de Boer (2000) and these two values are used to compute a weighted Euclidean distance D, like in de Boer (2000) and as shown in the following equation:

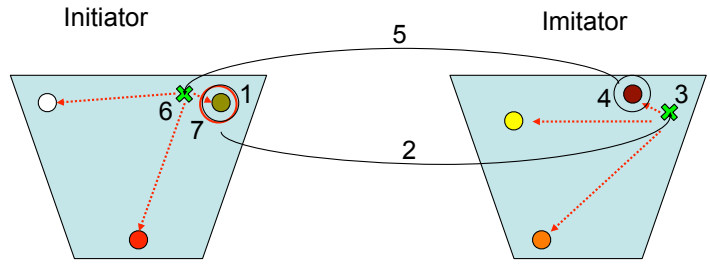
$$D = \sqrt{(F_1^a - F_1^b)^2 + 0.3(F_2^{a'} - F_2^{b'})^2} \tag{7.2}$$

7.4.4 Interactions

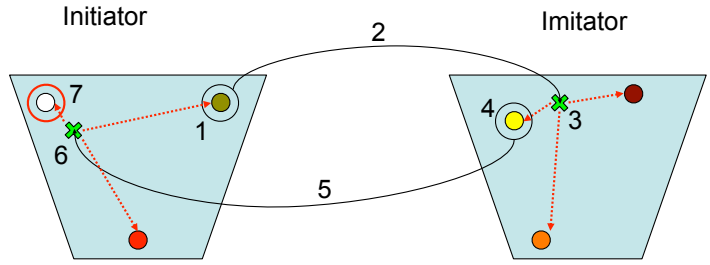
The agents in the simulation interact with each other by playing imitation games (de Boer, 2000). In every game, two agents are randomly selected from the population. One of them is the *initiator* and starts the interaction by selecting a random vowel from its prototype repertoire and producing this sound. The other agent is the *imitator* and perceives the vowel that the initiator has just produced: it finds its prototype that is closest to the sound the initiator produced. Next, the imitator produces the selected prototype and the initiator perceives this sound. The initiator then determines which prototype in its memory is closest to this sound. If this prototype is the same as the one which the initiator initially produced, the game is a success, if not, it is a failure. This information is communicated to the imitator non-verbally and both agents update their memory in response to the game. When an agent is selected for the first time, its repertoire is still empty. In the role of initiator it then creates a random sound from the articulatorily possible range and adds this prototype to its memory. An imitator with an empty repertoire adds a new prototype. In order to turn this new prototype into a close imitation of the heard sound, the agent adopts a rehearsal strategy in which the agent repeats the sound for itself and improves its pronunciation with a hill-climbing heuristic. In figure 7.2 both a successful (7.2a) and an unsuccessful (7.2b) game are illustrated.

7.4.5 Memory update steps

In response to an imitation game both players update their vowel memory. After each game the use and success counters for the vowel prototypes are updated for both the initiator and the imitator. Whenever the success/use ratio of a prototype moves below a predefined threshold (0.7) and it has been used often enough (a minimum of 5 times), it is removed from the vowel repertoire because this means it is not well aligned with the prototypes of the other agents in the population. Prototypes that have become too close to each other in either acoustic



a Imitation game success



b Imitation game failure

Figure 7.2: *Procedure of an imitation game: 1. selecting a random vowel prototype, 2. producing this vowel, 3. perception of this vowel in the acoustic space of the other agent, 4. finding the closest vowel prototype, 5. producing this vowel, 6. perception of this vowel in the acoustic space of the first agent, 7. finding the closest vowel prototype. In 7.2a the game results in a success because the recognised vowel in step 7 is the same as the one produced in step 1. In 7.2b it is a failure because the recognised vowel in step 7 is not the same as the one produced in step 1. Images adapted from animations by de Boer (2000).*

or articulatory space will be merged. In articulatory space they are too close when the Euclidean distance is smaller than 0.17, corresponding to a minimal difference of 0.1 for each articulatory parameter. In acoustic space the threshold θ depends on the relative difference in Bark $\Delta_{relBark}$ and the maximal acoustic noise ψ. For a multiplication by a factor of x in Hertz, the relative difference in Bark is given by:

$$\Delta_{relBark}(x) = 7\ln(x) \tag{7.3}$$

The threshold θ is then given by :

$$\theta = \Delta_{relBark}(1+\psi) - \Delta_{relBark}(1-\psi) \tag{7.4}$$

This threshold defines a noise-dependent just noticeable difference in acoustic space.

There is no meaning in this model, which means that the signals do not refer to anything. In order to simulate a pressure for having more

different signals in a repertoire, improving the expressivity, the agents can add a random vowel to their repertoire with a small probability, thus causing the repertoire of signals to grow.

In addition, there are adjustments that only the imitator makes. These steps depend on the outcome of the game. If it was a success, the imitator will move the prototype that it used closer to the sound that was heard. If the game failed, there are two alternatives: if the prototype that was used is not a very successful one (its success/use ratio is lower than 0.5), it is improved by shifting it closer to the heard signal with a predefined step size; if, on the other hand, the used prototype is a successful one, it remains unchanged (it has contributed to successful games with other agents) and the initiator adds a new prototype which is determined with the previously described rehearsal strategy. The details of how these steps were implemented exactly can be found in the original article by de Boer (2000).

7.4.6 Population dynamics

In the original model (de Boer, 2000) the population does not change during an experiment. de Boer and Vogt (1999) introduced a version in which the population does change. Agents die and new agents with empty vowel repertoires are born. The new agents have to learn the existing vowel system from the older agents. This introduces a component of vertical transmission which makes it possible to investigate what happens to a vowel system when it is passed from generation to generation. If the system is run for long enough, with a high enough replacement rate, at a certain point in the simulation none of the initial speakers are left in the population.

A necessary adjustment by de Boer and Vogt (1999) to the original model concerns the rehearsal strategy. In the original model agents were able to repeat the sounds for themselves an indefinite number of times but in reality this is impossible. The number of times that agents can rehearse is therefore limited to 10 steps per prototype.

de Boer and Vogt (1999) showed that the vowel systems are better preserved over time if the population has an age structure. In this case, agents that have been in the population for a shorter period can learn faster than the ones that have been around for longer. The older agents therefore make smaller changes to their vowel repertoires and these provide a more stable target for the new agents. The age structure is implemented by a variable step size with which the agents can shift the vowels in their repertoire. This step size decreases with age according to equation (7.5), where ε_t is the step size at time t, α is the speed of ageing and ε_∞ is the ultimate step size. In this case the step size decreases from $\varepsilon_0 = 0.03$ to $\varepsilon_\infty = 0.01$.

$$\varepsilon_t \leftarrow \varepsilon_{t-1} + \alpha(\varepsilon_\infty - \varepsilon_{t-1}) \tag{7.5}$$

Equation (7.5) defines an exponential decay following:

$$\varepsilon_\infty + (\varepsilon_0 - \varepsilon_\infty) e^{\ln(1-\alpha)t} \tag{7.6}$$

This equation was used to determine the value for α used in the experiments, based on the approximate number of steps that the 'sensitive period' of the agents had to last.

7.5 Simulations

In the previous section, the original imitation game model of de Boer (2000) and the extension including population dynamics of de Boer and Vogt (1999) were described. These models have been re-implemented, with the addition of comparing two different types of the age structure: a gradual decline of learning ability such as described above (from de Boer and Vogt (1999)) and a critical period with a strict cut-off moment. In the following the simulated experiments with this model are described as well as the results of these simulations.

7.5.1 Experiments

The experiments described in this section all start off with the same emerged vowel system that was the outcome of a run of 200 000 games of the original model by de Boer (2000) with 50 agents. This emerged vowel system is shown in figure 7.3.

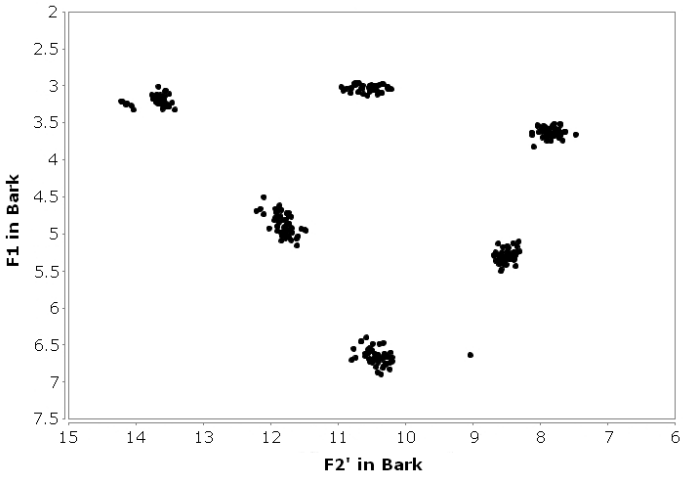


Figure 7.3: *Emerged vowel system after 200 000 games in a population with 50 agents. Each dot represents a vowel prototype in an agents memory. This system was used as the starting point of the experiments with changing populations.*

The games were continued with a replacement probability of 0.003 per game. With this probability old agents were replaced by new, empty agents. This replacement rate was chosen such that the agents had enough time to acquire the vowel system of the population before their replacement. Simulations were run under four different conditions. In two conditions there was no age structure and in one of these the step size, which determines the speed of learning, was large (0.03) and in the other small (0.01). In the other two conditions there was an age structure in which the learning step size decreased across the lifespan from $\varepsilon_0 = 0.03$ to $\varepsilon_\infty = 0.01$ according to equation (7.5). The speed of ageing, α differed for these two conditions. One population aged quickly ($\alpha = 0.02$) such that the sensitive phase was approximately $1/20^{th}$ of the expected life time. The other population aged slowly ($\alpha = 0.005$) so that agents were more sensitive for approximately $1/5^{th}$ of their life time.

For each of the conditions 30 000 games were played so that it was unlikely that agents from the first 'generation' would still be in the final population. All runs were repeated 100 times.

7.5.2 Measures

At the end of each run, measures were computed for a comparison of the results: the success, the size and the energy. *Success* is the average imitation game success over all games in a run, *size* is the average number of vowels of all agents in their repertoire at the end of a run and *energy* is the average over all agents of the energy of their vowel systems at the end of a run. The energy of a vowel system is calculated using the method from Liljencrants and Lindblom (1972), which was described in chapter 4 and measures perceptual contrast (a lower energy means a better contrast). In addition, two measures were used to compute the extent to which the resulting vowel systems resembled the initial system. A *similarity* measure is used which is based on the average communicative success between agents from the initial population and those of the final population. In addition a *distance* measure is used which clusters the vowel prototypes of all agents in a population using Leader-Follower clustering (Duda et al., 2001). Then, the average is computed of all Euclidean distances between each cluster centre in the final population, and the nearest cluster centres in the original population.

These two measures complement each other. The average communicative success reflects more realistically how one would assess the difference between natural languages because it uses mutual intelligibility. However, communicative success is remarkably robust to small changes in the vowel systems. In addition, communicative success depends on the size of the vowel system. With smaller systems the expected number of mistakes is smaller. The distance measure is able to detect small changes more easily, but a disadvantage is that the results of Leader-Follower clustering are sensitive to the order of data presentation and might not work desirably in cases of confusion such as

big differences in individual realisations of the same vowel. Using both measures therefore seemed like a good combination.

7.5.3 Results

The results of the experiments are presented in table 7.1, showing the computed measures for each of the four situations and averaged over the 100 runs.

Step size Ageing	$\varepsilon_0 = 0.01$ $\varepsilon_\infty = 0.01$ no ageing: $\alpha = 0$	$\varepsilon_0 = 0.03$ $\varepsilon_\infty = 0.01$ fast ageing: $\alpha = 0.02$	$\varepsilon_0 = 0.03$ $\varepsilon_\infty = 0.01$ slow ageing: $\alpha = 0.005$	$\varepsilon_0 = 0.03$ $\varepsilon_\infty = 0.03$ no ageing: $\alpha = 0$
Success:	0.910 $\pm$ 0.012	0.879 $\pm$ 0.009	0.878 $\pm$ 0.011	0.914 $\pm$ 0.012
Energy:	0.79 $\pm$ 0.26	1.02 $\pm$ 0.23	0.98 $\pm$ 0.26	0.65 $\pm$ 0.34
Size:	3.05 $\pm$ 0.30	3.43 $\pm$ 0.30	3.35 $\pm$ 0.31	2.74 $\pm$ 0.51
Similarity:	0.719 $\pm$ 0.033	0.774 $\pm$ 0.031	0.768 $\pm$ 0.031	0.709 $\pm$ 0.037
Distance:	0.792 $\pm$ 0.104	0.686 $\pm$ 0.085	0.692 $\pm$ 0.108	0.852 $\pm$ 0.126

Table 7.1: *Results of the experiments with and without age structure, showing the averages and standard deviations of the measures. ε is the learning step size and α the speed of ageing. Note that the similarity is higher and the distance lower for runs with the age structure: the vowel systems are preserved better. The average success is smaller in the two cases with ageing because vowel systems here are larger.*

From table 7.1 it is clear that when there is an age structure the similarity is higher than when there is none. Independent samples t-tests reveal that this difference is significant with $p < 10^{-29}$ for all four cases: comparing both ageing speeds with both the low ε_0 case and the high ε_0 case. The effect sizes (computed with Cohen's *d*), which indicate the discriminability or non-overlap between the two distributions, are for all four comparisons larger than 1.5, which is a very strong effect according to Cohen's (1992) scale. For all four comparisons the exact effect sizes and corresponding confidence intervals (computed using the bootstrapping method following Kelley (2005)) are shown in table 7.2. The distance measures of the four conditions also show a clear difference, with a lower distance in conditions with an age structure. The

No ageing		Fast ageing	Slow ageing
low ε_0	d	1.7118	1.5506
	CI	1.3694 – 2.1668	1.2086 – 1.9849
high ε_0	d	1.8840	1.7337
	CI	1.5407 – 2.3287	1.3842 – 2.1872

Table 7.2: *Similarity measure: Effect sizes (based on Cohen's d) and the corresponding 95% confidence intervals (CI).*

values for both ageing speeds are significantly different (with all four $p < 10^{-15}$) from both non-ageing conditions. The effect sizes were all larger than or almost equal to 1. The exact effect sizes and corresponding confidence intervals for the distance measure are shown in table 7.3.

No ageing		Fast ageing	Slow ageing
low ε_0	d	1.1174	0.94697
	CI	0.83322 – 1.48669	0.65448 – 1.27165
high ε_0	d	1.5433	1.3646
	CI	1.2742 – 1.8750	1.0840 – 1.6629

Table 7.3: *Distance measure: Effect sizes and confidence intervals.*

In addition to this, the size measures show that the age structure helps to preserve the original size of the vowel system. The size in conditions with age structure remains significantly higher than in both conditions without age structure (with all four $p < 10^{-16}$ and effect sizes larger than 1, as shown in table 7.4).

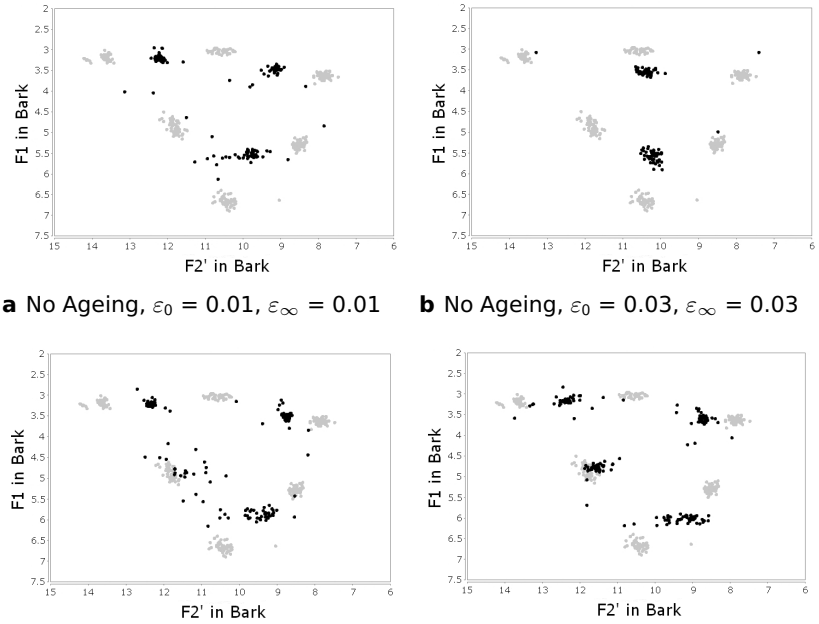
No ageing		Fast ageing	Slow ageing
low ε_0	d	1.2915	1.0052
	CI	0.96376 – 1.73162	0.7016 – 1.3856
high ε_0	d	1.6632	1.4560
	CI	1.3408 – 2.0171	1.1734 – 1.7888

Table 7.4: *Size measure: Effect sizes and confidence intervals.*

In figure 7.4 an example is plotted for each situation with the original vowel system in grey and the newly emerged vowel systems in black. These images also show that overall, the age structure helps to preserve the vowel systems' shape and complexity.

There is disagreement in the literature on the exact slope of the critical period for language acquisition and about whether it is a strict cut-off moment or a more gradual decline (Birdsong, 2005). In the context of the work presented in this chapter, it is not necessary to take a position in this discussion since it is not important for the simulation how exactly learning changes over age, as long as younger learners learn faster than older learners. The critical factor is that there is "a temporal span during which an organism displays heightened sensitivity to certain environmental stimuli" (Birdsong, 2005), in our case to language.

To illustrate that the exact shape of the age structure does not influence the current results, the experiments that were described above were repeated, but this time with a strict cut-off moment instead of a gradual decline in learning ability. All parameter values are the same except for the ones concerning ageing. Both an early critical period (strict cut-off



a No Ageing, $\varepsilon_0 = 0.01$, $\varepsilon_\infty = 0.01$ **b** No Ageing, $\varepsilon_0 = 0.03$, $\varepsilon_\infty = 0.03$

c Fast Ageing, $\varepsilon_0 = 0.03$, $\varepsilon_\infty = 0.01$ **d** Slow Ageing, $\varepsilon_0 = 0.03$, $\varepsilon_\infty = 0.01$

Figure 7.4: *Influence of the age structure on the preservation of vowel systems. An example is plotted for each situation with the original vowel system in grey and the newly emerged vowel systems in black. Note that although in all cases changes occur, they are smaller for the systems with the age structure.*

at $1/5^{th}$ of the expected life time) and a late critical period (at $1/2$ of the expected life time) at which the step size decreased at once to 0, were considered. Table 7.5 shows the results of these experiments. For both the similarity and the distance measure as well as the size, the differences between the conditions with and without critical period are significant (with every $p < 10^{-30}$ and effect sizes above 2, which is so clear that it makes the inclusion of tables for effect sizes and confidence intervals seem superfluous, therefore these have been omitted).

7.6 Discussion

This chapter described a simulation of the emergence of vowel systems, where the model of de Boer (2000) and de Boer and Vogt (1999) was used and the influence of the presence of an age structure on the culturally evolving vowel system could be measured. This model shows very clearly that vowel systems in a changing population remain more complex and are preserved better if agents learn faster when they are younger. The results presented here reproduce the findings in de Boer and Vogt (1999) and show that the same occurs in the case of a differently shaped age

Step size Ageing	$\varepsilon_0 = 0.00$ $\varepsilon_\infty = 0.00$ No CP	$\varepsilon_0 = 0.03$ $\varepsilon_\infty = 0.00$ CP at $\frac{1}{5}$ of lifetime	$\varepsilon_0 = 0.03$ $\varepsilon_\infty = 0.00$ CP at $\frac{1}{2}$ of lifetime	$\varepsilon_0 = 0.03$ $\varepsilon_\infty = 0.03$ No CP
Success:	0.922 $\pm$ 0.009	0.846 $\pm$ 0.010	0.895 $\pm$ 0.018	0.919 $\pm$ 0.010
Energy:	0.65 $\pm$ 0.16	1.29 $\pm$ 0.23	0.95 $\pm$ 0.25	0.59 $\pm$ 0.27
Size:	2.90 $\pm$ 0.11	3.79 $\pm$ 0.25	3.45 $\pm$ 0.33	2.67 $\pm$ 0.44
Similarity:	0.698 $\pm$ 0.020	0.804 $\pm$ 0.021	0.780 $\pm$ 0.029	0.688 $\pm$ 0.038
Distance:	0.835 $\pm$ 0.039	0.394 $\pm$ 0.068	0.567 $\pm$ 0.116	0.858 $\pm$ 0.105

Table 7.5: *Results of the experiments with a strict critical period (CP), going to step size (ε_0) of 0, showing the averages and standard deviations. As in the case with the gradual decline in learning ability, the similarity is higher and distance lower for runs with the age structure: vowel systems are preserved better.*

structure. Vowel systems in a population with age structure stay closer to the initial vowel system in terms of size and shape, and the mutual intelligibility between initial and final generations is higher.

One important basic design problem that is often encountered in the creation of autonomous machines that need to cope with the complexity of the real world, is the stability-plasticity dilemma (Carpenter and Grossberg, 1988): intelligent systems should be plastic enough to be able to adapt quickly to new circumstances and learn new things, while at the same time they should be stable enough to retain the knowledge and skills they already acquired. In language acquisition this dilemma is of equal importance. On the one hand language users should be able to learn a new language when moving to a different community, but on the other hand they need to keep the ability to identify and communicate with people from their own community as well. Carpenter and Grossberg (1988) suggest that mechanisms for self-stabilisation are needed to be able to function in the real world: “non-self-stabilizing learning systems are not capable of functioning autonomously in ill-controlled environments” (Carpenter and Grossberg, 1988). One learning system that possesses such self-organising properties, Hebbian learning (Hebb, 1949), has been applied to the problem of language acquisition (McCandliss et al., 2002; McClelland et al., 1999), showing how it can be easier for such a model to learn new representations in an earlier phase than in a later phase. For language acquisition therefore, it might be that our general self-stabilising learning system results in observable age sensitivity effects, causing enough plasticity to be able to adjust in foreign environments while being able to identify and communicate with people from home as well. The findings presented in this chapter suggest that the stabilisation reaches further than the level of the individual, and that it has additional advantageous consequences on the population level. It was found that vowel systems remain larger in populations with an age structure than in populations without an age structure. The fact that older agents do not adjust as fast as younger agents provides new

agents with a more stable target. This is expected to facilitate acquisition because it causes less confusion and ambiguity in the interactions. In consequence, this makes it possible for the agents to learn more complex systems. Apparently, an age structure can contribute to the (cultural) preservation of complexity of communication systems.

As mentioned before, the presented findings are corroborated by findings from the field of sociolinguistics, where it has been suggested that an age structure plays a role in the preservation of structure in linguistic change. Similar to the results presented here, these findings indicate that stability and regularity are enhanced by an unbroken chain of transmission from adults to children. Similar stability is not found in adult to adult contact only (Labov, 2007).

Although both the sociolinguistic finding by Labov (2007) and the simulation that was presented here point in the same direction it is hard to make a more in-depth comparison at this moment. Both sources clearly indicate that it is important to consider the consequences of language acquisition age effects in the study of language preservation and change, but there are also points of difference.

First, the examples of real-life linguistic change that were used in Labov (2007) to illustrate the difference between transmission and diffusion both involved language contact. In the computer simulations, the populations did not come in contact with other populations, but the changes in the vowel systems happened through cultural transmission only. It might be that language contact introduces other dynamics that would change the results in the current implementation. Therefore, it might be necessary to model the language contact situation as well. With a type of language game that is related to the imitation games, the 'Naming Game', similar experiments have been designed in which a spatial structure exists and language contact is modelled (Steels, 1997a). It remains to be discovered whether the integration of so many different aspects (several open populations, a spatial structure, an age structure and contact) in one simulation run will be suitable to provide much more clarification. This is because it becomes harder and harder to analyse simulations when several levels of complexity interact, but an integration is definitely worth the try.

Second, the linguistic changes in Labov (2007) all involve structures and constraints that are more complex than the level of vowel shifts. They deal with the borrowing of complex rules determining for instance differences in vowel quality connected to lexical constraints. Therefore, to be able to model the same phenomena of linguistic contact, a model with more complex repertoires of shared structures may be necessary. It is expected that the found effect of stabilisation is generalizable to more complex sound systems and other areas of linguistics. In future work, it would therefore be interesting to find out what would happen if the individuals in the population were able to learn more complex sound patterns, such as syllable systems. This could be done, for example, by

combining the experiments of de Boer and Vogt (1999) with syllable production and perception models from Oudeyer (2001, 2006).

Third, the literature indicates that the role children play in language systems is very important in the formation of structure and that their learning is not only faster but also of a different nature. Newport (1988) describes differences between adult and infant learning of American Sign Language and finds that adult learners use different generalisation strategies. The signs produced by adult learners have a more holistic relation to their meanings while infant learners make more use of generalisation and analyse words into smaller structural elements. Hudson Kam and Newport (2005) investigated this difference in more detail in artificial language learning experiments. Both adults and children in these experiments were taught an artificial language which contained 'unpredictable variation': inconsistencies in the language. It turned out that both groups could learn the language, but that children as opposed to adults regularised the inconsistencies. Adults showed probability- matching behaviour, exactly reproducing the variability in the input, while children found (whether they were actually there or not) more general patterns in the input and removed the inconsistencies (Hudson Kam and Newport, 2005). Although for language acquisition, differences in adult and child learning, whatever the underlying mechanisms, are observed to result in faster acquisition in childhood than in adulthood, a model in which adults and children only differ in their speed of acquisition might fail to grasp the exact nature of the actual differences.

In short, more complete models would be needed to simulate Labov's proposal more exactly because the current model might lack some important ingredients. Spatial structure, social structure, meaning and language contact, for instance, are not taken into consideration in the current simulations, which has the advantage of making it easier to analyse the behaviour of the model, but a disadvantage is the loss of realism. However, it is certainly interesting that this model, even with many abstractions, still points in the same direction as findings from real language change and indicates that the age structure in language learning ability plays an important role at the population level of language change and stability.

Using only one specific computer model to show the consequences of an age structure on a culturally evolving communication system may involve the risk of drawing conclusions about findings that are actually the result of behaviours of this particular model and not of the more general phenomenon that is investigated. This is another reason why it is important to replicate these results with other models, not only with more complete or complex versions of the current simulations, but also with different learning mechanisms and vowel representations.

This chapter investigated the consequences that an age structure has on a culturally evolving communication system. It was not shown, however,

where such an age structure might have originated from. The age structures in the model were built-in and therefore this study cannot explain what caused such a structure to come into being. Although it is demonstrated that a change of language learning ability with age has a clear population level benefit, this does not prove that it evolved for this reason. More plausible explanations have already been proposed. It could have been a by-product of evolution under the assumption that language provides a selective advantage (Hurford, 1991), it may be the result of general self-stabilising learning mechanisms (McCandliss et al., 2002; McClelland et al., 1999) or it might have originated from an interaction of several evolutionary and developmental influences.

In summary, the results presented in the preceding sections indicate how the presence of an age structure can have an effect on the cultural evolution of speech. It has been shown that an age structure causes preservation and stabilisation of sound systems in an open population. On a coarse scale, these findings are in line with the observations in real language evolution. However, to make a more legitimate comparison, more complete simulations are desirable.

Discussion

8

The studies described in this thesis have been conducted with the aim to contribute to an understanding of the evolution of speech. This research fits within a framework in which language is considered to be a complex adaptive system (Brighton and Kirby, 2001; Kirby, 2002; Steels, 1997b) which is shaped through complex interactions between different levels of organisation including that of the individual speaker and of the language system as a cultural phenomenon. The focus of the studies presented is on the emergence of combinatorial structure in speech sounds. Two different methods were used in the current research: computer simulations and laboratory experiments with human participants. In chapter 3 a first attempt at experimentally investigating the emergence of combinatorial structure in sound systems was described. In that experiment we could not observe the emergence of combinatorial structure due to limitations with the design of the sound-production interface that appeared to be very difficult to use. In order to deal with the issues that characterised this first experiment, the study described in chapter 4 was conducted. Here, participants produced sounds with a more intuitive device, namely a slide whistle. The improved experiment resulted in a clear emergence of combinatorial structure in artificial whistled languages, showing that this type of structure can emerge through cultural transmission. In chapter 5 a more elaborate analysis of the combinatorial structure in the emerged whistled languages from chapter 4 was presented. By means of additional experiments it was shown that human participants are aware of the regularities in the emerged systems and they are able to use it to distinguish between two different languages. Chapter 6 provided a description of a follow-up experiment involving artificial whistled languages in which the words referred to meanings. Also in this case, combinatorial structure at the level of the signal emerged reliably. Finally, complementing all these studies about emergence, adaptation and change, chapter 7 described a

This chapter contains parts that also appear in the following articles:

Verhoef, T., de Boer, B., & Kirby, S. (2012) Holistic or synthetic protolanguage: Evidence from iterated learning of whistled signals. In *The evolution of language: Proceedings of the 8th international conference (EVOLANG8)*. (pp. 386-375). Hackensack NJ:World Scientific.

Verhoef, T. (2013) Cultural evolution, compression and the brain. *The Past, Present and Future of Language Evolution Research* (to appear).

computational study in which it was shown how cultural phenomena can influence the preservation of structure over generations.

In chapter 1 the sharp contrast between different ideas that relate to the evolution of language was outlined. There are theories that assume language is innate in the form of a language-specific module unique to humans (Chomsky, 1976; Piattelli-Palmarini, 1989; Pinker and Bloom, 1990), which sometimes is assumed to have evolved in human biology through natural selection (Pinker and Bloom, 1990). Pinker and Bloom (1990) mention two criteria from evolutionary theory in support of their idea of language as a biological adaptation that has undergone natural selection. These criteria define when natural selection can be used as a valid explanation for a phenomenon: “complex design for some function, and the absence of alternative processes capable of explaining such complexity”. As we have seen, we currently are in fact aware of alternative processes that can account for the emergence of complex linguistic structures. The process of cultural evolution, in which languages are transmitted from one generation of naïve learners to the next, has been proven to be able to account for “the appearance of design without a designer” as well as biological evolution (Kirby et al., 2008). It has been proposed that language change needs an ‘invisible hand explanation’ (Keller, 1994), a term that originated from the field of economics and was introduced by Adam Smith. This is based on the idea that sometimes individuals, who are not directed to work towards a central goal, behave in a certain way ‘as if they were guided by an invisible hand’, which results in unintended structure at the level of the community. Language can be seen as one such structure as both Keller (1994) and Fitch (2007) pointed out. This view nicely illustrates the importance of considering both the individual micro level and the population macro level in language evolution. The view of cultural transmission as a key process in the way languages are shaped complies with this.

8.1 Main findings

The results presented in this thesis provide additional support for the idea that the process of cultural evolution is important in shaping linguistic structure, by showing that aspects of phonological organisation can emerge as the result of cultural transmission, as opposed to the assumption that humans are born with a set of innate phonetic features as Chomsky (1976) proposed: “languages have a partially determinate structure as a matter of biological necessity, much as the general character of bodily organs is fixed for the species. The theory of distinctive features is perhaps the most familiar case. It has been proposed that a certain set of features is available in principle for phonetic representation; each language must make its selection from among these. (...) it seems to me not unreasonable to approach

the study of language as we would the study of some organ of the body" (Chomsky, 1976, p.46). The results of chapter 4 and chapter 6 demonstrate that cultural evolution can cause a system of random continuous signals to become organised in a way that is very similar to how speech is organised: a small number of basic elements is combined into a larger number of signals, resulting in systems that are more constrained and show transmission-chain specific 'traditions'. The emergence of combinatorial structure in the presented experiments seems to be due to general properties of the way humans learn and generalise signals, which drives the systems to become more predictable and more learnable. The structure in the artificial whistled languages cumulatively developed in adaption to the process of being transmitted from participant to participant, without any influence of communication and without the invention of structure by individual participants. It therefore seems unnecessary to assume that phonetic features are innate.

With respect to current theories in evolutionary phonology, it has been proposed that the emergence of combinatorial structure was driven by vocabulary expansion and dispersion: the limits of the signal space were reached at some point and no more distinguishable holistic signals could be added. Similar ideas have been proposed by Abler (1989); Studdert-Kennedy and Goldstein (2003). However, as mentioned in chapter 2, there are reasons to believe this is not the complete picture, following for instance from the example of ABSL with its high functionality but still emerging combinatorial structure (Sandler et al., 2011). Also, other mechanisms have been proposed in reaction to the fact that dispersion theories do not explain consonant inventories well. Ohala (1980) for instance suggested that the organisation in speech sounds seems to follow a principle of "Maximal use of available distinctive features". This theory, as well as others that are based on similar ideas (Clements, 2003; Maddieson, 1995), focuses on principles of economy and the efficient reuse of basic elements. The experimental data presented in this thesis seems to be more in line with economy principles, showing that combinatorial structure emerged in sets of whistles that were culturally transmitted in absence of pressures from vocabulary growth. The vocabularies in the experiments contained only twelve whistles and in all conditions combinatorial structure emerged long before the signal space had been fully covered. Apparently, a good reason to have combinatorial structure, even for a very simple system, is that a system with such structure is easier to learn and reproduce.

In line with earlier findings on the dynamics of iterated learning (Kirby et al., 2008; Kirby and Hurford, 2002), the whistles that fit the structure and conform to people's cognitive biases are more likely to be preserved from generation to generation in cultural evolution. Combinatorial structure therefore potentially emerged within a gradual cultural evolutionary process. As follows from the results presented in chapter 5, different parallel chains result in whistle languages that are recognisably different in

terms of the specific rules, building blocks and constraints. This further supports the view that the emerging structure in the artificial languages is the result of conventionalisation and emerges through cultural transmission. In the following sections some more general points of discussion are presented involving implications, limitations and future plans for extensions of the work described in this thesis.

8.2 The protolanguage debate

As mentioned in chapter 2, there is an ongoing debate on the nature of a possible ancestral protolanguage. Did human protolanguage consist of holistic utterances (in the form of produced sounds or gestures) that were segmented into words (Arbib, 2005; Wray, 1998) or did it start with simple words that were combined into more complex structures (Bickerton, 1992; Tallerman, 2007)?

Arguing against holistic protolanguage, Tallerman (2007) suggested why holistic protolanguage would be unlikely to lead to compositional syntax. First of all, holistic protolanguage would by definition be irregular. In addition, if it is learned, it is expected to change rapidly over generations, just as modern language does. Tallerman therefore argues that this would not provide a sufficiently stable target for analysis into components. In response to this argument, it has been pointed out that learners often overgeneralise and that inconsistencies in the input do not necessarily prevent the discovery of regularities (Smith, 2008b; Fitch, 2010). This argument is backed up by computer simulations showing that regularisation can indeed happen in cultural transmission (Brighton and Kirby, 2001; Kirby and Hurford, 2002; Kirby et al., 2004), as well as by cultural learning experiments with humans that show emergence of compositional structure from initially holistic sets of utterances (Kirby et al., 2008).

Tallerman's (2007) response to these arguments and to the computational models is that they already assume that there are building blocks: the target words are built up out of a set of discrete segments. She assumes that the task of segmenting signals into relevant elements is impossible if there are no predefined segments, because in that case even the tiniest distinctions between signals should be considered potentially significant. She therefore does not find demonstrations that rely on pre-existing segments convincing.

The experimental results that were presented in chapter 4 and chapter 6 address the above-mentioned concerns. The results show that modern human learners quickly generate structure in initially structureless (holistic) sets of continuous signals. Because the initial set is too difficult to learn precisely, learners tend to overgeneralise the structure they (think they) observe. This introduces reuse of a small set of building blocks and increases the learnability of the set of signals.

Apparently, modern humans have no problems finding (apparent) structure in holistic, continuous utterances. Both experiments, with and without meaning, show that participants introduce combinatorial structure very rapidly and as a system independent of structure in the meaning space. Subsequently, as Kirby et al. (2008) have shown, signals built up of discrete elements with an associated meaning (but without a systematic form-meaning mapping) can transform into systems with a systematic form-meaning mapping.

As for the argument about rapid change: rapid change does happen in the experiments, but it leads to structure and better learnability. This is an example of how repeated introduction of naïve learners with acquisition limitations drives the development of a linguistic system towards being learnable (Zuidema, 2003). Therefore the argument that the changeability of a culturally transmitted holistic system would prevent emergence of structure is not supported by empirical evidence from modern human behaviour.

Of course, these observations do not necessarily generalise to ancestral hominids, who may have had very different cognitive adaptations. However, as Smith (2008b) has pointed out, research with cotton-top tamarins (Hauser et al., 2001) has shown that at least some non-human primates already have simple abilities for segmenting streams of speech. It is therefore possible that the ability to find regularities in speech is much older than the split between humans and the other apes. Given a pre-existing ability to analyse, we can expect that re-use of regularities could have been possible at the earliest stages of protolanguage.

Although the current findings refute arguments against holistic protolanguage, they may also imply that the idea of an extended holistic protolanguage phase is unlikely. It appears that such a system would perhaps not be stable for a very long time, because combinatorial structure at the signal level would emerge rapidly. I would also not exclude the possibility that the holophrases were concatenated into larger constructs simultaneously, following a synthetic scenario in parallel. In fact, it would be possible to have aspects of both holistic and synthetic protolanguage in one system: holistic phrases break up into smaller units, while at the same time words could be combined into short utterances. As Smith (2008a) suggested, there is no fundamental contradiction between these two points of view.

The main aim of this thesis was not to unravel the nature of protolanguage and the above mentioned ideas are still speculative and preliminary. However, it is exciting to note that iterated learning experiments can be used to empirically investigate some issues that are alive in the debate on holistic and synthetic protolanguage. It is therefore no longer necessary to base protolanguage theories solely on conjectures, because relevant data can be obtained even with modern humans.

8.3 Cultural transmission and efficient coding

As reviewed in chapter 2, the influence of cultural evolution on the way linguistic structure emerges is increasingly being studied with the use of laboratory experiments. This method has generated a vast amount of data in the past few years, including the results presented in this thesis. The emergence of compressible and predictable systems appears to be a prevalent result of cultural transmission experiments and the results in this thesis are no exception. The emergence of combinatorial structure in the sets of whistles forms an additional example. In these systems, whether they include meanings or not, discrete sets of basic building blocks could be identified in the sounds and these were reused and combined in a predictable way. Quantitatively, a cumulative decrease of entropy over the reuse of basic elements could be measured in the languages, indicating that equally large languages could be described using fewer basic elements. The whistled systems therefore became more constrained and more compressible.

In chapter 2 a short overview was presented of advances in the field of computational neuroscience in which principles of compression and simplicity in neural processing are studied. As we have seen, it can be proven that many features of natural stimuli are optimally efficiently encoded in both the visual and auditory cortices. The study by Smith and Lewicki (2006) in particular is interesting in the context of language evolution, because here it was shown that the auditory cortex of a cat optimally encodes the sounds of speech. This provides convincing evidence in favour of the view that the sounds used in language are adapted to the (mammalian) auditory cortex. This is in line with the suggestion that transmitted systems adapt to human biases and constraints (Christiansen and Chater, 2008; Deacon, 1997; Griffiths and Kalish, 2007; Kirby and Hurford, 2002). It is unlikely that cat auditory processing has evolved to efficiently encode human speech, therefore a more plausible assumption would be that the sounds used in speech have adapted to be efficiently coded by the brain. Likewise, it is expected that linguistic structure at other levels of organisation has adapted to general cognitive ‘simplicity’ biases and is shaped in such a way that it is compressible. The study by Smith and Lewicki (2006) provides an exciting example of more direct evidence of adaptation through cultural evolution. Even though this has so far only been shown for very early processing and sound primitives for speech, it is a promising avenue for further research. Following this direction we should try to formulate experiments and create biologically plausible models that can provide this kind of evidence for other levels of organisation in linguistic structure as well.

As Deacon (2009; 1997) argues, researchers have not been able to associate human linguistic behaviour with a unique change or difference in brain anatomy as compared to non-human ancestors. Instead it is likely that a large variety of systems, with perhaps different

functions in our ancestors, contributed to and are involved in modern human linguistic behaviour. The study by Smith and Lewicki (2006) is a brilliant example of how such a homologous system (involving auditory processing in this case) can be linked to efficient coding of speech sounds in non-human species. There may be other aspects of language processing and learning for which it is possible to demonstrate preferences or efficient coding in homologous systems inside non-human mammalian brains. The method of demonstrating such efficiency by predicting properties of measurable brain responses through computational modelling of optimally efficient coding is a path that deserves exploring. Especially in the case that we can show this effect for cognitive processing of (linguistic) compositional and combinatorial structure, this would be compelling evidence against language-specific biological adaptations and must indicate a strong influence of general cognition and cultural evolution. In addition, this may be a direction that can potentially reveal relevant differences between human and non-human processing. Perhaps it is therefore fruitful to consider an integrative framework combining the study of cultural transmission, the systems that emerge from it and the neuroscientific study of efficient coding in the brain.

8.4 Possible concerns

A possible concern with the current results, if we were to consider this experiment as a reconstruction of language evolution, could be that we use modern human participants who obviously have modern cognitive adaptations unlike our ancestors. This fact is shared among all language evolution experiments that make use of human participants, but should not necessarily be viewed as problematic. As Scott-Phillips and Kirby (2010) point out, the results of this type of work should not be interpreted as an attempt to reconstruct the emergence of linguistic structure, or, in this case of structure in speech sounds, but as a method to shed light on what mechanisms may be involved in this emergence. The current work is meant to illustrate how human cognitive biases influence a sound system when it is repeatedly transmitted to new learners and what role these biases play in the maintenance of combinatorial structure. In addition, when experiments such as the ones presented in this thesis are paired with results from computer simulations, a stronger point can be made. As reviewed in chapter 2, the iterated learning model has been studied with a variety of learning mechanisms (Kirby, 2002). In these simulations naïve agents are used that obviously do not have language built in and also do not have any experience with language prior to the model runs. Still, the results are very similar to what has been found with modern humans in the laboratory. The computer simulation that was presented in this thesis is not directly modelled after the specific experiments that were conducted for this research, but this is planned in the future continuation of this work.

Another concern that has been expressed in response to the experiment described in chapter 4 involves the lack of meanings conveyed by the whistled signals. The design of that early experiment abstracted away from full human semantic complexity by not having an explicit meaning connected to the whistles. As was pointed out in chapter 4, the systems do in a way acquire meaning when the experiment progresses because of the fact that participants have to reproduce a complete set of twelve whistles. The languages therefore have some degree of expressivity. As an adaptation to the learning constraint, the whistles evolve in a way that makes them share more and more features or building blocks. This makes it possible to remember the signals as subsets, which makes learning and recall easier. The idea that chunking of information in this way facilitates encoding more information in short-term memory is well established Miller (1956). Participants tended to categorise the whistles as subsets, such as 'the ones that all start with a falling slide'. This first investigation of combinatorial structure in a set of whistles without referents was necessary to be able to control for effects of semantics such as iconicity or compositional structure. With such influences present it would be harder to distinguish whether the emerging structure relates to the structure of the meaning space or whether they are truly meaningless units being recombined. In addition it would be harder to know what drove the emergence of structure. Chapter 6 presents experiments in which meaning was added and the analysis of the emerging whistled languages indeed proved to be non-trivial. However, together the two experiments have already provided interesting new data for studies about the emergence of phonology. The results conform to the idea that phonology is an autonomous system with generative power (Studdert-Kennedy and Goldstein, 2003) and they also show that combinatorial structure is not necessarily linked directly to vocabulary size or driven by signal distinctiveness.

It could be argued that the first whistle experiment has little to do with language and should instead perhaps be compared with musical systems because of its lack of meaning. I would like to stress that in my opinion, the observation that the systems emerging in the whistle experiments show characteristics of musical structure (whether this is actually the case or not) should not at all be a reason to conclude it is uninteresting for research on language. In fact, I would argue that the experiment would have been equally successful and equally relevant for theories on the emergence of phonology if the conclusion had been that a discrete set of basic notes or rhythmic primitives had emerged that were combined into different musical styles in the four chains. As Fitch (2006,2010) argues, music and (spoken) language share many structural characteristics, especially at the level of phonology. Music and phonology are both built from a discrete set of basic meaningless primitives that are combined in a generative way to construct an unlimited number of signals and both are also culturally transmitted (Fitch,2006; 2010). An important difference is the role of meanings,

which is much more prominent in language and part of a systematic organisation. In music the form has a more holistic and affective relation to meaning and subtle differences in for instance expressive timing (Honing, 2002) may lead to different interpretations. Another point of difference is the nature of the discrete elements. In music, features such as pitch and rhythm or other temporal features are the elements that are discretised, constrained and regularised, while these features tend to be more variable in speech (except for pitch in tonal languages). In speech, the most important elements of recombination tend to be vowels and consonants, which are not normally associated with music. However, in terms of their combinatorial structure and the way this structure forms part of a transmitted cultural tradition, music and phonology seem to be very parallel (Fitch,2006; 2010). This has also become clear from data on experiments that explicitly compare the two domains with neuroimaging and show there is significant overlap in the processing of language and music (Patel, 2003; 2012).

To further illustrate how subtle the boundaries are between music and phonology, perhaps Pirahã provides an interesting example of a language where phonological contrasts could perhaps be considered to involve music-like as well as language-like features. As Everett described for Pirahã, vowels and consonants seem to play a less important role in this language than patterns of tone, timing and stress (Everett, 1985). This is probably related to the fact that this language is used over several different channels, one of which is hummed speech. Humming is often used there in intimate, close-contact situations such as mother-child interaction. Children also apparently acquire control over the prosodic structure of the language earlier than the specific vowels and consonants (Everett, 1985). In absence of the knowledge that a rich, unrestricted repertoire of meanings can be conveyed with the hummed speech system, perhaps an outside observer may classify it as music based solely on its form.

In summary, sometimes it can be informative to abstract away from the full complexity of language and focus on a specific structural property to learn more about the way it can emerge. The fact that this particular type of structure is shared with other systems such as music (and many other culturally transmitted systems such as dance and art I would say) should not lead to the conclusion that the results are less interesting for language. Instead, as Fitch (2010) also proposed, the parallels between different domains should be used to our advantage in research about the evolution of language and other systems.

8.5 Plans for the future

The studies presented in this thesis together represent an exploratory investigation which involved the adaptation of existing methods that were recently introduced into the field to make them applicable to

questions on the evolution of speech. As reflected in the results that were presented in the different chapters, many issues were resolved and interesting insights obtained. However, many questions remain unanswered or require further investigations. In this section I summarise some of my plans for the future continuation of this work.

8.5.1 Experimental designs

One limitation of the iterated learning experiments described in this thesis is that the generations consisted of only one participant. A first problem with this is that it is obviously not realistic, because real speech communities necessarily consist of more than one speaker. Another problem is that this design ignores the importance of interaction and communication. Participants learn a language, but they do not use it for communication with others. This has been a very important abstraction to be able to demonstrate that linguistic structure can emerge as an adaptation to a transmission bottleneck, independent of communication and without the conscious creation of structure by individuals. It was therefore sensible and necessary to start this way. However, the lack of a pressure for communication is probably the reason why expressivity needs to be maintained in an artificial way in experiments with this design, such as the filtering technique used by Kirby et al. (2008) and my reproduction constraint. Changing the design therefore seems desirable.

Other designs for iterated learning experiments have been explored already, although these have not been applied to the study of structure in continuous auditory signals yet. Tamariz et al. (2012) for instance created transmission chains in which two participants formed each generation. They showed that, when these *dyads* interacted with each other and could negotiate to arrive at new versions of the artificial language together, the structure increased more than without such interaction. Tan and Fay (2011) also showed that interaction improves faithful transmission in chains where participants had to communicate a description of an event to another person. In addition, the iterated learning paradigm has now been extended by Caldwell and Smith (2012) to involve *microsocieties*. Here, groups of four participants communicate about meanings with drawings and there is a gradual progression from generation to generation in which the most experienced participant is replaced by a new participant. These designs allow to carefully investigate effects of both conventionalisation and cultural transmission on the way signalling systems are shaped and I expect that these innovations will become more widely used in the future.

The work in this thesis had a strong focus on speech, and therefore on sound systems and structure therein. It would however be fruitful to extend the work to other modalities as well, since combinatorial structure also clearly plays a role in for instance sign languages. del Giudice et al. (2010) used a different modality by studying the emergence of combinatorial structure in transmitted graphical systems.

A study on manual systems seems to be a promising addition to these existing studies. Such an experiment would study with non-signing participants whether and under what conditions a random set of gestures evolves into a system that shows regularities similar to those found in sign languages. This approach would address the role of embodiment in producing signals. Another advantage of this approach would be that it allows a more direct comparison of experimental results with data from Al-Sayyid Bedouin Sign Language, the only known language in which regularised combinatorial structure is still emerging. In addition, given the direct visual-to-visual mapping, sign languages are more conducive to iconic expressions (Perniss et al., 2010). The manual modality may therefore be more suitable and would perhaps allow for a more natural investigation into questions on iconicity.

8.5.2 Neuroscience-inspired computer model

Experimental work in language evolution is often modelled after designs and findings that were obtained with the use of computer simulations. Scott-Phillips and Kirby (2010) review examples of this. The two methods nicely complement each other. Human participants in an experiment are unavoidably modern humans who may not have the same cognitive abilities as our ancestors and may be biased by their linguistic experience. Biases of computer learners can be controlled, but computer models have been criticised to be less realistic. Computer models provide a possibility for taking a bottom-up approach to explore the cognitive biases needed to explain behaviours found in the laboratory. However, no simulation exists that is directly comparable to the iterated learning experiments presented in this thesis. The computer model presented in chapter 7 is less suitable for a direct comparison to the experimental findings because of the focus on a different population structure and sound system. The next step would therefore involve the design of a computational model that may be able to explain the observed patterns in iterated learning experiments with continuous signals. To be able to explain the emergence of efficient and compressible representations in languages, perhaps a link should be formed with neuroscience-inspired models on efficient coding strategies in the human brain, as reviewed in chapter 2 and mentioned above.

8.5.3 Conclusion

The possible follow-up studies proposed above cover only a small subset of all the possible open questions that deserve further investigation. With the work in this thesis I have merely set the first steps towards developing a paradigm that studies evolutionary phonology empirically. We now know that the same methods that were previously used successfully to study aspects of language that can be represented with discrete symbols such as compositional structure, are useful for studying combinatorial structure in sound systems as well. I expect that many more insights

can be gained in the future with an approach in which computational simulations and laboratory experiments are used hand in hand to unravel the origins of structure in linguistic systems of continuous signals.

References

Abler, W. L. (1989). On the particulate principle of self-diversifying systems. *Journal of Social and Biological Structures*, 12(1):1–13.

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

Arbib, M. (2005). From monkey-like action recognition to human language: An evolutionary framework for neurolinguistics. *Behavioral and brain sciences*, 28(2):105–124.

Arbib, M. A. and Bickerton, D., editors (2008). *Holophrasis vs Compositionality in the Emergence of Protolanguage*, volume 9, number 1 of *Interaction Studies*. Special Issue. John Benjamins.

Arvaniti, A., Ladd, D., and Mennen, I. (1998). Stability of tonal alignment: the case of greek prenuclear accents. *Journal of phonetics*, 26(1):3–25.

Baken, R. J. and Orlikoff, R. F. (2000). *Clinical measurement of speech and voice*. Thomson Learning, 2 edition.

Bakker, P. (1994). Pidgins. In Arends, J., Muysken, P., and Smith, N., editors, *Pidgins and Creoles: An Introduction*, pages 26–39. John Benjamins.

Barlow, H. (1961). Possible principles underlying the transformation of sensory messages. *Sensory communication*, pages 217–234.

Bates, D., Maechler, M., and Bolker, B. (2013). *lme4: Linear mixed-effects models using S4 classes*. R package: http://CRAN.R-project.org/package=lme4.

Beckner, C., Blythe, R., Bybee, J., Christiansen, M. H., Croft, W., Ellis, N. C., Holland, J., Ke, J., Larsen-Freeman, D., and Schoenemann, T. (2009). Language is a complex adaptive system: Position paper. *Language Learning*, 59(s1):1–26.

Bergen, B. (2004). The psychological reality of phonaesthemes. *Language*, 80(2):290–311.

Berrah, A.-R. and Laboissière, R. (1997). Phonetic code emergence in a society of speech robots: Explaining vowel systems and the muaf principle. In Kokkinakis, G., Fakotakis, N., and Dermatas, E., editors, *Eurospeech 97: 5th european conference on speech communication*

and technology, volume 4, pages 2395–2398.

Bickerton, D. (1992). *Language and species*. University of Chicago Press.

Bickerton, D. (2007). Language evolution: A brief guide for linguists. *Lingua*, 117(3):510–526.

Birdsong, D. (2005). Interpreting age effects in second language acquisition. In Kroll, J. and DeGroot, A., editors, *Handbook of Bilingualism: Psycholinguistic Perspectives*, pages 109–127. Oxford University Press, New York.

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

Brighton, H. (2002). Compositional syntax from cultural transmission. *Artificial Life*, 8(1):25–54.

Brighton, H. (2005). Compositionality, linguistic evolution, and induction by minimum description length. *The Compositionality of Meaning and Content: Applications to linguistics, psychology and neuroscience*, 2:13.

Brighton, H. and Kirby, S. (2001). The survival of the smallest: Stability conditions for the cultural evolution of compositional language. *Advances in artificial life*, pages 592–601.

Caldwell, C. A. and Smith, K. (2012). Cultural evolution and perpetuation of arbitrary communicative conventions in experimental microsocieties. *PloS one*, 7(8):e43807.

Carlson, N., Ming, V., and DeWeese, M. (2012). Sparse codes for speech predict spectrotemporal receptive fields in the inferior colliculus. *PLoS Computational Biology*, 8(7):e1002594.

Carpenter, G. A. and Grossberg, S. (1988). The art of adaptive pattern recognition by a self-organizing neural network. *Computer*, 21(3):77–88.

Chater, N. and Vitányi, P. (2003). Simplicity: A unifying principle in cognitive science? *Trends in cognitive sciences*, 7(1):19–22.

Chomsky, N. (1976). On the nature of language. *Annals of the New York Academy of Sciences*, 280(1):46–57.

Christiansen, M. H. (2000). Using artificial language learning to study language evolution: Exploring the emergence of word order universals. In *The evolution of language: 3rd international conference*, pages 45–48.

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

Clark, R. (1994). Kolmogorov complexity and the information content of parameters. *IRCS Technical Reports Series*, 163:1–43.

Clarke, E., Reichard, U. H., and Zuberbühler, K. (2006). The syntax and meaning of wild gibbon songs. *PLoS One*, 1(1):e73.

Clements, G. (2003). Feature economy in sound systems. *Phonology*, 20(3):287–333.

Cohen, J. (1992). Statistical power analysis. *Current Directions in Psychological Science*, 1(3):98–101.

Corina, D. and Sandler, W. (1993). On the nature of phonological structure in sign language. *Phonology*, 10(2):165–207.

Cornish, H., Kirby, S., and Christiansen, M. H. (2010). The emergence of structure from sequence memory constraints in cultural transmission. In Smith, A. D. M., Schouwstra, M., de Boer, B., and Smith, K., editors, *The Evolution of Language: Proceedings of the 8th International Conference*, pages 387–388. World Scientific Press.

Curtiss, S., Fromkin, V., Krashen, S., Rigler, D., and Rigler, M. (1974). The linguistic development of genie. *Language*, 50(3):528–554.

Cuskley, C. and Kirby, S. (2013). *Synesthesia, cross-modality, and language evolution*, chapter 43, pages 869–899.

de Boer, B. (1997). Self-organisation in vowel systems through imitation. In Coleman, J., editor, *Computational Phonology, Third Meeting of the ACL Special Interest Group in Computational Phonology*, pages 19–25.

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

de Boer, B. (2009). Acoustic analysis of primate air sacs and their effect on vocalization. *The Journal of the Acoustical Society of America*, 126(6):3329–3343.

de Boer, B. (2012). Loss of air sacs improved hominin speech abilities. *Journal of human evolution*, 62(1):1–6.

de Boer, B., Sandler, W., and Kirby, S. (2012). New perspectives on duality of patterning: Introduction to the special issue. *Language and Cognition*, 4(4):251–259.

de Boer, B. and Verhoef, T. (2012). Language dynamics in structured form and meaning spaces. *Advances in Complex Systems*, 15(3&4):1150021–1–1150021–20.

de Boer, B. and Vogt, P. (1999). Emergence of speech sounds in changing populations. In Floreano, D., Nicoud, J. D., and Mondada, F., editors, *Advances in Artificial Life*, Lecture Notes in Artificial Intelligence 1674, pages 664–673. Berlin Springer Verlag.

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

De Cheveigné, A. and Kawahara, H. (2002). Yin, a fundamental frequency estimator for speech and music. *The Journal of the Acoustical Society of America*, 111:1917–1930.

Deacon, T. (2009). The evolution of language systems in the human brain. In Kaas, J., editor, *Evolutionary Neuroscience*, pages 897–916. Academic Press.

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

del Giudice, A. (2012). The emergence of duality of patterning through iterated learning: Precursors to phonology in a visual lexicon. *Language and cognition*, 4(4):381–418.

del Giudice, A., Kirby, S., and Padden, C. (2010). Recreating duality of patterning in the laboratory: a new experimental paradigm for studying emergence of sublexical structure. In Smith, A. D. M., Schouwstra, M., de Boer, B., and Smith, K., editors, *The Evolution of*

Language: Proceedings of the 8th International Conference, pages 399–400. World Scientific Press.

Dingemanse, M. (2011). Ezra pound among the mawu. *Semblance and Signification*, 10:39.

Dingemanse, M. (2012). Advances in the cross-linguistic study of ideophones. *Language and Linguistics Compass*, 6(10):654–672.

Doupe, A. J. and Kuhl, P. K. (1999). Birdsong and human speech: common themes and mechanisms. *The Annual Review of Neuroscience*, 22:567–631.

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

Duda, R. O., Hart, P. E., and Stork, D. G. (2001). *Pattern Recognition*. New York: A Wiley-Interscience.

Evans, N. and Levinson, S. C. (2009). The myth of language universals: Language diversity and its importance for cognitive science. *Behavioral and Brain Sciences*, 32(05):429–448.

Everett, D. L. (1985). Syllable weight, sloppy phonemes, and channels in pirahã discourse. In *Proceedings of the Berkeley Linguistics Society*, volume 11, pages 408–416.

Fay, N. and Lim, S. (2010). From hand to mouth: An experimental simulation of language origin. In Smith, A. D. M., Schouwstra, M., de Boer, B., and Smith, K., editors, *The Evolution of Language: Proceedings of the 8th International Conference*, pages 401–402. World Scientific Press.

Feldman, J. (2000). Minimization of boolean complexity in human concept learning. *Nature*, 407(6804):630–632.

Fischer, O. and Nänny, M. (1999). *Introduction: Iconicity as a creative force in language use.*, pages 15–36. Benjamins.

Fisher, S. E. and Marcus, G. F. (2006). The eloquent ape: genes, brains and the evolution of language. *Nature Reviews Genetics*, 7(1):9–20.

Fitch, W. (2000). The evolution of speech: a comparative review. *Trends in cognitive sciences*, 4(7):258–267.

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

Fitch, W. T. (2006). The biology and evolution of music: A comparative perspective. *Cognition*, 100(1):173–215.

Fitch, W. T. (2007). Linguistics: an invisible hand. *Nature*, 449(7163):665–667.

Fitch, W. T. and Friederici, A. D. (2012). Artificial grammar learning meets formal language theory: an overview. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 367(1598):1933–1955.

Flege, J. E. (1999). Age of learning and second language speech. In Birdsong, D. P., editor, *Second language acquisition and the critical period hypothesis*, pages 101–131. Lawrence Erlbaum.

Flege, J. E., Munro, M. J., and MacKay, I. R. (1995). Factors affecting strength of perceived foreign accent in a second language. *The Journal of the Acoustical Society of America*, 97(5 Pt 1):3125–3134.

Fox, A. (1995). *Linguistic reconstruction: an introduction to theory and method*. Oxford University Press, USA.

Friston, K. (2010). The free-energy principle: a unified brain theory? *Nature Reviews Neuroscience*, 11(2):127–138.

Galantucci, B. (2005). An experimental study of the emergence of human communication. *Cognitive Science*, 29:737–767.

Galantucci, B., Kroos, C., and Rhodes, T. (2010). The effects of rapidity of fading on communication systems. *Interaction Studies*, 11(1):100–111.

Garrod, S., Fay, N., Lee, J., Oberlander, J., and MacLeod, T. (2007). Foundations of representation: where might graphical symbol systems come from? *Cognitive Science*, 31(6):961–987.

Garrod, S., Fay, N., Rogers, S., Walker, B., and Swoboda, N. (2010). Can iterated learning explain the emergence of graphical symbols? *Interaction Studies*, 11(1):33–50.

Gergely, G. and Csibra, G. (2006). Sylvia's recipe: The role of imitation and pedagogy in the transmission of cultural knowledge. *Roots of human sociality: Culture, cognition, and human interaction*, pages 229–255.

Goldin-Meadow, S., Mylander, C., and Butcher, C. (1995). The resilience of combinatorial structure at the word level: Morphology in self-styled gesture systems. *Cognition*, 56(3):195–262.

Goodman, N. D., Tenenbaum, J. B., Feldman, J., and Griffiths, T. L. (2008). A rational analysis of rule-based concept learning. *Cognitive Science*, 32(1):108–154.

Griffiths, T., Christian, B., and Kalish, M. (2008a). Using category structures to test iterated learning as a method for identifying inductive biases. *Cognitive Science*, 32(1):68–107.

Griffiths, T., Christian, B., and Kalish, M. (2008b). Using category structures to test iterated learning as a method for identifying inductive biases. *Cognitive Science*, 32(1):68–107.

Griffiths, T. and Kalish, M. (2007). Language evolution by iterated learning with bayesian agents. *Cognitive Science*, 31(3):441–480.

Griffiths, T., Kalish, M., and Lewandowsky, S. (2008c). Theoretical and empirical evidence for the impact of inductive biases on cultural evolution. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 363(1509):3503–3514.

Hall, R. (1962). The life cycle of pidgin languages. *Lingua*, 11(1):151–156.

Hare, M. and Elman, J. L. (1995). Learning and morphological change. *Cognition*, 56(1):61–98.

Harlow, H. F., Dodsworth, R. O., and Harlow, M. K. (1965). Total social isolation in monkeys. *Proceedings of the National Academy of Sciences of the United States of America*, 54(1):90–97.

Harnad, S., Steklis, H., and Lancaster, J. (1976). Origins and evolution of language and speech. *Annals of the New York Academy of Sciences*, (280).

Hauser, M., Newport, E., and Aslin, R. (2001). Segmentation of the speech stream in a non-human primate: statistical learning in cotton-top

tamarins. *Cognition*, 78(3):B53–B64.

Hebb, D. O. (1949). *The organization of behavior*. Wiley, New York.

Hess, E. H. (1958). Imprinting in animals. *Scientific American*, 198(3):81–90.

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

Honing, H. (2002). Structure and interpretation of rhythm and timing. *Tijdschrift voor Muziektheorie*, 7(3):227–232.

Horner, V. and Whiten, A. (2005). Causal knowledge and imitation/emulation switching in chimpanzees (pan troglodytes) and children (homo sapiens). *Animal cognition*, 8(3):164–181.

Hubbard, T. L. (1996). Synesthesia-like mappings of lightness, pitch, and melodic interval. *The American Journal of Psychology*, 109(2):219–238.

Hubel, D. H. and Wiesel, T. N. (1970). The period of susceptibility to the physiological effects of unilateral eye closure in kittens. *J Physiol*, 206(2):419–436.

Hudson Kam, C. L. and Newport, E. L. (2005). Regularizing unpredictable variation: The roles of adult and child learners in language formation and change. *Language Learning and Development*, 1(2):151–195.

Hurford, J. R. (1991). The evolution of the critical period for language acquisition. *Cognition*, 40(3):159–201.

Hurford, J. R. (2011). *The Origins of Grammar: Language in the Light of Evolution II*. Oxford: Oxford University Press.

Hurford, J. R. and Kirby, S. (1999). Co-evolution of language-size and the critical period. In Birdsong, D. P., editor, *Second language acquisition and the critical period hypothesis*, pages 39–63. Lawrence Erlbaum.

Imai, M., Kita, S., Nagumo, M., and Okada, H. (2008). Sound symbolism facilitates early verb learning. *Cognition*, 109(1):54–65.

Immelmann, K. and Suomi, S. J. (1981). Sensitive phases in development. In Immelmann, K., Barlow, G. W., Petrinovich, L., and Main, M., editors, *Behavioral Development: The Bielefeld Interdisciplinary Project*, pages 395–431. Cambridge University Press.

Israel, A. and Sandler, W. (2011). Phonological category resolution in a new sign language: A comparative study of handshapes. *Formational units in sign languages*, pages 177–202.

Janik, V. M. and Slater, P. J. (1998). Context-specific use suggests that bottlenose dolphin signature whistles are cohesion calls. *Animal behaviour*, 56(4):829–838.

Johnson, J. S. and Newport, E. L. (1989). Critical period effects in second language learning: the influence of maturational state on the acquisition of english as a second language. *Cognit Psychol*, 21(1):60–99.

Keller, R. (1994). *On Language Change: The Invisible Hand in Language*. Routledge.

Keller, R. (1998). *Theory of linguistic signs*. Oxford University Press.

Kelley, K. (2005). The effects of nonnormal distributions on confidence intervals around the standardized mean difference: Bootstrap and parametric confidence intervals. *Educational and Psychological*

Measurement, 65(1):51–59.

Keogh, E. and Pazzani, M. (2001). Derivative dynamic time warping. In *the 1st SIAM International Conference on Data Mining (SDM-2001), Chicago, IL, USA*.

Kirby, S. (2000). Syntax without natural selection: How compositionality emerges from vocabulary in a population of learners. pages 303–323.

Kirby, S. (2001). Spontaneous evolution of linguistic structure-an iterated learning model of the emergence of regularity and irregularity. *IEEE Transactions on Evolutionary Computation*, 5(2):102–110.

Kirby, S. (2002). Natural language from artificial life. *Artificial life*, 8(2):185–215.

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):10681–10686.

Kirby, S., Dowman, M., and Griffiths, T. L. (2007). Innateness and culture in the evolution of language. *Proceedings of the National Academy of Sciences*, 104(12):5241–5245.

Kirby, S. and Hurford, J. R. (2002). The emergence of linguistic structure: An overview of the iterated learning model. In *Simulating the evolution of language*, pages 121–147. Springer.

Kirby, S., Smith, K., and Brighton, H. (2004). From ug to universals: Linguistic adaptation through iterated learning. *Studies in Language*, 28(3):587–607.

Knowlton, B. and Squire, L. (1994). The information acquired during artificial grammar learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 20(1):79.

Komarova, N. L. and Nowak, M. A. (2001). Natural selection of the critical period for language acquisition. In *Proceedings: Biological Sciences*, volume 268, pages 1189–1196. The Royal Society.

Kuhl, P. K. (2000). A new view of language acquisition. *Proceedings of the National Academy of Sciences*, 97(22):11850–11857.

Labov, W. (2007). Transmission and diffusion. *Language*, 83(2):344–387.

Ladd, D. R. (2012). What is duality of patterning, anyway? *Language and cognition*, 4(4):261–273.

Lenneberg, E. H. (1969). On explaining language. *Science (New York, N.Y.)*, 164(880):635–643.

Lewicki, M. (2002). Efficient coding of natural sounds. *Nature neuroscience*, 5(4):356–363.

Liljencrants, J. and Lindblom, B. (1972). Numerical simulation of vowel quality systems: The role of perceptual contrast. *Language*, 48(4):839–862.

Lorenz, K. Z. (1937). The companion in the bird's world. *The Auk*, 54(3):245–273.

Lupyan, G. and Dale, R. (2010). Language structure is partly determined by social structure. *PloS one*, 5(1):e8559.

Maddieson, I. (1995). Gestural economy. In Elenius, K. and Branderud, P.,

editors, *Proceedings of the 13th International Congress of Phonetic Sciences. Volume 4*, pages 574–577. Stockholm: KTH Stockholm University.

Martinet, A. (1984). Double articulation as a criterion of linguisticity. *Language Sciences*, 6(1):31–38.

Mayberry, R. I. (1993). First-language acquisition after childhood differs from second-language acquisition: the case of american sign language. *Journal of speech and hearing research*, 36(6):1258–1270.

Mayberry, R. I. (2007). When timing is everything: Age of first-language acquisition effects on second-language learning. *Applied Psycholinguistics*, 28:537–549.

McCandliss, B. D., Fiez, J. A., Protopapas, A., Conway, M., and McClelland, J. L. (2002). Success and failure in teaching the [r]–[l] contrast to japanese adults: Tests of a hebbian model of plasticity and stabilization in spoken language perception. *Cognitive, Affective, and Behavioral Neuroscience*, 2:89–108.

McClelland, J. L., Thomas, A. G., McCandliss, B. D., and Fiez, J. A. (1999). Understanding failures of learning: Hebbian learning, competition for representational space, and some preliminary experimental data. *Progress in brain research*, 121:75–80.

Meadow, S. G. (1978). Review: A study in human capacities. *Science*, 200(4342):649–651.

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

Monaghan, P., Christiansen, M. H., and Fitneva, S. A. (2011). The arbitrariness of the sign: Learning advantages from the structure of the vocabulary. *Journal of Experimental Psychology*, 140(3):325.

Müller, F. M. (1861). The theoretical stage, and the origin of language. *Lectures on the Science of Language*, pages 108–122.

Newmeyer, F. J. (1992). Iconicity and generative grammar. *Language*, 68(4):756–796.

Newport, E. (1988). Constraints on learning and their role in language acquisition: Studies of the acquisition of american sign language. *Language Sciences*, 10(1):147–172.

Nowak, M. A., Krakauer, D., and Dress, A. (1999). An error limit for the evolution of language. *Proceedings of the Royal Society of London*, 266:2131–2136.

Ohala, J. J. (1980). Moderator's introduction to symposium on phonetic universals in phonological systems and their explanation. In *Proceedings of the 9th International Congress of Phonetic Sciences, Volume 3*, pages 181–185. Copenhagen, Institute of Phonetics.

Olshausen, B. and Field, D. (1996). Emergence of simple-cell receptive field properties by learning a sparse code for natural images. *Nature*, 381:607–609.

Olshausen, B. and Field, D. (1997). Sparse coding with an overcomplete basis set: A strategy employed by v1? *Vision research*, 37(23):3311–

3325.

Olshausen, B. and Field, D. (2004). Sparse coding of sensory inputs. *Current opinion in neurobiology*, 14(4):481–487.

Onnis, L., Roberts, M., and Chater, N. (2002). Simplicity: A cure for overgeneralizations in language acquisition? In *Proceedings of the Twenty-Fourth Annual Conference of the Cognitive Science Society*, pages 720–725. Hillsdale, NJ: Lawrence Erlbaum Associates.

Oudeyer, P. (2001). The origins of syllable systems: an operational model. In *proceedings of the International Conference on Cognitive Science*, pages 744–749. Lawrence Erlbaum.

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

Oyama, S. (1976). A sensitive period for the acquisition of a nonnative phonological system. *Journal of Psycholinguistic Research*, 5(3):261–283.

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

Patel, A. D. (2003). Language, music, syntax and the brain. *Nature neuroscience*, 6(7):674–681.

Patel, A. D. (2012). Language, music, and the brain: A resource-sharing framework. *Language and music as cognitive systems*, pages 204–223.

Payne, R. and Mcvay, S. (1971). Songs of humpback whales. *Science*, 173(3997):585–597.

Perniss, P., Thompson, R., and Vigliocco, G. (2010). Iconicity as a general property of language: evidence from spoken and signed languages. *Frontiers in Psychology*, 1(227).

Piattelli-Palmarini, M. (1989). Evolution, selection and cognition: From âĂIJlearningâĂİ to parameter setting in biology and in the study of language. *Cognition*, 31(1):1–44.

Pinker, S. and Bloom, P. (1990). Natural language and natural selection. *Behavioral and Brain Sciences*, 13(4):707–784.

Quinn, M. (2001). Evolving communication without dedicated communication channels. *Advances in Artificial Life*, pages 357–366.

Ramachandran, V. and Hubbard, E. (2001). Synaesthesia–a window into perception, thought and language. *Journal of Consciousness Studies*, 8(12):3–34.

Reali, F. and Griffiths, T. (2009). The evolution of frequency distributions: Relating regularization to inductive biases through iterated learning. *Cognition*, 111(3):317–328.

Roberts, G. and Galantucci, B. (2012). The emergence of duality of patterning: Insights from the laboratory. *Language and cognition*, 4(4):297–318.

Saffran, J. R., Aslin, R. N., and Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science*, 274(5294):1926–1928.

Sakoe, H. and Chiba, S. (1978). Dynamic programming algorithm optimization for spoken word recognition. *Acoustics, Speech and*

Signal Processing, IEEE Transactions on, 26(1):43–49.

Sandler, W., Aronoff, M., Meir, I., and Padden, C. (2011). The gradual emergence of phonological form in a new language. *Natural language & linguistic theory*, 29(2):503–543.

Sankoff, G. and Laberge, S. (1974). *On the acquisition of native speakers by a language*, pages 73–84.

Schmidhuber, J. (2009). Driven by compression progress: A simple principle explains essential aspects of subjective beauty, novelty, surprise, interestingness, attention, curiosity, creativity, art, science, music, jokes. *Anticipatory Behavior in Adaptive Learning Systems*, pages 48–76.

Schouwstra, M. (2012). *Semantic structures, communicative strategies and the emergence of language*. PhD thesis.

Schroeder, M. R., Atal, B. S., and Hall, J. L. (1979). Objective measure of certain speech signal degradations based on masking properties of human auditory perception. In Lindblom, B. and Öhman, S., editors, *Frontiers of speech communication research*, pages 217–229. Academic Press, London.

Schwartz, J., Boë, L., Vallée, N., and Abry, C. (1997). The dispersion-focalization theory of vowel systems. *Journal of Phonetics*, 25(3):255–286.

Scott-Phillips, T. and Kirby, S. (2010). Language evolution in the laboratory. *Trends in cognitive sciences*, 14(9):411–417.

Scott-Phillips, T., Kirby, S., and Ritchie, G. (2009). Signalling signalhood and the emergence of communication. *Cognition*, 113(2):226–233.

Senghas, A., Kita, S., and Ozyurek, A. (2004). Children creating core properties of language: Evidence from an emerging sign language in nicaragua. *Science*, 305(5691):1779–1782.

Shannon, C. E. (1948). A mathematical theory of communication. *The Bell System Technical Journal*, 27:379–423, 623–656.

Shrout, P. E. and Fleiss, J. L. (1979). Intraclass correlations : Uses in assessing rater reliability. *Psychological Bulletin*, 86(2):420–428.

Simner, J., Cuskley, C., and Kirby, S. (2010). What sound does that taste? cross-modal mappings across gustation and audition. *Perception*, 39(4):553.

Simoncelli, E. and Olshausen, B. (2001). Natural image statistics and neural representation. *Annual review of neuroscience*, 24(1):1193–1216.

Singleton, D. (2002). Age and second language acquisition. *Annual Review of Applied Linguistics*, 21:77–89.

Smith, A. (2008a). Protolanguage reconstructed. *Interaction Studies*, 9(1):99–115.

Smith, E. and Lewicki, M. (2006). Efficient auditory coding. *Nature*, 439(7079):978–982.

Smith, K. (2002). The cultural evolution of communication in a population of neural networks. *Connection Science*, 14(1):65–84.

Smith, K. (2008b). Is a holistic protolanguage a plausible precursor to language? *Interaction Studies*, 9(1):1–17.

Smith, K. (2009). Iterated learning in populations of bayesian agents. In *Proceedings of the 31st Annual Conference of the Cognitive Science Society*, pages 697–702. Austin, TX: Cognitive Science Society.

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., Smith, A. D. M., and Blythe, R. A. (2011). Cross-situational learning: an experimental study of word-learning mechanisms. *Cognitive Science*, 35:480–498.

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

Spalding, D. A. (1873). Instinct, with original observations on young animals. *Macmillan's Magazine*, 27:282–293.

Steels, L. (1997a). Language learning and language contact. In Daelemans, W., Van den Bosch, A., and Weijters, A., editors, *Workshop Notes of the ECML/MLnet Familiarization Workshop on Empirical Learning of Natural Language Processing Tasks*, pages 11–24, Prague.

Steels, L. (1997b). The synthetic modeling of language origins. *Evolution of communication*, 1(1):1–34.

Steels, L. (1997c). *The Talking Heads Experiment*. Laboratorium, Antwerpen.

Studdert-Kennedy, M. and Goldstein, L. (2003). Launching language: The gestural origin of discrete infinity. In Christiansen, M. and Kirby, S., editors, *Language Evolution*. Oxford University Press.

Suzuki, R., Buck, J. R., and Tyack, P. L. (2006). Information entropy of humpback whale songs. *The Journal of the Acoustical Society of America*, 119:1849–1866.

Tallerman, M. (2007). Did our ancestors speak a holistic protolanguage? *Lingua*, 117(3):579–604.

Tamariz, M. (2011). Could arbitrary imitation and pattern completion have bootstrapped human linguistic communication? *Interaction Studies*, 12(1):36–62.

Tamariz, M., Cornish, H., Roberts, S., and Kirby, S. (2012). The effects of generation turnover and interlocutor negotiation on linguistic structure. In *The evolution of language: Proceedings of the 9th international conference (EVOLANG9)*, pages 555–556.

Tamariz, M. and Smith, A. (2008). Regularity in mappings between signals and meanings. In *The evolution of language: Proceedings of the 7th international conference (EVOLANG7)*, pages 315–322.

Tan, R. and Fay, N. (2011). Cultural transmission in the laboratory: agent interaction improves the intergenerational transfer of information. *Evolution and Human Behavior*, 32(6):399–406.

Teal, T. and Taylor, C. (2000). Effects of compression on language evolution. *Artificial life*, 6(2):129–143.

Theisen, C., Oberlander, J., and Kirby, S. (2010). Systematicity and arbitrariness in novel communication systems. *Interaction Studies*, 11(1):14–32.

Thompson, R., Emmorey, K., and Gollan, T. H. (2005). âĂIJtip of the fingersâĂİ experiences by deaf signers insights into the organization of a sign-based lexicon. *Psychological Science*, 16(11):856–860.

Tolar, T. D., Lederberg, A. R., Gokhale, S., and Tomasello, M. (2008). The development of the ability to recognize the meaning of iconic signs. *Journal of Deaf Studies and Deaf Education*, 13(2):225–240.

Tomasello, M., Davis-Dasilva, M., Camak, L., and Bard, K. (1987). Observational learning of tool-use by young chimpanzees. *Human Evolution*, 2(2):175–183.

Ward, J., Huckstep, B., and Tsakanikos, E. (2006). Sound-colour synaesthesia: To what extent does it use cross-modal mechanisms common to us all? *Cortex*, 42(2):264–280.

Weber Fox, C. M. and Neville, H. J. (1996). Maturational constraints on functional specializations for language processing: Erp and behavioral evidence in bilingual speakers. *Journal of Cognitive Neuroscience*, 8(3):231–256.

Wray, A. (1998). Protolanguage as a holistic system for social interaction. *Language and Communication*, 18(1):47–67.

Wray, A. and Grace, G. W. (2007). The consequences of talking to strangers: Evolutionary corollaries of socio-cultural influences on linguistic form. *Lingua*, 117:543–578.

Yun, A. J., Bazar, K. A., and Lee, P. Y. (2004). Pineal attrition, loss of cognitive plasticity, and onset of puberty during the teen years: is it a modern maladaptation exposed by evolutionary displacement? *Medical hypotheses*, 63(6):939–950.

Zuberbühler, K. (2000). Referential labelling in diana monkeys. *Animal Behaviour*, 59(5):917–927.

Zuidema, W. (2003). How the poverty of the stimulus solves the poverty of the stimulus. *Advances in neural information processing systems*, pages 51–58.

Zuidema, W. and de Boer, B. (2009). The evolution of combinatorial phonology. *Journal of Phonetics*, 37(2):125–144.

Zuidema, W. and Westermann, G. (2003). Evolution of an optimal lexicon under constraints from embodiment. *Artificial Life*, 9(4):387–402.

List of Figures

3.1 Representation of the scribble to sound mapping. The trajectory that is shown in the figure would approximately sound like “iiiiiiuuuuuuaaaaa”. Note that participants did not see the axes or transcriptions, the scribble area on the screen was empty. 24
3.2 Meaning space 25
3.3 Scribbles produced by participants during the final test in chain one. The first row shows the trajectories for the random input sounds and each following row shows the output of a participant who received the data from the previous row as input. The darker border around the picture means that this item was part of the training set for the next person. The grey dots indicate the starting point of the trajectories. 28
3.4 Scribbles produced by participants during the final test in chain two. The first row shows the trajectories for the random input sounds and each following row shows the output of a participant who received the data from the previous row as input. The darker border around the picture means that this item was part of the training set for the next person. The grey dots indicate the starting point of the trajectories. 29
3.5 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 colour of the object. . . . 31
3.6 Chain two, generation two. Note that the location of the trajectory indicates the colour of the object in the meaning space. 31
3.7 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 colour of the object. . . . 32
3.8 Scribbles produced by participants during the final test in chain three. The first row shows the trajectories for the random input sounds and each following row shows the output of a participant who received the data from the previous row as input. The darker border around the picture means that this item was part of the training set for the next person. The grey dots indicate the starting point of the trajectories. 33

3.9 Produced scribbles of three successive generations for the donut shaped objects. All three participants follow the 'rule' that blue is high, red is low and green is middle, even though none of these participants were exposed to the green object. 34
3.10 Average distance between input and output for chain one 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). 35
3.11 Average distance between input and output for chain two 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). 36
3.12 Average distance between input and output for chain 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). 36

4.1 Plastic slide whistle from the brand Grover-Trophy 41
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. 44
4.3 An example of recombination in chain four: a whistle from the previous generation is combined with the second part of another whistle and a second version is added with a mirrored part. Note the co-articulation-like effect highlighted with circles: the final pitch of the first part influences the initial pitch of the second part. 46
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. 48
4.5 Fragment of the whistles plotted as pitch tracks in the last set of a chain. Basic elements can be identified that are systematically recombined. 49
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. 51
4.7 Entropy of the whistle sets over generations for all four chains, demonstrating that the combinatorial structure increases. . . 52

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. . . . 53

4.9 Dispersion measured as energy between whistles in the set for each generation. The whistles do not tend to become more dispersed (no decrease in energy) towards the end of the chains. On the contrary, for at least one of the chains there appears to be an increase of energy. 54

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. 55

4.11 Dispersion measured as energy between building blocks in each generation. The building blocks tend to become more dispersed (lower energy) towards the end of the chains. . . . 56

5.1 Screenshot of the UFO game. 63

5.2 Two experimental conditions: (1) the 'intact' condition, where each of the two alien species languages consisted of an intact emergent whistle set from the last generation of chain one and chain four of the experiment described in chapter 4. (2) the 'mixed' condition, where mixing sounds from both sets created the two languages. 64

6.1 Examples of novel objects used in the experiment. These objects were created by Smith et al. (2011) and were slightly modified. To reduce potential categorisation according to colours in the meaning space, all objects are in blue tone (transformed with a blue filter). 75

6.2 **a**: The next person in a chain was exposed to the exact pairs of whistles and objects that the previous person created. **b**: The next person in a chain was exposed to the exact set of whistles that the previous person created but from one person to the other the set of objects was replaced and the whistles were randomly paired with the objects. Two sets of 12 objects were alternated and each was used every other generation so that the odd-numbered generations saw one set, and the even-numbered generations the other set. 76

6.3 The initial whistle sets used in the experiment 78

6.4 Development of structure in a chain from the scrambled condition. Half of each of the whistles in the first row is borrowed and reused to form a new whistle. The left part of the smooth whistle is also reused and combined with existing whistles. These are then reproduced and all kinds of other variations on this appear. 80

6.5 Development of structure in a chain from the intact condition. The whistle on the first row seems to be an example for two new whistles in the next generation: one with one 'bump' and another with two. The 'two-bump' whistle is starting to be reused and combined with another pattern and in generation six both the one-bump and two-bump whistles are being reused, mirrored and recombined more widely. 81
6.6 Fragment from the whistle set produced by the last participant in a chain from the scrambled condition. Whistle sounds are plotted as pitch tracks on a semitone scale. Basic building blocks can be identified. 82
6.7 A structure where silences do not determine segment boundaries. 83
6.8 A structure where recombination is not solely sequential. . . 84
6.9 Examples of iconic whistle-object pairs in the data. The first shows how the holes in the object that are arranged from the bottom to the top and become bigger are iconically depicted as a sequence of notes in a rising pattern. The second shows how the shape of the object is mimicked in the pitch contour. The third shows how the orientation of the object is imitated in the pitch contour. 85
6.10 An example of iconicity that is lost over generations. 86
6.11 Recall error over generations in both conditions, showing the mean and standard error. Recall error decreases significantly in both conditions. 87
6.12 Recall error on the exact whistle-object pairs over generations in both conditions, showing the mean and standard error. Recall error on the exact whistle-object pairs decreases significantly in both conditions. 88
6.13 Entropy of the whistle sets over generations in both conditions, showing the mean and standard error. Entropy decreases significantly in both conditions. This suggests that the combinatorial structure increased over generations. 89

7.1 Graphical representation of the acoustic space (the trapezium), with a repertoire of vowel prototypes (the dots) 104
7.2 Procedure of an imitation game: 1. selecting a random vowel prototype, 2. producing this vowel, 3. perception of this vowel in the acoustic space of the other agent, 4. finding the closest vowel prototype, 5. producing this vowel, 6. perception of this vowel in the acoustic space of the first agent, 7. finding the closest vowel prototype. In 7.2a the game results in a success because the recognised vowel in step 7 is the same as the one produced in step 1. In 7.2b it is a failure because the recognised vowel in step 7 is not the same as the one produced in step 1. Images adapted from animations by de Boer (2000). 106

7.3 Emerged vowel system after 200 000 games in a population with 50 agents. Each dot represents a vowel prototype in an agents memory. This system was used as the starting point of the experiments with changing populations. 108
7.4 Influence of the age structure on the preservation of vowel systems. An example is plotted for each situation with the original vowel system in grey and the newly emerged vowel systems in black. Note that although in all cases changes occur, they are smaller for the systems with the age structure. . . . 112

List of Tables

5.1 Results of the UFO experiment. There is a significant difference between the scores in the two conditions, measured as d' and as the median score of correct classification. 65

5.2 Results of the follow-up UFO experiment. For both pairs of chains there are significant differences between the scores in the two conditions, measured as d' and as the median score of correct classification. 68

7.1 Results of the experiments with and without age structure, showing the averages and standard deviations of the measures. ε is the learning step size and α the speed of ageing. Note that the similarity is higher and the distance lower for runs with the age structure: the vowel systems are preserved better. The average success is smaller in the two cases with ageing because vowel systems here are larger. 110

7.2 Similarity measure: Effect sizes (based on Cohen's d) and the corresponding 95% confidence intervals (CI). 110

7.3 Distance measure: Effect sizes and confidence intervals. . . . 111

7.4 Size measure: Effect sizes and confidence intervals. 111

7.5 Results of the experiments with a strict critical period (CP), going to step size (ε_0) of 0, showing the averages and standard deviations. As in the case with the gradual decline in learning ability, the similarity is higher and distance lower for runs with the age structure: vowel systems are preserved better. 113

Appendix A: Scribbles

A.1 Instructions

Experiment: Scribble2Sound

Welcome!

This experiment is conducted in the context of a research project on the evolution of speech at the Amsterdam Center for Language and Communication (University of Amsterdam). Thank you very much in advance for your participation! Please read this short instruction carefully before you start the experiment. If you have any questions or comments, don't hesitate to let the experimenter know or contact t.verhoef@uva.nl afterwards.

The application you will start in a minute contains a `Scribble area' and a `Topic area'. The scribble area allows you to produce scribbles that will be transformed into sounds. The topic area will display different pictures that have a specific sound connected to them. First, you will get some time to practice by drawing some shapes in the scribble area to find out how it relates to the sounds. It is important that you familiarize yourself somewhat with the scribble area. After this you start the experiment. During the actual experiment you are going to **learn** these specific **combinations of pictures and sounds** and at the same time you will learn to **imitate** these sounds using the scribble area.

In the training phase, you will hear a sound, see the corresponding picture and are asked to reproduce it by drawing a scribble. Try your best to make your own produced imitation resemble the example sound as closely as possible. **This is not an easy task!** So, please do not get frustrated! A colored border surrounding the topic will indicate how close your imitation was: **the greener, the better**.

In the test phase, you will only see the picture and are asked to produce the corresponding sound with a scribble. There will be a total of three rounds of training and testing, with short breaks in between.

Good luck!

Figure A.1.1: *Written instructions given to participants in the Scribble to Sound experiment (described in chapter 3).*

A.2 User interface

Figure A.2.1: *Screenshot of the user interface for the Scribble to Sound experiment (described in chapter 3).*

A.3 Random scribble trajectory generation

To create the initial set of sounds that were given as input to the first participant in each chain of the scribble to sound experiment (described in chapter 3), the computer generated random trajectories. These trajectories were transformed into sound using the same scribble to sound mapping that was used in the rest of the experiment. The trajectories were not entirely unconstrained, so that they sounded as if they could have been created by a person controlling the mouse.

Points on the trajectories are *x,y* pairs that represent locations in the two-dimensional scribble area where *x* and *y* are values between 0 and 1. For the first point of a random trajectory, a uniformly distributed random value inside the scribble area is chosen for both *x* and *y*. Then, the choice of each next point is constrained, so that it (1) is not too far away from the previous point and (2) creates a line between the current point and the previous point that has a large enough angle with the line between the previous point and its predecessor.

(1) The new point is chosen such that the new *x* value is *x*+α, where α is a uniformly distributed random value between -0.025 and 0.025, with the additional constraint that it stays within the scribble area. The new value for *y* is computed in the same way and the new location needs to have a valid angle, determined as follows:

(2) The angle is computed by using the definition of the dot product of two vectors. The angle (θ) between two vectors can be computed as:

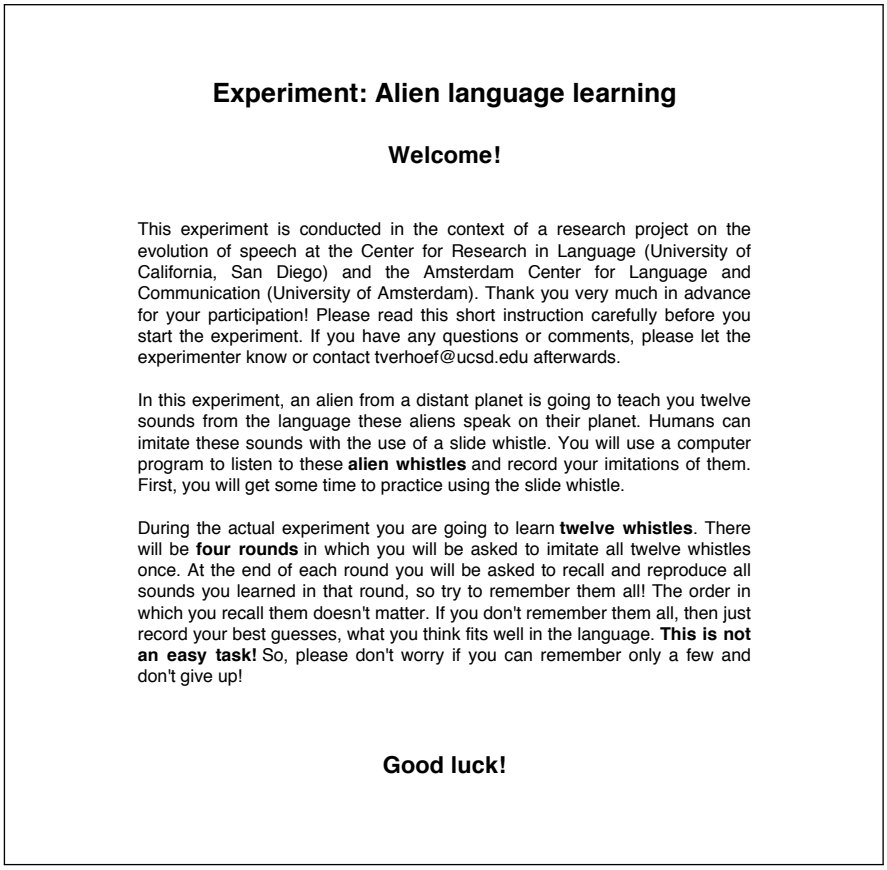
$$\theta = \arccos\left(\frac{\mathbf{A} \cdot \mathbf{B}}{\|\mathbf{A}\| \, \|\mathbf{B}\|}\right) \tag{1}$$

A is a vector representing the line between the current point and the previous point, **B** is a vector representing the line between the previous point and its predecessor, $\mathbf{A} \cdot \mathbf{B}$ is the dot product of **A** and **B**, $\|\mathbf{A}\|$ is the magnitude of **A** which is the Euclidean distance between the current point and the previous point and $\|\mathbf{B}\|$ is the magnitude of **B** which is the Euclidean distance between the previous point and its predecessor.

The angle is considered to be valid if it is larger than $\frac{3}{4}\pi$.

Appendix B: Whistles

B.1 Instructions



Experiment: Alien language learning

Welcome!

This experiment is conducted in the context of a research project on the evolution of speech at the Center for Research in Language (University of California, San Diego) and the Amsterdam Center for Language and Communication (University of Amsterdam). Thank you very much in advance for your participation! Please read this short instruction carefully before you start the experiment. If you have any questions or comments, please let the experimenter know or contact tverhoef@ucsd.edu afterwards.

In this experiment, an alien from a distant planet is going to teach you twelve sounds from the language these aliens speak on their planet. Humans can imitate these sounds with the use of a slide whistle. You will use a computer program to listen to these **alien whistles** and record your imitations of them. First, you will get some time to practice using the slide whistle.

During the actual experiment you are going to learn **twelve whistles**. There will be **four rounds** in which you will be asked to imitate all twelve whistles once. At the end of each round you will be asked to recall and reproduce all sounds you learned in that round, so try to remember them all! The order in which you recall them doesn't matter. If you don't remember them all, then just record your best guesses, what you think fits well in the language. **This is not an easy task!** So, please don't worry if you can remember only a few and don't give up!

Good luck!

Figure B.1.1: *Written instructions given to participants in the whistle experiment (described in chapter 4).*

B.2 User interface

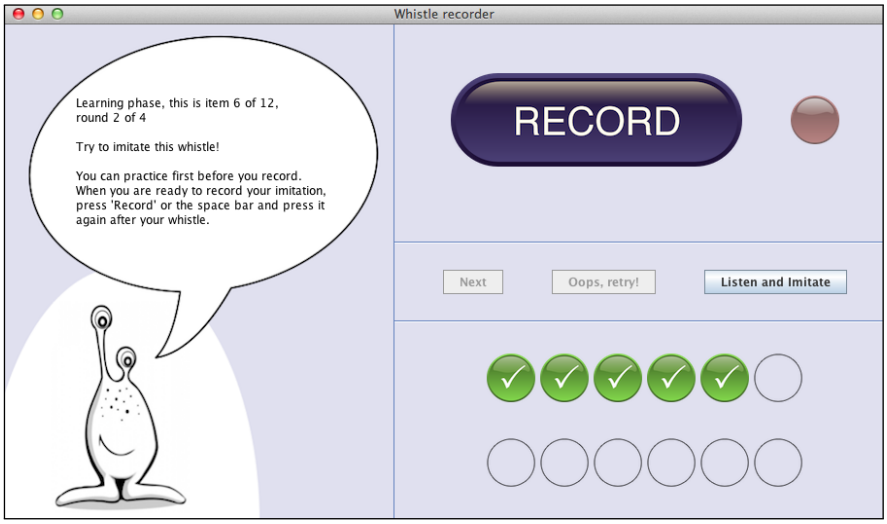


Figure B.2.1: *Screenshot of the user interface for the whistle experiment (described in chapter 4). The instructions appeared in the speech bubble and the green check marks helped to keep track of the progress, both in the imitation phase and in the recall phase.*

B.3 Transmission chains

This section shows the transmission chains that resulted from the experimental iterated learning experiment with whistled signals (described in chapter 4). Whistle sounds are displayed as pitch tracks on a semitone scale and the signals are organised in tables, spanning two pages for each chain, in which rows represent generations and columns the twelve different whistles. The first row shows the initial input set of whistle sounds and each following row represents the last recalled output of consecutive participants in the chain. Due to the fact that participants freely reproduced the whistles in the order they preferred, it was impossible to know exactly which signal from their input they attempted to recall. To organise the signals into the columns as displayed in the tables, the whistle distance measure as described in section 6.2.3 was used to find the best mapping between the whistle sets from two consecutive generations. Each whistle from one set was paired with a unique whistle from the other set and this was repeated in all possible ways to find the pairing for which the sum of distances was minimal. Based on this measured best fitting mapping, the whistles were displayed in the tables.

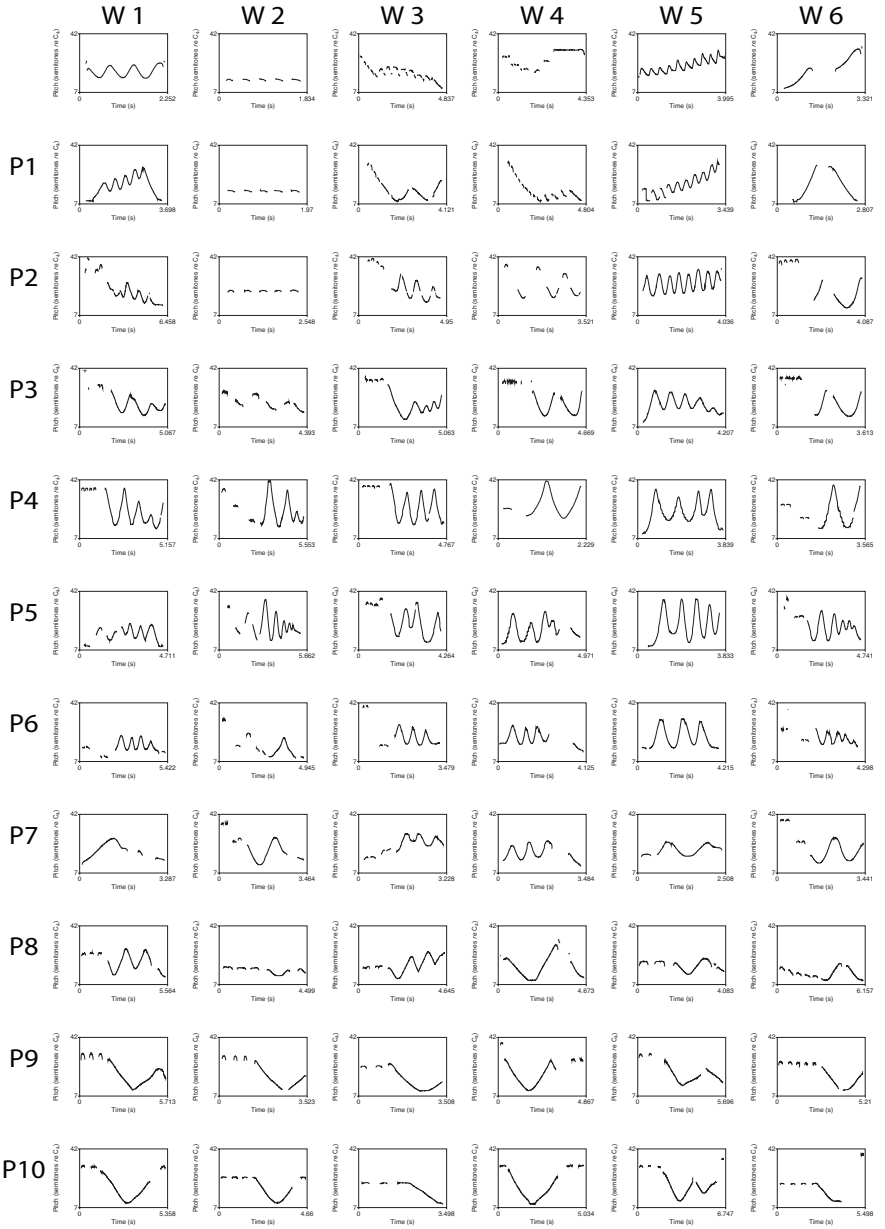


Figure B.3.1: *Transmission chain one of the whistle experiment (chapter 4). The first row shows the initial input set of whistle sounds (W 1 to 12) and each following row represents the last recalled output of consecutive participants (P 1 to 10) in the chain.*

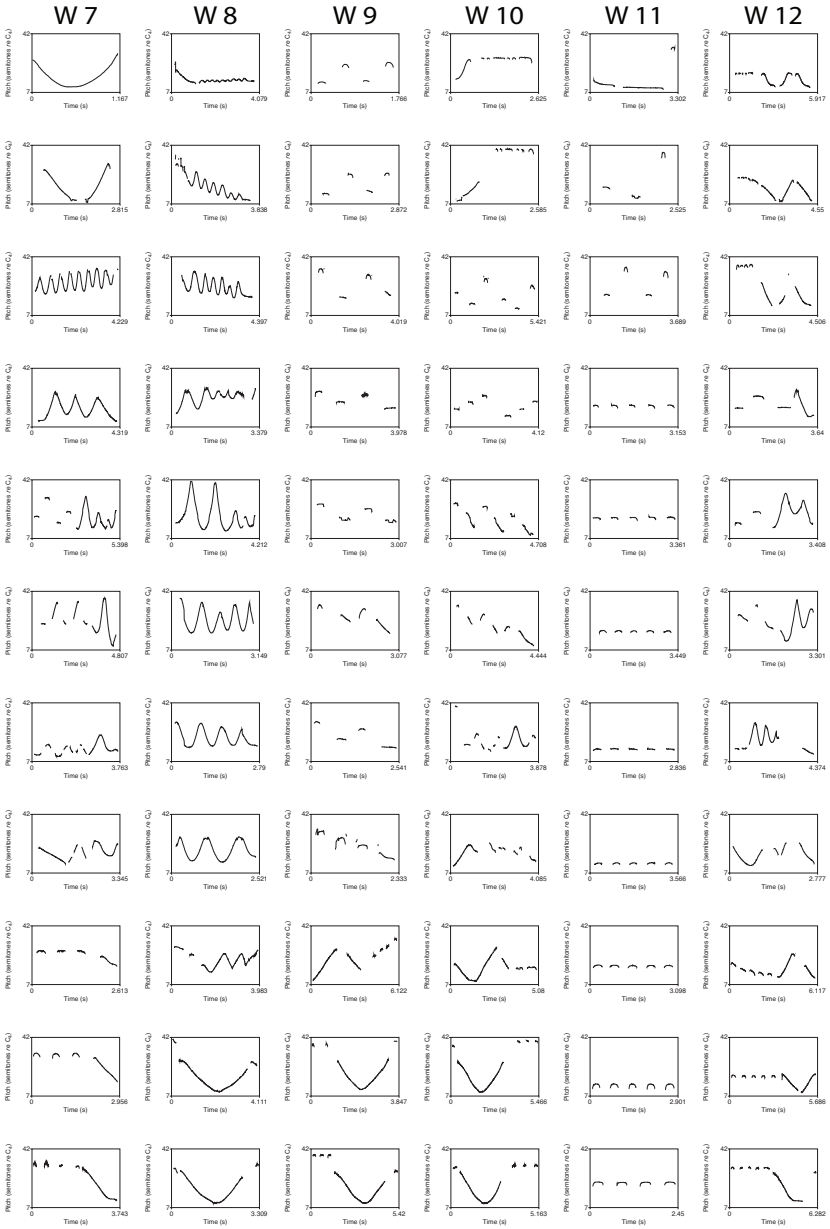


Figure B.3.2: *Chain one continued*

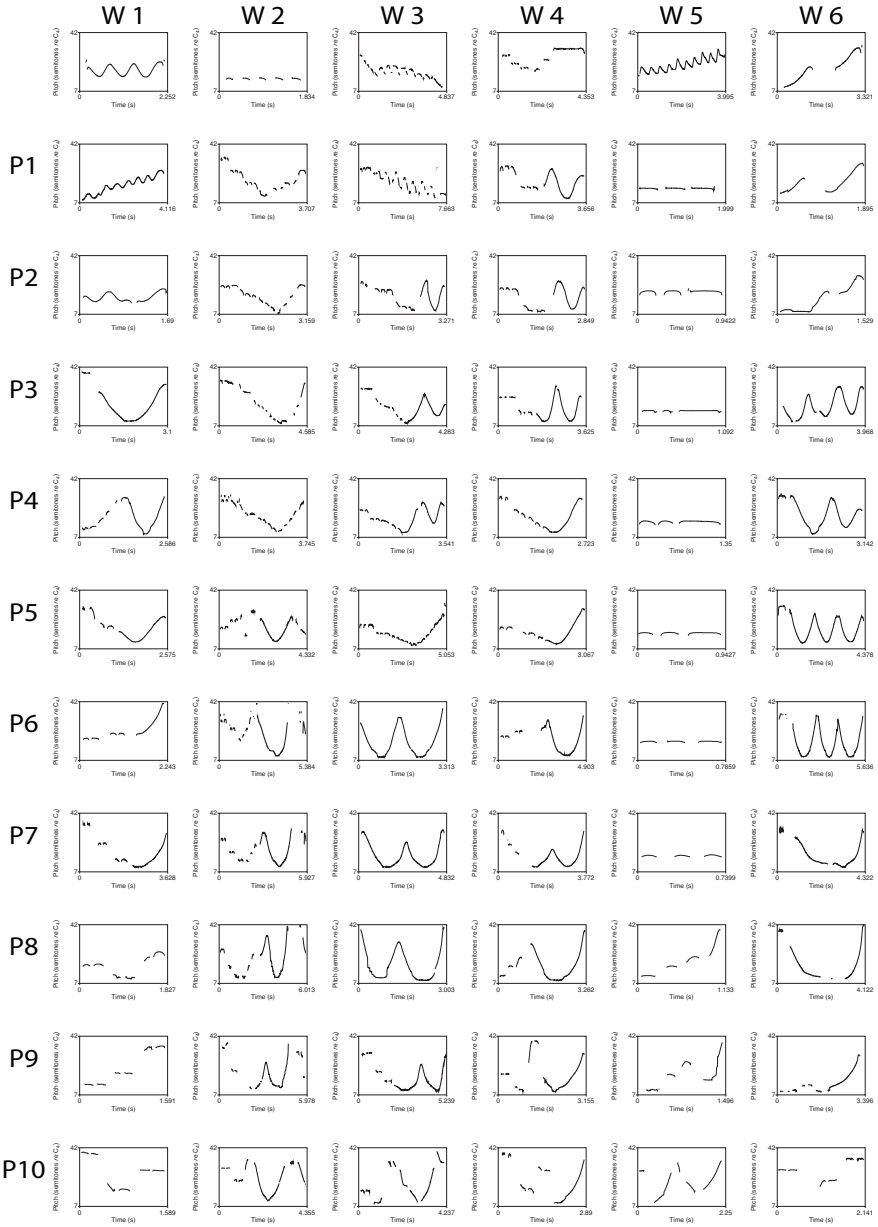


Figure B.3.3: *Transmission chain two of the whistle experiment (chapter 4). The first row shows the initial input set of whistle sounds (W 1 to 12) and each following row represents the last recalled output of consecutive participants (P 1 to 10) in the chain.*

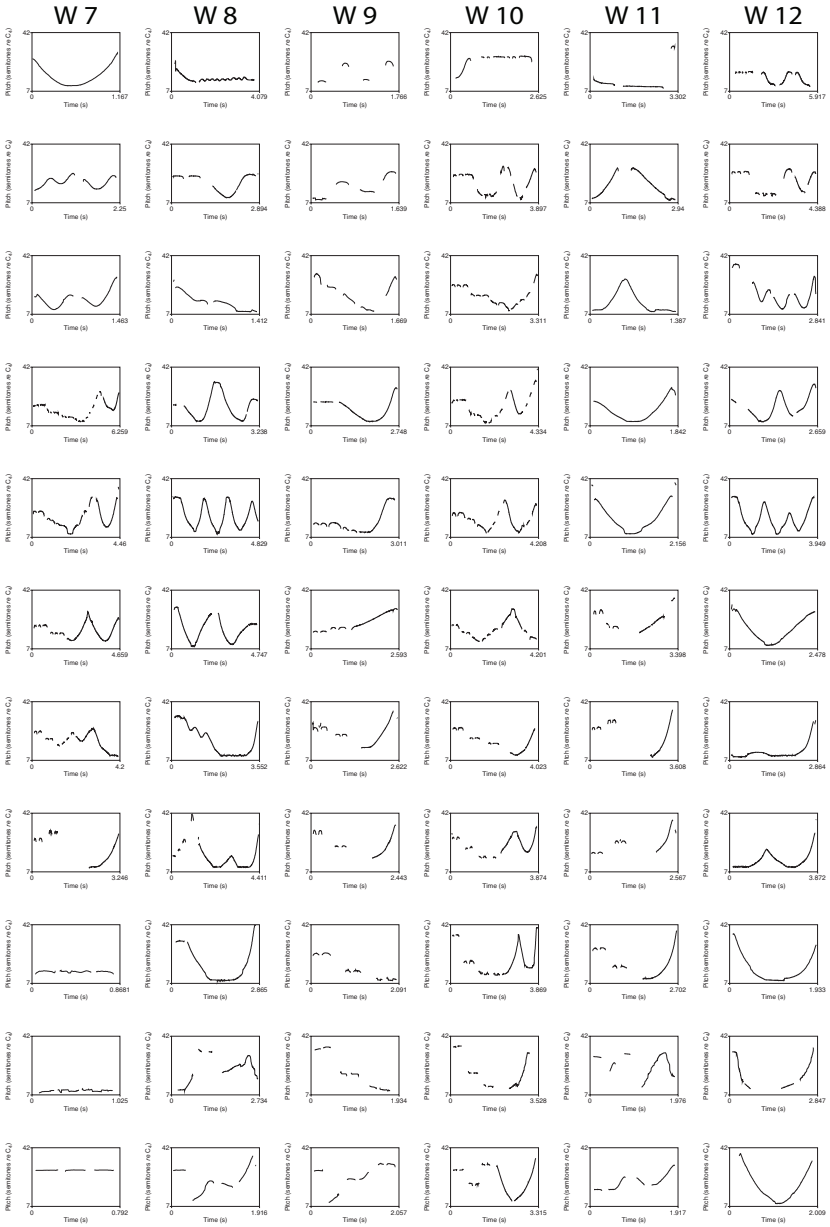


Figure B.3.4: *Chain two continued*

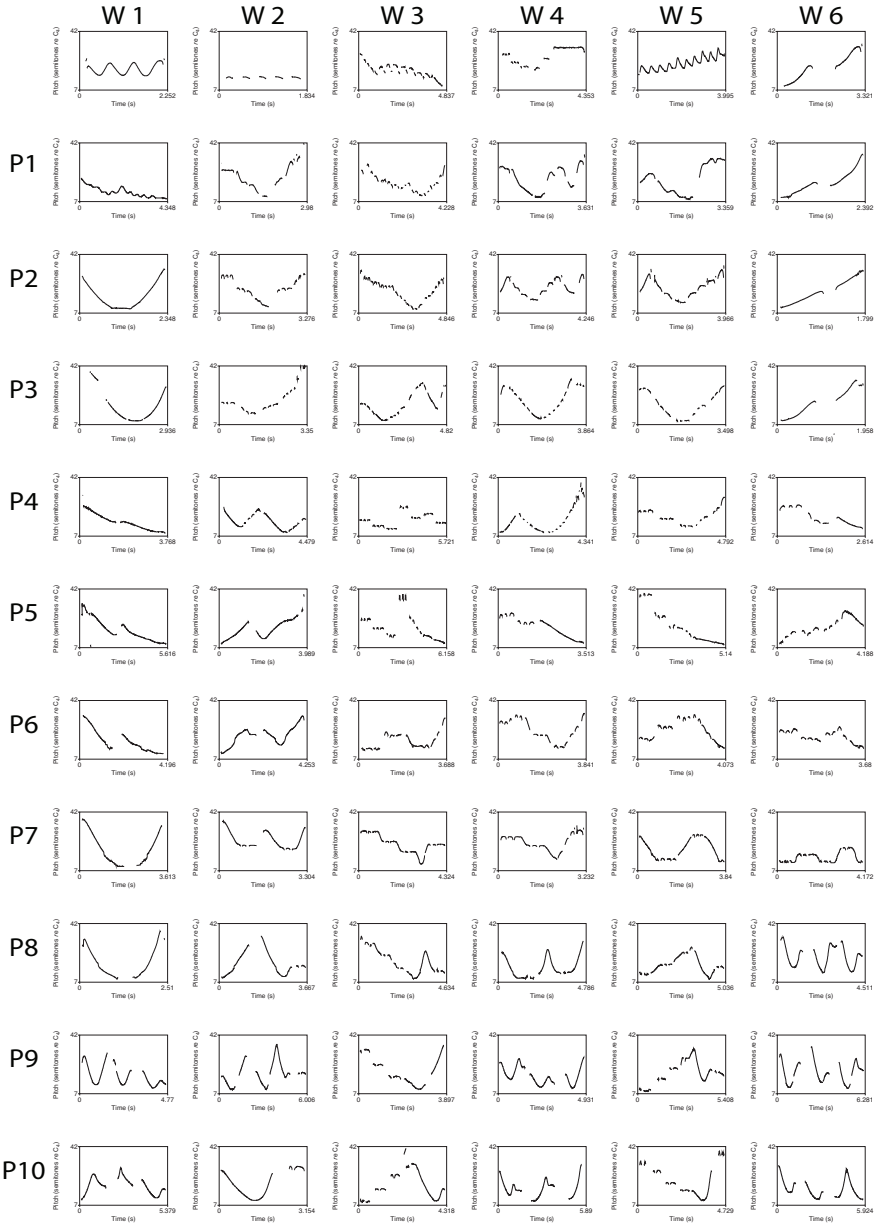


Figure B.3.5: *Transmission chain three of the whistle experiment (chapter 4). The first row shows the initial input set of whistle sounds (W 1 to 12) and each following row represents the last recalled output of consecutive participants (P 1 to 10) in the chain.*

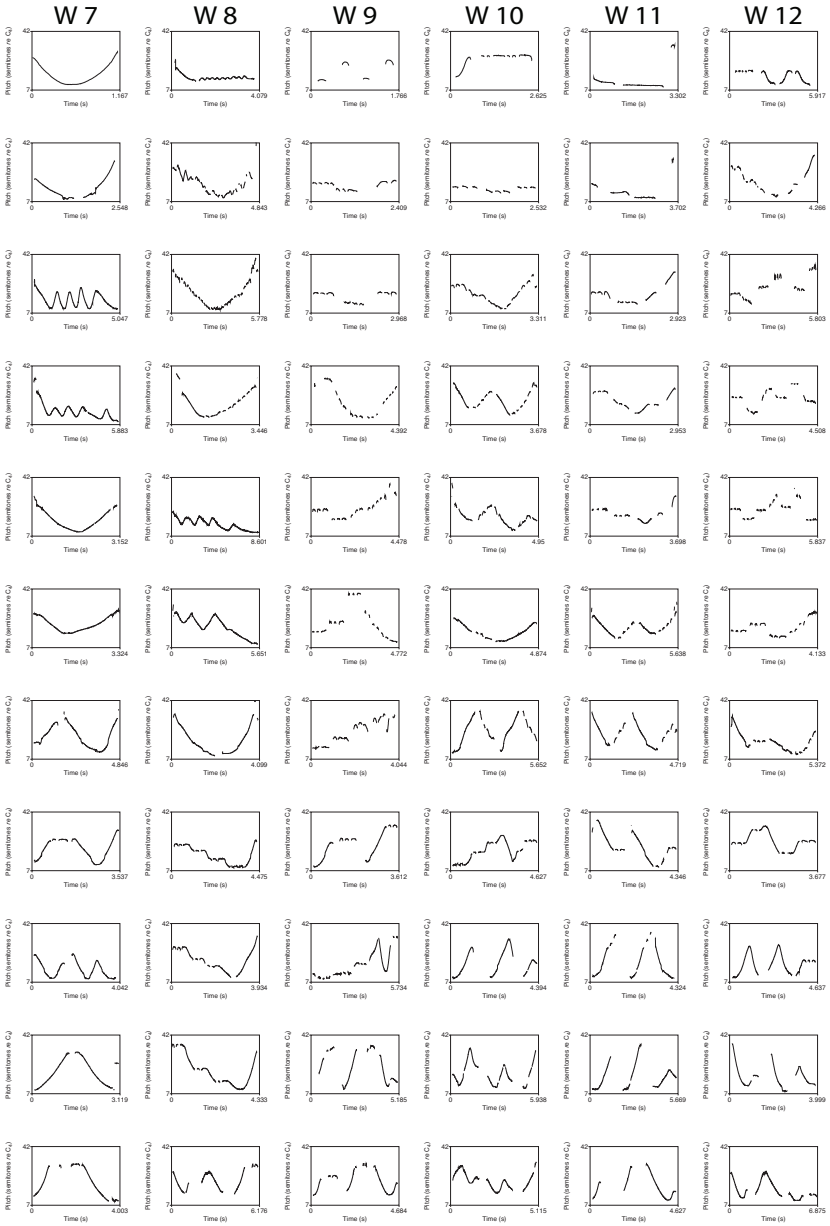


Figure B.3.6: *Chain three continued*

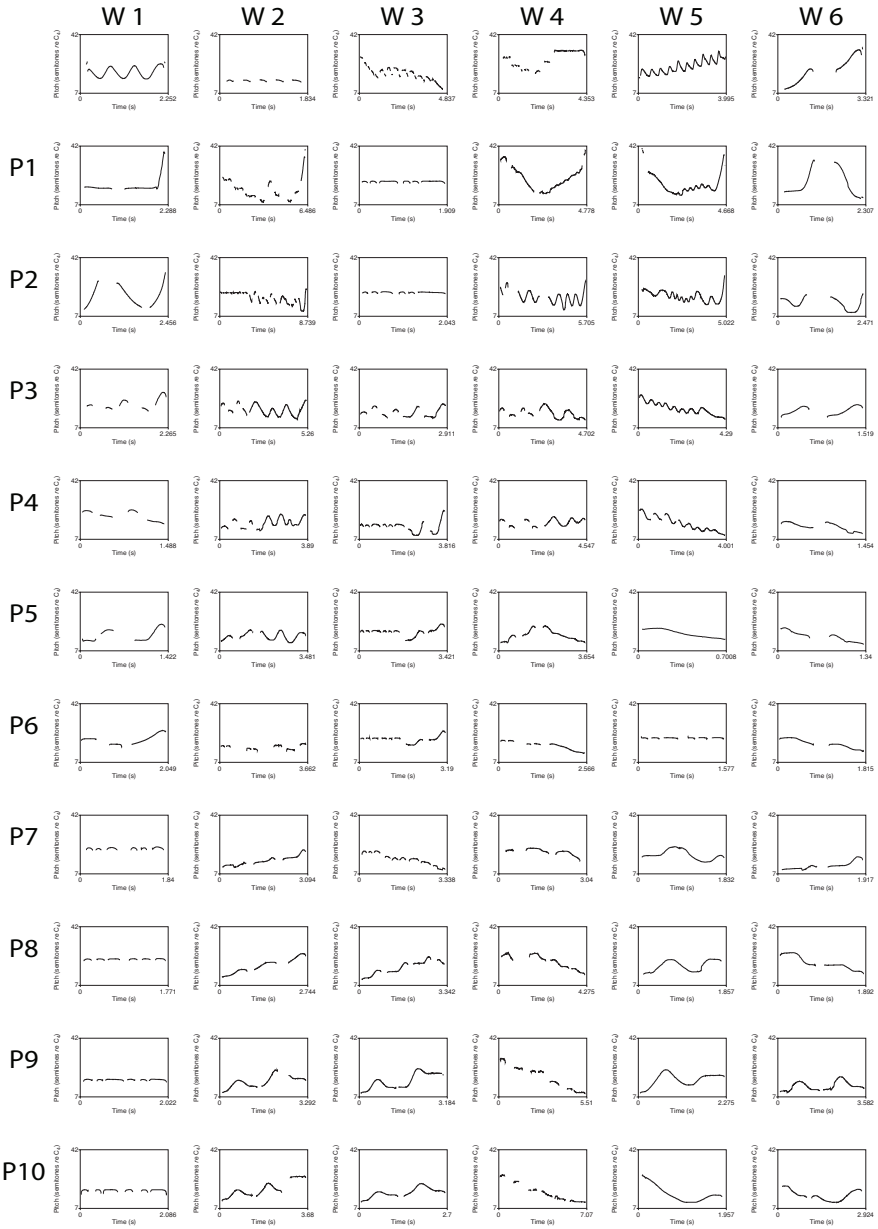


Figure B.3.7: *Transmission chain four of the whistle experiment (chapter 4). The first row shows the initial input set of whistle sounds (W 1 to 12) and each following row represents the last recalled output of consecutive participants (P 1 to 10) in the chain.*

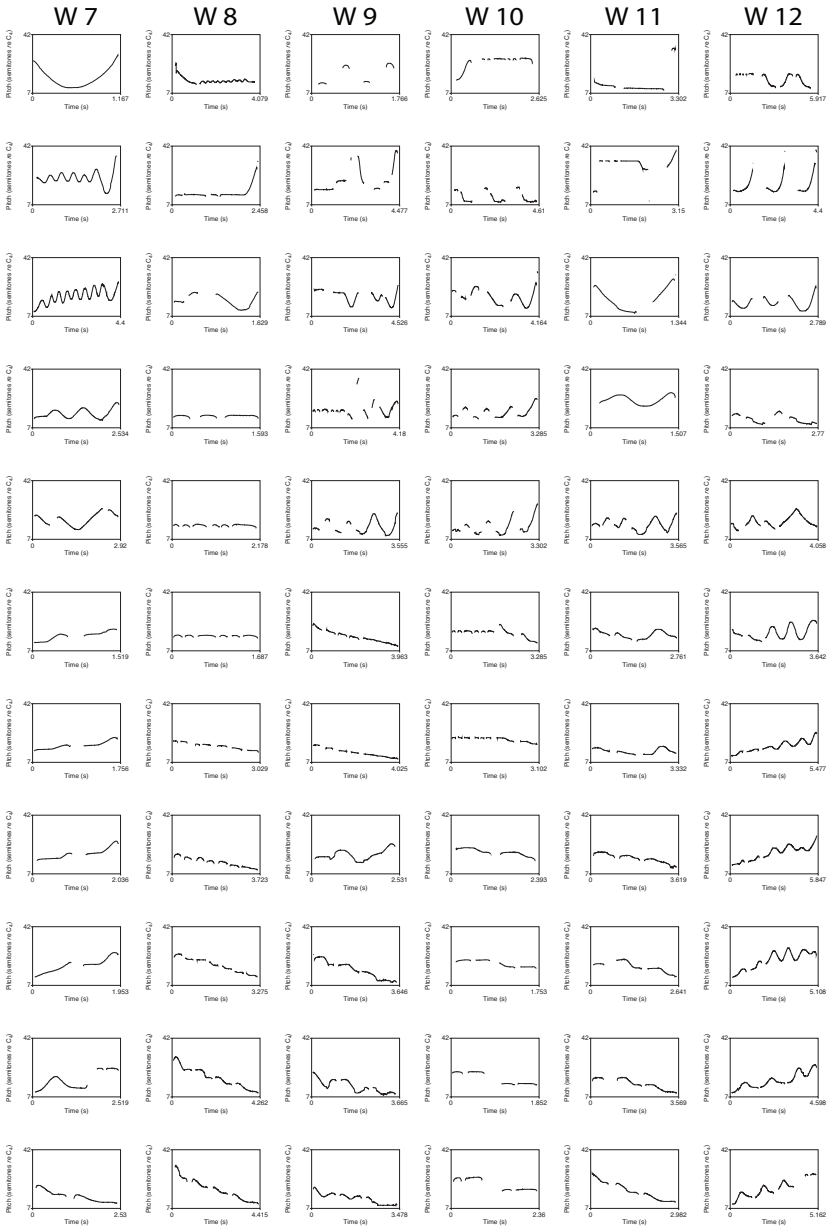


Figure B.3.8: *Chain four continued*

B.4 Analysis details

This section describes in detail how the whistle sound files were stored, preprocessed and analysed in order to compute the measures that were used in chapter 4 and chapter 6 for the analysis of learnability and structure of whistled languages.

B.4.1 Pre-processing of whistle sound files

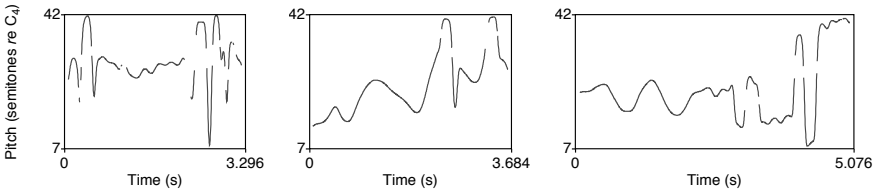
The user interface of the experiment records the whistle sounds as .au files. The first step in the analysis is the extraction of pitch and intensity data in Praat (Boersma, 2001). This was done with a script that processed each of the files, starting with pitch extraction, using the following settings:
`To Pitch (ac)... 0 200 15 no 0.05 0.45 0.01 0.35 0.14 3000`,
and intensity extraction, using the following settings:
`To Intensity... 200 0 yes`

A PitchTier object was created from the pitch track so that the pitch track could be adjusted in the way described in the next section. From the resulting pitch tier, the pitch track was extracted with a sample rate of 500 samples per second. With the same sample rate the intensity track was extracted. In addition, two tables were computed and stored, one from the pitch that was originally extracted from the sound, displaying voiced and unvoiced intervals and one from the intensity track with silent and sounding intervals. This collected data was then used for further processing in Java. When the experiment was conducted, some processing of the whistle sounds had to be done on the fly while the user interface of the experiment was running, for instance for the working of the reproduction constraint. In this case, the pitch was extracted directly in Java, using the Yin method (De Cheveigné and Kawahara, 2002).

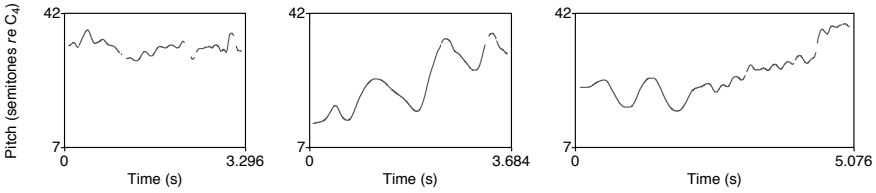
B.4.2 Jump removal

Due to the nature of the slide whistle, it happened often that there were unintentional jumps in the pitch tracks. With a change in the air pressure, sometimes overtones get more prominent, or the pitch gets under or over estimated which causes the measured pitch track to suddenly jump up or down. This distorts the pitch tracks, making it look as if the participant very rapidly moved the whistle plunger when this was not the case. A specific procedure was implemented to track down and fix these jumps. Figure B.4.1 shows a few example before and after jump removal.

Many of the jumps occurred at the beginning or end of a whistle segment (as the air pressure rises and falls). Therefore the silences between segments of sound were used in the first step. For this, a table was created which indicated the sounding and silent intervals, based on:
`To TextGrid (silences)... -25 0.03 0.06 m s`.



a Before



b After

Figure B.4.1: *Pitch tracks on a semitone scale before and after the jump removal procedure.*

For each silent segment in the intensity table, the points in the PitchTier between the start of the silent segment minus 0.03 *ms* and the end of the silent segment plus 0.03 *ms* were removed. Then, since all points above a pitch value of 3000 *Hz* or below 300 *Hz* were well outside of the range of the slide whistle, any (single) points outside this interval were removed as well.

After that the procedure would go on to search for jumps within the sounding segments and from one segment to the other. A table was created listing all the values on the PitchTier and their specific time stamp. Looping over all points, for each pair of consecutive pitch points, it was determined whether it was a 'long' (up to 0.3 *ms*, usually between segments) or a 'short' (up to 0.005 *ms*) interval. For long intervals, a difference in pitch value of 6 semitones or more was considered an unintended jump. When such a jump was found either all values preceding the current one were shifted on the PitchTier (in the case a shift of the current point would place it outside the slide whistle pitch range) or the current point and all points following it were shifted. For small intervals, a difference in pitch of 0.5 semitones would be considered unintended/inaccurate and the current value was shifted in this case. For any time interval of a longer duration than 0.3 *ms*, no pitch adjustments were made.

It could happen that, as part of the procedure described above, some part of the pitch would have gotten shifted outside the slide whistle range. If this was the case, the whole signal was shifted again to correct this.

The jump search procedure fixed the vast majority of problems in the pitch tracks for the collected data, but there were some exceptions for which the applied heuristic would make adjustments where they were not needed. For this reason, the script was executed both with and without the octave jump search part and the results were manually checked, restoring the original where the procedure messed it up.

B.4.3 Segmenting whistle sounds

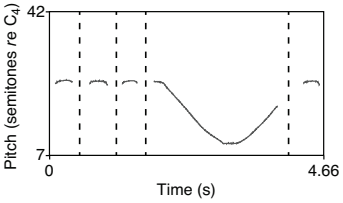
As part of the measures in the analysis of the entropy of whistle sets, whistle sounds had to be segmented. In chapter 4 only one way of segmenting was used, which will be described first, followed by two other methods that were only used in chapter 6.

The first method segmented the whistles on the basis of silences between sounding parts as segment boundaries. From Praat we got two sources of information about where the silences could be between the segments: the table with voiced/unvoiced intervals for the measured pitch and the table with sounding/silence intervals for the measured intensity. If the signal is very clear, the two results will give the same intervals and the segments can immediately be extracted on the basis of these. Sometimes however, the two tables do not entirely overlap, for instance because the pitch was not properly detected in one segment, or because the intensity was not strong enough to pass the threshold. In the case there were inconsistencies, a heuristic was used by consulting both tables to find out as well as possible where the segments actually are.

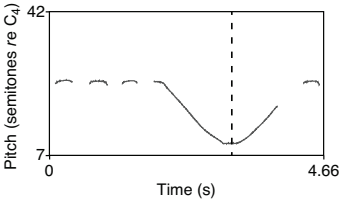
First, the pitch table tended to overestimate the number of segments more often and sometimes resulted in very short voiced segments where it thought it detected a pitch while it was not there. Therefore, all voiced or unvoiced segments that were really short (< 0.03 seconds) were removed so that the surrounding segments could be merged. If the pitch table still estimated a higher number of segments, it was inspected to see if there are any voiced intervals that are shorter than 0.04 seconds and these are also removed.

If these steps did not solve the difference, the remaining voiced segments were all checked again and if they were 0.1 second or shorter and the average intensity in the interval was much lower than the average intensity of the complete signal, the segment was also removed. Eventually the segments were extracted on the basis of the intervals from the (adjusted) pitch table.

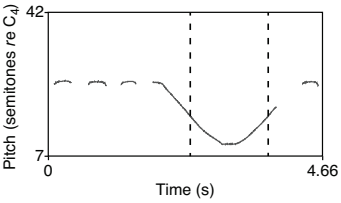
In chapter 6 two other methods were used for extracting segments. This time, the segment boundaries were not only based on the silences, but sometimes also on the minima and maxima in the plunger movement track and the points of maximal velocity. Figure B.4.2 shows illustrates the three different segmentations. To find the right intervals for these segmentations, two other procedures were used.



a Segmenting on the basis of silences



b Segmenting on the basis of movement maxima and minima



c Segmenting on the basis of movement maximal velocity

Figure B.4.2: *Three different ways of segmenting whistle signals.*

To compute the intervals for the segments based on the minima and maxima in the plunger movement tracks, the first derivative of the plunger movement track was used. First, the plunger movement track as computed from the pitch track following equation 2, where l is the length in cm between the mouthpiece and sliding stopper, c is the speed of sound at body temperature ($35,000 cm/s$) and f is the measured frequency in Hz.

$$l = \frac{c}{4f} \tag{2}$$

The first derivative was computed as described by Keogh and Pazzani (2001). Maxima and minima could then easily be found as the points where the first derivative crosses 0. Sometimes however, changes in direction would be very small and be accidental 'tremors' instead of real intended up and down movements. Therefore a threshold was used (min 0.5 cm) for the size of the plunger displacement.

To compute the intervals for the segments based on the points of maximal velocity in the plunger movements, the same procedure was used as above, but this time on the second derivative.

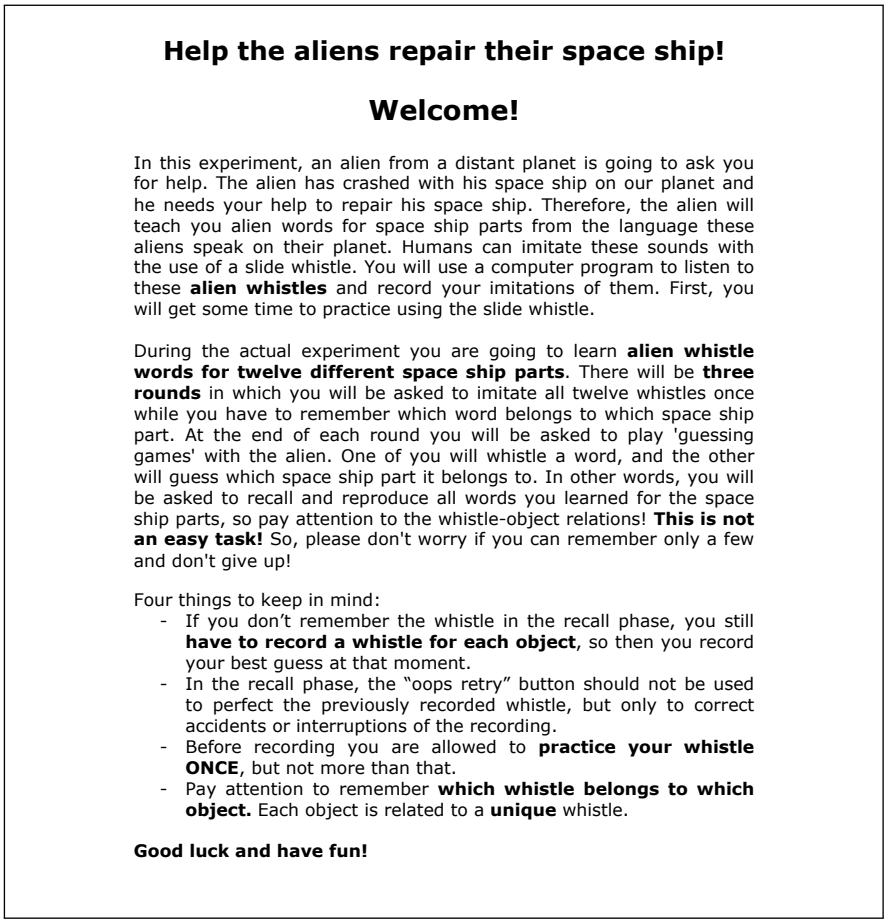
B.4.4 Dynamic time warping

Many comparisons between whistle sounds in the analyses of the experimental data in this thesis made use of Dynamic Time Warping (Sakoe

and Chiba, 1978). Here, the dynamic time warping distance between two sequences was computed using the original method described in (Sakoe and Chiba, 1978), using their step pattern Symmetric P1. For the computation of Derivative Dynamic Time warping, which was also used in the analyses, the same implementation for DTW was used, but the input signals were the derivatives of the signals computed in the way described by (Keogh and Pazzani, 2001). The signals all had different durations so to normalise for the differences in the lengths of the signals, the DTW distance was divided by the sum of the lengths of the signals as in (Sakoe and Chiba, 1978).

Appendix C: Meanings

C.1 Instructions



Help the aliens repair their space ship!

Welcome!

In this experiment, an alien from a distant planet is going to ask you for help. The alien has crashed with his space ship on our planet and he needs your help to repair his space ship. Therefore, the alien will teach you alien words for space ship parts from the language these aliens speak on their planet. Humans can imitate these sounds with the use of a slide whistle. You will use a computer program to listen to these **alien whistles** and record your imitations of them. First, you will get some time to practice using the slide whistle.

During the actual experiment you are going to learn **alien whistle words for twelve different space ship parts**. There will be **three rounds** in which you will be asked to imitate all twelve whistles once while you have to remember which word belongs to which space ship part. At the end of each round you will be asked to play 'guessing games' with the alien. One of you will whistle a word, and the other will guess which space ship part it belongs to. In other words, you will be asked to recall and reproduce all words you learned for the space ship parts, so pay attention to the whistle-object relations! **This is not an easy task!** So, please don't worry if you can remember only a few and don't give up!

Four things to keep in mind:

- If you don't remember the whistle in the recall phase, you still **have to record a whistle for each object**, so then you record your best guess at that moment.
- In the recall phase, the "oops retry" button should not be used to perfect the previously recorded whistle, but only to correct accidents or interruptions of the recording.
- Before recording you are allowed to **practice your whistle ONCE**, but not more than that.
- Pay attention to remember **which whistle belongs to which object.** Each object is related to a **unique** whistle.

Good luck and have fun!

Figure C.1.1: *Written instructions given to participants in the whistle experiment with meanings (described in chapter 6).*

C.2 User interface

Figure C.2.1: *Screenshot of the user interface for the whistle experiment with meanings (described in chapter 6). The instructions appeared in the speech bubble as well as the current topic (one of the space ship parts) in both the imitation phase (in a random order) and in the recall phase (in the order chosen by the participant by using the buttons at the bottom). In the guessing phases and in the recall phases the objects were chosen by clicking on the buttons at the bottom.*

C.3 Transmission chains

The tables printed on the following pages, spanning two pages each, display the transmission chains that resulted from the whistle experiment with meanings (described in chapter 6). The first four tables represent the languages that emerged in the intact condition. Here, the first row shows the meanings (the objects) and each row represents the last recalled output of the participants in the chain where whistles are printed as pitch tracks on a semitone scale. The first row of whistle sounds shows the initial input and then each row shows what the consecutive participant produced for the object in each column. The following four tables represent the languages that emerged in the scrambled condition. Here, the objects are not shown since these were randomly reassigned and replaced from generation to generation. The first row shows the initial input set of whistle sounds and each following row represents the last recalled output of consecutive participants in the chain. Columns represent the transmission of the specific whistle sounds: even though these were paired with different objects in the experiment, the next row shows what the next participant produced for the object that got paired with the whistle from the previous generation in the same column.

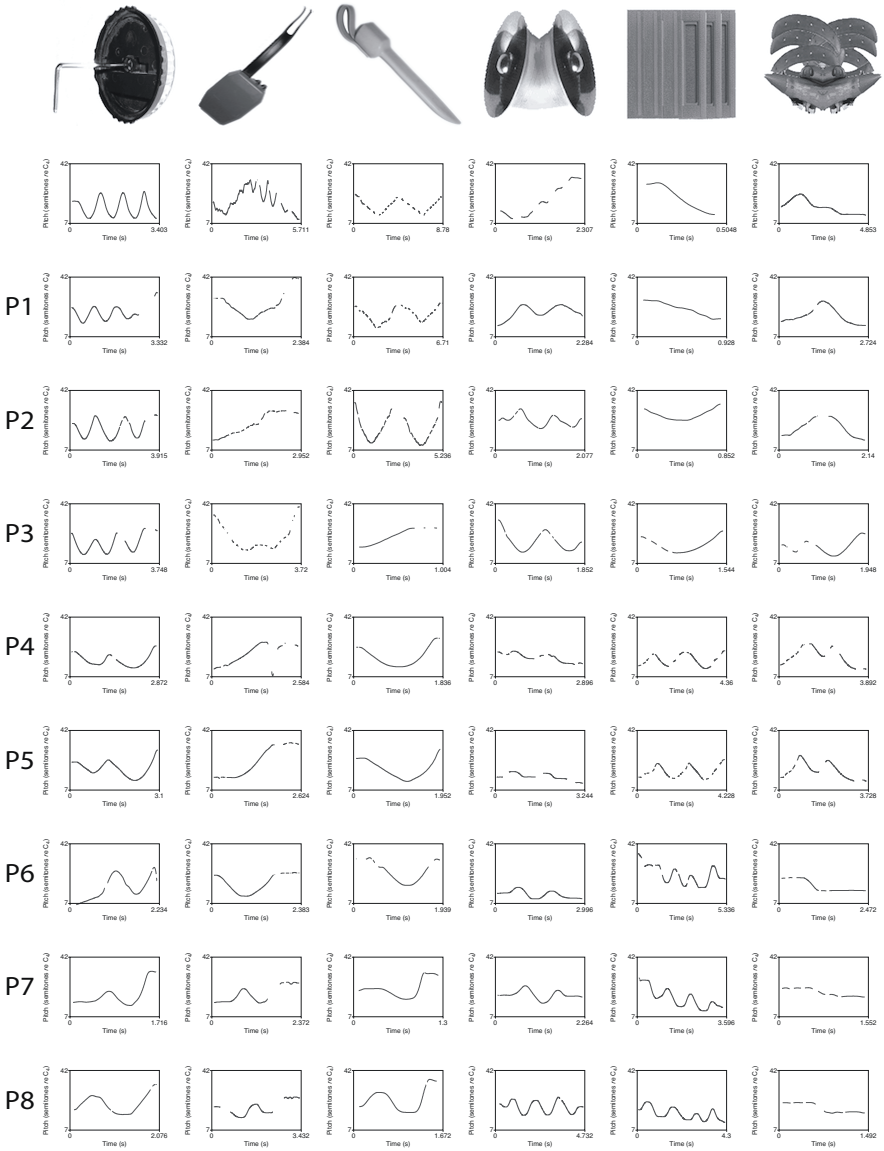


Figure C.3.1: *Transmission chain one of the intact condition in the whistle experiment with meanings (chapter 6). The first row shows the meanings (the objects) and each row represents the last recalled whistles in each generation. The first row of whistle sounds shows the initial input and then each row shows what the consecutive participants (P 1 to 8) produced for the object in each column.*

Figure C.3.2: *Chain one of the intact condition continued*

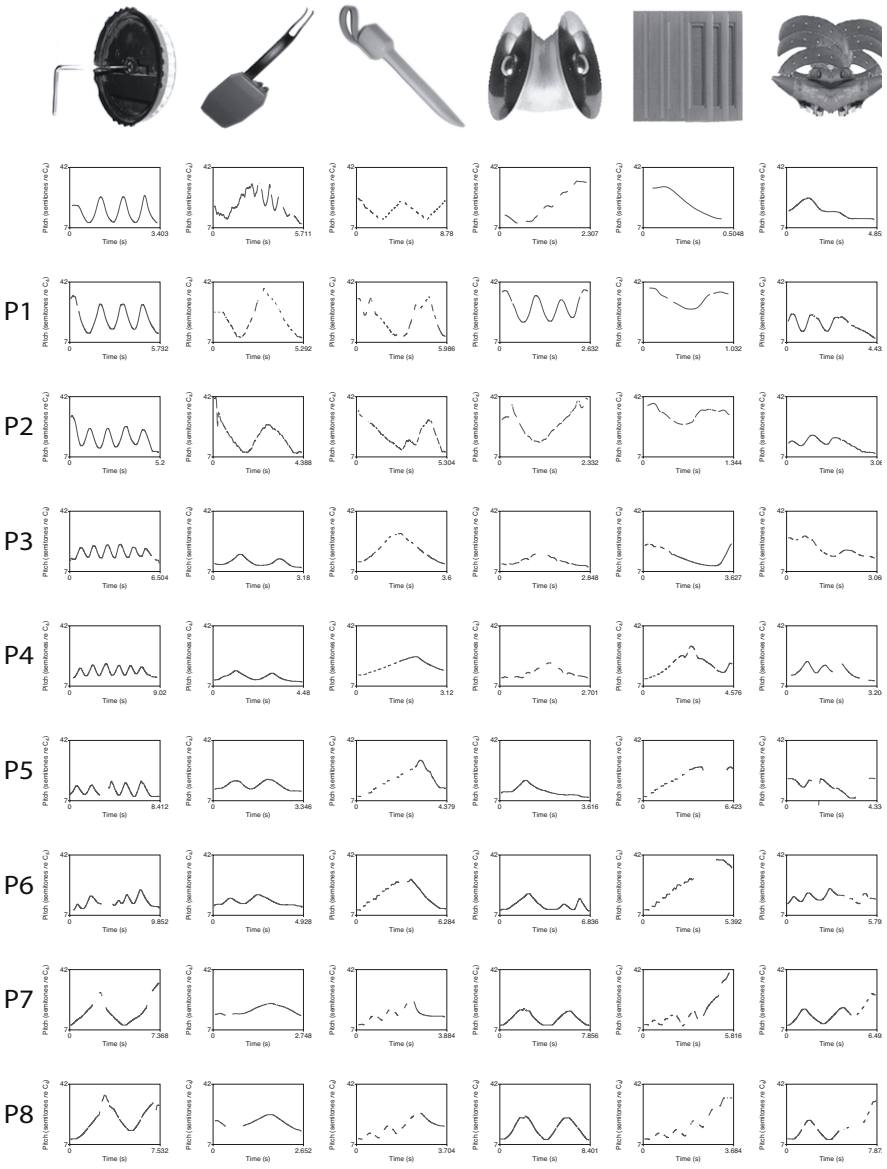


Figure C.3.3: *Transmission chain two of the intact condition in the whistle experiment with meanings (chapter 6). The first row shows the meanings (the objects) and each row represents the last recalled whistles in each generation. The first row of whistle sounds shows the initial input and then each row shows what the consecutive participants (P 1 to 8) produced for the object in each column.*

Figure C.3.4: *Chain two of the intact condition continued*

Figure C.3.5: *Transmission chain three of the intact condition in the whistle experiment with meanings (chapter 6). The first row shows the meanings (the objects) and each row represents the last recalled whistles in each generation. The first row of whistle sounds shows the initial input and then each row shows what the consecutive participants (P 1 to 8) produced for the object in each column.*

Figure C.3.6: *Chain three of the intact condition continued*

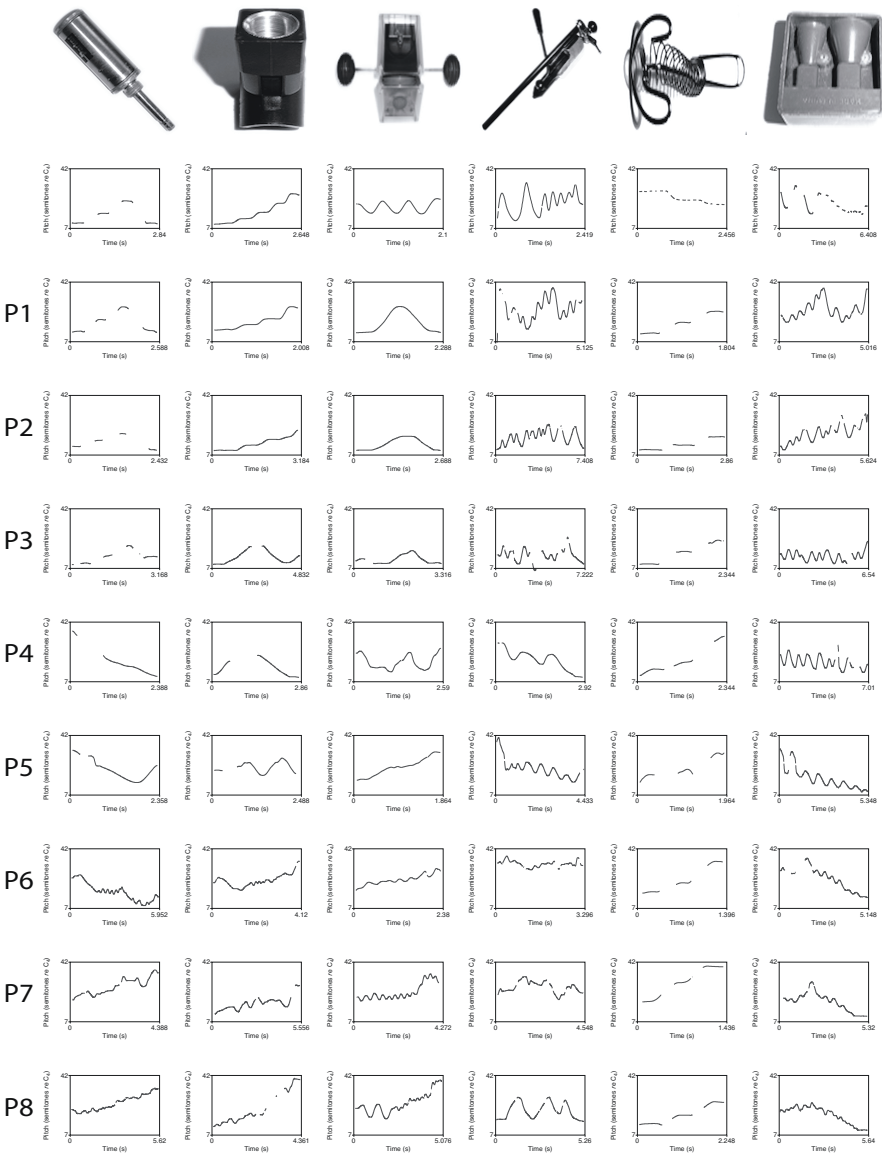


Figure C.3.7: *Transmission chain four of the intact condition in the whistle experiment with meanings (chapter 6). The first row shows the meanings (the objects) and each row represents the last recalled whistles in each generation. The first row of whistle sounds shows the initial input and then each row shows what the consecutive participants (P 1 to 8) produced for the object in each column.*

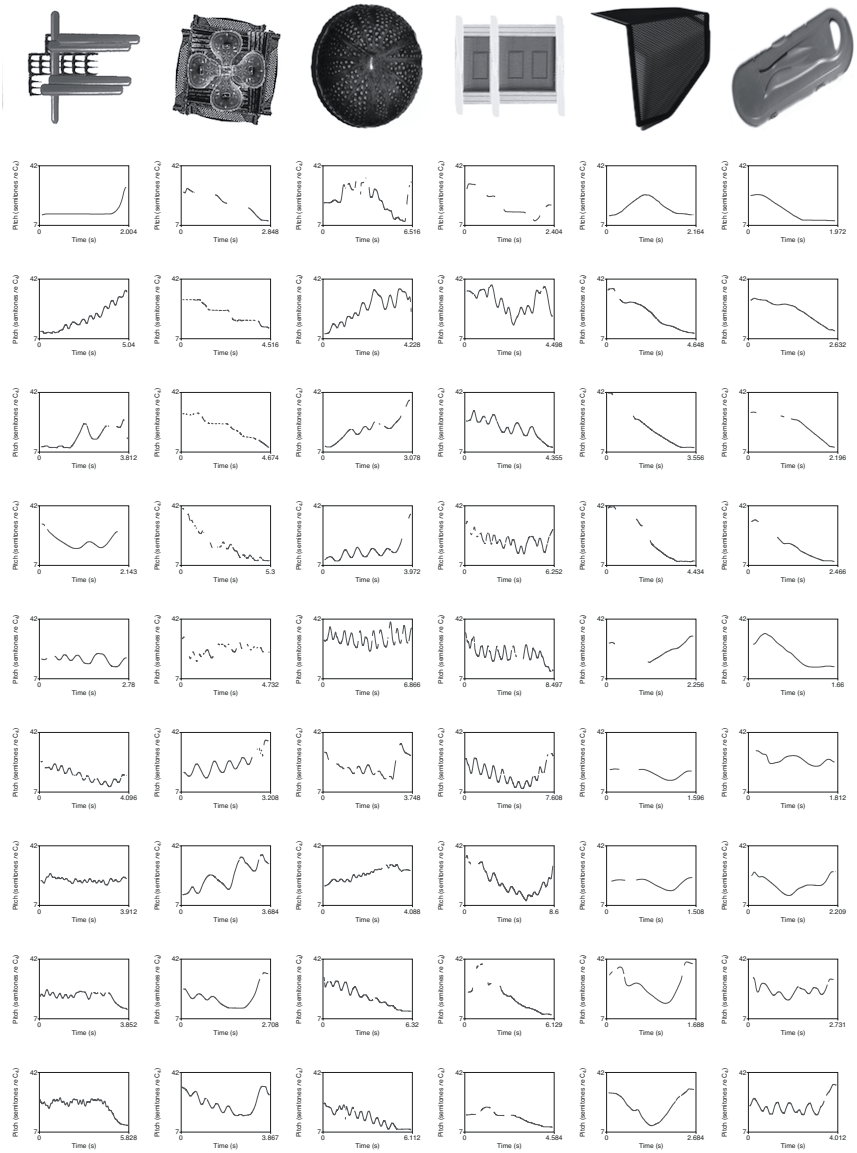


Figure C.3.8: *Chain four of the intact condition continued*

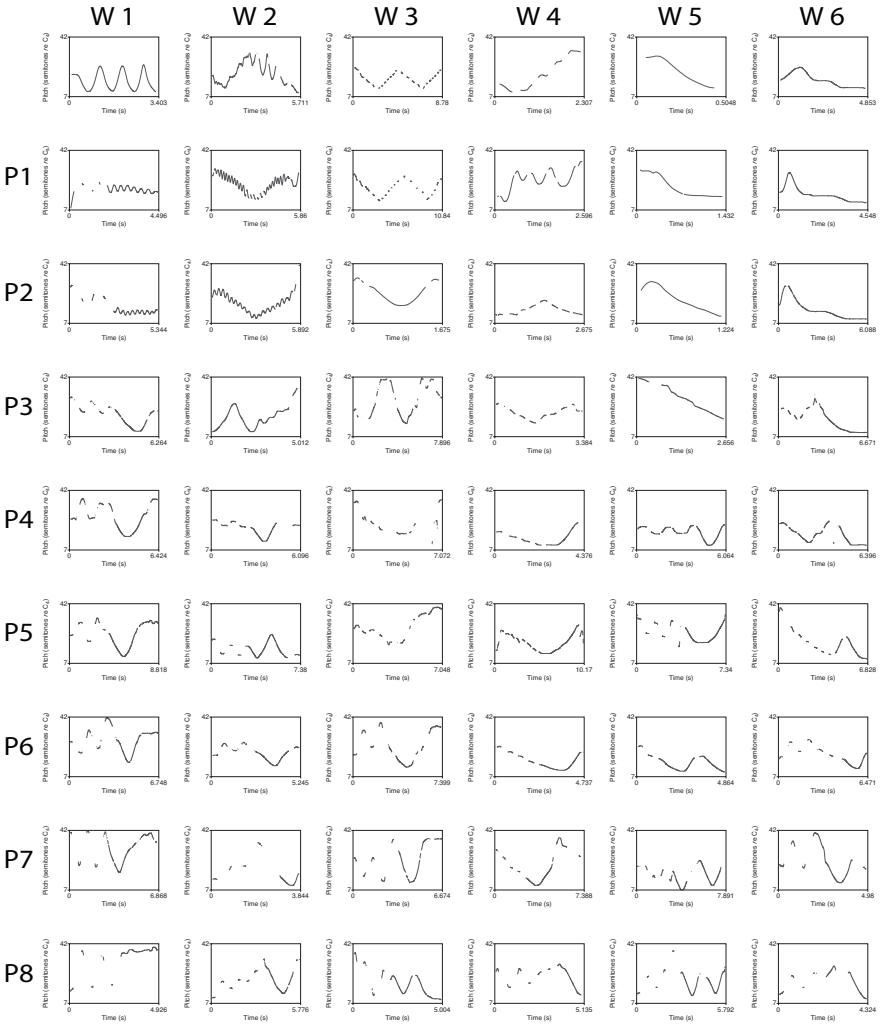


Figure C.3.9: *Transmission chain one of the scrambled condition in the whistle experiment with meanings (chapter 6). The first row shows the initial set of whistle sounds (W 1 to 12) and each following row shows the last recalled output of consecutive participants (P 1 to 8) in the chain. Columns represent transmission of specific whistle sounds: even though these were paired with different objects in the experiment, the next row shows what the next person produced for the object that got paired with the whistle from the previous generation in the same column.*

Figure C.3.10: *Chain one of the scrambled condition continued*

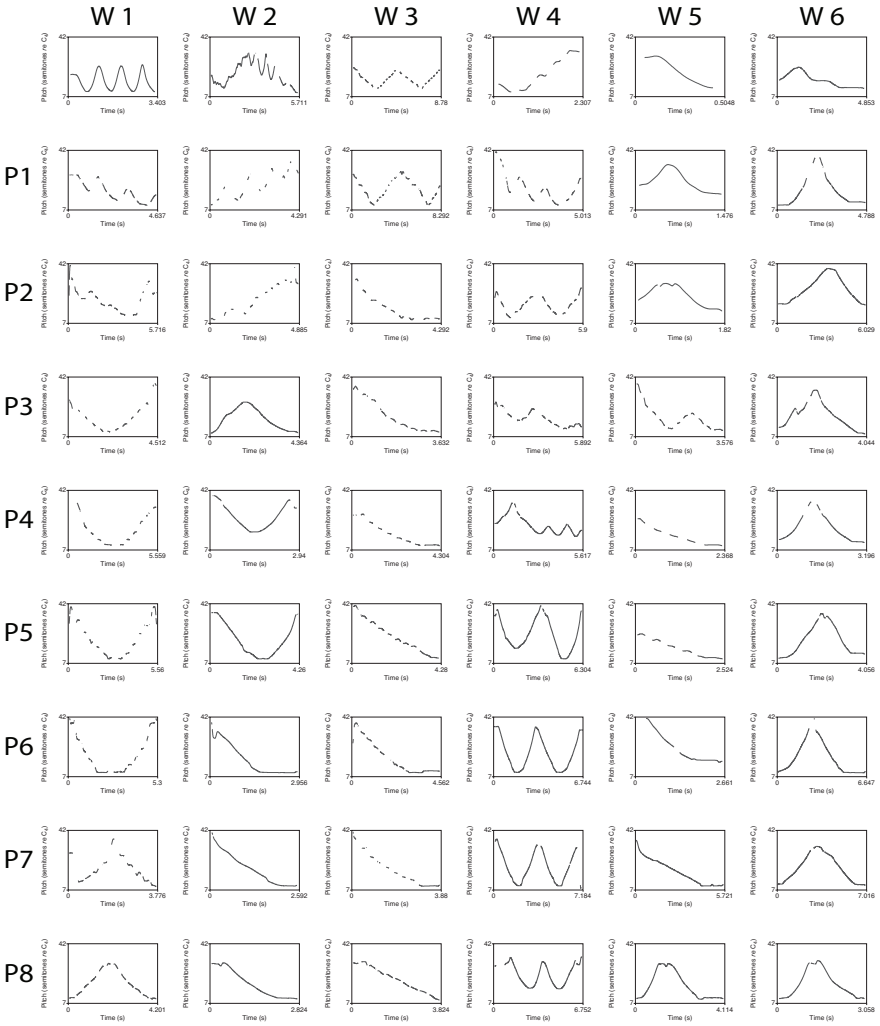


Figure C.3.11: *Transmission chain two of the scrambled condition in the whistle experiment with meanings (chapter 6). The first row shows the initial set of whistle sounds (W 1 to 12) and each following row shows the last recalled output of consecutive participants (P 1 to 8) in the chain. Columns represent transmission of specific whistle sounds: even though these were paired with different objects in the experiment, the next row shows what the next person produced for the object that got paired with the whistle from the previous generation in the same column.*

Figure C.3.12: *Chain two of the scrambled condition continued*

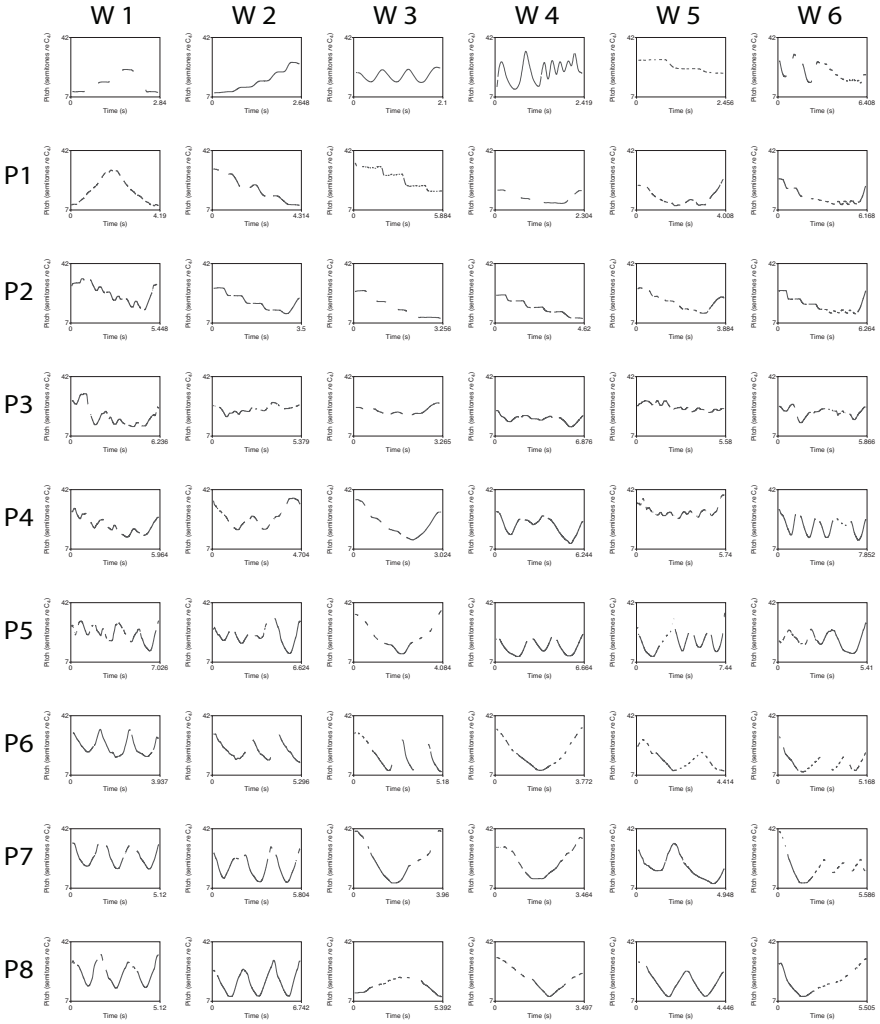


Figure C.3.13: *Transmission chain three of the scrambled condition in the whistle experiment with meanings (chapter 6). The first row shows the initial set of whistle sounds (W 1 to 12) and each following row shows the last recalled output of consecutive participants (P 1 to 8) in the chain. Columns represent transmission of specific whistle sounds: even though these were paired with different objects in the experiment, the next row shows what the next person produced for the object that got paired with the whistle from the previous generation in the same column.*

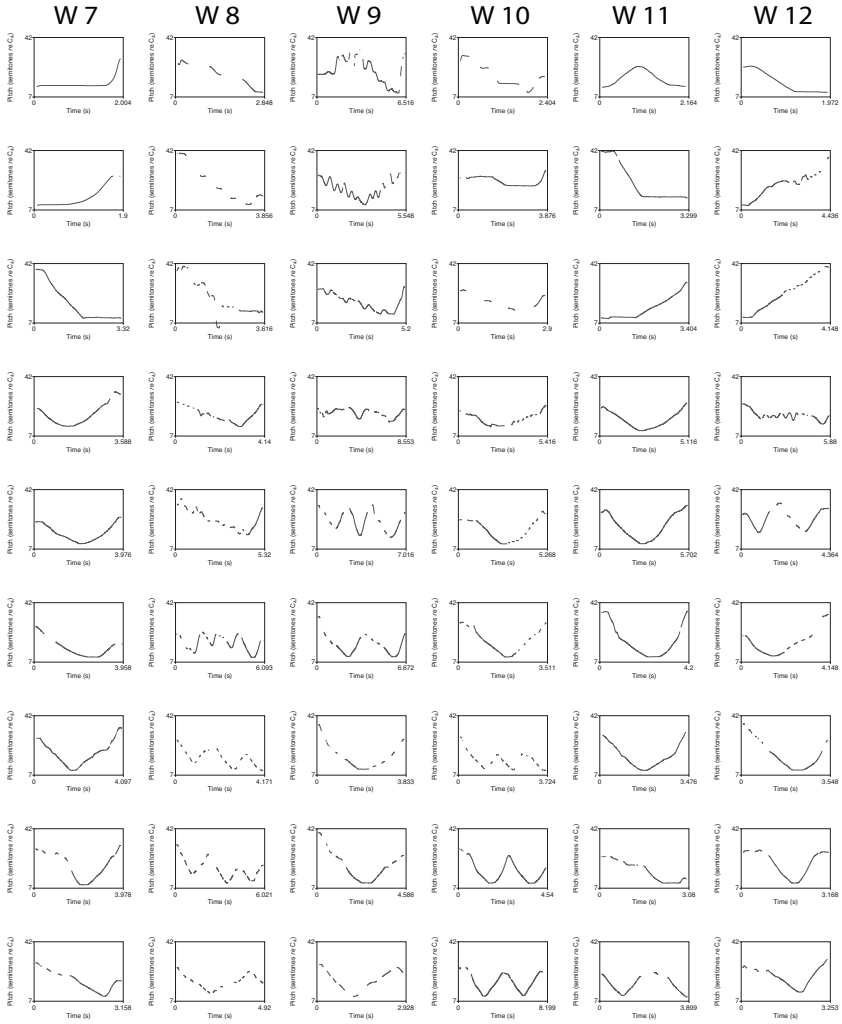


Figure C.3.14: *Chain three of the scrambled condition continued*

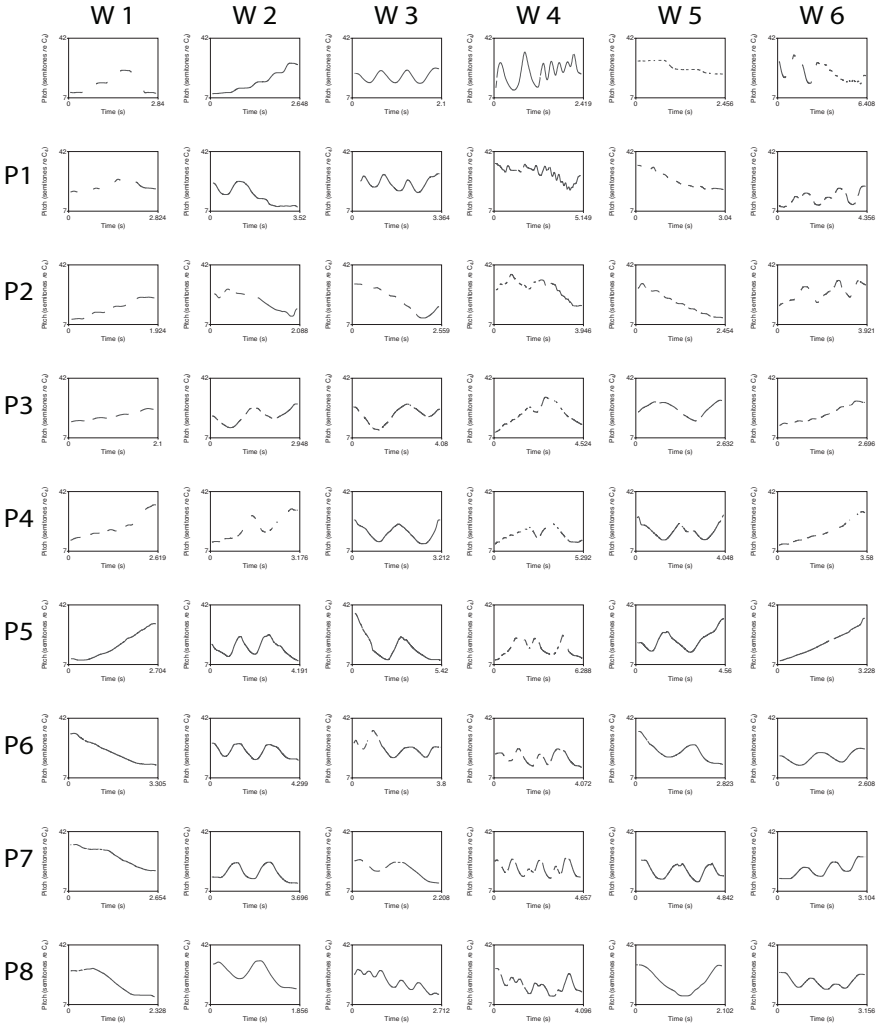


Figure C.3.15: *Transmission chain four of the scrambled condition in the whistle experiment with meanings (chapter 6). The first row shows the initial set of whistle sounds (W 1 to 12) and each following row shows the last recalled output of consecutive participants (P 1 to 8) in the chain. Columns represent transmission of specific whistle sounds: even though these were paired with different objects in the experiment, the next row shows what the next person produced for the object that got paired with the whistle from the previous generation in the same column.*

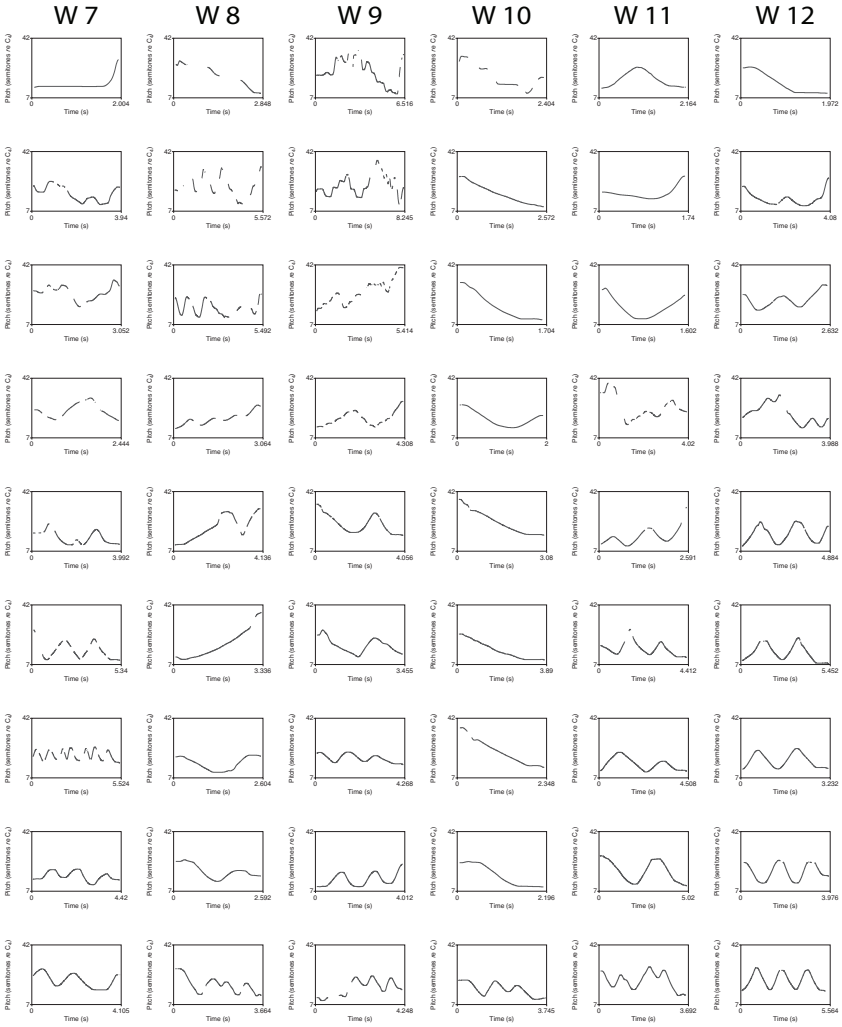


Figure C.3.16: *Chain four of the scrambled condition continued*

Summary

Language is one of the most important features that separate us humans from the rest of the animal kingdom. This is why there is a great interest in discovering how language arose. This is not easy to find out, because there is not much tangible evidence to be found in this area. For a long time scientists could do little more than to use their imagination to develop theories about the evolution of language.

In the meantime, a lot has changed. Many researchers in the field of language evolution collaborate with researchers from other fields of study and new empirical methods have been developed. Geneticists for instance now search for unique genes that may explain human linguistic behaviour; computer modellers analyse and simulate evolutionary scenarios and interactions between individuals; linguists head into the field and study newly emerging (sign) languages; cognitive scientists and psychologists conduct experiments in which human participants learn or invent artificial languages and more.

The main question that is the focus of this thesis is: How did structure in speech arise? Speech is made up of basic building blocks: meaningless sounds are combined into words. Complex rules determine which combinations of sounds are correct in a language and which are not. How this property of language, *combinatorial structure*, emerged is still unclear. Some researchers assume the driving force behind it has to do with signal distinctiveness: the sounds used in language need to differ from each other maximally, otherwise words would sound too similar and we would be less well able to understand each other. Other researchers have proposed that principles of efficient coding play a role: a small set of sound primitives are reused and combined in a maximally efficient way. The results presented in this thesis show that the first assumption alone is not sufficient and that the second indeed seems to play an important role.

Experiments with humans and virtual robots

In my research I use two methods: experiments with human participants and computer simulations. The experiments can de compared to the game of ‘Chinese Whispers’ (or ‘Telephone’, ‘Broken Telephone’). In this

game a person whispers a message in someone else's ear and this message is passed on from person to person until it reaches the last person. The last person then says the message out loud and usually it is very different from the initial message, with often funny alterations. This game demonstrates in miniature what happens when languages are learned and reproduced repeatedly and transmitted from generation to generation. We call this process *cultural transmission* and this can be simulated in the lab by doing Chinese Whispers with entire (artificial) languages instead of single messages. Participants learn an artificial miniature language and are asked to reproduce this language. These reproductions are then passed on by asking the next participant to learn them. In this manner a chain of transmission is created and the transmitted language can be investigated. This method is called *iterated learning*. In the computer simulations, individual *agents* (virtual robots) interact with each other. These agents can produce and perceive sounds and learn an artificial language by dynamically updating their memory in response to interactions with others.

Figure 1: *Participant during whistle experiment in the studio.*

Evolving whistled languages

Natural language already has structure and regularities. How then, can we investigate the emergence of such structure in the laboratory with modern humans? In the experiments described in this thesis participants do not learn an existing spoken language, but a fictional miniature language. Participants cannot use their voice, but use an alternative, non-linguistic, device for sound production. In chapter 3 this device consisted of a two-dimensional interface in which scribbles on the screen were transformed into sounds. This appeared to be quite difficult to use and led to the use of whistles in subsequent studies. In chapter 4 and 6 the words of the artificial languages are whistled with a slide whistle. These whistled languages were transmitted and evolved in the laboratory with iterated learning. Figure 1 shows a participant with a whistle in the studio.

In chapter 4 the languages did not have meanings. Participants simply had to memorise a set of sounds. If we examine one of those evolved whistle languages, we can see that the set of sounds has gained a type of structure that is reminiscent of what we see in real languages. After some transmissions, a few basic building blocks can be identified and these elements are reused and combined in a systematic way. Figure 2 shows some of the sounds of such a language, where the pitch is plotted against time. Basic elements are clearly visible.

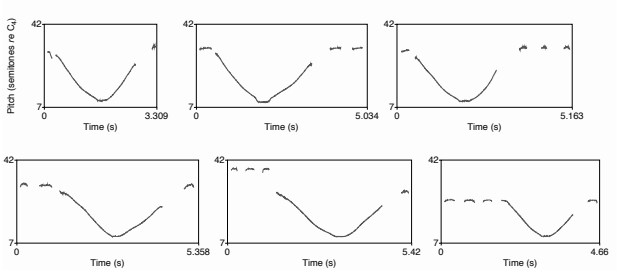


Figure 2: *Fragment of an evolved whistle language, plotted as pitch tracks. Basic elements, such as short level tones and falling-rising contours can be identified and they are systematically recombined.*

This is just one example, but the experiment was repeated four times. Each time a new chain of transmission was started with an unstructured source whistle language. Overall, the results show that the languages become easier to learn and more structured after a number of transmissions. This happens gradually and the participants are not aware of this.

In chapter 6 meanings were attached to the whistled signals. Participants were told that an alien space ship had crashed on earth and that the aliens needed help to repair their ship. To be able to help the friendly extraterrestrials, participants had to learn whistled words for alien space ship parts. Figure 3 shows a few examples of such ‘space ship parts’. In this experiment with meanings, the results of the first whistle experiment were replicated. Even when the signals refer to meanings, and signal-meaning pairings could potentially be iconic and holistic, combinatorial structure emerged in all transmission chains.

UFO game experiments

As part of the analysis of the emerging miniature languages in the whistle experiments, another experiment was conducted which is described in chapter 5. The data for this experiment was collected both online and in Science Center NEMO and it consisted of a game in which participants had to save or destroy UFO’s[1]. The UFO’s contained aliens who spoke either the language of the good kind, or the language of the evil kind.

[1]This game was created by Jelle Zuidema and Vanessa Ferdinand

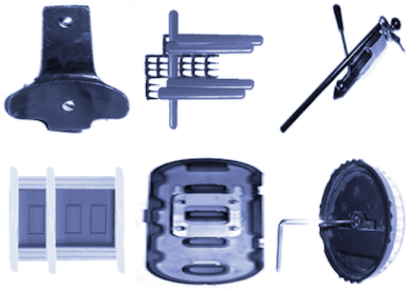


Figure 3: *Examples of 'space ship parts' used in the experiment.*

By listening to the sounds the aliens made, participants had to decide whether to shoot or save the UFO. Figure 4 shows a screenshot of the game.

The goal of this experiment was to investigate whether the evolved artificial languages from the previous experiments could be learned and distinguished by humans. The alien speech coming from the UFO's was constructed using the sounds from the whistle experiment. Two different conditions were created so that one group of participants got to listen to complete structures, while the others were exposed to random sounds, with no structure at all. If it is indeed true that the alien languages evolved in the experiments to become more learnable through an increase of structure, we would expect the first group, with exposure to the complete alien languages, to score much better at distinguishing between good and bad UFO's. This was indeed the case. On average, participants who could make use of the structure scored much higher than the other group.

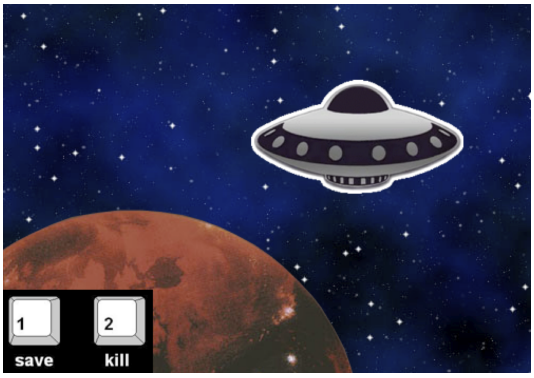


Figure 4: *Screenshot of the UFO game.*

Preservation of structure in populations of agents

Most of the chapters in this thesis focus on how linguistic structure emerges and develops when it is transmitted over generations. We have seen that the structures get simplified, more constrained and they become easier to learn. In addition to this focus on linguistic change, chapter 7 investigates how complexity and mutual intelligibility may be *preserved* over generations. The computer simulations described in that chapter involve experiments in which the emergence and development of artificial vowel systems is studied in populations of interacting agents. The main aim was to show how the preservation of complexity in vowel systems would be influenced when children learn faster than adults. Complexity appears to be preserved better over generations in populations where agents have such a *critical period*.

Conclusion

What this research can teach us is that structure in speech sounds can emerge as languages are repeatedly transmitted from generation to generation. Previously, the important influence of cultural transmission on structure in language has been studied for aspects of language such as syntax in detail, but for the study of complex compositional structure in phonology data was more limited. On the basis of those and other earlier findings, it has been pointed out that languages undergo their own evolutionary process and change gradually. Every generation of speakers changes the language a little bit without being aware of it. Within a language there is selection on structures that are learnable. Unnecessarily complicated rules or words will eventually disappear because speakers will not reproduce utterances they cannot learn. In this manner, cultural evolution causes languages to adapt to the human brain and become more learnable.

This is the first time iterated learning experiments have been conducted to study continuous sounds. My results provide additional support for the above mentioned ideas. Cultural evolution seems to be important in shaping not only compositional syntax, but phonological structure as well, because combinatorial structure in sounds can in principle emerge as the result of transmission and cognitive biases. The way in which structure emerges in the artificial languages in my experiments conforms with theories in evolutionary phonology that are based on principles of efficient coding. In general, iterated learning experiments appear to result in transmitted systems that become more compressible and more predictable. In future work theories on information-theoretic principles in neuroscience will be linked to these findings to gain a better understanding of the neurocognitive biases involved in the emergence of structure.

Samenvatting

Taal is een van de belangrijkste kenmerken waarin wij ons als mensen onderscheiden van de rest van het dierenrijk. Er is daarom grote interesse om erachter te komen hoe taal precies is ontstaan. Dit is niet gemakkelijk te achterhalen, want er is weinig bewijsmateriaal beschikbaar op dit gebied. Wetenschappers moesten daarom vroeger veelal op hun fantasie vertrouwen om theorieën te bedenken over het ontstaan van taal.

Inmiddels is er gelukkig veel veranderd. Veel wetenschappers in de taalevolutie werken samen met onderzoekers uit andere vakgebieden en nieuwe empirische onderzoeksmethoden zijn ontwikkeld. Momenteel wordt er bijvoorbeeld door genetici gezocht naar unieke genen die mogelijk menselijk taalgedrag kunnen verklaren; computerkundigen analyseren en simuleren evolutionaire scenario's en interacties tussen individuen; taalwetenschappers reizen de wereld rond om net nieuw ontstane (gebaren-)talen te onderzoeken; cognitiewetenschappers en psychologen doen experimenten waarbij proefpersonen kunstmatige talen leren.

De hoofdvraag in dit proefschrift is: Waar komt structuur in spraak vandaan? Spraak is opgebouwd uit bouwsteentjes: betekenisloze geluiden worden gecombineerd tot woorden. Complexe regels bepalen welke combinaties van geluiden correct zijn in een taal en welke niet. Hoe deze eigenschap van taal, *combinatorische structuur*, precies is ontstaan is nog onbekend. Sommige onderzoekers gaan ervan uit dat een belangrijke drijfkracht was om de geluiden die voor spraak worden gebruikt zo veel mogelijk van elkaar te laten verschillen, anders zouden woorden veel te veel hetzelfde klinken en zouden we elkaar niet goed verstaan. Andere onderzoekers zeggen dat principes van efficiënt coderen een belangrijke rol spelen: kleine sets van basisgeluiden worden gecombineerd en efficiënt hergebruikt.

Experimenten met mensen en virtuele robots

In mijn onderzoek gebruik ik twee methoden: experimenten met proefpersonen en computersimulaties. De experimenten kunnen we vergelijken met het spel doorfluistertje. In dit spel wordt een bericht van persoon tot persoon doorgefluisterd totdat het de hele kring rond is

geweest. De laatste persoon zegt vervolgens hardop wat er werd doorgefluisterd. Meestal is dit heel wat anders dan het oorspronkelijke bericht, met vaak grappige vervormingen. Dit spel demonstreert in het klein wat er gebeurt als talen herhaaldelijk worden geleerd en doorgegeven van generatie op generatie. Dit noemen we *culturele transmissie* en dit proces kan worden nagebootst in het laboratorium door doorfluistertje te doen met hele (kunstmatige) talen in plaats van losse berichten. Proefpersonen leren een kunstmatige mini-taal en worden daarna gevraagd om deze taal te reproduceren. Die reproducties worden vervolgens doorgegeven door de volgende persoon deze te laten leren. Zo ontstaat er een keten van transmissie en kan de overgedragen taal worden onderzocht. Deze methode heet *iterated learning*. In de computersimulaties maken *agents* (virtuele robots) deel uit van een populatie waarin ze met andere individuen interacteren. Deze agents kunnen geluiden produceren en waarnemen en leren door dynamisch hun geheugen aan te passen na interacties met anderen.

Figure 1: *Proefpersoon in het trekfluit experiment in the studio.*

Evoluerende fluittalen

Natuurlijke spraak heeft al regelmaat en proefpersonen kunnen al spreken. Hoe kunnen we dan het ontstaan van deze structuur onderzoeken? In de experimenten die beschreven staan in dit proefschrift leren proefpersonen geen bestaande gesproken taal, maar fictieve miniatuurtalen. Proefpersonen kunnen hun stem hierbij niet gebruiken, maar moeten een alternatief, niet-talig, apparaat gebruiken voor het produceren van geluiden. In hoofdstuk 3 bestond dit uit een tweedimensionaal vlak waarin muisbewegingen op het scherm werden vertaald naar geluid. Dit bleek lastig te zijn voor proefpersonen om te leren en daarom werd er in de studies daarna gebruik gemaakt van fluitsignalen. In hoofdstuk 4 en 6 worden de woorden van de kunstmatige taal gefloten met een trekfluitje. De evoluerende fluittalen werden doorgegeven van persoon tot persoon met *iterated learning*. Figuur 1 laat een proefpersoon zien met een trekfluit in de studio.

In het experiment dat beschreven staat in hoofdstuk 4 kregen proefpersonen een set betekenisloze geluiden te leren als fluittaal. Als we nu kijken naar de geëvolueerde fluittalen, dan zien we dat deze na een aantal keer ‘doorfluisteren’ een structuur hebben gekregen die veel lijkt op wat we zien in natuurlijke talen. Er is een set basiselementen ontstaan en deze elementen worden op een systematische manier hergebruikt en gecombineerd. In figuur 2 zijn de geluiden van zo’n taal afgebeeld, waarbij de toonhoogte is uitgezet tegen de tijd. De basiselementen zijn duidelijk van elkaar te onderscheiden.

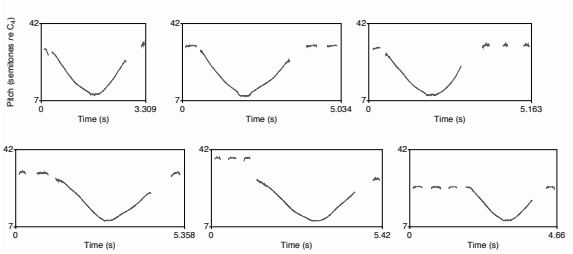


Figure 2: *Fragment van een geëvolueerde fluittaal, waarbij de toonhoogte is uitgezet tegen de tijd. Basiselementen zoals korte toontjes en contouren die dalen en stijgen worden gecombineerd.*

Dit is één voorbeeld, maar in totaal werd het experiment vier keer herhaald. Hierbij werd er vier keer een nieuwe keten van transmissie gestart met een onregelmatige bron-fluittaal. Als we kijken naar de globale resultaten dan zien we dat de fluittalen gemiddeld na een aantal transmissies makkelijker te leren worden en steeds meer structuur krijgen. Dit gebeurt geleidelijk en zonder dat de deelnemers zich hiervan bewust zijn.

In hoofdstuk 6 werd er betekenis aan de fluittalen verbonden. Proefpersonen kregen te horen dat een buitenaards ruimteschip was neergestort op aarde en dat de aliens hulp nodig hadden bij het repareren van hun UFO. Om de aliens te kunnen helpen moesten de proefpersonen fluitwoorden leren voor buitenaardse UFO onderdelen. In figuur 3 worden een aantal van deze onderdelen getoond. In dit experiment met betekenis werden de eerdere bevindingen gerepliceerd. Zelfs als de fluitsignalen aan een betekenis verbonden zijn, en de relatie tussen signaal en betekenis potentieel iconisch en holistisch kan zijn, ontstaat er snel combinatorische structuur in de transmissieketens.

UFO spel experimenten

Als onderdeel van de analyse van de evoluerende fluittalen, werd er nog een experiment uitgevoerd, zoals beschreven in hoofdstuk 5. De data voor dit experiment werd online verzameld en in Science Center NEMO. Proefpersonen moesten in dit experiment een spel[1] spelen, waarin zij

[1] Dit spel is gemaakt door Jelle Zuidema en Vanessa Ferdinand

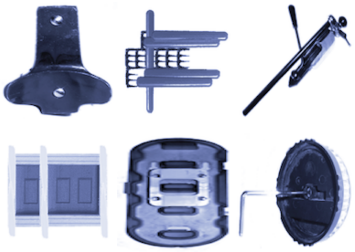


Figure 3: *Voorbeelden van 'UFO onderdelen' die gebruikt werden bij het experiment.*

UFO's moesten redden of vernietigen. In de UFO's zaten buitenaardse wezens die ofwel de taal van de 'goede' soort spraken, ofwel de taal van de 'kwade' soort. Door naar de spraakgeluiden te luisteren moet de speler beslissen of de UFO gered of vernietigd moest worden. Figuur 4 laat een screenshot zien van het spel.

Het doel van dit UFO experiment was om te onderzoeken hoe goed de experimentele buitenaardse talen door mensen te leren en te onderscheiden zijn. De spraakgeluiden die uit de UFO's kwamen in dit spel waren afkomstig uit het eerder uitgevoerde fluitexperiment dat hierboven besproken is. Hierbij werd er voor gezorgd dat een deel van de proefpersonen talen met een volledige structuur te horen kregen, terwijl de andere proefpersonen geluiden te horen kregen waar geen enkele structuur uit op te maken was. Als het inderdaad waar is dat de buitenaardse talen door de experimentele transmissie makkelijker te onthouden worden door een groei van de structuur, dan zou de eerste groep proefpersonen beter moeten scoren met het onderscheiden van goede en kwade UFO's. Dit was inderdaad het geval. Gemiddeld scoorden de proefpersonen die gebruik konden maken van de structuur hoger dan de andere groep.

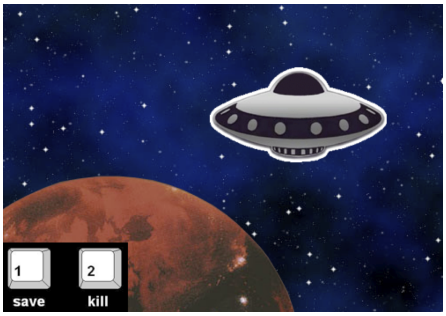


Figure 4: *Screenshot van het UFO spel.*

Behoud van structuur in populaties van agents

De meeste hoofdstukken van dit proefschrift gaan over het ontstaan en de ontwikkeling van structuur als taal wordt doorgegeven van generatie tot generatie. We hebben gezien dat structuren versimpelen, meer afgebakend raken en leerbaarder worden. Naast deze focus op verandering, wordt er in hoofdstuk 7 onderzocht hoe complexiteit en onderlinge verstaanbaarheid behouden blijven over generaties. De computermodellen simuleren experimenten waarin klinkersystemen ontstaan in een populatie van interacterende agents. Het doel was om te kijken hoe het behoud van complexiteit in de klinkersystemen beïnvloed zou worden wanneer kinderen sneller leren dan volwassenen. Het blijkt dat complexiteit beter behouden blijft over generaties in het geval dat agents zo'n *kritieke periode* hebben voor het leren van taal.

Conclusie

Wat we kunnen leren van dit onderzoek is dat structuur in taal kan ontstaan doordat de taal van generatie tot generatie wordt doorgegeven. Dat culturele evolutie een belangrijke invloed heeft op taal werd al vaker onderzocht, voornamelijk voor aspecten van taal zoals syntax, maar voor fonologische structuur was de beschikbare data beperkter. Op basis van die (en andere) eerdere bevindingen werd vastgesteld dat talen hun eigen evolutie doormaken. Elke generatie sprekers verandert de taal onbewust een klein beetje. Binnen de taal is er hierbij selectie op leerbare structuren. Onnodig ingewikkelde regels of woorden zullen op den duur verdwijnen omdat sprekers datgene wat ze niet kunnen leren ook niet zullen reproduceren. Op deze manier zorgt culturele evolutie er dus voor dat de taal zich aanpast aan wat het brein makkelijk vindt om te leren en onthouden.

Dit is de eerste keer dat *iterated learning* experimenten gedaan zijn met continue signalen in het auditieve domein. Mijn resultaten wijzen in dezelfde richting als de bovengenoemde ideeën. Culturele transmissie lijkt niet alleen belangrijk te zijn in de vorming van syntax, maar ook voor fonologische structuur. De experimenten laten namelijk zien dat combinatorische structuur in principe kan ontstaan als gevolg van herhaald doorgeven en cognitieve leerpatronen. De manier waarop structuur ontstaat in de kunstmatige talen in mijn experimenten is in overeenstemming met theorieën in de fonologie die gebaseerd zijn op principes van efficiënt coderen. In het algemeen lijken resultaten van *iterated learning* experimenten te leiden tot systemen die comprimeerbaarder en voorspelbaarder worden. In de toekomst zullen vermoedelijk gerelateerde ideeën in de neurowetenschappen helpen een beter beeld te krijgen van de cognitieve leerpatronen die een rol spelen bij het ontstaan van structuur.

Curriculum Vitae

Tessa Verhoef was born on March 3rd, 1985 in Maarssen, the Netherlands. She obtained her BSc (with honors) and MSc (cum laude) degrees in Artificial Intelligence from the University of Groningen in the Netherlands with a specialization in Robotics and Machine Learning. For her MSc thesis, she conducted research within the Affective Social Computing Lab at Florida International University in Miami, FL, USA. In January 2009, she started her PhD research at the Amsterdam Center for Language and Communication, University of Amsterdam, the Netherlands, which resulted in the present dissertation. During her PhD project she was affiliated with the Center for Research in Language at the University of California, San Diego (UCSD), USA, as a visiting scholar and was there to conduct research from June to October 2010, followed by two more shorter visits. In February and March 2013 she returned to UCSD to work as a staff research associate and from May 1st until July 15th 2013 she worked as a temporary postdoc at the Artificial Intelligence Lab at the Free University of Brussels, Belgium. She is a recipient of a Rubicon grant from the Dutch Science Foundation (NWO) and this gives her the opportunity to work as a postdoc for two years at UCSD, starting December 2013.