Birds and babies : a comparison of the early development in vocal learners

Haar, S.M. ter

Citation

Haar, S. M. ter. (2013, June 5). *Birds and babies : a comparison of the early development in vocal learners*. Retrieved from https://hdl.handle.net/1887/20932

Author: Haar, Sita Minke ter
Title: Birds and babies : a comparison of the early development in vocal learners
Issue Date: 2013-06-05

BIRDS AND BABIES

A COMPARISON OF THE EARLY DEVELOPMENT IN VOCAL LEARNERS

Sita M. ter Haar

Birds and Babies:

A comparison of the early development in vocal learners

Sita M. ter Haar

Sita M. ter Haar

Birds and babies: A comparison of the early development in vocal learners
Proefschrift Universiteit Leiden

ISBN/EAN: 9789088916397
Een elektronische versie van dit proefschrift is online verkrijgbaar op: https://openaccess.leidenuniv.nl

Gedrukt door Proefschriftmaken.nl
Omslag: vormgeving Sita ter Haar & Chris Klink

Birds and Babies:
A comparison of the early development in vocal learners

PROEFSCHRIFT

ter verkrijging van
de graad van Doctor aan de Universiteit Leiden,
op gezag van Rector Magnificus prof.mr. C.J.J.M. Stolker,
volgens besluit van het College voor Promoties
te verdedigen op woensdag 5 juni 2013
klokke 15.00 uur

door

Sita Minke ter Haar

geboren te Schiedam
in 1982

Promotiecommissie:

Promotor: Prof.dr. Carel J. ten Cate

Co-promotor: Dr. Clara C. Levelt

Overige leden:

Prof.dr. Julia Fischer
Deutsches Primatenzentrum GmbH (DPZ)

Prof.dr. René W.J. Kager
Utrecht Institute of Linguistics OTS

Prof.dr. Niels O. Schiller
Universiteit Leiden

Prof.dr Herman P. Spaink
Universiteit Leiden

Het in dit proefschrift beschreven onderzoek is gefinancierd door het LIBC (Leiden Institute for Brain and Cognition)

Contents

Voor mijn grootouders

Leven vol verwondering

J.L.A Mac Gillavry – van Santen (Ip)
3 sep 1911 – 28 sep 2012

&

Prof.dr. H.J. Mac Gillavry (Mac)
12 juni 1908 – 24 jan 2012

1 INTRODUCTION

One of the difficulties of interdisciplinary research is to define complex terms in a way that is comprehensible for researches of all fields involved. In order to combine such definitions and yet keep this thesis readable without too many interruptions for definitions, I included a glossary with definitions for terminology used throughout this thesis (appendix I, p 118). These terms are written in ***bold*** *in the introduction.*

“Hebban olla uogala nestas hagunnan hinase hi(c) (a)nda thu uuat unbidan uue nu” is a well known sentence in Old Dutch[1] dating from approximately the year 1100. Since then the Dutch language has changed substantially. There are different causes for language change, one of them being the cultural transmission of language, where language is transferred from one person to the next or from one generation to the next. Due to ‘noise’ in the signal transfer in perception or production, changes gradually accumulate. When populations get separated geographically, they go through their own cultural evolution resulting in linguistic variation. The cultural transmission of language exists by virtue of **vocal learning**, a feature that is quite rare in the animal kingdom. Human behavior is usually compared to that of other primates like the chimpanzee, because genetically they are closely related. In terms of language and speech however, the comparison is less appropriate, since non-human primates usually do not seem to learn their vocalizations (Egnor & Hauser, 2004); humans (*Homo sapiens*) are the only primates known to learn their vocalizations. Only a few other animal taxa in the animal kingdom are known to be vocal learners, including songbirds. The vital role of experience with auditory input in the vocal development of humans and songbirds led scientists to instigate comparative research on the two. This does not mean that language or speech and birdsong are the same, but rather that the mechanisms for learning, development and evolution show parallels.

In humans, the cultural transmission of language can lead to the variation that exists between different dialects and languages (Baronchelli et al, 2012). Similarly, in songbirds it can lead to song diversity between populations and individuals Petrinovich and Baptista 1984; Catchpole and Slater 1995 (Gammon et al, 2005). Despite the diversity found in both language and birdsong however, some aspects are remarkably similar between languages (Hyman, 2008; Jakobson, 1941; Jusczyk et al, 2002; Macken & Ferguson, 1981) and songs of the same species of songbirds (Marler & Pickert, 1984; Soha et al, 2009; Soha & Marler, 2000, Fig. 1.1). This may be explained by predispositions or ‘innate’ biases guiding the change of sounds and structures of languages in a direction that is preferred by the learner’s brain. During development, such predispositions may come about in the form of selective attention for certain sounds over others, influencing the learning process (Nelson & Marler, 1993; Soha & Marler, 2001). This would also explain why some sounds are more likely to occur in languages or birdsong populations than other sounds.

Taken together it seems likely that both experience independent processes and experience dependent processes play a role in the acquisition and the evolution of language and birdsong (Lachlan & Feldman, 2003; Yang, 2004). Considerable debate exists, however, on the relative roles of experience dependent and independent processes

---

[1] Possibly the sentence is Old English and not Old Dutch (De Grauwe, 2004). The meaning of the sentence in modern English is thought to be: “All birds have a nest except for me and you, what are we waiting for?”

on language learning and evolution. One of the difficulties of trying to disentangle these variables is that it is rather impossible and quite unethical to experimentally manipulate language exposure to human children during development. A solution to this problem is not to control the exposure but to measure it. This can be used to get a grip on the influence of experience dependent and independent processes on vocal learning and how they interact. Questions like the following can be addressed: How often does a certain syllable occur? If syllables occur equally often are they perceived similarly? If they are not, what causes the difference?

Another way to gain insight in the influence of these processes on vocal learning in general is by comparing different species of vocal learners, like songbirds and human beings. By using songbirds as a model for vocal learning it is possible to study the influence of experience dependent and experience independent processes in a more controlled experimental way.

In this thesis I demonstrate how both these processes influence vocal development in human infants and zebra finches (*Taeniopygia guttata)*. Mainly focusing on acquisition of phonemes and elements in human infants and zebra finches respectively, I show that different processes affect vocal acquisition at different developmental stages. In addition, I present an 'element typology' of zebra finch song in order to illustrate how developmental patterns relate to sound patterns found across populations of the same species. Implications for the development and evolution of language and birdsong are discussed.

1.1 Similarities and differences between language and birdsong

Many similarities have been described between birdsong and language (Berwick et al, 2011; Bolhuis et al, 2010; Doupe & Kuhl, 1999; Yip, 2006). Similar neural mechanisms have been thought to play a role in vocal learning and memorization, including possible hemispheric lateralization (Moorman et al, 2012; Nottebohm et al, 1976). Both language and birdsong have a hierarchical structure. Language consists of phonemes forming syllables and words, while words form sentences. Zebra finch elements (the smallest units of song) can form syllables and sequences of syllables form a motif (Fig. 1.2). Motifs can be repeated to form a song bout. The most important parallel in relation to this thesis is the vocal learning itself. More detailed aspects of development are similar as well, for instance the presence of a sensitive period and babbling. Babbling is called subsong in birds, which later develops into plastic song and eventually crystallized song (i.e. adult song, fig 1.3b, (Eales, 1985; Immelmann, 1969; Price, 1979; Zann, 1996)). Vocal acquisition seems to be influenced by both experience dependent and independent processes. Similar to dialects in human language, both variation and commonalities are present among vocalizations of birds from different populations of the same species.

In addition to parallels, there are some important differences between birdsong

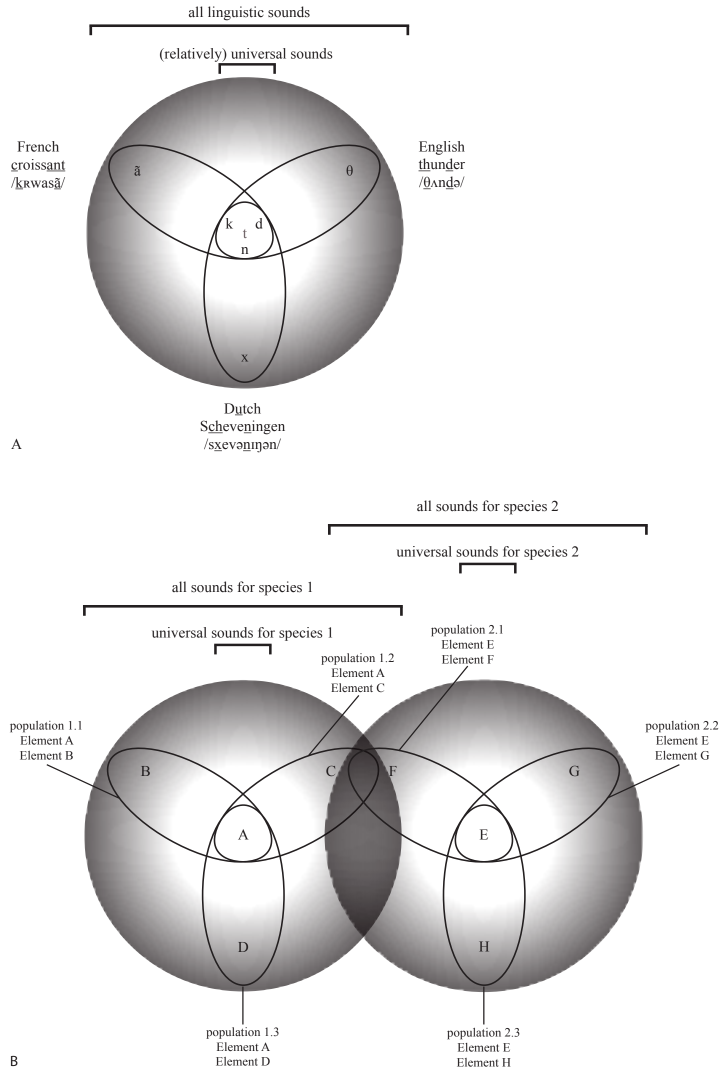


Figure 1.1. Schematic representation illustrating diversity and uniformity in language **(A)** and birdsong **(B).** In language, some sounds are common (relatively universal) across languages whereas others are more specific to certain languages. The sounds /ã/, /θ/ and /x/ are language specific, whereas /k/, /n/ and /d/ are more common. The segment /t/ is thought to be most common across languages **(A).** Similarly, sound diversity and uniformity can be found in birdsong. In contrast to human language, which occurs only in one species, birdsong is present in different species of song birds. Diversity and uniformity can be found across populations within species (as in humans) but also between species **(B)**.

and language. Although some songbirds have quite large repertoires, complex semantics or syntax as in human language has not been found in birdsong to date (Doupe & Kuhl, 1999; ten Cate & Okanoya, 2012). A lively debate is currently going on about whether songbirds can detect recursive patterns. Gentner et al. (2006) claimed that starlings are able to learn a rule of center imbedding and Abe et al (2012) have claimed this too. Both studies, however, have been criticized and alternative explanations for the results have been proposed. In the case of Gentner et al, the birds possibly used a more basic rule to solve the task (van Heijningen et al, 2009). In the study by Abe et al birds may have memorized the test items during training instead of having learnt an abstract syntactical rule (Beckers et al, 2012).

Whereas humans are the only primate species known to learn their vocalizations, the clade of songbirds consists of many different species. Thus, when it concerns birdsong, one has to realize that there is extraordinary diversity both *within and between* the species in what the songs sound like, how they are structured and how they have been shaped by development and evolution. The chipping sparrow (*Spizella passerina*), for instance, sings a relatively simple song consisting of one repeated syllable (Albrecht & Oring, 1995), whereas the nightingale has a repertoire of up to approximately 200 song types (Kipper et al, 2004).

The function of birdsong is also different from the function of human language. Even though language and birdsong are both used for communication, birdsong is used mainly for territory defense or mate attraction. Clear gender differences as found in

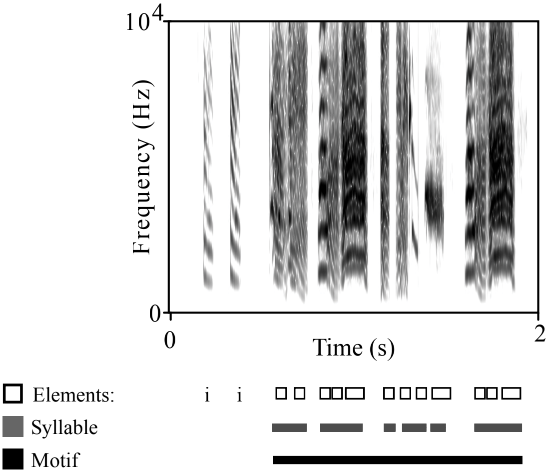


Figure 1.2. A spectrogram of a zebra finch song. A zebra finch song starts with a few introductory notes, followed by one or more motives. A motive consists of a sequence of elements. Elements are grouped into syllables (which are separated by silent intervals).

several songbird species, where most singing behavior is performed by males, are absent in human language.

1.2 Phonological acquisition in human infants

The first issue I address in this thesis is how experience dependent and independent processes influence phonological acquisition in infants. Phonology is the study of sound – phonemes – and how they are organized within and across languages. A **phoneme** is considered the smallest unit in speech. For instance, the consonant /b/ the vowel /ʊ/ and another consonant /k/ are the three phonemes that together form the sounds of the word ‘book’. In language, phonemes are contrastive: substituting phonemes in a word changes the meaning of the word. For example, when in the above mentioned example the phoneme /b/ is substituted by the phoneme /r/, the word turns into ‘rook’ instead of ‘book’. **Phonological typology** is a term used for the study of phonemes and their distribution across languages and dialects. Some phonemes, contrasts and **phonological processes** vary between languages, while others are similar across languages.

Despite linguistic diversity, infants can learn any language they grow up with. There is, on the one hand, considerable variation in language development among children, indicating that the learning process is flexible and experience dependent. On the other hand, certain common patterns can be found both in languages and in language development, indicating that experience independent influences play a role too. The question thus remains how experience dependent and experience independent processes interact during language development.

1.2.1 The relation between language acquisition and typology

Jakobson (1941) already suggested that children acquire “universal” sound contrasts first, regardless of their language environment or culture. He also emphasized the link between language development and patterns found in languages (typology), suggesting first that sounds occurring more often in languages across the world are also likely to be acquired early. Second, he introduced laws of “irreversible solidarity” (or implicational relations) based on cross-linguistic observations of phoneme inventories. These laws specify that one phonemic contrast implies the existence of another, whereby the implied contrasts are those that occur more generally in languages and are acquired first. This relation is not reversible (not every language that has a /t/ also has a /d/)

The observation that some sounds are more likely to appear in languages than others, are more frequent within languages, are acquired earlier, and are more likely to undergo than to trigger phonological processes, has been studied by many other phonologists. The notion of **markedness** is often used to describe these qualitative differences in sound; unmarked sounds are defined as more “universal” (i.e. more common across languages) and are acquired early in development, whereas marked sounds are defined as less common and are acquired later (De Lacy, 2006; Rice, 2007). For this thesis, the most important reason for studying markedness is that both typological as well as developmental data suggest a special status of unmarked sounds. The qualitative difference between marked and unmarked sounds might indicate differences in processing and learning and leads to two potentially contrasting hypotheses.

The first hypothesis is that the special status of unmarked sounds indicates a **predisposition**, a biological sensitivity to specific sounds that have not been heard before. The rationale behind this is that *if* there are predispositions involved in language learning, their effects are likely to show up early in development and therefore they are likely to be common across languages (universal). Alternatively, the second hypothesis is that the higher abundance of unmarked sounds cross-linguistically and in early child language, is a consequence of learning and thus of cultural transmission. This could be due to the fact that within languages the frequency of unmarked sounds is usually high, i.e. markedness correlates with frequency. In other words, the first hypothesis points towards experience independent processes during development and the second to experience dependent processes.

The link between child language data and cross-linguistic data has been further investigated by Zamuner et al (2005). Studying 35 languages, the authors conclude that in coda position (the last position of a word or a syllable) **sonorants** and **coronals** are more abundant across languages than obstruents and labials or dorsals (Zamuner et al, 2005). Child language data, however, show that sonorant and coronals are not the only segments acquired early. Dorsals (for instance /k/) are produced early in coda position as well. In English, sonorants and coronals are frequent but dorsals are as well. In contrast, markedness theories do not predict dorsals to be produced early, leading the authors to conclude that child language data are better explained by input frequency (Zamuner et al, 2005). Nevertheless, they do not claim that acquisition is based on input alone but rather that both theories should be taken into account. For instance, it is unclear whether acquisition of segments in onset position can be explained better by frequency than by markedness. Furthermore, the question remains if this is the case for the very initial set of acquired segments (Levelt & Van Oostendorp, 2007). More importantly, it is not clear why these sounds are so frequent within languages in the first place. Are they frequent due to accumulating cultural transmission or are there other reasons?

In order to further understand language acquisition and its relation to cross-linguistic sound patterns it is necessary to disentangle the effects of markedness, which point to experience-independent effects on acquisition, and the effects of frequency, which point to experience-dependent effects. In this thesis I study these effects separately, in the perceptual behavior of infants. The questions I address are: to which type of information are infants sensitive (experience dependent or –independent)? And how do these sensitivities develop with age? The data known from typology and child language production (markedness) can help to predict when and how development might be affected by experience-independent processes, whereas within language frequency data can give more insight in experience dependent processes.

1.2.2 EXPERIENCE DEPENDENT PROCESSES IN INFANT LANGUAGE ACQUISITION

Throughout this thesis, when I refer to **experience dependent** processes I mean the processes influenced by the learner's exposure to adult language, or adult birdsong in the case of birds. Other experience possibly influences the learning process as well, including practice, self perception, perception of other sounds than speech or birdsong and visual and social experience. The main focus of this thesis is however on auditory input provided by adults.

Several studies have demonstrated the relevance of experience dependent mechanisms to child language acquisition. Jusczyk et al. showed that 9 month old American infants perceptually prefer the most frequently occurring stress pattern of American English (strong- weak, (Jusczyk et al, 1993a). Additionally, unfamiliar words with native language sound patterns are preferred over those with non-native patterns by 9-month-olds (Jusczyk et al, 1993b). This is probably due to the phonotactic probability of the stimuli, since frequently occurring **phonotactic patterns** are also preferred over infrequent ones at 9 months of age (Jusczyk & Luce, 1994). Together these data suggest that around 9 months of age experience dependent processes start to show their effect since infants develop recognition and preference for frequent native language aspects during this period (fig 1.3a). Further evidence for experience dependent processes comes from phoneme discrimination experiments showing that up to the age of 6 months, infants discriminate phonemes from their native language as well as from non-native languages, whereas later in development the capacity to discriminate between non-native phonemes declines. The ability to discriminate native language patterns however increases over the course of development (Kuhl & Meltzoff, 1996; Tsao et al, 2004).

1.2.3 EXPERIENCE INDEPENDENT PROCESSES IN INFANT LANGUAGE ACQUISITION

Frequency of occurrence is often used as a measure of experience dependent processes (Jusczyk et al, 1994; Zamuner et al, 2004). Possible experience independent processes however, could also induce preferences, which in turn could lead to frequent occurrence of the preferred sound too. This means that whenever a frequent sound is preferred by infants, it is not always clear whether the frequency of occurrence in the language is indeed the (only) cause. Experience independent factors like predispositions or an interaction between different factors may cause the preference.

Indications for experience independent factors playing a role in acquisition can

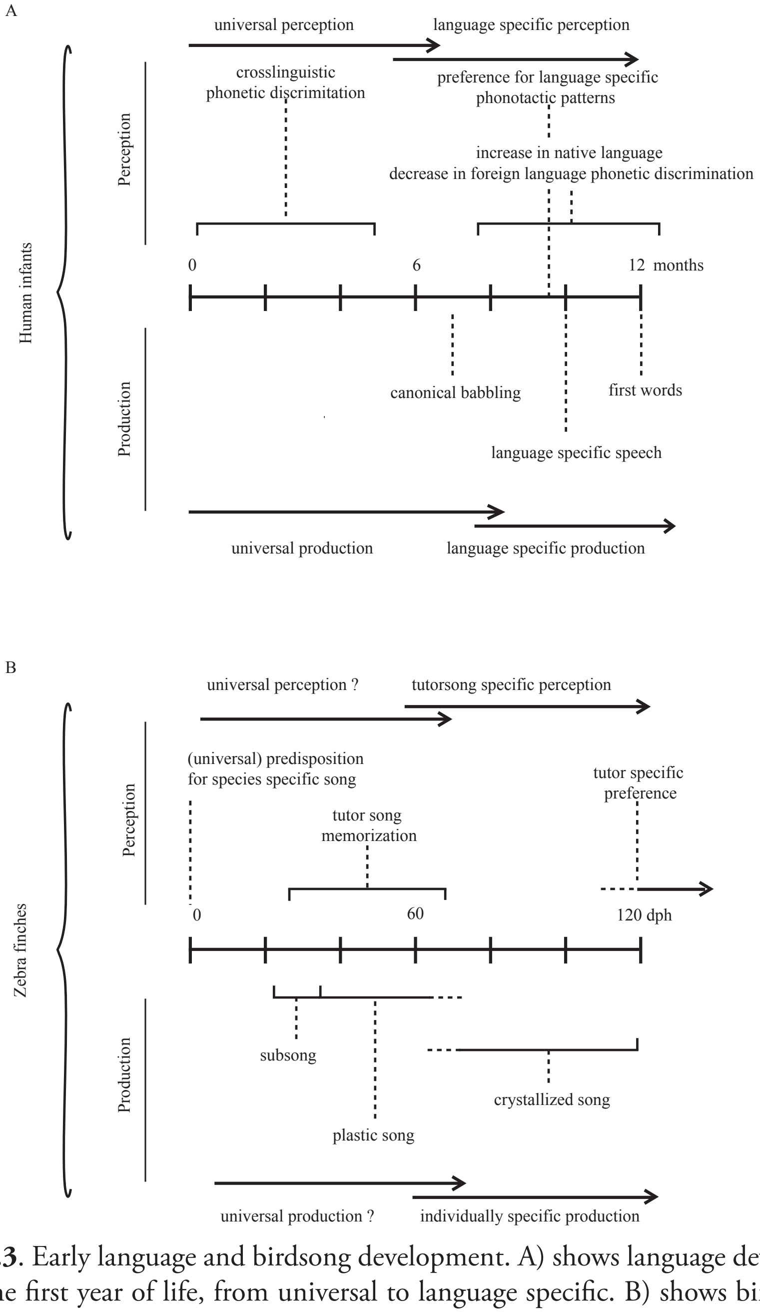


Figure 1.3. Early language and birdsong development. A) shows language development during the first year of life, from universal to language specific. B) shows birdsong development. The exact timeline varies between bird species. The timeline indicated here is representative for zebra finch development.

be found in child language production data. The order in which phonemes are acquired by Dutch children, for instance, cannot be explained by the frequency of these phonemes in the Dutch language (Levelt & Van Oostendorp, 2007). The first consonants that infants produce consistently during babbling can also not be explained by the frequency of occurrence in the speech of these infants' mothers (DePaolis et al, 2011). Furthermore, the way children initially pronounce words often does not match the actual adult pronunciation of the target words. For instance, children tend to omit the coda (the last consonant of a syllable, see example (1.1)) or substitute one phoneme in a word with another (1.2), which is clearly not an effect of exposure to language since these 'mispronunciations' are not available in the input.

(1)Production patterns in early child language production:

CVC →CV (Jacob 1;4/1;5 (Menn, 1976))

Target	*Child Production*
hat	[hæ]
nose	[do]
tape	[dæ]

Dorsal → Coronal (Ruth Hills 2;0 (Ingram, 1974))

Target	*Child Production*
kiss	[tɪ]
key	[ti:]
go	[dou]

Thus, in addition to learning by experience, experience independent processes seem to be involved in language acquisition. How can we explain these experience independent patterns? Constraints on the vocal apparatus (vocal chords, larynx, tongue, teeth and lips) could limit the sounds that can be produced. This is especially true since the infant and child anatomy is different from adult anatomy. Alternatively, phonological explanations for the patterns in 1.1 and 1.2 have been proposed. For instance, example 1.1 can be explained in terms of markedness, because CVC syllables are marked and CV syllables are unmarked, and they therefore replace the marked form. Similarly in example 1.2, dorsals are marked and coronals are unmarked and therefore coronals replace the marked dorsal segments.

These phonological explanations for experience independent patterns imply predispositions for unmarked forms. If predispositions are indeed the cause of these patterns, one might expect a preference for unmarked forms even before production starts. Thus, if infants show a perceptual preference before production and independent of experience, this may indicate the presence of a predisposition. As mentioned before, studies showing a perceptual change in phoneme discrimination and perceptual preferences for native language at 9 months of age point to the existence of experience dependent processes. However, they show that at 6 months of age no evidence is found for a preference for native language stress patterns or frequent phonotactic patterns. Moreover, discriminative abilities in infants younger than 6 months of age are less influenced by experience (however see Kuhl et al, 1992) since infants can still discriminate phonemic contrasts of non-native languages (Kuhl et al, 2006; Rivera-Gaxiola et al, 2005; Tsao et al, 2006). Thus, before 6 months of age perceptual properties seem to be more general across infants growing up in different language backgrounds.

In this thesis I use frequency within language independent of cross-linguistic markedness data to study experience dependent influences. A perceptual preference for frequent sounds within language is expected when experience plays a role. The 'universal' patterns in child and adult language (markedness), independent of within-language-frequency, are used to study possible experience independent predispositions. A perceptual preference for unmarked sounds is thus expected when predispositions play a role. Moreover, by studying the course of development of experience dependent and -independent preferences during the second half of the first year I investigate how experience and predispositions may interact.

1.3 Song acquisition in songbirds

The second aim of this thesis is to study experience dependent and independent processes during vocal acquisition in songbirds. Similar to human speech, birdsong consists of a sequence of different sound units. In zebra finches, the species studied in this thesis, the smallest sound units are called 'elements' (fig. 1.2). Elements are to a certain extent comparable to phonemes, since they are distinguishable separate units and the smallest vocal units described (Zann, 1993). However, because of the lack of meaningful words in birdsong, phonemic contrast is hard to define in birdsong. A similar approach as in the infant study is chosen, using 'typological' information to form predictions about possible experience dependent and independent processes. As with markedness in human language, commonly occurring elements in zebra finch song might indicate predispositions (experience independent) or they may be an effect of cultural transmission and thus experience dependent processes. In zebra finches however, it is possible to control the exposure to elements and thus to disentangle the experience dependent and –independent processes.

The zebra finch is an excellent model species for studying the acquisition of song and the influence of experience dependent and independent processes during this process. Zebra finches are easily maintained in the lab, they breed all year round and have a short developmental period (Doupe & Kuhl, 1999). It has been demonstrated that zebra finches have a sensitive phase for song learning, during which they have to be exposed to song in order to learn to produce normal song later in life (Eales, 1987; Pytte

& Suthers, 2000).

Male zebra finches sing but females do not. Females do produce calls for communicating about, for instance, presence of food or danger (Simpson & Vicario, 1990; Zann, 1996). Even though they do not learn to produce songs, they are able to recognize songs and to develop a preference for their **tutorsong** (Clayton, 1988; Riebel et al, 2002). The fact that females do not sing is convenient because this gives researchers the opportunity to have juvenile birds grow up with their mother only. This way they are raised without song exposure while maintaining relatively natural rearing conditions. Subsequently, these birds can be exposed to songs via a speaker, a method called **tape-tutoring** (nowadays using computer files rather than tapes). This way it is possible to manipulate and control the type of exposure, the amount of exposure and the timing of when the birds are exposed to song, creating a highly controlled setting for examining relative effects of experience dependent and independent processes.

1.3.1 Zebra finch typology

Studying variation in birdsong is not new. Several studies have shown dialects in birdsong (Kroodsma et al, 1999; Marler & Tamura, 1964; Petrinovich & Baptista, 1984). The vocal repertoire may differ between individuals of the same species, and different species differ in the types of elements and complexity. Variation in songs or repertoires of different species should not be confused with variation between different languages, since human language exists in only one species. Therefore it is relevant for drawing parallels between birdsong and language to also study song elements in different populations *within* the same species.

With respect to the zebra finch, relatively little is known about typology. A few studies have described the song elements of a specific population (Holveck et al, 2008; Leadbeater et al, 2005; Sturdy et al, 1999; Zann 1993), and one study describes the song elements in 4 different populations (Sturdy et al, 1999). The purpose for this latter study however, was to form a general and simple classification scheme for describing song of different populations in a similar way. Although the aim was not directly to give a detailed description of differences between populations, the study showed that even with this simple classification, population differences in the distribution of elements were evident. Nothing is known however, about how these within species population differences relate to developmental mechanisms. The same question can be asked for infants: how do experience dependent and independent processes relate to each other and to 'typological' patterns found across populations? In order to answer these questions properly I address developmental issues in this thesis and I present an analysis of the song elements in 13 zebra finch populations in order to come closer to a zebra finch element typology.

1.3.2 EXPERIENCE DEPENDENT PROCESSES IN SONG ACQUISITION

Songbirds learn their songs from adult input. Manipulating input or depriving birds from input severely affects song learning (see Woolley 2012 for review). Zebra finches copy songs to which they have been exposed during a sensitive phase early in life (Immelmann, 1969; Jones et al, 1996). In natural conditions they learn the song of their father or other adult males. In artificial settings it has been shown that zebra finches also learn from auditory input alone (Houx & ten Cate 1999a,b), though live tutoring is more effective (Eales, 1989).

Auditory input is necessary for song recognition and memorization, which induces changes in neural activation. Perceptual preferences for the tutor song have been shown in male and female zebra finches. Also, in response to tutor song, neuronal activation in NCM (caudomedial nidopallium) is positively correlated with the strength of song learning (measured as number of elements copied). This correlation is absent in response to novel song or the bird's own song (Bolhuis et al, 2001; Bolhuis et al, 2000; Terpstra et al, 2004), indicating a relation specifically between auditory experience induced activity and song production.

Even though learning is evident, the role of frequency of occurrence is not so clear. It has even been suggested that the amount of exposure is inversely correlated with song learning (Tchernichovski et al, 1999). Moreover it seems that zebra finches (and probably other songbirds) need only very little experience in order to learn to recognize song (Braaten et al, 2008). After only a day of exposure changes in spine turnover have been shown in the bird's neurons, at least after the first exposure of their life (Roberts et al, 2010). On the other hand, the only factor that seemed relevant for song learning in a study by Houx et al (2000) seemed to be the fact that elements are repeated within a song (and thus appeared more often).

It is thus unclear to what extent the amount of exposure matters, but the type of exposure has been shown to be relevant. Songbirds do not just copy any song but prefer to learn conspecific song. When zebra finches for instance hear both zebra finch song and Bengalese finch song, they will copy the zebra finch song.

1.3.3 EXPERIENCE INDEPENDENT PROCESSES IN SONG ACQUISITION

Previous research on songbirds has indicated how predispositions may play a role in song learning. In most studies, adult song production is used as a measure of learning or selective preference. Isolate birds – raised without song exposure - sing an aberrant of song in which the fine grained acoustic structure of natural song is absent, but some crude features remain (Price, 1979). Cross-fostering experiments for instance have shown that zebra finches copy their own species (**conspecific**) song more accurately than songs of Bengalese finch foster fathers (Clayton, 1989).

Only few studies have examined perceptual predispositions in naïve birds. An experiment in which juvenile zebra fiches could elicit exposure to either conspecific or heterospecific song by hopping on a perch showed more hopping on the perch that generated conspecific song than on the one for heterospecific song (Braaten & Reynolds, 1999). Also, white-crowned sparrow (*Zonotrichia leucophrys*) fledglings produced more begging calls in response to conspecific song than to heterospecific song (Nelson & Marler, 1993; Soha & Marler, 2001) Preference for the own subspecies over other subspecies was not confirmed, but exposure to songs of the own subspecies lead to better discrimination than experience with another subspecies song. This outcome suggests that the perceptual system is more attuned to acoustic features of the own subspecies (Nelson, 2000). Moreover, universal white-crowned sparrow's introductory whistles have been shown to function as a cue for song learning, since songs (even heterospecific ones) are better copied when they contain these universal whistles (Soha & Marler, 2000). This suggests that in addition to a preference for conspecific over heterospecific vocalizations, there are preferences for certain *within-species* element types (independent of song exposure). This has also been shown in a study on grasshopper sparrows, showing that naïve female fledglings respond more to one conspecific song type ('buzz' simple structure and uniform across individuals) than another conspecific one ('warble', complex and possibly individually specific, (Soha et al, 2009).

In summary, similar to what is known for human infants, it seems that both experience dependent and –independent processes are involved in song acquisition. The questions remain however, how and when the different processes are relevant during song learning. In addition, there is not much known about the within-species sensitivities for song features in zebra finches. The questions I address for human infants are asked for zebra finches as well: To which type of information do the young birds show sensitivity and how does this develop over age?

If experience independent processes play a role in early vocal development, preferences are expected for 'universal' within-species properties of song before juveniles are exposed to song. If experience dependent processes are involved, the subsequent exposure to song should have an cumulative effect on these preferences.

1.4 Thesis overview

Significant parallels between birdsong and language have been found. There is evidence for both experience dependent and independent processes involved in vocal development. Several questions remain however, specifically with regard to the relation between predispositions and experience, and to how this could affect uniformity and variety in language and birdsong. In the present thesis I examine the development of perceptual preferences in zebra finches and human infants, aiming to disentangle these effects. Furthermore, I present an analysis of zebra finch songs across populations in order to gain a more detailed view of zebra finch 'typology'.

CHAPTER 2 is aimed at studying the development of perceptual preferences in human infants at 9 and 12 months of age. By carefully balancing the stimuli, we were able to disentangle experience dependent and independent influences on attentional preference. A preference for frequently occurring syllables (in Dutch) is thought to reflect experience dependent mechanisms, whereas preferences for unmarked syllables could indicate experience independent processes. In **homorganic syllables** (which can be viewed as unmarked) the consonant and the vowel share a place of articulation feature. Since these syllables have been shown to appear early in child language and babbling cross-linguistically, a preference for homorganic syllables in perception could also indicate possible experience independent processes.

In CHAPTER 3 I test the preference of male zebra finches for certain types of elements while they are naive to song in order to identify preferences independent of song experience. The stimuli are either common ('universal') or uncommon song elements, similar to unmarked and marked phonemes in human language. Subsequently the birds are tutored with common and uncommon songs and tested for their preference again later in development to examine the effects of experience. Furthermore, an analysis of the produced songs of these same birds is performed to be able to link the perceptual data to the elements that the birds eventually produce.

To take the step towards a zebra finch typology CHAPTER 4 provides an extensive acoustic computational clustering analysis of zebra finch song collected form 13 populations. The distribution of elements across populations and continents is analyzed. In addition an analysis of **phonological syntax** is performed on the same songs to investigate constraints on element sequences. Phonological syntax is a term used in birdsong research to describe the sequence or order of elements or syllables in birdsong. This should not be confused with **syntax** in human language since here it refers to grammar, including grammatical classes. Meaning is usually considered to be dependent on word order here. CHAPTER 4 provides insight in uniformity and variation of elements and sequences across zebra finch populations.

In CHAPTER 5 I discuss in depth the different results in the light of the developmental and evolutionary issues presented above. Findings in human infants and songbirds are compared to determine to what extent new parallels between the two species can be drawn.

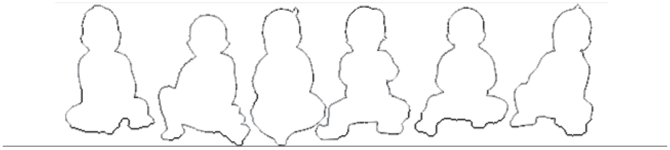
2 The role of frequency and phonology in the perceptual attention of nine- and twelve-month-old infants.

S.M. ter Haar*[1,2,3], C. ten Cate[a1,3], C.C. Levelt[1,2]

[1] Leiden Institute for Brain and Cognition, Leiden University, Postzone C2-S, P.O.Box 9600, 2300 RC, Leiden, The Netherlands

[2] Leiden University Centre for Linguistics, Leiden University, P.O. Box 9515, 2300 RA, Leiden, The Netherlands

[3] Behavioral Biology, Leiden University, P.O. Box 9505, 2300 RA, Leiden, The Netherlands



Abstract

In production, early phonological development shows a change from language universal, unmarked patterns to language specific, marked, patterns. In addition, a change from early production of homorganic consonant-vowel sequences to additional heterorganic ones has been observed. In the present study we test whether these developmental patterns can also be found in the perceptual biases of infants. Input frequency and phonological markedness are highly correlated in languages. Here we disentangle their influence by studying their effects separately. Listening preferences were tested in nine- and twelve-month-old infants for stimuli contrasting either in frequency, markedness or what we will call homorganicity. Nine-month-olds preferred homorganic syllables, while twelve-month-olds preferred heterorganic patterns. No effect for frequency or segmental markedness was found. These results indicate that similar to production, perception shows a developmental path from larger to smaller units of representation. These findings are relevant for discussions about both language acquisition and language evolution.

INTRODUCTION

In early child language productions, several phonological regularities have been observed. For example, around seven months of age infants produce “canonical babbling”, i.e. repetitive Consonant Vowel (CV) syllable patterns (Koopmans-van Beinum & Van der Stelt, 1986; MacNeilage & Davis, 1990). This initial bias towards CV patterns has been reported cross-linguistically and is continued in early word production (Davis & Macneilage, 1995), where CVC target words initially tend to be mispronounced as CV (Moskowitz, 1970; Menn, 1976; Ingram, 1978; Fikkert, 1994; Demuth, 1995; Levelt, Schiller & Levelt, 1999; Levelt & van de Vijver, 2004). With respect to early segmental productions, asymmetric substitution patterns have been noticed. For example, in onset position, target fricatives often become stops, but target stops do not become fricatives in early word productions (‘stopping’ (Ingram, 1976)), while target dorsal consonants tend to become coronal, but not vice versa (‘fronting’, (Ingram, 1974b)).

(1) Production patterns in early child language production
CVC → CV (Jacob 1;4/1;5, Menn 1976)

Target	Child Production
hat	[hæ]
nose	[do]
tape	[dæ]

Patterns like these have been accounted for in terms of markedness: unmarked aspects of language are acquired before marked aspects, and up until these marked aspects are acquired, they tend to be replaced by their unmarked counterparts (Ingram, 1976; Macken, 1980). Markedness plays an important role in accounting for phonological processes like neutralization, epenthesis and deletion; language neutralization always goes in a particular direction, segments that are the result of neutralization are also the segments that show up in epenthesis, or tend to be the targets of deletion. Processes like these can be used as diagnostics for markedness (De Lacy, 2006): segments that are the result of neutralization, that can be epenthesized, or are deletion targets, carry unmarked feature values, while their counterparts, that are neutralized or resistant to deletion, carry marked feature values. The sounds that are diagnosed as ‘unmarked’ are indeed also the sounds that are acquired early in production.

Jakobson (1941) already suggested that children acquire “universal” sound contrasts first, regardless of the language environment or culture. He also formulated laws of “irreversible solidarity” based on cross-linguistic observations of sound inventories. These laws specify that one phonemic contrast implies the existence of another, whereby the implied contrasts are those that occur more generally in languages. These implied contrasts also occur as the earliest contrasts in child language productions. In short, it appears that infants’ early language acquisition shows a typical developmental path where more universal – unmarked - aspects of the native language are acquired before the language specific – marked - aspects.

The first question we address here is whether these early patterns are specific to production or whether they result from more general language processing biases (Jusczyk, 1998): are prelingual infants sensitive to phonological markedness in perception? We investigate this issue for three phonological features: place of articulation (PoA), vowel height, and voicing.

Place of Articulation (PoA)
Cross-linguistically, [Coronal] is considered to be the unmarked feature, while both [Labial] and [Dorsal] are marked (Lahiri & Evers, 1991), (2))

(2) Dorsal → Coronal (Ruth Hills 2;0, (Ingram, 1974a))

Target	Child Production
kiss	[tɪ]
key	[ti:]
go	[dou]

Vowel height
Both [+low], i.e. /a/, and [+high], i.e. /i/, vowels are considered to be unmarked, while the mid vowels, [-high, -low], i.e. /e/, /o/, are marked. Both low and high vowels appear early in child word productions and with a relatively low error rate, while mid vowels appear later and have a relatively high error rate (Levelt, 1994). Mid vowels imply their low and high counterparts (Jakobson, 1941).

(3) Mid vowel → low/high vowel (Levelt, 1994)

Target	Adult Production	Child Production	
beer (‘bear’)	/beɹ/	[‘bi:]	(Tirza 1;8.5)
Ernie (name)	/ɛɹni/	[‘nanɑ]	(Jarmo 1;8.12)
pop (‘doll’)	/pɔp/	[‘pup’]	(Noortje2;2.21)

Voicing
When studying transcriptions of early English and Dutch child language productions it can be confusing to note that the first plosives in English appear to be [b], [d], [g], while in Dutch they are [p], [t] and [k]. This stems from the fact that English is an aspiration language, while Dutch is a pre-voicing language. In Dutch, voice onset time is approximately -80 ms for voiced consonants (/b/, /d/) and between 0-25 ms for voiceless consonants (/p/, /t/, /k/, (4), (van der Feest, 2007)). It has been proposed that the phonological contrast in pre-voicing languages is in terms of the feature [±voice], while for aspiration languages like English the contrast is in terms of the feature [±spread glottis] (Kager, Van der Feest, Fikkert, Kerkhoff & Zamuner, 2007). In both English

and Dutch child language, stops are usually first produced with a short-lag, positive VOT value, denoting phonologically [-voice] consonants in Dutch, transcribed /p/, /t/, /k/, but denoting phonologically [-spread glottis] consonants in English, transcribed as /b/, /d/, /g/. Since our participants are Dutch, the contrast we are interested in is, thus, in terms of [±voice], whereby cross-linguistically [-voice] is considered to be the unmarked feature (van der Feest, 2007).

(4) Voiced → Voiceless (CLPF 1;10-2;1, Van der Feest 2007)

Target	Adult Production	Child Production
douche ('shower')	/duʃ/	[tus]
bus ('bus')	/bʏs/	[pʏs]
bootje ('little boat')	/botjə/	[pɔtjə]

Another pattern that has been observed in early child language productions is the tendency to produce utterances with homorganic syllables, where consonants and vowels in the utterance share their place of articulation (PoA), also referred to as CV co-occurrence patterns (Davis & Macneilage, 1995; Levelt et al., 1999; MacNeilage, Davis, Kinney & Matyear, 2000) or consonant-vowel harmony (Seidl & A Buckley, 2005). Examples are in (5):

(5) Co-occurrence patterns (Eva 1;6, Levelt, 1994)

Target	*Adult production*	*Child Production*
Brood ('bread')	/brot/	[bop]
Schoen ('shoe')	/sxun/	[pum]
bed ('bed')	/bɛt/	[dɛt]

Two types of accounts have been put forward for this pattern, a motor account and a phonological account. MacNeilage & Davis present the Frame-Content model (MacNeilage & Davis, 1990; MacNeilage, 1998). In this model, homorganic production patterns result from a mandibular oscillation, the frame, which gives a consonant vowel alternation, combined with a fixed tongue-position throughout the mandibular movement, the content. For the alternative explanation, Levelt (1994; 1995) and Fikkert & Levelt (2008) build on work by, among others, Waterson (1971) and Ferguson & Farewell (1975), and propose that phonological representations are initially 'holistic' in the sense that rather than individual segments, whole words or syllables are specified for a PoA feature. A word unit represented for Labial will end up containing labial consonants and round vowels, a Coronal word will consist of coronal consonants and front vowels, and a Dorsal word will contain dorsal consonants and back vowels. Words with combinations of different PoA features (heterorganic) appear later in child language productions, when individual segments rather than larger chunks can be specified for

PoA features.
A difference between Davis & MacNeilage's findings and those by Levelt are the CV combinations with labials; Davis & MacNeilage found that labial consonants occurred together with central vowels whereas Levelt found that labial consonants occurred preferably with round vowels. This could be due to the fact that Levelt studied Dutch, which has more pronounced roundedness in the back vowels /o/, /ɔ/ and /u/ than English. In the present study labials with round vowels are used as homorganic Labial stimuli.

For both the 'holistic' and the Frame-Content approach, a perceptual account could be considered. In line with the idea of a 'holistic' representation, speech may be perceptually processed at the level of the syllable rather than at the segmental level. Therefore homorganic CV patterns may be preferred perceptually over heterorganic ones. For the Frame-Content theory a perceptual explanation has been proposed from an embodiment perspective, where 'intrinsic' (self-produced) information may play a role in perceptual organization (Davis & MacNeilage, 2000). Thus either point of view suggests a possible perceptual differentiation between homorganic (co-occurring) syllables and heterorganic ones. However, as far as we know it has never been tested if homorganic patterns are perceptually preferred over heterorganic ones. The second question we thus address in the present study is whether infants prefer homorganic syllables over heterorganic ones.

Up until now, early preferences have been found for legal versus illegal patterns, or occasionally frequent versus infrequent patterns in infants' native language. Preferences for legal (native) over illegal (non-native) sound patterns have been shown (Friederici & Wessels, 1993; Jusczyk et al, 1993; Sebastian-Galles & Bosch, 2002) in nine- and 10-month-old infants. Within native language sound patterns, nine-month-olds prefer to listen to frequently occurring phonotactic patterns over infrequent ones (Jusczyk & Luce, 1994), in contrast to six-month-olds, who don't show this preference. This suggests that infants become more and more aware of native language patterns in the second half of their first year of life.

One explanation for the early appearance of and preference for phonologically unmarked patterns in production might be that most of these unmarked sounds are also frequent within languages (Zamuner, Gerken & Hammond, 2005). Thus, just by hearing these sounds more often, infants could acquire them first. However, the patterns found in children's early speech productions cannot always be explained by frequency in the input. For instance, consonant harmony, a phenomenon encountered in child language productions, cannot be explained by input frequency since it hardly ever occurs in adult speech. Similarly at least one of the homorganic patterns found in early word productions in, the labial C with round V combination is highly infrequent in Dutch (Levelt et al., 1999; Fikkert & Levelt, 2008). Segmental frequency also does not explain the order of segmental development either; for instance /b/ and /d/ are more frequent in Dutch than /p/ and /t/, but the latter are acquired earlier (Levelt & Van Oostendorp, 2007). Furthermore, there is one study showing an initial perceptual preference for nasal place assimilation (a form of markedness) in infants (Jusczyk et al, 2002). This indicates that other factors than frequency influence the acquisition process as well. In short, even though markedness and language specific frequency are correlated, not every frequent sound is unmarked and not every infrequent sound is marked. In the experiments below we strive to disentangle these factors, by carefully balancing markedness and frequency in the different conditions.

We test nine- and twelve-month-old infants' perceptual preferences, as indicated by their looking time while being presented with lists of CV syllables contrasting in frequency – balanced for markedness - , phonological markedness – balanced for frequency - , or PoA structure (homorganic or heterorganic) – balanced for frequency.

A preference for frequent syllables can be expected at nine months of age because a preference for frequently occurring sounds has been shown previously (Jusczyk & Luce, 1994). This preference is expected to increase with age (by 12 months of age), as attention focuses more and more on the native language. If there is a general early speech processing bias for unmarked, universal aspects of sounds, then we expect to find a preference for unmarked syllables in the nine-month-olds. We also expect the preference to be stronger for the nine-month-olds than for the twelve-month-olds, since a development from unmarked, universal to marked, language-specific has been shown for production, and the perceptual sensitivity to language-specific patterns increases between six- and 12-months of age. If infants start out with representational units larger than the segment, then it can be expected that nine-month-olds prefer homorganic syllables to heterorganic syllables. This finding would, in turn, predict that sensitivity to differences in *segmental* markedness cannot be detected at this age.

EXPERIMENT 1

The first experiment was aimed to identify the listening preferences of nine-month-old infants. Preferences for homorganic versus heterorganic, phonologically unmarked versus marked (at the segmental level) and frequent versus infrequent conditions were tested.

Method

Participants

Dutch nine-month-old infants from a monolingual background were tested (n = 40, 21 males, 19 females; mean age 9.02 months; age range 8.45 – 9.50 months). Caregivers reported that the infants developed normally and had no neurological or auditory problems. 19 additional infants were tested but were excluded from further analyses because they did not complete the test (n=6), because they were more than 3 weeks preterm (n=2), because of dyslexia in the family (n=1) or because of experimental errors (n=4). Infants receiving bilingual linguistic input in their home environment were excluded because this could affect the relative frequencies of sounds to which they are exposed (n=6). Individual trials were rejected from the results when not reliable (see the statistics section for details). All caregivers gave written consent for the infants to participate in this study.

Stimuli

Two sets of natural stimuli were used; one spoken by a female voice and one by a male voice (21 and 19 infants tested per voice respectively). The reason we chose two sets of stimuli was to verify that the possible effects were independent of basic acoustic features like voice characteristics. Stimuli were pronounced in a monotonous way and were recorded in a sound attenuated room using Adobe Audition (version 1.5, build 4124.1) and a Sennheiser mkh 416t microphone. Ten sets of Dutch CV syllables were constructed, pairs of which contrasted in frequency (high/low) or markedness (marked/unmarked segments), or were either homorganic or heterorganic. All stimuli are presented in table 2.1. All consonants were stops, since these appear early in child language productions (Boysson-Bardies & Vihman, 1991). Each set of 10 syllables contained 5 different syllables, which were recorded and pseudorandomly ordered in two blocks of 5, using Praat (version 5.1.25, (Boersma & Weenink, 2009),). The syllables were separated by 500ms of silence. The sequences of 10 syllables were presented auditorily while a (motionless) checkerboard pattern was shown on a screen. The sets of syllables were presented in two different orders. There were no large differences in syllable duration or fundamental frequency in any of the stimuli of interest (table 2.1) and stimuli were rms-equalized. A trained phonetician and a trained phonologist listened to the stimuli and judged them to be representative Dutch syllables.

Frequency

Two sets contrasted in frequency. The syllable frequency was based on an infant directed speech corpus by Van de Weijer (Weijer, 1999). Frequencies were calculated for the stressed CV syllables in infant directed speech, because it has been shown that 8- and nine-month-old infants pay attention mostly to stressed syllables (Jusczyk, Cutler & Redanz, 1993; Johnson & Jusczyk, 2001). The frequency values of the frequent and infrequent set are .296% and .049% respectively (relative to the total number of syllables in the database, table 2.2B).

Markedness

The sets of stimuli contrasting in phonological markedness were constructed on the basis of phonological features as described in the introduction.

Voicing

For the [±voice] condition, the consonants in the two sets of syllables contrasted in being either marked [+voice], i.e. /b, /d/, or unmarked [-voice], i.e. /p/, /t/.

Place of Articulation (PoA)
Syllable sets in the PoA condition contrasted marked labials, like /p/, with unmarked coronals, like /t/. We did not use dorsal consonants in the sets, because they have a relatively low frequency in Dutch, and we wanted to avoid a potential confound with frequency.

Vowel height
Two sets of syllables contrasted in vowel height. Both [+low], i.e. /a/, and [+high], i.e. /i/, vowels are considered to be unmarked, while the mid vowels, [-high, -low], i.e. /e/, /o/, are marked.

Homorganicity
Two sets contrasted at the level of the syllable. The consonant and vowel were homorganic in one set or heterorganic in the other. In case they were homorganic, the entire syllable could be labeled [Labial], like /po/ with a labial consonant and a labial (i.e. round) vowel, or [Coronal], like /ti/ with a coronal consonant and a coronal (i.e. front) vowel. In heterorganic syllables the consonant and the vowel carried different PoA features, like /pi/ (labial consonant, coronal vowel) or /to/ (coronal consonant, labial vowel). Homorganic structures have traditionally not been categorized in terms of markedness (though see Seidl & Buckley 2005). As mentioned in the introduction, however, independent of the language, homorganic syllables occur in the earliest productions (MacNeilage & Davis, 2000). Moreover, there seems to be a tendency for homorganic syllables to be more common cross-linguistically than heterorganic ones (MacNeilage et al, 2000), although this view has been criticized (Albano, 2011)). Thus it is an open question whether homorganicity is a form of markedness or not, but most importantly predictions are in the same direction: unmarked and homorganic are expected to be preferred over marked and heterorganic syllables early in development.

Balancing
Care was taken to balance all the sets contrasting in markedness or homorganicity for frequency, and the sets contrasting in frequency for markedness and homorganicity. For frequency values we used syllable frequency as described above. Since segments can be marked or unmarked with respect to several features, stimuli contrasting in markedness for one feature, were balanced for all other features. Due to the complexity of the study, balancing everything perfectly turned out to be impossible. However, stimuli were balanced in such a way that the only feature contrasting in all syllables in the stimuli set, was the contrast of interest and was therefore the most likely to explain a potential difference in looking time. Table 2.2 shows all stimuli with their markedness (A) and frequency values (B) used for balancing. For instance, sets contrasting in consonant voicing were balanced for frequency (f), and other marked features (m) (PoA of the consonant, homorganicity, and vowel height), whereas all 5 stimuli contrasted in the feature

[+-voice]. If we take the first voiceless syllable /pi/ this has a markedness value of 2 because it is marked for two out of four features: labial (marked) heterorganic (marked) voiceless (unmarked) and has a high vowel (unmarked). The first voiced syllable /bi/ has a markedness value of 3 because it is labial (marked) heterorganic (marked) voiced (marked) and has a high vowel (unmarked). A total markedness difference between the voiced (14) and voiceless (9) set of stimuli is thus 5 and is only caused by the difference in voicing. The average markedness values are 1.8 (0-3) for voiceless and 2.8 (1-4) for voiced stimuli. The average frequency values for the voiceless and voiced sets are 173,6 (15-692) versus 201,4 (11-423) respectively. Thus the main difference between the voiced and voiceless stimuli is indeed the feature [+-voice].

Table 2.1. *Acoustic Properties of the stimuli*

		female voice						male voice					
		F0	range		dur	range		F0	range		dur	range	
poa	coronal	200	197	204	388	261	142	133	129	142	377	223	223
	labial	205	201	209	399	292	141	137	132	141	365	192	466
voice	voiceless	203	198	206	337	292	141	136	132	141	311	192	360
	voiced	203	209	209	444	387	137	134	129	137	412	324	471
hom	homorganic	201	197	206	362	261	143	137	130	143	352	223	466
	heterorganic	203	198	209	389	310	138	136	129	138	358	257	428
vh	high/low	203	197	210	355	310	138	134	130	138	353	257	471
	mid	200	197	203	404	350	143	137	130	143	371	326	420
freq	frequent	202	197	209	426	353	143	136	130	143	389	326	420
	infrequent	204	200	206	368	200	454	134	129	142	355	223	466

Acoustic properties of the two sets of stimuli; male voice and female voice. Pao: place of articulation, vh: vowel height, feq: frequency, dur: mean syllable duration in ms, F0: mean fundamental frequency in Hz.

Table 2.2 – frequency and markedness information for all stimuli
A

PoA	**unmarked** coronal	m	hom	PoA (C)	vow	voic	**marked** labial	m	hom	PoA (C)	vow	voic
	do	3	1	0	1	1	**bo**	3	0	1	1	1
	da	1	0	0	0	1	**ba**	2	0	1	0	1
	to	2	1	0	1	0	**po**	2	0	1	1	0
	ti	0	0	0	0	0	**pi**	2	1	1	0	0
	de	2	0	0	1	1	**be**	4	1	1	1	1
total		**8**						**13**				
average		**1.6**						**2.6**				
voice	voiceless	m	hom	PoA (C)	vow	voic	voiced	m	hom	PoA (C)	vow	voic
	pi	2	1	1	0	0	**bi**	3	1	1	0	1
	po	2	0	1	1	0	**bo**	3	0	1	1	1
	to	2	1	0	1	0	**do**	3	1	0	1	1
	pe	3	1	1	1	0	**be**	4	1	1	1	1
	ta	0	0	0	0	0	**da**	1	0	0	0	1
total		**9**						**14**				
average		**1.8**						**2.8**				
homorganicity	**unmarked**	m	hom	PoA (C)	vow	voic	**marked**	m	hom	PoA (C)	vow	voic
	po	2	0	1	1	0	**to**	2	1	0	1	0
	ti	0	0	0	0	0	**ki**	2	1	1	0	0
	ko	2	0	1	1	0	**ke**	3	1	1	1	0
	de	2	0	0	1	1	**be**	4	1	1	1	1
	ba	2	0	1	0	1	**do**	3	1	0	1	1
total		**8**						**14**				
average		**1.6**						**2.8**				
vowel height	**unmarked**	m	hom	PoA (C)	vow	voic	**marked: vowel: o,e (mid)**	m	hom	PoA (C)	vow	voic
	da	1	0	0	0	1	**de**	2	0	0	1	1
	ki	2	1	1	0	0	**ke**	3	1	1	1	0
	ka	1	0	1	0	0	**ko**	2	0	1	1	0
	ta	0	0	0	0	0	**to**	2	1	0	1	0
	bi	3	1	1	0	1	**bo**	3	0	1	1	1
total		**7**						**12**				
average		**1.4**						**2.4**				
frequency	**frequent**	m	hom	PoA (C)	vow	voic	**infrequent**	m	hom	PoA (C)	vow	voic
	bo	3	0	1	1	1	**bi**	3	1	1	0	1
	be	4	1	1	1	1	**ba**	2	0	1	0	1
	de	2	0	0	1	1	**do**	3	1	0	1	1
	ta	0	0	0	0	0	**ti**	0	0	0	0	0
	ko	2	0	1	1	0	**pe**	3	1	1	1	0
total		**11**						**11**				
average		**2.2**						**2.2**				

B

	unmarked					**marked**				
PoA	coronal					labial				
		f	f%	bf	bf %		f	f%	bf	bf %
	do	19	0.01	269	0.12	**bo**	423	0.19	584	0.27
	da	190	0.09	1619	0.74	**ba**	17	0.01	34	0.02
	to	64	0.03	65	0.03	**po**	15	0.01	50	0.02
	ti	34	0.02	154	0.07	**pi**	52	0.02	187	0.09
	de	357	0.16	413	0.19	**be**	364	0.17	1200	0.55
total		**664**	**0.30**	**2520**	**1.15**		**871**	**0.40**	**2055**	**0.94**
average		**132.8**	**0.06**	**504**	**0.23**		**174.2**	**0.08**	**411**	**0.19**
voice	voiceless					voiced				
		f	f%	bf	bf %		f	f%	bf	bf %
	pi	52	0.02	187	0.09	**bi**	11	0.01	18	0.01
	po	15	0.01	50	0.02	**bo**	423	0.19	584	0.27
	to	64	0.03	65	0.03	**do**	19	0.01	269	0.12
	pe	45	0.02	63	0.03	**be**	364	0.17	1200	0.55
	ta	692	0.32	737	0.34	**da**	190	0.09	1619	0.74
total		**868**	**0.40**	**1102**	**0.50**		**1007**	**0.46**	**3690**	**1.68**
average		**173.6**	**0.08**	**220.4**	**0.10**		**201.4**	**0.09**	**738**	**0.34**
homorganicity	**unmarked**	f	f%	bf	bf %	**marked**	f	f%	bf	bf %
	po	15	0.01	50	0.02	**to**	64	0.03	65	0.03
	ti	34	0.02	154	0.07	**ki**	70	0.03	73	0.03
	ko	273	0.12	313	0.14	**ke**	102	0.05	468	0.21
	de	357	0.16	413	0.19	**be**	364	0.17	1200	0.55
	ba	17	0.01	34	0.02	**do**	19	0.01	269	0.12
total		**696**	**0.32**	**964**	**0.44**		**619**	**0.28**	**2075**	**0.95**
average		**139.2**	**0.06**	**192.8**	**0.09**		**123.8**	**0.06**	**415**	**0.19**
vowel height	**unmarked: i,a (high/ low)**					**marked: vowel: o,e (mid)**				
		f	f%	bf	bf %		f	f%	bf	bf %
	da	190	0.09	1619	0.74	**de**	357	0.16	413	0.19
	ki	70	0.03	73	0.03	**ke**	102	0.05	468	0.21
	ka	335	0.15	457	0.21	**ko**	273	0.12	313	0.14
	ta	692	0.32	737	0.34	**to**	64	0.03	65	0.03
	bi	11	0.01	18	0.01	**bo**	423	0.19	584	0.27
total		**1298**	**0.59**	**2904**	**1.33**		**1219**	**0.56**	**1843**	**0.84**
average		**259.6**	**0.12**	**580.8**	**0.27**		**243.8**	**0.11**	**368.6**	**0.17**
frequency	**frequent**					**infrequent**				
		f	f%	bf	bf %		f	f%	bf	bf %
	bo	423	0.19	584	0.27	**bi**	11	0.01	18	0.01
	be	364	0.17	1200	0.55	**ba**	17	0.01	34	0.02
	de	357	0.16	413	0.19	**do**	19	0.01	269	0.12
	ta	692	0.32	737	0.34	**ti**	34	0.02	154	0.07
	ko	273	0.12	313	0.14	**pe**	45	0.02	63	0.03
total		**2109**	**0.96**	**3247**	**1.48**		**126**	**0.06**	**538**	**0.25**
average		**421.8**	**0.19**	**649.4**	**0.30**		**25.2**	**0.01**	**107.6**	**0.05**

Table 2.2 A: markedness of the stimuli, B: frequency of the stimuli. m: markedness: Numbers represent in how many features a syllable is marked (for instance /be/ is voiced + labial + mid vowel + heterorganic: 4). f: frequency, syllable frequencies based on stressed syllables from the Van de Weijer database (absolute numbers and % of total nr of syllables in the corpus). bf: biphone frequencies from the Van de Weijer database (absolute numbers and % of total nr of syllables in the corpus), represented for comparison with Jusczyk et al. (Jusczyk et al., 1994).

Apparatus

The experiment was performed in a sound attenuated booth. A chair was placed approximately 90 cm from a 104 cm Philips flat-screen. The stimuli were played in stereo from speakers on both sides of the screen. The screen was connected to a computer outside the booth. Habit X software (Cohen et al., 2000) was used to present the stimuli. Under the screen behind a panel, a camera and a microphone were placed to monitor the infant's behavior and eye movements from outside the test booth. The video recordings were used for off-line analysis of looking time.

Procedure

An adjusted version of the visual-fixation-based auditory preference paradigm was used (Cooper & Aslin, 1990) to test the infants' listening preferences. During the experiment the infant was seated on the caregiver's lap, in front of the screen presenting the stimuli. Caregivers listened to a mix of classical music and backward speech through headphones, to mask the stimuli. A red blinking light was presented on the screen to catch the infant's attention before each trial started. The 10-second syllable stimuli were presented auditorily, while the infant watched a checkerboard pattern on the screen. Between each auditory trial, the checkerboard was presented again for 10 seconds, but a melody was played in order to avoid habituation. The experiment started with two pre-test trials in which all the syllables were presented once. This was to avoid a primacy effect and to let the infant get used to the setup. After the pre-test and at the end of the whole experiment, a movie of a fish was presented to monitor general attention. Trials were presented in a different order for each infant.

Scoring

Total looking time was scored for each trial. A difference in looking time between auditory trials is thought to reflect a difference in attentional preference (Colombo & Bundy, 1981; Nelson, Jusczyk, Mandel, Myers, Turk & Gerken, 1995). Video files recorded during the experiment were analyzed frame by frame using ELAN software (version 3.7.2). All analyses were performed off-line by three trained scorers who were blind to the stimuli. During each 10-second trial, the looking behavior of the infant was scored. Looks were scored when the infant looked at the center of the screen. Total looking time (TL) within the 10-second trial and duration of the longest look (LL) were used as variables. For a reliability estimate, a subset of all data was scored by one of the other experimenters trained in rescoring. The average Pearson's correlation was 0.9 and the average reliability score (intraclass correlation coefficient) was 0.8 (p=0.001 and p=0.0001 respectively).

Statistics

A Repeated measures ANOVA was performed for each contrast separately and for LL and TL separately. Looking times for each contrast were within subject variables, and stimulus voice and stimulus order were between subject variables. Since the data was not distributed normally (see appendix A for normality tests), transformed data were used for statistical analysis because ANOVAs are based on the assumption that data are normally distributed. LL data were log-transformed because of positive skew and TL data were square-root transformed. After transformation the data were no longer significantly different from normal distribution (appendix A). Results on raw data are reported in appendix B.

The number of infants may be different between conditions because individual trials were rejected when they were not reliable. Individual trials were excluded when the infant was (temporarily) crying or fussy, when the caregiver interrupted or distracted the infant, when the infant did not look within the first 0,5 second of the trial or when the infant's eyes were not visible. Data from an infant were not included when more than half of the trials had to be excluded or when general attention during the final fish movie had dropped to less than 50% compared to attention to the first fish movie.

Results

Repeated measures ANOVAs revealed an effect for homorganicity in the TL data (F(1,32) =11.14, p= .002, $\eta^2 = .258$, fig. 2.1) and a marginally significant effect for stimulus voice (F(1,32) = 4.20, p= .049, $\eta^2 = .12$)) but there were no interactions between homorganicity and voice group nor were there effects or interactions for stimulus order (all p's> .05, appendix B1). The LL ANOVAs also showed an effect for homorganicity (F(1,32) = 24.61, p< .001, $\eta^2 = .44$), but no effects for or interactions with stimulus voice or stimulus order (all p's> .05). Neither in TL data nor in LL data any effects were found for frequency or segmental markedness (vowel height, voicing or PoA (all p's> .05, appendix B1)).

Discussion

The nine-month-old infants tested in this experiment looked significantly longer while listening to CV syllables with homorganic PoA than to those with heterorganic PoA, regardless of the stimulus voice or order. This corroborates the findings in early infant and child language productions, and may point to a general speech processing preference.

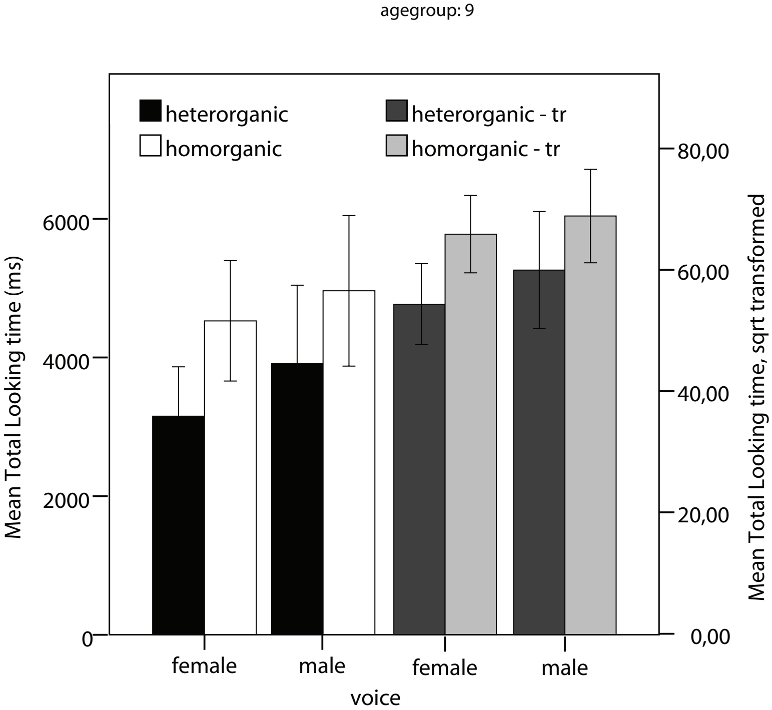


Fig. 2.1 Results for homorganicity in nine-month-olds for stimuli set spoken by different voices. Mean of raw and square-root transformed TL data for homorganicity for nine-month-olds listening to female voice stimuli (left) and male voice stimuli (right). White bars indicate raw TL for homorganic stimuli and black bars indicate TL for heterorganic stimuli. Light grey bars indicate transformed (tr) TL data for homorganic stimuli and dark grey bars transformed TL data for heterorganic stimuli.

Alternatively, the preference in perception could result from hearing one's own speech production i.e. babbling. This will be further discussed in the general discussion.

The results are also in line with the idea of an initial "holistic" representation (Fikkert & Levelt, 2008) since homorganic syllabic units were preferred, and no significant results were found for sensitivity to segmental markedness (consonantal PoA, vowel height and voicing). It could be that sensitivity to markedness at the segmental level emerges at a later stage. This means that infants start with a supra-segmental feature representation at the syllable (or word) level, which later develops into a segmental representation. In experiment 2, below, we test twelve-month-old infants to see if they are indeed more responsive to segmental markedness differences at this age.

Since no interaction was found between the stimulus voice and the homorganicity effect, the difference in the voice of the stimuli cannot explain this effect, validating that it is based on something else than just basic acoustic differences. For the TL data an effect of voice was found (no interaction) indicating that voice may have an effect on the infants' perceptual attention in general. Even though this difference is not relevant for the current results, it is relevant to consider in future research.

The lack of an effect for frequency was not expected, given the findings by Jusczyk & Luce (1994) who showed that infants listen longer to lists of syllables with high probability phonotactic patterns than to those with low probability (Jusczyk & Luce, 1994). One explanation for this difference in results is that the frequency sets in the present study were carefully balanced for markedness. Thus the preference found by Jusczyk et al. might be partly caused by the fact that frequent sounds are often also unmarked. Alternatively, since Jusczyk et al. did not explicitly control for markedness, their frequent to infrequent ratio might be higher than in the present experiment The frequent to infrequent ratio in the present study might be too subtle for nine-month-old infants to be noticed. The frequent to infrequent ratio for syllable frequency here is 16.7 but this ratio is not reported by Jusczyk et al. The ratio for biphone frequency in the study by Jusczyk et al. is not much different from the ratio in the present study. In Jusczyk et al.'s experiment 3, which is the most balanced and thus most comparable to the present study, a ratio of 5.8 is reported for biphone frequency and a ratio of 2.8 for positional phoneme probability for adult directed language. These ratio's were and 4.5 and 1.9 for infant directed speech. In the present study the biphone frequency ratio was 6.0 and the positional phoneme frequency ratio was 1.1, based on an infant directed speech corpus. Further details regarding these differences are presented in the general discussion.. Possibly more exposure to the native language is necessary to become sensitive to the relative frequencies of the syllables used in this experiment. This is another reason to test a group of twelve-month-olds, in experiment 2.

EXPERIMENT 2

The same experimental procedures were used as in experiment 1, but twelve-month-old infants were tested to see whether a change in sensitivity to frequency and to segmental markedness could be found.

Methods

Participants

Dutch twelve-month-old infants from a monolingual background were tested (n = 24, 13 males, 11 females; mean age 12.08 months; age range 11.54 – 12.69 months). Caregivers reported that the infants developed normally and had no neurological or auditory problems. Twenty-three additional infants had to be excluded from the analyses due to crying or fussiness (n=6), bilingual input (n=4), because they were 3 weeks or more preterm (n=7), because being at risk for dyslexia (n=2) or because of experimental errors (n=4). All caregivers gave written informed consent for the infants to participate in this study.

Stimuli

The same lists and contrasts were used as in experiment 1. Since the voice used for stimuli did not show any interactions we only used the female voice stimuli.

Apparatus, Procedure, Scoring

Apparatus, procedure and scoring were the same as in experiment 1.

Statistics

Data transformations were the same as experiment 1 (appendix A). To test if there was an effect of age, data from experiment 1 and 2 were combined and repeated measures ANOVAs were performed for each contrast separately, and for LL and TL separately. LL and TL were within subject variables, age was the between subject variable. The number of infants may be different between conditions because individual trials are rejected when they are not reliable. The criteria for rejecting trials were similar to those in experiment 1.

Results

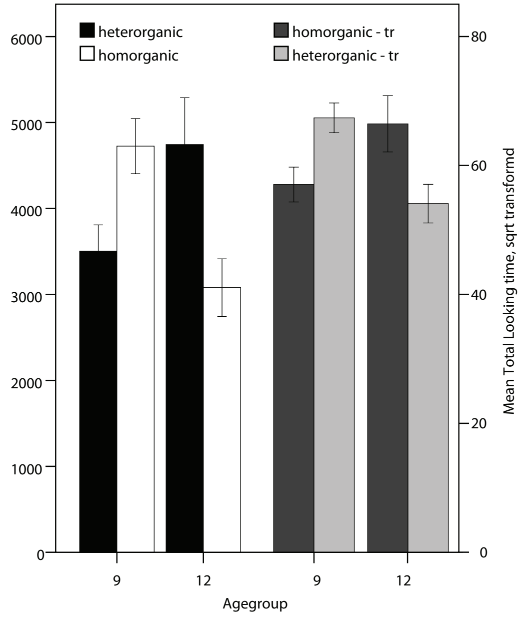
Repeated measures ANOVAs on LL and TL for grouped data from experiment 1 and 2, show significant interactions between homorganicity and age (age*homorganicity TL: $F(1,51)=24.27$, $p<0.001$, ŋ^{2}=.322, LL: $F=25.35$, $p<0.001$, ŋ^{2}= .332). In contrast to the nine-month-olds, the twelve-month-olds showed longer looking times for heterorganic than for homorganic syllables (fig. 2.2). No significant effects or interactions were found for frequency, vowel height, voicing or PoA (all p's > .05, appendix B2).

Discussion

This experiment shows that twelve-month-olds looked longer at stimuli with different PoA features for the consonant and the vowel within a syllable, in contrast to the nine-month-olds who preferred the homorganic stimuli. This corroborates findings in the production data of slightly older children (Fikkert & Levelt, 2008).

Against expectation, no significant effect or interaction for frequency was found for the twelve-month-olds. This result might again be due to the effort of balancing of stimuli for markedness, which was not explicitly performed in the study by Jusczyk & Luce (1994). This will be further discussed in the general discussion.

No significant differences in looking time were found for the segmental contrasts in markedness (vowel height, voicing and consonantal PoA). One interpretation of this lack of effect is that at 12 months of age, infants in fact still have no genuine



Error bars: +/- 1 SE

Fig. 2.2 Results for homorganicity in nine- and twelve-month-olds. Mean of raw and sqrt transformed Total Looking time (TL) data for homorganicity for each age group. White bars indicate raw TL data for homorganic stimuli and black bars indicate raw TL for heterorganic stimuli. Light grey bars indicate transformed (tr) TL data for homorganic stimuli and dark grey bars transformed TL data for heterorganic stimuli.

segmental representation. Infants' first word productions, between approximately 12 and 17 months of age are often also still 'holistic' (Fikkert & Levelt, 2008). However, the shift in preference from homorganic to heterorganic syllables that was found for the twelve-month-olds suggests that they have become aware of representational units below the larger, syllabic or word, unit. This apparent discrepancy in the results will be further discussed in the general discussion.

General Discussion

The most important result in this study is the finding that, independent of frequency, nine-month-olds show a preference for homorganic syllables, which changes to a preference for heterorganic syllables in twelve-month-olds. This result corroborates findings in young children's early word productions, where a clear initial preference for (target) words containing homorganic consonants and vowels is found, while words containing heterorganic consonants and vowels appear only later (Fikkert & Levelt, 2008).

The timing is different, however; the twelve-month-olds show a preference for, or at least sensitivity to, heterorganic syllables in perception, while the early word productions, starting around 12 months, are homorganic. Fikkert et al. (2008) found that heterorganic word productions occurred only by the age of approximately 17 months, suggesting that somewhere between 12 and 17 months, infants' representations become segmental. Thus in this case, at 12 months of age infants still make use of larger, supra-segmental units of representation for production, but they start to shift their focus to the segmental level perceptually. This enables a subsequent shift to a segmental representation in production.

If the shift in preference of the twelve-month-olds indicates a shift from a 'holistic' representation to a segmental representation in perception, we still need to understand why we don't find any sensitivity to segmental markedness in this group. The answer might actually be found in early production data. The first step in the segmentalization process in production is that vowels become separate units from consonants. It takes a while, however, before consonants within a word can be individually and independently represented (Levelt, 1994; Costa, 2008). In Levelt (1994) 4 developmental stages are recognized in the development of representational units, (1) the entire word is the unit, (2) the vowel can be specified separately from the consonants – consonants cannot be specified separately, (3) the consonant at the word onset can be specified individually, (4) all segments are individually specifiable. It could thus be that the twelve-month-olds in this study have become sensitive to consonants and vowels as separately specifiable units, i.e. stage (2), but that it is too early to measure sensitivity to segmental markedness in individual segments. This implies that the infants should have representations at the vowel level, but no evidence for markedness (vowel height) at the vowel was found in the present study. The question is however, whether vowel perception in infants develops before consonants in terms of salience. It has been shown that 16- and 20-month-old infants can discriminate words when the contrast is based on a consonantal feature but not when based on a vocalic feature (Havy et al., 2009). Thus, although infants are able to distinguish between the vowels auditorily (Martinez, 2008), the difference might not be salient enough to evoke a difference in preference.

No effect for frequency was found in the present study. This result is in contrast with findings by Jusczyk et al. (1994), who did find an effect for frequency at nine months of age. Several aspects may account for these different findings. For one thing, different methods were used to calculate frequency. Here, we used syllable frequency, for which the difference ratio was 16.7. Jusczyk et al. used a combination of biphone probability and positional frequency, while syllable frequency was not reported, which makes the comparison in this respect more difficult. When we calculate the biphone probability in the Van de Weijer database for the data that were used in the frequent/infrequent syllable sets in the present experiment, the frequent to infrequent ratio is 6.0. This difference ratio is higher than the 4.7 ratio Levelt & van de Vijver (2004) calculated to be noticeable by young children. Moreover, the biphone probability ratio for the lists used by Jusczyk et al., in their experiment 3 (which was the most balanced) was 4.5 based on infant directed speech corpora and 5.8 for adult directed speech. Thus, based on biphone frequency measures, the ration in the present study is actually higher. Positional phoneme frequency difference ratios were low in both studies, 1.9 and 2.8 in Jusczyk et al. for infant directed and adult directed respectively and 1.1 in the present experiment. Taken together, biphone frequency is the strongest contrasting in both studies and is actually higher in the present study.

Alternatively, the more balanced design in the present study might be part of the explanation for the difference between studies. Even though experiment 3 in Jusczyk et al's study is balanced for vowel quality between frequent and infrequent lists, it was not balanced for consonants. The infrequent lists in their study contained much more fricatives and affricates than the frequent lists, whereas plosives were more abundant in frequent lists than in infrequent ones. Fricatives have been considered more marked than plosives. Indeed, plosives have been shown to be common in early speech and babbling (Gildersleeve-Neumann et al., 2000) and to appear in first words before fricatives (Alvater-Mackensen 2010). Thus, in addition to a frequency difference, the frequent and infrequent stimuli in Jusczyk et al's study differ in markedness. This suggests a possible effect of markedness on the preference in their study. We found no difference in preference for markedness in the present study, but we did not include manner of articulation (plosive/fricative) in our stimuli since all stimuli were plosives. Future research taking a similar approach as the present study but including manner of articulation might give insight into this difference.

Alternatively the infrequent items in Jusczyk et al's study are so infrequent that they are perceived as similar to non-native. A few other studies have reported perceptual preferences for legal over illegal or native over non-native phonotactics (Friederici & Wessels, 1993; Jusczyk et al, 1993; Sebastian-Galles & Bosch, 2002). This implies a familiarity preference rather that sensitivity to a frequency difference. One would expect a similar mechanism to cause a preference for frequent items as for native over non-native items. This might, however, implicate a more discrete rather than linear relation between preference and linguistic input.

One could argue that in Dutch, homorganic syllables as a class could be more common than heterorganic syllables and therefore, indirectly, frequency would have an effect on the preference for homorganicity. When frequencies are calculated for homorganic and heterorganic syllables classes in Dutch (according to the classification used for the stimuli) however, we see only a small difference in the opposite direction: heterorganic syllables are as a class *more* frequent than homorganic ones (39.1% and 32.1% respectively, based on token frequencies of all syllables in a spoken Dutch corpus: Corpus Gesproken Nederlands). In addition, in order to be able to calculate the frequencies for each class, infants would have to be able to categorize syllables into homorganic and heterorganic ones, requiring some pre-existing sensitivity to this distinction.

The preference for the specific heterorganic syllables in 12 month olds in the present experiment is independent of frequency, since the syllable sets were balanced for frequency. The shift in preference between nine and 12 months however, may have been caused by experience. If at the age of nine months, infants attend to homorganicity and have 'learnt' these syllables before 12 months of age, possibly they then start to move their focus of attention towards heterorganic syllables. This explanation and the explanation regarding a shift from a supra-segmental to a segmental representation are not necessarily mutually exclusive.

The finding of a bias for homorganic syllables at nine months of age sheds new light on the discussion about the basis of the early appearance of these syllables in child language. The frame/content model of MacNeilage and Davis (2000) is motor-based but the present data indicate that perceptual factors may play a role as well. It is currently unclear, however, if the bias for homorganic syllables at nine months is a consequence of infants listening to their own productions (DePaolis, Vihman & Keren-Portnoy, 2011). The infant's perception of his or her own productions may be a relevant form of input, matching a motor pattern with the auditory input, and thus activating sensory-motor feedback loops (Davis & MacNeilage, 2000).

Alternatively, many studies have shown examples of perception preceding and predicting production in language development (Kuhl & Meltzoff, 1996; Tsao, Liu & Kuhl, 2004). The finding that the perceptual bias of twelve-month-olds has shifted towards heterorganic patterns, while twelve-month-olds usually still produce homorganic patterns, is another likely instance of perception preceding production. It would be worthwhile to test infants of a younger age, i.e. before they start canonical babbling, in order to disentangle these two possibilities. Collecting and analyzing both production patterns and perceptual preferences of infants would be another option to elucidate this issue. For now we deem it likely that both factors play a role, since auditory feedback mechanisms must rely on both auditory and motor input while the infant is speaking or babbling. The interaction between perception and production has also been pointed out by Davis and MacNeilage (2000) from an embodiment perspective, suggesting a mechanism where intrinsic perception, i.e. perception of own productions is matched with extrinsic perception, i.e. perception of the environment.

The observed homorganicity bias is in contrast with an earlier study on nine-month-olds investigating learnability of marked and unmarked patterns (Seidl & A Buckley, 2005). The aim of the study was to see whether phonetically grounded patterns (homorganic) were learnt differently from 'arbitrary' ones. Their stimuli were sets of CVCV patterns of which the first syllable was either homorganic or heterorganic. Infants were tested to see whether they generalized homorganic sets more easily than 'arbitrary' ones, which included both homorganic and heterorganic stimuli. Their results indicated that infants were able to learn a rule with homorganic patterns, but also with the arbitrary sets. The authors concluded that learnability of unmarked patterns is not different from that of marked ones. The approach was slightly different from the present study, which may explain the different results. First of all, the paradigm by Seidl et al. tests what infants can *learn,* while the present study is testing what infants naturally attend to. However, we would also expect that infants learn more easily if they are more attentive. Alternatively, a methodological issue may explain the results: a looking time difference between the generalization phase and a test phase with novel items. This is a correct setup for testing generalizations, however the consonants in the test items were also novel which, rather than the novel rule, may have caused the longer looking time in the test phase. Nevertheless, a difference might be expected between the homorganic and arbitrary condition, which was not found. However, the test items used by Seidl et al. were less homorganic than the items used in the present experiment: only the first syllable in their bisyllabic items was homorganic. If the results found here do reflect a preference for homorganicity (i.e. larger units than the segment) it would be expected to have an effect only if the consonants and vowels of the whole test item are of similar PoA. Further research is needed to elucidate this issue.

As mentioned before, the idea of infants processing units larger than the segment would also correspond to the findings for the nine month-old infants, who do respond to homorganicity but not to differences in segmental markedness, indicating a lack of segmental awareness. It should be noted though, that a lack of preference does not necessarily indicate a lack of discrimination. Previous research also indicated that infants at this age are capable of discriminating subtle segmental differences, at least in word initial position (Eimas, Siquelan, Jusczyk & Vigorito, 1971; Werker & Tees, 1984; Zamuner, 2006). This implies that infants can discriminate at the segmental level in a habituation-dishabituation setting, though spontaneous attention as measured in the present study does not reveal any difference because to the infant the one side of the contrast is not more salient or attractive than the other. For example, /t/ and /d/ can be discriminated by infants but this doesn’t entail that /t/ is a more salient or attractive segment for infants than /d/.

The finding of an early perceptual bias changing over time is also interesting with respect to language evolution. Computational linguists have shown that cultural transmission can lead to universals by amplifying weak innate biases (Kirby, Dowman & Griffiths, 2007). Kirby et al. also suggest that cultural evolution can possibly override innate predispositions. The present data show an initial preference indicating a possible predisposition, either for homorganic syllables or for processing units at the syllable level. The change in preference at 12 months of age could possibly be an indication of cultural evolution overriding these initial biases through experience. However, at this point it is not possible to distinguish between development due to maturation or due to cultural transmission and experience.

The question if a bias in perception precedes a bias in production on the evolutionary scale is relevant as well. If production were first, sensory biases matching these production patterns may have emerged as a consequence. Another possibility is that, in case of a perceptual predisposition, it could be a sign of sensory exploitation; a predisposition for a specific feature shapes the evolution in the direction of this feature. More specifically, a pre-linguistic bias for homorganicity in larger units might have caused communication sounds to change in this direction. In linguistics a similar idea has been proposed (Christiansen & Chater, 2008), suggesting that many aspects of language may not have evolved due to linguistic adaptations, but rather emerge from general learning and processing capacities already present before language emerged. This has been suggested for categorization of speech sounds for instance. Earlier, categorization of speech sounds was thought to be a uniquely human adaptation to language. Studies on chinchillas have shown that they categorize voiced and voiceless consonants in a way comparable to humans (Kuhl & Miller, 1978). Recently, in a study with songbirds, it was shown that they discriminated vowels using the same acoustic distinctions that human listeners make. Together, these studies suggest a pre-existing perceptual mechanism (Ohms, Gill, Van Heijningen, Beckers & ten Cate, 2009).

Taken together the present results shed new light on both development and evolution of linguistically relevant features. Although the exact role of the initial bias for homorganic syllables still needs to be clarified, the initial bias and developmental change suggest an important function for perceptual mechanisms in phonological development.

Acknowledgements

We thank Joost van de Weijer for providing us with the frequency data. We also thank Jeanette van der Stelt for valuable advice. We thank all student assistants for helping with collecting and analyzing data and administrative work.

APPENDIX A – Normality tests

Total looking time

9 months

Shapiro-Wilk	raw data			Sqrt transformed data		
	Statistic	df	Sig.	Statistic	df	Sig.
frequent	.821	19	.002*	.891	19	.034*
infrequent	.940	19	.267	.960	19	.581
poaM	.907	19	.065	.941	19	.280
poaUM	.887	19	.029*	.965	19	.666
heterorganic	.968	19	.731	.918	19	.105
homorganic	.908	19	.069	.983	19	.973
voiceM	.965	19	.664	.975	19	.876
voiceUm	.970	19	.780	.941	19	.277
vowM	.940	19	.265	.947	19	.350
vowUM	.948	19	.365	.957	19	.524

Group data (9- and 12 months)

Shapiro-Wilk	raw data			Sqrt transformed data		
	Statistic	df	Sig.	Statistic	df	Sig.
frequent	.851	25	.002*	.925	25	.067
infrequent	.931	25	.094	.962	25	.452
poaM	.923	25	.059	.972	25	.703
poaUM	.915	25	.039*	.974	25	.745
heterorganic	.983	25	.937	.962	25	.448
homorganic	.917	25	.045*	.987	25	.982
voiceM	.984	25	.949	.978	25	.854
voiceUm	.978	25	.849	.946	25	.203
vowM	.970	25	.637	.953	25	.300
vowUM	.959	25	.395	.975	25	.777

Longest look
9 months

Shapiro-Wilk	raw data			Log transformed data		
	Statistic	df	Sig.	Statistic	df	Sig.
frequent	.651	20	.000*	.915	20	.078
infrequent	.946	20	.311	.952	20	.398
poaM	.870	20	.012*	.982	20	.958
poaUm	.629	20	.000*	.982	20	.953
heterorganic	.943	20	.277	.919	20	.094
homorganic	.794	20	.001*	.917	20	.088
voiceM	.946	20	.309	.961	20	.565
voiceUm	.844	20	.004*	.939	20	.234
vowM	.938	20	.217	.983	20	.965
vowUm	.627	20	.000*	.972	20	.789

Group data (9- and 12 months)

Shapiro-Wilk	raw data			Log transformed data		
	Statistic	df	Sig.	Statistic	df	Sig.
frequent	.697	29	.000*	.949	30	.160
infrequent	.900	29	.010*	.982	30	.874
poaM	.891	29	.006*	.951	30	.184
poaUm	.699	29	.000*	.977	30	.742
heterorganic	.868	29	.002*	.938	30	.079
homorganic	.781	29	.000*	.932	30	.057
voiceM	.926	29	.045*	.973	30	.611
voiceUm	.842	29	.001*	.950	30	.166
vowM	.951	29	.191	.983	30	.907
vowUm	.668	29	.000*	.982	30	.874

Shapiro-Wilks tests of normality were performed for Longest Look (LL) and Total Looking time (TL) raw data. Part of the TL data was significantly different from normal distribution (indicated by *). TL data were square root transformed (sqrt). Normality test on transformed data show that data are mostly normally distributed after sqrt transformation. LL data were also not always normally distributed before transformation. For LL data sqrt transformations were not sufficient to reach normality so log transformations were used. After log transformation no significant differences from normality were found in LL data.

APPENDIX B – Statistics for transformed and raw data

The following tables show statistics for all ANOVAs. Transformed data for nine-month-olds (table B1), for nine- and twelve-month-olds grouped (table B2), raw data for nine-month-olds (table B3) and raw grouped data (table B4). Significant results are indicated by asterisks: * p< .05, ** p< .01, *** p< .001. LL: longest look, TL: total looking time, order: within stimulus order, SV: stimulus voice, ^{a}: df = 1.

Table B1. Analysis of variance at 9 months (transformed data)

	LL data			TL data		
N=35	F	p	η_p^2	F	p	η_p^2
	within subject			within subject		
hom (homoriganicity)[a]	24.610	.000***	.435	9.866	.004**	.236
hom x order[a]	1.798	.189	.053	1.081	.306	.033
hom x SV[a]	2.570	.119	.074	.135	.715	.004
error df (mean square)	32	(.048)		32	(2138193.837)	
	between subjects			between subjects		
order[a]	.431	.516	.013	3.735	.062	.105
SV[a]	1.082	.306	.033	4.856	.035*	.132
error df (mean square)	32	(.094)		32	(4353301.388)	
	LL data			TL data		
N=35	F	p	η_p^2	F	p	η_p^2
	within subject			within subject		
frequency (fr) [a]	3.862	.058	.108	2.078	.159	.061
fr x order[a]	.021	.886	.001	.077	.783	.002
fr x SV[a]	1.452	.237	.043	.645	.428	.020
error df (mean square)	32	(.041)		32	(129.368)	
	between subjects			between subjects		
order[a]	.003	.955	.000	.008	.931	.000
SV[a]	.189	.666	.004	.016	.899	.001
error df (mean square)	32	(.120)		32	(445.830)	
	LL data			TL data		
N=32	F	p	η_p^2	F	p	η_p^2
	within subject			within subject		
poa[a]	1.594	.217	.052	.014	.907	.000
poa x order[a]	3.309	.079	.102	2.756	.108	.087
poa x SV[a]	1.883	.180	.061	3.821	.060	.116
error df (mean square)	29	(.051)		29	(2006295.269)	
	between subjects			between subjects		
order[a]	.282	.600	.010	3.449	.073	.106
SV[a]	.555	.462	.019	.902	.350	.030
error df (mean square)	29	(.199)		29	(6843071.499)	
	LL data			TL data		
N=32	F	p	η_p^2	F	p	η_p^2
	within subject			within subject		
voicing[a]	1.825	.187	.059	.707	.407	.024
voicing x order[a]	.142	.709	.005	.613	.440	.021
voicing x SV[a]	.715	.405	.024	.998	.326	.033
error df (mean square)	29	(.080)		29.000	(3087225.174)	
	between subjects			between subjects		
order[a]	.244	.625	.008	.169	.684	.006
SV[a]	.306	.584	.010	.303	.586	.010
error df (mean square)	29	(.077)		29	(5409023.925)	
	LL data			TL data		
N=31	F	p	η_p^2	F	p	η_p^2
	within subject			within subject		
vow[a]	1.347	.256	.046	.645	.429	.023
vow x order[a]	1.549	.224	.052	1.023	.321	.035
vow x SV[a]	1.493	.232	.051	.228	.636	.008
error df (mean square)	28	(.048)		28	(3442421.502)	
	between subjects			between subjects		
order[a]	1.277	.268	.044	.265	.610	.009
SV[a]	1.714	.201	.058	.030	.864	.001
error df (mean square)	28	(.127)		28	(5575748.443)	

Table B2. Analysis of variance for 9- and 12-month-olds (transformed data)

	LL data			TL data		
N=53	F	p	η_p^2	F	p	η_p^2
	within subject			within subject		
hom (homoriganicity) [a]	.000	.990	.000	.196	.660	.004
hom * age group [a]	21.393	.000***	.296	24.274	.000***	.322
error df (mean square)	51	(1938516.833)		51	(126.366)	
	between subjects			between subjects		
age group [a]	.319	.575	.006	.267	.607	.005
error df (mean square)	51	(3290029.287)		51	(336.080)	
	LL data			TL data		
N=50	F	P	η_p^2	F	p	η_p^2
	within subject			within subject		
freq [a]	3.719	.060	.072	2.367	.130	.047
freq * age group [a]	.157	.693	.003	.158	.693	.003
error df (mean square)	48	(1839593.228)		48	(141.177)	
	between subjects			between subjects		
age group [a]	.300	.587	.006	.467	.497	.010
error df (mean square)	48	(3895229.100)		48	(378.207)	
	LL data			TL data		
N=45	f	p	η_p^2	f	p	η_p^2
	within subject			within subject		
poa [a]	2.368	.131	.052	.865	.358	.020
poa * age group [a]	.106	.746	.002	.829	.368	.019
error df (mean square)	43	(.049)		43	(160.683)	
	between subjects			between subjects		
age group [a]	.447	.507	.010	.128	.722	.003
error df (mean square)	43	.175		43	(436.747)	
	LL data			TL data		
N=51	F	p	η_p^2	F	p	η_p^2
	within subject			within subject		
voicing [a]	3.445	.069	.066	.101	.752	.002
voicing * age group [a]	.291	.592	.006	.557	.459	.011
error df (mean square)	49	(1932989.355)		49	(149.691)	
	between subjects			between subjects		
age group [a]	.785	.380	.016	.127	.723	.003
error df (mean square)	49	(2536864.346)		49	(425.467)	
	LL data			TL data		
N=51	F	p	η_p^2	F	p	η_p^2
	within subject			within subject		
vow [a]	.362	.550	.007	.099	.754	.002
vow * age group [a]	1.543	.220	.031	.167	.684	.003
error df (mean square)	49	(1887031.980)		49	(2835713.303)	
	between subjects			between subjects		
age group [a]	1.929	.171	.038	.599	.443	.012
error df (mean square)	49	(3467520.437)		49	(5119881.788)	

Table B3. Analysis of variance at 9 months (raw data)

N=35	F	p	η_p^2	F	p	η_p^2
	within subject			within subject		
hom[a]	16.249**	.000***	.337	9.866	.004**	.236
hom x order[a]	2.593	.117	.075	1.081	.306	.033
hom x SV[a]	3.325	.078	.094	.135	.715	.004
error df (mean square)	32	(1855289.361)		32	(2138193.837)	
	between subjects			between subjects		
order[a]	1.594	.216	.047	3.735	.062	.105
SV[a]	2.711	.109	.078	4.856	.035*	.132
error df (mean square)	32	(2709103.976)		32	(4353301.388)	

	LL data			TL data		
N=35	F	p	η_p^2	F	p	η_p^2
	within subject			within subject		
frequency (fr) [a]	4.204	.049*	.116	1.334	.257	.040
fr x order[a]	.206	.653	.006	.037	.849	.001
fr x SV[a]	1.099	.302	.033	.704	.408	.022
error df (mean square)	32	(2339205.666)		32	(2785600.980)	
	between subjects			between subjects		
order[a]	.208	.652	.006	.004	.949	.000
SV[a]	.232	.633	.007	.008	.928	.000
error df (mean square)	32	(4489089.071)		32	(7883493.624)	

	LL data			TL data		
N=32	F	p	η_p^2	F	p	η_p^2
	within subject			within subject		
poa[a]	.546	.466	.018	.014	.907	.000
poa x order[a]	4.086	.053	.124	2.756	.108	.087
poa x SV[a]	4.600*	.040*	.137	3.821	.060	.116
error df (mean square)	29	(1063970.288)		29	(2006295.269)	
	between subjects			between subjects		
order[a]	1.190	.284	.039	3.449	.073	.106
SV[a]	.211	.650	.007	.902	.350	.030
error df (mean square)	29	(5066129.462)		29	(6843071.499)	

	LL data			TL data		
N=32	F	p	η_p^2	F	p	η_p^2
	within subject			within subject		
voicing[a]	2.131	.155	.068	.707	.407	.024
voicing x order[a]	.018	.893	.001	.613	.440	.021
voicing x SV[a]	.742	.396	.025	.998	.326	.033
error df (mean square)	29	(2602963.807)		29.000	(3087225.174)	
	between subjects			between subjects		
order[a]	.114	.739	.004	.169	.684	.006
SV[a]	.134	.717	.005	.303	.586	.010
error df (mean square)	29	(2176303.873)		29	(5409023.925)	

	LL data			TL data		
N=31	F	p	η_p^2	F	P	η_p^2
	within subject			within subject		
vow[a]	2.299	.141	.076	.645	.429	.023
vow x order[a]	.015	.902	.001	1.023	.321	.035
vow x SV[a]	.182	.673	.006	.228	.636	.008
error df (mean square)	28	(1910666.963)		28	(3442421.502)	
	between subjects			between subjects		
order[a]	.405	.530	.014	.265	.610	.009
SV[a]	.368	.549	.013	.030	.864	.001
error df (mean square)	28	(3470713.258)		28	(5575748.443)	

Table B4. Analysis of variance for 9- and 12-month-olds (raw data)

	LL data			TL data		
N=53	F	p	η_p^2	F	p	η_p^2
	within subject			within subject		
hom (homoriganicity)[a]	.000	.990	.000	.593	.445	.011
hom * age group[a]	21.393	.000***	.296	25.148	.000***	.330
error df (mean square)	51	(1938516.833)		51	(1968067.560)	
	between subjects			between subjects		
age group[a]	.319	.575	.006	.194	.661	.004
error df (mean square)	51	(3290029.287)		51	(5057541.534)	

	LL data			TL data		
N=50	F	p	η_p^2	F	p	η_p^2
	within subject			within subject		
frequency (fr)[a]	3.719	.060	.072	1.291	.262	.026
fr * age group[a]	.157	.693	.003	.006	.940	.000
error df (mean square)	48	(1839593.228)		48	(2627114.837)	
	between subjects			between subjects		
age group[a]	.300	.587	.006	.568	.455	.012
error df (mean square)	48	(3895229.100)		48	(6504985.021)	

	LL data			TL data		
N=45	F	p	η_p^2	F	p	η_p^2
	within subject			within subject		
poa[a]	1.622	.210	.036	.650	.425	.015
poa * age group[a]	.193	.662	.004	.721	.400	.016
error df (mean square)	43	(1106286.381)		43	(2012263.162)	
	between subjects			between subjects		
age group[a]	.030	.863	.001	.188	.666	.004
error df (mean square)	43	(4555172.719)		43	(6714953.688)	

	LL data			TL data		
N=51	F	p	η_p^2	F	p	η_p^2
	within subject			within subject		
voicing[a]	3.445	.069	.066	.183	.671	.004
voicing * age group[a]	.291	.592	.006	.787	.379	.016
error df (mean square)	49	(1932989.355)		49	2301958.284	
	between subjects			between subjects		
age group[a]	.785	.380	.016	.066	.798	.001
error df (mean square)	49	(2536864.346)		49	(6254148.598)	

	LL data			TL data		
N=51	F	p	η_p^2	F	p	η_p^2
	within subject			within subject		
vow[a]	.362	.550	.007	.099	.754	.002
vow * age group[a]	1.543	.220	.031	.167	.684	.003
error df (mean square)	49	(1887031.980)		49	(2835713.303)	
	between subjects			between subjects		
age group[a]	1.929	.171	.038	.599	.443	.012
error df (mean square)	49	(3467520.437)		49	(5119881.788)	

3 The interplay of perceptual predispositions and experience on song preferences and development in zebra finches (Taeniopygia guttata)

Sita M. ter Haar[1,2,3,],
Wiebke Kaemper[1], Koen Stam[1], Clara C. Levelt[2,3,], Carel ten Cate[1,3]

[1] Institute Biology Leiden (IBL), Leiden University, Leiden University, P.O. box 9505, 2300 RA, Leiden, The Netherlands

[2] Leiden University Centre for Linguistics (LUCL), Leiden University, P.O. box 9515, 2300 RA, Leiden, The Netherlands

[3] Leiden Institute for Brain and Cognition (LIBC), Leiden University

Abstract

Vocal acquisition in songbirds and humans show many similarities, one of which is that both are driven by a combination of experience and perceptual predispositions (Bolhuis et al, 2010; Doupe & Kuhl, 1999). Among languages, some speech sounds are shared, while others are not. This could reflect a perceptual predisposition in young infants for learning some speech sounds over others, which combines with an exposure based preference to guide learning. Similarly, in songbirds some sounds are general across populations while others are more specific to populations or individuals. Here we examine whether this is also due to perceptual preferences for certain species specific element types in naïve juvenile birds and how such preferences interact with exposure to guide subsequent song learning. We show that young male zebra finches lacking previous song exposure perceptually prefer more common (general) over less common elements of species specific songs. This indicates a bias for within-species vocalizations, independent of exposure. Next we demonstrate that subsequent exposure to either common or to less common elements alters the birds' perceptual preference, resulting in a preference for tutor song elements. In adulthood, birds tutored with more common elements showed a higher song similarity to their tutor song, indicating that the early bias influenced song learning. Our findings help to understand the maintenance of vocal similarities, as well as the presence of differences, among birds' songs, their dialects and human languages.

INTRODUCTION

Vocal learning is essential for spoken language as well as for bird song, and the learning processes involved show many parallels (Bolhuis et al, 2010; Doupe & Kuhl, 1999). One interesting parallel is that both processes are guided by perceptual predispositions (i.e. perceptual biases independent of perceptual experience) that interact with experience. While presence of this interaction is broadly accepted, debate and discussions concern the nature of the predispositions and of the interaction in shaping vocal production (Adret 2004; Woolley 2012). In this paper we address these questions for a songbird species, the zebra finch.

The presence of predispositions in vocal learning is suggested by the distribution of sound patterns within and between populations. In human languages, some sounds are more or less universally shared, suggesting a possible predisposition for such sounds, while others are more language specific. Similarly, different populations of the same songbird species can share elements, but also sing different 'dialects' and song elements (Kroodsma et al, 1999; Marler & Tamura, 1964; Petrinovich & Baptista, 1984). In addition to geographic variation, song can also differ between individuals. Zebra finch song, for instance, consists of different types of elements (fig 3.1) and birds may vary in which elements are used and how they are combined. Some element types are more common across individuals than others. A recent study of song elements present in 13 different zebra finch populations showed that, although all element types occur in all populations, the proportion of some elements varied between populations whereas other elements are more equally present across populations. Also, individuals within a population can differ substantially in which elements they share or not (this thesis, Chapter 4). These differences on population and individual level have implications for understanding both development and evolution of learned vocalizations. Patterns or elements that are common across individuals and populations might indicate species general predispositions facilitating selective learning and constraining vocal variation. The population or individually specific elements might not be based upon such predispositions but arise from plasticity in the learning process, allowing deviating elements to develop and be learned and maintained by cultural transmission. In such a scenario, predispositions for common elements would likely be present early in development before song exposure, guiding the learner to particular conspecific sounds. Preferences for less common elements are expected to emerge later on as a result of experience.

Evidence for predispositions so far has mainly come from experiments showing a preference to learn conspecific sounds over heterospecific ones in studies involving isolate rearing and tape tutoring. In most studies adult song production is used as a measure of learning or selective preference. Only few studies have examined perceptual predispositions in naïve birds. An experiment in which juvenile zebra fiches (*Taeniopygia guttata*) could elicit exposure to either conspecific or heterospecific song by hopping on a perch showed that birds hopped more on the perch generating conspecific song than on the one for heterospecific song (Braaten & Reynolds, 1999). In another species, the white-crowned sparrow (*Zonotrichia leucophrys*), fledglings produced more begging calls in response to conspecific song than to heterospecific song (Nelson & Marler, 1993; Soha & Marler, 2001). Preference for the own subspecies over other subspecies was not confirmed, but exposure to songs of the own subspecies lead to better discrimination than experience with another subspecies' song. This outcome suggests that the perceptual system is more attuned to acoustic features of the own subspecies (Nelson, 2000). Moreover, the universal white-crowned sparrow's introductory whistle has been shown to function as a cue for song learning, since songs (even heterospecific ones) are better copied when they contain these universal whistles (Soha & Marler, 2000). This suggests that in addition to a preference for conspecific versus heterospecific vocalizations, there are predispositions for certain *within-species* element types. This has also been shown in a study on grasshopper sparrows, showing that naïve female fledglings respond more to one conspecific song type ('buzz' simple structure and uniform across individuals) than another conspecific one ('warble', complex and possibly individually specific) (Soha et al, 2009).

While the above mentioned studies suggests the presence of *within species* perceptual predispositions in male songbirds, clear demonstrations of their presence are lacking, as well as insight in the effect of later exposure on perceptual preferences. Moreover, little is known about the relative effects of predispositions and exposure on later song production. In the present study we address these questions. Perceptual preferences are tested in juvenile male zebra finches at different stages of development. By testing the birds before hearing song and by manipulating subsequent exposure, we disentangle the effects of possible predispositions and auditory song experience. We also examine the similarity of the acquired songs to those heard during exposure and discuss how predispositions and vocal learning relate to the distribution of song elements over populations and individuals.

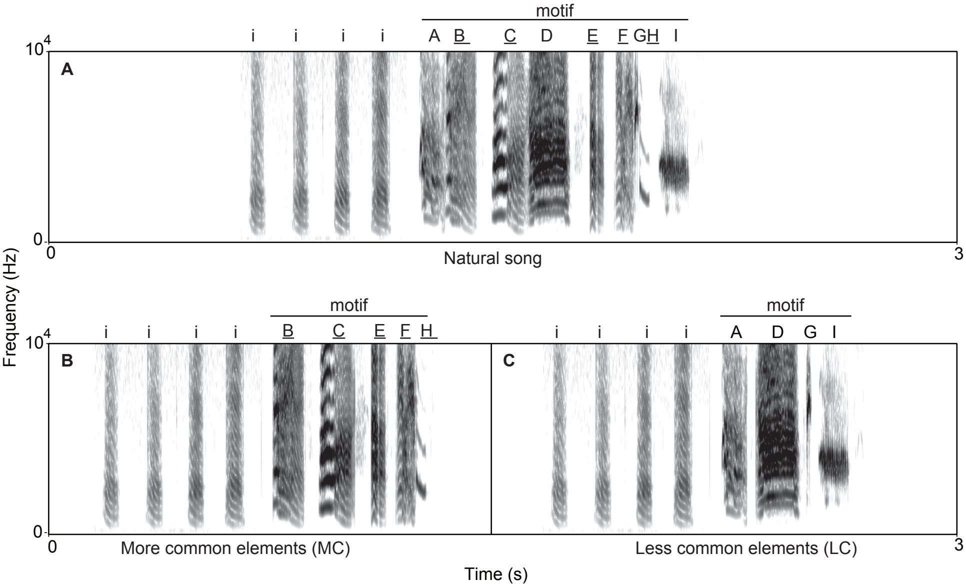


Figure 3.1. Examples of one pair of stimuli, constructed from one original song (**A**). From the original song, more common (MC) element types were selected (indicated by underlined letters) and combined into an artificial 'common song' stimulus (**B**) and similarly 'uncommon' song stimuli were constructed (**C**) using less common (LC) elements of the same original song. Both stimulus types started with four introductory notes from the original song (indicated by 'i').

Results

Perceptual preference for more common element types in male birds naive to song
Juvenile male zebra finches were reared by their mother only, from approximately 8 dph (days post hatch), well before the start of the sensitive phase for song learning (Eales, 1985; Eales, 1987; Jones et al, 1996a). As female zebra finches do not sing, we thus created relatively natural rearing conditions where zebra finches are not exposed to song. The birds were tested at 37 dph (+-2d) when they were naive to song, to see if they have a preference for more common or less common elements (see methods for details).

The classification of elements as common or not, was based on literature describing the element types found across several populations (Holveck et al, 2008; Leadbeater et al, 2005; Sturdy et al, 1999; Zann 1993) or reporting rare elements ((Leadbeater et al, 2005), see methods and SI table 1). Based on this classification, we created two paired stimuli from each one of eight original natural songs. The more common elements of a song were selected for constructing a 'common song'. The less common elements of the same original song were used to construct the 'uncommon song' (fig 3.1). Each bird was tested with four pairs of common and uncommon songs on the same day.

The results show that juvenile males significantly prefer common songs over uncommon ones (n=12, deletion $p < 0.01$, SI, fig 3.2). So, initially, before song exposure, males have a bias for more common zebra finch song element types.

Preference for tutor song at 57 dph

After the 37 dph preference tests the birds were exposed to (tutored with) either a common or an uncommon song until approximately 67 dph. The tutor songs were selected from the same set of stimuli used for the preference test and each bird was tutored with a different song. The preference tests were repeated at 47 (+-2) dph and 57 (+-2) dph in order to test the effect of the subsequent exposure on the birds' preferences. Each preference test again consisted of four blocks (four pairs of common versus uncommon song), one of which included the tutor song. By comparing preference for the pair including

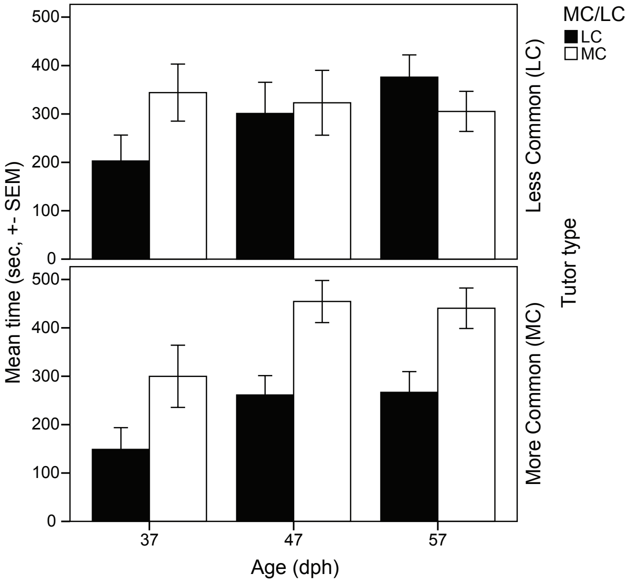


Figure 3.2. Preferences independent and dependent of song exposure. The preference measured is time in seconds (+-SEM) near the speaker broadcasting songs constructed with more common elements (MC) or songs constructed with less common elements (LC) for zebra finches at 37 dph (before tutoring, see methods). This preference is higher for more common elements (white bar) then for less common elements (black bar). At 57 dph (after tutoring) birds prefer the type of sounds they have been tutored with; birds tutored with more common elements (lower panel, MC) prefer songs with more common elements and birds tutored with less common elements (top panel, LC) preferred less common element types.

the tutor song to the other three pairs, we could test whether the preference at 57 dph was specific for the tutor song only or that other songs of the same type (common/uncommon) are preferred.

A significant 4-way interaction was found between preference (common/uncommon song), age (37,47, 57 dph), tutor type (common/uncommon song) and block (tutor/non-tutor). This indicates that the preference changes with age, depending on the type of tutoring and whether it is the tutor song or not (n=16, deletion $p < 0.05$, fig 3.2, SI). Since 4-way interactions can be hard to interpret and to confirm the interactions at lower levels, separate analyses at 57dph, when the interactions at expected. This revealed a significant interaction between preference and tutor type at 57 dph for the blocks including tutor songs (n=13, deletion $p < 0.01$), but not for the other three blocks (n=16, deletion $p > 0.05$), suggesting that the later preference was specific to the tutorsong. When we look at the two tutor groups separately the interaction between preference and block (tutor/non-tutor) was significant for both the birds tutored with common songs (n=8, deletion $p< 0.01$) and birds tutored with uncommon songs (n=8, deletion $p < 0.05$, fig. 3.3). These findings indicate that birds specifically prefer their tutor song (and thus song exposure) at 57 dph, in line with previous findings in adult birds (Houx & ten Cate, 1999a,b, Riebel et al, 2002).

Experience affects song learning.

After the last preference test at 57 dph the birds were tutored in isolation for 10 more days. Adult birds' songs were recorded (approximately 120 dph or older (Jones et al, 1996b)) and similarity between the pupils song and the tutor song was measured using Sound Analysis Pro software (SAP (Tchernichovski et al, 2000)).These measurements revealed that the pupils' song similarity with the tutor song was higher than similarity with a control song of the same type (common/uncommon song, paired Wilcoxon: n=16, $p < 0.05$), confirming earlier findings that experience leads to song copying. We also compared similarity with the tutor song to similarity with its counterpart, derived from the same original natural song but belonging to the other song category (fig 3.1). Thus for song from a pupil tutored with common songs (fig 3.1b), a comparison was made for similarity with the common tutor song (fig 3.1b) and the uncommon counterpart of that song (fig 3.1c, and vice versa for the pupil tutored with uncommon song). This comparison revealed an overall tendency for higher similarity with the tutor song than with its counterpart (paired Wilcoxon: n=16, $p = 0.06$). Most noticeable, however, was the tutor group difference found for this latter comparison. Pupils tutored with common songs showed higher similarity with the tutors than with their counterparts (paired Wilcoxon: n=8, $p = 0.02$), whereas this difference was not significant for the group tutored with uncommon songs (n=8, $p > 0.05$, group difference: Kruskal-Wallis: n=16, $p= 0.02$). In other words, evidence for similarity with the tutor song is stronger for pupils tutored with common song.

Discussion

Our findings suggest that juvenile birds naive to song have perceptual predispositions, making some elements more attractive than others. In addition to this, we show that these perceptual preferences can be modified by exposure, even resulting in a preference for initially non-preferred elements. Furthermore, while young birds can incorporate both types of song elements in their later songs, common elements are more likely to be copied.

Altogether, these observations provide evidence of a mechanism that may explain the species wide presence and maintenance of particular types of elements in a vocal learning species as follows. The perceptual bias present in naïve juvenile males can guide the learning process, directing the learners' attention towards particular conspecific vocal elements. This is likely to result in including these element types in the birds' later song production. The result of this process will be that these elements are more likely to be maintained in a population and hence become, or stay, more common. Over generations this process is likely to cause stabilization of vocal patterns containing these common features. There is an interesting parallel here with a mechanism that has been proposed for language which has described in terms of markedness, suggesting unmarked

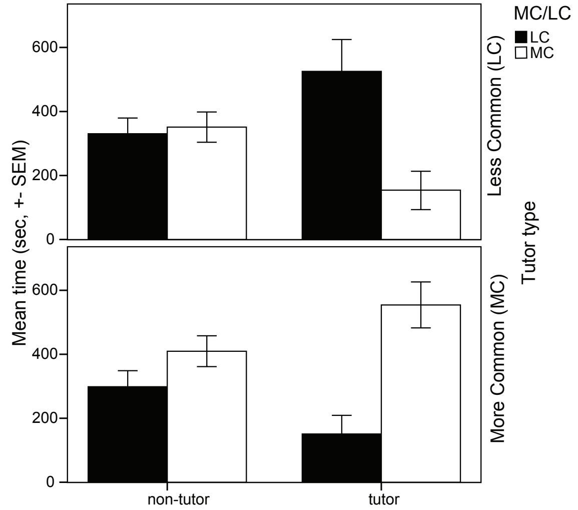


Figure 3.3. Specific preference for tutor song at 57 dph. Birds tutored with songs containing less common elements (top panel, LC) as well as birds tutored with songs containing more common elements (lower panel MC) prefer their tutor song. Birds lost their initial preference, thus they do not show a preference for common or uncommon songs that were not their tutor song.

('universal') sounds are acquired early in development and marked ones later. Unmarked sounds are more likely to (re)occur and to be maintained in languages. Acquisition of unmarked sounds (De Lacy, 2006), similar to acquisition of common song elements, might be driven by processes independent of linguistic input (Jakobson, 1941).

In addition to processes resulting in song conformity, if a young bird is exposed to elements for which there is initially no perceptual bias, this exposure can still result in copying such elements. This may result in song variation, and the appearance and maintenance of these elements in a population might depend on local factors and chance (drift) affecting cultural transmission.

Interestingly, our study provides empirical support for a mechanism suggested by a study by Feher et al., that elegantly showed a process of vocal convergence on more common, species specific, song features over generations (Feher et al, 2009). Zebra finch males reared without song exposure which produce aberrant song were used as tutors for a second 'tutor generation'. These latter birds were again used as tutors for a third tutor generation and so on. By three to four generations, songs had evolved towards songs with wild type characteristics. The tutees copied most of the elements of the aberrant song, but also induced alterations to their tutor song. Due to accumulation of these alterations over tutor generations, the songs in the later generations became more similar to wild type songs. We suggest that the biases we demonstrated affected the direction of element deviations, driving them towards becoming more similar to preferred (and also more normal and common) elements. While the observed bias is independent of song exposure, we cannot fully exclude that it is fully independent of any acoustic exposure, as it may have been affected by the mothers' vocalizations during rearing. Even though females do not sing they do produce calls that may affect the auditory preferences. There is some evidence for perceptual preferences being formed before 35 dph (Clayton, 1988; Roper & Zann, 2006). The female's call has some features in common with some of the more common elements (stacks) but certainly not all of them (for instance slides). Future research could elucidate this issue by examining in more detail which specific acoustic features make certain element types attractive and by using muted females to rear the subjects. Regardless of the cause of the sensitivity, the effect of the bias is independent of *song* exposure and thus remains relevant in terms of development and evolution and may result in maintenance of such sounds in a species.

While the present experiment can provide insight in the evolutionary consequences of the perceptual biases and the developmental processes involved, less can be said about the evolutionary origin of the perceptual biases. Nevertheless, our findings may be interesting for future research on mate attraction. In zebra finches songs are supposed to function primarily in mate choice and pair bonding. There is ample evidence that female zebra finches prefer specific songs or song features over others (Riebel, 2009). These preferences may concern the presence of certain general features of songs. For example, naïve and normally reared female zebra finches prefer conspecific song (Braaten & Reynolds, 1999; Lauay et al, 2004), which might help to maintain the species specificity of male songs. On the other hand, females prefer tutored song (normal quality) song over untutored song (abnormal quality, (Lauay et al, 2004)) and larger repertoires over smaller ones (Holveck & Riebel, 2007). This might drive the use of additional uncommon song elements and thus male song plasticity, since tutored song requires experiences and larger repertoires are more likely to include more uncommon elements in addition to the common ones. However, little is known about specific element types being preferred by females and how the presence of these contributes to attractiveness of the song as a whole. It is known that females don't systematically prefer songs with expiratory elements (classified as common) over songs with inspiratory high notes (here classified as uncommon, (Leadbeater et al, 2005)), but it would require more specific tests to examine whether female zebra finches differentially prefer (songs with) common or uncommon elements, to get insight in the evolutionary dynamics from which the current male songs have arisen.

The finding that both an initial bias for more common elements and later experience affect song learning has a striking parallel in human infants. In infants, early phonemic discrimination is universal and becomes more language specific later on (Tsao et al, 2006). A similar developmental change can be observed for acquisition of syllable structure. In early language productions the first syllables are of the CV (consonant-vowel) type which is common across different languages. Subsequent development of novel syllable types is influenced by frequency of occurrence and may therefore also be experience dependent (Levelt et al, 2000). Although there is a clear parallel, the distinction between more and less common elements is somewhat different from that between universal versus non-universal speech sounds in humans. In zebra finches there are clear individual differences within populations, while differences in sound inventories between populations are less clear (this thesis, chapter 4). In contrast, human speech sound inventories differ between languages and people speaking the same dialect usually make use of approximately the same phoneme inventory. It should be noted however, that the use of different analytical methods for human language and birdsong makes a direct comparison difficult. Thus the developmental mechanism may be similar (attention changing from more common to less common, i.e. from internal biases to external influences), but the eventual effect of the developmental plasticity due to the vocal learning may differ between humans and songbirds. If the developmental mechanism is indeed the same for birds and humans, the implications described above may also hold for language evolution. Initial biases could maintain the universals in languages whereas additional plasticity allows for learning language-specific patterns and facilitates cultural evolution.

Material and methods

Subjects and housing

For the experiments 16 male wild-morph domesticated zebra finches from an out bred breeding colony at Leiden University, the Netherlands, were used. The birds were kept at 20-22 °C and 55-65 % humidity on a light dark schedule of 13.5:10.5 hours. Food, water and a cuttlebone were available ad libitum.

At the age of 8 dph (± 2) the young birds and their mother were moved into a room where no adult males were present. At the age of 37 dph (± 2) the birds received the first preference test. Each bird was moved to the preference cage the day before the test in order to acclimatize to the new cage and to isolation. After the test, the birds were isolated in sound attenuated chambers for song exposure. In the sound attenuated chambers food, water and cuttlebone was available ad libitum. The light dark schedule in the sound attenuation chambers was 13.5:10.5 hours with a temperature of 21-24 °C and a humidity of 50-55 %. The experiment was approved by the Leiden University Committee for Animal Experimentation (DEC) under proposal number 10043.

Stimuli

The classification of elements as common or not, was based on literature describing the element types found across several populations (Holveck et al, 2008; Leadbeater et al, 2005; Sturdy et al, 1999; Zann 1993) or reporting rare elements (Leadbeater et al, 2005). These data are summarized in SI table 1 which was used to estimate which elements were more or less common among populations or individuals. As can be seen in SI table 3.1, the frequencies are rather a continuum than discrete distinction between common and uncommon. However we can estimate that stacks, slides, short slides and tones are in general more common elements and high notes (especially inspiratory ones), trills, high sweeps and noisy elements and elements that do not clearly fall into a category, are less common.

Elements were used to construct 'common' and 'uncommon' songs from natural songs produced by normally reared birds in the Leiden University zebra finch colony. From each of 8 natural songs, two versions of a motif were created: one 'common' version, using the common elements of the song and one 'uncommon' version using the uncommon elements from the same original song (fig 3.1). This way, individual factors like voice characteristics cannot be the cause of the difference in preference between common and uncommon stimuli. Each stimulus song consisted of 4 introductory notes followed by 5 motifs.

The stimuli used for the preference tests and for exposure were modified by using Praat sound analysis software (version 5.1.41 for windows) and had a mean motif duration of 0.385s (range 0.284-0.519 s) for common songs and 0.379 s (range 0.276-

0.548 s) for uncommon songs. All 16 stimuli were rms equalized.

Exposure

The birds remained in isolation in sound attenuated rooms while tutored with either 'common' or 'uncommon' song via a speaker. The amount of exposure for all birds was the same (approximately 180 bouts per day, of five motifs per bout).

After the first day of preference testing, the birds were moved to a sound attenuated isolation chamber where exposure (tutoring) started the next day. Every bird was tutored for approximately 20 times per hour (random timing) during 7.15 h and 13.15 h and 10 times per hour during 13.15 h and 19.15 h. Thus the amount of exposure for all birds was the same ((6hx20=) 120 + (6hx10=) 60)=180 bouts per day, of five motifs per bout). Birds were tutored each day from age 37 dph to 67 dph, with exception of the days of preference testing.

Preference tests

Birds were tested for their preference at 37 dph (before tutoring started), 47 dph and 57 dph, each time using the same four sets of stimuli (i.e. the same blocks, see below), including the stimuli to which the birds were exposed during the tutor phase.

Preferences were measured using a phonotaxis setup (Holveck & Riebel, 2007); a cage with one speaker on each side, alternating song playbacks with more common element types from one speaker and less common element types from the other. The time spent on the left and the right side of the cage was used as a measure of preference (measured from the first response after playback). When the birds were in the centre of the cage (a neutral zone) this was not included in the response time. Each test consisted of 4 blocks on one day (always in the morning when birds were most active), each block with a different pair of common and uncommon stimuli. Thus two sets of 8 stimuli (4 pairs of common and uncommon song) were used, 8 birds were tested with the first set and 8 birds with the second set.

One block consisted of 14 minutes alternating each minute between common song from one speaker and uncommon song from the other speaker. Each minute contained 7 identical songs. The order of the type of stimulus and side from which they were played back was counterbalanced between blocks and between subjects. Songs were broadcast at approximately 70db. After each block, the bird had a break of 45 minutes before the next block started.

Video recordings of each test were analyzed while blind to the stimuli using ELAN software (version 3.8.1 http://www.lat-mpi.eu/tools/elan/, Max Planck Institute for Psycholinguistics, The Language Archive, Nijmegen, The Netherlands, (Sloetjes & Wittenburg, 2008)). When birds did not show any response during a given block, this block was excluded from further analysis. Due to the lack in response in all 4 blocks, 4 birds had to be excluded from the 37 dph analysis, thus this analyses was based on 12 birds.

Song analysis

Birds' songs were recorded at age 120 dph or older (when they did not sing at 120 dph). From each bird the predominant motif was selected and similarity measures between the subject's song (pupil) and the artificial tutor song were measured using Sound Analysis Pro (SAP2011, (Tchernichovski et al, 2000)).

Statistics

All statistical analyses were performed using R version 2.11.0. Linear mixed effect models were performed for preference test data using the nlme package for R, version: 3.1-96 (Pinheiro et al, 2009). Subject was included as a random factor, with 'block' (the four songs tested per bird per age) nested within bird. Deletion p-values were accomplished by comparing models with and without the variables of interest using the anova method in R. Model assumptions (normally distributed errors and lack of heteroscedasticity) were always verified after model selection.

Statistical analyses for song similarity measures were performed using Wilcoxon Signed Rank tests and Kruskal-Wallis rank sum tests.

Acknowledgements

We would like to thank Harald van Mil, Caroline van Heijningen and Erwin Ripmeester for help with the statistical analyses and Ofer Tchernichovski for feedback with Sound Analysis Pro.

Supplementary Information

Sturdy et al. 1999	JH	QB	R	QP	Leadbeater et al. 2005		Zann 1993		Holveck et al. 2008		Price 1979	
	*	*	*	*		#		*		*		*
combination note	36	3	4	14	combination note	60	distance call	16	-		long call	27
slide note	9	35	22	16	slide note	95	downslur	2	down sweep	53	-	
flat note	9	20	21	10	flat note	90	tone	1	tone	17	medium call	9
							stack	17	stack	9		
short slide	37	35	44	44	-		introductory element	24	-		short call	31
							introductory diad	2				
high	9	7	9	16	inspiratory high	45	high	10	high	3	non-call type	32
					expiratory high	60						
									high sweep	1		
-					noise	10	noise-noise	7	noisy	11		
							noise-structure	*4*	short noisy	1		
							ladder-noise	*4*				
							tone-niose	*3*				
							noise-DC	*3*				
							noise-tone	*3*				
							n-n-DC	*1*				
-					buzz	5	-		trill	3		

SI Table 1. **Elements described in the literature**. The different studies used different classification systems causing some categories to fall into two categories in one study and into one category in another study (for instance inspiratory and expiratory high notes in Leadbeater et al., are grouped in to 'high notes' according to Sturdy et al.'s classification). In order to compare studies, Sturdy et al. is used as a reference point and element types on the same row are expected to be similar to a certain extend. This table was used as an estimate to classify elements into 'more' or 'less' common for constructing the stimuli. Frequency of note types from different colonies are given in %, (* : % of total number of elements, #: % of motives containing a specific element). The investigated colonies are from Pennsylvania, U.S.A. (Price, 1979), Alice Springs and Murray River, Australia (Zann, 1993; Zann, 1996), Utah, U.S.A. (Leadbeater et al, 2005), Leiden, the Netherlands (Holveck et al, 2008) and the following 4 colonies by Sturdy et al. (Sturdy et al, 1999): JH: John Hopkins, QB: Queen's biology, R: Rockefeller, QP: Queen's psychology.

Statistical details

Preference test 35dph

16 birds were tested, of which 4 did not show a response and had to be excluded. When a bird sat on the same perch during a whole test this was counted as a lack of response and the test was excluded. Several blocks (one of the 4 tests on one day) also had to be excluded due to lack of response. Statistics were performed on the remaining 12 birds with 27 blocks in total (blocks

were excluded when birds showed no response at all).

A linear mixed effect model was performed with time spent on each side of the cage as dependent variable, stimulus type (common/uncommon) as fixed factor, subject as random factor and block nested within subject.

Results for model comparison with (model2) or without (model1) stimulus type are listed below:

	Model	df	AIC	BIC	logLik	Test	L.Ratio	p-value
model1	1	4	740.2076	748.1635	-366.1038			
model2	2	5	735.3988	745.3437	-362.6994	1 vs 2	6.808736	0.0091

Preference test from age 35dph to 55dph

A linear mixed effect model was performed with time spent on each side of the cage as dependent variable, stimulus type (preference test common/uncommon), tutor type (common/uncommon), age (35,45,55dph) and tutor/nontutor as fixed factors, subject as random factor and block nested within subject.

Model comparison for an interaction between stimulus type x tutor type x age x tutor/non-tutor:

	Model	**df**	**AIC**	**BIC**	**logLik**	**Test**	**L.Ratio**	**p-value**
model1	1	25	3444.145	3532.181	-1697.072			
model2	2	27	3441.838	3536.918	-1693.919	1 vs 2	6.306308	0.0427

Age 55

Since a 4-way interaction was found, data were split in order to inspect the 55dph in more detail. A linear mixed effect model was performed with time spent on each side of the cage as dependent variable, stimulus type (common/uncommon) and tutor/nontutor as fixed factors, subject as random factor and block nested within subject.

Model comparison for an interaction between stimulus type x tutor type x tutor/non-tutor:

	Model	**df**	**AIC**	**BIC**	**logLik**	**Test**	**L.Ratio**	**p-value**
model1	1	4	1606.046	1617.06	-799.023			
model2	2	11	1598.321	1628.611	-788.1607	1 vs 2	21.7246	0.0028

We further spilt up the data to see if the difference between tutor song and non-tutor song stimuli was similar for birds tutored with common song and those tutored with uncommon song.

Model comparison for the group tutored with common songs testing for an interaction between stimulus type x tutor/non-tutor:

	Model	**Df**	**AIC**	**BIC**	**logLik**	**Test**	**L.Ratio**	**p-value**
model1	1	4	778.9096	787.011	-385.4548			
model2	2	7	772.4299	786.6074	-379.2149	1 vs 2	12.47972	0.0059

Model comparison for the group tutored with uncommon songs testing for an interaction between stimulus type x tutor/non-tutor:

	Model	**df**	**AIC**	**BIC**	**logLik**	**Test**	**L.Ratio**	**p-value**
model1	1	4	835.0433	843.4207	-413.5216			
model2	2	7	831.7837	846.4441	-408.8919	1 vs 2	9.259555	0.026

4 Zebra finch song phonology and phonological syntax across populations and continents

Sita. M ter Haar[1,2], Caroline A. A. van Heijningen[1,2], Rob F. Lachlan[3] and Clara C. Levelt[2,3], Carel ten Cate[1,2]

[1]Behavioural Biology, Institute of Biology Leiden, Leiden University
Sylvius Laboratory, Sylviusweg 72, P.O. Box 9505, 2300 RA Leiden, The Netherlands

[2]Leiden Institute for Brain and Cognition, Postzone C2-S, P.O. Box 9600, 2300 RC, The Netherlands

[3]Leiden University Centre for Linguistics (LUCL), Leiden University, P.O. box 9515, 2300 RA, Leiden, The Netherlands

[4]Department of Biology, Duke University, P.O. Box 90338 Durham, NC, 27708-0338, USA

CAA van Heijningen and SM ter Haar share first authorship.

Abstract

The zebra finch (*Taeniopygia guttata*) has been a widely used as model-species. Many different neuro- and behavioural biology laboratories maintain a local population of these song birds. In this paper we present an extensive study of the phonology and phonological syntax (restrictions in sequences of elements) from 13 different populations from Europe, North-America and Australia. Our aim was to explore whether, similar to language, in addition to the well-established individual variation in songs there are more or less species-wide 'universals' as well as systematic variation among populations. By using a computational approach, we found 11 element types occurring among all 13 populations. The only commonality in terms of sequences of elements seems to be the position of long flat elements towards the end of the song. Despite these commonalities there is considerable variation between populations in the distribution of element type frequencies. For syntax, 8 out of 13 populations had a significant amount of structural constraints in their songs. However, these constraints were absent on a species-wide and continent level. These results shed light on the apparent variation in song characteristics between populations of a widely used model species for language evolution and vocal learning. These findings may be important for interpretation of results from different populations and other (bird) species used in experimental research. Moreover, they are suggestive of cultural differentiation being at least as important as genetic diversification processes in causing large scale changes in song characteristics.

Introduction

Birdsong and language exhibit many parallels. First of all both communication systems rely on learning. Human infants as well as young songbirds must be exposed to adult vocalizations in order to properly learn the sounds. Second, predispositions for species specific aspects of the sound seem to play a role in the learning process. Third, both song and language learning go through a babbling stage in which auditory feedback shapes vocal development. And as a last parallel, songbirds and humans posses specialized brain mechanisms for vocal production and learning. Together, this evidence suggests that there may be similar mechanisms underlying the learning of birdsong and language (Bolhuis et al, 2010; Doupe & Kuhl, 1999).

Another similarity, and relevant to this paper, is that both songs and language consist of rapid sequences of differently sounding vocal elements, organized according to specific patterns. In this paper we present an extensive study of the phonology and phonological syntax (restrictions in sequences of elements, independent of differences in meaning) of a songbird species, the zebra finch. Our aim is to explore whether, similar to language, underneath the well-established individual variation in songs there are more or less species-wide 'universals' as well as systematic variation among populations.

Studying both universal and non-universal song features may tell us more about the relative roles of experience dependent and independent factors on vocal development and on the way sound patterns are organized. Birds of different populations can have different 'dialects' if birds from the same population have a more similar song (or element repertoire) than birds from neighbouring populations often related to geographic variation. Differences between populations have been found in for instance, white-crowned sparrows (Marler & Tamura, 1964; Petrinovich & Baptista, 1984) and black-capped chickadees (*Poecile atricapillus* (Kroodsma et al, 1999)) showing distinct sound patterns between neighbouring populations, related to the geographical distance and social interaction between individuals. Over generations such populations may diverge even further, changing their vocalizations because of reduced exchange between populations combined with cultural transmission with slight modifications within populations. Cultural transmission processes may, however, also help to conserve sound patterns, and this, combined with biological factors such as perceptual biases and production constraints, may limit the types of patterns that are possible, leading to some song features that may occur in many (but not necessarily all) populations ('universals') whereas others occur in a few.

Zebra finches are an excellent model system for studying how song patterns may start to differentiate between populations, for which parameters, and which mechanisms cause these changes. Since the 1960's, domesticated zebra finches (*Taeniopygia guttata*) have been an increasingly popular model species for neurobiological and behavioral research. Instead of repeatedly acquiring wild individuals directly from Australia to maintain a stock, research laboratories and commercial breeders around the world have successfully bred many generations of domesticated zebra finches. This process thus creates an unprecedented large scale experiment that allows us to examine both the universal song features present across all populations as well as how strongly song features have diverged among populations. Here, we take a first step towards addressing this issue by providing a detailed description of the variation among 13 populations.

The process of domestication had consequences for the genetic and cultural evolution of these (laboratory) 'populations'. The genetic relatedness between 18 of these laboratory populations from North America and Europe and two wild populations from Australia was analyzed by Forstmeier et al. (2007). They found that populations differed significantly according to broad geographic relationships: the most pronounced genetic differences occurred between the three continents. In addition, domesticated birds were found to differ phenotypically from wild zebra finches, being physically larger than wild zebra finches. In conjunction with this genetic differentiation over the years, cultural evolution may have led to differentiation in learned song characteristics. Repeated (and at least partly isolated) breeding at research laboratories might have lead to the formation of dialects comparable to dialects in isolated populations in other bird species. Comparing the genetic analyses with the variation in songs among different populations and continents can provide insight in whether songs in genetically more similar populations are also more similar in song features or not.

Furthermore, it is yet unknown whether the song of domesticated zebra finches still reflects the natural variation of wild populations or whether domestication has lead to differences in song characteristics compared to those of wild birds. For instance, it is known for Bengalese finches (*Lonchura striata* var. *domestica*) that the process of domestication of white-backed Munia's (*Lonchura striata)* originating from China into Bengalese finches by Japanese breeders of (starting approximately 260 years ago) resulted in more variable intra-individual element transitions and hence much more variable songs (Honda & Okanoya, 1999).

In the current study we examine the variation and universal patterns of phonology and phonological syntax in zebra finch song from wild and laboratory populations and whether the differentiation among populations has parallels with the genetic differences found between continents.

Phonology

Many studies on song variability have focused on 'phonology'. By phonology we mean the study of the smallest units of which a song consists: 'elements' or 'notes', which are usually categorized by visual inspection of spectrograms (fig. 4.1 See Method section for an overview of the different categorization schemes in the zebra finch song literature). In our study, we follow the usual convention of studies on song bird phonology (Lachlan et al, 2010; Prather et al, 2009; Sturdy et al, 1999a; Sturdy et al, 1999b). See also Ten Cate , Lachlan & Zuidema, in press.) by focusing on the signal itself instead of the phonological/categorical perception that has been demonstrated for some species (Nelson & Marler, 1989). Phonological variation between local populations in the type of elements has been shown in several species, for instance in white crowned sparrows (Nelson, 1998). For the zebra finch, some comparative element type studies have been done (Lachlan et al, 2010; Slater & Clayton, 1991; Sturdy et al, 1999b; Zann, 1993a; Zann, 1993b; Zann, 1996). These studies were limited in scope, however. In the present study we examine the presence and distribution of element types on a much wider scale and rather than relying on visual inspection and categorization, we use an advanced computerized element categorization algorithm.

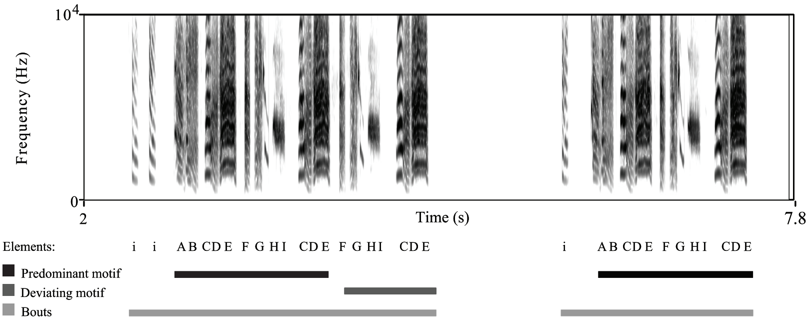


Fig. 4.1. Example of a zebra finch song. A song consists of bouts (light grey) of which the predominant motif (black) is selected for further analyses. The predominant motif is the version occurring most often in an individual's repertoire, compared to a less often occurring one like the deviating motif presented here (medium grey). Introductory elements (i) were not taken into account in this study.

Phonological syntax

Syntax in the context of bird research can be defined in general as restrictions in the sequencing or position of structural units (e.g Lachlan et al., 2010). The degree of variability in the sequence of song elements differs greatly between bird species. For example, Chipping sparrows (*Spizella passerina*) have a single song, consisting of one repeated element (Albrecht & Oring, 1995) while on the other side of the complexity spectrum, the nightingale is a wide known for it's relatively high individual sequence variability (Todt & Geberzahn, 2003; Todt & Hultsch, 1998). Various methods have been employed to describe and analyze the patterns in variability, such as transition diagrams (Honda & Okanoya, 1999) and (higher order) Markov Models (Hailman & Ficken, 1986; Lemon & Chatfiel.C, 1973); see also ten Cate and Okanoya (2012) for an overview). But most such methods have been concerned with investigating variability within individual repertoires, rather than variation between different populations.

In zebra finches, individual males were previously thought to sing in bouts of largely invariant motifs (Glaze & Troyer, 2006; Immelmann, 1969; Zann, 1996). But the current general opinion is that although males sing relatively stereotyped songs, they can vary on this pattern by adding, deleting or sometimes altering elements in the motif (Helekar et al, 2003; Helekar et al, 2000; Scharff & Nottebohm, 1991; Sossinka & Bohner, 1980; Sturdy et al, 1999b; Volman & Khanna, 1995). Additionally, Helekar et al., (2003; 2000) found relatively many repeated elements in songs from their lab population, a phenomenon that has been observed in isolate song (raised without a tutor) as well (Helekar et al, 2000; Jones & Slater, 1996).

On the inter-individual (or population) level, there is far more variability, more specifically in where element types can occur in the song. Lachlan et al., (2010) found high variability in the syntactic patterns as all element types can be followed by each other. This is in contrast to an earlier study by Zann (1993 a, b & 1996), who found patterns for element types relative to the position in the song (start, middle or end). Zann suggested that song elements are not equally distributed over the song: the first section consists mostly of introductory elements, the middle section consists of a 'high' element and 'stacks' and 'distance calls' occur more towards the end of the song (Zann, 1996). Thus, previous studies have shown that both variation and common patterns can be found in zebra finch song.

In the present study we provide an extensive species-wide analysis including both phonology and phonological syntax on data from 12 laboratory populations and one wild population across three continents (populations grouped according to geographical proximity. Therefore, we grouped the Auckland population from New Zealand with the wild population sample from Australia). Computational bioacoustic analyses are performed on zebra finch song to answer the following questions for both phonology and syntax:

1) What are the species-wide element types and are they present in all populations (i.e. are they universal) or do populations differ in this?
2) What is the degree of species-wide phonological syntax and how consistent is this across populations (i.e. are they universal) or do populations differ in this?
3) Are there differences between continents and how do these relate to population differences?

Methods

Song analysis by computer software: Luscinia

Zebra finch song is known for it's relatively difficult segmentation and annotation in element types and/or syllables. Except for Lachlan et al., (2010) all previously mentioned articles have been based on visual inspection for assessing element categorization. Even though the use of several observers can reduce the variability of visual classification (Jones et al, 2001), such studies are limited in the sample size of songs they can analyze, and the degree to which quantitative measures of inter-element similarity can be provided (as opposed to same/different binary scoring). It is therefore difficult to statistically address questions of phonological and syntactical organization without computational methods.

Using the software package Luscinia (http://luscinia.sourceforge.net) for acoustic and statistical analysis Lachlan et al., (2010) compared zebra finch song from four laboratory populations as part of a larger comparison between several songbird species. Because we had relatively few suitable songs from Australian regions (Alice Springs and Northern Victoria) we decided to merge these to one population. Moreover, because of the high level of dispersion in wild zebra finches, there are no clear dialects found in wild zebra finch song (Zann, 1996), nor large genetic differences between wild populations (Forstmeier et al, 2007). See table 4.1 for a list of all populations in this analysis.

Sound Analysis in Luscinia

Recently, several software packages (e.g. Sound Analysis Pro, Tschernikovski et al., 2000, Luscinia, Lachlan et al., 2010) have been developed that allow the acoustic analysis of large datasets of birdsongs. By using computational algorithms to compare songs, quantitative comparisons between large numbers of elements can be made, with little or no observer bias. In Luscinia, measures of several acoustic parameters are taken at each spectrogram time slice throughout each element. Luscinia then uses an implementation of Dynamic Time Warping (DTW) to compare elements based on these time-series vectors. (See the SI for the DTW parameters chosen for this study). The output of this comparison is a matrix of the dissimilarities between each pair of elements in the dataset. This matrix can be ordinated into Euclidean space using nonmetric multidimensional scaling. The resulting principal components can then serve as the basis for cluster analyses in statistical software packages like R (and open-source package for statistical analysis available from www.r-project.org). With this species-wide clustering solution it is possible to identify element clusters specific to populations or universal to all populations.

Motif selection

As shown in fig 2.1, zebra finch songs consist of motifs, preceded by a number of introductory elements. Repeated motifs are called a song bout. Motifs can vary slightly within birds. We selected the predominant motif of each bird for analysis (i.e. the motif with the highest probability for that bird (Helekar et al, 2000; Sturdy et al, 1999a). Introduction notes were excluded unless they recurred in motifs later in the bout. Motifs were selected by two observers (StH and CvH). Each observer selected half of the motifs for each population reducing the possible effect of observer bias. For a subset of the songs both observers selected the motif in order to test reliability. Since this reliability between the observers was relatively low (0.60 (see SI for more details)), the songs that were ambiguous were re-examined by both observers and the motif for which consensus was reached was selected.

Element segmentation

Throughout the years of zebra finch song research, multiple methods of element segmentation have been published (Lachlan et al, 2010; Scharff & Nottebohm, 1991; Sturdy et al, 1999a; Sturdy et al, 1999b; Zann, 1993a; Zann, 1993b). These are not all in agreement. In the present study we segmented motives into elements based on the criteria of silent gaps and/or abrupt changes in frequency and amplitude. The segmentation was performed manually by StH and CvH, using Luscinia, after several years of experience with zebra finch song analysis and multiple training rounds on other songs. Each observer performed element segmentation on half of the motifs from each population reducing the effect of observer bias. For an estimate of reliability a subset of the data was segmented by both observers (see SI for more details). Average agreement between observers was 86 %.

Cluster analysis

A *cluster analysis* was performed with Mclust (package 'mclust' was built under R version 2.12.2) (Fraley & Raftery, 2002) within R (version 2.13.0, available at http://www.r-project.org) on the output from Luscinia (the first ten PCO values which explained more than 92% of the variation in the data set). Mclust clusters data using Gaussian Mixture Modeling. This approach allows for clusters of varying size and shape to be discovered (Fraley & Raftery, 1998). Mclust then uses the Bayesian Information Criterion approach to determine which model best fits the data, allowing a statistical assessment of the number of clusters present.

Having placed elements into clusters, further statistical phonological analyses were performed on the zebra finch songs annotated according to the element types calculated by Mclust.

Relative position

Some element types may be restricted to certain positions in the song. For instance the songs of chaffinches (*Fringilla coelebs*) show relatively strict patterns (Slater & Ince, 1979). The relative position of elements in a motive was estimated by ranking each element according to its position in the song (low numbers for elements early in the song and higher numbers for those at the end). Because songs differ in total number of elements, a relative measure of position was calculated by dividing the position number by the total number of elements per song. For example: the 8th element in a 10-element song has a relative position of '0.8'. A relative position was determined for each element in each song. The mean relative position was calculated for each cluster of elements (as classified by the Mclust algorithm in R). This thus reveals if there is a higher chance for certain types of elements to in a specific region of the song. This does not mean that the elements are restricted to that area and cannot occur in other areas, but rather that a certain area (the last part of the song for instance) contains more elements of a certain type (for instance cluster 1) than another area.

Species-wide syntax analysis
The amount of structure, or syntax, was measured with a Sliding Window Match Length (SWML) entropy estimator that analyses the sequences of elements in the zebra finch songs (Suzuki et al, 2006). 'Entropy', in this context, is a measure developed in the field of information theory to describe the unpredictability in a signal. This theory has been applied before to find patterns in animal vocalizations, for instance by Lemon and Chatfield (1973) in rose-breasted grosbeaks (*Pheucticus ludovicianus*), by Suzuki et al., (2006) in humpback whales (*Megaptera noveangliae*) and by Briefer at al (2010) in skylarks (*Alauda arvensis*).

The higher the information entropy, the lower the degree of structure or constraints in the data is. For instance, a random sequence of different element transitions results in a very high entropy value. In contrast, messages with lexicographical, grammatical and contextual constraints result in a lower entropy value as not all combinations of elements are present in the message (Suzuki et al, 2006).

One limitation with entropy-based measures of syntax is that they require constituent elements to be placed into discrete categories. Different categorizations of elements may result in very different entropy estimates. A solution to this problem was proposed by Lachlan et al., (in prep). This method searches for the partitioning of elements that minimizes entropy (that is maximizes syntax). There are two steps in this method. In the first, elements are assigned to clusters based on their acoustic similarity (as determined from the DTW comparison, above) to a predetermined number of *k* cluster centroids. In the second step, entropy is estimated using the first order Markov estimator for the clustered data. The algorithm proceeds by searching for centroids that minimize the entropy estimator in a similar way to the k-medoids algorithm (from which this algorithm is adapted).

Once the partitioning with most structure had been identified, we recalculated entropy for that partitioning using the SWML estimator. The benefits of this type of entropy estimator are that it is applicable to a relatively broad range of data, as it has less a-prior data assumptions than other syntax inducing algorithms. Also, it is relatively robust to departures from these assumptions. And lastly, even if the sample sequence is not long enough to achieve asymptotic convergence, the resulting entropy estimates are an upper bound on the source entropy (Suzuki et al, 2006). Therefore, even with relatively small sample sizes this method can estimate the amount of structure present in zebra finch song reliably. It is, however, too slow to use during the clustering part of the algorithm (above). We set the window length for this algorithm to be the length of the individual motifs. We then rescaled entropy as 'redundancy', calculated as $(H_{baseline}-H_{swml})/H_{baseline}$. We used $H_{baseline}$, the entropy calculated from randomly sequenced sequences instead of the maximum entropy (H_{max}) that Suzuki et al. (2006) used as $H_{baseline}$ turned out to be a more stable parameter. Our redundancy estimates then served as our measures of syntactic structure. Redundancy is a way of scaling entropy were '0' means that element transitions are random (like randomly shuffling the sequences of elements in each song) and '1' means that sequences are entirely predictable and non-variable. We estimated confidence limits around our estimates using a bootstrapping technique (Suzuki et al, 2006). We estimated redundancy for values of *k* between 2 and 10 and selected the highest redundancy value as our estimate of syntactical structure for the particular level of analysis (species-wide, per laboratory or per continent).

Population level syntax analysis

The individual redundancy scores per population were calculated and bootstrapped to arrive at confidence intervals around these scores to indicate presence or absence of structural constraints, or redundancy (ρ) per population.

The redundancy analysis returns one value per population, so in order to compare the populations with each other for in their degree of syntactical structure, a deviation scores for each individual in each population was calculated and analyzed in an ANOVA. The deviation (calculated from the SWML distances, converted to PCO values) is a measure for how different song syntax is between one song sequence and it's population average (centroid), an adaptation of the method described for variation tests by Anderson (2006) and Anderson et al., (2006). If songs are on average further away from the centroid, they also on average vary more from each other, and hence have a higher entropy (and are therefore less structured or constrained in terms of syntax).

Continent level syntax analysis

To calculate the amount of syntax in each continent, pairwise ρ scores for within-continent populations were calculated instead of ρ scores per continent. The latter scores are less suitable in this context because the Australian continent contained data of only two populations (one wild population from Australia and 1 laboratory population from Auckland, New Zealand). The other populations contain 6 (North America) and 5 populations (Europe) and the ρ is influenced by the number of populations in the comparison. We therefore calculated pairwise ρ scores between populations within continents to estimate the amount of syntax within continents and bootstrapped the results to calculate confidence intervals. We then averaged over the pairs within continents to arrive at an average and confidence interval per continent.

Results

Phonology

To cluster all elements according to acoustic parameters we measured them in Luscinia

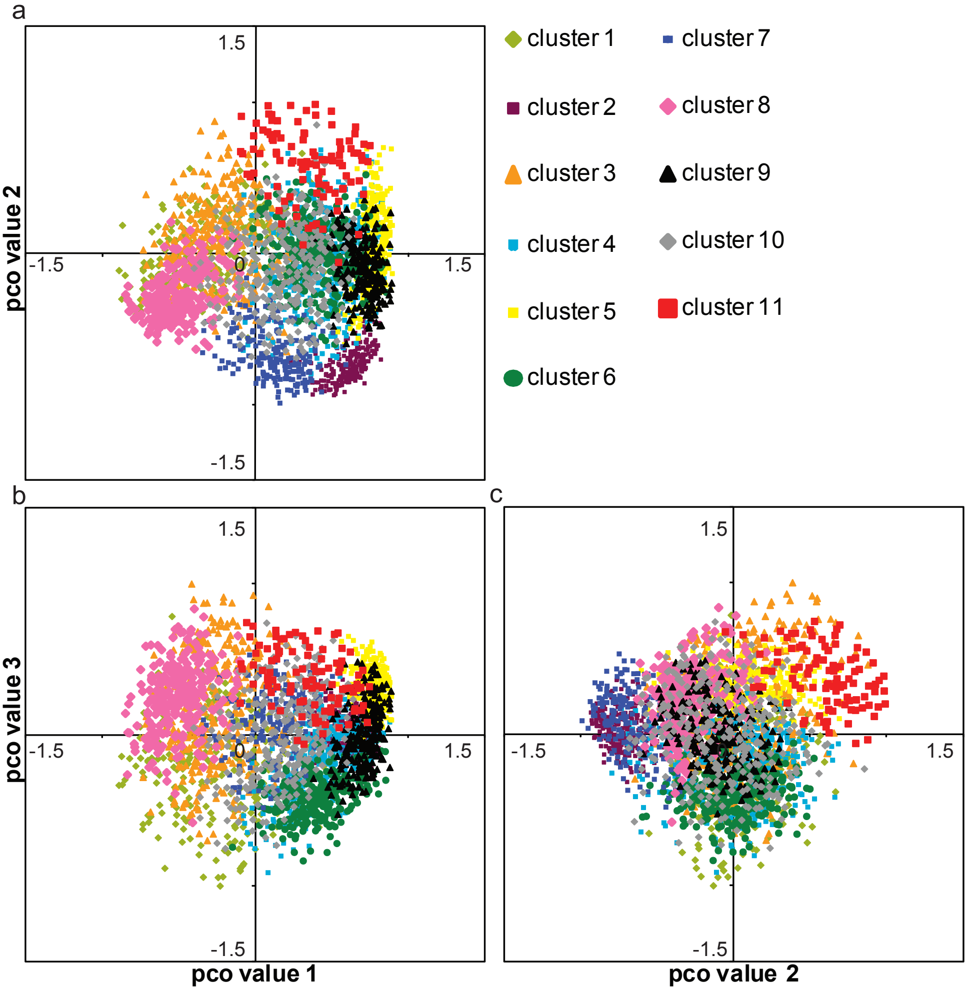


Fig. 4.2. Clustering graph 11 clusters all populations together. A) PCO value 1 vs 2. B) PCO value 1 vs 3. C) PCO value 2 vs 3. Each data point is an element. Different colors (and shapes) represent the different element types (clusters as modeled by Mclust).

and performed a DTW analysis which was followed by an NMDS and subsequent PCO analysis based on 10 components. These 10 components explained 92% of the variation (Kruskal stress value of 0.051, indicating a good fit (Ozturk et al, 2009).

The Mclust algorithm calculated an optimal model of 11 clusters (see fig. 4.2a, b and c) in a VEV configuration (Ellipsoidal, variable volumed, equal shaped and variably oriented) and a BIC value of 8821.2. The next best fitting models had BIC values of 8820.8 with 14 clusters and 8777.1 with 13 clusters respectively. We chose the 11-clus-

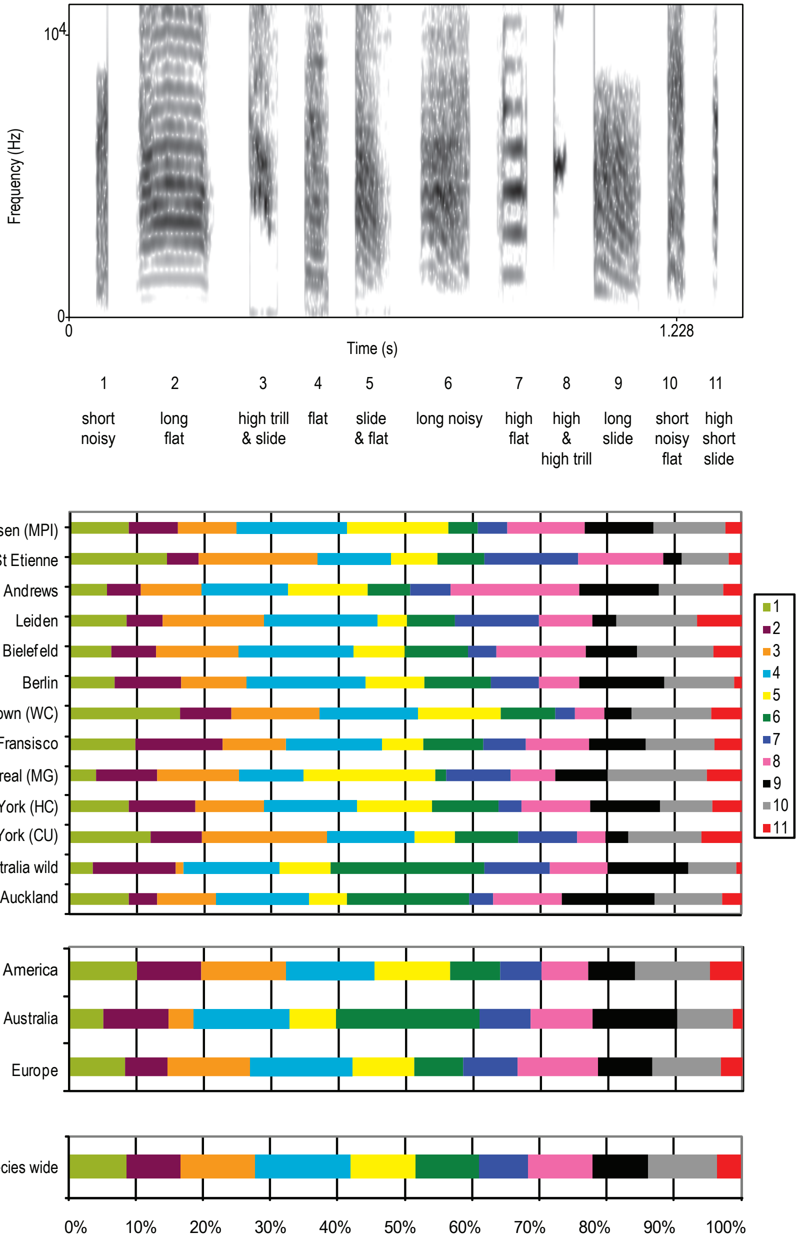


Fig. 4.3. Element types and their distributions per population, continent and species-wide level. Top panel provides representative examples of the element types found in the Mclust clustering analysis. Numbers indicate individual element types (clusters). Lower panels show element type distributions in percentages per population, continent and species-wide.

ter model as this one had the highest BIC and it is the most parsimonious model where the number of clusters is concerned. To estimate the overall tendency for the data to cluster, we compared the one cluster solution BIC value (-2283.4) with the value Mclust gives for the optimal clustering model. The large difference between the two indicates a very high propensity of the data to fall into clusters.

To give an example for each of our clusters, we selected the first three elements from the densest area of each cluster to characterize the element type. If the first three elements would not give a conclusive characterization, we analyzed four elements (in the case of cluster 3 and 8). For labeling of the element clusters, see fig. 4.3.

The distribution of the 11 clusters per laboratory, per continent and species-wide is shown in fig. 4.3. All clusters occur in every population, indicating that all element types occur in all populations and therefore that there are no population-specific element types. Yet, we found that the elements from the different clusters are not distributed evenly over the populations using a X^2 permutation test in which individuals were permuted between populations (with 10000 permutations, $X^2 = 441.8$ $p < 0.001$, Cramér's effect size 0.13). We also found that continents differed in element type distribution by X^2 permutation test in which individuals were permuted between continents (with 10000 permutations, $X^2=162.9$, $p <0.01$, Cramér's effect size 0.17).

Number of elements per song

There was variation in the number of elements per song between the different populations, see fig. 4.4. (F=2.344, deletion p = 0.008, based on 188 df, with a glm model with quasi-poisson link). However, when aggregated over continents, there were no differences in element number per song (F=1.736, deletion p > 0.05, based on 188 df, with a glm model with quasi-poisson link).

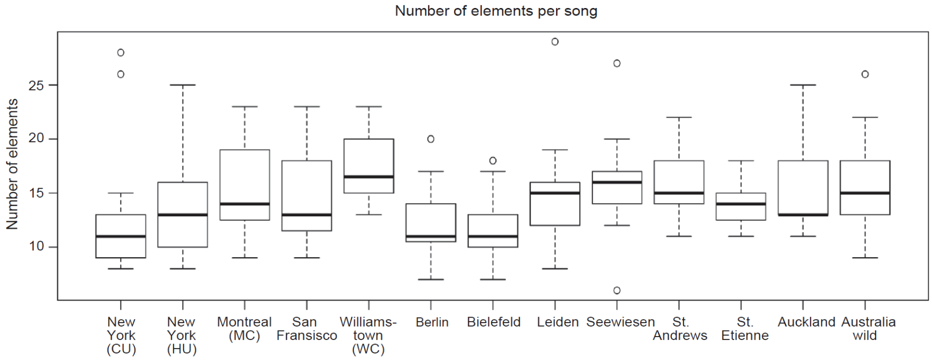


Fig. 4.4. Boxplot distribution of the number of elements per song, per population.

Relative position of elements within a song

The mean relative position of each of the 11 element types is presented in fig. 4.5. The overall Kruskal-Wallis test was significant (K-W X^2=166.84, df=10, p <0.001 and therefore the element types are not occurring evenly over the positions in the song. Further testing showed that on a species-wide level clusters 2 (two-sample Wilcoxon test, W=158762, p <0.001, and 7 (two-sample Wilcoxon test W=198835, p <0.001) occurred in the end region of the song (see fig. 4.5). (All tests were still significant after Bonferroni correction for multiple testing). Roughly the same distribution pattern can be observed on continent level: For all continents, the overall K-W tests were significant: X^2= 32.4, df = 10, p < 0.001 (Aus), K-W test X^2= 125.3, df = 10, p < 0.001 (N-Am); K-W test X^2= 52.8, df = 10, p < 0.001. After post-hoc tests, we found similar patterns compared to the species-wide pattern, except for Australia, where cluster 9 was also occurred relatively to the end (K-W X^2= 6847, p = 0.006, still significant after Bonferroni correction) and cluster 7 was not.

Phonological syntax

Species-wide level syntax

After analyzing the songs with the k-medoids/Markov chain modeling and SWML estimator algorithm in Luscinia, we found a mean redundancy (ρ, amount of structural constraints) that did not significantly differ from 0 (-0.016 with maximum redundancy at k= 8 syntactical clusters CI 95%: -0.065 to 0.030). This implies that there is no evidence for syntactic structure, in terms of sequences of specific element types, on a species-wide level in zebra finch song.

Per population syntax

When the data from each individual population were analyzed separately for the amount of structure (redundancy), we found that the bootstrapped estimate of ρ, averaged over all populations was 0.109. This is higher than when all populations were analyzed together on the species-wide level. For eight of the populations, there was significant evidence of syntactic constraints (see table 4.1 with pop mean ρ's, k=nr of clusters and bootstrap p value). Estimates of ρ were not influenced by sample size in any detectable way when comparing between populations, indicating that these sample sizes were adequate (see fig. SI 4.1).

We compared the syntactical structuring between the different populations by calculating the deviations, but there were no significant differences found (ANOVA, df 188, p-value 0.19), so there were no significant differences between the lab populations in the amount and variability of syntax in the songs of that population.

Table 4.1. List of populations with amount of syntax scores (rho).

lab population	continent*	sample size	rho	highest k	bts p-value	sign
ALL	-	189	-0,016	8	p> 0.5	
Auckland	Australia	9	0,108	4	p<0.025	*
Bielefeld	Europe	18	0,082	4	p>0.025	
Berlin	Europe	15	0,169	6	p<0.0025	*
New York (CU)	N. America	14	0,018	3	p>0.025	
New York (HC)	N. America	15	0,146	10	p<0.025	*
Leiden	Europe	15	0,050	6	p>0.025	
Australia wild	Australia	17	0,128	4	p<0.025	*
Montreal (MG)	N. America	15	0,121	10	p<0.025	*
St Andrews	Europe	14	0,102	9	p>0.025	
St Etienne	Europe	15	0,138	2	p<0.025	*
San Fransisco	N. America	15	0,178	2	p<0.0025	*
Seewiesen (MPI)	Europe	13	0,061	7	p>0.025	
Williamstown (WC)	N. America	14	0,096	7	p<0.0025	*

* 'continent' refers to grouped in terms of geographical distances. CU=Columbia University,

HC=Hunter College, MG=McGill University, MPI=Max Plank Institute for Ornithology, WC=Williams College. Bts: bootstrap

Per continent syntax

Estimating continent-wide syntactical structure was complicated by the relatively low number of Australian populations (two: 1 wild, 1 lab from Auckland) compared to the others (five and six). Based on the population level analysis (above), this would be expected to artificially boost the estimate from Australia compared to the other continents. To avoid this effect we therefore made pairwise comparisons between all possible pairs populations within each continent and calculated deviation scores. The mean deviation values averaged over the pairs of populations in these continents were: 0.047 (North America, with CI 95%: -0.053 to 0.140) 0.03 (Australia with CI 95%;-0.012 to 0.175) and 0.036 (Europe with CI 95%: -0.052 to 0.122). However, none of these scores were significantly different from 0, indicating that there is no evidence for structure on a continent level. Therefore there is also no variation between continents.

General discussion

Our computational analysis of zebra finch song from 13 populations indicates that in terms of phonology, there is evidence for clustering of elements into 11 types on a species-wide scale. All element types occur in all populations but the distribution of element types differs significantly between populations as well as continents. In terms of syntax, or structural constraints found in the sequences of elements in zebra finch song, there is variation in the amount of syntax on population level, but these effects are absent when looking at continent and species-wide scale (all results from this study are summarized in table 4.2).

Table 4.2. Results overview

	phonology	**phonological syntax**		**relative position**	**tot nr of elements**
level of analysis	**element types**	**amount of syntax (ρ)**	**deviation**		
species-wide	11 clusters	0	-	2 and 7 in end region	-
continent	variation in distribution of these 11 clusters	0	-	2 (Aus,Eur, N-Am), 7 (Eur, N-Am), 9 (Aus) in end region	no variation
population	variation in distribution of these 11 clusters	8/13>0	not sign	-	variation

Phonology

The 11 element categories as obtained by the clustering algorithm analysis of all 13 zebra finch populations together is different from earlier findings using a comparable analysis, but data from three populations only. This resulted in just two categories; high notes and all low frequency elements together (Lachlan et al, 2010). The difference may be explained by the number of populations used. The difference in clustering method however, may also account for some of the differences, since the Mclust method used here also takes into account different cluster shapes and sizes, whereas the simple agglomerative clustering method used in the previous study assumed only perfect spherical clusters. Other studies have described different note types as well, using visual inspection of spectrograms of laboratory birds' songs (Price, 1979; Sturdy et al, 1999b; Williams & Staples, 1992), or from wild individuals only (Zann, 1993a; Zann, 1993b). Zann reported 14 element types, Sturdy 5 types and Price also 5 types. There is some overlap in categories described previously and the ones found in the present study. For instance high notes and slides are describes as separate categories by Zann (1993a and b) and Sturdy et al. (1999b). This is also valid for the present study, however multiple categories of slides and stacks are found. Sturdy et al.'s study including four laboratory populations also found differences in distribution of element types between populations but every category was found in all populations (Sturdy et al, 1999b). This seemingly corroborates our present findings. It should be noted however, that the classification by Sturdy at al. was aimed at providing generalizable information about zebra finch song. Thus Sturdy et al. probably deliberately classified elements (or note types) into more universally applicable categories. Differences in classification methods between studies make a more direct comparison with the present results difficult. For instance some of the previous authors used silence as a segmentation criterion, and therefore their classification includes larger units such as combination elements. In our study these combination elements are segmented into multiple elements as we also included abrupt changes in frequency as a segmentation criterion.

This brings us to an important point of consideration for interpreting the current findings: the analysis is directed at the level of elements instead of syllables, since we also use abrupt frequency change as a criterion for segmentation. At this point it is not entirely clear what level of representation is that birds use themselves while processing songs. Ten Cate and Slater (1991) as well as Williams and Staples (1992) reported that songs are copied mostly as chunks separated by silent intervals, which may include several elements and/or syllables. The same studies, however also demonstrated that elements and syllables within chunks could also be copied separately, indicating a more detailed representation. Therefore, in the present study we chose an analysis on the element level.

None of the 11 element clusters found in the present study is specific to any of the populations (see fig. 2.3) and specifically each cluster appeared in each population. The species-wide distribution of element clusters does show significant variation between populations, however. For instance, cluster 11 seems more consistent in presence among populations compared to cluster 3, which appears to show much more variation in abundance. Although perception and production might go in parallel, it is yet unknown whether the zebra finches would perceive the same 11 clusters we found as separate perceptual categories. Since distribution of categories is different between populations in the present study, it is also possible that category boundaries are different across populations (Prather et al, 2009). This might be the case for zebra finches as well and therefore the clustering itself might differ between populations. Using clustering methods similar to those in the present study but rather on population level, could give insight in where these boundaries may lay (however for the present study, this was not possible due to sample size constraints). But ultimately, perceptual experiments with zebra finches are required to support the conclusions of computational and statistical studies like this one.

Variation in distribution of element clusters is also found when populations are grouped per continent, implying that there is some continent specific phonology (see fig. 2.3). The effect sizes from both the population and continent analysis are rather similar (0.13 and 0.17 respectively). The study by Forstmeier et al. demonstrates genetic differentiation between zebra finches in different continents, but nevertheless considerable variation within continents exists (Forstmeier et al, 2007). When we compare our phonology results to the genetic relatedness in the unrooted dendrogram from Forstmeier et al, (2007), we found a slightly different pattern, namely that the variation between continents was more or less equal to the variation between populations. In the Forstmeier study (2007) instead, continent variation was larger than population variation. Nonetheless, the roughly equal variation in song elements between populations and continents suggests that cultural evolution might have an equally important role in the diversity of birdsong. To analyze this in more detail, one would also need information on the cross-laboratory exchange history of birds. On the other hand, the fact that similar element types exist among populations that have been separated for several generations suggests that the cultural variation might to a certain extent be constrained by perceptual or learning biases. That certain constraints shaping the song into a certain direction are present, has also been indicated by an intriguing experiment by Feher et al., (2009). They used birds that were raised without exposure to song, and therefore developed aberrant isolate song and used these birds as tutors for the next generation. After approximately three generations the song features were already similar to wild-type song. Moreover, a recent study in our lab suggests that biases for acoustic features common to zebra finch song may be present in juvenile birds before they learn song (Chapter 3, this thesis).

Total number of elements per song

There is significant variation in the total number of elements per song between the different lab populations (see fig. 2.4), but this was not the case when aggregated over continents. Slater and Clayton (1991) did not find such a difference when analyzing three populations. However, they analyzed the number of *different* elements per individual song. It would therefore be interesting to re-analyze the total number of element data according to element type in a follow-up study. It is important though to keep in mind the difference between within song element differences (Slater & Clayton, 1991) and analysis based on element *types*: two elements in a song may belong to the same basic type, but still be noticeably different. We also did not find clear evidence for the occurrence of repeats, as was found by Helekar (2000 and 2003). It would be interesting to further analyze the dataset for within song immediate repeats and subsequently whether these can be linked to specific element types.

Relative position of element types in song

First, we examined whether there are associations between element type and relative position in the song. When all lab populations are averaged, the element type from cluster 2 (long flat), and to a lesser extend also cluster 7 (short noisy flat), are positioned towards the end of the song (see fig. 4.5). The same pattern can be observed for North America and Europe. Interestingly, in Australia, also cluster 9 (and not cluster 7) has a tendency to occur in the end region of the song (see fig. SI 4.2). An explanation why this did not lead to a higher ρ in the redundancy syntax analysis, might be that the long flat element do not necessarily occur only exactly as the last element of the song, but on average in the end region (mean position 0.7) and can be flanked by any type of other element. The link between long flat notes and the end region of the song has been observed by other researchers as well (Zann 1993 a, b and 1996).

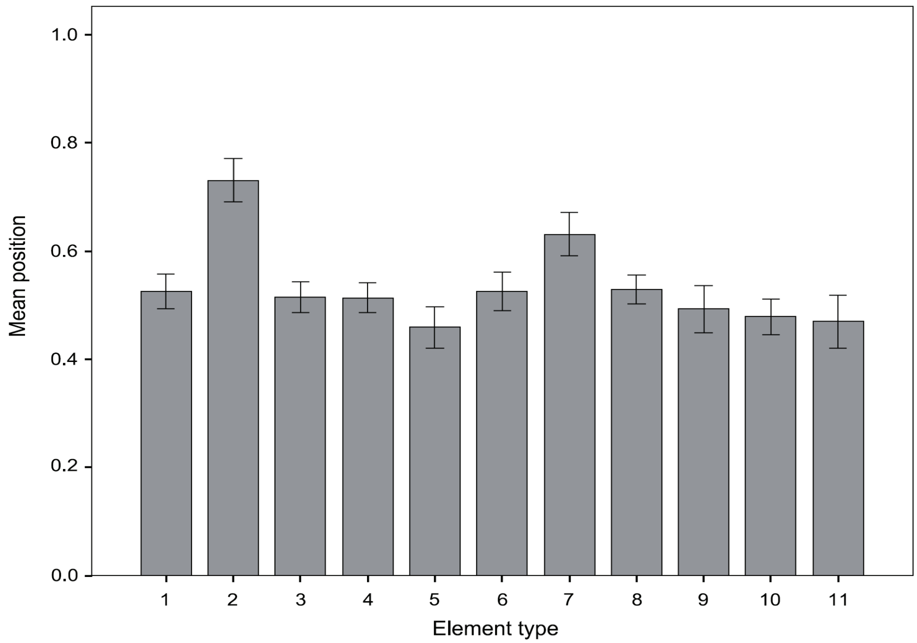


Fig. 4.5. Relative element position on species-wide level. A mean position of 0.5 indicates that the element type does not particularly occur at the beginning or end of a song. A mean position of 0.7 means that that element type occurs relatively often at in towards the end of the song. Error bars represent 95% CI.

Phonological syntax

Species-wide, continent and population level syntax

On a species-wide and continent level, we did not find evidence for species-wide constraints in the sequences of zebra finch song. The estimates of redundancy found for zebra finches in this study were, for example, much lower than a recent study of chaffinches, which found population-level redundancy values of approximately 0.5 (Lachlan et al., in prep). Lachlan et al., (2010) found no evidence for clear structural constraints based on data combined from three lab populations (from different continents), similar to our results.

Nevertheless, on population level we did find evidence for the presence of syntactical constraints in 8 of the 13 populations. So in a number of populations we found restrictions in what kind of song sequences the local zebra finches sing, while in others there is no evidence for such restrictions or syntax.

At first sight this might seem to conflict with the finding that on a species-wide scale, we found no evidence for such constraints in zebra finch song. This is not due to small sample sizes, which potentially might affect the SWML estimator (Suzuki et al., 2006). As shown in fig. SI 4.1, the estimates of ρ were not influenced by sample size in any detectable way when comparing between populations. On the other hand, we found a significant difference in the deviations when aggregated for all populations compared to the individual deviations calculated for each population separately (paired t test, t= 13.35, p <0.001). Therefore, although there was syntactical structure within populations, it varied between populations. Thus when the populations were aggregated and analyzed together, there was no evidence for species-wide structure.

The most likely explanation is that there might be an effect of tutor sharing leading to frequently occurring sequences in particular populations. Zebra finches can incorporate chunks (short sequences of elements) from their tutor(s) in their song (ten Cate and Slater 1991) and can also from learn other males nearby (Williams and Staples 1992). The most obvious way that two individuals could share elements was if they were siblings. Although the aim was to keep the number of brothers in the laboratory datasets as minimal as possible, this was not always possible to avoid, especially when birds are allowed to breed in aviaries and the paternity cannot be established easily. We could not control whether males shared ancestors in earlier generations, such as grandfathers. Nevertheless, such a process would resemble the formation of dialects known from other song birds and human literature (Marler & Tamura, 1964 and Tserdanalis & Wong, 2004 respectively).

What is important to keep in mind here is that is hard to quantify syntactical structure; different methods make different assumptions, and consequently capture different aspects of syntactical structure. It is possible that we missed patterns on a higher level such as syllable- and chunk sequences and nonadjacent dependencies. Furthermore, as a first effort, we chose to include one motif per individual, but it might be interesting to take into account the intra-individual variation in future studies as well.

Domestication

The differences we have observed in the relative frequencies of element types and syntax between different populations might be in part an effect of domestication. Effects of domestication in songbirds have been observed in a closely related species, the Bengalese finch (Honda & Okanoya, 1999) that has had a longer history of domestication (approximately 260 years, compared to approximately 150 years in zebra finches (Rogers, 1979; Sossinka, 1970). It has been shown that song of domesticated Bengalese finches is less constrained and has different acoustic characteristics compared to those of wild strain birds (Honda & Okanoya, 1999). It might be that after a longer period of domestication, zebra finch populations show more phonological diversity than they do now and that the commonalities are partly a result of cultural transmission as such. However, the history of zebra finch domestication, if any, longer than 150 years ago is mostly unknown (Sossinka, 1970) so we cannot be certain if this comparison is realistic.

Domestication effects can be caused by different factors, such as founder/common tutor effects in small populations (more random drift effects) or human/breeder selection choices such as the number and size of generations and whether specific traits were selected for, such as bird size and clutch size. Such effects probably differ between populations. In case of aviary breeding female choice for specific male (song) traits may play a role (see Riebel 2009) for a review on female preference in zebra finches). Also for disentangling the effects of domestication or drift, more should be known about the breeding history of several populations.

Taken together, this may have consequences for how representative data from one population is for zebra finches in general, since different research laboratories may not always be able to replicate each other's results. Domestication may have different effects on different populations, depending on the size of populations; influx of new birds from other populations and breeding schedules. It remains difficult to disentangle the relative effects between cultural transmission, biological evolution and domestication on the evolution of song. Nevertheless, the methods in the current study present a useful set of tools to investigate such issues in more detail.

Comparison with human language

Songbirds are often used as a model for studying vocal learning and language or speech. The current findings may offer some insight in how biological and cultural evolution may relate to difference in sound patterns across populations worldwide. A direct comparison with language variation among populations of humans is difficult. In terms of phonology usually one language or dialect has a different phoneme inventory then another, but all individuals within a population make use of the same inventory. This is different in zebra finch populations first of all because individual differences between birds in element use are abundant (thus only a subset of the elements of 11 clusters, are present in all individual birds, but all element clusters are present in all populations). Second, the present findings suggest that all phoneme categories are present in all populations. However the different distribution of elements may still tell something about cultural evolution, since some categories seem more flexible than others. For instance studies on language change suggest that some phonemes are more prone to change than others, although the cause of such change is a subject of debate (Moreton, 2008). Additionally, language and speech research shows that some aspects of languages are (initially) universal in both perception and production, indicating biological constraints on the possible variation of sounds (Jakobson, 1941). It would be interesting to see in a longitudinal study whether the more flexible element categories found in the present study indeed show more change over time than the less flexible ones.

Conclusion

By using a computational approach, we found 11 element types occurring among all 13 populations. The only commonality in terms of sequences of elements seems to be the position of long flat elements towards the end of the song. Despite these commonalities there is considerable variation between populations in the distribution of element type frequencies. For syntax, 8 out of 13 populations had a significant amount of structural constraints in their songs. However, these constraints were absent on a species-wide and continent level. These results shed light on the apparent variation in song characteristics between populations of a widely used model species for language evolution and vocal learning. These findings may be important for interpretation of results from different populations and other (bird) species used in experimental research. Moreover, they are suggestive of cultural differentiation being at least as important as genetic diversification processes in causing large scale changes in song characteristics.

Acknowledgements
We would like to thank Peter Snelderwaard for help with the software and database installation.

SUPPLEMENTARY INFORMATION

Luscinia DTW Settings list

Settings and parameters we included in the DTW analysis were: weight by amplitude; log transform frequencies; time, fundamental frequency, fundamental frequency change and harmonicity; compression factor 0.25; SD ratio 0.5.

Motif Selection Reliability

For reliability estimate, a subset of the motifs was selected by two observers. Percentage of motifs that were the same for both observers was used as a measure of reliability and is reported in the manuscript

Element Selection Reliability

For reliability a subset of the data was segmented by both observers. Agreement was calculated as the percentages of elements that were segmented similarly by the two observers. Since the total amount of elements was slightly different for each observer, two percentages were calculated; one for each observers' total number of elements per motive (83 % and 89 %). The average of these two numbers was used as average percentage agreement and is reported in the manuscript

Fig. SI 4.1. No relationship between population size and estimated ρ.

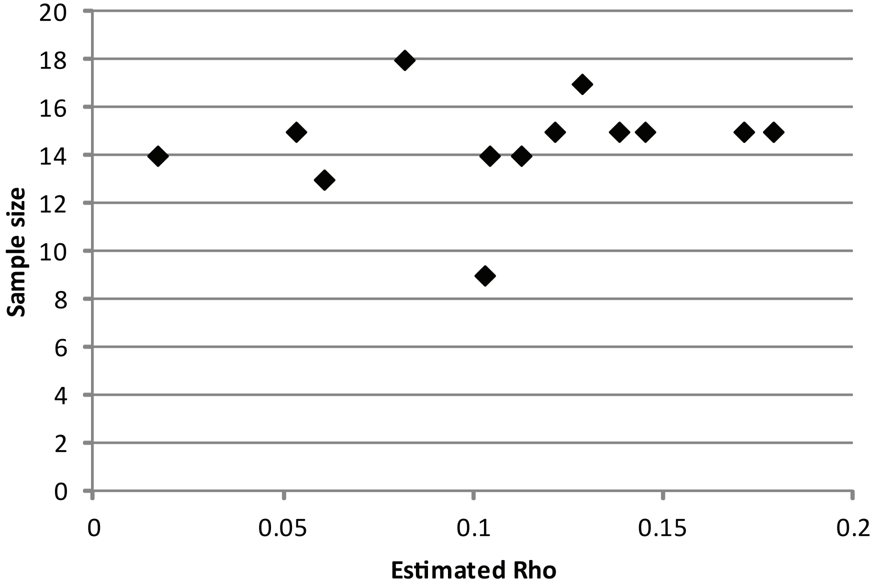
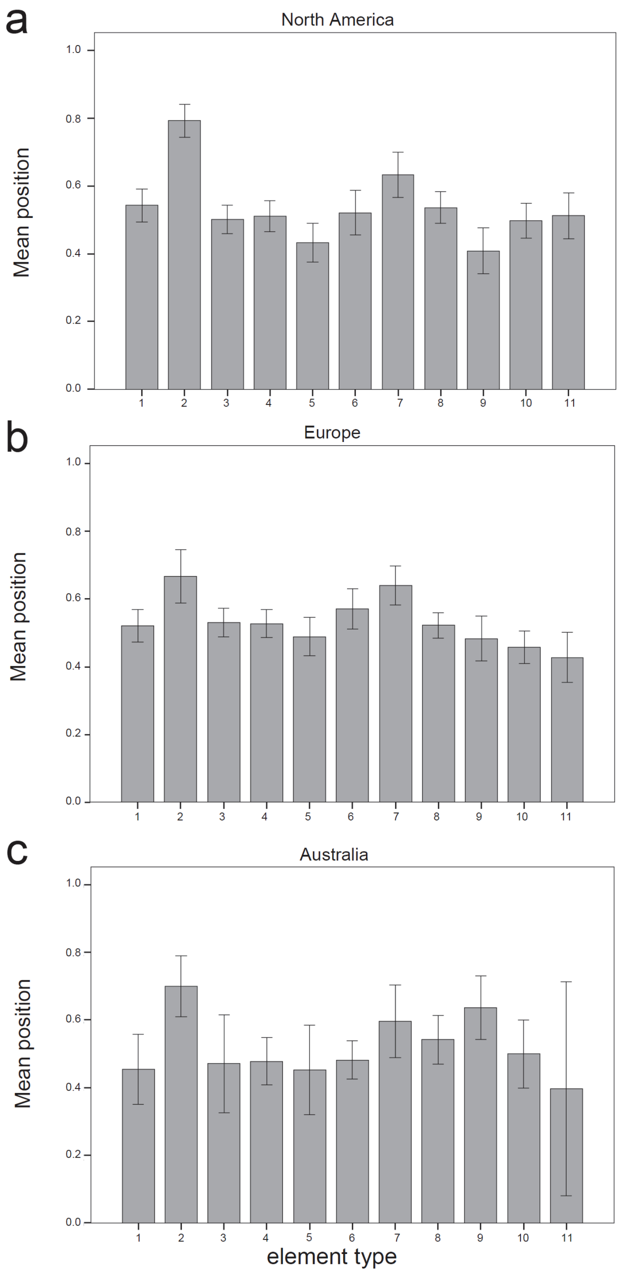


Fig. SI 4.2. Relative element position in the song on continent level. Error bars represent 95% CI. A) North America. B) Australia. C) Europe.



5 Summary and General discussion

5.1 Summary

The aim of this thesis is to disentangle experience dependent and –independent mechanisms in vocal learning and to give insight in their relation to typological patterns in language and birdsong. In perceptual experiments on human infants and juvenile zebra finches I tested to which sounds learners attend and what drives this attention.

In Chapter 2, I tested human infants for their preferences independent of exposure by balancing stimuli for frequency of occurrence in the ambient language (in this case Dutch). The results show that 9-month-olds prefer homorganic syllables to heterorganic ones independent of frequency of occurrence. This probably reflects a processing of larger units (words or syllables) than segments, as has been suggested for early child language production. An early suprasegmental representation would also explain the lack of preference for unmarked segments at 9 months of age. The preference for heterorganic syllables in 12-month-olds possibly indicates a perceptual shift from processing larger units to segmental processing. Whether this is driven by biological maturation or exposure to language is currently unclear, however a combination of both factors is likely. Homorganic consonant vowel sequences have been suggested to be more common across languages, which could be partly caused by the early bias. The evidence for this cross-linguistic pattern in the literature is mixed, however, thus more insight in this is necessary to reveal if and how the homorganicity bias affects cross-linguistic patterns. Initial perceptual preferences in zebra finches (Chapter 3) were independent of song exposure because the birds were raised by the female parent only, which does not sing. These birds, naive to song, prefer common over uncommon song elements, somewhat similar to unmarked and marked sounds in human language. These preferences indicate that, independent of exposure, perception is selective for certain within species elements over other within-species elements.

In addition I tested experience dependent factors in both species. In human infants (Chapter 2) I tested whether frequent syllables are perceptually preferred over infrequent ones, independent of the possibly interfering factor markedness. Effects for frequency were found neither in 9-month-olds nor in 12-month-olds. This result contradicts earlier findings of infant preferences for frequent phonotactic patterns. However, the earlier study did not control for markedness, which demonstrates how important it is to study these effects separately. I do not deny that frequency or other forms of experience play a role, but I do want to emphasize the possibility of interaction with other factors such as markedness or homorganicity. More research is necessary to find out how exactly experience dependent and independent factors interact and at which stage in development.

Experience in zebra finches (Chapter 3) was manipulated by tape-tutoring the birds with songs containing different types of song elements (common or uncommon). Birds tutored with common songs preferred common songs at a later age as well, but now only preferred the specific common song they had been exposed to. Birds tutored with uncommon song changed their preference towards uncommon song and also specifically preferred their (uncommon) tutor song. This clearly demonstrates the effect of experience, which alters the preference towards the sounds they have been exposed to. The similarity between the pupil's song and the tutor song is higher for birds exposed to common songs than for birds exposed to uncommon songs, revealing an interaction between the early bias and later exposure. This interaction may explain why variation and uniformity in vocal signals exist.

The classification of elements into common and uncommon in the experiment with juvenile songbirds in Chapter 3 was based on literature describing zebra finch elements. These studies are valuable, but the methods used in the different studies are quite different and information on different populations is limited. In order to gain more insight in this, I studied the distribution of song elements and sequences across 13 zebra finch populations to approach a zebra finch typology (Chapter 4). Little evidence was shown for cross-population patterns of element sequences. Some populations however, showed weak evidence for some preferred sequences. Furthermore, one type of element seems more likely to be placed near or at the end of the song than at the beginning. Data on element distribution, however, did show that all element clusters occur in all populations. Nevertheless, variation between populations was found in the frequency of occurrence of certain elements within populations. A clear link with the experimental data in juvenile zebra finches could not be shown.

Further research should reveal what the long lasting effect of biases and experience is on variation and uniformity in birdsong and language. The two experimental studies do show a link between early biases and later production. The experiment on juvenile zebra finches shows how experience dependent and independent mechanisms interact. Furthermore both studies show that different factors affect perception at different developmental stages. The findings in this thesis, together with those in previous research, suggest that similar mechanisms operate in songbirds and human infants. More specifically, early biases are independent of experience and later preferences are more influenced by experience. Both factors seem to affect later production patterns, but the relation to cross-population and cross-linguistic patterns is not clear yet.

5.2 Experience independent processes in infants and birds

In Chapter 2 and 3 I demonstrate that both in human infants and in juvenile male songbirds, initial preferences are present independent of adult exposure. Using a 'looking while listening' paradigm, I show that human 9-month-old infants prefer homorganic syllables to heterorganic ones, independent of frequency. These syllables, of which the consonant (C) and the vowel (V) are pronounced with a similar place of articulation (PoA), are abundant in child language production. The early perceptual bias is interpreted as an indication of processing at a suprasegmental level, which has been suggested for child language production as well. Canonical babbling (repeated CV sequences like bababa) usually consists of homorganic syllables. In such sequences the syllable is called a 'frame' and only later the 'content' (segments) will develop (MacNeilage, 1998b). When young children start to produce their first words, these are also predominantly homorganic forms (Fikkert & Levelt, 2008). As shown in chapter 2, by 12 months of age (approximately the onset of word production) perceptual preference has shifted to heterorganic over homorganic syllables, which is interpreted as a first step in a developmental change from a suprasegmental to a segmental representation. The first heterorganic word productions have been found only around 17 months of age (Fikkert & Levelt, 2008) following the same chronology as perception but later. Thus the same developmental path is shown in perception and production, although perception precedes production. This would also be in line with the finding in chapter 2 that infants showed no sensitivity to segmental markedness contrasts at 9 months of age. If infants have no segmental representation yet, it is not likely that they will show a preference for segmental markedness. At 12 months of age however, the infants did not show a segmental markedness preference either, whereas their preference for heterorganic syllables suggest a segmental representation. A possible explanation is that a segmental representation only *begins* to form around 12 months of age but is not mature enough to allow for segmental markedness sensitivity. Further research is necessary to determine if infants perceptually prefer unmarked segments at a different age.

The finding of a perceptual preference independent of frequency suggests that phonological development is influenced by more than just language input. Chapter 3 demonstrates that also in male zebra finches early biases independent of song exposure are present. In this experiment the males were raised without song, which makes interpretation of the data slightly easier. A phonotaxis setup (not to be confused with the term phonotactics in the linguistic literature) was used in this experiment to measure preferences. A measure for preference is the time a bird spends on the side of the cage where a stimulus is played compared to the other side of the cage, where another stimulus is played. This indicated that birds naive to song, preferred 'typologically' common zebra finch elements to uncommon ones. Note that common and uncommon does not refer to how often the birds have heard the elements, because they haven't been exposed to song. Conceptually, the terms uncommon and common are more comparable to universal and non-universal, or unmarked and marked. Chapter 4 however, suggests that strictly speaking there is no evidence for non-universals. Nevertheless, looking at song described in the literature it seems that some elements are more likely to appear in songs (common) than others (uncommon). Thus before song exposure, male zebra finches attend more to common elements than uncommon ones.

For zebra finches we cannot completely exclude the role of exposure to vocalizations other than song. Females do not sing but they do produce calls, which may affect the perception of the nestlings. This possibility is discussed in the next section but we consider it unlikely that this would explain the early bias fully. For the following discussion I will therefore assume that this bias is a predisposition.

Birds tutored with common elements sing songs that are more accurately copied than those of birds tutored with uncommon elements. This shows again a link between the early bias and later production, similar to what is found for human infants. In previous literature a preference for conspecific over heterospecific songs was shown that was independent of song experience (Braaten & Reynolds, 1999; Nelson, 2000; Nelson & Marler, 1993; Soha & Marler, 2001) and this was hypothesized to be indicative of a selective learning mechanism. The perceptual system selectively directs attention to conspecific sounds, catalyzing song learning of the appropriate song elements. Marler et al suggest it may work as a template (see Adret 2004 of a review) of features for conspecific sounds pre-encoded in the brain. This results in selective neural activity triggered by auditory exposure corresponding to the conspecific features represented in the template (Marler & Peters, 1977; Marler & Sherman, 1983).

The findings in this thesis more specifically show that also *within species* selection is taking place. This also sheds light on the template hypothesis. Since not all within species elements are preferred initially, it is possible that the template does not cover features for the whole repertoire but only features for common elements and the rest is learnt. In other words, the more acoustic features match the template, the more it is activated and the stronger the bias. Thus, heterospecific elements match the template poorly whereas uncommon zebra finch elements match slightly better and common zebra finch elements show the best match and therefore trigger the most activation. This activation will enhance learning initially and subsequent exposure to non-matching elements will gradually adjust to or complement the initial template. This latter idea would also support the finding of a more gradual distribution of common and uncommon elements rather than a strict separation between the two. It is at this point unclear how the template is affected by experience and what is exactly encoded in the template. Nevertheless, the finding of an initial bias for within species vocalizations, suggest that not all conspecific vocalizations are pre-encoded equally.

A similar mechanism has been suggested for human infants as well. It has been suggested that (a sensitivity to) a set of unmarked phonological features is present in infants before they start the language acquisition process (Jakobson, 1941). This could be similar to the pre-encoded template for conspecific common elements in songbirds. However, we did not find a perceptual preference for unmarked sounds in infants in the presents study. Hence, at this point I cannot draw the conclusion that infants have a ‘template’ for unmarked features. Infants do have a bias for homorganicity (larger units than the segment) on the other hand (which has been suggested to be unmarked in itself) and this might be pre-encoded. A bias for processing units larger than the segment might interact with a bias for segmental phonological features and could possibly mask an effect at the ages tested in the experiments described in chapter 2. A preference for marked segments in older children at the stage of segmental processing, or a preference for homorganic syllables with unmarked PoA over homorganic syllables with a marked PoA could be a sign of a 'template' for unmarked features in infants. This has so far not been tested, but this would be interesting future research.

5.2.1 Alternative explanations for the early biases

The specific perceptual bias demonstrated in this thesis cannot be explained by a difference between native and foreign language since all stimuli were Dutch syllables. The bias cannot be explained by within language frequency either, since homorganic-heterorganic stimulus sets were balanced for frequency. However, self perception may play a role, since canonical babbling starts around 7 months of age and is thus quite common at 9 months of age. One preliminary finding (unpublished data from the same children tested in the experiment in Chapter 2) suggests that the babbling in 12-month-olds is more homorganic than in 9-month-olds, whereas the perceptual data show the opposite. This is in line with the idea that perception precedes production. The data on children's first words are also more in favor of the thought of production after perception, since heterorganic word productions are found first by 17 months of age (Fikkert & Levelt, 2008). More research is necessary however to unravel the exact interaction between perception and production.

For the initial bias for common elements found in zebra finches we cannot completely rule out the role of any type of exposure (Chapter 3). Females do not sing but they do produce calls (Zann 1996), which may affect the perception of the nestlings. Even though there is some similarity between female calls and the elements classified as common, not all of these elements resemble female calls. It is thus unlikely that female calls would explain the bias as such, but they may have an amplifying effect for instance. More importantly, for both human infants and juvenile zebra finches, we can conclude that the early biases cannot be explained by, respectively, adult language input and song input alone (Chapter 2 and 3 of this thesis). It is, thus, important to take other factors than language and song exposure into account when studying vocal acquisition.

5.3 Experience dependent processes in infants and birds

Previous research has shown that 9-month-old infants' discriminative abilities are affected by exposure to language (Best & McRoberts, 2003; Best et al, 1995; Kuhl et al, 2006; Werker & Tees, 1984) and that infants prefer to listen to their native language (Jusczyk et al, 1993a; Jusczyk et al, 1993b).

In Chapter 2 the role of experience is also specifically addressed by testing if infants prefer frequent over infrequent syllables independent of the possibly correlating factor markedness**.** Using the same paradigm as for experience independent factors, it was shown that 9-month-old infants do not prefer frequent over infrequent syllables. At 12 months of age, infants still do not prefer frequent over infrequent syllables. This is in contrast with earlier findings by Jusczyk et al., who demonstrated a perceptual preference for high probability phonotactic patterns within the native language in 9-month-olds. This can probably be explained by the selection of stimuli. The stimuli used by Jusczyk et al. were not balanced for markedness. First of all this resulted in a stronger frequency difference between frequent and infrequent stimulus sets, because it is easier to find highly frequent syllables or phonemes when it is no problem that they are unmarked. The extremely infrequent stimuli used by Jusczyk et al. might actually have been completely unfamiliar and perceptually comparable to foreign language to infants. Second, markedness itself might have played a role, interacting with frequency. That is, syllables that are both frequent AND unmarked could elicit a stronger response than frequent-marked or infrequent-unmarked syllables and the weakest response would be expected for infrequent-marked syllables. This means that the findings in the study by Jusczyk et al. might be confounded since not only frequency but also markedness may have caused the preference. Markedness alone, however, cannot explain the different results in the two studies since in this thesis the 9-month-olds showed no preference for unmarked segments (neither did the 12-month-olds). Further research is necessary to explore if and how markedness and frequency interact.

Experience does play a role in later perceptual development in zebra finches, as clearly indicated by the shift in perceptual preference in the group of zebra finches tutored with uncommon elements. These birds initially did not have a preference for uncommon elements, but after 20 days of exposure to uncommon elements, the initial preference for common elements declined and the birds preferred the song they had been exposed to. Also in the birds exposed to common elements, the preference became more specific for their tutor song: for common songs other than their tutor song, a preference was no longer found at 55 days post hatch. Several earlier studies have shown the relevance of auditory experience to song learning in birds (see Woolley 2012 for review) and have shown preferences for the tutor song specifically (Clayton, 1988; Houx & ten Cate, 1999; Houx & ten Cate, 1999b; Riebel & Smallegange, 2003; Riebel et al, 2002). Little is known however, on how predispositions and exposure interact. In chapter 3, I show that experience can override initial biases and change perceptual preferences. Nevertheless, the results on song production also indicate that even though production is affected by experience, there still remains an effect of the early bias on production as well. The perceptual preference is equally affected by exposure to common as well as uncommon elements, but the song quality (song similarity with the tutor) is better for birds tutored with common elements than for birds tutored with uncommon elements. Even though the lack of effect for frequency on infant preferences in Chapter 2 does not confirm experience dependent effects in human infants, obviously experience has to play a role at some point. Previous research also shows developmental changes from experience independent towards experience dependent effects. For instance, infants can discriminate non-native phonemes early in development whereas in the second half of the first year, this discriminative ability declines, while native language discrimination increases during this period. Whether the shift in preference from homorganic to heterorganic syllables (or from suprasegmental to segmental representations) between 9 and 12 months of age (Chapter 2) is driven by the same mechanism remains an open question. In human infants, it is often hard to distinguish between developmental change caused by auditory experience or by maturational processes. This is still an issue for the change in preference from homorganic to heterorganic syllables shown in chapter 2. Since the frequency for homorganic and heterorganic stimuli was balanced, this could not directly explain the shift. Obviously, 12-month-olds have 3 months more experience, which possibly gives them time to pass through the homorganic stage, and facilitates focusing on a new learning stage. Previous research comparing full term infants with preterm infants gives some insight in maturational processes, because their linguistic input is similar but the brain maturation stage is different (Pena et al, 2010). It would be interesting to test pre-term and full-term infants on their preferences for homorganicity.

In summary, combining knowledge from previous studies and the present research it becomes clear that in both human infants and zebra finches experience plays a role in vocal development and may partly override initial biases. It seems that initially in both species experience independent processes dominate, whereas over the course of development experience dependent processes take over. As is especially evident in the songbirds, however, there are interactions between initial biases and later experience.

5.4 Effects of development on language and birdsong typology

What can vocal development tell us about the evolution of sounds? Cultural evolution of signals is only possible by virtue of learning. Even though humans are ‘open ended’ learners, meaning learning is still possible in adulthood, the major part of language learning occurs early in development. It is thought that humans and zebra finches have a sensitive phase during which they have to be exposed to language for proper language acquisition to be possible. Moreover, biological aspects of development may influence the learning process during this stage and affect the end product: the actual speech or song produced in adulthood, which will be the input for the next generation. This implies that it is highly relevant to study vocal development from an evolutionary perspective too. Studying the effects of experience dependent and independent processes on the adult end product, could shed light on the interaction between cultural evolution and biological processes.

The experience independent biases in songbirds and human infants may affect the final vocal productions of song and speech by constraining learning to a subset of the signal. The bias for homorganic patterns in infants seems to affect at least later child language productions, since homorganic syllables predominate in babbling and first words (Davis & Macneilage, 1995; Fikkert & Levelt, 2008; Levelt, 1994; Levelt, 1995). As described above, it is unclear yet whether the origin of the bias lies in perceptual or motor mechanisms or both, it is most likely independent of auditory experience. Previously, MacNeilage and Davis have suggested that CV-co-occurrence (similar to homorganicity) is also widespread across adult languages. These authors mention that historical linguistic research also indicates that these homorganic (or CV-co-occurrence) patterns may be part of a 'protolanguage' (MacNeilage & Davis, 2000). This may have an evolutionary background in sucking and chewing (MacNeilage, 1998a; MacNeilage, 1998b), resulting in repetitive jaw movements. These movements, together with phonation, form the first babbling sounds. According to MacNeilage and Davis this could have led to the common occurrence of such patterns in present language. The bias for homorganicity may thus have effects not only on child language but also on adult language. The present data demonstrate that there is an (additional) perceptual bias for homorganicity. The proposed account, an initial suprasegmental representation, might explain the findings on adult language data as well; even though adults have a segmental representation, if language acquisition starts out with a suprasegmental representation, this may increase the chance that homorganic syllables arise in language. The claim by MacNeilage and Davis that homorganicity is frequent across languages, however, has been challenged by Albano et al. (2011). Moreover, heterorganic patterns are also observed cross-linguistically, which complicates the matter. The analyses showing higher frequency cross-linguistically for homorganic patterns (or CV-co-occurrence) are based on observed to expected ratios within language, a calculation that has not been used for more general typological analyses. This aspect needs to be further investigated to better understand the relation between developmental patterns and cross-linguistic patterns.

In zebra finches, a direct effect of the initial bias on adult vocalizations is shown. Birds tutored with more common songs (for which they have an initial bias) produced songs that more accurately resembled the tutor song. This shows there is also a bias for producing these common types of elements, resembling what is found in child language. One study on zebra finch song learning in naive birds, does suggest that biases can lead cultural evolution in a certain direction (Feher et al, 2009). In this study, naive birds were used as tutors for other naive birds, which were thus tutored with isolate songs. This iterative learning was repeated for a few 'cultural generations'. After approximately 4 of these generations the birds' songs were similar to wild type song and no longer resembled isolate song. This suggests that predispositions guide the learning towards wild type song. The initial bias found in chapter 3 might be one of those biases affecting song production patterns.

Possibly this can lead to increasing frequency of occurrence and/or stabilization of such preferred elements within and across populations. If this is indeed the case, the effect should become evident in Chapter 4, where the distribution of elements within and across populations is analyzed. Chapter 4 shows how elements are distributed across 12 laboratory populations and one wild population. It is difficult to directly compare the types of elements classified as common or uncommon in chapter 3 (the juvenile zebra finch experiment) with element clusters in chapter 4, since the classification in chapter 3 was based on visual interpretation of the spectrograms and the one in chapter 4 by means of acoustic analyses. Thus it is not known whether the elements classified as common (chapter 3) would match certain clusters of the acoustic analysis (chapter 4). In chapter 3, high notes (especially inspiratory ones), trills, high sweeps and noisy elements are classified as uncommon. Subjectively, these are most comparable with the clusters 8, 3, 11, respectively, for the first three, while 1, 6 and 10 possibly resemble noisy elements (see fig 4.3a in Chapter 4). The elements classified as more common in Chapter 3 are stacks, slides, short slides and tones, which are probably covered by the clusters 2, 4, 5, 7 and 9 (see fig 4.3a in Chapter 4). Objective acoustic analyses should reveal how well this subjective classification fits. Regardless of this classification, it is clear that none of the elements really dominates. On the other hand, none of the elements is absent from any of the populations either, indicating that in this analysis all elements are ‘universal’. There is variation between populations in how the elements are distributed across populations, in the sense that some elements are more abundantly present in one population than in the other. This variation in distribution, however, does not show a clear relation to the classification of elements into universal or non-universal used for the juvenile zebra finch stimuli. Therefore, it is at this point unclear to what extent the early bias affects sound patterns across and within populations. It should be taken into account that acoustic and clustering analyses were performed on elements of all populations together. Slightly different patterns might occur when analyses would be performed on each population separately. For instance it is possible that in the range of one cluster in the present analysis, two clusters are formed in a one-population analysis (fig. 5.1).

This makes the comparison between this analysis and a typological linguistic analysis more difficult as well. Phonological typology in human language is usually based on categories that can be verified by speakers of that language. Unfortunately, we cannot ask the birds directly to verify which categories they use, and as far as we know the elements in birdsong are not contrastive. Discrimination and categorization tests in songbirds have been performed previously (Prather et al, 2009; Sturdy et al, 1999a; Vicario et al, 2001). A study on zebra finches revealed that they have a categorization that is somewhat comparable to the classification used by Sturdy et al. (Sturdy et al, 1999b), which is again different from the clusters found in Chapter 4. (Although no evidence for categorical perception was shown). This direction of research should be further explored to find out how the clusters found in chapter 4 are perceived by zebra finches.

Apart from how birds categorize song, it is hard to predict what an acoustic clustering analysis similar to the one in Chapter 3 would result in when performed on human speech. It is likely that some clusters will be found when analyzing sound pronunciations using phonological or phonetic information in human speech (Miller, 1989). For instance, clustering has been shown for vowel distribution, at least within a particular language. The distribution into clusters with high frequency in the centre of the cluster and low frequency around the borders of the cluster (bimodal distribution) has been thought to be a possible cue for infants to learn phonetic categories. Usually however, these comparisons are based on the features known to be the cue for a segmental contrast (for instance formant frequency (Miller, 1989)). Furthermore, phoneme clusters seem to overlap too, even within a language, thereby blurring cluster boundaries. Nevertheless, some language-specific features are so different from others (for instance click sounds (Herbert 1990)) that it seems likely they would form a cluster that does not occur in other languages. If this is correct, it follows that there is more diversity among languages than among songs in different zebra finch populations.

In summary, early perceptual biases affect production patterns later in development in songbirds as well as in human children. Whether this effect of the bias on song and speech production results in patterns cross-linguistically or across songbird populations is not clear yet.

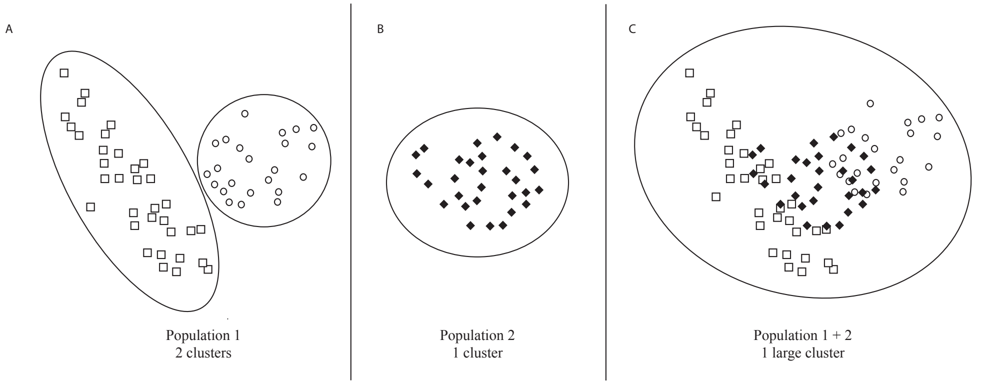


Figure 5.1. A schematic representation of how grouping populations in the analysis might blur some cluster boundaries. If a separate analysis on population 1 reveals 2 clusters (A) and a separate analysis of population 2 reveals one (intermediate) cluster, an analysis where population 1 and 2 are analyzed together might blur the cluster boundaries and group all three clusters into one large cluster (C).

5.5 Conclusion

This thesis has demonstrated that phonological development in both human infants and in zebra finches is affected by factors dependent on, as well as independent of adult input. Early biases have been shown in infants to be independent of frequency of occurrence, and in zebra finches biases were present before song exposure. Such biases can guide the learning process early in development. Whether these biases have an effect on the sound patterns within and between populations and can explain mechanisms like sound change and/or evolution is at this point unclear. No direct relation is found between the early biases in zebra finches and population wide acoustic clustering analyses, whereas in chapter 3 the birds' songs probably were affected by the initial bias. Further research is needed to elucidate this issue. What the acoustic clustering analysis does show is that, similar to language, population differences are found in the way elements are distributed, some elements occurring more often in one population than in the other.

The influence of experience dependent factors is very clear in the study on juvenile zebra finches where the type of elements birds are tutored with, determines which elements they prefer to listen to after tutoring. The exposure also affects their song copying (as expected) but the copying is slightly more accurate in birds tutored with common elements (for which the birds have an early bias) indicating an interaction between the early bias and subsequent exposure. In infants the effect of experience is less clear, since no effect of frequency was found. The preference changes over the course of development as well, but if this change is driven by experience or because of maturation is at this point unclear. Obviously infants are influenced by experience in one way or another, so an interaction between experience dependent and independent processes is likely to explain these results. The perceptual data fit the child language production data very well, which also shows an early preference for homorganic syllables, while heterorganic syllables occur only later. Exploring the interaction between experience dependent and experience independent processes in vocal learning in more detail is a challenge for future research.

REFERENCES

Abe K, Watanabe D (2011) Songbirds possess the spontaneous ability to discriminate syntactic rules. *Nature Neuroscience* **14:** 1067-1047

Adret P (2004) In search of the song template. In *Behavioral Neurobiology of Birdsong*, Zeigler HPMP (ed), Vol. 1016, pp 303-324.

Albano E (2011) On the interpretation of consonant-vowel co-occurrence frequency biases In *The 17th International Congress of Phonetic Sciences (ICPhS XVII)*. Hong Kong

Albrecht DJ, Oring LW (1995) Song in chipping sparrows, spizella-passerina - structure and function. *Animal Behaviour* **50:** 1233-1241.

Alvater-Mackensen, N (2010) Do manners matter? Asymmetries in the acquisition of manner of articulation features. PhD Dissertation, Radboud Universiteit Nijmegen.

Anderson MJ (2006) Distance-based tests for homogeneity of multivariate dispersions. *Biometrics* **62:** 245-253

Anderson MJ, Ellingsen KE, McArdle BH (2006) Multivariate dispersion as a measure of beta diversity. *Ecology Letters* **9:** 683-693

Beckers GJL, Bolhuis JJ, Okanoya K, Berwick RC (2012) Birdsong neurolinguistics: songbird context-free grammar claim is premature. *Neuroreport* **23:** 139-145

Berwick RC, Okanoya K, Beckers GJL, Bolhuis JJ (2011) Songs to syntax: the linguistics of birdsong. *Trends in Cognitive Sciences* **15:** 113-121

Best CC, McRoberts GW (2003) Infant perception of non-native consonant contrasts that adults assimilate in different ways. *Language and Speech* **46:** 183-216

Best CT, McRoberts GW, Lafleur R, Silverisenstadt J (1995) Divergent developmental patterns for infants perception of 2 nonnative consonant contrasts. *Infant Behavior & Development* **18**

Boersma, P., & Weenink, D. (2009). [Computer Program] Praat: doing phonetics by computer (Version 5.1.25).

Bolhuis JJ, Hetebrij E, Den Boer-Visser AM, De Groot JH, Zijlstra GGO (2001)

Localized immediate early gene expression related to the strength of song learning in socially reared zebra finches. *European Journal of Neuroscience* **13:** 2165-2170

Bolhuis JJ, Okanoya K, Scharff C (2010) Twitter evolution: converging mechanisms in birdsong and human speech. *Nature Reviews Neuroscience* **11:** 747-759

Bolhuis JJ, Zijlstra GGO, den Boer-Visser AM, Van der Zee EA (2000) Localized neuronal activation in the zebra finch brain is related to the strength of song learning. *Proceedings of the National Academy of Sciences of the United States of America* **97:** 2282-2285

Boysson-Bardies d, B., Vihman MM (1991) Adaptation to Language: Evidence from Babbling and First Words in Four Languages. *Language* **67:** 297-319

Braaten RF, Miner SS, Cybenko AK (2008) Song recognition memory in juvenile zebra finches: Effects of varying the number of presentations of heterospecific and conspecific songs. *Behavioural Processes* **77:** 177-183

Braaten RF, Reynolds K (1999) Auditory preference for conspecific song in isolation-reared zebra finches. *Animal Behaviour* **58:** 105-111

Briefer E, Osiejuk TS, Rybak F, Aubin T (2010) Are bird song complexity and song sharing shaped by habitat structure? An information theory and statistical approach. *Journal of Theoretical Biology* **262:** 151-164

Christiansen MH, Chater N (2008) Language as shaped by the brain. *Behavioral and Brain Sciences* **31:** 489-558

Clayton NS (1988) Song discrimination learning in zebra finches. *Animal Behaviour* **36:** 1016-1024

Clayton NS (1989) The effects of cross-fostering on selective song learning in estrildid finches. *Behaviour* **109:** 163-175

Cohen LB, Atkinson DJ, Chaput HH. (2000) Habit 2000: A new program for testing infant perception and cognition (version 2. 2. 5c) [computer software]. University of Texas, Austin.

Colombo J, Bundy RS (1981) A method for the measurement of infant auditory selectivity. *Infant Behavior & Development* **4:** 219-223

Cooper RP, Aslin RN (1990) Preference for infant-directed speech in the 1st month after birth. *Child Development* **61:** 1584-1595

Davis BL, Macneilage PF (1995) The articulatory basis of babbling. *Journal of Speech and Hearing Research* **38:** 1199-1211

Davis BL, MacNeilage PF (2000) An embodiment perspective on the acquisition of speech perception. *Phonetica* **57:** 229-241

Da Costa, T. (2010). The acquisition of the consonantal system in European Portuguese: focus on Place and Manner features. PhD Dissertation, University of Lisbon.

De Grauwe L (2004) Are olla-vogala Flemish, or is Dutch philology harboring a cuckoo›s egg in its nest? *Tijdschrift Voor Nederlandse Taal-En Letterkunde* **120:** 44-56

De Lacy P (2006) *Markedness: reduction and preservation in phonology*, Cambridge: Cambridge University Press.

Demuth K (1995) Markedness and the development of prosodic structure. In *Proceedings of the North East Linguistic Society 25*, Beckman J (ed), pp 13-25. Amherst, MA: GLSA, Univ. of Mass.

DePaolis RA, Vihman MM, Keren-Portnoy T (2011) Do production patterns influence the processing of speech in prelinguistic infants? *Infant Behavior & Development* **34:** 590-601

Doupe AJ, Kuhl PK (1999) Birdsong and human speech: Common themes and mechanisms. *Annual Review of Neuroscience* **22:** 567-631

Eales LA (1985) Song learning in zebra finches - some effects of song model availability on what is learnt and when. *Animal Behaviour* **33:** 1293-1300

Eales LA (1987) Song learning in female-raised zebra finches: another look at the sensitive phase. *Animal Behaviour* **35:** 1356-1365

Eales LA (1989) The influences of visual and vocal interaction on song learning in zebra finches. *Animal Behaviour* **37:** 507-508

Egnor SER, Hauser MD (2004) A paradox in the evolution of primate vocal learning. *Trends in Neurosciences* **27:** 649-654

Eimas PD, Siquelan E, Jusczyk P, Vigorito J (1971) Speech perception in infants. *Science* **171:** 303-306

Feher O, Wang HB, Saar S, Mitra PP, Tchernichovski O (2009) De novo establishment of wild-type song culture in the zebra finch. *Nature* **459:** 564-568

Ferguson CA, Farwell CB (1975) Words and sounds in early language acquisition. *Language* **51:** 419-439

Fikkert P (1994) On the acquisition of prosodic structure. Leiden University, The Hague

Fikkert P, Levelt CC (2008) How does place fall into place? The lexicon and emergent constraints in the developing phonological grammar. In *Contrast in phonology: Perception and Acquisition*, P. Avery BED, amp, Rice K (eds). Berlin: Mouton

Forstmeier W, Segelbacher G, Mueller JC, Kempenaers B (2007) Genetic variation and differentiation in captive and wild zebra finches (Taeniopygia guttata). *Molecular Ecology* **16:** 4039-4050

Fraley C, Raftery AE (1998) How many clusters? Which clustering method? Answers via model-based cluster analysis. *Computer Journal* **41:** 578-588

Fraley C, Raftery AE (2002) Model-based clustering, discriminant analysis, and density estimation. *Journal of the American Statistical Association* **97:** 611-631

Friederici AD, Wessels JMI (1993) Phonotactic knowledge of word boundaries and its use in infant speech-perception. *Perception & Psychophysics* **54:** 287-295

Gammon DE, Baker MC, Tipton JR (2005) Cultural divergence within novel song in the Black-capped Chickadee (Poecile atricapillus). *Auk* **122:** 853-871

Gentner TQ, Fenn KM, Margoliash D, Nusbaum HC (2006) Recursive syntactic pattern learning by songbirds. *Nature* **440:** 1204-1207

Gildersleeve-Neumann CE, Davis BL, MacNeilage PF (2000) Contingencies governing the production of fricatives, affricates, and liquids in babbling. *Applied Psycholinguistics* **21:** 341-363
Glaze CM, Troyer TW (2006) Temporal structure in zebra finch song: Implications for motor coding. *Journal of Neuroscience* **26:** 991-1005

Hailman JP, Ficken MS (1986) Combinatorial animal communication with computable syntax - chick-a-dee calling qualifies as language by structural linguistics. *Animal Behaviour* **34:** 1899-1901

Havy M, Nazzi T (2009) Better Processing of Consonantal Over Vocalic Information in Word Learning at 16 Months of Age. *Infancy* **14:** 439-456
Helekar SA, Espino GG, Botas A, Rosenfield DB (2003) Development and adult phase plasticity of syllable repetitions in the birdsong of captive zebra finches (Taeniopygia guttata). *Behavioral Neuroscience* **117:** 939-951

Helekar SA, Marsh S, Viswanath NS, Rosenfield DB (2000) Acoustic pattern variations in the female-directed birdsongs of a colony of laboratory-bred zebra finches. *Behavioural Processes* **49:** 99-110

Holveck M-J, de Castro ACV, Lachlan RF, ten Cate C, Riebel K (2008) Accuracy of song syntax learning and singing consistency signal early condition in zebra finches. *Behavioral Ecology* **19:** 1267-1281

Holveck M-J, Riebel K (2007) Preferred songs predict preferred males: consistency and repeatability of zebra finch females across three test contexts. *Animal Behaviour* **74:** 297-309

Honda E, Okanoya K (1999) Acoustical and syntactical comparisons between songs of the white-backed munia (Lonchura striata) and its domesticated strain, the Bengalese finch (Lonchura striata var. domestica). *Zoological Science* **16:** 319-326

Houx BB, ten Cate C (1999) Do stimulus-stimulus contingencies affect song learning in zebra finches (Taeniopygia guttata)? *Journal of Comparative Psychology* **113:** 235-242

Houx BB, ten Cate C (1999) Song learning from playback in zebra finches: is there an effect of operant contingency? *Animal Behaviour* **57:** 837-845

Houx BB, ten Cate C, Feuth E (2000) Variations in zebra finch song copying: An examination of the relationship with tutor song quality and pupil behaviour. *Behaviour* **137:** 1377-1389

Hyman LM (2008) Universals in phonology. *Linguistic Review* **25:** 83-137

Immelmann K (1969) Song development in the zebra finch and other estrildid finches. In *Bird Vocalizations*, Hinde RA (ed), pp. 61–74. Cambridge: Cambridge University

Press.

Ingram D (1974a) Fronting in child phonology. *Journal of Child Language* **1:** 233-241

Ingram D (1974b) Phonological rules in young children. *Journal of Child Language* **1:** 49-64

Ingram D (1976) Sensorimotor intelligence and language development. . In *Action, gesture and symbol: the emergence of language.*, Lock A (ed). New York: Academic Press

Ingram D (1978) The role of the syllable in phonological development. In *Syllables and segments*, A. Bell JH (ed), pp 143-156. New York: Elsevier-North Holland

Jakobson R (1941) *Child Language, Aphasia and Phonological Universals (Translated by A. Keiler, 1968),* Den Haag-Parijs: Mouton.

Johnson EK, Jusczyk PW (2001) Word segmentation by 8-month-olds: When speech cues count more than statistics. *Journal of Memory and Language* **44:** 548-567

Jones AE, Slater PJB (1996a) The role of aggression in song tutor choice in the zebra finch: Cause or effect? *Behaviour* **133:** 103-115

Jones AE, ten Cate C, Bijleveld CJH (2001) The interobserver reliability of scoring sonagrams by eye: a study on methods, illustrated on zebra finch songs. *Animal Behaviour* **62:** 791-801

Jones AE, ten Cate C, Slater PJB (1996b) Early experience and plasticity of song in adult male zebra finches (Taeniopygia guttata). *Journal of Comparative Psychology* **110:** 354-369

Jusczyk PW (1998) Constraining the search for structure in the input. *Lingua* **106:** 197-218

Jusczyk PW, Cutler A, Redanz NJ (1993a) Infants preference for the predominant stress patterns of english words. *Child Development* **64:** 675-687

Jusczyk PW, Friederici AD, Wessels JMI, Svenkerud VY, Jusczyk AM (1993b) Infants sensitivity to the sound patterns of native language words. *Journal of Memory and Language* **32:** 402-420

Jusczyk PW, Luce PA, Charles-Luce J (1994) Infants sensitivity to phonotactic patterns in the native language. *Journal of Memory and Language* **33:** 630-645

Jusczyk PW, Smolensky P, Allocco T (2002) How English-Learning Infants Respond to Markedness and Faithfulness Constraints. *Language Acquisition* **10:** 31 – 73

Kager R, Van der Feest S, Fikkert P, Kerkhoff A, Zamuner T (2007) Representations of [voice]: Evidence from acquisition. In *Voicing in Dutch*, van de Weijer JM, van der Torre EJ (eds), pp 41-80. Amsterdam: John Benjamins

Kipper S, Mundry R, Hultsch H, Todt D (2004) Long-term persistence of song performance rules in nightingales (Luscinia megarhynchos): A longitudinal field study on repertoire size and composition. *Behaviour* **141:** 371-390

Kirby S, Dowman M, Griffiths TL (2007) Innateness and culture in the evolution of language. *Proceedings of the National Academy of Sciences of the United States of America* **104:** 5241-5245

Koopmans-van Beinum FJ, Van der Stelt JM (1986) Early stages in the development of speech movements. In *Precursors of early speech*, Zetterström BJL, R (eds), pp 37-50. Basingstoke, England: Macmillan Press Ltd.

Kroodsma DE, Byers BE, Halkin SL, Hill C, Minis D, Bolsinger JR, Dawson JA, Donelan E, Farrington J, Gill FB, Houlihan P, Innes D, Keller G, Macaulay L, Marantz CA, Ortiz J, Stoddard PK, Wilda K (1999) Geographic variation in Black-capped Chickadee songs and singing behavior. *Auk* **116:** 387-402

Kuhl PK, Meltzoff AN (1996) Infant vocalizations in response to speech: Vocal imitation and developmental change. *Journal of the Acoustical Society of America* **100:** 2425-2438

Kuhl PK, Miller JD (1978) Speech-perception by chinchilla - identification functions for synthetic VOT stimuli. *Journal of the Acoustical Society of America* **63:** 905-917

Kuhl PK, Stevens E, Hayashi A, Deguchi T, Kiritani S, Iverson P (2006) Infants show a facilitation effect for native language phonetic perception between 6 and 12 months. *Developmental Science* **9:** F13-F21

Lachlan RF, Feldman MW (2003) Evolution of cultural communication systems: the coevolution of cultural signals and genes encoding learning preferences. *Journal of Evolutionary Biology* **16:** 1084-1095

Lachlan RF, Peters S, Verhagen L, ten Cate C (2010) Are There Species-Universal Categories in Bird Song Phonology and Syntax? A Comparative Study of Chaffinches (Fringilla coelebs), Zebra Finches (Taenopygia guttata), and Swamp Sparrows (Melospiza georgiana). *Journal of Comparative Psychology* **124:** 92-108

Lahiri A, Evers V (1991) Palatalization and Coronality. In *The special status of coronals: Internal and external evidence. (Phonology and Phonetics 2.), 79–100*, Paradis C & Prunet, J-F (ed). San Diego: Academic Press.

Lauay C, Gerlach NM, Adkins-Regan E, Devoogd TJ (2004) Female zebra finches require early song exposure to prefer high-quality song as adults. *Animal Behaviour* **68:** 1249-1255

Leadbeater E, Goller F, Riebel K (2005) Unusual phonation, covarying song characteristics and song preferences in female zebra finches. *Animal Behaviour* **70:** 909-919

Lemon RE, Chatfiel.C (1973) Organization of song of rose-breasted grosbeaks. *Animal Behaviour* **21:** 28-44

Levelt CC, van de Vijver R (2004) Syllable Types in Cross-Linguistic- and Developmental Grammars. In *Fixing Priorities: Constraints in Phonological Acquisition*, Kager RJP, Zonneveld W (eds). Cambridge University Press

Levelt CC, Van Oostendorp M (2007) Feature co-occurrence constraints in L1 acquisition. In *Linguistics in The Netherlands 2007*, Los B & van Koppen, M (ed), pp 162-172. Amsterdam/Philadelpia: John Benjamins Publishing Co.

Levelt CC (1994) On the acquisition of place. PhD Dissertation, University of Leiden University, Leiden

Levelt CC (1995) The segmental structure of early words: articulatory frames or phonological constraints. In *Twenty-seventh annual child language research forum*, Vol. 27. Stanford

Levelt CC, Schiller NO, Levelt WJM (1999) A developmental grammar for syllable structure in the production of child language. *Brain and Language* **68:** 291-299

Levelt CC, Schiller NO, Levelt WJM (2000) The acquisition of syllable types. *Language Acquisition* **8:** 237-264

Macken MA (1980) The Acquisition of Stop Systems: A Cross-linguistic Perspective. In *Child Phonology: Perception and Production I*, Yeni-Komshian G, Kavanagh JF, Ferguson CA (eds), pp 143-168. New York: Academic Press

Macken MA, Ferguson CA (1981) Phonological Universals in Language-Acquisition. *Annals of the New York Academy of Sciences* **379:** 110-129

MacNeilage PF (1998) The frame/content theory of evolution of speech production. *Behavioral and Brain Sciences* **21:** 499-546

MacNeilage PF, Davis B (1990) Acquisition of Speech Production - Frames, Then Content. *Attention and Performance***:** 453-476

MacNeilage PF, Davis BL (2000) On the origin of internal structure of word forms. *Science* **288:** 527-531

MacNeilage PF, Davis BL, Kinney A, Matyear CL (2000) The motor core of speech: A comparison of serial organization patterns in infants and languages. *Child Development* **71:** 153-163

Marler P, Peters S (1977) Selective vocal learning in a sparrow. *Science* **198:** 519-521

Marler P, Pickert R (1984) Species-universal microstructure in the learned song of the swamp sparrow (*Melospiza-georgiana*). *Animal Behaviour* **32:** 673-689

Marler P, Sherman V (1983) Song structure without auditory-feedback - emendations of the auditory template hypothesis. *Journal of Neuroscience* **3:** 517-531

Marler P, Tamura M (1964) Culturally transmitted patterns of vocal behavior in sparrows. *Science* **146:** 1483-1586

Martinez A, Eisenberg L, Boothroyd A, Visser-Dumont L (2008) Assessing speech pattern contrast perception in infants: Early results on VRASPAC. *Otology & Neurotology* **29:** 183-188

Menn L (1976) Pattern, Control, and Contrast in Beginning Speech: A Case Study in the Development of Word Form and Word Function. PhD Dissertation Thesis, University of Illinois, Champaign

Miller JD (1989) Auditory-perceptual interpretation of the vowel. *Journal of the Acoustical Society of America* **85:** 517-531

Moorman S, Gobes SMH, Kuijpers M, Kerkhofs A, Zandbergen MA, Bolhuis JJ (2012) Human-like brain hemispheric dominance in birdsong learning. *Proceedings of the National Academy of Sciences of the United States of America* **109:** 12782-12787

Moreton E (2008) Analytic bias and phonological typology. *Phonology* **25:** 83-127

Moskowitz A. (1970) The acquisition of phonology. Report nr. WP-34. University of California at Berkeley, Language and Behavior Research Lab.

Nelson DA (1998) Geographic variation in song of Gambel's white-crowned sparrow. *Behaviour* **135:** 321-342

Nelson DA (2000) A preference for own-subspecies' song guides vocal learning in a song bird. *Proceedings of the National Academy of Sciences of the United States of America* **97:** 13348-13353

Nelson DA, Marler P (1989) Categorical perception of a natural stimulus continuum - birdsong. *Science* **244:** 976-978

Nelson DA, Marler P (1993) Innate recognition of song in white-crowned sparrows - a role in selective vocal learning. *Animal Behaviour* **46:** 806-808

Nelson DGK, Jusczyk PW, Mandel DR, Myers J, Turk A, Gerken L (1995) The head-turn preference procedure for testing auditory-perception. *Infant Behavior & Development* **18:** 111-116

Nottebohm F, Stokes TM, Leonard CM (1976) Central control of song in canary, serinus-canarius. *Journal of Comparative Neurology* **165:** 171-192

Ohms VR, Gill A, Van Heijningen CAA, Beckers GJL, ten Cate C (2009) Zebra finches exhibit speaker-independent phonetic perception of human speech. *Proceedings of the Royal Society B-Biological Sciences* **277:** 1003-1009

Ozturk I, Orhan H, Dogan Z (2009) Comparison of Principal Component Analysis and Multidimensional Scaling Methods for Clustering Some Honey Bee Genotypes. *Journal of Animal and Veterinary Advances* **8:** 413-419

Pena M, Pittaluga E, Mehler J (2010) Language acquisition in premature and full-term infants. *Proceedings of the National Academy of Sciences of the United States of America* **107:** 3823-3828

Petrinovich L, Baptista LF (1984) Song dialects, mate selection, and breeding success in white-crowned sparrows. *Animal Behaviour* **32:** 1078-1088

Pinheiro JC, Bates DM, DebRoy S, Sarkar D, Team atRC. (2009) nlme: Linear and Nonlinear Mixed Effects Models. R package version 3.1-96, http://www.R-project.org.

Prather JF, Nowicki S, Anderson RC, Peters S, Mooney R (2009) Neural correlates of categorical perception in learned vocal communication. *Nature Neuroscience* **12:** 121-228

Price PH (1979) Developmental determinants of structure in zebra finch song *Journal of Comparative and Physiological Psychology* **93:** 260-277

Pytte CL, Suthers RA (2000) Sensitive period for sensorimotor integration during vocal motor learning. *Journal of Neurobiology* **42:** 172-189

Rice K (2007) *Markedness in phonology*
The Cambridge Handbook of Phonology: Cambridge University Press.

Riebel K (2009) Song and Female Mate Choice in Zebra Finches: A Review. *Advances in the Study of Behavior, Vol 40* **40:** 197-238

Riebel K, Smallegange IM (2003) Does zebra finch (Taeniopygia guttata) preference for the (familiar) father's song generalize to the songs of unfamiliar brothers? *Journal of Comparative Psychology* **117:** 61-66

Riebel K, Smallegange IM, Terpstra NJ, Bolhuis JJ (2002) Sexual equality in zebra finch song preference: evidence for a dissociation between song recognition and production learning. *Proceedings of the Royal Society of London Series B-Biological Sciences* **269:** 729-733

Roberts TF, Tschida KA, Klein ME, Mooney R (2010) Rapid spine stabilization and synaptic enhancement at the onset of behavioural learning. *Nature* **463:** 948-952

Rogers CH (1979) *Zebra finches*, Edlington, Horncastle, Lincs: K&R Books Ltd.

Roper A, Zann R (2006) The onset of song learning and song tutor selection in fledgling zebra finches. *Ethology* **112:** 458-470

Scharff C, Nottebohm F (1991) A comparative-study of the behavioral deficits following lesions of various parts of the zebra finch song system - implications for vocal learning. *Journal of Neuroscience* **11:** 2896-2913

Sebastian-Galles N, Bosch L (2002) Building phonotactic knowledge in bilinguals: Role of early exposure. *Journal of Experimental Psychology-Human Perception and Performance* **28:** 974-989
Seidl A, A Buckley E (2005) On the Learning of Arbitrary Phonological Rules. *Language Learning and Development* **1:** 289 - 316

Simpson HB, Vicario DS (1990) Brain pathways for learned and unlearned vocalizations differ in zebra finches. *Journal of Neuroscience* **10:** 1541-1556

Slater PJB, Clayton NS (1991) Domestication and song learning in zebra finches taeniopygia-guttata. *Emu* **91:** 126-128

Slater PJB, Ince SA (1979) Cultural-evolution in chaffinch song. *Behaviour* **71:** 146-166

Sloetjes H, Wittenburg P (2008) Annotation by category – ELAN and ISO DCR. . In *6th International Conference on Language Resources and Evaluation (LREC 2008).*

Soha JA, Lohr B, Gill DE (2009) Song development in the grasshopper sparrow, Ammodramus savannarum. *Animal Behaviour* **77:** 1479-1489

Soha JA, Marler P (2000) A species-specific acoustic cue for selective song learning in the white-crowned sparrow. *Animal Behaviour* **60:** 297-306

Soha JA, Marler P (2001) Cues for early discrimination of conspecific song in the white-crowned sparrow (Zonotrichia leucophrys). *Ethology* **107:** 813-826

Sossinka R (1970) Domestikationserscheinungen beim Zebrafinken Taeniopygiaguttata castanotis (GOULD). *Jahrbücher Abteilung für Systematik Okologie und Geographie der Tiere* **97:** 455-521

Sossinka R, Bohner J (1980) Song types in the zebra finch poephila-guttata-castanotis. *Zeitschrift Fur Tierpsychologie-Journal of Comparative Ethology* **53:** 123-132

Sturdy CB, Phillmore LS, Price JL, Weisman RG (1999a) Song-note discriminations in zebra finches (Taeniopygia guttata): Categories and pseudocategories. *Journal of Comparative Psychology* **113:** 204-212

Sturdy CB, Phillmore LS, Weisman RG (1999b) Note types, harmonic structure, and note order in the songs of zebra finches (Taeniopygia guttata). *Journal of Comparative Psychology* **113:** 194-203

Suzuki R, Buck JR, Tyack PL (2006) Information entropy of humpback whale songs. *Journal of the Acoustical Society of America* **119:** 1849-1866

Tchernichovski O, Lints T, Mitra PP, Nottebohm F (1999) Vocal imitation in zebra finches is inversely related to model abundance. *Proceedings of the National Academy of Sciences of the United States of America* **96:** 12901-12904

Tchernichovski O, Nottebohm F, Ho CE, Pesaran B, Mitra PP (2000) A procedure for an automated measurement of song similarity. *Animal Behaviour* **59:** 1167-1176

ten Cate C, Lachlan RF, Zuidema W (2013) Analyzing the structure of bird vocalizations and language: finding common ground. In *Birdsong,speech and language: converging mechanisms*, Bolhuis JJ, Everaert M (eds). Cambridge, MA: MIT press

ten Cate C, Okanoya K (2012) Revisiting the syntactic abilities of non-human animals: natural vocalizations and artificial grammar learning. *Philosophical Transactions of the Royal Society B-Biological Sciences* **367:** 1984-1994

ten Cate C, Slater PJB (1991) Song learning in zebra finches - how are elements from 2 tutors integrated. *Animal Behaviour* **42:** 150-152

Terpstra NJ, Bolhuis JJ, den Boer-Visser AM (2004) An analysis of the neural representation of birdsong memory. *Journal of Neuroscience* **24:** 4971-4977

Todt D, Geberzahn N (2003) Age-dependent effects of song exposure: song crystallization sets a boundary between fast and delayed vocal imitation. *Animal Behaviour* **65:** 971-979

Todt D, Hultsch H (1998) How songbirds deal with large amounts ofserial information: retrieval rules suggest a hierarchical song memory. *Biological Cybernetics* **79:** 487-500

Tsao FM, Liu HM, Kuhl PK (2004) Speech perception in infancy predicts language development in the second year of life: A longitudinal study. *Child Development* **75:** 1067-1084

Tsao FM, Liu HM, Kuhl PK (2006) Perception of native and non-native affricatefricative contrasts: Cross-language tests on adults and infants. *Journal of the Acoustical Society of America* **120:** 2285-2294

Tserdanalis G, Wong We (2004) *The language files 9th edition: Materials for an introduction to language and linguistics*, Columbus, USA: Ohio stateuniversity press.

Van der Feest SVH (2007) Building a lexicon:The acquisition of the Dutch voicing contrast in perception and production. PhD Thesis, Radboud university, Nijmegen

Van Heijningen CAA, de Visser J, Zuidema W, ten Cate C (2009) Simple rules can explain discrimination of putative recursive syntactic structures by a songbird species. *Proceedings of the National Academy of Sciences of the United States of America* **106:** 20538-20543

Vicario DS, Naqvi NH, Raksin JN (2001) Sex differences in discrimination of vocal communication signals in a songbird. *Animal Behaviour* **61:** 805-817

Volman SF, Khanna H (1995) Convergence of untutored song in group-reared zebra finches (taeniopygia-guttata). *Journal of Comparative Psychology* **109:** 211-221

Waterson N (1971) Child Phonology: A Prosodic View. *Journal of Linguistics* **7:** 179-211

Weijer van de, JC (1999) Language input for word discovery. Max Planck Institute for Psycholinguistics, Nijmegen

Werker JF, Tees RC (1984) Cross-language speech-perception - evidence for perceptual reorganization during the 1st year of life. *Infant Behavior and Development* **7:** 49-63

Williams H, Staples K (1992) Syllable chunking in zebra finch (taeniopygia-guttata) song. *Journal of Comparative Psychology* **106:** 278-286

Woolley SMN (2012) Early experience shapes vocal neural coding and perception in songbirds. *Developmental Psychobiology* **54:** 612-631

Yang CD (2004) Universal Grammar, statistics or both? *Trends in Cognitive Sciences* **8:** 451-456

Yip MJ (2006) The search for phonology in other species. *Trends in Cognitive Sciences* **10:** 442-446

Zamuner TS (2006) Sensitivity to word-final phonotactics in 9-to 16-month-old infants. *Infancy* **10:** 77-95

Zamuner TS, Gerken L, Hammond M (2004) Phonotactic probabilities in young children's speech production. *Journal of Child Language* **31:** 515-536

Zamuner TS, Gerken L, Hammond M (2005) The acquisition of phonology based on input: a closer look at the relation of cross-linguistic and child language data. *Lingua* **115:** 1403-1426

Zann R (1993a) Structure, sequence and evolution of song elements in wild Australian zebra finches. *Auk* **110:** 702-715

Zann R (1993b) Variation in song structure within and among populations of Australian zebra finches. *Auk* **110:** 716-726

Zann R (1996) *Zebra Finch: A Synthesis of Field and Laboratory Studies*, Oxford: Oxford University Press.

APPENDIX I - DEFINITIONS

conspecific: of the same species

coronal: frontal place of articulation (see appendix II) in Dutch: /t/,/d/,/s/,/z/ and /n/ (and for some accents /r/) are coronal consonants and /i/ and /e/ are front vowels.

experience dependent: this thesis, when I refer to **experience dependent** processes I generally mean the processes influenced by the learner's exposure to adult language, or adult birdsong in the case of birds.

heterospecific: of a different species.

markedness: The notion of **markedness** is often used to describe the observed asymmetry; unmarked sounds are defined as more "universal" (i.e. more common across languages) and are acquired early in development, whereas marked sounds are defined as less common and acquired later.

motif: a sequence of birdsong elements or syllables, which can be repeated within a bout.

phoneme: the smallest unit in speech. Phonemes are contrastive: interchanging phonemes in a word changes the meaning of the word.

phonotactic patterns: combinations of phonemes.

phonological processes: the change of one phoneme or structure into another under certain linguistic circumstances/context. For instance, in Dutch the pronunciation of the 'n' in the word 'inpakken' changes into the labial /m/ because it is followed by a labial /p/

phonological typology: the study of phonemes and their distribution across languages and dialects

phonological syntax: a term used mainly in songbird literature to describe patterns in the order of song elements or syllables and thus a 'syntax' in terms of structure rather than meaning.

predisposition: In this thesis: a (biological) perceptual sensitivity to sounds not heard before.

tutoring: In this thesis: Auditory presentation sound to a juvenile bird providing input for song learning.

tutorsong: the song that a juvenile bird listens to and uses as a model to copy song.

tape-tutoring: an experimental method of exposing a juvenile bird to a song through a speaker (originally a tape recording, these days usually sound files) in order for the bird to copy the song.

s ongbout: a motif or series of motifs repeated directly after each other. Different bouts are separated by relatively long silent intervals (+- >0.5ms).

sonorant: sounds that are produced with a continuous airflow in the vocal tract. In Dutch: /m/,/n/, /ŋ/,/w/,/j/,/l/ and al vowels.

suprasegmental: on a level larger than the phoneme.

syllable:

Linguistic: a speech unit comprised of a vowel or a combination of consonants and a vowel, forming (part of a) word.

Birdsong: an element or combination of elements separated by relatively short (+- < 0.1ms) adjacent silent intervals .

syntax: the structure of a language, in terms of the order of words in a sentence, which is related to the meaning of the sentence.

vocal learning: the process of learning to use vocal communication sounds, by being exposure to the auditory signal produced by adults.

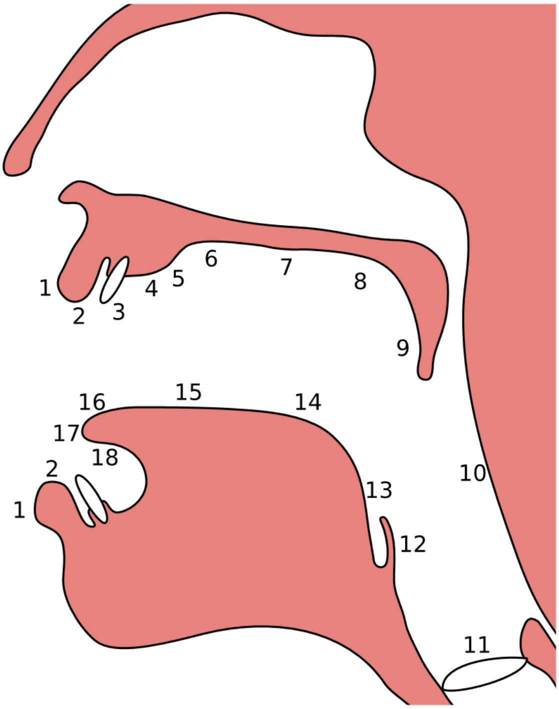
APPENDIX II - PLACES OF ARTICULATION

(active and passive)

Labial:
Exo-labial (outer part of lip)
Endo-labial (inner part of lip)

Coronal:
Dental (teeth)
Alveolar (front part of alveolar ridge)
Post-alveolar (rear part of alveolar ridge & slightly behind it)

Dorsal:
Pre-palatal (front part of hard palate that arches upward)
Palatal (hard palate)
Velar (soft palate)
Uvular (a.k.a. *Post-velar*; uvula)
Pharyngeal (pharyngeal wall)
Glottal (a.k.a. *Laryngeal*; vocal folds)
Epiglottal (epiglottis)
Radical (tongue root)
Postero-dorsal (back of tongue body)
Antero-dorsal (front of tongue body)
Laminal (tongue blade)
Apical (apex or tongue tip)
Sub-laminal (a.k.a. *Sub-apical*; underside of tongue)



source: https://en.wikipedia.org/wiki/Place_of_articulation

NEDERLANDSE SAMENVATTING

"Hebban olla uogala nestas hagunnan hinase hi(c) (a)nda thu uuat unbidan uue nu[1]", is een bekende Oud Nederlandse zin die werd opgeschreven rond het jaar 1100. Sindsdien is de Nederlandse taal aanzienlijk veranderd. Wat een belangrijke rol speelt bij deze veranderingen is 'culturele transmissie'. Hierbij wordt een eigenschap, in dit geval taal, door sociaal contact overgedragen van persoon tot persoon. Bij deze overdracht kunnen onzuiverheden in spraak en waarneming ontstaan, waardoor taal geleidelijk verandert. Culturele transmissie van taal kan alleen plaatsvinden doordat mensen vocaal kunnen leren; de gesproken taal die men hoort kan ook worden nagebootst. Vocaal leren is een eigenschap die vrij zeldzaam is in het dierenrijk. Menselijk gedrag wordt vaak vergeleken met dat van andere primaten, zoals de chimpansee, vanwege de sterke genetische verwantschap. Maar voor het leren van taal en spraak gaat de vergelijking niet op, aangezien niet-menselijke primaten niet vocaal kunnen leren; hun communicatiegeluiden ontwikkelen zich zonder dat ze die gehoord hoeven te hebben. Zangvogels daarentegen, kunnen wel vocaal leren en moeten zang horen tijdens de ontwikkeling om later normaal te kunnen zingen. De cruciale rol van vocaal leren bij zowel zangvogels als mensen hebben onderzoekers er toe gebracht om het leerproces van beide te vergelijken. De manier waarop de ontwikkeling verloopt, blijkt sterk overeen te komen. In dit proefschrift ga ik op een specifieke vergelijking in namelijk die van het leren van klanken. De kleinste klankeenheden in taal zijn fonemen, bijvoorbeeld /b/, /o/ en /m/ in het woord 'boom'. In zebravinken zang is de kleinste klankeenheid een 'element' en een reeks elementen vormt een liedje (fig 1.2, p 6).

Talen kunnen sterk verschillen in de klanken die worden gebruikt. Ondanks de aanmerkelijke variatie zijn er ook klanken die overeenkomen tussen talen (fig 1.1 p 5). Juist deze 'universele' klanken komen ook veel voor in de eerste spraakklanken die jonge kinderen systematisch gebruiken. Dit zou verschillende oorzaken kunnen hebben (die elkaar niet noodzakelijk uitsluiten): 1) universele klanken zijn makkelijker te produceren, 2) universele klanken komen vaak voor binnen een taal, dus het kind hoort deze klanken vaker dan niet-universele klanken 3) kinderen hebben meer aandacht ('voorkeur') voor bepaalde eigenschappen van de universele klanken, onafhankelijk van hoe vaak ze voorkomen. De laatste mogelijkheid wordt ook wel selectief leren genoemd; alleen de klanken waar het kind al meer aandacht voor heeft worden geleerd, aan het begin van het leerproces althans. In hoofdstuk 2 wordt een reeks experimenten beschreven waarbij de twee perceptuele verklaringen, 2 en 3 tegen elkaar worden afgezet en getest. De aandacht werd gemeten van baby's van 9 en 12 maanden, die dan zelf meestal nog niet praten. Terwijl de baby naar een beeldscherm met een schaakbordpatroon keek, werden er reeksen gesproken lettergrepen afgespeeld. Er werd gemeten hoe lang de baby

---

[1] De zin wordt vrij vertaald als: "Hebben alle vogels nesten begonnen behalve ik en jij, waarop wachten we nu?

naar het patroon bleef kijken - een maat voor hoeveel aandacht een baby heeft voor de reeksen. De reeksen contrasteerden ofwel in fonologische gemarkeerdheid (ongemarkeerde klanken zijn klanken die in veel talen voorkomen en ook in vroege kindertaal worden gevonden), dan wel in hoe vaak de lettergrepen in het Nederlands voorkomen (frequentie). Vanwege het vroege voorkomen van homorganische lettergrepen in productie, waren de ongemarkeerde lettergrepen ook homorganisch. Dit zijn lettergrepen waarbij de klinker en de medeklinker op dezelfde plek in de mond worden uitgesproken. Bijvoorbeeld de /t/ en de /i/ in lettergreep /ti/ worden op dezelfde plek uitgesproken: voorin de mond rond de tanden. In de gemarkeerde reeksen waren ze heterorganisch. De reeksen die verschillen in frequentie, verschillen niet in gemarkeerdheid zodat alleen de frequentie een verschil in aandacht kan verklaren. Andersom zijn de reeksen die verschillen in gemarkeerdheid juist weer gelijk in frequentie. De resultaten laten zien dat baby's van 9 maanden meer aandacht hebben voor het type lettergreep dat veel te vinden is in de vroegste gesproken kindertaal (ongemarkeerd), namelijk homorganische lettergrepen. De aandacht voor homorganische lettergrepen is onafhankelijk van hoe vaak de lettergrepen voorkomen in het Nederlands en is dus een vorm van selectieve perceptie die invloed kan hebben op hoe kinderen zich de lettergrepen eigen maken. De 12 maanden oude baby's daarentegen hebben juist meer aandacht voor de heterorganische lettergrepen waarbij de klinker en de medeklinker niet op dezelfde plek worden uitgesproken (zoals /to/). Deze verandering heeft waarschijnlijk te maken met de eenheid waarin baby's fonologische eigenschappen verwerken. Bij homorganiciteit hoort een fonologische eigenschap – plaats van articulatie – bij een grotere eenheid dan een klank, namelijk bij een hele lettergreep. Bij heterorganische lettergrepen is deze eigenschap voor iedere klank uit de lettergreep apart gerepresenteerd. Het zou dus kunnen dat het taalsysteem van baby's eerst bij voorkeur grotere eenheden dan het segment verwerkt, en later pas fonologische eigenschappen op segmenteel niveau representeert. Een zelfde verandering is ook waargenomen in gesproken kindertaal waarbij eerst homorganische lettergrepen worden geproduceerd en later pas heterorganische. Dus naast selectiviteit zien we ook een ontwikkeling in selectiviteit. Een tweede opvallende bevinding is dat baby's geen verschil in aandacht vertonen tussen reeksen lettergrepen die vaak en reeksen lettergrepen die niet vaak in het Nederlands voorkomen. Wat hiervan precies de oorzaak is, moet nog verder onderzocht worden, maar het geeft aan dat de rol van frequentie niet zo eenvoudig is als eerder gedacht.

Bij dit soort onderzoek bij baby's is het erg lastig om ervaringsafhankelijke en –onafhankelijke invloeden uit elkaar te houden. Bij zangvogels is dit makkelijker omdat er meer controle mogelijk is over de factoren tijdens de ontwikkeling. Door vergelijkbaar onderzoek uit te voeren bij jonge kinderen en bij zangvogels kunnen we meer inzicht krijgen in de algemene processen die een rol spelen bij ontwikkeling en specifiek bij vocaal leren. Hoofdstuk 3 laat een aantal experiment zien waarmee de wisselwerking wordt onderzocht tussen ervaring (luisteren naar zang) en selectiviteit bij de zebravink, een zangvogel. Evenals bij mensentaal is er bij zebravinkenzang zowel sprake van variatie als overeenkomst tussen verschillende populaties. Daarnaast zingen zebravinken ieder een eigen liedje en bestaat er dus variatie tussen individuen. Selectiviteit is ook aangetoond bij zebravinken, aangezien ze de zang van hun eigen soort kiezen boven die van een andere soort om van te leren. De vraag is nu of er net als bij mensenbaby's ook *binnen de soort* selectiviteit bestaat. De zangvoorkeuren zijn gemeten door zebravinken individueel in een grote kooi te zetten met aan beide kanten van de kooi een luidspreker en te meten bij welke luidspreker ze het meeste tijd doorbrengen. De luidsprekers spelen zebravinkenliedjes af die zijn opgebouwd uit ofwel universele elementen (klanken) dan wel niet-universele elementen. De resultaten laten zien dat jonge zebravinken die geen zang hebben gehoord, een voorkeur hebben voor zangelementen die 'universeel' zijn voor zebravinkenzang. Aangezien ze geen zang hebben gehoord, is deze voorkeur net als bij de mensenbaby's, onafhankelijk van ervaring. Na deze eerste test werden de jonge zebravinken gedurende 20 dagen blootgesteld aan zang die door een luidspreker werd afgepeeld. Eén groep luisterde naar liedjes opgebouwd uit universele elementen en een andere groep naar liedjes opgebouwd uit niet-universele elementen. Tijdens en na deze leerperiode (tutorfase) zijn de voorkeuren van de jonge vinken opnieuw gemeten. Hieraan is te zien dat in de loop van de ontwikkeling de voorkeur verandert naar wat ze gehoord hebben tijdens de tutorfase. De vogels die getutord zijn met niet-universele liedjes, hadden namelijk voor de tutorfase een voorkeur voor universele liedjes, maar daarna een voorkeur voor niet-universele liedjes. Net als bij mensenbaby's vindt er dus een verandering plaats in selectiviteit gedurende de ontwikkeling. Van dezelfde zebravinken is ook de zang die ze als volwassen vogels produceerden opgenomen en vergeleken met het liedje dat ze hebben gehoord tijdens de tutorfase (tutorliedje). De overeenkomst tussen het leerlingliedje en het tutorliedje is een maat voor leren. Dieren die tijdens de tutorfase een universeel liedje hebben gehoord, leren dus beter dan dieren die niet-universele liedjes hebben gehoord. Dit wijst erop dat zowel ervaring als (ervaringonafhankelijke) selectiviteit een rol spelen bij vocaal leren, in ieder geval voor het leren van zang in zebravinken.

Het indelen van elementen in universeel en niet-universeel in hoofdstuk 3 was gebaseerd op literatuuronderzoek. Er zijn echter maar enkele studies die zebravinkenzang beschrijven op klankniveau en er zijn verschillende methoden gebruikt bij die verschillende studies, wat de vergelijkbaarheid niet ten goede komt. Om deze kennis uit te breiden is in hoofdstuk 4 de zang van 13 zebravinkpopulaties onderzocht. Het doel was om te zien of en hoe klanken en structuur (de volgorde van klanken) verschilden tussen populaties. In tegenstelling tot veel eerdere studies gebruikten we een relatief objectieve methode van vergelijken en analyseren door gebruik te maken van automatische computationele analyses. Er zijn geen sterke aanwijzingen gevonden voor universele patronen in de structuur (elementvolgorde) van liedjes uit verschillende populaties. Binnen sommige populaties was er wel aanwijzing voor enige overlap in elementvolgorde, maar dit gold niet voor iedere populatie. Daarnaast wijzen de resultaten erop dat een van de typen elementen vaker voorkomt aan het einde van een liedje dan aan het begin. De relatieve frequentie van de verschillende typen elementen verschilde wel tussen populaties. Ieder type element werd in iedere populatie gevonden, maar de frequentie van sommige typen elementen was in de ene populatie hoger dan de andere.

Conclusie

De bevindingen in dit proefschrift laten zien dat zowel ervaring als ervaringsonafhankelijke processen zoals selectiviteit een rol kunnen spelen bij vocaal leren. Zowel in baby's als in jonge zebravinken is er vroeg in de ontwikkeling sprake van voorkeuren die onafhankelijk zijn van zang of spraak van de ouders. In de loop van de ontwikkeling veranderen de voorkeuren in beide soorten en bij zebravinken is dit duidelijk gerelateerd aan (auditieve) ervaring. Daarnaast is zowel bij vinken als bij mensen een verband tussen de vroege perceptuele voorkeuren en de latere productie te zien. Dezelfde voorkeuren en dezelfde verandering tijdens ontwikkeling zijn waargenomen in perceptie bij baby's en vroege taalproductie van kinderen. Bij zebravinken blijken zowel de vroege selectiviteit als het latere leren van een liedje een invloed te hebben op hoe goed ze het liedje van de tutor kopiëren. Daarnaast geeft dit proefschrift meer inzicht in de variatie en uniformiteit van zebravinkenzang, wat van belang is voor het voortzetten van 'fonologisch' onderzoek bij zebravinken en de parallel met mensentaal. De bevindingen in dit proefschrift laten een nieuwe kant zien van de vergelijking tussen vocaal leren bij mensen en bij zangvogels en dat hoogstwaarschijnlijk dezelfde processen van belang zijn voor leren.

ACKNOWLEDGEMENTS

I would like to thank all the birds, babies and their parents for participating in the studies in this thesis. Without them this work would not have been possible. In addition to those mentioned in the chapters' acknowledgements there are several other people who have contributed to this work. First of all, I would like to thank my supervisors Carel ten Cate and Clara Levelt. Thank you for your tremendous effort and advice and for your patience with the project and with me. I would also like to thank all my coauthors and specifically Rob Lachlan for his very important contributions to Chapter 4 and for his hospitality for Caroline and me during our visit at Duke University. The different bachelor and master students who have helped me with several projects each made important contributions. Therefore I would like to thank Desiree, Samantha, Wiebke, Jacqueline, Esben, Koen, Eline, Anne-Marie, Chang, Patricia and Annelieke. In this interdisciplinary working environment there are many colleagues to thank, within LIBC, LUCL, IBL and psychology thank you all for the feedback and great discussions. Most importantly I would like to thank the ones I worked with directly. From behavioral biology I would like to thank Caroline, Verena, Erwin, Wouter, Errol, Jiani, Saeed, Hans, Katharina, Hennie, Niels, Aukje, Peter en Sabine. From the babylab and linguistics I would like to thank Monique, Mirjam, Renee, Marijn, Kathrin, Margarita, Caroline, Simone, Annika, Stef, Szilvia, Kalinka and the members of the phonological reading group. Thijs, Jos, Rinus, Frits, Rob and Ap thank you for your valuable technical support and Barbara, Erwin, Caroline and Harrald for statistical advice. I also highly appreciated the mind boggling discussions during the language and evolution reading group with Marijn, Marieke, Carel, Jelle, Arie and occasional others. My family and friends, thank you for your mental support and the necessary distractions. A special thanks to my brother Gijs and colleague and friend Caroline for acting as my paranymphs. Most of all, thank you Chris! Thank you for your scientific and non-scientific advice, for being there for me in so many ways and for keeping me sane with jokes, music, surprises, wisdom and everything. Thank you for being so brilliant and yet never brag about it.

CURRICULUM VITAE

Born 4th of June 1982 in Schiedam, the Netherlands

1994 - 2001 VWO Schravenlant, Schiedam

2001 - 2005: Bachelor's Degree in Biology, Utrecht University

2005 - 2007: Master's Degree in Animal Biology, Utrecht University

Research project at the Behavioural Biology department of Utrecht University:
Sex differences in perceptual processing of long calls in the zebra finch (*Teaniopygia guttata*).
Supervised by dr. Sharon M.H. Gobes and Prof. dr. Johan J. Bolhuis

Research project at Music Cognition research group at the Max Planck Institute for Human Cognitive and Brain Sciences, Leipzig:
Auditory perception of emotionally salient acoustic features: an fMRI study
Supervised by dr. Daniel Mietchen and Prof. dr. Stefan Koelsch

2007 - 2012: PhD research – Leiden University, Behavioural Biology, Institute of Biology Leiden (IBL), Leiden University Centre for Linguistics (LUCL), Leiden Institute for Brain and Cognition (LIBC).

Vocal development in songbirds and human infants.

Promotor: Prof. dr. Carel J. ten Cate
Copromotor: Dr. Clara C. Levelt

The LIBC provided the grant for this project.

I received funding from the LUF (Leids Universiteits Fonds) to attend the summer school Linguistic Institute (Linguistic Society of America, LSA)

At the 12th congress of the International Association for the Study of Child Language (IASCL) I received the poster award

Sept 2011 – Dec 2011: Research Assistant Babylab Leiden, Leiden University.

June 2012 – July 2012: Animal care taker Leiden University.

October 2012 – December 2012: Language teacher at Sagenn

December 2012 - present: post-doc at the University of Liege, GIGA neurosciences

The influence of hormones on neural plasticity in developing songbirds

PUBLICATIONS

Ter Haar, S.M., ten Cate, C., Levelt, C.C. (in preparation). The role of frequency and phonology in the perceptual attention of nine- and twelve-month-old infants.

Ter Haar, S.M., Levelt, C.C., ten Cate, C. (in preparation). The interplay of perceptual predispositions and experience on song preferences and development in zebra finches (taeniopygia guttata).

Ter Haar, S.M., Van Heijningen C.A.A., Lachlan, R.F., C.C. Levelt, ten Cate, C. (in preparation). Zebra finch song phonology across populations and continents.

Ter Haar, S.M., Van Heijningen C.A.A., Lachlan, R.F., C.C. Levelt, ten Cate, C. (in preparation). Zebra finch song phonological syntax across populations and continents.

Gobes, S.M.H., ter Haar, S.M., Vignal, C., Vergne, A.L., Mathevon, N. & Bolhuis, J.J. (2009). Differential responsiveness in Brain and Behavior to Sexually Dimorphic Long Calls in Male and Female Zebra Finches. Journal of Comparative Neurology 516: 312–320.