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

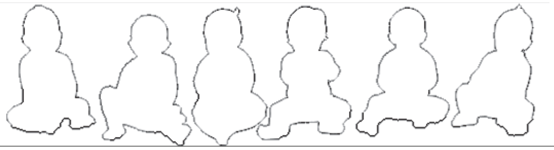
2 The role of frequency and phonology in the perceptual attention of nine- and twelve-month-old infants.

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

	un-marked						**marked**					
PoA	coronal	m	hom	PoA (C)	vow	voic	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	voice-less	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	**un-marked**	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	**un-marked**	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 ($F(1,32) =11.14$, $p= .002$, $ŋ^2 = .258$, fig. 2.1) and a marginally significant effect for stimulus voice ($F(1,32) = 4.20$, $p= .049$, $ŋ^2 = .12$)) 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 ($F(1,32) = 24.61$, $p< .001$, $ŋ^2 = .44$), 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.

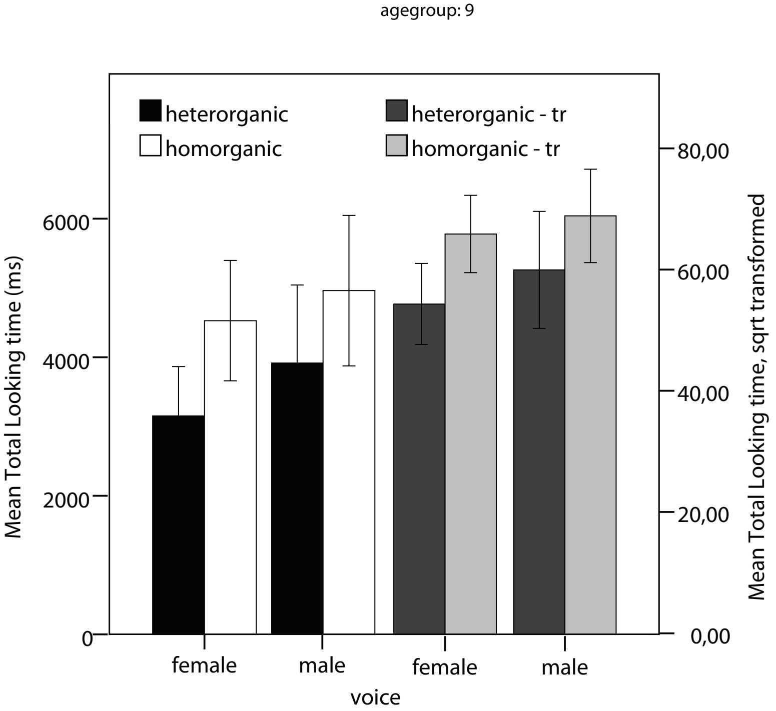


Fig. 2.1 Results for homorganicity in nine-month-olds for stimuli set spoken by different voices. Mean of raw and square-root transformed TL data for homorganicity for nine-month-olds listening to female voice stimuli (left) and male voice stimuli (right). White bars indicate raw TL for homorganic stimuli 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: $F(1,51)=24.27$, $p<0.001$, $\eta^2=.322$, LL: $F=25.35$, $p<0.001$, $\eta^2= .332$). 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

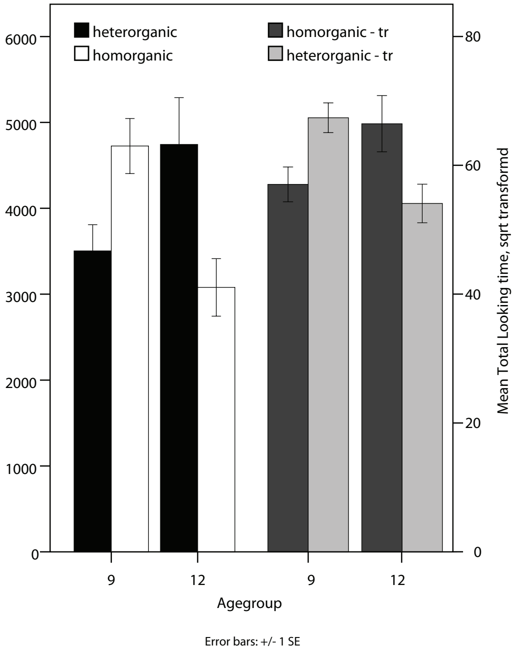


Fig. 2.2 Results for homorganicity in nine- and twelve-month-olds. Mean of raw and sqrt transformed Total Looking time (TL) data for homorganicity for each age group. White bars indicate raw TL data for homorganic stimuli and black bars indicate raw TL for heterorganic stimuli. Light grey bars indicate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