Phonetic and phonological development of stress in German

Von der Philosophisch-Historischen Fakultät der Universität Stuttgart
zur Erlangung der Würde eines Doktors der Philosophie (Dr. phil.)
genehmigte Abhandlung

Vorgelegt von

Britta Lintfert

aus Waiblingen

Hauptberichter:	Prof. Dr. Bernd Möbius
1. Mitberichter:	Prof. Dr. Grzegorz Dogil
2. Mitberichter:	Prof. Yvan Rose, PhD

Tag der mündlichen Prüfung: 14.12.2009

Institut für Maschinelle Sprachverarbeitung
Universität Stuttgart
2010

Danke!

Mein herzlicher Dank geht an die vielen Menschen, die mich die letzten Jahre während der Entstehung dieser Dissertation unterstützt haben:

Bernd Möbius für die sehr angenehme Zusammenarbeit der letzten Jahre, dessen Resultat nun auch diese Dissertation ist. Das Enstehen dieser Dissertation war verbunden mit zahlreichen konstruktiven Kritiken von seiner Seite und einer tollen Betreuung wie sie besser nicht sein konnte. Danke auch für die Anregungen für neue Projekte, sowie für fachliche und sonstige Gespräche.

Grzegorz Dogil danke ich dafür, dass er sich als zweiter Gutachter zur Verfügung gestellt hat. Außerdem dafür, dass er mein Interesse an der Phonetik, insbesondere der Prosodie und des Spracherwerbs, geweckt hat. Und für die Möglichkeit, ein neues Forschungsfeld an seinem Institut zu eröffnen.

Yvan Rose for being a great external assesor with lots of comments on fearlierr versions of this thesis. Thanks for travelling to Stuttgart for my oral defence and for a great time together. And for developing a great tool: PHON!

Ein großer Dank auch an **Wolfgang Wokurek** für die fachliche Beratung bezüglich Aufnahmetechniken bei Kindern zu Hause, akustischen Analysen, Vokalqualität und Statistik. Außerdem Danke für die kritischen Anmerkungen zum Methodenkapitel.

Dieser Arbeit entstand im Rahmen des DFG-Projektes **Ein exemplartheoretisches Modell zum Erwerb der akustischen Korrelate der Betonung** (MO 579/1-1). Dank der Unterstützung der DFG konnten wir in Stuttgart einen umfangreichen Kindersprachkorpus für prosodischen Analysen aufbauen. Innerhalb diese Projektes durfte ich mit **Katrin Schneider** und **Ursula Vollmer** zusammenarbeiten. Mein besonderer Dank gilt Ursula, ohne deren Hilfe ich die Vokalellipsen und deren Flächenberechnung wohl nicht so anschaulich hinbekommen hätte. Außerdem Danke an Ines, Felix und Steven für ihr unermüdliches Labeln von Kindersprache und an Manuel und Bernd für das Programmieren der Analysetools.

Ein riesengroßes **Danke** auch an die Kinder und deren Eltern, die über drei Jahre hinweg geduldig meine Besuche und Aufnahmen mitgemacht haben. Ohne die Bereitschaft der Kinder und deren Eltern und deren Geduld wäre ein toller Korpus und somit auch diese Arbeit nicht entstanden.

Für die Unterstützung der **Phonetikgruppe am IMS** möchte ich mich ebenso bedanken bei: **Sabine Dieterle**, der guten Seele des Lehrstuhls, bei **Kati Schweitzer** für ihre gute Laune, bei **Mike Walsh** und **Petra Augurzky** für das Korrekturlesen sowie bei **Antje Schweitzer** für eine tolle Zusammenarbeit.

Und natürlich bei meinen Zimmergenossinnen der letzten Jahren: Bei **Silvia Lipski**, die mich schon im Studium begleitet hat. Bei **Giusy Rota**, meinem italienischen Sonnenschein. Und bei **Uta Benner**, die mich in den letzten zwei Jahren sehr toll unterhalten und mit großer Bereitschaft alles von mir Geschriebene korrigiert hat.

Danke auch an meinen Bruder mit Familie und an all meine Freunde. Ihr habt mich in den letzten Jahren immer unterstützt und netterweise nicht so oft nachgefragt, wann ich denn nun endlich fertig bin. Dafür wart ihr immer bereit, meine Kinder (und meinen Mann) zu unterhalten, wenn ich mal wieder keine Zeit hatte.

Ein besonders großes **Danke** geht an **Dorothée**, die in all den Jahren mit großer Geduld und Sorgfalt alle von mir bisher auf Englisch geschriebenen Artikel und Kapitel korrigiert und verbessert hat. Ohne sie wäre diese Arbeit nie soweit gekommen.

Ohne die liebevolle Unterstützung meiner Eltern und deren festen Glauben an mich wäre ich nicht fähig gewesen, diese Arbeit zu vollenden. Nun ist es vollbracht! Danke für alles!

Ein großes **Danke** auch an **Peter**, **Lena** und **Emma** für ihre Liebe und Geduld, vor allem, wenn ich mal wieder das ganze Wochenende durchgearbeitet habe, nicht ansprechbar war und keine Zeit zum Spielen hatte.

Contents

List of Figures

List of Tables

1 Introduction

1.1 Motivation

Speech is one of the most important keys for communication with the environment. Infants acquire their native language with astonishing speed and ease. Even by the end of their first year of life children begin to produce their first words without any explicit instruction. Before meaningful speech is produced, infants are sensitive to the frequency of occurrence of the sound patterns in the native language (Jusczyk et al., 1994). The acquisition of the sound structure of the native language is the basis for a successful acquisition of the native language (Peperkamp, 2003).

From birth the infant sensory system continuously records speech inputs (Vihman et al., 2004) and infants are able to encode the salient properties of the speech signal (Pierrehumbert, 2003a). Each perceived lexical representation contains language–specific echoic traces including the acoustic properties and voice quality and leaves a unique memory trace (Goldinger, 1997, 1998). The persistence of this trace over time is however a matter of debate. On the basis of their sensitivity to the statistical distribution of the patterns in the input language (Saffran et al., 1996; Maye et al., 2002; Anderson et al., 2003), with increasing age these remembered token are organised as enriched representations in different categories. A cognitive map is built up in such a way that highly similar instances are closer together than dissimilar ones (Maye et al., 2008).

This cognitive map consequently expresses the actual range of variation in the physical realisation of each category and has to be developed during speech acquisition (Pierrehumbert, 2001, 2003a). In the first months of life the discrimination of speech sounds seems to be universal as the infants have not gained sufficient experience with the phonotactic and phonetic distribution of the native language. Effects of experience can be observed with a functional reorganisation of native contrasts and the development of language–specific categories and the decline of nonnative contrasts (e.g. Werker and Tees, 1984; Kuhl et al., 1992; Werker and Polka, 1993; Bosch and Sebastián-Gallés, 2003; Kuhl et al., 2006).

Stress plays an important role for the segmentation of the speech stream and infants pay attention to stress in the input and use this information to posit initial word boundaries (Jusczyk et al., 1993a; Echols, 2001; Christophe et al., 2003a; Thiessen and Saffran, 2003, 2007). For infants learning metrical stress languages, like English, it is assumed that stress information in the native language not only shapes the representational landscape used by infants in segmentation, but that this information is also encoded in the representations of parsed sequences (Lindfield et al., 1999; Saffran et al., 1999; Curtin, 2002; Thiessen and Saffran, 2004; Curtin et al., 2005). Even after several months of exposure to the target language, in the infants' perceptual space stressed and unstressed syllables are represented in different ways (Curtin et al., 2005). But in languages with lexical stress a kind of "deafness" for stress seemed to exist (Dupoux and Peperkamp, 2001; Peperkamp and Dupoux, 2002; Peperkamp, 2004). Peperkamp and Dupoux conclude for this case that no phonological encoding for stress is built up as stress is encoded in the lexical lemma. But this is subject to much methodological and theoretical controversy.

Based on this findings the phonological and phonetic development of stress is analysed. Therefore, the frequency of occurrence of metrical patterns and syllable structures as well as the acoustic realisation of stress is described. A goal of this dissertation is to explain the variability and developmental pattern observed in the acquisition of stress and to develop an exemplar–based model of acquisition of stress in German.

1.2 Research questions

The aim of this dissertation is the development of an exemplar–based model of acquisition of acoustic correlates of stress in German.

In an exemplar–based model production and perception are closely linked to each other in a perception–production loop. Based on the assumption of a perceptual space in long–term memory, all categories of a language are arranged as clouds of perceived exemplars. During the perception of language the encoded phonetic parameters of the input localise this classification in the parameter space. During production a speaker has access to previously stored enriched exemplars. These enriched exemplars serve as references for production. To develop an exemplar–based model of the acquisition of stress, the effect of the stress on the syllable structure as well as on the acoustic realisation of stress should be examined.

This thesis investigates when and how children start realising lexical stress in German, and whether they tend to produce stress by actively using the same acoustic correlates of stress as their parents do. In order to create an exemplar–based model of the acquisition of stress, it is assumed that there would be a gradual development of representations for stressed and unstressed syllables. In word production the children adopt the features that their parents use to indicate lexical stress. They probably start with the most salient parameter they can find. Parameters that need a closer analysis of the speech sample will be perceived later and are then also used to indicate stress. During babbling different prominence can also be produced but the children's productions are fairly limited by their motor abilities even as their perception of the different acoustic cues is initially unconstrained. In the pre–linguistic stage no preference for a special acoustic parameter of stress can be found due to the poor ability of motor control.

The following research questions are addressed in this thesis for develop an exemplar–based model of the acquisition of stress in German:

1. investigate phonological development by analysing the acquisition of metrical and syllabic structures. Explore the importance of proto–syllables which occur during the babbling phase with respect to their role in word production.

2. investigate the phonetic development of stress by analysing the acquisition of the acoustic correlates of stress. Compare the findings with the adult use of the parameters.

3. describe the development of vowel spaces for stressed and unstressed vowels.

1.3 Corpus and Method

Longitudinal speech samples from children from 5 months up to 36 months of age are recorded and analysed to create a prosodically annotated speech corpus of babbling, first words and meaningful speech from German speaking children. In Chapter 3.1 a detailed description of the corpus and the individual participants is given. The parents are recorded and annotated as well.

The obtained prosodically annotated speech corpus containing longitudinal data as well as parental speech is the basis for the conducted analyses. First of all the phonological development of stress is analysed. Therefore the produced metrical structures as well as the development of different syllable structures depending on stress between 5 and 36 months of age is described.

First the crucial role of stress in the development of different syllable structures is described. Then the acoustic realisation of stress is analysed. Therefore, the correlates of German word stress are analysed for children between 5 and 36 months of age and their parents.

To illustrate a cognitive map of representations in the perceptual space the development of the F1–F2 and F2–F3 vowel space depending on stress is described. The development of the articulatory–acoustic relationship seems to be nonlinear and depending on an increased control of motor abilities.

The results of the different analyses are integrated in a neural model of speech production and speech acquisition following the ideas of Guenther's neural model of sensorimotor control of speech production (Guenther, 1994, 2003).

1.4 Overview

The goal of this thesis is to describe the development of stress in child languages and its evolution over the first few years of life. Within the results of the conducted analyses an exemplar–based speech production model of the acquisition of stress should be developed. To this end the phonological and phonetic development of stress is analysed as well as the development of vowel space. The development of different perceptual spaces is addressed in the context of stress development in German children.

Chapter 2 gives an overview of recent research and an introduction to exemplar theory and its use in language acquisition.

In Chapter 3 the speech corpus, participants, annotation and the used automatic analysing tools are described in detail. The individual analysed participants are introduced and an overview of the period of recordings and number of analysed tokens are given.

In Chapter 4 the main focus lies on describing the phonological development of stress. Therefore the prosodic development of stress as well as the development of syllable types depending on stress is analysed for the children between 5 and 36 months of age.

In Chapter 5 the phonetic development of stress is described. The main focus of the analysis lies on the acoustic realisation of stress. The correlates of German stress, vowel duration, intensity, fundamental frequency, first four formants and vowel quality are analysed for children and parents. The results are presented for each participant separately as well as the child compared to their recorded parents.

In Chapter 6 the development of vowel spaces for stressed and unstressed vowels is described. For this analysis further cross–sectional data for children between 5 and 10 years of age are also used. Therefore the development from first centralised 'vocalic' segments during babbling to the vowel space of an adult speaker is described.

In Chapter 7 the general discussion of the results and the attempt to build up an exemplar–based speech production model based on the acquisition of stress concludes this thesis.

2 Background

2.1 Introduction

How do infants learn language? Two fundamental questions in human language research pertain to which language–specific features are perceived at which age and from which age children can produce what they have perceived. First of all, infants have to learn to identify the units of sounds that comprise their language. Then they have to learn to produce the sounds. An interaction between prosody and statistics in the segmentation of fluent speech suggests that prosody acts as a filter to find out possible word–like sequences in the speech stream (Shukla et al., 2007). The infants have to build up phonetic representations of a language–specific set of sound categories. The development and organisation of the acoustic–phonetic space depending on speech perception are essential for learning to speak.

The ability to segment fluent speech into words emerges during the first year of life (for an introduction see Jusczyk, 1996). Before infants can begin to identify the syntactic structure of sentences they have to identify individual words in the speech stream. The segmentation of the speech stream into words plays a central role in the acquisition of speech. Although there are no clear–cut pauses in the speech stream the child must segment the stream of continuous speech to discover the individual words. From word segmentation studies (Jusczyk et al., 1994; Echols and Marti, 2004; Christophe et al., 2003b; Maye et al., 2008; Nazzi et al., 2008) we know that infants are sensitive to the relative frequencies of different phonemes and segments in the native language and are able to use this information to parse sequences of speech. Infants use this knowledge to develop a cognitive map comprising many dimensions. Furthermore before meaningful speech is produced, infants are sensitive to the frequency of occurrence of the stress patterns in the native language (Jusczyk et al., 1994), and they are able to encode stress information and build up cognitive representations and categorisation of these patterns. Each perceived lexical representation contains language–specific echoic traces including the acoustic properties and voice quality and leaves a unique memory trace in the cognitive map (Payne, 2006). This cognitive map expresses the actual range of variation in the physical realisation of that category (Pierrehumbert, 2001). These findings support the possibility that stressed syllables are represented in a different perceptual space than their unstressed counterparts and that children learn stress on the basis of exemplars stored in memory and not on the basis of rules (Daelemans et al., 1994; Gillis et al., 2000). This supports an exemplar–based theory of representations (Goldinger, 1998), in which details from the input are retained in memory for each token.

Exemplar Theory (Nosofsky, 1986; Lacerda, 1995; Kirchner, 1999; Pierrehumbert, 2001) assumes that speech perception and production are closely linked to each other in a perception–production loop. The similarity of the input to the various cohorts of exemplars within the lexicon helps to distinguish phonemes or lexical items more than invariant phonetic properties. In this point of view each perceived lexical representation contains language–specific echoic traces including the acoustic properties and voice quality and leaves an unique memory trace in the cognitive map. Aspects of vowel perception, sex identification and speaker variability (Johnson, 1997), duration of phones (Fougeron and Keating, 1997; Cho and Keating, 2009), context and frequency effects (Wade et al., 2010), place of articulation (Friedrich et al., 2008), and talker identification (Perrachione et al., 2010) are known to be encoded in the perceived representations and stored in the mental lexicon. E.g. duration of phones showed to vary consistently as a function of the segment's position in a prosodic hierarchy (Fougeron and Keating, 1997; Cho and Keating, 2009). In this point of view an exemplar model would entail that longer consonants are connected with consonant–initial words and shorter consonants occur more in (re–syllabified) coda consonants. New exemplars are classified according to their similarity to stored exemplars. Percepts of speech events are stored in memory as exemplars with fairly long stretches in a perceptual space. This space can be represented as a cognitive map comprising many dimensions, which code the phonetic and phonological properties of the exemplars. Percepts of nearly identical instances are located on the map in close vicinity to each other, whereas percepts of less similar instances are located in different regions. Thus, perceived realisations of speech events form clouds of exemplars on the map and are linked to the context in which they originally occurred. These exemplar clouds represent the categories of a given language. Within each category the distribution of exemplars indicates the range of variation of the parameters that characterise the respective category. The optimal location of an exemplar prototype (assuming there is one) does not have to be represented by an existing exemplar token as the exemplar locations may represent idealised, abstract prototypes. So the phonetic inventory of a language is a set of perceived probability distributions over the phonetic space. From this point of view speech categories do not rely on abstract rules and processes but are best described with fine phonetic details. Therefore categories of speech are represented by encountered instances of the categories and specified in detail. This echoic memory is available from the earliest infancy (Vihman et al., 2004, 2006), although segmental representation is not (for arguments of echoic traces see also Jusczyk, 1997). This means that children have enriched lexical representations and stress, especially its acoustic correlates, is a prominent cue for parsing out potential words (Lindfield et al., 1999; Thiessen and Saffran, 2003, 2007) The role of stress is particularly relevant to prosodic development since early truncated productions are determined by syllable salience.

But how and when do infants learn to produce these different representations? We know there is a mismatch between perception and production. Even newborn infants have a sufficient sampling of exemplars to make language–specific generalisations about the distribution of values in the auditory dimension of the phonetic space (Vihman et al., 1998; Beckman, 2003; Nazzi et al., 2006; Burns et al., 2007; Thiessen and Saffran, 2007). This means that with the beginning of perception during fetal life at about 26 weeks, infants begin to establish language–specific exemplars in their memory (Vihman, 2002). The production of these exemplars depends on the development of motor abilities and follows with a lag.

In the process of acquisition, first, infants have to develop a cognitive map comprising numerous dimensions. Each perceived exemplar is stored as enriched representation in memory and based on frequency effects different categories are built–up. For the process of speech production, the stored exemplars serves as plans of articulation and the frequency of the exemplars plays an important role. A multi–level exemplar–based process seemed to be necessary for the selection of the best exemplar for production (Schweitzer and Möbius, 2004; Walsh et al., 2010). For a successful production first the necessary categories based on a critical number of perceived exemplars has to be established. Given the perception–production loop, the child's ability to produce a new category depends on the number and quality of input elements and the development of motor abilities. For speech production an output plan of neuromuscular command to the vocal tract has to be computed for an input representation (Kirchner, 1999). With the beginning of production each perceived exemplar which has formed a perceptual representation in the memory has now connected to the articulatory–acoustic gesture for a successful production.

Babbling seems to be one key mechanism that permits babies to discover and produce the perceived structures of the native language and to connect them to articulatory–acoustic gestures. With the beginning of meaningful speech, children can produce CV–syllables using their proto–syllables which they have learnt and produced in the babbling phase. They begin to align their productions with the prevailing values heard in the input (de Boysson-Bardies, 1999; Vihman and Velleman, 2000). With the beginning of word production, context plays more and more a role for the selection of the best exemplar. So even if the speaker has previously produced a few tokens of this particular word, these particular tokens may be an inappropriate plan for pronunciation in the current context. For production not only a simple semantic–to–articulation adaption is necessary but a systematic knowledge of the mapping between perceptual targets and their articulatory realisations. These mappings has to be learnt. A continuous development of control and coordination of particular structures in producing sound sequences based on self–feedback during babbling is necessary. Frequency of occurrence and frequency of experience is thus a crucial factor in the development of speech. In this point of view, the production lag may depends on the critical threshold to build a new category with exemplars marked for their position in a prosodic hierarchy. An organisational hierarchy is involved in the competition between the different units at neighboring levels (Walsh et al., 2010). For a good production the similarity of the original context with the relevant neighboring sound in the current production has to be taken into account and learnt by each child during babbling and first word production.

In word production the perception–production loop helps a child to perceive significant differences between his representations of exemplars and the representational exemplars of the target word. For the production of more adult like exemplars a reorganisation of the built–up representations takes place (Dogil, 2010).

2.2 Development of speech

Children must acquire knowledge about the sound structure of their native language. They acquire this knowledge from the linguistic input of their environment (a detailed description of this development can be found in van de Weijer, 1998). A representation of the sounds has to be developed via the acoustic input (Christophe et al., 2003b). However the representations of the sounds vary – depending on speaker, speech rate or phonetic context – in their acoustic realisation *(phonemic variability)*. The listener can, and must, assign the different realisations of a sound to the same category or to the appropriate category *(phonemic constance)*.

Infants in their first year of life acquire the ability to distinguish the native language from other languages with the help of their phonetic (Jusczyk et al., 1993b; Friederici and Wessels, 1993) and phonotactic structures (Jusczyk et al., 1993a; Nazzi et al., 2000). In their first months of life babies are able to distinguish a big span of vowel contrasts and consonant contrasts, including contrasts not relevant for their mother tongue. From birth children possess the ability to perceive contrasts between two phonetic categories better than differences within a phonetic category. The ability to discriminate non–native consonant contrasts get lost between the eighth and tenth month of life (Werker and Tees, 1984). During that time the perception and discrimination of vowel contrasts (Werker and Polka, 1993) develops.

Long before children are able to produce their first sounds, they can distinguish speech sounds and categories. In fact, speech perception begins during the last trimester of intrauterine development when the auditive system is functional. The foetus extracts invariant patterns from the complex auditory input that is filtered through the amniotic fluid. The amniotic fluid filters out high frequency components but preserves prosodic characteristics like the rhythmic properties of the native language (Locke, 1993; Lecanuet and Schaal, 2002). The foetus learns to recognise the melody and rhythms of language, as intonation contours and stress patterns, and becomes familiar with its native language and its prosodic structure. So newborn children can already recognise their native language with the help of their prosody (Christophe et al., 2003a) learnt during gestation. Prosody helps them to collect information about the structure of their native language. And with the help of this quite early acquired information of the tonal and temporal structure of the mother tongue they can parse the linguistic signal into segments (*bootstrapping from the signal*), acquire these segments after time and produce it (Echols, 2001).

In contrast to perception which begins before birth, production takes longer to develop. Within the first year production can be divided into different phases (Vihman, 1996; de Boysson-Bardies, 1999) depending on age and on the development of the child's vocal tract (Kent and Miolo, 1995). The vocal tract of the newborn is not simply a miniature of the adults but is characterised by a proportionally shorter pharynx, with a relatively larger oral cavity. As the mass of the tongue is situated more in the front the movements are limited. The larynx lies directly at the soft palate and seals the nasal cavities. At this time the major task of the larynx is closing the trachea to prevent a suffocation while drinking and the vocal tract does not allow the infant to produce articulate sounds (Kent and Murray, 1982). Besides the infant can not control its breathing, which feeds the production of sounds. This stage is called pre–babbling phase (Kent and Miolo, 1995). With increasing age the vocal tract length increase and the palate becomes lowered and moves forward, also the tongue body is moved backwards and lower (Vorperian et al., 2005). The control of breathing, necessary for phonation increase, and at about five months, when babies are capable of controlled breathing, the first vocalisations with opening and closing the mouth occur. Now the so–called babbling phase begins. Words are generally first produced between twelve and twenty months of age (Vihman, 1996; de Boysson-Bardies, 1999).

A large discrepancy between what an infant can perceive and what it can produce can be observed. This observation actually raises a number of questions, some of which are addressed in the subsequent chapters, mainly what are the constraints on prosodic acquisition? The application of prosodic structure applies to contexts larger than just one syllable. Hence pre–linguistic infants require perceptual sensitivity to the elements and domain of prosody and translate these elements into fine motor adjustments over more than one syllable.

However, the development of production of the different possible prosodic correlates of stress in German has not been examined. The significance of prosody for language acquisition has not been examined sufficiently either. Is the major constraint perceptual (detecting the domain and parameters of prosody), or physiological (coordination of the articulatory apparatus), or a combination of the two?

The choice of the speech sounds during the babbling phase depend on the articulatory abilities as well as on the nature of the toddler's linguistic experience (MacNeilage, 1997). Outward perception, self–perception by means of feedback mechanisms and the production of sounds are linked with each other (Vihman, 1996). Children learn to move their articulatory organs in a certain way while listening to their own articulation and to realise therefore different sound changes. Besides the common phonemes, words and word orders of their first language, they have to understand the prosodic structure and domain of the language.

The acquisition of different phonemic contrasts is very different across individuals within a language (Vihman et al., 1985). Nevertheless, a general rule seems to be that the phonemes which require a very precise temporal sequence from glottal and supra–glottal movements (e.g. aspirated or glottalised plosives) are acquired very late. Such phonemes where a precise positioning of the articulators without tactile self–control (fricative and liquids – in contrast to plosives) is important, are also acquired very late (Menn and Stoel-Gammon, 1995). The temporal sequence of the acquisition of the phonemic contrasts depends on the development and control of the articulation apparatus. But also the *frequency of occurrence* of sounds and sound patterns of the surrounding language plays a considerable role at this early time of linguistic perception and production. With the help of these frequencies the children must develop associative connections between the distinctive patterns of a phoneme and its phonotactic conditions in order to acquire the phoneme inventory of their language. This dependency on the ability to control the articulation apparatus and the frequency of a sound pattern is also expected in the acquisition of prosody, from the word stress up to entire discourse structure. However, from the production of lexical word stress up to the production of contrastive stress there are great temporal differences between the perception and the production of these sound patterns. All possible acoustic parameters can be used by the children to mark stress and for stressed and unstressed syllables different acoustic realisations has to be established (Atkinson-King, 1973; Allen and Hawkins, 1980; Pollock et al., 1993; Kehoe et al., 1995; Schwartz et al., 1996; Lee et al., 1999; Altmann and Kabak, 2000; Vogel and Raimy, 2002).

Learning to speak requires increasing control over the articulators. With increasing experience and practise, different articulatory–acoustic gestures are built up and gradually improved (Davis and MacNeilage, 1995). Even the child's earliest productions show an influence of frequency and are influenced by the phonetic and phonotactic structure of the native language (de Boysson-Bardies et al., 1989; de Boysson-Bardies and Vihman, 1991).

2.3 Phonological representations

Research about the prosodic structure, probabilistic phonotactic and allophonic variation showed that listeners are sensitive to the different phonological and acoustic–phonetic details in the acoustic signal and weighting the multiple segmentation strategies at multiple levels of linguistic organisation in a hierarchical framework (Mattys et al., 2005; Wade et al., 2010; Walsh et al., 2010). This hierarchy attempts to capture the fact that, though each cue presumably has an independent effect on the activation of lexical candidates, some cues trump others when multiple cues are available to the listener. This weighting fluctuates depending on the saliency or availability of other cues at any given point in the signal.

Prosody links semantic information, syntactic and morphological structure as well as segmental sequences into a consistent set of address frames like syllables, metrical feet, phonological word and intonation phrases in different levels of this prosodic hierarchy (Levelt, 1989) (see also Dogil (2003)). With this prosodic frame babies and toddlers can recognise the single segments of the speech stream during the acquisition

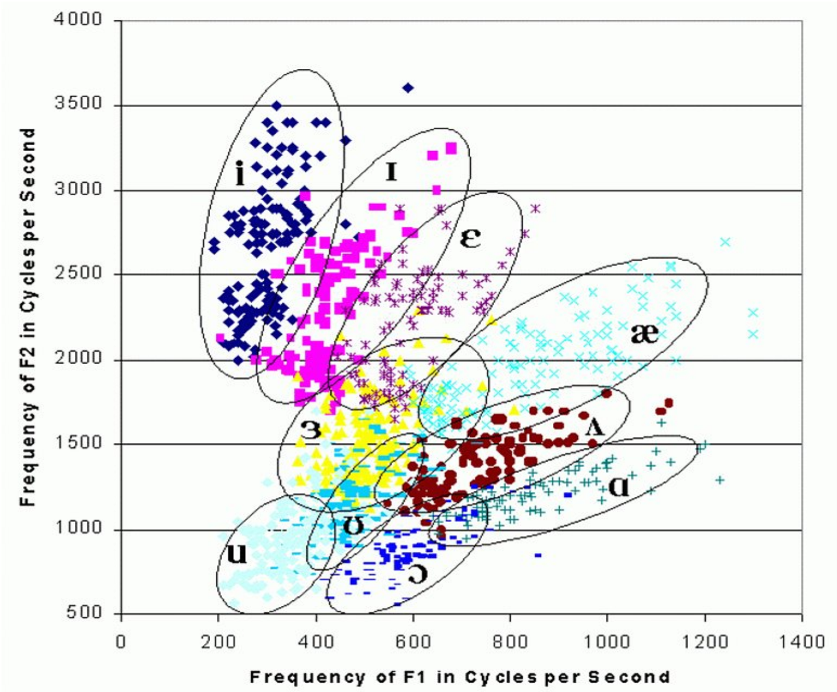


Figure 2.1: The F1–F2 vowel space. Data, taken from Peterson and Barney (1952) after (Pierrehumbert, 2003b)

process. They can select and integrate the information about the semantic, syntactic and morphological structure of the surrounding language and build up different levels of exemplars with enriched representations. The phonetic inventory of a language is a set of labelled probability distributions over the phonetic space. The acoustic and articulatory parametrisation of speech characterise this phonetic space. Vowels, for example, can be viewed as density distributions over F1 – F2 space, as shown in Figure 2.1. Each vowel occupies a continuous region of the space. Each vowel is more frequently instantiated by values near the centre of its distribution than by values near the edges of its distribution. These exemplar clouds represent the categories of a given language. Within each category the distribution of exemplars indicates the range of variation of the parameters. The parameters are linked to articulatory movements (e.g. tongue position, lip opening, etc.) that characterise the respective category.

With increasing control of the articulation apparatus vocalisation improves gradually. The articulatory movements must be linked to the acoustic signal. The more a child babbles, the more frequent are the opportunities to develop a good feedback loop which is necessary for the production and control of language. Within an exemplar–based theory of speech processing, it is proposed that sound categories are acquired bottom–up from the acoustic signal by means of statistical learning procedure. External feedback by a perception–production loop as well as internal phonological feedback helps to build up a mental representation of every single speech event in the memory (Pierrehumbert, 2003a,b). Percepts of speech events are stored in memory as exemplars in a perceptual space. This space can be represented as a cognitive map comprising many dimensions, which code the phonetic and phonological properties of the exemplars. Percepts of nearly identical instances are located on the map in close vicinity to each other, whereas percepts of less similar instances are located in different regions. The probability distributions over the phonetic space with the language specific realisation in the acoustic space have to be acquired to establish the phonetic encoding system of a language (Pierrehumbert, 2000; Pierrehumbert et al., 2000). Even when a certain similarity between two languages can be realised, the relationship of categorical label to the parametric level may not be the same in the two languages. During the acquisition of speech, phonetic and phonological categories are built up from perceived exemplars. The *frequency of occurrence* of the perceived exemplars plays an important role (Pierrehumbert, 2001) as a certain number of exemplars is necessary for an accurate generalisation and categorisation (Pierrehumbert, 2003a). In this point of view a category is defined as language–specific relations between discrete level and a parametric phonetic level. During speech acquisition the infant has to learn the probability distributions over the parametric phonetic space and incremental updating of the categories is critical for the children's speech performance. First infants only store the perceived exemplars in the memory without the ability to produce them. Therefore, initially, they do not produce everything they perceive (Fisher et al., 2004). This phenomenon is called *mismatch* or *lag* between perception and production during acquisition. The mismatch between perception and production during speech acquisition can be explained within an exemplar–based model of acquisition. Even newborn infants get enough input from their parents to have certain expectations of the prosodic structure of their native language (a good overview is given by Mehler et al., 2000). Based on their experience with a number of perceived exemplars they build up speech–specific distributions of internal models. They generalise over these distributions of the stored internal sound patterns to be able to extract meanings or syntactic demands of single words from the surrounding speech stream despite all variability (cf. Saffran et al., 1996; ten Bosch and Cranen, 2007).

In the traditional, abstractionist view, this variability of the speech is kept out of the lexicon via a process of normalisation to reduce storage requirements in the lexicon. Therefore the lexicon is built up with maximally underspecified representations (cf. Chomsky and Halle, 1968; Werker, 2000). Nevertheless studies of adult speech (Pierrehumbert, 2001) as well as language acquisition (Curtin, 2002; Beckman, 2003;

Fisher et al., 2004; Zamuner et al., 2004) suggest the possibility that the phonological representations which form the basis of the mental lexicon are not abstract but that context–sensitive fine phonetic–acoustic details are part of the representations. Detailed and context–sensitive representations of words help both to learn the language–specific phonetics and phonology and to learn to identify and compensate for variations in dialect, accents, and social register (Fisher et al., 2004). For the development of a lexicon the child has to establish a perceptual system and knowledge of general phonotactic rules.

Similar to the introduced model of short–term memory by Levelt (1989) for the understanding and production of words, Baddeley (1996) developed the concept of *Working Memory*. According to this model a working memory in the brain consists of a central executive, a visual–spatial part and a phonological loop. The phonological loop contains a phonological short–term memory, where the phonological information for a time span of 1 – 2 seconds is stored and connected to articulatory rehearsal component that can revive the memory traces. Any auditory verbal information is assumed to enter automatically into the phonological store. This activation depends on how the speech data were taken up (within a conversation, while reading, passive listening, etc.). During both perception and production the phonological loop is activated and within this activation the available articulatory representations are compared with the perceived realisation and the phonological loop is upgraded with this realisation. The phonological loop may play a key role in the acquisition of vocabulary, particularly in the early childhood years (Baddeley et al., 1998). However, even if the child has learnt to articulate certain phonemes and acoustic realisations within a word, the transfer of these articulatory gestures to other forms which contain this phoneme depends on the frequency with which this phoneme appears in the same phonetic surroundings (Beckman and Edwards, 2000).

2.4 The acquisition of prosodic structure

In order to understand phonological acquisition it is necessary to ascertain the type of information encoded in children's lexical representations. There is a controversial discussion in the acquisition literature whether the underlying representation of a word is quite close to the adult/target form (Smith, 1973; Kehoe and Stoel-Gammon, 1997) or initial representations are in fact fairly impoverished (Pater, 2004; Fikkert and Levelt, 2008). Growing evidence suggests that context–sensitive acoustic–phonetic details are part of these representations (cf. Curtin, 2002; Pierrehumbert, 2003a; Swingley, 2003; Burns et al., 2007; Shukla et al., 2007; Thiessen and Saffran, 2007; Zamuner, 2009). These enriched lexical representations give evidence for an exemplar (or episodic) theory of representation (Goldinger, 1997, 1998).

Speech production involves several cognitive processes, such as conceptual preparation, lexical access, word form encoding, and articulation. Based on the speech production model of Levelt et al. (1999b) (see also Levelt, 1989; Levelt and Wheeldon, 1994) the metrical pattern of a word consists of the number of syllables and the location of main stress. For frequently occurring syllables the underlying learnt motor programs are stored in a phonetic mental syllabary. The temporal domain for lexical stress encoding seems to be the time window identified for phonological encoding in language production (Schiller, 2006). Therefore phonological encoding is necessary for the production of metrical stress. The phonological encoding is part of the word form encoding, and can start right after the word form of a lexical item has been retrieved from the mental lexicon. For the production of metrical stress the stored stress patterns in the mental lexicon have to be activated and compared. The mental lexicon consists of perceived enriched representations of syllables and their metrical pattern. The acoustic speech signal has therefore first to be analysed by the peripheral auditory system where acoustic parameters and cues of metrical stress are extracted. The syllable structures based on metrical stress, as well as their different features, are estimated and matched by an internal analysis–by–synthesis process (Schneider et al., 2006).

To build up exemplars the listener has to perform a mapping between acoustic, articulatory, and feature–based representations and to build up abstract linguistic representations (Pierrehumbert, 2003a,b) explaining higher order phonological knowledge, such as knowledge of phonotactics. At the beginning of perception children begin to develop enriched representations in their mental lexicon. A category can be produced when enough exemplars are perceived. During the babbling phase the child produces, depending on its articulatory abilities, the most frequent sound patterns of the surrounding language (de Boysson-Bardies, 1999; Vihman and Velleman, 2000). In the phase in which the first words are produced together with babbling phrases, a continuous improvement of the articulation movements of consonants and a better control of the tongue position with the vowel production (Fikkert, 2005; Fikkert and Levelt, 2008) take place. In addition to syllabic duration, phonotactic structure must also be acquired and refined in this acquisition process (Vihman et al., 2006).

In her study on the mental representations of stress Curtin et al. (2005) showed that stress is encoded in the representation of an analysed sequence. Therefore stressed and unstressed syllables are organised differently. With 5– to 7–month–old children stress is apparently a part of their proto–lexical representation. This coincides with an exemplar–based model according to which exemplars are stored together with a feature and context description. Different prosodic representations have to be developed before word learning as these representations seem to be the frame for the acquisition of the syntactic and semantic information (Fikkert et al., 2005). The view that prosody is the frame for grammatical morphemes is supported by a study conducted by Demuth and Tremblay (2008). They studied the prosodically–conditioned variability in children's production of French determiners which supports the notion that grammatical morphemes first appear in prosodically licensed contexts, and suggest that some of the early variability in morphological production is systematic and predictable.

These different aspects of speech production are associated with the development of a variety of anatomical and neurophysical structures and functions of the vocal tract and nervous system and develop individually. It has been shown that during acquisition a mental lexicon has to be built up by an internal analysis by synthesis.

But how and when do infants build up these mental lexicon? And how is it trained and refined? To answer this question the development of syllable structure will be described. It is hypothesised that the mental syllabary is built up based on perceived representations. The production of syllables depends on the articulatory abilities learnt during babbling and has to be built up gradually based on the core syllable CV.

2.5 Neural mapping and mental syllabary

First linguistic phrases occur with the onset of babbling at about five months of age. Do these first 'syllables' build up the mental lexicon or is the purpose of babbling to train the articulatory apparatus in order to produce meaningful words?

Babbling seems to establish a neural network of auditory–articulatory mappings as articulatory movements produce acoustic signals that are feedback to the neural network (Guenther, 2001). In a given language each phoneme has a unique phonetic target region and this region has to be established during babbling (Perkell et al., 2000; Guenther, 2003) by mappings between abstract phonemes and their corresponding auditory targets. The production therefore is perception–oriented. A computational simulation of this view is embodied in the DIVA model (Directions Into Velocities of Articulations, Guenther (1995)). With the onset of babbling different motor speech patterns are stored as exemplars in a so–called *protosyllabary* (MacNeilage and Davis, 2000).

From this point of view babbling could be described as 'training' for the articulation and connects articulatory movements to acoustic realisations. For the production of babbling no phonological encoding is necessary. The output of the infants is influenced by the rhythmic mandibular cycle and therefore produced by random. The articulatory–acoustic gestures build up a first mental lexicon. In the so–called *protosyllabary* speech motor patterns are stored without any phonological encoding or lexical meaning. In contrast to babbling where no encoding is necessary, for the production of speech phonological encoding is necessary. With the onset of meaningful speech the infants' have to recognise a semantic relation between output and lexical concept. Therefore the child starts to connect some particular babble to some particular lexical concept (Levelt et al., 1999b) or generates utterances which are sufficiently speech–like to provoke a linguistic response from the environment (Messum, 2007; Howard and Messum, 2007). Word production emerges once the child has coupled two initially independent systems, a conceptual system and an articulatory motor system building the protosyllabary. With increasing word production the previously built up articulatory schemes in the protosyllabary are connected to phonological encoding and therefore with metrical patterns. The underlying output plan of neuromuscular commands to the vocal tract must also be mediated by the underlying phonological encoding.

In summary, during babbling infants have to learn the acoustic consequences of articulatory gestures as well as the auditory–perceptual consequences of acoustic properties of speech. Based on self–feedback a neural network of auditory–articulatory gestures has to be built up during the production of babble utterances. At the same time the infants also establish exemplars in the perceptual space based on internal synthesis of the surrounding input. With these exemplars they build up language–specific expectations on the input. With the beginning of meaningful speech the production process has to be enlarged. The main difference between babbles and words seems to be that for the production of words the mental syllabary and semantics have to be taken into account. A phonological encoding takes place and the child has to find out which phonemic mapping and which trained articulatory–acoustic targets fit best to produce an adult–like target. For the production of words therefore self–feedback by the perceptual loop as well as feedback by the environment is taken into account to enhance the production.

Based on these assumptions the following developmental patterns are described to shed light on the development of a mental lexicon as well as on the development of stress in German:

- to investigate the phonological development of stress, the development of metrical pattern, as well as the development of different syllable structures depending on stress, are analysed for children between 5 and 36 months of age.

- to describe the phonetic development of stress, the development of different acoustic parameters of German word stress (e.g. vowel length, loudness, pitch, formants, voice quality) are analysed for children between 5 and 36 months of age, as well as the parental use of these parameters.

- to describe the development of exemplars stored in a cognitive map in the perceptual space, the development of the vowels and their language specific probability distribution over the phonetic space is analysed for children between 5 months and 10 years of age, as well as for adult speakers.

3 Methods

3.1 The Stuttgart Child Language Corpus

The present study is based on longitudinal speech samples which were collected for a study on the acquisition of stress in German. For this project children from 6 months up to 15 years of age were recorded and analysed to create a speech corpus of babbling, first words, and meaningful speech from German speaking children. This corpus was annotated for prosody. The core of the project was to develop an exemplar-based model of stress acquisition in German. In Table 3.1 an overview of all children who took part in this study is given.

In this dissertation the focus lies on the acquisition data of the six youngest children (3 boys and 3 girls) of the described corpus. In Appendix A.1 an overview of all recordings for each of these six children between 5 and 36 months of age is provided. The parents were recorded and analysed as well. In Appendix A.2 an overview of the parental data is given.

3.1.1 Data collection

The parents and children who participated in this longitudinal and cross-sectional study were recruited by one of the investigators of the study in a small village near Stuttgart over a period of three years. The first recorded children (MS, TS, LL, and EL) are the children of the investigators themselves and therefore were involved in the study from the beginning. As HH, ED and OZ were well known to one of the investigators since their birth they were also involved in the study from a very early point. The other children were recruited from a local family initiative in February 2005 (BW, NB, FZ, RL, JB, PH). All recorded children live in monolingual German-speaking families and had no unusual prenatal, sensory or developmental concerns or hearing problems. Most of the children were cared for mainly by their mothers. However, in the cases of HH and BW their fathers took parental leave for the first two years and therefore shared child-care with the children's mothers.

At the age of 12 and 24 months the speech development of the infants was tested (parental questionnaire for early recognition of children at risk (Grimm and Doil, 2004)). The older children were tested at preschool age (Heidelberger preschool screening (Brunner et al., 2005)). For the data collection different recording tasks had to be developed depending on age.

Table 3.1: Profiles of the children and period of recordings.

child	date of birth	sex	siblings	first recordings	Period of recordings (year; months)
MS	08.09.98	male	firstborn younger brother (TS)	05/2002	3;9 – 8;0
CZ	07.01.99	male	firstborn younger brother (OZ)	01/2005	6;0 – 7;8
LL	23.03.99	female	firstborn younger sister (EL)	06/2002	3;3 – 7;6
TS	19.11.00	male	second born older brother (MS)	05/2002	1;6 – 5;1
ED	21.02.02	female	third born two older sisters (born 1998 and 2000)	05/2004	2;3 – 4;7
RL	21.02.02	male	firstborn no siblings	04/2005	3;2 – 4;7
JB	08.04.02	male	firstborn younger brother (born 2005)	04/2005	3;0 – 4;5
EL	22.06.02	female	second born older sister (LL)	11/2002	0;5 – 4;3
PH	06.07.03	male	firstborn no siblings	01/2005	1;6 – 3;2
HH	02.08.03	female	firstborn no siblings	11/2003	0;4 – 3;1
OZ	01.03.04	male	second born older brother (CZ)	11/2004	0;9 – 2;9
BW	22.04.04	male	firstborn younger brother (born 2006)	03/2005	0;11 – 2;11
NB	20.08.04	male	firstborn younger sister (born 2006)	04/2005	0;7 – 2;8
FZ	05.09.04	female	firstborn younger brother (born 2006)	09/2005	1;0 – 2;6

3.1.2 Emergence of speech

Between 5 and 18 months of age speech data from six children (3 boys, 3 girls) were collected. The infants were audio-recorded every 6 to 8 weeks starting between five and seven months of age, when CV-syllable productions first occur. When the children were older, at the word stage, they looked at and talked about a picture book, which was the same for all children at this age group.

In this phase the development of six children was monitored by recording their utterances from the early production phase at the age of about 5 months, until an approximate production of 100 words and the production of first sentences. These recordings include the babbling phase, the occurrences of first words, first one– and two–word sentences, and the development of first phonological rules.

Figure 3.1: Picture cards: *Pinguin* [ˈpɪŋguːiːn] *penguin*, *Banane* [baˈnaːnə] *banana*, *Krokodil* [krokoˈdiːl] *crocodile*.

3.1.3 Speech becomes language

Between 18 and 36 months of age speech data were collected every 6 to 8 weeks. For all children the recordings were carried out until their third birthday in the same manner as described for the babbling phase, i.e., playing with their parents (or older siblings) or looking at picture books (the same picture book was given to all children by the author).

In this phase word productions increased. Stress placement on words became more sophisticated, and the development of the phonological lexicon was prepared (de Boysson-Bardies, 1999). These recordings include the estimated production of 250 to 300 words and the production of first two- and three-word sentences.

With the occurrences of first words the children born 2004 (OZ, BW, NB, FZ) were additionally tested every recording session on the use of stress in their word truncations by representing two- and three-syllable German words with stress on the first, second and third syllable.

Therefore picture cards representing two– and three–syllable German words were given to the children (see Figure 3.1). The words contain stress on the first, second or third syllable. In Appendix A.4 a full list of the words is given.

With this task the development of phonological patterns can be tested. For example, the manner in which different stress patterns are realised and produced by the children can be analysed. Since the same pictures were always presented to the children, a good overview of their production of the target words with different stress and varying degrees of truncation were recorded. The acquisition of syllable structure and stress patterns can be analysed and compared with the adult productions in their phonological and phonetic representations.

3.1.4 Contrastive stress production

From the age of 36 months recordings were made each 10 to 12 weeks. According to the TAKI task design proposed by Allen (1981), we created five pairs of animal toys. The names within each pair differ only in the position of stress. These names were the target words in parts of the study and the participants were to use them to refer to the animal toys during the recordings. The target words (Appendix A.5) are bi– or trisyllabic and consist of consonant–vowel (CV) syllables. Only speech sounds that German children typically acquire first were used, e.g. the vowels /a i o/ and the consonants /b d m n/. The CV syllables are phonetically similar to the reduplicated babbling and to a child's first words. Animal pair names differ only with respect to the position of word stress, e.g. /ˈbimo/ vs. /biˈmo/ (contrastive stress). All stress positions in the target words are possible stress patterns in German. Contrastive stress is supposed to be acquired rather late in language development (Pollock et al., 1993; Altmann and Kabak, 2000) but it provides conditions for comparing stressed and unstressed vowels due to the identical segmental context. The different position of stress in the target words follow typical stress patterns in German, which were estimated from the CELEX database (Baayen et al., 1995). According to a CELEX–based corpus analysis performed by Féry (1998) the majority of German disyllabic (73%) or trisyllabic (51%) words are stressed on the penultimate syllable. But when considering only words with two full vowels the frequency distribution of initial and final stress changed (Domahs et al., 2008). In 61% of disyllabic words, those with two full vowels were stressed on the final syllable. For trisyllabic words antepenultimate (29%) as well as ultimate stress (53%) becomes more frequent.

Motivated by these results, the production of contrastive stress was analysed between the first and the second syllables of disyllabic words, and between the first and the third syllables of trisyllabic words.

3.2 Recordings

The recordings took place at children's homes in familiar play situations with their parents. No person unfamiliar to the child needed to be present during the recordings. During a recording session one parent played with the child. Before the recordings started the parents were instructed on how to use the recording equipment.

Phonetic transcription and analysis requires high quality recordings. To ensure high quality of the recordings as well as high comfort for the children during their play, the children wore a cordless Lavalier microphone (NADY LT–4 (Lavalier) E–701 (600 Ohm)) clipped on to a vest. The wireless transmitter was attached to the children's clothes at the back. With this wireless microphone the children could move through the room and play freely without the risk of low recording levels which would occur if the microphone was placed in a fixed position in the room. For the recordings a Digital

Audio Tape Recorder (DAT–Recorder) Sony TCD–D100 with a sampling rate of 48 kHz (16 Bit linear) was used for one–track recordings.

Because the parents' speech was also required for this study, two–track recordings from children and parents were also made at some recording sessions with a Marantz PMD670 Flash Recorder with 2 GB CF–Card and sampling rate of 48 kHz and wireless microphone AKG CK 97–L. With this equipment high–quality recordings were obtained, which could be analysed acoustically using standard speech analysis tools.

The children were recorded during naturalistic interaction with their caretakers while looking at picture books or playing with toys. Therefore the data represent spontaneous productions of the children. However the setting was controlled to some degree because the caretakers were always offered the same picture book during child babbling and first word productions to motivate comparable productions from the children. As the children became more productive with age it became possible to elicit trisyllabic words. Therefore the picture cards were introduced into the spontaneous interaction.

The parental interaction with the children turned out to be helpful to identify the attempted lexical word mainly in the phase where babbling sequences and meaningful words were produced.

3.3 Annotation

All recordings were transferred to a workstation, downsampled to 16 kHz and manually annotated with WaveSurfer [1]. Figure 3.2 shows an example for the annotation of speech data with seven different levels.

All utterances by children aged 18 months or younger were transcribed and analysed following general guidelines for transcribing child speech samples (Stoel-Gammon, 2001). With the onset of meaningful words, the first hundred words and all target words of the picture task were transcribed in a narrow phonetic transcription according to XSAMPA. The adult speech was also transcribed. In the children's sound inventory it is sometimes difficult to identify the exact segment produced but it is possible to determine certain features of that segment. Therefore until the age of 18 months a broad transcription of the segments was also carried out. In the broad transcription the consonantal structure and the vocalic structure were described with different symbols based on place and manner of articulation.

As the segmental structure of the utterances was also important for the phonetic and phonological analysis of the data, the following segmental levels had to be annotated:

[1] http://www.speech.kth.se/wavesurfer/

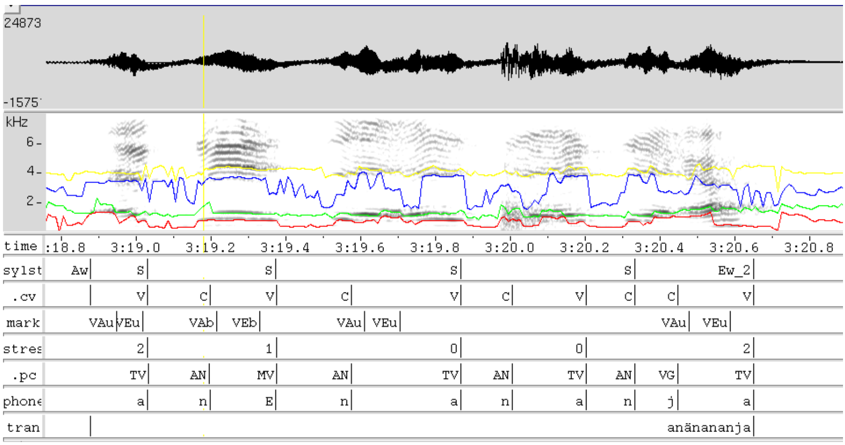


Figure 3.2: Annotation of utterance [aˈnɛnananˌja] (Nils 0;7:11).

cv : Annotation of vocalic and consonantal structures.

stress : Perceptual prominence for each syllable: no prominence (0), most prominent (1), prominent but not most (2).

marks : Beginning and end of stable phase of vowel (marked with beginning (VA) and end (VE) of second formant). No influence of the surrounding context should be audible. The stable parts within a vowel are characterised by the parallel course of formants (visible in the spectrogram) and constant course of amplitude (visible in time signal).

sylstr : Syllable structure, beginning and end of word as well as of the syllable. Speech development rating (Stoel-Gammon, 1989) at the end of an utterance.

phones : XSAMPA transcription

trans : orthographic transcription

Table 3.2: Adult and child parameter for automatic acoustic analyses

	f_0 min	f_0 max	window duration	frame step
adults	60Hz	500Hz	0.015s	0.005s
children	150Hz	700Hz	0.0075s	0.005s

3.3.1 Reliability

It is important to assess the reliability of phonetic transcriptions by examining agreement across transcribers. As the goal of this study is to analyse stress, the beginning and end of the vowels and the stress labels are very important. Therefore the inter–transcriber reliability for phone, stress, and mark labels was measured

Two trained transcribers were responsible for the transcripts. For the first transcriber (IF) about 1798 mark labels (beginning and end of vowels), 2058 phone labels (vowel identity) and 943 stress labels were cross–labelled by a third transcriber (the author). The inter–transcriber reliability between IF and BL was for the marks labels 92.5%, for phones 93.9% and for stress 96.5%.

From the second transcriber (FW) 841 mark labels, 2476 phone labels and 861 stress labels were cross–labelled by the author. The inter–transcriber reliability between FW and BL was 85.7% for the mark labels, 95.35% for phones and 93.7% for stress.

As acceptable reliability measurements for transcriptions of children speech are typically 60–80% (Stoel-Gammon, 2001), no alternative assessments of transcription agreement were necessary.

3.4 Speech analysis

In order to process the large number of speech samples, different automatic tools for analysing speech were used.

The analysis tools were evaluated for acoustic measurements of adult speech data. In order to use these tools for children's speech data the parameters of the tools had to be adapted and evaluated (see Section 3.4.3) to the conditions of child speech. The window length in particular had to be changed in order to meet the distinctly higher fundamental frequency (f_0) of child speech. In Table 3.2 the measurements of the parameters for the analysis of adult speech and the adaptation to measurements of child speech are given.

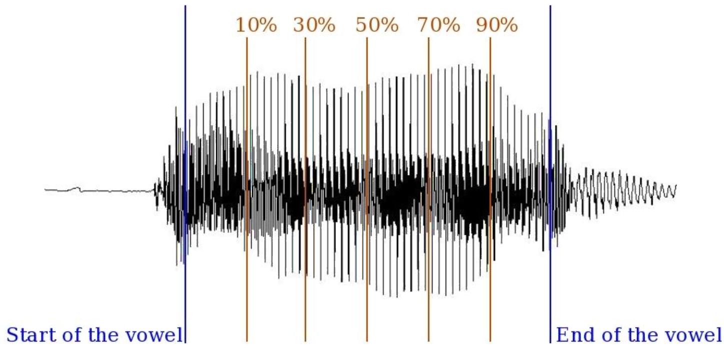


Figure 3.3: Position of measure points over labelled vowel length.

3.4.1 Acoustic parameters

In order to extract the fundamental frequency, first four formants and RMS (root mean square) Intensity an automatic fundamental frequency (*get_f0*) and formant–tracking (*formant*) programme in the ESPS signal–processing package by Entropic Research Laboratory was used. The software utilises a dynamic programming technique to select the best formant tracks from raw formant estimates. Each speech waveform was internally downsampled to 10 kHz and processed using a 12th order linear–prediction analysis, and a pre–emphasis factor of 0.7 and a Hamming window with the described parameters.

Apart from the well known acoustic parameters of stress additional voice quality measurements were carried out. In order to calculate the corresponding voice quality parameter from the acoustic signal an analysis tool, developed by Pützer and Wokurek (2006), was used.

For the automatic measurement fundamental frequency, frequency and bandwidth of the first four formants and narrow–band spectra were estimated from the speech signal also using the ESPS signal–processing package. In the first part of the measurement process the strongest spectral peaks of the first two harmonics were measured and logged. Based on these acoustic values different voice quality parameters corresponding to fundamental and formant frequency, their bandwidth, and spectral peaks, were calculated. In Chapter 5.3 a more detailed description of the calculated parameters and their use is introduced.

All acoustic parameters were extracted at five points of the vowel, at 10, 30, 50, 70, and 90% of the marked vowel duration (Figure 3.3) to capture parameter changes over the duration of the speech sound (Shih et al., 1999; Möbius, 2004).

3.4.2 Vowel duration

The durations of all marked vowels (marked with beginning and end of second formant) were extracted from the corresponding label files (produced by manual segmentation) and measured. Vowel duration was measured as the difference in time between the marked beginning and end of the vowel. As vowel length is also dependent on vowel quality, the vowel durations were normalised with respect to the corresponding mean duration and standard deviation (see Section 3.4.4).

Since the automatic measurement procedure sometimes erroneously yielded excessively short or long vowel durations, the raw duration data were refined using exploratory data analysis methods. The initial data were grouped separately for all subjects. For each recording extreme values were calculated for each vowel depending on stress. Data points that fell outside the range of 1.5 of the interquantile range were removed (in Appendix A.6 the results are given). Despite the efforts to remove erroneous values, it is possible that the refined data still includes some under– and over–estimated values.

3.4.3 Evaluation of the acoustic measurements

It has been shown (Kent and Murray, 1982; Vallabha and Tuller, 2002) that formant measurements are particularly difficult to perform in high–pitched voices. The large distance between adjacent harmonics can lead to undersampled spectra and this should be kept in mind for LPC analyses, in which formant measurements are greatly influenced by the closest harmonic (Vallabha and Tuller, 2002). However, LPC analysis is the procedure used in extensive studies of acoustic characteristics of child speech (Peterson and Barney, 1952; Hillenbrand et al., 1995; Lee et al., 1999; Ménard et al., 2007b). To avoid formant measurement errors the performance of the automatic program was evaluated using hand–measured formant values of 80 vocalic segments from one girl aged 22 months. The manually estimated formant frequencies were compared to the automatically computed ones. Mean differences (standard deviation) between the automatically and manually estimated values in Hz were 19.47 (318.88) for *F1*, and 17.94 (269.53) for *F2*. The differences between automatic and manually measured formant values for this girl are less than reported for male adult speakers (Möbius, 2001) and comparable to differences found in other acoustic studies of child speech (Lee et al., 1999; Ménard et al., 2007b). Although the automatic formant–tracking program yielded reasonable estimates for the first formant frequency in most cases, higher formants were often inaccurate for vowels produced by young children due to poor spectral resolution. This low quality resolution at high frequencies was caused by wider harmonic spacing and breathy voicing, which confused the formant tracker. To minimise statistical biases due to erroneously estimated formant frequencies and intensities, the raw formant and intensity data were refined using the exploratory data analysis method described earlier. All fundamental and formant frequencies as well as intensities were controlled and the data points that fell outside the range of 1.5 of the interquantile range were removed (see Appendix A.6).

3.4.4 Normalisation

Vowel production is influenced by different types of information, as phonemic, i.e., the intended phonemic identity, the anatomical conditions and sociolinguistic information (Ladefoged and Broadbent, 1957) systematically affect formant frequencies (e.g. Peterson and Barney, 1952; Pols et al., 1973). In order to eliminate these influences, different vowel normalisation procedures can be performed (e.g. Lobanov, 1971; Syrdal and Gopal, 1986; Miller, 1989; Traunmüller, 1990). For a good comparison of vowel normalisation procedures see Adank et al. (2004).

In German, vowel duration is also dependent on vowel identity (Jessen, 1997, 1999) and should also be normalised when comparing the durations of different vowels. For the normalisation of F0 and formants the Bark scale was chosen. For normalisation of vowel duration, RMS Intensity and all voice quality values zscores were calculated and used for analysis.

Bark scale

The Bark scale is a psychoacoustical scale and is designed to represent the tonotopic outputs of the critical filters in the ear. It is useful to model the critical–band functions based on the natural division of the audible frequency range of the ear. It has been suggested (Syrdal and Gopal, 1986; Carter, 2002) that the Bark scale provides a useful tool for comparison across speakers and across sexes when measurements are given in Bark differences. At the very least, analysis of formants in terms of Bark differences provides a fairly crude first pass at a normalisation for the differing sizes of vocal tracts between speakers which leads to different acoustic natural resonance characteristics and hence differences in the absolute placement of formants in the vowel space. Therefore the formant frequencies were converted to the Bark scale, since this scale models the ear's integration of Hz frequency, following the formula (Eq. 3.1) found in Traunmüller (1990) [2].

$$F_{Bark} = \frac{26.81}{1 + \frac{1960}{f_{Hz}}} - 0.53 \tag{3.1}$$

[2] if result < 2 add 0.15*(2-result), if result > 20.1 add 0.22*(result-20.1)

Z–transformation

Vowel duration, RMS Intensity and the voice quality parameters based on formants are vowel–specific as well as speaker–dependent. To eliminate the speaker–dependent and vowel–specific effects a normalisation is done. The z–transformation (Lobanov, 1971) eliminates the specific effects by replacing absolute values by their difference to the specific mean and dividing this difference by the specific standard deviation of this parameter (Eq. 3.2).

$$z = \frac{x - mean(x)}{sd(x)} \tag{3.2}$$

3.5 Data selection

In order to develop an exemplar–based model of the acquisition of stress in German, various analyses were conducted in the following chapters.

The segmental annotation describing the consonantal and vocalic development as well as the development of syllable structure was the basis for describing the phonological development. The developmental stages between 5 and 36 months of age were studied for six children. In Chapter 4 the results of this analysis are presented. The normalised data of the automatic speech analysis was the basis for describing the phonetic development of stress in German. For this the acoustic data from the six children taking part were analysed, as well as their parents' use of the acoustic parameters of stress. In Chapter 5 the results of this analyses are presented.

The normalised formant values reached by the automatic speech analysis were used to describe the development of vowel space for stressed and unstressed vowels. As the development of vowel space is not finished by the age of 36 months, cross–sectional data for children in the Kindergarten and primary school were analysed alongside the six children who had already participated in this research in this chapter. In Chapter 6 the development of vowel space and their language specific probability distribution over the phonetic space is described for children between 5 months and 10 years of age, as well as for adult speakers.

4 Phonological development of stress

4.1 Introduction

The prosodic domain of word stress is the syllable. On the phonetic level, stress typically occurs on multi-segmental strings (Beckman, 1986). For a comprehensive study about the acquisition of stress, it is therefore necessary to look at both the development of the syllable and the processing of German word stress.

Stress is a structural, linguistic property of a word that specifies which syllable is more prominent than the others and is determined by the language system (Liberman, 1975; Sluijter, 1995). Each language uses different acoustic cues to produce stress, e.g. length or overall intensity, which have both been claimed as a correlate of stress in dynamic stress languages such as English and German (for a a good overview of the languages of Europe see van der Hulst, 1999). The relationship between the phonetic cues of stress in a given language and the hierarchical prosodic and intonational representation must be learnt by each infant. During babbling it is assumed that the unit of speech is the rhythmic mandibular cycle, simulating consonants in the closure phase, vowels in the open phase. Syllables are considered to be a consequence of the consonant-vowel alternations. During their development, as the infants get increasing control over the coordination of the articulators, segmental content elements such as consonants and vowels become separate entities (MacNeilage, 1998; MacNeilage and Davis, 2000; Davis et al., 2002). With the beginning of babbling the produced open–closure phases representing the syllable are the frame of the productions. These frames are filled with melodic content represented by the produced sounds. The first produced pre–linguistic sounds are represented by *protosyllabaries* based on four sequential sound patterns (three CV and one CVC sequences) (MacNeilage and Davis, 2000). During the second year of life content typically starts to be inserted into the frames to form the child's first words. The development of syllables plays a critical role in the acquisition of fluent speech (Packman et al., 2007). In early words, it is assumed that children tend to retain both stressed and final syllables in their output forms (Echols, 1993; Fikkert, 1994; Grimm, 2008). In recent research (Allen, 1981) it is has been thought that children show a trochaic bias in syllable acquisition. However, newer studies (Demuth and Fee, 1995; Demuth, 1996b; Vihman et al., 1998) show that the proposed trochaic template is too rigid and arguments against the trochaic bias are given (Rose and Champdoizeau, in press). A constraint–based analysis (Prince and Smolensky, 2004) gives evidence of stress as the unmarked case in a lexical item and shows that it is therefore learnt extremely early, even before the onset of the first words (Demuth, 1996a). As can be seen in Figure 4.1 the prosodic domain of stress,

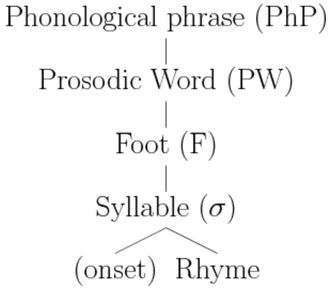


Figure 4.1: Universal prosodic hierarchy.

the syllable is part of universal prosodic hierarchy (Selkirk, 1984; McCarthy and Prince, 1996; Blevins, 1995). One function of stress is the hierarchical organisation of rhythmic units. The syllables are grouped into feet and the feet are grouped into prosodic words. This grouping determines the potential position of primary stress in a word as the strong syllables of a strong foot always gets the main stress.

It has been shown (e.g. Fikkert (1994) for Dutch; Demuth and Fee (1995); Kehoe and Stoel-Gammon (1997) for English; Grimm (2008) for German) that the acquisition of syllable structure goes along with the development of the metrical foot and depends on stress as stressed syllables show a strong tendency to be retained. With the beginning of the production of first words, the development of stress and syllable can be described in four different stages (Dresher and Kaye, 1990; Fikkert, 1994). The transition from one stage to the next stage often involves the extension of the child's metrical template or a change in one or more parameter values. For children learning German, Grimm (2008) found three developmental steps in the production patterns of simplex words. German children seem to preserve the main-stressed syllable rather than the rightmost stressed one. Evidence for this assumption arises from words which have their main stress at the beginning but also have a heavily accented final syllable, i.e. a syllable with secondary stress (Grimm, 2008). Truncation of the medial weak syllables as found in Dutch (Fikkert, 1994) and English (Kehoe, 2000) could not be observed for the German data by Grimm (2008). The stages found by Grimm (2008) for the production pattern of German simplex words are as follows (examples for production patterns taken from Grimm (2007)):

- Single foot stage.

  At the initial stage, all target words are reduced to one single monosyllabic foot or to a disyllabic trochee. Simplex words preserve the main-stressed syllable in truncated form. The prosodic word consists of a single or a two–syllabic trochaic foot.

syll	target word	phonetic target	child output
2	Kamel (*camel*)	/kaˈmeːl/	[mel]
3	Papagei (*parrot*)	/ˌpapaˈga͡ɪ/	[ga͡ɪ]
3	Giraffe (*giraffe*)	/giˈrafə/	[ˈhafə]
4	Marmelade (*jam*)	/ˌmaɐməˈlaːdə/	[ˈhaːtə]

- Intermediate stage.

  For German children, two different strategies in extending the prosodic word size are feasible.

  Foot-based strategy:

syll	target word	phonetic target	child output
2	Ballon (*balloon*)	/baˈlɔŋ/	[lɔm]
3	Giraffe (*giraffe*)	/giˈrafə/	[ˈkrafə]
4	Schokolade (*chocolate*)	/ˌʃokoˈlaːdə/	[ˌsɔkoˈlaːtə]

  A second foot is added to the existing one. Words with two foots are produced only if the target word is stressed on the first syllable. Initially unfooted syllables are reduced to a single foot.

  Syllable-based strategy:

syll	target word	phonetic target	child output
2	Kamel (*camel*)	/kaˈmeːl/	[kʰiˈmiː]
3	Papagei (*parrot*)	/ˌpapaˈga͡ɪ/	[kaˈka͡ɪ]
4	Marmelade (*jam*)	/ˌmaɐməˈlaːdə/	[ˈlali]

  Preservation of two syllables from the right edge of the word. Using this strategy children only produce disyllables at this stage. The expansion of the size of words is based on syllables.

- Target-like prosodic shapes.

syll	target word	phonetic target	child output
2	Kamel (*camel*)	/kaˈmeːl/	[kʰaˈmel]
3	Giraffe (*giraffe*)	/giˈrafə/	[giˈʁafə]
3	Papagei (*parrot*)	/ˌpapaˈga͡ɪ/	[ˌkakaˈka͡ɪ]
4	Schokolade (*chocolate*)	/ˌʃokoˈlaːdə/	[ˌlɔkəˈlaːlə]

  At this stage adult–like prosodic words are produced by children following the foot–based strategy. Children following syllable–based strategy have to exceed the restriction of disyllabic prosodic words.

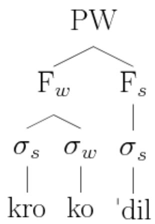


Figure 4.2: Metrical structure of *Krokodil* (crocodile).

It is assumed that during the pre–linguistic period the early vocal production patterns are grounded in biomechanical constraints. The infants have to learn to control acoustic cues underlying different elements of prosody. Based on computational models (Guenther, 1994; Guenther et al., 1998; Guenther and Perkell, 2004) it can be claimed that the infant needs enough input to create a new category. Until this category is successfully implemented, the infant has to register the acoustic cues of a category and needs to attempt to produce them. During babbling, based on a wide variability in different articulatory gestures which reach nearly the same target, different production schemes are trained. Based on these trained articulatory gestures the child has to refine his production when beginning with word production. The transfer of these articulatory gestures to other forms which contain the same prosodic shape, depends on the frequency of this prosodic shape (Beckman and Edwards, 2000). The acoustic parameters of a certain phoneme also vary and appear with varying frequencies in different positions.

4.2 Prosodic representation of stress

Following metrical stress theories (Liberman, 1975; Selkirk, 1984; Giegerich, 1985; Nespor and Vogel, 1986; Hayes, 1995; Trommelen and Zonneveld, 1999) stress gives the rhythmic units of a language a hierarchical organisation. Stress is a relationship between prominent and weak syllables and is produced with different acoustic parameters (van der Hulst, 1999).

On the level of the prosodic word, stress can be organised in metrical feet, which assigns strong and weak syllables to metrical patterns. The number of syllables, syllable quantity and the construction direction of feet (rightward or leftward) define different foot–types as either trochees (initial stress) or iambs (final stress) (Hayes, 1995). The resulting metrical structure of the underlying prosodic word consists of a two–level hierarchy under the prosodic word. The prosodic word consists of strong and weak feet and the feet consist of strong and weak syllables. The syllable weight is determined by the unit of sounds (**mora**).

Table 4.1: Target prosodic shapes of picture naming task.

syll	target shape	Example
2	SW	*'Spiegel* (mirror), *'Fliege* (fly)
2	WS	*Ka'mel* (camel), *Pa'ket* (package)
3	sWS	*ˌPapa'gei* (parrot), *ˌEle'fant* (elephant)
3	WSW	*Ba'nane* (banana), *Gi'raffe* (giraffe)
3	SWW	*'Pinguin* (penguin), *'Trampolin* (trampoline)

A syllable containing one mora is said to be monomoraic (light syllable); one with two moras bimoraic (heavy syllable). An example of the resulting metrical structure is given for the final stressed *Kroko'dil* (crocodile) in Figure 4.2.

As the crucial prosodic parameter of a language is its foot type (iamb or trochaic), Modern Standard German remains a trochaic language (e.g. Jessen, 1999) as the minimal German word consists of a bimoraic syllable (e.g. *Wort*, 'word') or of two monomoraic syllables (e.g. *Mama* 'mummy'). Despite the fact that there is a preference for specific stress patterns in German, by parsing syllables into feet, the main stress can wander. In German, the position of stressed syllables is not fixed but often predictable (Wiese, 1996) and the German word stress system is often suggested to have a close similarity to the stress pattern of English. In this way German word stress is often discussed as similar and quantity–sensitive consisting of long vowels as well as heavy closed syllables (Giegerich, 1985; Vennemann, 1986; Féry, 1998). On the other hand re–examinations of German stress suggest that the patterns of main stress are more comparable to those of Dutch, where long syllables do not count as heavy in the assignment of stress (Eisenberg, 1991; Wiese, 1996; Jessen, 1999; Janßen, 2003). In this case, it is assumed that stress on the penultimate syllable is the dominant pattern but also more marked patterns occur. Even German shows metrical stress the main stress pattern of disyllabic utterances is trochaic (Féry, 1998, 2008) (stressed on the first syllable). But iambic patterns can also occur depending on syllable weight (see Table 4.1).

In summary, it can be generalised that stress in German has to fall on one of the last three full syllables [1] of a prosodic word and that stress alignment operates from the right edge of the word to the left. If stress does not fall on the final or penultimate syllable, it has to fall on the antepenultimate syllable, independent of the internal structure. A regular deviation from this restriction occurs only if the antepenultimate syllable contains a schwa, as stress on the schwa is ruled out. If this is the case, stress generally occurs four syllables from the right. For polysyllabic words in German, word stress falls on one of the last three syllables of an utterance and final (US), penultimate (PU), as well as antepenultimate (AP) stress patterns occur. Féry (2008) also stated that there is a dominance

[1] syllables containing /ə/ are not counted as full syllables.

Table 4.2: Possible metrical structure of trisyllabic German simplex words.

stress pattern	example	metrical structure
final	Ele'fant (elephant)	$(\sigma\sigma)^{\text{'}}(\sigma)$
penultimate	Gi'raffe (giraffe)	$\sigma^{\text{'}}(\sigma\sigma)$
antepenultimate	'Pinguin (penguin)	$^{\text{'}}(\sigma\sigma)(\sigma)$

of penultimate over ultimate stress. Therefore it can be concluded that penultimate stress represents the most frequent pattern of German disyllabic and trisyllabic words and as a consequence should be learnt first by children. From this point of view antepenultimate stress is exceptional because the syllabic structure is rather inhomogeneous. In a re–analysis considering only words with full vowels where final syllables can also be stressed, the frequency of stress pattern changed (Domahs et al., 2008). The type of stress pattern depends not only on the frequency of a specific stress pattern but also on the structure of the final syllable. In this case, during the development of disyllabic and trisyllabic words there should be different stress patterns occurring depending on the final syllable. New empirical findings on the perception of German word stress using event–related brain potential as the dependent measure (Domahs et al., 2008) suggest that for the processing of word stress, lexical information as well as both (quantity–sensitive) information about the metrical structure, and the hierarchical ordering of syllable and feet is necessary. These results strongly suggest that German word stress relies on structural properties, e.g. its foot structure, as well as on the distinction between default stress and lexically–specified stress. The potential position of primary stress in a word depends on the grouping of syllables into feet and on the grouping of feet into prosodic words. Strong and weak syllables assign metrical feet and build up the metrical pattern of an utterance.

Based on the approach of the organisation of prosodic words, different metrical structures for the three possible stress patterns can be found (Table 4.2). German words with final or antepenultimate stress consist of two prosodic feet including a trochee $(\sigma\sigma)$ and a non–branching foot (σ), stressed on the final foot (final stress) or on the first foot (antepenultimate stress). For words with penultimate stress it is assumed that they consist of only one final trochee and the leftmost syllable of these words either builds a foot on its own or is unfooted and often noted extrametrical. These two options have their own implications as unfooted syllables will never be stressed while footed syllables will receive at least secondary stress. To conclude, for German different stress positions are derived from the way in which the syllables are parsed into feet. When the syllables are grouped into a metrical structure, the structure of the final syllable play a leading role. Only words with an open final syllable tend to be stressed on the penultimate syllable whereas words with closed final syllables are stressed either on the final or the antepenultimate syllable (Janßen, 2003; Domahs et al., 2008).

To describe the phonological development of stress in German, two different analyses are conducted. Firstly, the development of the prosodic structure from babbling to words mainly for disyllabic and trisyllabic utterances are described. Secondly, as it has been stated that German word stress also depends on the syllabic structure – mainly of the last syllable – the acquisition pattern of syllabic structures are also described. The aim is to find out whether syllable structures depend on the stage of production (babbles or words) and whether different syllabic structures depend on stress, in order to build up a mental syllabary (Levelt, 1989; Levelt and Wheeldon, 1994; Levelt et al., 1999b). Furthermore it will be observed if there are any recurring samples in the stress pattern depending on syllable structure.

4.3 Development of prosodic structure

Based on Allen and Hawkins' (1980) study of early productions by young English learners, it has often been proposed that children have an inborn bias in favour of a trochaic (strong–weak) foot shape in phonological development (Fikkert, 1994; Archibald, 1995; Gerken, 1996). However Vihman et al. (1998) failed to find a trochaic bias for English and French learning infants. The development of prosodic structure during the word production stage is guided by the expansion of the prosodic domain from the single foot stage up to the production of polysyllabic words. During this development period, children display a variety of prosodic processes in their production such as deleting and adding syllables in certain positions more than in others, and altering stress pattern in a systematic fashion. The underlying syllabification routine has to be learnt by children in the stages of word production. During babbling, no restrictions occur to the metrical pattern that can be produced as it is claimed that infants, even in their pre–linguistic utterances, are able to produce acoustically trochaic (strong–weak), iambic (weak–strong), and spondaic (strong–strong) patterns (Vihman et al., 1998; Davis et al., 2000; DePaolis et al., 2008).

4.3.1 Method

For the analysis of the prosodic structure acquisition of babble and words, all labelled utterances, including words and non–words or babble vocalisations, were taken from each of the six children during the observation period from 5 months up to 3 years of age. With the beginning of word production, the picture cards representing two and three syllable German words were presented to the children. Table 4.1 describes the target prosodic shapes of the simplex words occurring in the database. In order to look at the development of prosodic structure, the production data were compared by age, stage (babble or words) and word stress pattern. The data therefore include first babbles as well as first words up to trisyllabic words.

Table 4.3: Overview of age groups.

group	**age (in months)**	**stage**
1	5–9	babbling
2	12–18	babbling and first words
3	19–24	mainly words occur but also some babbles
4	25–30	first word stage
5	31–36	meaningful speech

As stress is a structural, linguistic property of a word that specifies which syllable is more prominent, each syllable is marked for its perceptual prominence in the utterance. Three levels of prominence are given: no prominence perceivable marked with 'W', most prominent syllable in the utterance ('S'), and less prominent syllable in the utterance ('s'). In the case of word production where the target word was a compound, this was marked as 'comp'. In German word compounds, the initial constituent obtains main stress and the other constituent obtains secondary stress, providing the compounds were derived from monosyllabic or trochaic constituents. As the data are non–metrical and nominal, cross tabulations were used for descriptive analysis. A cross tabulation (often abbreviated as cross tab) displays the joint distribution of two or more variables. They are usually presented as a contingency table in a matrix format. Whereas a frequency distribution provides the distribution of one variable, a contingency table describes the distribution of two or more variables simultaneously. Each cell shows the number of respondents who gave a specific combination of responses, that is, each cell contains a single cross tabulation. Furthermore, the frequency of intersection of the variable categories was counted. The standardised residual is found by dividing the difference of the observed and expected values by the square root of the expected value. They provide information about which cells contribute to a significant chi–square result. The expected value of the standardised residual is 0, in our case this means that the acoustic stress patterns would be equal for words and babbles. Negative values describe less than the expected values, a positive value exists in the case of more productions of a special pattern than expected.

4.3.2 Results

First of all, cross tabulations were carried out for all utterances produced by all children. For a better overview, the children were grouped in five different groups according to their age. In Table 4.3 the grouping is given.

For all utterances the distribution of all produced different acoustic stress patterns were described using cross tabulation. The variable stress patterns produced in the different age groups and different stages (babbles or words) were analysed.

During babbling, a great variety of possible stress patterns occurred. Furthermore, utterance length was not limited as babbling utterances up to 22 syllables were produced. In Appendix A.7 utterances of babbling with more than six syllables are listed. On the other hand, utterances of words were more limited in their prosodic structure than babbles. The longest utterances classified as words contained six syllables[2].

In the first group, a great variety of stress patterns were observed. The production of different stress patterns were arbitrary as no preference pattern were observed. Even with increasing control over their articulation and the production of first words (group two), a great variety in the production of babbling were still observed. However, when words were produced in this group, they tended to be mainly monosyllabic words. The first produced disyllabic words had generally a trochaic stress pattern.

Between 19 and 24 months of age (group 3) more and more words were produced and less babbles with less variability. The children were paying more attention to the stress pattern of their input, trying to adapt their productions of words as well as babbles to adult patterns. The one–syllable words were extended and more and more two–syllable words with trochaic pattern were produced. There were also instances of iambic intonation. These prosodic patterns were extended with increasing age (group 4 and 5). As the main prosodic pattern produced by all children contain one to three syllables, a closer look at the development of prosodic structure of disyllabic and trisyllabic utterances was taken to describe how babbles and words differ or are comparable in their prosodic structure. Therefore, a reanalysis of all disyllabic and trisyllabic utterances was done. In Figure 4.3 all disyllabic utterances are classified according to their perceptual stress. Weak–strong stress patterns are grouped under iambic whereas strong–weak stress patterns are grouped under trochaic and strong–strong stress patterns under spondaic. As in some productions of the children no weak syllable were perceived, level stress (s) was supposed. However, the stress pattern s–S (like iambic) and S–s (like trochaic) were also observed. At the words stage compSs is included, as this stress pattern were observed for disyllabic productions of words based on German compounds. For German compounds it is assumed that the second part of the compound bears level stress (s).

In Figure 4.4 all trisyllabic utterances are classified. All utterances perceived as stressed on the third syllable are grouped as stressed on the antepenultimate syllable, utterances stressed on the middle syllable are grouped as penultimate stress and utterances stressed on the last syllable are grouped as ultimate stress.

The prosodic stress patterns produced during babbling were extremely variable for disyllabic (Figure 4.3) as well as for trisyllabic (Figure 4.4) utterances. During babbling, the first disyllabic utterances already showed that all possible acoustic stress patterns were produced when babbling begins at 5 months. However, the frequency of occurrence varies between utterances stressed on the first syllable (produced with trochaic like pattern: SW as well as Ss) and utterances stressed on the second syllable (produced with

[2] Hanna (3;0): $[[[\sigma\sigma]_F[\text{'}\sigma\sigma]_F]_{PW}[[\text{,}\sigma\sigma]_F]_{PW}]_{PW}]]$; prosodic shape: sWSW–sW; /pɔlɪˈtsɪstənˌaʊtoː/; polizistenauto

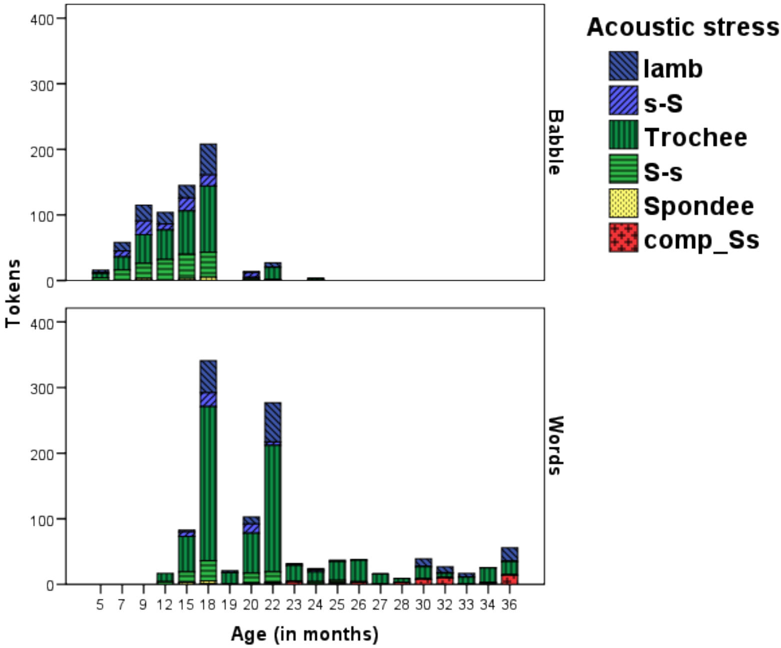


Figure 4.3: Token frequency of disyllablic utterances classified acoustically as stressed on the last (iambic), or first (trochaic) syllable, or on both syllables (spondaic) for babble and words of all six children.

iambic like pattern: WS as well as sS). Spondaic stress patterns were rarely perceived in the first productions but also occurred. During the whole babbling phase, an increase in the trochaic productions were observed. Spondaic productions mainly occurred between 9 and 18 months of age. The iambic productions during babbling showed a stable occurrence. To sum up, during babbling the German children in this study showed a preference to produce stress on the first syllable (trochaic stress pattern) but were also able to produce stress on the second syllable (iambic). Real equal stress of both syllables (spondaic stress) was rarely produced. The reduction of the unstressed syllable was not complete as most syllables were perceived as level stressed (sS and Ss) and not as unstressed. The picture changed for word production. Up to the age of 22 months, the word productions were mainly produced with a trochaic stress pattern, and only few with an iambic stress pattern. Spondee word production occurred, too, but mainly during

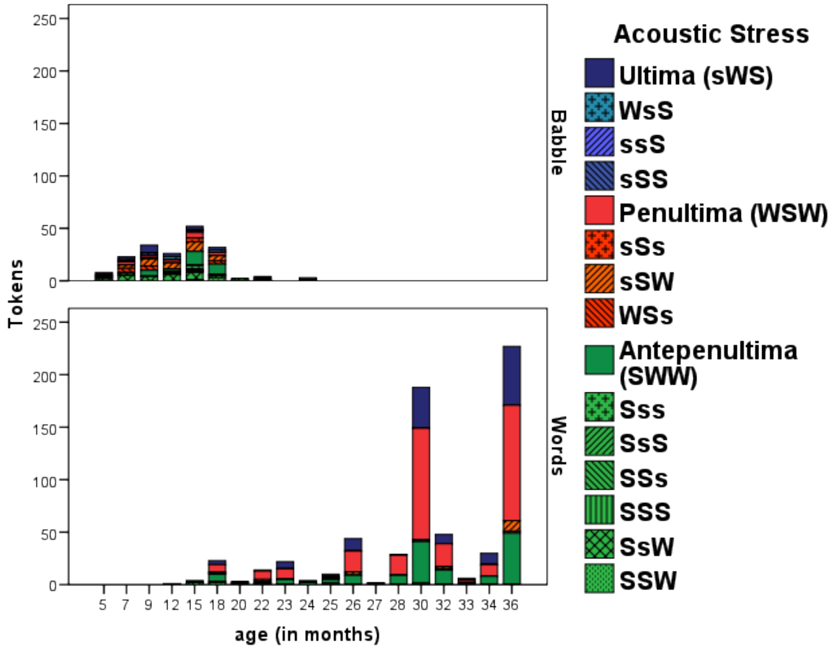


Figure 4.4: Token frequency of trisyllabic utterances classified acoustically as stressed on the antepenultimate syllable, penultimate syllable, or ultimate syllable for babble and words of all six children.

the period where babbling and words were produced side by side. As word production increased, all possible stress patterns were produced, with a preference for a trochaic pattern. As the children increased in age, they attempted more and more compounds, as can be seen in increasing compSs–stress patterns. A similar pattern were observed for trisyllabic utterances. During the babbling stage, a clear preference for antepenultimate stress were seen, which is in line with the assumption that the first syllable in German bears main stress. The other possible stress patterns were also produced, but not so frequently. The reduction of stress was difficult for trisyllabic utterances. In the babbling stage, more level stress patterns were perceived, an not so much unstressed syllables. With increasing control over the acoustic cues of stress, the reduction of weak syllables to really unstressed syllables was managed with greater success. Gradually, more and more unstressed syllables were produced by the children and the adult production tar-

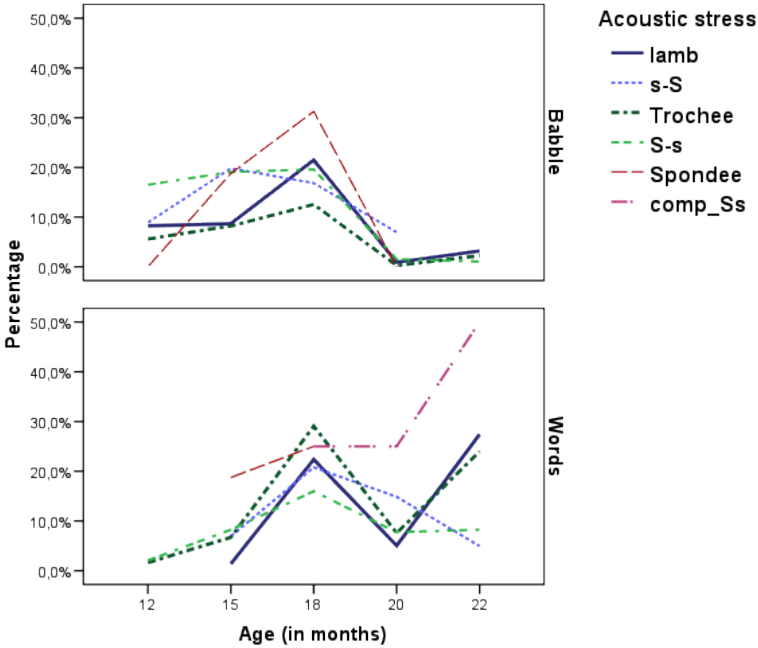


Figure 4.5: Frequency of occurrence (in %) of disyllables classified acoustically as iambic, trochaic, or spondaic for babble and words of all six children.

get was reached with more accuracy. The variance of different stress patterns correlated negatively with age.

In an examination of the frequency of the different stress levels (stressed, level stress, unstressed), at the beginning of babbling more syllables are produced as stressed or as containing level stress. Only about 30% of all produced syllables were unstressed. This percentage remained nearly unchanged during the entire babbling phase. In contrast, in word production the frequency of syllables produced as unstressed was nearly equal to the number of syllables produced as stressed. With increasing ability to control the acoustic parameter to mark stress, more unstressed syllables were produced by the children. At the stage where babbles as well as words were produced, the frequency of the various acoustic stress patterns depending on production stage were analysed. Therefore, Pearson's χ^2 Test was carried out. Disyllabic (χ^2(2,1853)=103.79; p<0.001) as well as trisyllabic (χ^2(18,1057)=289,63; p<0.001) utterances showed significantly dif-

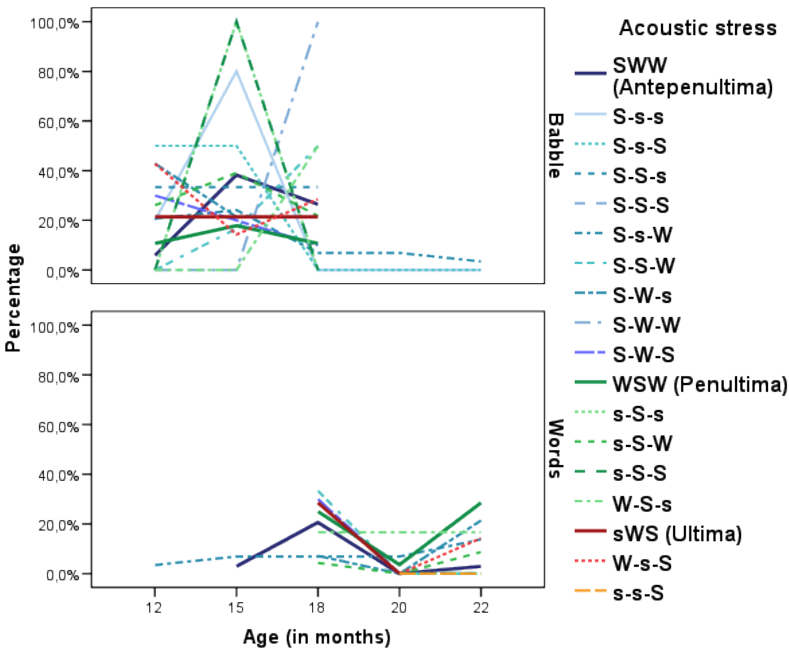


Figure 4.6: Frequency of occurrence (in %) of trisyllabic utterances classified acoustically as stressed on the antepenultimate syllable, penultimate syllable, or ultimate syllable for babble and words of all six children.

ferent frequency patterns for babbles and words. Cross tabulations with standardised residuals for disyllabic (Appendix B.1) and trisyllabic (Appendix B.2, B.3) utterances showed the distribution of word stress according to production stage and age. In Figure 4.5 the various frequencies of the different production patterns at each production stage for disyllabic utterances and in Figure 4.6 for trisyllabic utterances between 12 and 22 months of age (where babbles as well as words are produced by the children) are given.

The stress patterns produced in babbles occurred in nearly equal numbers, with an increase in the production of trochaic as well as iambic pattern at the age of 18 months. In contrast, the first disyllabic words showed significant preference for trochaic patterns. The first production of words only contained trochaic patterns. Iambic patterns also occurred from 15 months of age, but a clear preference for trochaic pattern were observed.

The stress patterns for trisyllabic utterances produced in babbling showed a variable pattern with no clear preference for antepenultimate or penultimate stress. In contrast, the first trisyllabic words were produced with stress on the first syllable. The reduction of the following syllables was not complete as the second or third syllables were not perceived as unstressed. Furthermore, the first trisyllabic production of words showed a preference for antepenultimate stress, whereas the frequency of this pattern was eventually matched by ultimate stress patterns as the children's age increases.

The main problem in the children's utterances was their undeveloped ability to reduce stress in a word, to enable them to produce unstressed syllables rather than level stressed. This ability was developed with age, and the level–stress patterns decrease. The children managed to produce more and more adult–like stress patterns. The children had usually acquired this ability to reduce the acoustic correlates and produce perceivable unstressed syllables at about 30 months of age.

4.3.3 Summary

In the babbling stage, variable word stress pattern can be observed. Only about 30% of all produced syllables in babbling are produced as unstressed; the majority of syllables are perceived by the hearer as stressed and level stressed. The children cannot control their acoustic cues to mark stress. During babbling infants first have to learn the acoustic consequences of articulatory gestures as well as the auditory–perceptual consequences of acoustic properties of speech. They experiment with the parameters in order to find out how they can be used to achieve the desired result (Perkell et al., 2000; Guenther, 2003). The unmarked case seems to be the production of stressed syllables as they were produced most frequently, followed by syllables perceived as less prominent but not weak. Babble phrases show variable stress patterns with nearly the same dispersion during the whole babbling phase with a slight preference for the production of trochaic patterns with most prominence on the first syllable. The other syllables can also be perceived as less prominent and not unstressed. In the production of words, a clear preference for trochaic patterns can be observed for the production of first disyllabic and trisyllabic words. Even in the phase where babbling and words are both produced, syllable structures produced during babbling are not produced simultaneous for word production. In the production of words the infants are restricted. The first word productions show a preference for trochaic patterns. The other possible metrical patterns have to be developed for word productions even when the metrical pattern occurs for babbling phrases. A significant difference between stress patterns for words and babbles occurs, which supports the conclusion that for the development of prosodic competence an increased linguistic knowledge is required (Behrens and Gut, 2005; Snow, 2006). When meaningful speech begins, a restriction to few word stress patterns can be observed. During the babbling stage, the possibility to produce variable patterns is given. But the implementation of these patterns and their underlying articulatory gestures to word production patterns depends on the frequency of the acoustic stress pattern in word productions as the most frequent metrical stress pattern is produced first (Beckman and Edwards, 2000).

4.4 Development of syllable structures

In all true stress–based languages, the syllable appears to be the stress–bearing unit. Without the recognition of the phonological constituent 'syllable', it is not possible to acquire word stress. As the internal structure of syllables differs from language to language (Blevins, 1995), children have to develop language–specific syllable structure when entering the word production stage. As languages allow syllables with different degrees of complexity, the most unmarked syllable type is CV as all languages have CV syllables (but see Blevins (1995) for potential exceptions). It can be stated that if a language has possible initial and final syllabic clusters containing a particular number *n* of consonants, then clusters with less consonants are also possible. The number of potential clusters is language–specific and has to be learnt by children from input.

During babbling no limitations on syllable structures occur, but at the beginning of the word production stage, initial limitations on syllables are reported (Fikkert, 1994; Demuth and Fee, 1995; Kehoe and Stoel-Gammon, 1997; Fikkert and Freitas, 1997; Rose, 2000; Grimm, 2008) for Dutch, English, French, European Portuguese and German. Studies describing the acquisition of syllable types mostly focus on the development of primary stressed syllables in words (Levelt et al., 2000, 2003). As differences in the production of stress patterns between babbling and words could be observed, the question arises whether differences in the syllable structures of babbling and words also occur. As the acquisition of stress is the main core of this thesis, not only the development of primary stress syllables but also the development of unstressed and secondary stressed syllable patterns will be analysed.

4.4.1 Method

In order to analyse the developmental pattern of syllable structure in babbling as well as in words, the same data from 5 months up to 3 years were used as in Chapter 4.3. This section looks at the development of syllable type during babbling as well as at the stage of meaningful speech for stressed, unstressed and secondary stressed syllables.

Firstly, the syllable structure, in terms of sequences of C(onsonant) and V(owel), was determined. The stage at which the utterance was produced is also marked in terms of b(abbling) and w(ords). The study concentrates only on the most common syllable types which follow the Sonority Sequencing Principle; because of this, clusters with flat or unexpected sonority profiles were discarded. Such considerations should not have any implications for exemplar–based approaches, aside from considerations that pertain to frequency, maybe.

The German language contains a systematic distinction between long and short vowels. Therefore, it is not possible to pinpoint a stage of acquisition in which vowel length is totally random (Kehoe and Lleó, 2003) so these vowels have to be handled separately. The vowel length was therefore marked and long (V:) and short (V) vowel types were treated as different syllable structures to observe. Specifically, it was observed when long vowels were produced and if the production of syllable structures with long vowels depended on the successful production of this syllable type with short vowel.

This left the following syllables types in terms of Cs and Vs (in order of quantity and independent of vowel length): CV, CVC, V, CCV, VC, CCVC, CVCC, CCVCC, VCC.

In an analysis of the acquisition of syllable structures in Dutch, four different steps in the development of syllable structures were found (Levelt et al., 2000) and ranked with a set of universal, violable constraints on output representations (Prince and Smolensky, 2004; Pater, 2004) in a language–specific way. As the acquisition of prosodic pattern in German is comparable to Dutch, it can be assumed that the acquisition of syllable structures should also be comparable. The four stages found by Levelt et al. (2000) for the acquisition of syllable structures will be evaluated in this section.

4.4.2 Results

Two different types of statistical analysis were conducted. As the data is non–metrical and nominal, cross tabulations and Pearson's χ^2–Test were used for descriptive analysis to describe the different developmental patterns in babbling and word production. A developmental order of acquisition of syllable structures of words was then deduced by aligning the CV–structures on a Guttman scale. Guttman scaling is a procedure for obtaining an order and for seeing to what extent this order is followed.

Cross Tabulations

In order to ascertain the children's different developmental patterns, the syllable structures were cross–tabulated and arranged according to their frequency of occurrence. With the help of these cross tabulations, a ranking of the CV–syllable structures was done depending on developmental stage. For clarity, the children were grouped according to their age in five different groups as described in Table 4.3. For Ollie no data from the period 19 to 30 months were analysed. Emma and Rike's data were only analysed up to the age of 30 months.

During the babbling stage, a great variability in the produced syllable structures were observed. The focus was on the described syllable types, however, other syllables types were also produced during babbling but often only once and therefore only by chance. Furthermore, once the syllable structures were analysed using Pearson's χ^2 Test, no significant differences in the production of syllable types depending on stress were observed. This production of variable syllable structures lends evidence to the assumption that during babbling no restriction on the production of different syllable structures exist. No target pattern has been found during babbling production and the considered 'syllables' are more the consequences of the consonant–vowel alternations than the consequences of target syllable productions.

With increasing control over the coordination of articulators, segmental content elements viz. as consonants and vowels become separate entities. With the referential ability to associate these segmental strings to particular meanings first words were produced.

Between 12 and 18 months of age, syllables ranked as babbles showed nearly the same variety of syllable structures as observed for babbled utterances between 5 and 9 months of age. Syllables ranked as words were restricted in their syllable structures as Pearson's χ^2 Test for syllable type depending on developmental stage was significant ($\chi^2(43,10150)=1462.819$; $p<0.001$). Not all syllable structures which were produced during babbling can simultanious be produced in words, even at the same developmental stage. For babbling ($\chi^2(20,3537)=90.244$; $p<0.001$) as well as for words ($\chi^2(18,3537)=177.444$; $p<0.001$), stress have significant effect on the produced syllable types. Stress has a restricting influence on the production of different syllable structures. The greatest variability of syllable structures were found on stressed syllables. Unstressed syllables showed a lesser variability in the produced syllable structures. For the production of the unstressed and level stress structures the stressed counterpart has to be produced first. First word production showed no evidence for the assumed core syllable 'CV' (Fikkert, 1994; Levelt et al., 2000) as first words contained (when stressed) CV as well as CVC syllables. The first unstressed syllable produced was CV. Only one child (Nils) was able to produce distinctively long and short vowels in this phase: CV as well as CV: syllable types were observed.

Babbling decreased with age, as more and more words were acquired. Between 19 and 24 months of age only two children still produced babbles as well as words. From this time on, long vowels were also produced, but only in word production. Emma was the only child with no distinction between short and long vowels. The other children produced more and more syllable structures with different vowel lengths. First, the syllable type occurred with short vowel. When the syllable type was acquired with short vowel, the distinction between vowel lengths were acquired. With increasing age also the syllable structures expand. A significant effect of stress on the production of different syllable structures still exists ($\chi^2(32,1505)=804.558$; $p<0.001$).

Scaling

Although the data were categorised into five age–dependent groups, they still contain different stages of development. Children of the same age can still be at different stages of development. In order to compare the syllabic development of the six children, the data were arranged on a Guttman scale for syllable type. The Guttman scale serves to establish an order, which helps in observing to what extent this order is followed (Barton, 1976). As there does not appear to be a restriction on syllable structures in babbling, however, this data were not scaled.

Table 4.4 shows the Guttman scale for all stressed syllables found in utterances produced up to 18 months of age, for all children. As far as syllable structure was concerned, stressed syllables were acquired first and their unstressed counterparts could only be produced later. In order to see the developmental order better, only the derived Guttman

Table 4.4: Guttman scaling until 18 months of age.

	cv	cv:	cvc	v	vc	ccv	ccvc	cvcc	ccvcc	vcc
Ollie	+		+							
Hanna	+		+	+						
Nils	+	+	+	+	+	+				
Emma	+		+	+	+	+	+			
Rike	+		+	+	+	+	+			
Bennie	+		+			+	+			

scale for stressed syllable structures is shown. In Appendix B.4 the derived Guttman scales for stressed, unstressed and syllables with secondary stress structures for this age were given. Syllable structures are arranged from left to right, whereas participants are arranged from top to bottom. A "+" indicates that a participant has produced the syllable type at least twice in the period under supervision. When the syllable type was produced only once or has not been found in the data, the box is left empty. The participants are arranged according to their acquisition rate and therefore can change position in the various scales. This indicates that some participants acquired syllable structures at a higher rate than others. The rows and columns are arranged depending on the empty boxes to the left. The fewer empty boxes to the left of the line, the better the fit.

CV and CVC were the first syllable structures to be acquired. In the German data, the assumed core syllable CV were not found as CV and CVC were both produced by the children in their first word productions.

Two developmental patterns can be observed as Bennie has empty boxes for V and VC whereas the other children went in line with the scaling. The developmental pattern of Bennie was different from the other children. Therefore, the participants are split in two different groups. Group A describes the acquisition of syllable types of Hanna, Nils, Emma and Rike. Bennie's developmental pattern is described separately in group B. However, as only one child showed a different developmental pattern, the underlying order of acquisition has to be handled with care. The same holds true for the development of unstressed syllable types. The unstressed counterpart can only be developed once the stressed syllable type has acquired. The unstressed counterparts succeed the production of stressed syllable types and follow the same developmental order as stressed syllable types. However, no specific developmental order for syllables with secondary stress were observed. E.g., Nils produced CV and VC with secondary stress but not CVC or V. In Rike's case, she produced both CV and CCV also with secondary stress at this stage but not CVC, V, or VC, which all precede the production of CCV in developmental order. However, both children acquired these syllable structures in their stressed versions. It could be that they were able to use the existing syllable structures to produce the syllables with secondary stress, and did not have to create a separate category for secondary stress.

Appendix B.6 shows the scaling for both unstressed and secondary stressed syllable

Table 4.5: Guttman scaling until 24 months of age.

	cv	cv:	cvc	cv:c	v	v:	vc	ccv	ccv:	ccvc	cvcc	ccvcc	vcc
Hanna	+	+	+	+	+		+						
Nils	+	+	+		+		+	+					
Emma	+		+		+		+	+		+			
Rike	+	+	+	+	+	+	+	+	+	+	+		

Group A

	cv	cv:	cvc	cv:c	ccv	ccvc	v	vc	cvcc	vcc	ccvcc
Bennie	+	+	+	+	+	+	+				

Group B

Table 4.6: Guttman scaling until 30 months of age.

	cv	cv:	cvc	cv:c	v	v:	vc	ccv	ccv:	ccvc	ccv:c	cvcc	ccvcc	vcc
Hanna	+	+	+	+	+		+	+	+	+		+		
Nils	+	+	+	+	+	+		+	+	+		+		
Emma	+		+		+		+	+		+		+	+	
Rike	+	+	+	+	+	+	+	+	+	+	+	+	+	

Group A

	cv	cv:	cvc	cv:c	ccv	ccvc	v	v:	vc	cvcc	vcc	ccvcc
Bennie	+	+	+	+	+	+	+	+	+	+		

Group B

Table 4.7: Guttman scaling until 36 months of age.

	cv	cv:	cvc	cv:c	v	vc	ccv	ccv:	ccvc	ccv:c	cvcc	cv:cc	ccvcc	vcc
Ollie	+	+	+	+		+								
Nils	+	+	+	+	+	+	+	+	+	+	+	+		
Hanna	+	+	+	+	+	+	+	+	+	+	+		+	

Group A

	cv	cvc	ccv	ccvc	v	vc	cvcc	vcc	ccvcc
	cv:	cv:c	ccv:	ccv:c	v:	v:c	cv:cc	v:cc	
Bennie	+	+	+	+	+	+	+	+	

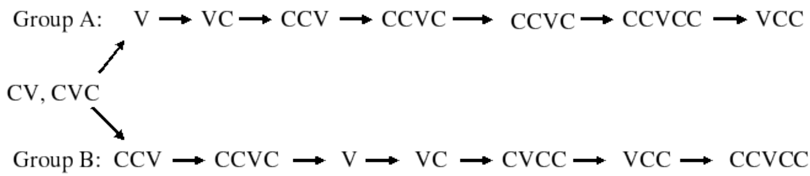
Group B

structures. The production of syllable structures containing long vowels do not to be dependent on stress as, in some cases, long vowels first occur in unstressed syllables types and not in stressed ones.

Table 4.7 shows the Guttman scale up to the age of 36 months. There are only a few empty boxes left now, mainly for syllable types with long vowels. The same holds true for the development of unstressed and secondary stressed syllables as given in Appendix B.7. Some empty boxes still remain for the production of secondary stressed vowels. At this stage, it is clearly shown that it is not necessary to create a new category for the production of secondary stressed vowels. For the production of secondary stressed vowels, the syllable type for stressed vowels are the base.

The missing syllable structures for Ollie could be due to the fact that it was only possible to analyse one recording session of the picture naming task for him. There was not adequate chance to produce all the syllable structures.

In summary, from the scales above, two different orders of acquisition of syllable structures can be deduced, which is in line with the statistical approach of (Levelt et al., 1999a, 2000) (but also see (Rose, in press) for a criticism of aspects of this work).



For long vowels, first the syllable type is acquired with the short vowel before vowel length distinction can be produced for this syllable type.

These developmental steps also occur in the production of unstressed syllables. The unstressed syllable type is only acquired once its stressed counterpart has been acquired.

As the frequency of input plays an important role for the development of a new category in exemplar theory, the developmental steps were compared with the syllable frequencies in German. The available corpus for this kind of analysis is mainly derived from dictionaries and newspapers (e.g. CELEX for German).

As child–directed speech may have a different frequency distribution of syllable structures than CELEX, short stories especially written for children were used for this frequency count. In the Appendix B.8 the resulting syllable structures and corresponding proportion of all syllables can be found.

Differences in vowel length cannot be produced before the syllable type with a short vowel can be produced. To conclude, the production of different vowel lengths does not depend on the frequency of the individual syllable, but rather on the overall frequency independent of vowel length. To compare the syllable structure frequency with the analysed developmental order, independent of vowel length, the frequencies found for syllables with short vowels and those with long vowels are added together. In Table 4.7 the developmental order of the nine most commonly used syllable structures in the short stories is given.

The developmental order corresponds quite well with the underlying frequency order of the different syllable structures in German. Variation is found for syllables containing

Figure 4.7: Syllable structure and corresponding proportion of all syllable tokens

Syllable	Tokens	Frequency	Developmental order
CV	9526	35,25	1
CVC	7610	28,16	1
VC	1751	6,48	3(A)/5(B)
V	1245	4,61	2(A)/4(B)
CCV	962	3,56	4(A)/2(B)
CVCC	957	3,54	6
CCVC	832	3,08	5(A)/3(B)
VCC	780	2,89	8(A)/7(B)
CCVCC	206	0,76	7(A)/8(B)

clusters. Both groups first acquire onset clusters before clusters at the coda. This can be due to the fact that the four syllable structures have nearly the same frequency of occurrence. The main developmental order was that first V and VC was acquired and then complex onset clusters. Only one boy acquired complex onset clusters before V and VC. Complex coda, which is less frequent in German, is acquired last in the developmental orders from both groups. The development of syllable structures depends on stress. Complex codas are relevant for stress placement in German. Complex onsets are universally irrelevant for stress (Goedemans, 1998). The stressed syllable structure has to be acquired first before the unstressed counterpart can be produced.

One possible explanation for the different developmental steps found for just one boy could be that his parental input consisted of more words with complex onset clusters than the parental input of the other children. The parents of this boy grown up in a different dialectal region (so called ’Russlanddeutsche’) to the other parents. Maybe this played a role. However, this is only a speculation as the analysis of the entire input of the children is beyond this work.

As the findings are based on a naturalistic corpus, these findings should be confirmed experimentally, as naturalistic data studies can only ever provide correlational evidence and will provide the impetus for future studies.

4.4.3 Summary

During the babbling stage, no restrictions on the production of different syllabic structures exist. Up to the age of nine months, where only babble phrases were produced, a great variety of different syllable types – including clusters containing more than two consonants – could be observed. No difference in the production of syllable structures depending on stress could be observed. The production of syllables during babbling does not be restricted by the underlying stress pattern.

With increasing age, when babbles and words are produced side by side, the picture changes for babbles and words. For babble phrases, a great variety of syllable structures can still be observed. It has been shown that stressed syllable structures are acquired first and therefore can be classified as unmarked. The unstressed counterparts are acquired later in the same developmental order as the stressed syllable structures. The developmental order of stressed and unstressed syllable structure corresponds well with the frequency of the various syllable structures in German.

The most frequent syllable structures in the input are acquired first and the less frequent and more complex syllable structures relatively late. This is in line with an exemplar–based model of acquisition as the frequency of input plays an important role for the production. During development, a kind of mental syllabary has to be built up for the most frequent syllable structures in German. As the development of unstressed syllable structures shows the same order as for stressed ones, the question arises how stress is represented in the mental lexicon.

4.5 Discussion

In this chapter, both the prosodic and the syllabic development of stress is described. Both words and non–words as well as babble vocalisation from six children were compared according to age, stage of production and word stress pattern. A language specific trochaic bias in the development of prosodic structure during word production stage is assumed. During babbling, no restriction of the syllable pattern seems to occur.

During babbling, every possible acoustic pattern was produced, with different frequencies of occurrence. The picture changes with the production of the first words. Generally, during the production of first words the child associates stress with initial syllables. Furthermore, it has been shown that it is not the production of stressed syllables that has to be learnt, rather syllable reduction, i.e. the ability to produce unstressed rather than level–stressed syllables. To get to this level, children have to achieve control over the acoustic parameters of stress.

As babbling and words show significantly different word stress patterns, it can be concluded that a certain amount of linguistic knowledge is required for the production of words and the underlying prosodic pattern.

In the second part of this chapter the acquisition of syllables – more specifically, the acquisition of different syllable structures – as the stress–bearing unit, was described for stressed, unstressed and level–stressed syllables.

During the babbling stage, a great variety of different syllable structures tend to be produced. The syllable structures depend more on the rhythmic mandibular cycle and the resulting consonant–vowel alternations than on underlying targets. As the children becomes older, they learn to divide the segmental content of a syllable into separate entities and with the beginning of word production, they have to develop first syllable structures. There was no restriction on syllable structure observed in the babbling phase. Only the production of words seemed to follow an ordered syllable structure development. With the help of the Guttman scale, a developmental order was deduced for stressed as well as unstressed syllables. The developmental pattern which was found corresponds well with the frequency of the different syllable structures and goes along with an exemplar–based model of acquisition. No developmental pattern for syllables perceived as secondary stressed can be found as not all syllable structures have to be acquired for the production of secondary stressed syllable structures. For the production of different stressed syllable first the stressed syllable structure has to be acquired.

To conclude, for stressed as well as unstressed syllables, a cognitive map of the various syllable structures has to be developed as first the stressed syllable types can be produced before the unstressed counterpart follows. Following the statistical approach the most frequent syllables in the input were acquired first as the most frequent perceived exemplars were stored. Different categories for stressed and unstressed syllable structures are built up during acquisition for the direct production (Levelt et al., 1999b; Varley and Whiteside, 2001). For secondary stressed syllables no separate category for the different syllable structure is build up as they are not frequent enough in German to build an own category. In the case of secondary stressed syllables a different phonological encoding has to take place.

5 Phonetic development of stress

5.1 Introduction

Knowledge about the prosody of a language facilitates the segmentation of the speech stream. Prosody acts as a filter and is based on phonological representation as well as on their acoustic realisation in the speech signal. Sensitivity to prosodic cues may be helpful for word segmentation (Shukla et al., 2007) as words are invariably aligned with phrasal prosodic edges and help building up a language–specific lexicon (Swingley, 2005). However, prosody does not only help children to segment words (Echols and Marti, 2004) but also gives rise to the predominant stress pattern of the native language(Jusczyk et al., 1993a). Even pre–linguistic infants have reported to be sensitive to aspects of stress and encode stress information as part of their proto–lexical representations (Curtin et al., 2005). These proto–lexical representations are established in the pre–linguistic period. During babbling these perceived representations are linked to different possible neuromuscular commands to the vocal tract (Kirchner, 1999). With the beginning of word production The perception–production loop helps to build up more adult–like representations which are used when children begin to establish a lexicon (Werker and Fennell, 2004). For the development of representations word stress plays a crucial role.

Nevertheless, children do not only have to learn the acoustic patterns occurring in their language (Maye et al., 2002). With the beginning of meaningful speech they also have to discover the connection between the acoustic/prosodic cues and the underlying syntactic constituents (Behrens and Gut, 2005; Thiessen and Saffran, 2007). As soon as these language–specific patterns are discovered, children can segment words better and more easily from the speech stream, because a combination of features is more reliable for segmentation than a single feature (Christiansen et al., 1998).

Sensitivity to metrical stress is necessary for the detection of word boundaries and word recognition. By accessing metrical stress, subsequent detailed phonological representations have to be built up during acquisition. When children and adults have difficulties with constructing, maintaining, and retrieving phonological representations, this deficit can results in dyslexia, as phonological deficits have been assumed to lie at the heart of developmental dyslexia (de Bree et al., 2008). They found that as at–risk children are less sensitive towards metrical stress,they have problems constructing phonological representations from early on. Van Alphen et al. (2007) compared the role of word stress for the comprehension and production of three–year–old children with a familial risk of dyslexia and normally developed children. They showed that a reduced ability to use word stress for the identification of word boundaries is related to phonological problems dyslexic children have in general with perception and production. Children and adults with dyslexia display difficulties in the construction of phonological representations.

To conclude, metrical stress is one cue for building up a mental lexicon. Each perceived phonological representation is stored in the mental lexicon as an exemplar. As the categories represented by exemplar distribution are plastic, they can be modified by perceived exemplars. If this system cannot be created, or if problems occur during the development of this neural network, phonological deficits like developmental dyslexia can emerge (van Alphen et al., 2007; de Bree et al., 2008). Infants discover the specific prosody properties of their native language very early, as pre–linguistic infants are sensitive to the rhythm of their native language (Ramus, 2002; Christophe et al., 2003b). They are sensitive to the predominant stress pattern (Jusczyk et al., 1993a) as well as to phrase structure (Jusczyk et al., 1992). The target language have a meta–phonological influence even on later babbling but only subliminally in an incompletely developed form, as precursors (de Boysson-Bardies et al., 1984). But this underlying influence is neither completely determinable auditorily nor via acoustical measurement (Engstrand et al., 2003). One possible explanation could be that babbling utterances are temporally misaligned to the relevant channels in the phonetic score of the adult listeners (House, 1996).

During the time from the production of first babbling to the onset of language–specific sounds the infants have to discover, organise and consolidate the input of their native language on the basis of their own perceptual and articulatory capabilities (Davis et al., 2002). For reaching adult–like production of prosody, the pre–linguistic infant has to require fine motor adjustments affecting fundamental frequency, timing and intensity over more than one syllable. Control of the acoustic cues from first babbling to the onset of word use is essential for the prosodic learning mechanisms. Every child has to learn the language–specific relationship of the phonetic cues with their underlying hierarchical prosodic (and intonational) representation (Pierrehumbert, 2003b). During the babbling phase, the different prominences of syllables often reflect a combination of biological predispositions rather than segmental constraints as in word production. For an integration of prosody and segmental features, additional lexical experience appears to be necessary (DePaolis et al., 2008). When word production begins to overtake babbled utterances in infants' vocalisations, the integration of adult–like prosodic cues into the underlying segmental patterns has to be learnt and activated (Behrens and Gut, 2005).

5.2 Background

Stress seems to be one reliable cue for segmenting the speech stream. To recognise the different stress patterns, children must be sensitive to the correlates of stress in the acoustic signal. Infants are sensitive to the distribution of stress in the phrase, and distribution effects alter their segmentation strategies (Thiessen and Saffran, 2007). English–learning infants about seven to nine months of age incorporate information about stress into their representations of word forms. For instance, after hearing “DObita””, infants recognise that “doBIta” is different (Curtin et al., 2005). Their knowledge about stress influences subsequent word recognition and word segmentation, as the stress pattern of the parsed sequence was encoded, too. In addition to focusing on perception, we should also consider the production of stress in more detail, as only few studies explore children’s acquisition of the phonetic features of stress (Allen and Hawkins, 1980; Pollock et al., 1993; Kehoe et al., 1995; Davis et al., 2000). Given the perception–production–loop, the successful production of a category depends on the number and quality of input exemplars (Pierrehumbert, 2003a) and their connections to articulatory–acoustic gestures.

In order to study stress development, it is necessary to define what is meant by “stress”. Stress refers to the prominence of a syllable (word stress) or of a word in a sentence (accent) and is associated with a number of acoustic features. Acoustic and perceptual studies of adult speech reveal that the stressed syllable in German speech is marked with increased syllable duration, increased fundamental frequency, increased intensity, often measured as spectral tilt, and changes in the quality of the vowel (Pompino-Marschall, 1990; Heuft and Portele, 1994; Jessen et al., 1995; Heuft et al., 1995; Dogil and Williams, 1999). Despite an array of studies with German–speaking adults, there are few reports on the acoustic correlates of stress in young children. Most of than deal with children whose native language is English. They deal with children at the word stage (Allen and Hawkins, 1980; Allen, 1981; Pollock et al., 1993) or at the babbling stage (Kent and Murray, 1982; de Boysson-Bardies et al., 1989; Halle et al., 1991; Whalen et al., 1991; Robb and Tyler, 1995), dealing with only one prosodic cue of stress. Only a few studies examine the prosodic correlates of stress in babbling (Davis et al., 2000) or the onset of word productions (Kehoe et al., 1995; Engstrand et al., 2003; DePaolis et al., 2008). Infants are able to produce different prosodic patterns by varying the acoustic correlates of prosody at an early age. The control of pitch appears to be limited in the pre–linguistic period but by the time multiword combinations are produced, a re–organisation of F_0 takes place (Snow, 2006), and F_0 is produced in a more adult–like manner. The predominant pattern measured across three–, seven–, and nine–month–old infants was a falling F_0 contour (Kent and Murray, 1982). This holds true for English reduplicate babbling of pre–linguistic infants. For French–learning infants falling and rising patterns were equally produced (Whalen et al., 1991). These differences in intonation pattern seem to depend on the prosodic differences between the two languages.

A study by Davis et al. (2000) concerns the correlates of stress in babbling. They wanted to quantify the impression that pre–linguistic babbling often seems remarkably speech–like because of the adult–like prosody involved. Therefore, they compared disyllabic sequences from five infants and five adults in terms of the use of fundamental frequency, intensity and duration to mark stress. Infants in English–speaking environments produce adult–like stress patterns before they produce stress–specific lexical items. All acoustic correlates of stress were used redundant in a manner largely similar to adults, although children did not display a trochaic stress bias. Based on this findings Davis et al. (2000) concluded that English–learning infants use similar acoustic correlates to produce adult–like stress patterns even before they produce lexical items with the underlying stress placement. They showed that even in the pre–linguistic canonical babbling period, infants demonstrated evidence of coordination of the respiratory, laryngeal and supralaryngeal system for the production of asymmetrical stressed disyllables.

Allen and Hawkins (1980) conducted acoustic analyses of about 100 utterances by three children who were approximately three years of age. The utterances were extracted from natural speech samples, were all two or more words in length, and contained an identifiable nuclear syllable. Fundamental frequency and duration from stressed and unstressed syllables in different phrase positions were measured. In non–final position, pitch change primarily marked stress, whereas in final position, stress was marked primarily by duration. In another series of studies, investigating the perception and production of stress–accent (Allen, 1981), nonsense words were constructed to allow for control over stress placement. The English–speaking children, aged between 3;7 and 6;7 years, had little difficulty in perceiving the difference between words identical in segmental content but differed in stress placement (see Section 3.1.4). However, some of the children had difficulties in producing words with second–syllable stress, which were produced with inappropriate stress or by deleting the initial unstressed syllable. Thus, Allen and Hawkins (1980) claim that stress development involves learning to reduce the acoustic correlates to produce unstressed syllables.

A study by Pollock et al. (1993) supports this argument. In this study the acoustic parameters peak F_0, peak intensity, and absolute duration of stressed and unstressed syllables in two–, three–, and four–year–old children's imitated productions of nonsense words differing in stress placement were examined. The two–year–old children only used duration to mark stress. In contrast, the older children used higher F_0, greater intensity and longer duration to mark stressed syllables. Although all observed age groups were able to use vowel duration to differentiate stressed and unstressed syllables, a continued improvement in the ability to reduce duration for unstressed syllables with age can be observed. The authors conclude that significant developmental changes in the use of acoustic features to mark word stress and in the accuracy of stress placement appear to occur between two and three years of age. For three–year–old children a significant greater accuracy in stress placement were revealed than for two–year–old children.

The acoustic correlates of stress in children's productions of familiar words were examined by Kehoe et al. (1995). In contrast to the previously reported studies, which used non–words, familiar words were used. Children may control the phonetic correlates of stress sooner in productions of familiar words than in imitated production of nonsense words (Hochberg, 1988). They found that children as young as 18 months differentiate stressed and unstressed syllables by exploiting all analysed acoustic parameters. Nevertheless, an increase of intensity to mark stress with increasing age are indicated. As Pollock et al. (1993) had analysed nonsense words, Kehoe et al. (1995) came to the conclusion that children's control over phonetic parameters may emerge sooner in natural speech than in imitated forms in an experimental task. However, in contrast to Pollock et al. (1993)'s study Kehoe et al.'s (1995) study exclusively focused on stress marking in trochaic disyllabic words. In Pollock et al. (1993), both trochaic and iambic disyllabic words were examined. As the most frequent pattern in English is trochaic, it can be concluded that phonetic parameters are more adequately controlled in these types of words and therefore acquired first by children learning English. Furthermore, Schwartz et al. (1996) presented unfamiliar disyllabic words differing only in the placement of primary stress to two–year–old children. The child's output were compared with the input by the ratios of the stressed and unstressed syllables. No adult–like patterns were found. These findings indicate that the magnitude of acoustic differences underlying the distinction between stressed and unstressed syllables were smaller for children than for the observed adult speaker. The acoustic differences need to increase during development for a better differentiation between stressed and unstressed. These findings do not support the view that children simply have to learn to produce reduced unstressed syllables but that they have to learn different acoustic realisations for stressed and unstressed syllables.

In summary, the results of these studies suggest that children as young as 18 months of age are able to control the phonetic features of stress well enough to yield perceptually identifiable stress patterns in most of their productions. This holds true in the production of novel words where they are capable of producing unstressed syllables distinctly from segmentally identical stressed syllables. Nevertheless, children's productions are still considerably variable and are often perceived as stressed incorrectly. Stress development can be characterised as involving increased control over acoustic parameters to derive perceptually identifiable stress contrasts over time.

Given this background, the aim of this chapter is to investigate the use of the different acoustic correlates of stress depending on their development and increasing motor control. In an exemplar–based model of speech development, the input, i.e. parents' speech, has an influence on the development of representations as well as production. Therefore, the parents' input was analysed in addition to children's speech. I compared the childlike and parental use of the different correlates of stress.

The following questions were intended to shed further light on the variable use of the acoustic correlates of stress:

- What are the preferred correlates of stress marking depending on age and child?
- Does the input affect stress marking? Are those utterances directly affected by parental input comparable to the parents' stress patterns?
- What are the preferred correlates of stress marking in the parental data?
- Does the specific speech task (adult–directed or child–directed) affect parents' production of stress ?

5.3 Acoustic correlates of word stress

In German speech, the most reliable parameter to mark stress is vowel duration , followed by intensity (in this section measured as RMS), fundamental frequency, and formants.

Based on this findings, in this chapter Root mean square (RMS) intensity, F_0 contour and the first four formants of each marked vowel at five points of the vowel, at 10, 30, 50, 70 and 90% of the marked vowel duration were analysed using LPC analysis methods (for detailed description see Chapter 3.4) to capture the acoustic realisation of word stress.

In addition to these acoustic parameters, some of the voice quality parameters also have been proposed to be used to mark word stress. Spectral tilt, measured as *Skewness* and *Rate of Closure*, was found to be a good correlate of stress in Dutch (Sluijter, 1995), in German adult speech (Jessen et al., 1995; Claßen et al., 1998) as well as in child–directed speech (Schneider and Möbius, 2007). When children between 2;3 and 6;1 years of age were tested in the TAKI–task an use of the spectral tilt parameters to mark stress could be found (Schneider and Möbius, 2006).

It has been reported (Kent and Murray, 1982; Vallabha and Tuller, 2002) that formant measurements are particularly difficult to perform in high–pitched voices. Different categories of vowels emerge at the age of 20 months (Ishizuka et al., 2007), and a stable production of F1 is observed at the beginning of first word productions at about 15 months of age (see Chapter 6).

From 15 months of age, formants can be categorised to different vowels and determined as vowel–specific. Therefore, the present study also included analyses of voice quality parameters for utterances from the tested children older than 15 months of age. For calculating the voice quality parameters based on the acoustic signal, an analysis tool based on spectral differences of the corresponding harmonics and amplitudes of the first four formants was used (Pützer and Wokurek, 2006).

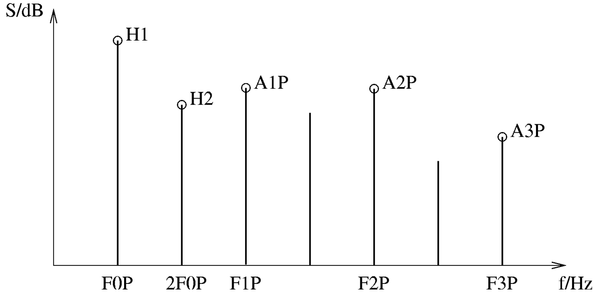


Figure 5.1: Schematic spectral lines from narrow band spectrum and LPC based formant measurements and formant model for the estimation of source spectrum.

5.3.1 Voice quality measurement

The used analysing tool extract vowel quality parameters from the acoustic signal as the glottal vibration can be classified with acoustic measurements of spectral differences (Stevens and Hanson, 1994; Hanson, 1995, 1997). The common structure of these estimators are ratios of spectral peak amplitudes of the harmonics. Based on these spectral estimates, different aspects of voice quality can be estimated as certain differences. In the first part of the measurement, the strongest spectral peaks of the first two harmonics were measured and logged. Figure 5.1 shows where the logged measured values are found in a schematic line spectrum.

Stevens and Hanson's parameters are based on spectral differences of the corresponding harmonics (H) and amplitudes (A). Formant corrections were restricted to F1 and F2. The presence of that correction is denoted by appending a star (*) to the variable name. Stevens and Hanson did not include bandwidth (B) in a consistent manner in their calculation. To describe the incompleteness of closure of the vocal folds, only the bandwidth of the first formant is used. The main parameters for calculations are:

F0	fundamental frequency
F1, F2, F3, F4	formant frequencies
B1, B2, B3, B4	bandwidth of the formants
F0P, 2F0P	frequency of the spectral peaks of the first two harmonics
H1, H2	amplitudes of the first two harmonics
F1P, F2P, F3P, F4P	frequency of the spectral peaks of the formants
A1P, A2P, A3P, A4P	amplitudes of the spectral peaks

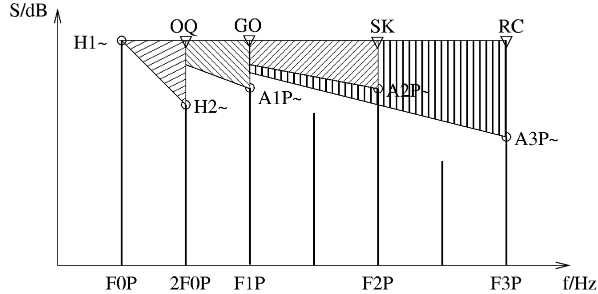


Figure 5.2: Definition of the voice quality parameters as gradients of the spectral height (amplitude ratios in dB).

To reduce the influence of vowel quality, Pützer and Wokurek (2006) elaborated upon the work of Stevens and Hanson and calculated spectral descent gradients instead of the difference of amplitude. Figure 5.2 shows the definition of the spectral descent gradients. The first triangle describes the differences between the first and the second harmonics and is correlated with the open quotient *OQ* (Stevens and Hanson, 1994; Sluijter, 1995; Claßen et al., 1998). The second triangle describes the gradient from the first harmonics and the amplitude of the highest harmonics near to the first formant. This triangle is correlated with the glottal opening *GO* (Stevens and Hanson, 1994; Sluijter, 1995; Claßen et al., 1998). The third triangle describes the gradient from the first harmonic to the amplitude of the highest harmonics near the second formant. This gradient is introduced as rate of skewness *SK* (Claßen et al., 1998). The fourth and final triangle describes the gradient from the first harmonic to the amplitude of the highest harmonics near the third formant. This gradient is described as rate of closure *RC* (Claßen et al., 1998). The bandwidth *B1* corresponds to the parameter completeness of closure and depends on F1. A normalised version of this parameter, in which the influence of the first formant on the *CC* values is minimised, is calculated as incompleteness of closure *IC*. The effect of the formants on the amplitude of the first and second harmonics is corrected (Wokurek and Pützer, 2003). The presence of that correction is denoted by appending a tilde (˜) to the variable name. The resonance gain of each of the four formants is subtracted. In this calculations frequencies and bandwidths were considered. Table (5.1) illustrates, how the parameters are calculated from the acoustic measurements in the work of Stevens and Hanson (1994); Hanson (1995, 1997) and the differences in the work of Pützer and Wokurek (2006).

Table 5.1: Voice quality parameters and their calculation according to Stevens and Hanson (1994) and further development by Pützer and Wokurek (2006).

acoustic	Voice quality parameter	calculation by Stevens/Hanson (1994)	calculation by Pützer/Wokurek (2006)
$2 * F_0$	opening quotient (OQ)	H1* -H2*	$OQ\tilde{}= H1\tilde{}- H2\tilde{}$
$F1$	glottal opening (GO)	H1* -A1	$GO\tilde{}= H1\tilde{}- A1\tilde{}$
	spectral tilt	H1* -A3*	
$F2$	skewness (SK)		$SK\tilde{}= H1\tilde{}- A2\tilde{}$
$F3$	rate of closure (RC)		$RC\tilde{}= H1\tilde{}- A3\tilde{}$
$F4$	$T4$		$T4\tilde{}= H1\tilde{}- A4\tilde{}$
–	completeness of closure (CC)	$B1$	$CC = B1$
–	incompleteness of closure (IC)		$IC = B1/F1$

Open quotient (OQ)

The Open Quotient *OQ* indicates the time during which the glottis is open, defined in the time domain as a fraction of the total period. *OQ* can be described via considering the differences between the first two harmonic peaks with the formant influence removed.

The primary acoustic manifestation of a narrow glottal pulse, i.e. of a decrease in open time, is a reduction of the amplitude of the fundamental component in the source spectrum relative to adjacent harmonics.

```
OQ˜ =  H1˜ - H2˜
```

If either the spectral peak of fundamental frequency is under 70 Hz and thus too low, or the octave distance between the fundamental wave and the second harmonic varies more than 10%, the first error condition ERR1 is produced. In this case, the harmonic structure is not reliably represented in the actual spectrum, and the affected values are removed from further analysis.

The Open Quotient *OQ* is considered to be a correlate of sentence accent and is possibly caused by an additional modulation of muscular tension (Sluijter, 1995; Lintfert and Wokurek, 2005).

Glottal opening (GO)

The parameter Glottal Opening (*GO*) describes the degree of opening over the entire glottal cycle. The amplitude of the first formant (A1) is influenced by the glottal aperture during the open phase. A correction is made for the effect of F1–F4 on H1 and F2–F4 on A1 (Stevens and Hanson, 1994).

```
GO˜ =  H1˜ - A1˜
```

If the frequency of the second harmonic is too close to or above the first formant, the second error condition ERR2 is produced. The affected values are removed from further analysis.

Glottal opening *GO* has been shown to be a correlate of stress in German adults (Jessen et al., 1995; Claßen et al., 1998). In adult speech, stressed vowels are produced with lower GO values than unstressed ones.

Spectral tilt parameters

At a higher vocal effort, for example during word stress, the formation of the glottal impulse demonstrates a distinct asymmetrical form when producing voiced signals. The stronger the vocalisation, the faster and steeper is the subsequent phase of the glottal cycle. This effect seems to depend on the increase of the sub–glottal pressure of the vocalisation: A difference in the intensity can be observed. Stressed vowels demonstrate a higher signal amplitude in the middle– and high–frequency spectrum than unstressed vowels. The ratio of the distribution between the higher and lower frequencies of the spectrum can be considered as correlates of word stress in German (Jessen et al., 1995; Claßen et al., 1998). Spectral tilt describes the slope of the spectrum by comparing the decrease of the amplitudes of the higher frequencies with respect to the amplitudes of the lower frequencies in the spectrum. The parameter *Skewness*, which corresponds to the abruptness of the glottal closure, and the parameter *Rate of closure*, which corresponds to the velocity of the glottal closure, are known to be good indicators for spectral tilt (Sluijter, 1995; Claßen et al., 1998). Unstressed vowels tend to have greater spectral tilt values than stressed ones. To sum up, for the production of stressed vowels the vocal folds are closed faster and more abruptly than unstressed vowels, which results in higher spectral tilt values for unstressed vowels compared to stressed ones (Jessen et al., 1995; Claßen et al., 1998).

- **Skewness (SK)**
  The abruptness of glottal closure influences the spectrum of the glottal waveform at mid and high frequencies.

  ```
  SK˜ =  H1˜ - A2˜
  ```

  Stressed vowels are expected to be produced with more abrupt closures of the vocal folds. Stressed vowels demonstrates smaller values of *SK* than do unstressed vowels.

- **Rate of closure (RC)**
  Duration of the closing part, which directly influences the skewness of the glottal pulse.

  ```
  RC˜ =  H1˜ - A3˜
  ```

  Stressed vowels are expected to be produced with faster closed vocal folds and show smaller values of *RC* than unstressed vowels.

- **T4**
  As spectral tilt appears to be a reliable correlate of word stress in German, the set of voice quality parameters was extended by the influence of the fourth formant.

  The parameter *T4* is an extrapolation of the parameters *GO*, *SK* and *RC* and was motivated by the results from Sluijter (1995) and Claßen et al. (1998). He corresponds to the calculation of the parameters *GO*, *SK* and *RC* as the next highest formant (F4) and its amplitude is involved.

  ```
  T4˜ =  H1˜ - A4˜
  ```

  The *T4* values for stressed and unstressed vowels were expected to be comparable to the values found for the parameters *SK* and *RC*.

Glottal leakage paramters

Completeness of Closure (CC) corresponds to the energy loss in the F1 range, adding significantly to the F1 bandwidth (B1) when the glottis is not completely closed during phonation.

```
CC = B1
```

Stressed vowels demonstrate higher CC values corresponding to a greater F1 bandwidth (B1) (Sluijter, 1995) as stressed vowels have a greater glottal leakage than unstressed vowels. CC appeared to be vowel specific. For this reason a normalised version of this parameter called *Incompleteness of Closure* (IC) is also calculated, in which the influence of the first formant on the CC values is minimised.

```
IC = B1/F1
```

The parameter *IC* is a good indicator for word stress in German, at least in adults (Schneider and Möbius, 2007). Unstressed vowels are produced with a greater glottal leakage than stressed vowels.

5.4 Results

The present chapter investigates the development of young children's use of fundamental frequency, vowel duration, intensity and different voice quality parameters to indicate word stress between 5 and 36 months of age. In order to contrast child speech with adult stress realisation, the acoustic parameters in the parents' utterances were additionally analysed.

Method

Each vowel in polysyllabic utterances (babbles, non–words, and words) of the six examined children between 5 and 36 months of age as well as their parents was analysed using the above–mentioned LPC–based analysis method. For Nils's and Bennie's parents, child–directed speech (spontaneous and picture naming task) and adult–directed speech (picture naming task) were analysed. For Hanna's parents only child–directed speech (spontaneous and picture naming task) was obtained; no adult–directed speech recordings are available. For Emma's mother, two different recording tasks are available for analysis: first, child–directed speech during the interaction with her own child; second, as Emma's mother had also recorded the evaluation data in the primary school (Chapter 6.3), child–directed speech to older children (six years of age and not her own). For Ollie's and Rike's mothers, only child–directed data were analysed. During babbling, there are no clear preferences for marking stress. Therefore, the perceived prominence of the produced syllables were annotated. First all prominent syllables were separated from syllables without any salience (unstressed). Second, prominent syllables were categorised depending on their strength as syllables bear primary stress (most prominent) or secondary stress (less prominent). In this section, only bi– and polysyllabic utterances were analysed, because in monosyllable utterances, the default production is stressed.

Statistics

For each parameter, analyses of variance using the General Linear Model, Univariate procedure in SPSS were carried out, with the analysed parameters as the dependent variable and stress and the child's age as factors on the significance level of $p<0.05$. For vowel duration as well as RMS intensity, the position of the syllable in the utterance (initial or final syllable) was also used as a factor, as vowel position in the utterance can have an effect on both of these parameters (Allen and Hawkins, 1980; Schneider and Möbius, 2007).

As voice quality parameters are influenced by vowel identity, the z–transformed values were analysed. When no significance of the normalised data occurred the raw data were also analysed to several at least tendencies. The results of the statistical analyses are discussed separately for each subject. Significant effects were further analysed by carrying out the Tamhane post hoc test.

5.4.1 Vowel duration

Vowel duration (as well as the overall syllable duration and the durations of other parts of the syllable) is regarded as the primary cue to German word stress (Jessen et al., 1995; Dogil and Williams, 1999). Vowel duration of stressed syllables as well as the duration of closure in plosives preceding vowels are significantly longer than in unstressed syllables.

Previous results on the production of contrastive stress showed that vowel duration is the most reliable cue for marking stress in older children (Schneider and Lintfert, 2005; Lintfert and Schneider, 2006; Schneider et al., 2006; Schneider and Möbius, 2006) as well as in adults (Jessen et al., 1995; Claßen et al., 1998) and in child–directed speech (Schneider, 2007).

Vowel duration is vowel–specific and can vary according depending on the recording sessions. Therefore only normalised data (see 3.4.4) were analysed to describe the development of this parameter to mark word stress for the children and their parents.

Previous results for children younger than two years (Lintfert and Schneider, 2006; Lintfert and Vollmer, 2007) showed that for these children, vowel duration does not constitute a stable parameter for marking word stress as the children varied in the use of this parameter. Hence, the development of the parameter vowel duration is reported for all six children separately.

Children's use of vowel duration

Vowel duration has been demonstrated to be the most reliable cue for word stress in German adults (Jessen et al., 1995; Claßen et al., 1998; Dogil and Williams, 1999).

Univariate ANOVAs were carried out for each child with vowel duration as dependent variable and stress, vowel position and child's age as factors. Table 5.2 shows results of the ANOVAs for each child.

For all children a main effect of stress on vowel duration was observed. Except in two children, the interaction of stress and age was also significant, which provides evidence for a developmental pattern and variable use of vowel duration to mark stress.

In summary, stressed vowels were produced with longer duration than unstressed vowels and secondary stressed vowels, but not in the same way at all ages. The mean duration of all vowels increased until the age of 12 months and then decreased. The children generally did not produce significant differences in vowel duration between unstressed and secondary stressed vowels.

Thus, children do not only have to learn to lengthen the stressed syllables, since to produce stressed and unstressed syllables in an utterance, the distinction between the two must be clearcut. To produce this difference, the underlying acoustic pattern has to be learnt. To conclude, in order to produced unstressed syllables, children rather have to learn the reduction of syllable length than the lengthening of a stressed vowel.

In German speech, the reduction of vowel length often goes along with reduced vowel quality. Unstressed vowels are more affected by the reduction in vowel quality than

Table 5.2: Individual results of the ANOVAs for parameter **duration**.

child	source	degree of freedom	F Statistic	p Value
Hanna	stress	2,1124	6.872	**0.001***
	position	2,1124	1.298	0.273
	age	8,1124	0.741	0.655
	stress*age	16,1124	1.284	0.199
	stress*position	4,1124	2.012	0.091
	stress*age*position	25,1124	1.405	0.089
Nils	stress	2,2015	5.308	**0.005***
	position	2,2015	15.373	**0.000***
	age	8,2015	2.103	**0.032***
	stress*age	14,2015	3.040	**0.000***
	stress*position	4,2015	0.419	0.795
	stress*age*position	25,015	1.550	**0.040***
Bennie	stress	2,1731	7.899	**0.000***
	position	2,1731	16.290	**0.000***
	age	7,1731	1.521	0.156
	stress*age	14,1731	2.066	**0.011***
	stress*position	4,1731	2.174	0.070
	stress*age*position	25,1731	2.528	**0.000**
Rike	stress	2,1237	3.920	**0.020***
	position	2,1237	4.416	**0.012***
	age	8,1237	1.101	0.360
	stress*age	14,1282	3.003	**0.000***
	stress*position	4,1237	0.660	0.620
	stress*age*position	24,1237	1.587	**0.036**
Emma	stress	2,1958	11.708	**0.000***
	position	2,1958	11.078	**0.000***
	age	7,1958	0.719	0.656
	stress*age	14,1958	1.166	0.295
	stress*position	4,1958	0.223	0.926
	stress*age*position	28,1958	1.012	0.448
Ollie	stress	2,377	6.920	**0.001***
	position	2,377	5.267	**0.006***
	age	4,377	0.532	0.712
	stress*age	8,377	3.487	**0.001***
	stress*position	4,377	1.720	0.145
	stress*age*position	15,377	0.510	0.935

stressed vowels, as only full vowels can carry word stress. The amount of this reduction depends on the context. Therefore, to produce a perceivable difference between stressed and unstressed syllables in an utterance, children do not only have to learn to shorten the vowel but also to reduce vowel quality. sThe vowels in the final syllable were mostly produced with longer duration than the vowels in initial and medial syllables. Only one boy showed no tendencies for final syllable lengthening. This production pattern provides evidence for final syllable lengthening for five of the six children.

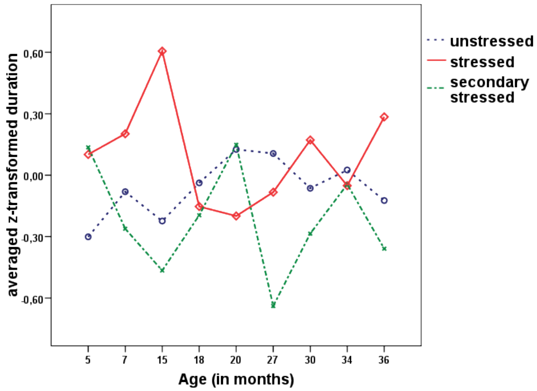


Figure 5.3: Z–transformed duration of **Hanna** depending on her age.

Hanna

For Hanna, univariate ANOVAs showed a significant main effect of stress. The factors vowel position and child's age were not significant. Stressed and unstressed syllables were produced with significantly different vowel durations (Tamhane post hoc test), but not in a consistent way.

In Figure 5.3 the averaged and normalised mean duration for all prominences depending on the child's age is given. At the age of 27 months, stressed vowels were produced with a lower average duration than unstressed vowels. At the age of 34 months, averaged vowel durations were very similar. Apart from this exceptions, Hanna produced stressed vowels with a longer duration than unstressed vowels. Vowels perceived as secondary stressed are mostly produced with the shortest vowel duration.

A post–hoc test (Tamhane) did not show any significant differences in the vowel duration between unstressed and vowel with secondary stress, but both differed significantly from stressed vowels. Hanna produced vowel with secondary stress with almost the same vowel duration as unstressed vowels.

Considering final syllable lengthening, Hanna produced vowels in final syllables with a significant increase in duration as opposed to initial or medial syllables. This tendency could be observed for stressed as well as for unstressed vowels but not for secondary stressed vowels.

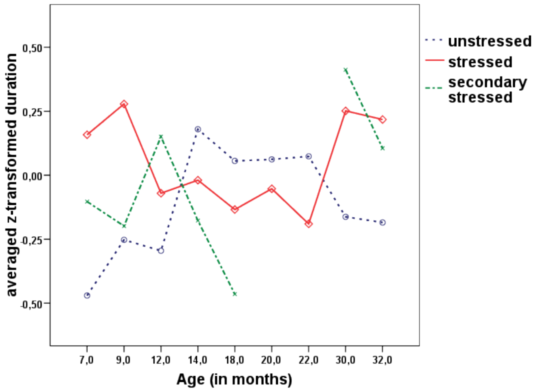


Figure 5.4: Z–transformed duration of **Nils** depending on his age.

Nils

For Nils, the factors stress, child's age, as well as vowel position revealed a significant main effect. A developmental pattern was indicated by the significant interaction between age and stress as well as between stress, child's age and vowel position.

Nils produced syllables bearing secondary stress with significantly different vowel duration as compared to stressed and unstressed syllables. Between stressed and unstressed syllables, no significant difference were found (Tamhane post hoc).

In Figure 5.4 the averaged z–transformed duration for all prominences depending on the child's age is shown. Nils's use of vowel duration varied depending on age. Further post hoc tests for each age group were carried out. If vowel duration was used to mark word stress, then unstressed vowels were significantly shorter as stressed ones. First, secondary stressed vowels were produced more like unstressed vowels. However, with the beginning of word production, vowel duration of vowels with secondary stress became more like stressed vowels. Between 12 and 22 months of age, vowel duration was not used to differentiate between stressed and unstressed vowels.

The factor vowel position was also significant. Initial syllables were produced with the shortest vowel duration, and final syllables with the longest duration independent of word stress. Thus, Nils seemed to produce final syllable lengthening.

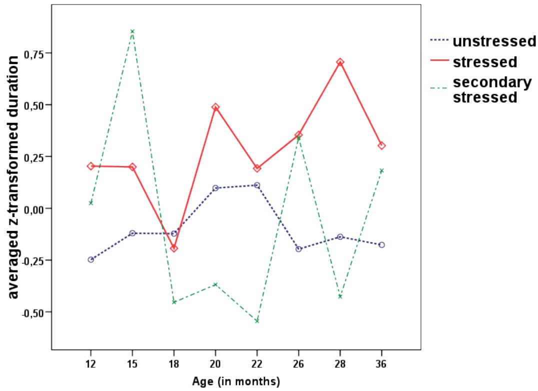


Figure 5.5: Z–transformed duration of **Bennie** depending on his age.

Bennie

For Bennie, the univariate ANOVA revealed a significant main effect of stress and vowel position. No main effects of the factor child's age were observed. A developmental pattern can be seen, as age had a significant influence on stress. The interaction of stress*age*position was also significant.

Unstressed vowels were produced with a significantly shorter vowel duration than stressed ones. For vowels with secondary stress, vowel length did not differ to the stressed and the unstressed versions (Tamhane post hoc). In Figure 5.5 the developmental pattern of vowel duration as a function of age is illustrated. A variable use of vowel duration to mark word stress can be seen. Except at 18 and 26 months of age, stressed vowels demonstrated longer averaged durations than unstressed ones. Vowels with secondary stress did not show a clear pattern. When babbles as well as words were produced, vowel duration was not used to differentiate between stressed and unstressed vowels. With increasing age, the difference of vowel duration for stressed and unstressed vowels becomes significant. From the age of 26 months on, Bennie used vowel duration in a consistent way to distinguish between stressed and unstressed vowels.

The factor vowel position was significant. Initial syllables were produced with the shortest vowel duration independent of stress. Bennie produced vowels in the final syllables with longest duration, but until the age of 36 months with no significant differences to vowels in medial syllables. Final syllable lengthening was not produced in a significant way.

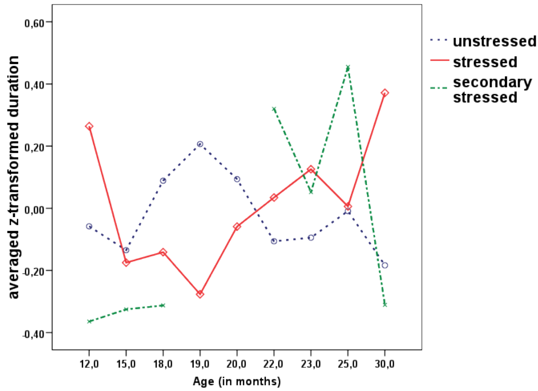


Figure 5.6: Z–transformed duration of **Rike** depending on her age.

Rike

For Rike, a significant influence of stress and vowel position in the utterance as well as of age on vowel duration could be observed. The developmental pattern was reflected in the significant interaction of stress, age, and position.

Vowel with secondary stress and stressed vowels were produced with significantly different vowel durations as well as stressed and unstressed vowels (Tamhane post hoc). No significant difference could be observed between secondary and unstressed vowels.

In Figure 5.6 this developmental pattern is shown.

Only at 12 and 30 months of age, Rike produced unstressed vowels with significantly shorter vowel duration than stressed vowels. For the other periods, no significant differences in vowel duration could be observed but a tendency: With increasing age, vowel duration was used to mark stress in the expected way. Vowels in initial syllables were produced with significantly shorter duration than vowels in final syllables.

Emma

For Emma, univariate ANOVA showed a significant main effect of stress and vowel position. No further main effects nor interactions reached significance. Emma produced unstressed vowels with significantly shortest vowel duration. No significant difference in vowel duration between stress and secondary stress was produced (Tamhane post doc). In Figure 5.7 this developmental pattern is given for Emma.

In the pre–linguistic phase, vowels with secondary stress were mainly produced like unstressed vowels with shorter vowel duration than stressed vowels. This picture

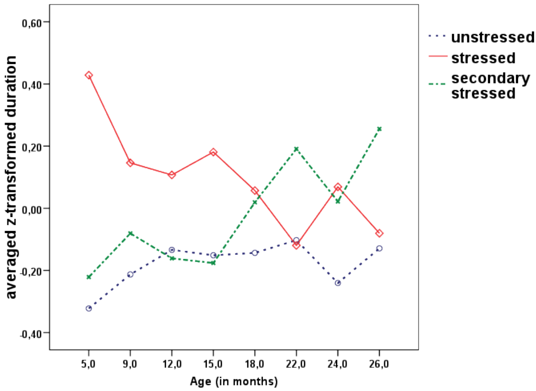


Figure 5.7: Z–transformed duration of **Emma** depending on her age.

changed at 18 months of age. Vowels with secondary stress were now produced with comparable vowel duration as stressed vowels. At 26 months of age, vowels with secondary stress were produced with the longest vowel duration.

The factor vowel position was significant. Until the age of 12 months and with the beginning of meaningful speech at about 26 months, vowels in final syllables were produced with the longest duration in the utterance. Emma thus seemed to have produced final syllable lengthening.

Ollie

For Ollie, a significant main effect of stress as well as of the factor vowel position can be found. The factor child's age was significant only in the interaction with stress and not as main factor. Post hoc tests did not reveal any significant differences between the observed stress patterns. In Figure 5.8 it is illustrated that unstressed vowels were produced with shorter vowel duration than stressed vowels. However, until the age of 33 months, the variation was too great for a significant use of vowel duration to be observed. Until this age, vowel duration was not used in an adult–like way to mark stress.

The factor vowel position was significant. Vowels in final and initial syllables demonstrated significantly different vowel durations. In final syllables, longer vowel durations were observed similar to initial syllables. No significant difference between initial and medial or medial and final position could be observed. Ollie seems to have produced final syllable lengthening.

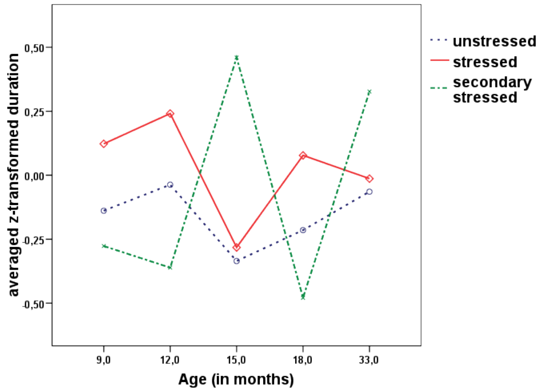


Figure 5.8: Z–transformed vowel duration of **Ollie** depending on his age.

Parents' use of vowel duration

Univariate ANOVAs were carried out with vowel duration as an independent variable and stress and vowel position as factors for each adult. An additional factor was age of the listener. It was necessary to explore whether parents made significant differences depending on the speech task (child–directed speech or adult–directed speech) and on the child's age. Vowel duration was used by all parents in this study as the main parameter to mark word stress in German. Figure 5.9 shows the z–transformed vowel durations for all parents. Stressed vowels were produced with the longest duration. Vowels with secondary stress were mainly produced like stressed syllables and not like unstressed ones. To conclude, reduction of vowel duration is the marked case and has to be learnt for producing different stress patterns, signalling that a syllable is reduced and produced with no salience. All parents tended to produce the vowel in the final syllable with the longest duration in the utterance.

Hanna's parents (CH, HH)

Univariate ANOVAs of vowel duration as dependent variable showed a significant influence of stress for Hanna's mother (CH) ($F(1,107)= 4.499$; $p<0.05$) as well as for her father (HH) ($F(2,49)= 3.565$, $p<0.05$). The factors vowel position as well as child's age showed no main effect for Hanna's parents. Both of them produced stressed vowels with longer vowel duration than unstressed vowels, independent of Hanna's age. For vowels with secondary stress no clear use of vowel duration could be observed. For Hanna's mother (CH), vowel position showed no main effect nor interaction.

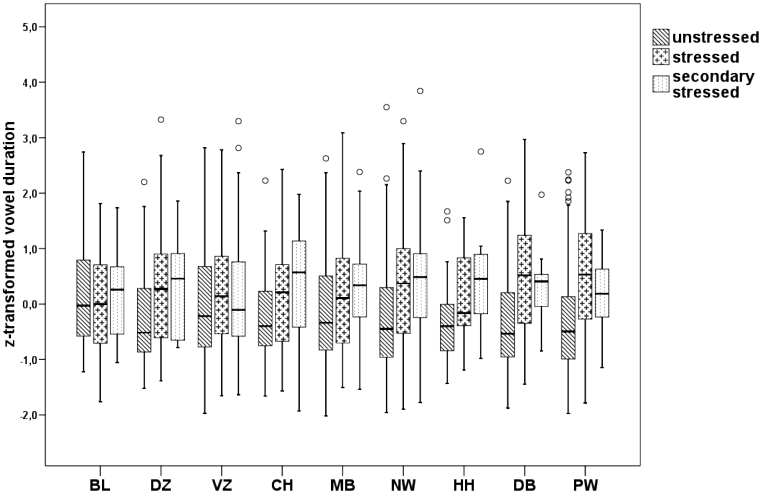


Figure 5.9: Z–transformed durations for each adult.

With regard to Hanna's father (HH), the interaction of stress*position was significant ($F(2,39)=5.531$; $p<0.05$). He produced vowels in the final syllable with a longer duration than vowels in initial and medial syllables. Adult–directed speech data were not available for these parents.

Nils' parents (MB, DB)

For Nils' mother (MB), stress ($F(2,745)=3.162$; $p<0.05$), vowel position ($F(2,745)=6.403$, $p<0.05$), and the interaction of stress*position ($F(3,745)=2.655$, $p<0.05$) were significant. For the father (DB), stress ($F(2,277)=40.351$; $p<0.05$), vowel position ($F(2,277)= 7.818$; $p<0.05$) and the interaction of stress*position ($F(4,277)=1.024$; $p<0.05$) were significant. The factor child's age was not significant for Nils' parents.

Nils' father (DB) as well as his mother (MB) produced stressed vowels with a significantly durational increase as opposed to unstressed vowels. Vowels with secondary stress were produced similarly to stressed vowels by both the father and the mother (Tamhane post hoc). Neither child's age nor the interaction of age with stress were significant. No differences between child–directed and adult–directed speech in the use of vowel duration to mark stress were observed for Nils' parents. Syllable position showed significant influence on vowel duration. Both parents produced stressed and unstressed vowels with a significant increase in segment length in final position.

Bennie's parents (NW, PW)

The univariate ANOVA was significant for stress for both the mother (NW) ($F(2,853)=9.539$; $p<0.05$) and the father (PW) ($F(2,417)=43.041$; $p<0.05$). Neither child's age nor vowel position were significant for Bennie's parents, or any interaction of the observed factors.

Both parents produced stressed syllables with significantly longer vowel durations than unstressed vowels. Bennie's father (PW) produced unstressed vowels with the shortest vowel duration, followed by vowels with secondary stress. Stressed vowels were produced with the longest duration. Bennie's mother (NW) did not display any significant difference between the production of stress and secondary stress. Unstressed vowels were produced with the significantly shortest duration. Both parents produced vowel duration independent of the child's age. In addition, no differences in vowel duration depending on the speech task could be observed. Their use of vowel duration to mark word stress did not differ between child–directed and adult–directed speech. The factor vowel position was not significant for both parents; tendencies for position–dependent vowel durations were observed (post hoc Tamhane): Vowels in final syllables were produced with longer duration than in initial syllables by both parents.

Rike's mother (VZ)

For Rike's mother (VZ), a significant influence of stress ($F(2,729)=8.214$; $p<0.05$) was found. There was a significant interaction between the factors vowel position and stress ($F(4,729)=6.234$; $p<0.05$). The factor child's age and vowel position did not show any no significant main effect.

VZ's use of vowel duration to mark word stress was independent of her child's age and of the vowel position in the utterance. She produced unstressed syllables with a significantly shorter vowel duration than stressed vowels. Vowels with secondary stress were produced without a significant difference in vowel duration to stressed or unstressed vowels. There was a significant interaction between the factors vowel position and stress. Vowel duration in final syllables was longer than in initial or medial positions.

Emma's mother (BL)

The univariate ANOVA was significant for stress ($F(1,214)=11.243$; $p<0.05$) and for the position in the utterance ($F(2,214)=8.128$; $p<0.05$). The factor child's age showed no effect. Independent of the children's age, Emma's mother produced stressed vowels with a significantly longer duration than unstressed vowels. No significant difference to vowels with secondary stress could be observed. Since syllable position was also significant, the vowels of the final syllable were produced with the longest duration in the utterance.

Ollie's mother (DZ)

The univariate ANOVA was significant for stress ($F(2,142)=5.046$, $p<0.05$) and for vowel position ($F(2,118)=7.224$; $p<0.05$). Age was not significant, and neither was any interaction between the factors.

Stressed vowels were produced with a significantly longer vowel duration than unstressed vowels. Vowel duration was only used to differentiate between stressed and unstressed vowels. Vowels in the initial syllable were produced with a significantly shorter duration than in medial or final syllables. Vowels in final syllables were produced with the longest vowel duration.

5.4.2 RMS–Intensity

Intensity has been reported to be a reliable cue to mark stress. In German, stressed vowels are produced with more vocal effort that result in higher intensity (loudness) than unstressed vowels. However, intensity is generally considered to be a weak cue in the perception of linguistic stress, as the contrast between the intensity of stressed and unstressed vowels is weaker than for other cues, e.g. duration (Isačenko and Schädlich, 1966; Dogil, 1995; Jessen et al., 1995).

One problem with intensity to mark stress is that it can be influenced by shifting formant frequencies due to stress. For an increase in intensity, an increase in the physiological correlate of intensity, the subglottal pressure, is necessary (Borden et al., 1994; Dogil and Williams, 1999). However, if a speaker produces more vocal effort, the higher frequencies increase more than the lower frequencies (Sluijter, 1995).

Therefore, differences in intensity, measured as RMS Intensity, depend on the fine–tuning of the subglottal pressure. To use different intensities to mark word stress, children have to learn to produce vowels with differences in subglottal pressure. They must reduce as well as strengthen subglottal pressure depending on the stress pattern.

Previous results analysing contrastive stress have shown that older children as well as some of the analysed mothers varied intensity based on syllable position in the word. No comparable variation was found for word stress (Schneider and Möbius, 2006, 2007).

As intensity is vowel–specific and also depends on the particular recording session, only normalised data were analysed to describe the development of this parameter to mark word stress for the children and their parents.

Table 5.3: Individual results of the ANOVAs for parameter **RMS Intensity**.

child	source	degree of freedom	F Statistic	p Value
Hanna	stress	2,5191	16.881	**0.000***
	position	2,4509	0.170	0.844
	age	8,5191	4.176	**0.000***
	stress*age	16,5191	7.632	**0.000***
	stress*position	2,4509	0.718	0.488
	stress*age*position	14,4509	2.856	**0.000***
Nils	stress	2,9720	18.485	**0.000***
	position	2,9720	4.893	**0.008***
	age	8,9720	3.009	**0.002***
	stress*age	14,9720	5.588	**0.000***
	stress*position	4,9720	3.817	**0.004***
	stress*age*position	24,9720	5.549	**0.000***
Bennie	stress	2,8379	49.928	**0.000***
	position	2,8379	0.027	0.973
	age	7,8379	4.747	**0.000***
	stress*age	14,8379	9.159	**0.000***
	stress*position	4,8379	1.960	0.098
	stress*age*position	24,8379	2.701	**0.000***
Rike	stress	2,5763	34.008	**0.000***
	position	2,5763	7.158	**0.001***
	age	8,5763	3.615	**0.000***
	stress*age	14,5763	6.315	**0.000***
	stress*position	4,5763	11.098	**0.000***
	stress*age*position	24,5763	7.544	**0.000***
Emma	stress	2,7963	31.149	**0.000***
	position	2,7963	34.956	**0.000***
	age	7,7963	8.840	**0.000***
	stress*age	14,7963	11.041	**0.000***
	stress*position	4,7963	3.481	**0.008***
	stress*age*position	27,7963	4.697	**0.000***
Ollie	stress	2,1829	10.094	**0.000***
	position	2,1829	0.439	0.645
	age	4,1829	2.829	**0.024***
	stress*age	8,1829	7.485	**0.000***
	stress*position	4,1829	1.082	0.364
	stress*age*position	14,1829	3.292	**0.000***

Children's use of intensity

Univariate ANOVAs were conducted on children's speech data, with intensity as dependent variable and stress, the child's age and vowel position as factors. For a better overview, the results of the ANOVAs for each child are shown in Table 5.3.

For all children a highly significant main effect of stress and child's age was found. In addition, age–dependent use of intensity can be observed, as the interaction between stress and child's age was also significant for all children. In summary, unstressed vowels were mainly produced with a lower intensity than stressed vowels. For the production of unstressed vowels, a reduction of intensity was realised.

The production of vowel with secondary stress varied depending on the child. Two children (Bennie and Ollie) produced vowels with secondary stress that resembled unstressed vowels rather than stressed ones. These two boys produced stressed vowels with

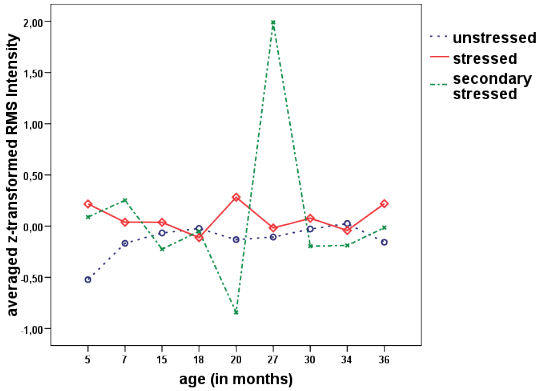


Figure 5.10: Z–transformed intensity of **Hanna** depending on her age.

the highest intensity in the utterance. Hanna produced vowels with secondary stress comparably to stressed vowels. Nils, Emma and Rike distinguished significantly between unstressed, secondary stressed and stressed vowels. Nils and Rike displayed the same tendency to produce vowel with secondary stress: with intensity values between stressed and unstressed vowels. Only Emma produced vowels with secondary stress with the highest intensity values in the utterance. Bennie and Ollie did not show an influence of vowel position on intensity. Apart from Rike, the other children tended to produce vowels in medial position with the highest intensity. Rike produced vowels in initial syllables with the highest intensity and showed a continuous decrease to vowels in final syllables.

Hanna

An univariate ANOVA with intensity as dependent variable showed a significant main effect of stress and age. Significant interactions of stress and child's age provided evidence for a developmental pattern and a variable use of this parameter to mark word stress. The factor vowel position showed no main effect but was significant in interaction with the factors stress and child's age. Apart from the age of 18 and 34 months, stressed vowels were produced with higher intensity than unstressed vowels (see Fig. 5.10).

Vowel with secondary stress were produced more similar to stressed vowels than to unstressed vowels. A post–hoc test (Tamhane) revealed no significant differences between vowels with secondary stress and stressed vowels but between secondary stress and unstressed vowels. Stressed vowels and secondary stress differed significantly from

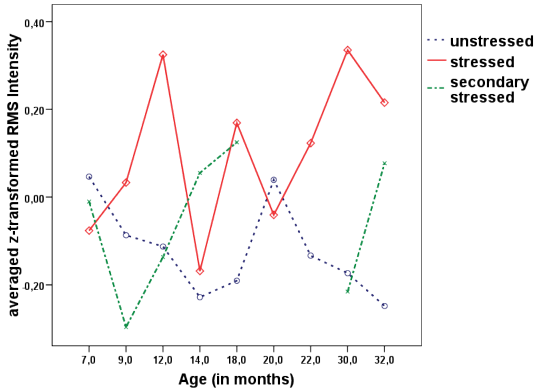


Figure 5.11: Z–transformed intensity of **Nils** depending on his age.

unstressed vowels. Vowels in initial syllables were produced with less intensity. Moreover, an increase in intensity in the medial position could be observed, followed by a decrease in the final position.

Nils

An univariate ANOVA with intensity as a dependent variable and stress, age and position as factors showed significant main effect for all factors as well as significant interactions between all factors.

For Nils, intensity was a reliable cue to mark stress: a post–hoc test revealed significant differences between all three classes of stress. Unstressed vowels demonstrated the lowest normalised values, followed by vowels with secondary stress. In Figure 5.11 it can be seen that stressed vowels had the highest normalised values of intensity except at the ages of 7 and 20 months. A closer look at each age group does not reveal a significant use of intensity to mark stress at the ages of 7, 14, 20 and 22 months. However, the use of intensity to mark stress also depends on vowel position. Vowels in final syllables were produced with less intensity. Vowels in the medial syllables were produced with the highest intensity in the utterance.

Bennie

The univariate ANOVA demonstrated significant main effect of stress and age. The factor vowel position showed no main effect. The interaction between stress, child's age, and vowel position was significant. Stressed vowels were produced with significantly

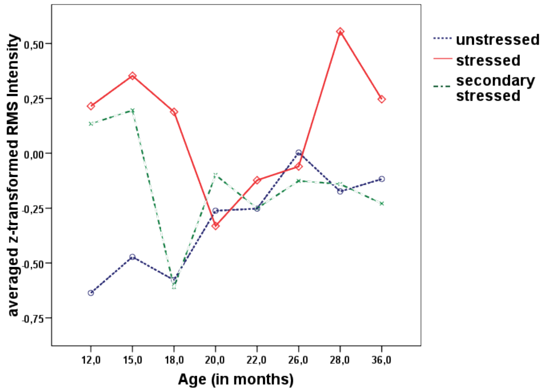


Figure 5.12: Z–transformed intensity of **Bennie** depending on his age.

higher intensity than unstressed and secondarily stressed vowels. Vowels with secondary stress were always produced with medium intensity between stressed and unstressed syllables (post hoc Tamhane). In Figure 5.12 a variable use of intensity is shown. Between 20 and 26 months of age, Bennie produced unstressed vowels with a higher intensity than stressed ones, but not with significant differences.

No differences in the production of intensity depending on the vowel position were observed. Independent of the vowel position in the utterance, stressed vowels were mainly produced with the highest intensity, and unstressed vowels with the lowest intensity.

Rike

For Rike, the univariate ANOVA on intensity as dependent variable revealed significant main effect of all analysed factors as well as interactions between factors.

Further post hoc tests showed that unstressed vowels were produced with significantly lower intensity values than vowels with secondary stress. Vowel with secondary stress were produced with significantly lower intensity values than stressed vowels.

In Figure 5.13 the dependence of intensity on child's age reveals in more detail. During babbling, vowels with secondary stress were mainly produced like stressed vowels and with a higher intensity than unstressed vowels. At stage of the production of polysyllabic words, the production scheme of vowels with secondary stress was comparable to unstressed vowels and produced with less intensity values than stressed vowels. At 18 and 19 month of age, Rike did not use intensity in a significant manner; she tended to produce stressed vowels with a higher intensity than unstressed ones. As the factor

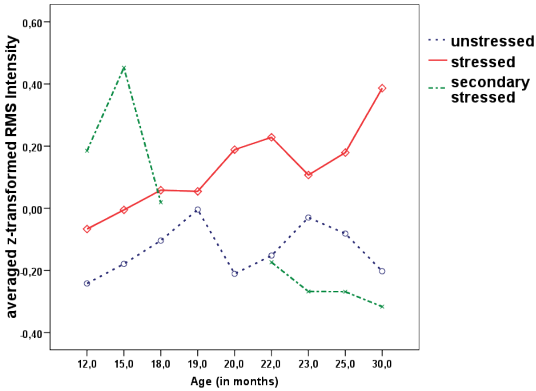


Figure 5.13: Z–transformed intensity of **Rike** depending on her age.

vowel position had a significant main effect on intensity, a decrease in intensity from the beginning to the end of an utterance was observed. Vowels in final syllables were produced with less intensity, and vowels in initial syllables were produced with the highest intensity in an utterance.

Emma

For Emma, the analysis showed significant main effects for all factors as well as for the interactions between factors. Unstressed vowels were generally produced with significantly lower intensity values than stressed vowels. Vowel with secondary stress were produced with significantly higher intensity values than stressed vowels.

In Figure 5.14 a detailed picture of the use of intensity depending on child's age is provided. Only at nine months of age, intensity was not used to mark stress: no significant differences in intensity was observed at this age. During babbling, Emma appeared to play with the parameter, as no clear preference was observed. At five months of age, vowels with secondary stress and unstressed vowels were both produced with less intensity than stressed ones. At nine months, vowels with secondary stress and stressed vowels were produced comparably. At twelve months, Emma used intensity to differentiate vowels with secondary stress from stressed and unstressed vowels. With the beginning of meaningful speech at about 22 months, a stable pattern emerged: unstressed vowels were produced with less intensity and vowels with secondary stress with the highest intensity. Vowel position showed a significant main effect. Emma produced vowels in medial syllables with a higher intensity than in initial and final syllables.

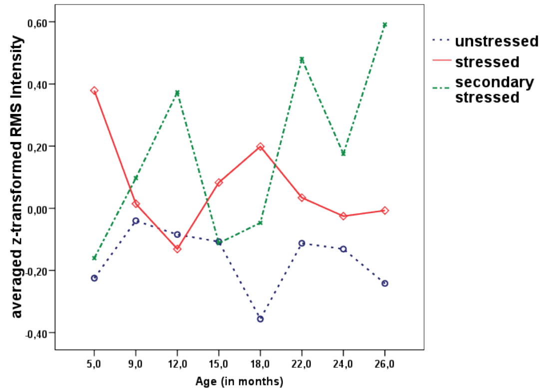


Figure 5.14: Z–transformed intensity of **Emma** depending on her age.

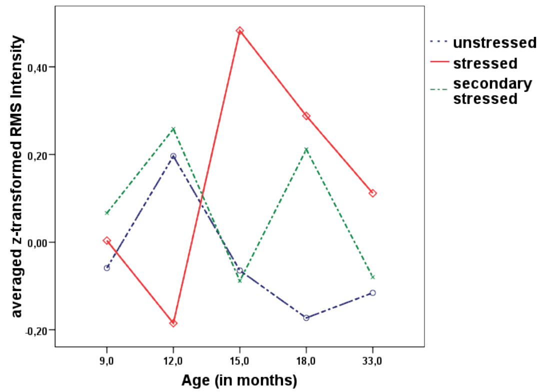


Figure 5.15: Z–transformed intensity of **Ollie** depending on his age.

Ollie

The univariate ANOVA for intensity as dependent variable showed significant main effect of the factors stress and child's age but not for vowel position. A significant interaction between stress and child's age reflects a developmental pattern. Further post hoc tests showed that stressed and unstressed vowels were produced with significant differ-

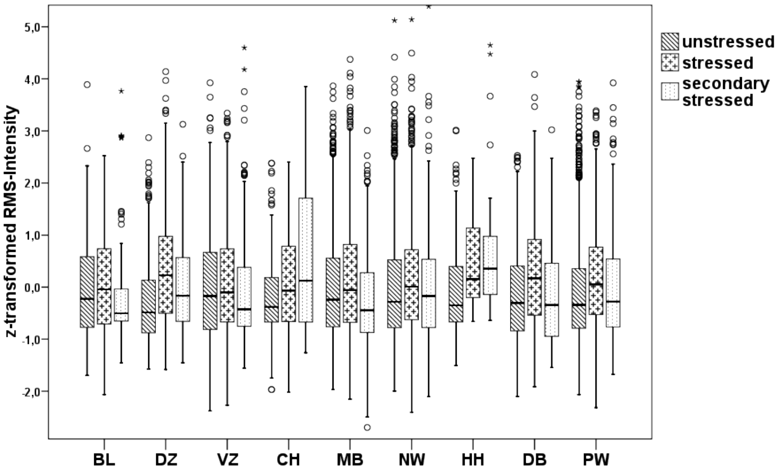


Figure 5.16: Z–transformed intensity for each adult.

ences in intensity. Vowels with secondary stress were produced with intensity values between stressed and unstressed vowels. They showed a comparable tendency to unstressed vowels. In Figure 5.15 the developmental pattern is illustrated. At the beginning of babbling, i.e. at nine months of age, no significant use of intensity was observed. With twelve months of age, stressed and unstressed vowels were produced with significant intensity differences which were contrary to what would have been expected. Stressed vowels were produced with lower intensity values than unstressed vowels. With increasing age, this picture changed. With 15 months of age, stressed vowels were produced with a higher intensity than unstressed vowels. With increasing age, Ollie managed to improve his production of vowels with secondary stress and unstressed syllables with less intensity than stressed ones. Syllables with secondary stress tended to be produced like unstressed syllables. No influence of vowel position in the utterance was observed.

Parents' use of intensity

In Figure 5.16 the parental use of the parameter intensity to mark word stress is illustrated. Vowels with secondary stress were mainly produced like unstressed vowels, although the use of intensity to mark stress depended on the child's age. The parents did not use intensity to mark word stress at every age of the child.

In some cases, vowels with secondary stress were produced with greater intensity than

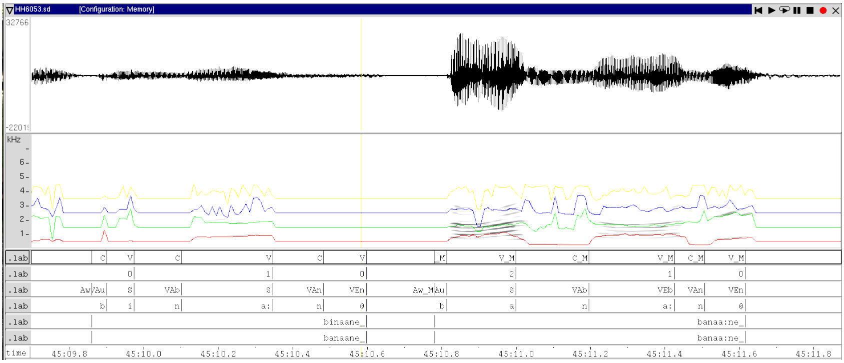


Figure 5.17: Waveform and label files from the wrong child's output [biˈnanə] and the mother's correction with the highest intensity on the child's wrong vowel production [ˌbaˈnanə].

stressed and unstressed vowels, and therefore, further analysis were examined. During the course of acquisition, secondary stress is one cue to keep the child's attention on a wrong production and to correct the child's output (see Figure 5.17). Such emphasised vowels were mainly produced with higher intensity values for the wrong production in the child's output (e.g. wrong vowel quality or a wrong phoneme). Vowels with secondary stress were produced with the highest intensity to keep the child's attention on a wrong production. The use of this parameter therefore depended on the age of the child. To conclude, parents marked syllables with secondary stress with highest intensity value of the vowel to emphasise the right production of stress or, alternatively, to keep the child's attention on a wrong vowel or stress pattern in the child's production. In the other cases, vowels with secondary stress were produced like unstressed vowels. Previous findings of dependency on vowel position but not of word stress (Schneider and Möbius, 2007) are not supported in this data on spontaneous speech. If intensity was used to mark stress, then stressed vowels were produced with significantly greater intensity than unstressed vowels. But intensity was not used in a consistent way to mark stress. Thus, intensity seemed to have been used to emphasise.

Hanna's parents (CH, HH)

For Hanna's mother, an univariate ANOVA showed a significant influence of stress ($F(2,632)=9.186$; $p<0.00$), vowel position ($F(2,632)=15.528$; $p<0.00$), child's age ($F(3,632)=3.396$; $p<0.05$) as well as significant interaction between the factors stress*child's age ($F(5,632)=5.637$; $p<0.00$) and stress*position ($F(4,632)=25.468$; $p<0.00$). Unstressed, stressed, and vowels with secondary stress were produced with significantly distinguishable intensities. Unstressed vowels were produced with less intensity than stressed ones. In general, Hanna's mother used intensity to mark secondarily stressed vowels. She seemed to have used intensity in this case to keep the child's attention on a wrong vowel production. Depending on the vowel position in the utterance, vowels in the final syllable were produced with less intensity.

For Hanna's father, there was also a significant main effect of stress ($F(2,227)=4.583$; $p<0.05$) and vowel position ($F(2,227)=3.912$; $p<0.05$). He produced unstressed syllables with significantly lower intensity values than stressed and secondarily stressed vowels. No significant differences between vowels with secondary stress and stress were produced, in which the latter tended to be produced with the highest intensity. As only data at 36 months of age were obtained for Hanna's father, no conclusion can be drawn on the use of intensity to keep the child's attention on a wrong production. Finally, Hanna's father produced vowels in final position with less intensity as opposed to vowels in medial or initial positions.

Nils' parents (MB, DB)

For Nils' mother, an univariate ANOVA revealed a significant main effect of stress ($F(2,4219)=4.626$; $p<0.05$) and child's age ($F(6,4219)=8.214$; $p<0.00$). Vowel position was not significant. The interaction between stress and age ($F(10,4219)=7.856$; $p<0.00$) as well as between stress, age, and position ($F(11,4219)=8.145$; $p<0.00$) was significant.

Nils's mother used intensity to mark stress with an increasing age of her child. She produced stressed syllables with the highest intensity. Vowels with secondary stress were produced with the lowest intensity. In case of adult–directed speech, Nils's mother used intensity exclusively for marking secondary stress: no significant difference in intensity between stressed and unstressed vowel occurred.

For Nils's father, an univariate ANOVA was significant for stress ($F(2,1429)=31.830$; $p<0.05$). The factors child's age and vowel position showed no effect. The interaction between stress and child's age was significant ($F(2,1429)=5.915$; $p<0.00$).

When Nils's father spoke to him, he produced stressed syllables with the highest intensity and with significant differences in secondary stress and unstressed. When he talked to adults, there was no significant difference between vowels with secondary stress and stressed syllables. Only the difference between stressed and unstressed vowel production remained significant.

Bennie's parents (NW, PW)

For Bennie's mother, no main effect was significant; however, the interaction between stress and child's age was significant ($F(13,5131)=4.979$; $p<0.00$).

Analyses of her child–directed and adult–directed speech showed a dependency of her use of intensity on speech direction. In NW's adult–directed speech, no significant use of intensity to mark stress was found. In the interaction with Bennie, his mother used intensity to mark stress with increasing age of the child. Further analysis of the use of intensity suggested that Bennie's mother used intensity to keep the child's attention on the correct production of given words. Bennie's mother used intensity to differentiate between stressed and unstressed vowels only in the case of the picture naming task.

For Bennie's father, a significant main effect of stress was found ($F(2,2886)=12.759$; $p<0.05$). The interaction between child's age and stress was also significant ($F(7,2886)=2.168$; $p<0.05$). PW used intensity to mark stress in adult–directed speech as well as in child–directed speech: Different uses of intensity were found. In adult–directed speech, vowels with secondary stress were produced with significantly lower intensity than unstressed and stressed vowels. No significant difference between stressed and unstressed vowels were observed. In child–directed speech, Bennie's father produced stressed vowels with significantly greatest intensity values. In the case of emphasis, syllables bear secondary stress with the greatest intensity.

Rike's mother (VZ)

For Rike's mother, no main effect of stress was observed. However, a significant effect of child's age ($F(7,3444)=6.025$; $p<0.000$) and vowel position ($F(2,3444)=10.270$; $p<0.000$) was shown. An interaction of age and stress ($F(14,3444)=5.814$; $p<0.000$) as well as an interaction between child's age and vowel position ($F(16,3444)=6.054$; $p<0.000$) was significant. Stressed vowels were produced with significantly highest intensity, followed by unstressed and secondary stress.

Emma's mother (BL)

For Emma's mother, an univariate ANOVA showed a significant main effect of stress ($F(2,1212)=3.169$; $p<0.05$). Neither the factor child's age nor vowel position reached significance.

However, there was a significant interaction between child's age and stress ($F(10,1212)=3.421$; $p<0.00$) as well as between stress and vowel position ($F(8,1212)=2.524$; $p<0.05$).

When intensity was used to differentiate stress, stressed vowels were produced with higher intensity than unstressed vowels but not significantly. The use of intensity to mark stress depended on the child's age. Emma's mother used intensity to emphasise the correct stress pattern or to keep the child's attention on the correct vowel quality.

Ollie's mother (DZ)

For Ollie's mother, the univariate ANOVA did not show a significant main effect of stress. The factor child's age (F(2,713)=4.718; p<0.05) and vowel position (F(2,713)=14.861; p<0.000) showed a significant main effect. The interaction between stress and position (F(4,713)=6.394; p<0.000) and stress and child's age (F(3,713)=6.025; p<0.000) were also significant. Unstressed vowels were produced with significantly lower intensity values than stressed vowels. No significant difference between vowels with secondary stress and stressed vowels were observed. Moreover, intensity was not used to emphasise stress. Vowels in final syllables were mainly produced with less intensity than in medial or initial syllables.

5.4.3 Fundamental frequency

In addition to duration and intensity, the phonetic realisation of word stress also involves fundamental frequency changes. In German, an increase of F_0 for stressed vowels was previously reported (Jessen et al., 1995; Dogil, 1995). Some studies suggest that for German adult speakers, F_0 is vowel–specific. Therefore, they may not rely on F_0 to mark word stress (Claßen et al., 1998; Dogil and Williams, 1999). As sentence intonation also depends on changes in fundamental frequency, these different results may depend on the different methods of analysis. The principle mechanism for changing the fundamental frequency is to stretch the vocal folds through contraction of the cricothyroid muscle. As an increase in fundamental frequency correlates with an increase of cricothyroid muscle activity (Borden et al., 1994; Sawashima and Hirose, 1983). Children have to learn how to apply more or less longitudinal tension to the vocal folds via the cricothyroid muscles and by varying subglottal pressure for controlling their use of F_0.

Children's use of fundamental frequency

Univariate ANOVAs were carried out on fundamental frequency as dependent variable and stress and child's age as factors. The results of the ANOVA for all children are given in Table 5.4. When age was significant, a regression analysis was conducted to show whether there was an age–dependent significant decrease of F_0 or not. The univariate ANOVA showed a highly significant main effect of stress and of the interaction between stress and age. With increasing age, a tendency towards a better control of fundamental frequency was observed. Regression analyses showed a significant decrease of fundamental frequency with increasing age, apart for Emma. When fundamental frequency was used to mark stress, unstressed vowels were produced with lower fundamental frequency than stressed ones. Whereas some of the children produced vowels with secondary stress similar to stressed ones, others produced them more similar to unstressed vowels. Therefore, no clear production pattern for vowels with secondary stress was observed.

Table 5.4: Individual results of the ANOVA for $\mathbf{F}_0$.

child	source	degree of freedom	F Statistic	p Value
Hanna	stress	2,4710	32.659	**0.000***
	age	8,4710	52.541	**0.000***
	stress*age	16,4710	7.347	**0.000***
Nils	stress	2,9469	65.541	**0.000***
	age	8,9469	179.990	**0.000***
	stress*age	14,9469	25.997	**0.000***
Bennie	stress	2,8179	61.418	**0.000***
	age	7,8179	47.441	**0.000***
	stress*age	14,8179	10.385	**0.000***
Rike	stress	2,5611	48.784	**0.000***
	age	8,5611	76.718	**0.000***
	stress*age	14,5611	9.023	**0.000***
Emma	stress	2,8091	1.035	0.355
	age	7,8091	145.066	**0.000***
	stress*age	14,8091	8.199	**0.000***
Ollie	stress	2,1724	20.185	**0.000***
	age	4,1724	130.243	**0.000***
	stress*age	8,1724	11.527	**0.000***

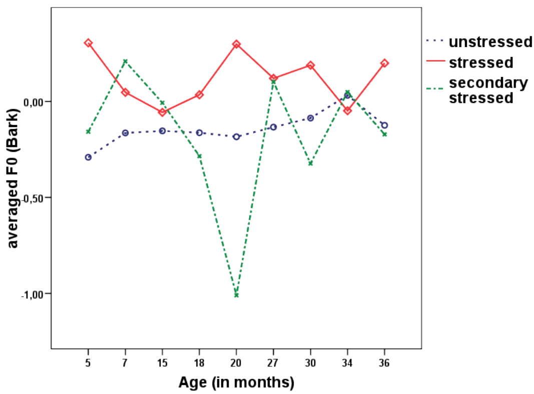


Figure 5.18: Averaged F_0 (in Bark) of **Hanna** depending on her age.

Hanna

The univariate ANOVA demonstrated a significant main effect of stress and child's age. Moreover, the interaction between stress and age was significant. Further analyses showed that stressed vowels were produced with significantly higher fundamental frequency than unstressed vowels. In addition secondarily stressed vowels were

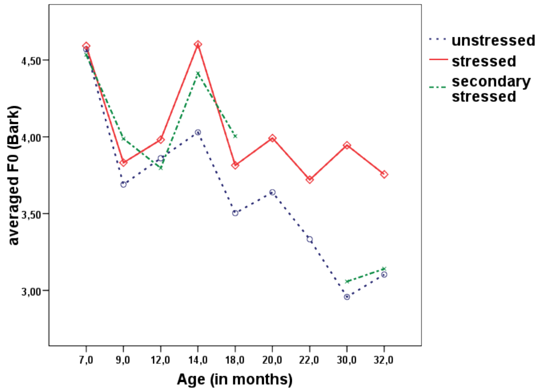


Figure 5.19: Averaged F_0 (in Bark) of **Nils** depending on his age.

produced with significantly lower fundamental frequency. As age was determined to be significant, a regression analysis was carried out for F_0 dependent on child's age. Hanna showed only a slight decrease of fundamental frequency with increasing age ($R=0.067$; $R^2=0.004$; $B=-0.04$; $p<0.000$). In Figure 5.18 it is illustrated that, except at the age of 34 months, Hanna produced stressed vowels with higher fundamental frequency values than unstressed vowels, but not in a consistent manner. An increase in fundamental frequency was observed at 30 and 36 months of age. The production of vowels with secondary stress was inconsistent until the age of 18 months. With the emergence of real words, these vowels were mainly produced by the lowest fundamental frequency.

Nils

For Nils, there were significant main effects for stress and age. The interaction between stress and age was also significant. Further analyses showed that unstressed vowels were produced with the lowest fundamental frequency values, followed by stressed vowels. Vowels with secondary stress were produced with the highest fundamental frequency. The regression analysis demonstrated a significant decrease of fundamental frequency with increasing age ($R=0.325$; $R^2=0.105$; $B=-0.034$; $p<0.000$). Nils always produced unstressed syllables with the lowest F_0 values in the utterance, followed by stressed vowels (see Figure 5.19). During babbling, vowels with secondary stress were produced with the highest F_0. With the beginning of producing words, the production scheme for vowels with secondary stress changed. Significant difference in F_0 between unstressed and secondarily stressed vowels were no longer produced.

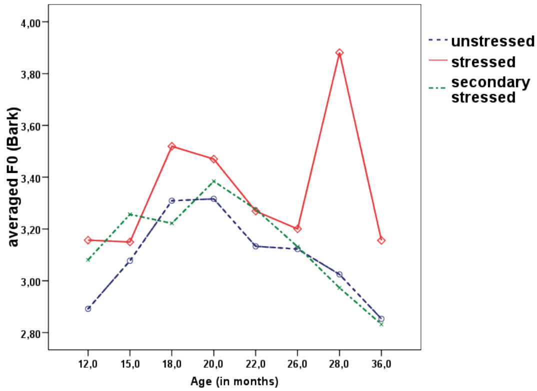


Figure 5.20: Averaged F_0 (in Bark) of **Bennie** depending on his age.

Bennie

Univariate ANOVAs revealed significant main effect for stress and child's age. The interaction between stress and age was also significant. Stressed vowels were always produced with significantly higher fundamental frequency than unstressed vowels. Vowels with secondary stress were produced with fundamental frequency values between stressed and unstressed vowels. In Figure 5.20 the dependency of fundamental frequency on age is shown. Stressed vowels were always produced with higher fundamental frequency values than unstressed vowels. Until 20 months of age, vowels with secondary stress were produced more similar to stressed vowels. After this age, the fundamental frequency of secondarily stressed vowels were more like unstressed vowels. As age as factor was significant, a regression analysis was conducted. In this analysis a significant decrease of fundamental frequency with increasing age was observed for Bennie ($R=0.213$; $R^2=0.045$; $B=-0.014$; $p<0.000$).

Rike

Rike showed a significant main effect of stress and age on F_0. The interaction between stress and child's age was also significant. She produced unstressed vowels with the lowest fundamental frequency, followed by secondarily stressed vowels. Stressed vowels were produced with the highest fundamental frequency. As age had a main effect on fundamental frequency, a regression analysis was conducted. Dependent on age, Rike also showed a significant decrease of F_0 with increasing age ($R=0.239$; $R^2=0.057$; $B=-0.023$; $p<0.001$). In Figure 5.21 the averaged fundamental

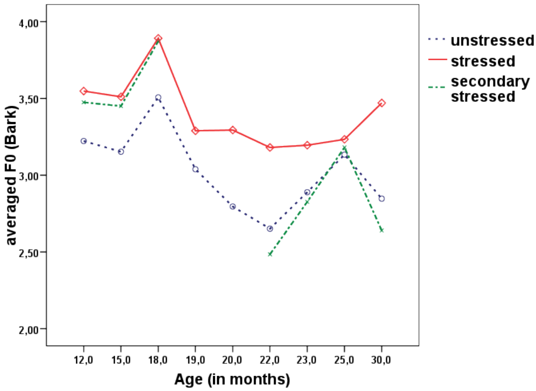


Figure 5.21: Averaged F_0 (in Bark) of **Rike** depending on her age.

frequency related to child's age is given. Unstressed vowels were produced with lowest F_0, and stressed with highest F_0. During babbling, vowels with secondary stress were mainly produced with a similar F_0 than stressed vowels. With the production of polysyllabic words, secondary stress remain less F_0 and were produced like unstressed vowels in order of F_0.

Emma

For Emma, stress had no main effect on F_0 but child's age was significant. The interaction between stress and age was significant. When Emma used F_0 to mark word stress than she produced unstressed vowels with lower F_0 than stressed and secondarily stressed vowels. Vowels with secondary stress were produced like stressed vowels. With 12 months of age, Emma produced stressed vowels with lowest fundamental frequency, followed by secondarily stressed vowels. At this age unstressed vowels were produced with the highest F_0. In Figure 5.22 her inconsistent production pattern of F_0 is shown. Although, Emma had the lowest F_0 at 26 months, no clear pattern could be observed throughout the entire analysis period. The regression analysis showed no significant decrease (R=0.009; R^2=0.000; B=-0.001; p=0.400).

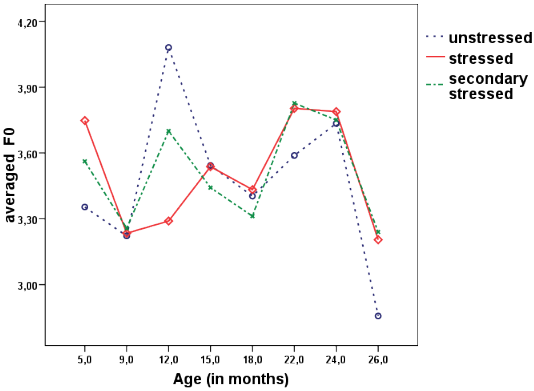


Figure 5.22: Averaged F_0 (in Bark) of **Emma** depending on her age.

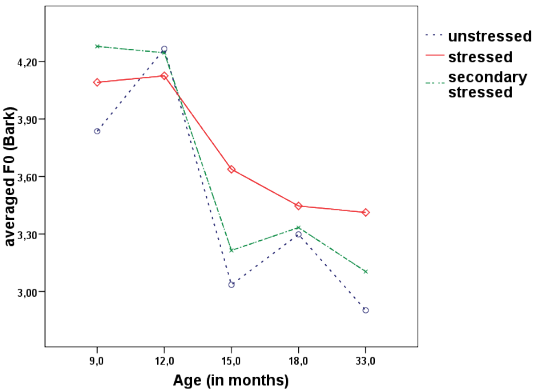


Figure 5.23: Averaged F_0 (in Bark) of **Ollie** depending on his age.

Ollie

The univariate ANOVA showed significant main effect of stress and child's age. Moreover, the interaction between stress and age was significant. Unstressed vowels were produced with significantly lower F_0 values than vowels with secondary stress. Vowels with secondary stress were produced with significantly lower F_0 values than stressed vowels.

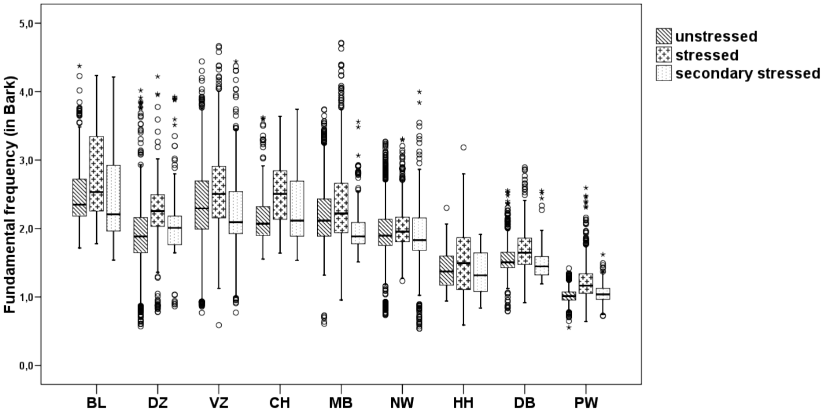


Figure 5.24: Averaged F_0 (in Bark) for each adult.

In Figure 5.23 a clear developmental pattern of the averaged fundamental frequency can be observed. Apart from the age of 12 months fundamental frequency was used to mark stress. Conducted regression analysis of fundamental frequency and age showed a significant decrease of F_0 with increasing age ($R=0.397$; $R^2=0.157$; $B=-0.037$; $p<0.001$).

Analyses of parents' use of F_0

Univariate ANOVAs were carried out for fundamental frequency as dependent variable and stress and child's age as factors for each adult.

All parents produced unstressed vowels with lower F_0 than stressed vowels. The production of vowel with secondary stress depended on the adult speaker, but were mainly produced more similar to unstressed than stressed vowels (Figure 5.24).

When the factor child's age was either as main effect or as between interaction significant, regression analyses were conducted. For the mothers of Hanna, Bennie, Rike and Ollie a significant decrease of F_0 with increasing age of the child were observed.

Hanna's parents (CH, HH)

For Hanna's mother, the univariate ANOVA showed significant main effect of for stress ($F(2,545)=24.089$; $p<0.001$) and child's age ($F(3,599)=4.283$; $p<0.05$). The interaction between stress and child's age was also significant ($F(5,599)=9.159$; $p<0.001$). CH produced stressed vowels with significantly higher F_0 than vowels with secondary stress and unstressed vowels. A regression analysis with the factor child's age showed a significant decrease in F_0 with increasing age of the child ($R=0.186$; $R^2=0.035$; $B=-0.013$; $p<0.001$).

For Hanna's father the univariate ANOVA showed no significant main effect nor interactions of stress and age.

Nils' parents (MB, DB)

For Nils' mother the univariate ANOVA showed a significant main effect of stress ($F(2,3763)=20.368$; $p<0.001$) and child's age ($F(6,3873)=16.818$; $p<0.001$). The interaction between stress and age was significant ($F(12,3873)=13.755$; $p<0.001$).

MB produced stressed vowels with the highest fundamental frequency values and vowels with secondary stress with the lowest F_0 values. Stressed vowels lied between.. The regression analysis with the factor child's age showed no significant decrease of F_0 with increasing age ($R=0.030$; $R^2=0.001$; $B=-0.003$; $p=0.080$).

Nils's mother made no difference in her use of F_0 depending on the speech task. The factor speech task showed no significant main effect.

For Nils' father, the univariate ANOVA showed significant main effects of stress ($F(2,1302)=37.573$; $p<0.001$) and child's age ($F(2,1343)=12.752$; $p<0.001$). The interaction between stress and age was significant ($F(3,1343)=25.070$; $p<0.001$).

DB produced stressed vowels with the highest fundamental frequency. No significant differences between vowels with secondary stress and unstressed vowels were observed.

The factor speech task showed a main effect. Nils's father used different F_0 values depending on the listener. When he had spoken to an adult he used a lower fundamental frequency than when he had spoken to Nils (mainly to the boy when he was younger than 30 months). The regression analysis showed no significant decrease of fundamental frequency dependent on the child's age ($R=0.072$; $R^2=0.005$; $B=-0.012$; $p=0.078$).

Bennie's parents (NW, PW)

For Bennie's mother, a significant main effect of stress ($F(2,3831)=12.744$; $p<0.001$) and child's age ($F(6,3831)=35.152$; $p<0.001$) was found. The interaction between stress and age was also significant ($F(11,3831)=3.913$; $p<0.001$).

NW produced unstressed vowels with lower F_0 values than stressed vowels. Vowels with secondary stress were produced similar to unstressed vowels. The regression analysis with the factor child's age showed a significant decrease of fundamental frequency with increasing age ($R=0.194$; $R^2=0.038$; $B=-0.009$; $p<0.001$). Bennie's mother used higher F_0 values during the babbling phase and lowered her fundamental frequency when Bennie mainly produced words. The fundamental frequency when speaking to the older child is nearly the same as speaking to an adult.

For Bennie's father a significant main effect of stress ($F(2,2046)=41.654$; $p<0.001$) and child's age ($F(4,2046)=15.293$; $p<0.001$) were observed. The interaction between stress and child's age was significant ($F(7,2046)=20.018$; $p<0.001$).

Bennie's father produced unstressed vowels with lowest fundamental frequency values. Stressed vowels were produced with highest F_0 values. The values of vowels with secondary stress lied between the values for stressed and unstressed vowels. A significant main effect of the factor speech task were observed. In adult–directed speech the fundamental frequency was significantly lower than when speaking to the child. A regression analysis for the factor child's age showed a decrease but not significantly (R=0.065; R^2=004; B=-0.003; p=0.006).

Rike's mother (VZ)

For Rike's mother, the univariate ANOVA showed significant main effect of stress ($F(2,3380)=11.625$; $p<0.001$) and child's age ($F(7,3380)=6.030$; $p<0.001$). Also the interaction between stress and age was significant ($F(14,3415)=12.708$; $p<0.001$). VZ produced vowels with secondary stress with the lowest F_0 values, followed by unstressed vowels. Stressed vowels were produced with the highest F_0 values. A significant decrease were observed by the conducted regression analysis for factor child's age (R=0.080; R^2=0.006; B=-0.010; $p<0.001$).

Emma's mother (BL)

For Emma's mother, the univariate ANOVA showed significant main effect of stress ($F(2,1114)=38.072$; $p<0.001$) and child's age ($F(5,1114)=220.695$; $p<0.001$). The interaction between stress and age was also significant ($F(9,1114)=5.481$; $p<0.001$).

BL produced stressed vowels with highest F_0 values than unstressed vowels. Vowels with secondary stress were produced similar to unstressed vowels. Her F_0 values depended on the speech tasks. She used significant different F_0 values when talking to Emma until the age of 26 months and when talking to school children (Tamhane post hoc).

Regression analysis of the age of Emma showed no significant decrease (R=0.007; R^2=0.000; B=-0.002; p=0.854). Her differences in fundamental frequency only depended on the child she talk to and not on Emma's age.

Ollie's mother (DZ)

For Ollie's mother, the univariate ANOVA showed significant main effect of stress ($F(2,671)=3.469$; $p<0.05$) and child's age ($F(2,671)=76.288$; $p<0.001$). The interaction between stress and child's age was also significant ($F(3,671)=6.558$; $p<0.001$).

DZ produced unstressed vowels with the lowest F_0 values. She produced no significant difference in her F_0 values used for stressed and secondarily stressed vowels. A significant decrease of F_0 depending on age was revealed by the regression analysis (R=0.402; R^2=0.161; B=-0.052; $p<0.001$).

5.4.4 Formants

Children's use of formants

Univariate ANOVAs were carried out on the first four formants as dependent variables and stress and child's age as factors. The results of the ANOVAs for all children can be found in Appendix C. All children produced stressed and unstressed vowels with significant different values for F1 and F3. With increasing age also F2 was controlled and stress showed a significant main effect. Stressed and unstressed vowels differ in articulatory organisation (Dogil, 1995) as they show different vowel dispersions in the perceptual space. To conclude, different representations for stressed and unstressed vowels need to be developed by the children. In Chapter 6 a further analysis of the development in vowel space is described. As secondarily stressed vowels were neither produced with significant differences to unstressed or stressed vowels no further analysis of the development is necessary.

Hanna

Univariate ANOVAs for the first four formants showed significant main effect of stress and age. The interaction between stress*age was significant. Hanna tried to produce stressed vowels with significantly higher formant values than unstressed vowels. She produced vowels with secondary stress mainly like unstressed vowels with no significant differences.

Nils

Univariate ANOVAs for the first and third formants showed significant main effect of stress and child's age. The interactions between stress and age were significant. With increasing age Nils managed to produced stressed vowels with different formant values than unstressed vowels. Vowels with secondary stress showed no clear tendency to be produced in a different way than stressed vowels.

Bennie

Stress and age as main factors as well as the interaction between stress and age were significant for all formants except stress for F4. A closer look at the production of vowels in dependency of stress showed that Bennie always produced unstressed vowels with lower formant values than stressed ones. F1 appeared to be the first formant to be produced with consistency, relative to stress. Bennie displayed a significant difference in F1 for the whole observation period, for stressed and unstressed vowels. With increasing age the values of F3 also seemed to be more and more dependent on stress. At 24 months of age it appears that F2 was now also produced in a consistent way.

No significant different formant values for vowels with secondary stress were observed. They were produced similar to unstressed vowels.

Rike

Rike showed a significant main effect of stress (except F4) and age as well as the interaction between stress and age. Apart from the age of 22 months, Rike's values of F1 dependent to stress. Different vowel values were also mainly produced for F3. In contrast to the other children, Rike tend to produced stressed vowels with lower formant values than unstressed ones. No tendencies were observed for vowels with secondary stress.

Emma

For Emma, the univariate ANOVAs with the formants as variables showed significant main effect of stress only for F1 and F3. The factor child's age was significant for all formants. The interaction between stress and age was significant for all formants. Depending on age, Emma produced unstressed vowels with significantly lower F1 values than stressed vowels. First, she produced vowels with secondary stress similar to unstressed vowels. With increasing age she produced vowels with secondary stress with formant values similar to stressed vowels. Her used for stressed and unstressed vowels different F2 and F3 values but not significant and not consistent.

Ollie

For Ollie, the ANOVAs showed a significant main effect of stress on F1 and F3. Age as main effect was significant for all formants as well as the interaction between stress and age. Until the age of 15 months, Ollie produced unstressed vowels with higher F1 values than stressed ones. At 18 and 33 months of age the picture changed as stressed vowels were now produced with higher F1 values than unstressed ones. Ollie produced Vowels with secondary stress always with the highest F1 values. At the age of 15 months, he produced stressed vowels with significant lower F3 values than unstressed ones. Apart from the age of 18 months, he produced stressed vowels with lower F2 values than unstressed ones.

Parents' use of formants

Univariate ANOVAs were carried out on the first four formants as dependent variables and stress and child's age as factors. In Appendix C the results of the ANOVAs for all parents are given. For almost all parents stress showed a significant main effect on the formant values of F1, F2 and F3. Stressed and unstressed vowels were produced with significantly different formant values by all parents. Further analysis showed that unstressed vowels were produced with less variability and more centralised formant values than stressed ones. Vowels with secondary stress showed no separate formant values. Some parents produced vowels with secondary stress similar to unstressed vowel space, some similar to stressed ones. No own vowel space were observed for vowels with secondary stress. In Chapter 6 a detailed analysis of the different vowel spaces for stressed and unstressed vowels in adult speech is given.

5.4.5 Voice quality

The Open Quotient (*OQ*) is considered to be a correlate of sentence accent and is not expected to be significant for word stress (Sluijter, 1995; Lintfert and Wokurek, 2005). Glottal Opening (*GO*) is regarded to be a correlate of word stress as stressed vowels are produce with lower (*GO*) values than unstressed ones in adult speech (Jessen et al., 1995). Equally Skewness (*SK*) and Rate of Closure (*RC*) depend on word stress for adult speech (Claßen et al., 1998) as well as for child–directed speech (Schneider and Möbius, 2007). Unstressed vowels are expected to be produced not so abruptly and with a slower closing of the vocal folds. Therefore, unstressed vowels should have higher spectral tilt values than stressed vowels. The same tendency is expected for (*T4*). On the other side should stressed vowels show greater glottal leakage and therefore higher Completeness of Closure (*CC*) and Incompleteness of Closure (*IC*) values (Schneider, 2007).

Children's use of voice quality parameters

Univariate MANOVAs were carried out for the raw and normalised data on the voice quality parameters as dependent variable and stress and child's age as factors. Further analyses were done with post hoc Tamhane. In Appendix C the detail results of the univariate MANOVAs for raw and normalised data is given. In Table 5.5 an overview of the use of the observed voice quality parameter to mark stress is given for each child.

Although the parameter *OQ* is considered to be a correlate of sentence accent and is not expected to be significant for word stress, a significant main effect of stress on this parameter were found for three children. These children produced stressed vowels with lower *OQ* values than unstressed vowels. The parameter was not used to distinguish vowels with secondary stress. As most analysed utterances of the children were one–word sentences an influence of intonation cannot be turned out. For all children, an increase of *OQ* values were observed with increasing age. *OQ* is considered to be a correlate of sentence accent and not for word stress. Therefore, an increase in *OQ* values could be a sign for a more adult–like intonation contour. Higher *OQ* values correlate with additional muscular tension in the larynx, the control of this muscle advances with increasing age of the children.

The parameter *GO* is considered to be a correlate of word stress. In adult speech, stressed vowels were produced with lower *GO* values than unstressed ones (Jessen et al., 1995). For the children, the data showed a significant main effect of stress as well as child's age on *GO*. If the children used this parameter to mark word stress then it was mainly to emphasise as vowels with secondary stress are significant. Stressed and unstressed vowels were mainly produced with same *GO* values.

The spectral tilt parameters *SK*, *RC* and *T4* demonstrated a significant main effect of stress only for Nils, Rike and Ollie. The interaction between stress and age was significant for all children. The use of these parameters depends on age. With increasing age a more consistent use of the parameters were observed. The main tendency for all children

Table 5.5: Overview of the use of voice quality parameters to mark stress. Significance is marked with (*), no significance with (–). If a clear tendency were observed a main use of the parameters to differ stress is given (Tamhane post hoc). It is described wether higher (>) or lower (<) values were observed.

vq	source	Hanna	Nils	Bennie	Rike	Emma	Ollie
OQ	stress	–	–	–	*	–	*
	age	–	–	–	*	–	*
	stress*age	–	*	–	*	*	–
	Tamhane		0>1	0>1,2	0,2>1		
GO	stress	*	*	*	*	*	–
	age	*	*	*	*	*	*
	stress*age	*	*	*	*	*	*
	Tamhane	(2>0,1)	1<0	2=/1,0	2<0,1	2>0,1	2>0,1
SK	stress	–	*	–	*	–	*
	age	–	–	–	–	–	–
	stress*age	–	*	*	*	*	*
	Tamhane	0>1,2	0,2>1	0>1,2	0>=2>=1		0>1
RC	stress	–	*	–	*	–	*
	age	–	–	–	–	–	–
	stress*age	–	*	*	*	*	*
	Tamhane	(0>1,2)	0,2>1	0>1,2	0>=2>=1		0>1
T4	stress	–	*	–	*	–	*
	age	–	–	–	–	–	–
	stress*age	*	*	*	*	*	–
	Tamhane	(0>1,2)	0,2>1	0>1,2	0>=2>=1		0>1
CC	stress	*	–	–	–	–	–
	age	–	–	–	–	–	–
	stress*age	*	–	*	–	–	*
	Tamhane	2>1,0	2<0	2>0,1	2<0,1		
IC	stress	*	–	*	–	–	–
	age	–	–	–	–	–	–
	stress*age	*	–	*	*	–	*
	Tamhane	2>1,0		2>0,1	2<0,1		

was that stressed vowels were produced with lower values of the spectral tilt parameters than unstressed ones. All children tended to produce unstressed syllables with stronger spectral tilt. For vowel with secondary stress no tendency could be observed as the use of the parameter for this purpose differed between the children.

For the glottal leakage parameters *CC* and *IC* stress showed some significant effect in interaction with age. The use of this parameter to mark stress varied between children and age. The children used these parameters not at each age to mark word stress. When these parameters were significant for word stress then secondarily stressed vowels were produced with significantly different values than stressed and unstressed vowels.

To conclude, children between 15 and 36 months of age were able to use different voice quality parameters to produce word stress. As age showed significant main effect, it can be concluded that the use of these parameters were variable and not stable until the age of 36 months. The children's use of the voice quality parameters varied and depends on the child. The parameters describing the vibration behaviour of the vocal folds were used by all children more or less consistent to produce word stress. They managed for stressed vowels a more abrupt and faster closure than for unstressed vowels with the

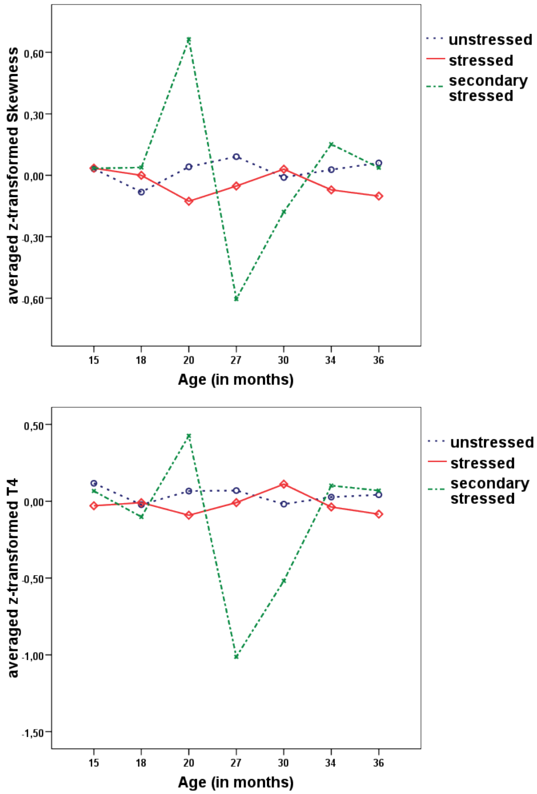


Figure 5.25: Z–transformed values of *SK* and *T4* for **Hanna** depending on her age.

beginning of word production. The parameters *GO*, *IC* and *CC* were mainly used to distinguish vowels with secondary stress. Stressed and unstressed vowels showed no significant differences for these parameters.

Hanna

For the normalised data of *OQ* no significant effect were found. Looking at the raw data, a developmental pattern were observed: Hanna produced lower OQ values with increasing age. For *GO* a significant main effect of stress and child' s age and for the interaction between stress and child's age were found. The parameter *GO* is mainly used

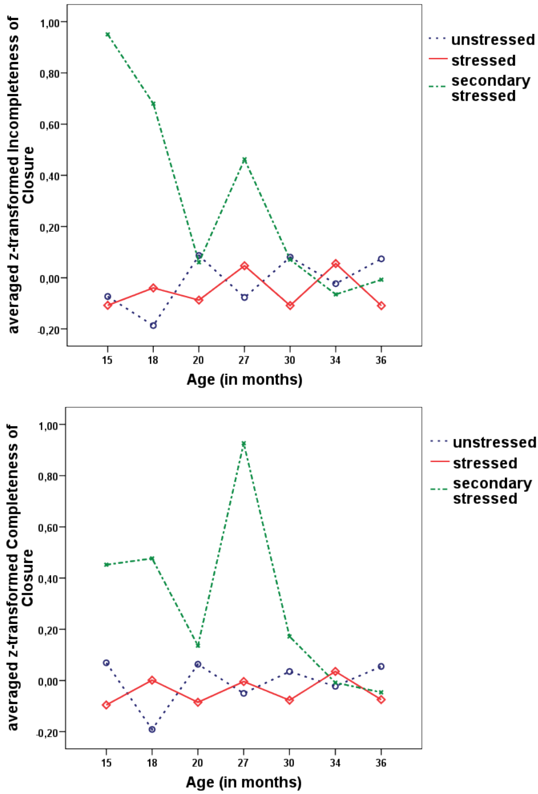


Figure 5.26: Z–transformed values of *IC* and *CC* of **Hanna** depending on her age.

to produce vowels with secondary stress with significantly higher values than stressed and unstressed vowels. No significant differences were produced between stressed and unstressed vowels. Hanna produced lower *GO* values with increasing age. For the parameters *SK*, *RC* and *T4* the normalised data showed no significant main effect of stress and age. The interaction between stress and age was significant for *T4* and showed tendencies for *SK*. Analyses of the raw data revealed a significant interaction between stress and age for *RC*. Figure 5.25 illustrates the averaged normalised values for *Skewness* and *T4*. When Hanna used the spectral tilt parameters for marking word stress then she produced unstressed vowels with the highest *Skewness*, *Rate of Closure* and *T4* values. She produced vowels with secondary stress and stressed vowels more abrupt and faster closing of the vocal folds than unstressed vowels. For *CC* and *IC* the univariate MANOVAs showed a significant main effect of stress but not of age. The interaction between stress and child's age was significant for *CC* and *IC*. Hanna produced stressed and unstressed vowels mainly with the same IC and CC. She used the glottal leakage parameters only for distinguish vowels with secondary stress. These vowels were produced with significantly higher values of *CC* and *IC* than unstressed and stressed vowels (see Figure 5.26). Without including vowels with secondary stress in the analysis, only the interaction between stress and age for *IC* was still significant ($F(6,2461)=2.293$; $p<0.05$). These findings supports the claim that Hanna used *IC* and *CC* to differentiate vowels with secondary stress from stressed and unstressed syllables. Vowels with secondary stress behave like stressed vowels in the case of all other voice quality parameters. Hanna produced secondarily stressed vowels with the greatest glottal leakage to distinguish these word stress pattern from stressed and unstressed vowels.

Nils

The normalised data of *OQ* revealed no significant main effect of stress or age; however, as the interaction between stress and age was significant further analyses for each age group were done. Until the age of 30 months, Nils used the parameter *OQ* to distinguish vowels with secondary stress. Stressed and unstressed vowels showed nearly the same *OQ* values. As from 30 months of age, Nils used *OQ* to distinguish between stress and unstressed and no longer for vowels with secondary stress. Stressed vowels were now produced with significantly lower *OQ* values than unstressed vowels. For vowels with secondary stress no production tendency were observed. At some ages the produced values were similar to stressed vowels and at others more like unstressed vowels. The normalised data of *GO* showed significant main effects for stress and child's age. The interaction between stress and age was also significant. Nils use of *GO* dependent on age. When he used *GO* to mark word stress then stressed vowels were produced with lower *GO* values than unstressed ones. For the spectral tilt parameters it SK, *RC* and *T4* a significant main effect of stress but not of child's age were found. The interaction between stress and age was significant. Nils produced stressed vowels with significantly lower values of *SK*, *RC* and *T4* than vowels with secondary stress and unstressed vowels. For the production of stressed vowels, Nils closed his vocal fold faster and more abruptly than for the production of unstressed or vowels with secondary stress (see Figure 5.27). For *CC* and *IC* only tendencies were observed. Both parameters showed no main effect of stress or age for the normalised data. The interaction between stress and age was significant for *CC*. Nils mainly produced vowels with secondary stress with significantly lower values of *CC* than unstressed vowels.

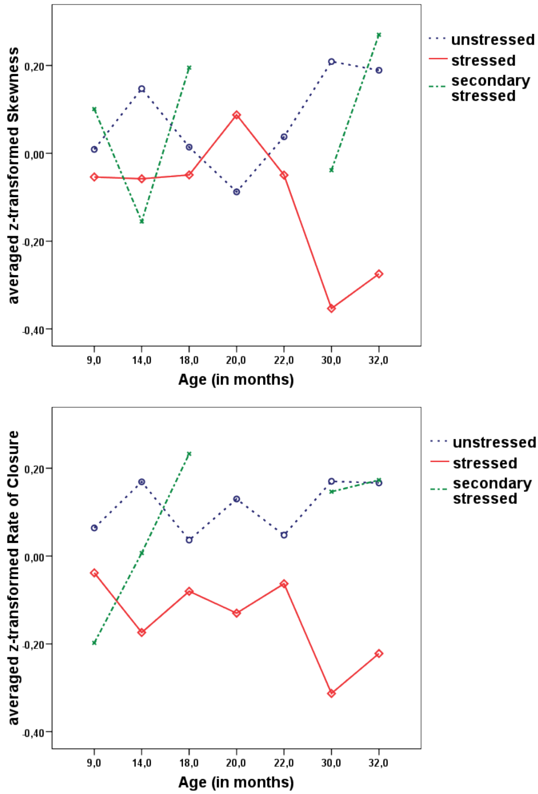


Figure 5.27: Z–transformed *SK* and *RC* of **Nils** depending on his age.

Bennie

For the normalised data of *OQ* no significant main effect of stress or age were observed. For normalised *GO* a significant main effect of stress and age was determined. Bennie used the parameter *GO* mainly to distinguish vowels with secondary stress from stressed and unstressed vowels. The use of voice quality parameters describing the spectral tilt for marking stress depended on age. Stress showed no significant main effect but was significant in interaction with age for all three spectral tilt parameters. Bennie produced unstressed vowels with significantly higher *SK*, *RC* and *T4* values than stressed vowels and vowels with secondary stress. The latter two prominences were produced in the same

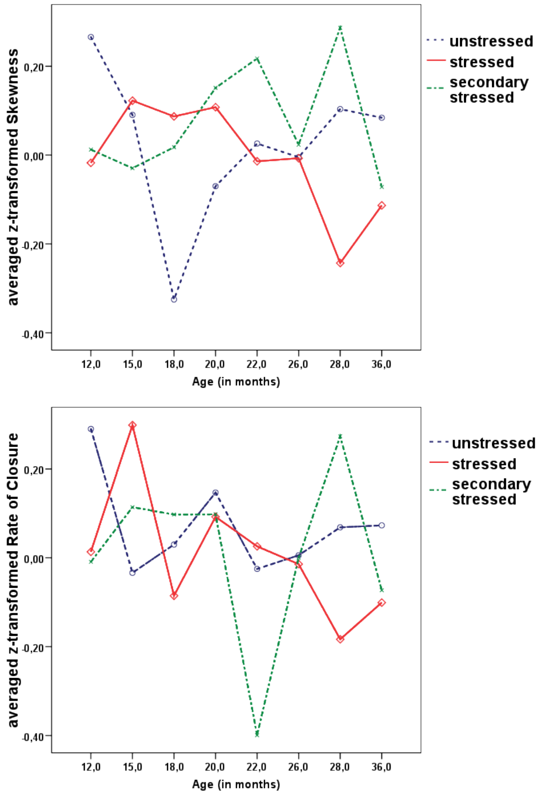


Figure 5.28: Z–transformed *SK*, and *RC* of **Bennie** depending on his age.

manner, with more abrupt and faster closing of the vocal folds than unstressed vowels. Figure 5.28 illustrates the averaged normalised values of the spectral tilt parameters according to age. An age–dependent use of the parameters were observed. A significant main effect of stress on the parameter *IC* but not on *CC* were found. No main effect of age were show for the normalised data. The interaction between stress and age stress was significant for both parameters. Bennie produced vowels with secondary stress with the highest *CC* and *IC* values. The glottal leakage parameters *CC* and *IC* were used to differentiate vowels with secondary stress from stressed and unstressed vowels, as they were produced with the greatest glottal leakage.

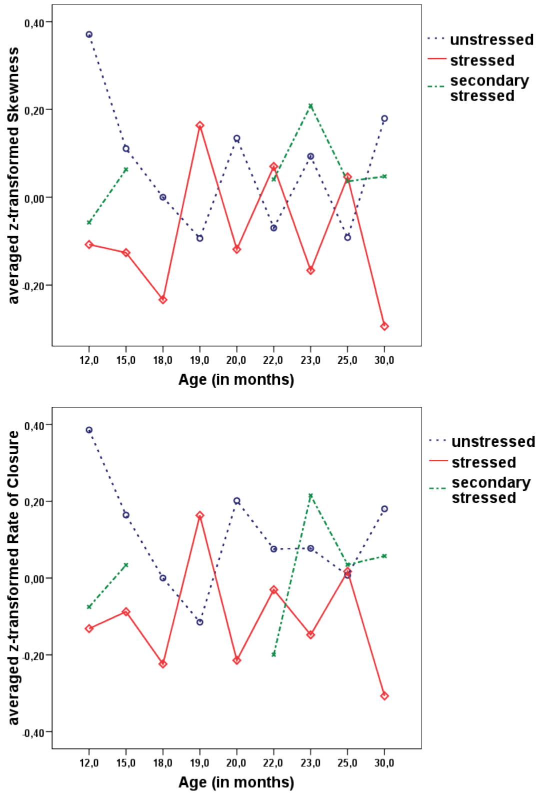


Figure 5.29: Z–transformed *SK* and *RC* of **Rike** depending on her age.

Rike

The normalised data of *OQ* showed a significant main effect of stress and age. The interaction between stress and age was significant. Therefore, a developmental pattern were assumed. Rike tended to produce unstressed vowels with the highest *OQ* values. For *GO* the normalised data showed a significant main effect for stress and age. The interaction between stress and age was also significant. When the parameter *GO* is used for marking word stress then Rike produced vowels with secondary stress with the lowest *GO*. All observed spectral tilt parameters showed a significant main effect of stress. No main effect of age was show on the normalised data; however, the interaction between stress and age was significant for all parameters. Rike produced stressed vowels mainly with significantly lower *SK*, *RC* and *T4* values than unstressed vowels. Vowels bearing secondary stress were produced with spectral tilt values between stressed and unstressed vowels. But the use of these parameters was not consistent (see Figure 5.29). When Rike used spectral tilt parameters to mark stress then she produced stressed vowels more abruptly and with faster closing vocal folds than unstressed and secondarily stressed vowels. For the parameter *CC* no significant main effect of stress and age were observed for the normalised data. For the parameter *IC* the interaction between stress and age was significant for normalised data. Rike used the glottal leakage parameters *CC* and *IC* for distinguish vowels with secondary stress, as they were produced with the lowest values.

Emma

For the normalised data of *OQ* no significant main effect of stress and age were observed. The interaction between stress and age was significant. A slight increase in *OQ* with increasing age were observed. *GO* showed a significant main effect of stress and age. The interaction between stress and age was significant. Emma used, with increasing age, *GO* to mark vowels with secondary stress. She produced these vowels with significantly higher *GO* values than stressed and unstressed vowels. The normalised values of the spectral tilt parameters showed no main effect. The interaction between stress and age was significant for all three parameters. At 15 months of age, Emma produced unstressed vowels with the lowest values of *RC*, *SK* and *T4*. With 18 months, the production pattern changed. From this age on, Emma produced the spectral tilt parameters in the expected way. She produced unstressed vowels with higher *SK*, *RC* and *T4* values than stressed ones. Between 20 and 26 months of age, she did not use the spectral tilt parameters in a significant way to mark word stress (see Figure 5.30) For the normalised *CC* and *IC* no significant main effect was found.

Ollie

The normalised data of the parameter *OQ* showed a significant main effect of stress and age. The interaction between stress and age was not significant. Ollie produced stressed vowels with lower *OQ* values than unstressed and secondarily stressed vowels. With increasing age an increase of the *OQ* values were observed. For the normalised data of *GO* a significant main effect of age was found. No main effect of stress was found. The interaction between stress and age was significant. This provides evidence for an age–dependent use of this parameter to mark stress. Ollie produced vowels with secondary stress with significantly higher *GO* values than unstressed and stressed vowels. All normalised spectral tilt parameters demonstrated a significant main effect of stress but not of age. For the parameters *SK* and *RC* the interaction between stress and age was also significant. Ollie produced stressed vowels with significantly lower spectral tilt parameters than unstressed vowels. Vowels with secondary stress were mainly produced similar to unstressed vowels. Figure 5.31 illustrates the developmental pattern of the spectral tilt parameters for Ollie depending on age. Stressed vowels were produced with

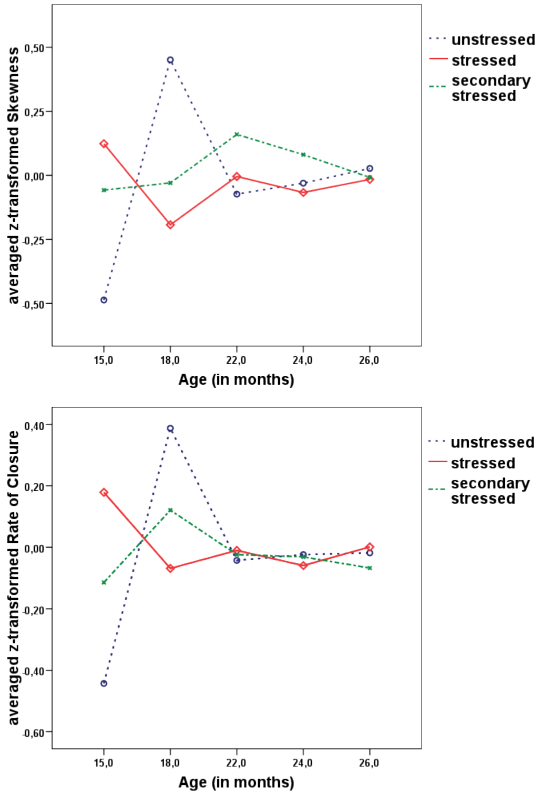


Figure 5.30: Z–transformed *SK* and *RC* of **Emma** depending on her age.

the lowest spectral tilt parameters and, therefore, with faster and more abrupt closure of the vocal folds. For the parameters describing the glottal leakage, *CC* and *IC*, no main effect of stress or age were found. The interaction between stress and age was significant. At the age of 15 months, Ollie produced stressed vowels with the lowest *CC* and *IC*. In contrast, at 18 months of age stressed vowels were produced with the highest *IC* and *CC*. No differences in the production of stress for the parameters *CC* and *IC* could be observed at 33 months of age. Vowel with secondary stress were produced similarly to unstressed vowels.

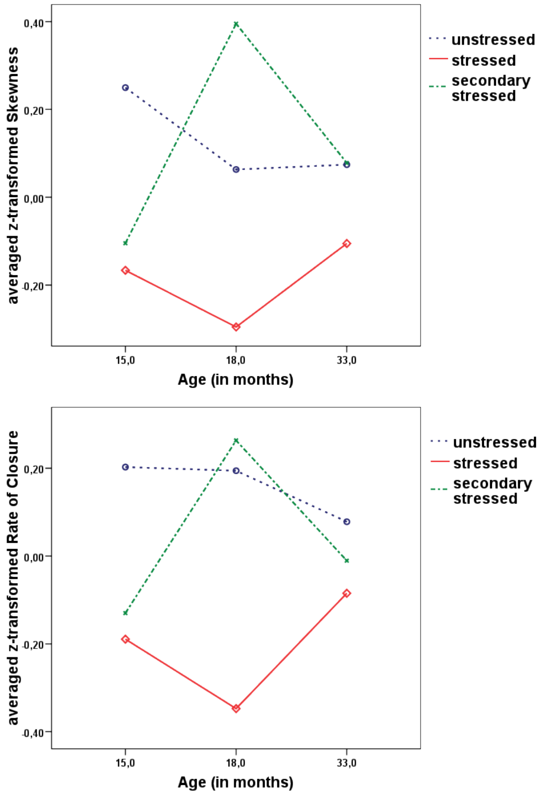


Figure 5.31: Z–transformed *SK* and *RC* of **Ollie** depending on his age.

Parents' use of voice quality parameters

Univariate MANOVAs were carried out for the voice quality parameters as dependent variable and stress and child's age as factors. The detailed results can be found in Appendix C. For some adults the parameters *Open Quotient* and *Glottal Opening* showed a significant main effect of stress. The parents used these parameters to emphasise (secondary stress) and not to distinguish prominence. When the parents used spectral tilt parameters to mark word stress than they produced unstressed vowels mainly with higher spectral tilt values, resulting in a steeper spectral slope (see Figure 5.32).

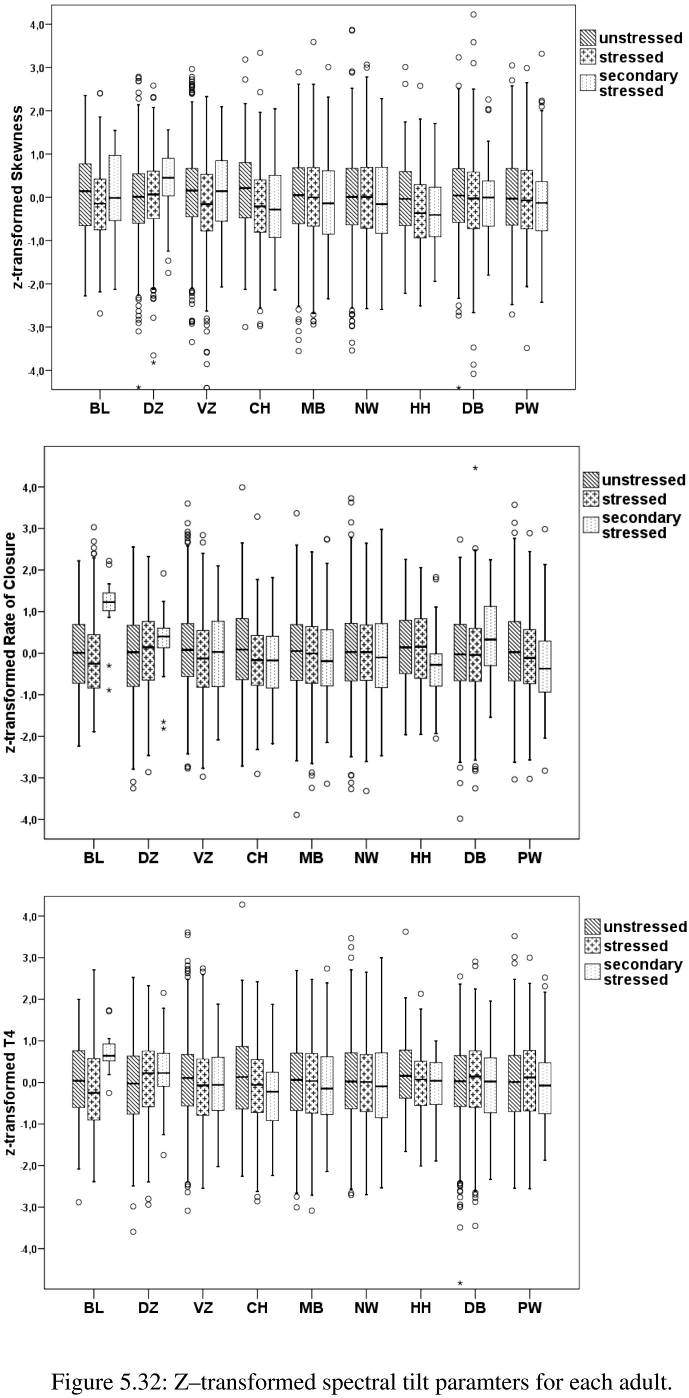


Figure 5.32: Z–transformed spectral tilt paramters for each adult.

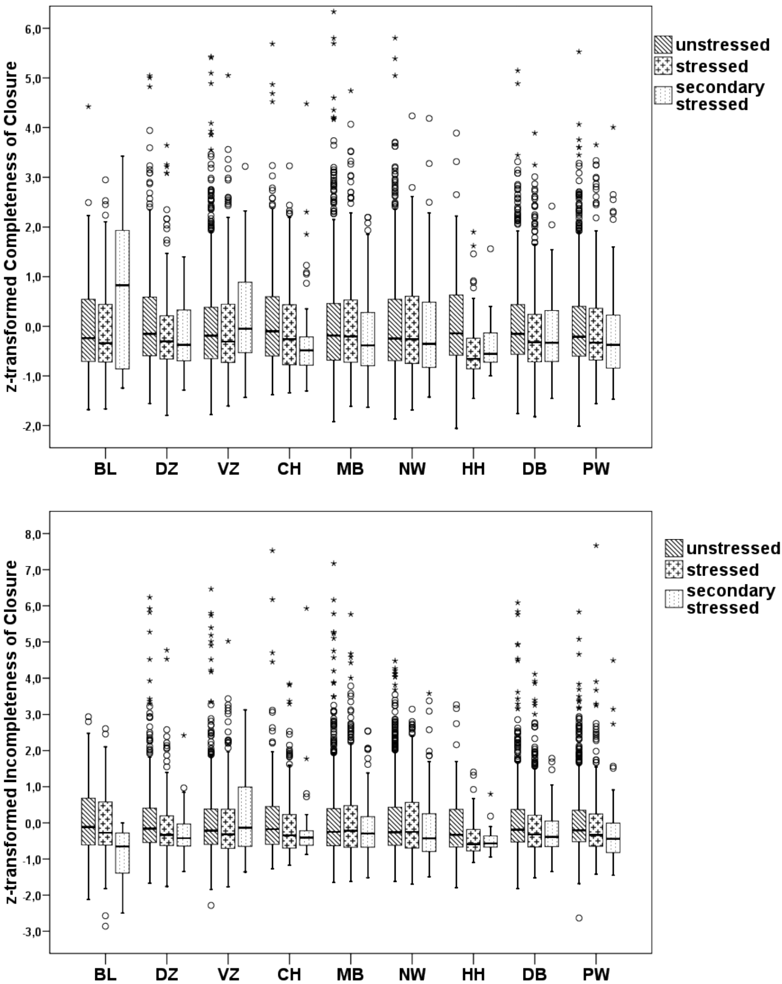


Figure 5.33: Z–transformed glottal leakage parameters for each adult.

Vowels with secondary stress showed no clear production scheme, their realisations varied between the parents. Some produced them mainly with the same spectral tilt parameters as stressed vowels, some the other way round. Only Ollie's mother produced vowels with secondary stress with significantly higher spectral tilt parameters than unstressed and stressed ones. As the effect of stress depends on the increase and decrease of the subglottal pressure it can be concluded that for the production of unstressed vowels a more regulated decrease of the subglottal pressure is necessary and have to be learnt by the children. The parameters describing the glottal leakage were mainly used to differentiate vowels with secondary stress from stressed and unstressed ones (see Figure 5.33). If *CC* and *IC* were significant for stress it is mainly because of secondary stress.

Hanna's parents (CH, HH)

For Hanna's mother, the univariate ANOVA showed no significant main effect of stress but a significant main effect of the factor child's age. The interaction between stress and child's age was significant. A significant main factor of stress and age was found for the parameter *GO*. The interaction between stress and age was also significant. Hanna's mother produced vowels with secondary stress with lower *GO* values than stressed and unstressed ones. The spectral tilt parameters *SK* and *RC* showed a significant main effect of stress but not on age. For the parameter *T4* only the interaction between stress and age was significant. Hanna's mother produced unstressed vowels with lower spectral tilt values than stressed ones. Secondary stress was mainly produced similar to stressed vowels. The parameters describing the glottal leakage were both significant for stress, age and their interaction. She produced unstressed vowels with significantly higher values than stressed vowels and secondary stress.

For Hanna's father, only one recording session with Hanna at 36 months of age were analysed; therefore, the factor child's age was not analysed. A significant main effect of stress on *OQ* were found. Hanna's father produced secondary stress with the lowest *OQ* values. *GO* showed no significant main effect of stress, nor did the spectral tilt parameters *SK*, *RC* and *T4*. Hanna's father did not use these parameters to mark word stress. He used *CC* and *IC* as these showed a significant main effect of stress. He produced unstressed vowels with significantly higher glottal leakage values than stressed vowels. Secondary stress were produced similar to stressed vowels.

Nils' parents (MB, DB)

For Nils's parents, child–directed as well as adult–directed speech were analysed.

For Nils's mother, a significant main effect of stress on *OQ* was found. She used this parameter to distinguish secondary stress. Vowels with secondary stress showed significantly lower *OQ* values than stressed and unstressed vowels. A significant main effect of stress on the parameter *GO* was found but no main effect of age. The interaction between stress and age was significant. Nils' mother distinguish secondary stress by producing significantly higher *GO* values than for unstressed and stressed vowels but not in all cases. No main effect was found for the spectral tilt parameters. The interaction between stress and age was significant for *SK*, *RC* and *T4*. Different production patterns depending on the child's age were observed. Nils's mother used these parameters to distinguish unstressed vowels. A significant difference mainly occurred in child–directed speech and not in adult–directed speech. In child–directed speech, in some cases, secondary stress were characterised by lower spectral tilt values to emphasise stress. Stress had a significant main effect on the parameter *CC*. The interaction between stress and age was significant for both parameters. Secondary stress were produced with higher *CC* as well as *IC* than stressed and unstressed vowels. Nils's mother mainly used the parameters describing the glottal leakage to emphasise secondary stress.

For Nils' father, *OQ* showed a significant main effect of stress. In the same way as the mother, he produced secondary stress using lower *OQ* values than for unstressed and stressed vowels. The parameter *GO* showed no main effect of stress but on age. Also the interaction between stress and age was significant. A different use of *GO* in adult and child–directed speech were observed. In adult–directed speech, Nils's father produced secondary stress by lower *GO* values than for unstressed and stressed vowels. In child–directed speech, no significant use of this parameter to mark stress were observed. No main effect was found for the spectral tilt parameters. The interaction between stress and age was significant for the parameter *RC*. Nils's father used *RC* in child–directed speech to mark secondary stress. In this case he produced higher *RC* values than for stressed and unstressed vowels. Neither *CC* nor *IC* demonstrated significant main effect of stress or age. However, when he used these parameters to mark stress then unstressed vowels were produced with the highest glottal leakage.

Bennie's parents (NW, PW)

For Bennie's parents, child–directed as well as adult–directed speech were analysed.

For Bennie's mother, a significant main effect of stress and age was found for *OQ*. She used this parameter to distinguish secondary stress. For the parameter *GO* no main effect was found. The interaction between stress and age was significant. Bennie's mother used this parameter similar to *OQ*, to distinguish secondary stress. She produced vowels with secondary stress with higher *GO* values than unstressed and stressed vowels. But the use of this parameter to distinguish secondary stress depended on the child's age. No significant main effect of stress or age for the spectral tilt parameters were found. Bennie's mother used the spectral tilt parameters only during the child's babbling phase to mark stress. During this stage, she produced unstressed vowels with the highest values of *SK*, *RC*, and *T4*. The parameters describing the glottal leakage showed no significant main effect of stress but of age. She used the glottal leakage parameters dependent on the speech tasks (adult– and child–directed speech). In child–directed speech, she used *CC* and *IC* for marking stress. But only with the beginning of the picture–naming task between the ages of 22 and 26 months. Between this age range, she produced unstressed vowels with higher *CC* and *IC* values. By using these parameters she she kept the child's attention on the different stress patterns. At 36 months of age, she did no longer used these parameters in the described matter. In adult–directed speech no use of these parameters was observed.

For the father, no main effect on *OQ* was found. A significant main effect of stress on the parameter *GO* was given. The interaction between stress and age was also significant for the parameter *GO*. Bennie's father used *GO* to distinguish secondary stress by producing highest *GO* values. For the spectral tilt parameters, only *RC* showed a main effect of stress. Age had no main effect. He used the spectral tilt parameters for marking stress as unstressed vowels tended to be produced with higher values of *SK*, *RC* and *T4*. Bennie's father used *CC* and *IC* when he introduced the picture–naming task for marking stress. With increasing age of the child, he no longer used these parameters.

Rike's mother (VZ)

For Rike's mother, no significant main effect of stress on *OQ*, but of age was found. The interaction between stress and age was significant. At certain times, she distinguished secondary stress with the highest *OQ* values. For *GO* no main effect of stress was found. Age was significant and this was shown by age–dependent values of *GO*. For the spectral tilt parameters *SK* and *RC* a significant main effect of stress was shown. No further effects were found. Rike's mother produced stressed vowels with lower *RC*, *SK* and *T4* values than unstressed vowels. Secondary stress showed a similar pattern as unstressed vowels. For the production of stressed vowels a more abrupt and faster closure of the vocal folds was necessary than for the production of unstressed vowels. For the glottal leakage no main effects were found for the normalised data. The raw data demonstrated a significant main effect of stress for both parameters. Rike's mother used *CC* and *IC* to distinguish secondary stress by using highest *CC* and *IC* values.

Emma's mother (BL)

For Emma's mother, no main effect on *OQ* was found. For the parameter *GO* a significant main effect of stress was given. She used *GO* mainly in interaction with Emma and not in interaction with the school children. When speaking to the younger child, she used the parameter *GO* for distinguish secondary stress and to kept the child's attention on the correct vowel quality. Only the spectral tilt parameter *RC* showed a significant main effect of stress. For the parameter *T4* the interaction between stress and age was significant. Emma's mother mainly produced stressed vowels with the lowest *SK*, *RC* and *T4* values. She produced stressed vowels with more abrupt and faster closing of the vocal folds than unstressed vowels, but not significantly. She distinguished secondary stress with the highest *SK*, *RC* and *T4* values from stressed and unstressed vowels. The spectral tilt parameters were used to differentiate the different possible stress patterns. The parameters describing the glottal leakage, *CC* showed a significant main effect of stress. For the parameter *IC* the interaction between stress and age was significant. Emma's mother used the glottal leakage parameters to distinguish secondary stress, characterised by the highest *CC* and the lowest *IC* values.

Ollie's mother (DZ)

For Ollie's mother, *OQ* showed no main effect. The parameter *GO* showed a significant main effect of stress. She used *GO* to distinguish secondary stress, characterised by higher *GO* values than unstressed and stressed vowels. All spectral tilt parameters showed a significant main effect of stress. Ollie's mother produced stressed vowels with the lowest spectral tilt values, but not significantly. However, she distinguish secondary stress with the highest *SK*, *RC*, and *T4* values. The glottal leakage parameters *CC* and *IC* showed a significant main effect of stress. Ollie's mother produced unstressed vowels with significantly higher *CC* and *IC* values than secondarily stressed and stressed vowels.

5.5 Summary

Appendix C gives an overview of the use of the different observed parameters, grouped for each child and the corresponding adult speakers. For the adult data, vowel duration was the most reliable cue to mark stress. Parents always produced stressed vowels an increase in duration as compared to unstressed vowels. Vowel duration was not used to distinguish secondary stress. No production category for secondary stress based on duration were established. Vowels with secondary stress were either produced similar to stressed vowels or with duration values between unstressed and stressed vowels.

Intensity was not used by all parents to mark stress. When intensity was used, it was mainly used to emphasise stress to keep the child's attention on wrong outputs.

F_0 was used by all parents, besides Hanna's father. For all other adults, unstressed vowels were produced with lower F_0 values than stressed ones. No explicit production pattern was found for secondary stress. Three parents used F_0 to differentiate the stress patterns. Four adults produced secondarily stressed vowels comparable to unstressed ones, one mother produced secondarily stressed vowels similar to stressed ones.

OQ was used by one father to distinguish between all observed stress patterns. Three other adults used this parameter to mark secondary stress as this stress pattern was produced with significantly lower *OQ* values. *GO* was mainly used to distinguish secondary stress. Stressed vowels are mainly produced with higher spectral tilt parameters *SK*, *RC*, and *T4* than unstressed vowels. No clear production pattern can be observed for secondarily stressed vowels. Only Hanna's father and Ollie's mother produced unstressed vowels with higher *CC* and *IC* values than stressed and secondarily stressed vowels. The other adults preferred these parameters to mark secondarily stressed vowels as they are produced mainly with higher values than unstressed and stressed vowels. For the production of the different stress patterns all acoustic parameters are used to differentiate with the preference on duration, intensity and F_0 to distinguish between stressed and unstressed vowels. Secondary stress mainly differ in the fine–tuning of the vocal folds described with the different voice quality parameters. Based on these results it can be assumed that for secondary stress no own category has to be build up. The status of secondary stress should be verified in further research.

5.5.1 Hanna and parents

Appendix C.16 gives an overview of Hanna's and her parents' use of the acoustic parameters to mark stress. With the beginning of babbling, Hanna was able to produce vowels and syllables with different perceived prominence. She realised different prominence patterns via varying intensity. With increasing age, vowel duration was also used more frequently. Unstressed vowels were produced with shorter vowel duration than stressed vowels. Stressed vowels were mainly produced with higher F_0 values than unstressed vowels. The voice quality parameters *GO*, *CC* and *IC* were mainly used to distinguish secondary stress. The spectral tilt parameters were used in the expected way: unstressed vowels were produced with the highest spectral slope. However, the use of the different parameters to mark word stress was variable until the age of 36 months. Hanna did not have established an adult–like use of the acoustic parameters for marking stress until 36 months. Her use was still inconsistent. Both parents used duration and intensity to mark stress, and with increasing age, Hanna also used this parameter accordingly. Her parents used the voice quality parameters in different ways. Hanna's father used *OQ* and the parameters describing the glottal leakage for the production of stressed and unstressed vowels and no spectral tilt parameters.

In contrast, Hanna's mother preferred fundamental frequency and the spectral tilt parameters to differentiate between stressed and unstressed vowels, and *CC* and *IC* only to emphasise stress.

Therefore, Hanna was exposed to different acoustic realisations of stress. Hanna used all possible acoustic parameters to mark stress. She did not prefer specific acoustic parameters comparable to her parents. She used fundamental frequency and the spectral tilt parameters comparable to her mother. Like her mother, Hanna also used the glottal leakage parameters to distinguish secondary stress. But unlike her mother, secondary stress was characterised by the highest glottal leakage values and not with the lowest values. Hanna's use of the parameters *OQ* and *GO* were the opposite of the use of these parameters of her parents.

5.5.2 Nils and parents

Appendix C.17 presents an overview of Nils's and his parents' use of the acoustic parameters to mark stress.

Nils's preferred parameter to mark stress varies with age and, therefore, on the developmental phase. He did not use vowel duration and intensity in the developmental phase from 14 to 22 months where both babbling and words occurred. When vowel duration was used to mark word stress than in the expected was as unstressed vowels showed shorter vowel duration and lower intensity than stressed vowels. Independent of Nils's age, his parents used vowel duration to mark stress. With increasing age of Nils also intensity was used to emphasise the correct stress production. With increasing age, Nils showed a more adult–like performance as far as control over the parameters duration and intensity is concerned. Nils always used F_0 to mark stress. In his productions, unstressed vowels were always produced with the lowest F_0. Both parents produced unstressed vowels with lower fundamental frequency values than stressed ones. Nils also used F_0 to distinguish secondary stress. Vowels with secondary stress were produced with the highest F_0, in contrast to his mother. Nils' parents used the voice quality parameter mainly to emphasise stress. Significantly different voice quality parameters were used to distinguish secondary stress from unstressed and stressed vowels. Nils's father also tended to produce unstressed vowels with higher *IC* and *CC* values compared to stressed ones. He produced unstressed vowels with greater glottal leakage than stressed ones. Nils was variable in the use of the parameters and differs in their use from his parents.

5.5.3 Bennie and parents

Appendix C.18 gives an overview of Bennie's and his parents' use of the acoustic parameters to mark stress. With increasing age, Bennie used vowel duration more and more in an adult–like manner as both parents used vowel duration to mark stress. Bennie most reliably marks stress via intensity as he always produced unstressed syllables with less intensity. The parents mainly use intensity as a further parameter to keep the child's attention on different stress patterns. Bennie and his parents used F_0 in a comparable manner for marking stress: unstressed vowels were produced with lower F_0 than stressed vowels. Comparable to his father, Bennie distinguish secondary stress with significantly higher F_0 values. Analogous to his mother, Bennie tended to use the parameter *GO* to distinguish secondary stress. As his mother preferred to use the spectral tilt parameter only during babbling but his father also during the picture naming task, Bennie also tried to use these parameters in an adult–like manner, i.e. with higher values of *SK*, *RC* and *T4* for unstressed vowels. Bennie used the glottal leakage parameters *CC* and *IC* to distinguish secondary stress. Neither did the mother nor the father displayed the same tendency. Bennie's mother used the voice quality parameters depending on Bennie's age. In contrast, his father used the observed voice quality parameters independently of the child's age. Bennie used all possible acoustic parameters to mark stress, but his parents did not use all parameters. For Bennie, no preference in the use of special parameters, similar to his parents, were observed.

5.5.4 Other children

For the other three children, the analysis was restricted to recordings of the mothers. This was due to the fact that these children spent most of the time with their mothers and were thus influenced to a greater degree by their mothers than by their fathers. Unfortunately, we do not have evidence to prove this. Appendix C.19 gives an overview of Rike's and her mother's use of the parameters to mark stress, Appendix C.20 for Emma and her mother, and Appendix C.21 for Ollie and his mother .

For Rike, a reliable influence of vowel duration on stress marking was restricted to the period between 12 and 30 months of age. Rike's mother used vowel duration as the main cue to differentiate between stressed and unstressed syllables. Rike produced stressed syllables with higher intensity in contrast to her mother, who only used intensity to keep the child's attention on the correct vowel quality and to emphasise secondary stress. F_0 was used in a comparable manner by mother and child. Differences occur only in the use of F_0 to mark secondarily stressed. During babbling, Rike produced vowels with secondary stress similarly to stressed vowels. With the production of meaningful words, the use of F_0 to mark secondary stress changed in the mother's use. At this stage secondarily stressed vowels were produced with lowest F_0 by Rike's mother. Rike's use of nearly all voice quality parameters to mark stress, except *CC* or *IC*, shows that she cannot control these parameters in an adult–like manner: her mother mainly used spectral tilt parameters to mark stress and *CC* and *IC* to distinguish secondary stress.

Emma used all parameters to mark word stress, but in a very variable way. No clear preference in the use of special parameters was observed until the age of 26 months. Her mother preferred vowel duration, F_0, and the spectral tilt parameters to mark stress; *CC* and *IC* were mainly used to mark secondarily stressed vowels.

Ollie preferred F_0 and spectral tilt parameters as well as *GO* to mark stress. Comparable to his mother, he produced unstressed vowels with the lowest F_0, the highest spectral tilt parameter *SK*, *RC* and *T4* values. He also used *GO* like his mother to distinguish secondary stress. Ollie's use of vowel duration and intensity to mark word stress depended on age. Control over these parameters progressed with increasing age.

5.6 Discussion

The children's speech production data analysed in this section provide clear evidence that with the beginning of babbling, at five months of age, children are able to produce perceptibly different stress patterns. However, the implementation and usage of the parameters that contribute to marking word stress appear to be inconsistent. Each measurable acoustic parameter for marking word stress can be observed, and with advancing age, the use of these parameters is very variable and depends on the child. Based on our hypothesis that perception precedes production, we compared the acoustic correlates of word stress produced by the children to those produced by their parents.

The most salient cue for syllabic stress in German is evidently vowel duration. All adults use this parameter to differentiate between stressed and unstressed vowels. In the children, a developmental pattern can be observed. With increasing age, the children are more and more able to reduce vowel duration for unstressed vowels and, therefore, produce perceptible differences in vowel duration for stressed and unstressed syllables.

For the children, intensity and F_0 seem to be the most reliable correlates to mark word stress. A pitch rise may be caused by reflexive tensing in the vocal folds or due to the fact that the increased subglottal air pressure of stressed vowels causes the vocal fold closure to occur more quickly. As intensity as well as F_0 seem to present the same picture, it can be concluded that the increase of both is caused by an increase in subglottal pressure without the necessary muscular adjustments to the vocal folds. Therefore, children have to learn the fine–tuning of the articulators to separate an increase in intensity from an increase in F_0. Apart from one child, a significant decrease in F_0 with increasing age was found.

Although previous studies have reported evidence that intensity is a correlate of word stress in German, our data do not completely support these findings. The use of intensity to mark stress depends on the child's age. When intensity is used to mark stress, stressed vowels are produced with significantly greater intensity than unstressed vowels, but intensity is not used in a constant way to mark stress. It appears that intensity is used more to emphasise stress.

In our data, F_0 was found to be a correlate of word stress, as all parents and children produced unstressed vowels with lower F_0 than stressed vowels. Four mothers demonstrated dependency of F_0 on the age of the child, as a significant decrease with increasing age was observed.

F_0 thus seems to be able to serve as a correlate of word stress in German. As the analysed utterances were produced in spontaneous mother–child and father–child interactions and therefore not spoken in isolation or in otherwise controlled contexts, an effect of sentence intonation and even discourse or dialogue structure cannot be excluded.

The parameter *OQ* is considered to be a correlate of sentence accent and is possibly caused by an additional modulation of muscular tension; some parents also used this parameter to distinguish secondary stress. *GO* has been shown to be a correlate of stress in German adults (Jessen et al., 1995). Our findings did not support this assumption, as no significant difference between stressed and unstressed vowels occurred. *GO* is mainly used to emphasise stress as secondarily stressed vowels are produced with higher values than unstressed and stressed ones.

The use of the spectral tilt parameters *SK*, *RC*, and *T4* confirm previous findings for adult speech that stressed vowels in German are produced with a faster and more abrupt closure of the vocal folds than are unstressed ones. In the children's data a variable and age–dependent use of these parameters was found.

The parameters describing the glottal leakage *CC*, corresponding to a greater F1 bandwidth, and *IC*, a normalised version of this parameter as the influence of the first formant is minimised, were used by two male adults to differentiate between stressed and unstressed vowels. The other adults tended to use *CC* as well as *IC* to distinguish secondary stress. Children used these glottal leakage parameters very variably with the tendency to use them in an adult–like manner to mark secondary stress.

During babbling, children experiment with these parameters in order to find out how they can be used with minimal effort to achieve the desired result based on their articulatory capabilities. These are trained during babbling. A variable use of the possible acoustic realisation of stress can be observed. With the beginning of the production of first words, a reorganisation of the use of the acoustic parameters has to be done; however, until the age of 36 months, the parameters are still not used in an adult–like way for stress production. The children are still too inconsistent in their use of the parameters to mark word stress. They first have to build up speech categories and therefore have to acquire motor control over their articulation for a more adult–like production.

With the beginning of word production, children have to build up speech categories. Therefore, they have to acquire motor control over their articulation, which was trained during babbling. With the beginning of word production, the trained patterns were compared to the input and self–monitored production to build up more adult–like production patterns. During babbling such comparison is not necessary as no production template has to be filled. Given the perception–production loop (Pierrehumbert, 2001), the children's abilities to produce a new category is a function of the number and quality of input elements, and depends on the continuing development of the control and coordination of particular structures in producing stress; and therefore the babbling phase with its proto–syllables is important. The most important cue to mark distinguishable different stress patterns is to learn to reduce vowel duration for the production of unstressed vowels. The children have to learn fine–tuning of the different acoustic parameters to mark word stress.

As input plays an important role, we compared the adult use of correlates of word stress in German with the child's use of the parameter. All parents use vowel duration to mark stress, and also fine–tuning differences in the use of the other parameters can be observed. Not all parameters are used by the parents to mark stress; an individual acoustic pattern for each adult can be seen.

This is not the case for the children. Although each child has different input, all children tend to use all possible acoustic parameters to mark stress. Children thus were playing with all parameters in order to find out how they can be used with minimal effort to achieve the desired result. The output then was compared with the perceived exemplars in their cognitive map. However, until 36 months of age, there is no constant use of the parameters and no preferential use as in their parents. The children were still too inconsistent in their use of the parameters to mark word stress. With increasing age, a more adult–like use of the parameters was achieved.

5.7 Conclusion

The children's data presented in this section suggest that children acquiring German as their native language use different strategies to produce stress. The most reliable correlate of word stress in German children during babbling seems to be intensity and F_0. Vowel duration as the most reliable correlate to mark word stress for adult speakers was also used by the children. However, until the age of about 36 months it was not used in a constant way. With increasing age, children show tendencies to use all the stress cues reported for adults, such as vowel duration, intensity, glottal opening, skewness, rate of closure and completeness of closure, in their parents–like usage pattern, but not consistently.

For the adult data, vowel duration was the most reliable correlate of word stress in German. All speakers produced unstressed vowels with reduced vowel duration. Vowels with secondary stress were mainly produced like stressed vowels. Final syllable lengthening can, but does not have to , occur.

Stressed vowels were also produced with higher intensity than unstressed vowels. Intensity did not seem to be a reliable correlate of word stress. The main use of intensity was to emphasise stress and to kept the child's attention on a wrong vowel or stress pattern in the child's production.

F_0 was used to mark stress in a consistent way, as unstressed vowels were produced with lower F_0 than stressed ones. Vowels with secondary stress were mainly produced like unstressed vowels. An age–dependency of F_0 with a decrease of F_0 with increasing age of the child could be observed for four of the six mothers but not for the fathers.

The voice quality parameters describing spectral tilt are reliable correlates to mark word stress in German. The effect of stress depends on the increase and decrease of subglottal pressure and the fine–tuning of the vocal fold vibration with muscular adjustments.

The parameters *GO* as well as *OQ* were mainly used to emphasise secondarily stressed.

Two of the male adults prefer ed the parameters describing the closing phase of the production (*CC* and *IC*) to mark word stress in German. The other adults used these parameters only to distinguish secondary stress.

To conclude, vowel duration as well as F_0 and spectral tilt parameters are used to differentiate stressed and unstressed syllables. Stressed vowels are produced with longer vowel duration, higher F_0 values and with faster and more abrupt closure of vocal folds than unstressed vowels. No clear tendencies can be observed for the production of secondary stress. In German, secondary stress is mainly used in compounds. In German word compounds, the initial constituent receives the main stress and the other constituent the secondary stress, when compounds are derived from monosyllabic or trochaic constituents. However, no clear use of vowel duration, F_0 or spectral tilt parameters to mark secondary stress was found.

To distinguish secondary stress, the parameters *OQ*, *GO* as well as *CC* and *IC* were used by the parents. This provides evidence that the production of secondary stress mainly depends on the use of the voice quality parameters. The fine–tuning of the vocal fold, described with the different voice quality parameters, is responsible for emphasis stress.

6 Vowel development

6.1 Introduction

The nucleus of the syllable is the vowel. Children are able to produce specific acoustic patterns, which are associated with a given articulatory template. During childhood, these relationships were affected by anatomical modifications. As motor control skills gradually develop, the vowel space has to expand and different categories of vowels have to be formed. As stressed and unstressed vowels differ in articulatory organisation (Dogil, 1995), it can be assumed that stressed and unstressed syllables show different dispersions of the vowel categories in the perceptual space. Therefore it can be concluded that different representations for stressed and unstressed vowels need to be developed. The relative position of the vowels in the acoustic space depends on age and motor control. It requires a recalibration of the articulatory–acoustic links to reach a perceptual template.

The acoustic characteristics of vowels produced by English and French learning infants have been described in a number of previous studies (Kent and Murray, 1982; de Boysson-Bardies et al., 1989; Lee et al., 1999; Davis et al., 2000; Rvachew et al., 2006; Ménard et al., 2007b). These studies used cross sectional data to describe vowel development and are therefore limited in describing the individual development (Smith and Kenney, 1998).

Major differences between the vocal tracts of infants and adults can be noted. Infants have a much shorter vocal tract than adults, a shorter pharynx relative to the overall length, a wider vocal tract relative to its overall length, and a very flat oral cavity due to the fact that infants have no teeth (Goldstein, 1980). In order to acquire the vowel system of their native language, the infants have to overcome these physiological limitations imposed by the immature speech motor control as the vocal tract morphology does not prevent the infants from achieving the perceptual targets representing the native vowels (Goldstein, 1980; Buhr, 1980; Ménard et al., 2004).

Due to this fact, it can be postulated that the anatomy of the vocal tract does not prevent infants from speaking, rather their limited neuromuscular control as well as the development of proprioception. A gradual increase in the range of vowels produced with age can be observed (Buhr, 1980; Lieberman, 1980; Selby et al., 2000; Rvachew et al., 2006; Vallabha et al., 2007). This increase is constrained by the sequential emergence of motor control, which gives rise to limited articulation abilities for infants and children (MacNeilage, 1998; MacNeilage and Davis, 2000; Lacerda, 2003).

Apart from the motor control development, the structure of the human vocal tract and its important modifications during development (Vorperian et al., 2005) play an important role in the production of the full range of vowels used in the native language. In order to gain language–specific vowel space (Ménard et al., 2004), children have to recalibrate their articulatory–acoustic links restricted by their gradual mastery of motor control as well as anatomical modifications. These anatomical modifications cause transformations of the learnt articulatory template to reach an acoustic target based on the production–perception links (Ménard et al., 2007a). The perceived speech categories contain enough acoustic structure for explicit unsupervised acquisition of a number of categories and their respective distributions in the native language (Vallabha et al., 2007). The children have to develop articulatory routines to reach a specific target and to associate a given articulatory strategy with specific acoustic patterns (Dogil, 2010). These production–perception links are encoded by the speaker and specify the speech task. Auditory feedback is essential for the development of different articulatory gestures (Rvachew et al., 1996) as the correction of feed–forward commands is error–based (Lane et al., 2005).

These different aspects of speech production are associated with the development of a variety of anatomical and neurophysical structures and functions of the vocal tract and nervous system developed for each child individually and not necessarily within the same time frame for each child's development. Based on the non–uniform vocal tract growth (Fant, 1960; Vorperian et al., 2005) the development of the infants' articulatory skills demands the mastery of a complex sensory–motor system. During his first babbles, the output of an infant is characterised by a neutral configuration based on the rhythmic mandibular cycle, simulating 'consonants' in the closure phase and 'vowels' in the open phase (MacNeilage, 1998; Davis et al., 2002). A strong preference for central vowels with very little developmental change in the location can be observed. A gradual expansion of the range of vowels, first associated with different F1 values, can be observed as the child grows in age (Buhr, 1980; Kent and Murray, 1982; Rvachew et al., 1996; Gilbert et al., 1997; Robb et al., 1997; Serkhane et al., 2007).

6.2 Background

The first study on the acoustics and perception of vowels was conducted by Peterson and Barney (1952) shortly after the introduction of the sound spectrograph. In addition to men and women, they also recorded and analysed two repetitions of ten vowel productions from 15 children. Unfortunately, the results were not reported separately for the different genders and ages and no information is given about the age or gender of the children. Since this study, the developmental change in spectral resonant peaks such as formants and fundamental frequency has been intensively investigated. These acoustic features reflect the acquisition of manners of articulation as well as the maturity of the articulatory organs such as the length of the vocal tract, the descent of the larynx, and the size of the pharynx. Studies on the acoustic development of vowel formants mainly covered the age between 3 and 18 years (Hillenbrand et al., 1995; Busby and Plant, 1995;

Palethorpe et al., 1996; Smith and Kenney, 1998, 1999; Lee et al., 1999; Whiteside, 2001; Yildrim et al., 2003) and have shown that children's speech is characterised by higher pitch and formant frequencies, longer segmental durations, and greater spectral and temporal variability compared to adults speech. These results also confirmed that due to the maturation of the vocal tract length, the formant frequencies decrease with age and that maturity continues through the middle and late teen years (Huber et al., 1999).

Between birth and 36 months of age, the vocal tract structures exhibit accelerated growth (Vorperian et al., 2005). This means that in the first few years, the acoustic features of speech are expected to change greatly. It has been shown that the vowel categories become perceptually more separated in the vowel space between three to five months of age (Kuhl and Melztzoff, 1995, 1996). Children try to imitate the vowels they hear, despite the immaturity of their articulation organs, by using their articulation skills. By analysing the resonant frequencies of infant vocalic utterances at 3, 6, and 9 months of age, Kent and Murray (1982) showed an expansion of the *F1* and *F2* values with age from 3 to 9 months.

It has been reported that formant values remain unchanged from 15 to 24 months of age, and decrease significantly from 24 to 36 months (Buhr, 1980; Rvachew et al., 1996; Gilbert et al., 1997; Robb et al., 1997). However, by describing the vowel spaces produced by 10–to 18–months–old infants exposed to either Canadian English or Canadian French, Rvachew et al. (2006) found a language–specific decline of F1 for French–learning infants and F2 for English–learning infants. A developmental expansion of the vowel space into the high–front and high–back regions was also evident. The influence of the native language on the development of the vowel space was also observed in a cross–cultural investigation of the influence of the target language on babbling (de Boysson-Bardies et al., 1984, 1986, 1989; Halle et al., 1991; Rvachew et al., 2006). This finding implies that the native language influences the development of the internal representations and that the infants try out different articulatory patterns for the configuration of the vocal tract (MacNeilage et al., 2000).

To summarise, an expansion in the vowel space with increasing age can be observed. As speech development shows a great inter–speaker variability, especially in the early stages, these cross–sectional studies can only show the global process of the development of vowel space within a limited age range. In order to look at the development process, longitudinal studies and analyses are needed.

Lieberman (1980) collected speech data from five infants between 4 and 60 months of age and showed in his analysis of formant–frequency plots over time the emergence of a well–developed vowel triangle. Buhr (1980) reported a gradual growth of the vowel space between 4 and 16 months of age for a single infant mainly in the front–back dimension depending on the increased ability to control the tongue, independent of the jaw height. Rvachew et al. (1996) described the vowel development from 18 infants from 6 to 18 months of age. In this latter study, the auditory environment was also taken into account. The development of vowel space from infants who have had otitis media in the first six months of life was compared with infants who have no auditory restrictions.

A similar developmental pattern to that reported by Kent and Murray (1982) was found with a significantly smaller increase in F2 for infants suffering from otitis media in the first six months of life. Problems in the auditory system can result in a restricted vowel space.

Ishizuka et al. (2007) describe the vowel development of two Japanese infants (one male, one female) in terms of spectral resonant peaks and cover the age of 4 to 60 months. They provided detailed findings on the developmental process of speech production and analysed labelled as well as unlabelled data. For labelled as well as unlabelled data they found categorisation around the age of 20 months. At this age, different categories can be determinded using linear discriminatory analysis methods.

To conclude, children are able to produce specific acoustic patterns, which are associated with a given articulatory template. During childhood, these relationships were affected by anatomical modifications. As motor control skills gradually develop, the vowel space also has to expand and different categories of vowels have to be formed.

The relative position of the vowels in the acoustic space depends on age and motor control (Ménard et al., 2007b) and requires a recalibration of the articulatory–acoustic links to reach a perceptual template.

This chapter aims to investigate when and how infants become able to produce categorically separated vowels as well as to study the internal organisation of the vowel system during vocal tract growth.

The following questions were asked concerning morphological differences:

- What are the articulatory strategies used by the infants throughout their development to produce adult–like vowels?
- Which strategies do they need to enable them to produce equal acoustic targets as they increase in age.

In this chapter, it is demonstrated that the production of German vowels, from initial canonical babbling to the first real words, is guided by acoustic targets in the F1, F2, and F3 space which require different articulatory strategies during the course of development.

6.3 Development of acoustic vowel spaces

The reported studies were mainly based on monosyllabic utterances and the analysed vowels can be stated as stressed. It can be assumed that stressed and unstressed syllables create different categories (see Chapter 4.4) in the perceptual space. Stressed and unstressed syllables in German seem to differ in articulatory organisation (Dogil, 1995). Based on this assumption the different representations of stressed and unstressed vowels in the perceptual space should be described as well as the the development of these different vowel spaces. For the analysis of the acoustic vowel space the data were taken from a database of all 15 German vowels uttered by 27 speakers ranging from 5 months of age to adulthood. The occurrences of all 15 German vowels /i, ɪ, y, ʏ, u, ʊ, e, ø, o, ɛ, œ, ɔ, a, ə, ɐ/ in natural speech context produced by speakers at various ages were analysed. Between 5 months and 3 years of age, the analysed data come from the described longitudinal recordings of the children. The analysed adult data consist of the corresponding parents' data. As the longitudinal data only went up to the age of 3 years and development has not yet finished at this age, cross–sectional data was also analysed. For this data, additional recordings were done in kindergarten and primary school using the picture naming task described in Section 3.1.3. In Appendix A.3 an overview of the recordings is given. In summary, it was possible to analyse child speech recordings up to the age of 10. Therefore alongside the existing sample of six recorded children from 5 up to 36 months of age, further analysis was done at 5, 6, 9, and 10 years. of age.

Nine age groups were analysed (years;months):

- 0–year–old group: averaged age: 0;7 (from 0;5 to 0;9)
- 1–year–old group: averaged age: 1;3 (from 1;0 to 1;11)
- 2–year–old group: averaged age: 2;4 (from 2;0 to 2;9)
- 3–year–old group: averaged age: 3;0 (from 2;10 to 3;0)

For these age groups the longitudinal recordings were grouped together.

- 5–year–old group: averaged 5;4 (from 5;2 to 5;10), four subjects
- 6–year–old group: averaged 6;7 (from 6;4 to 7;0), four subjects
- 9–year–old group: averaged 9;3 (from 9;4 to 9;11), two subjects
- 10–year–old group: averaged 10;1 (from 10;0 to 10;1), two subjects

Initial analysis of the formants shows an influence of stress (see 5.4.4). To describe the development of different vowel acoustic spaces depending on stress, vowels categorised as stressed and unstressed were analysed separately. The categorisation of level stress is not clear as neither in the development of syllables types (see 4.4) nor in the acoustic realisation (see 5.3) a own category for this stress patterns emerge. It is not clear whether the category for level stress had not emerged yet for other reasons or whether they were present in representation but not in production. Perhaps the frequency of level stressed vowels (and syllables) was too low to as a critical number of such exemplars must be perceived to form a new category.

6.3.1 Dispersion ellipses

In order to describe the development of the acoustic vowel space for stressed and unstressed vowels, typical dispersion ellipses[1] are calculated for the F1–F2 as well as F2–F3 vowel space. The data cover 9 speakers from the described age groups, using the perceptual Bark scale for frequencies (see Chapter 3.1). The overall sizes of the shown vowel spaces were aligned for a better comparison of the different distributions of the dispersion ellipses in the vowel space. In Appendix D.1 the colour code for the different vowels is given. For the longitudinal data, only one boy (Bennie) showed a nearly adult-like vowel space at 36 months of age. The other five children still showed great variability and overlapping of the different vowel spaces. Remarkably for the parents of Bennie were the production of a clearcut vowel space with nearly no overlapping of the different dispersion ellipses. As a notice, maybe this helped the boy to detect the right distribution in the vowel space before the other children. The other parents did not showed such a clear distribution in the vowel space. More overlapping of different distribution ellipses were observed. Unfortunately the longitudinal data finished at about 36 months of age. Therefore, no conclusion about their further vowel development could be made within this study but will provide the impetus for future studies.

To conclude, the acoustic values for each vowel differ with stress. Stressed and unstressed vowels show different dimensions of the dispersion ellipses as well as a different expansion in the vowel space. Unstressed vowels were more centralised – even in adult production – as can be seen in Figure 6.9. Furthermore, the acoustic values for each vowel differ among the speakers as can be seen in the various expansions in the vowel space. These differences reflect the non–uniform vocal tract growth and development of articulatory control (Hillenbrand et al., 1995; Lee et al., 1999).

The variability amongst the speakers, as revealed by the size of the dispersion ellipses, decreases over age, in spite of the Bark correction. By about 3 years of age, the children have mastered the production of all vowels and try to produce the four cardinal vowels /i/, /u/, /a/ and /y/ at the extreme limits of the F1 vs. F2 vowel space. No clear separation in the F2 vs. F3 vowel space can be observed.The most important cues to identify vowel quality seem to be the F1 vs. F2 vowel space.

[1] with radii corresponding to $\pm$ 1.5 standard error of the individual data around the mean

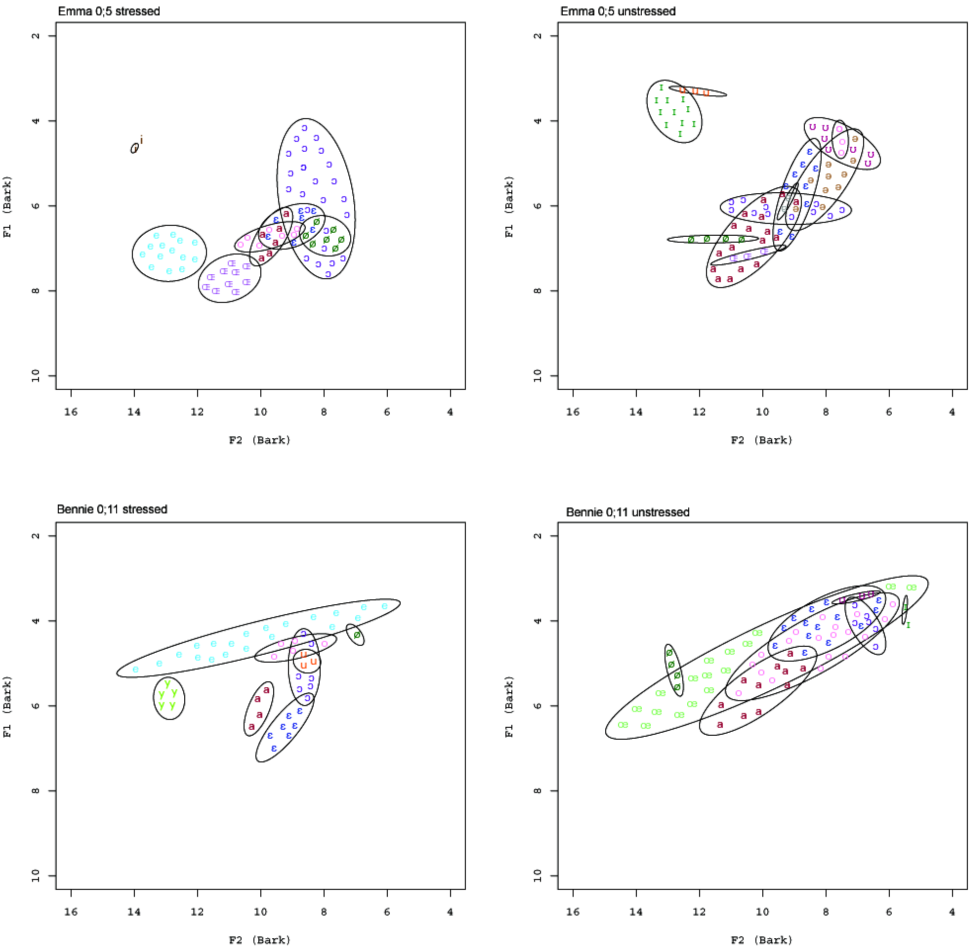


Figure 6.1: Dispersion ellipses (± 1.5 standard errors from the mean) in the F1 vs. F2 space in Bark for **stressed** (left) and **unstressed** vowels of German. Data are presented for one speaker representative for the age group 0, and 1 year.

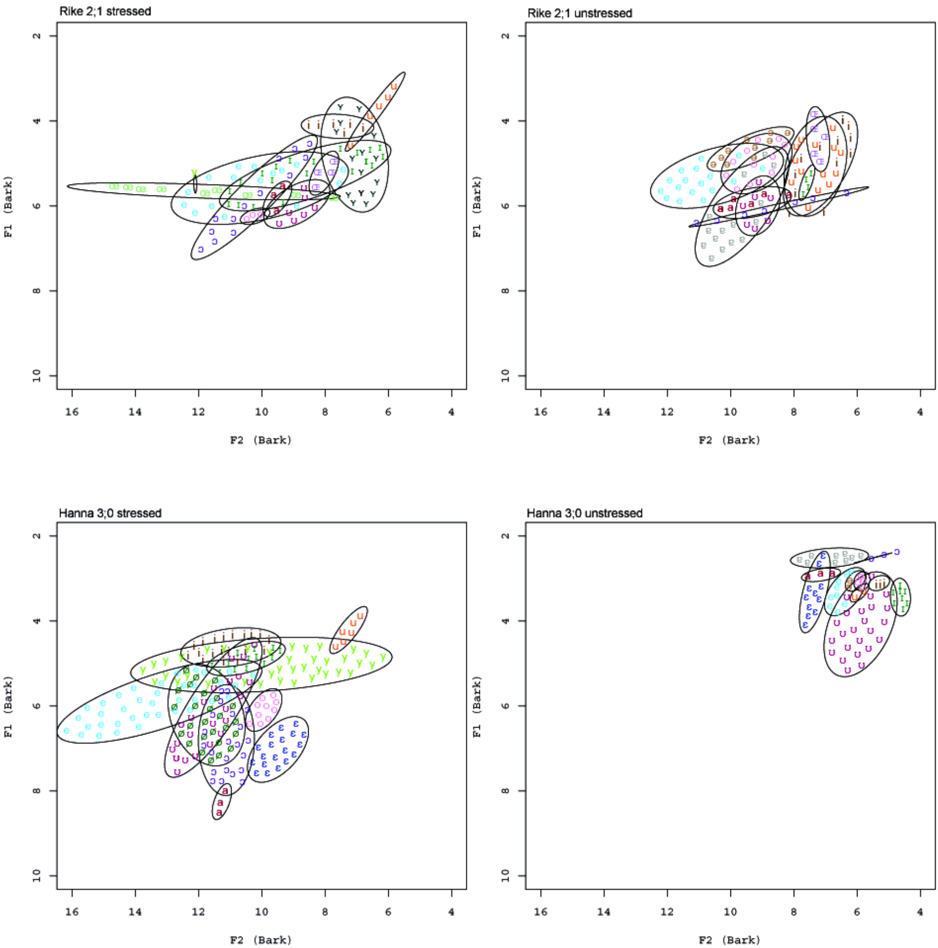


Figure 6.2: Dispersion ellipses (± 1.5 standard errors from the mean) in the F1 vs. F2 space in Bark for **stressed** (left) and **unstressed** (right) vowels of German. Data are presented for one speaker representative for the age group 2 and 3 years.

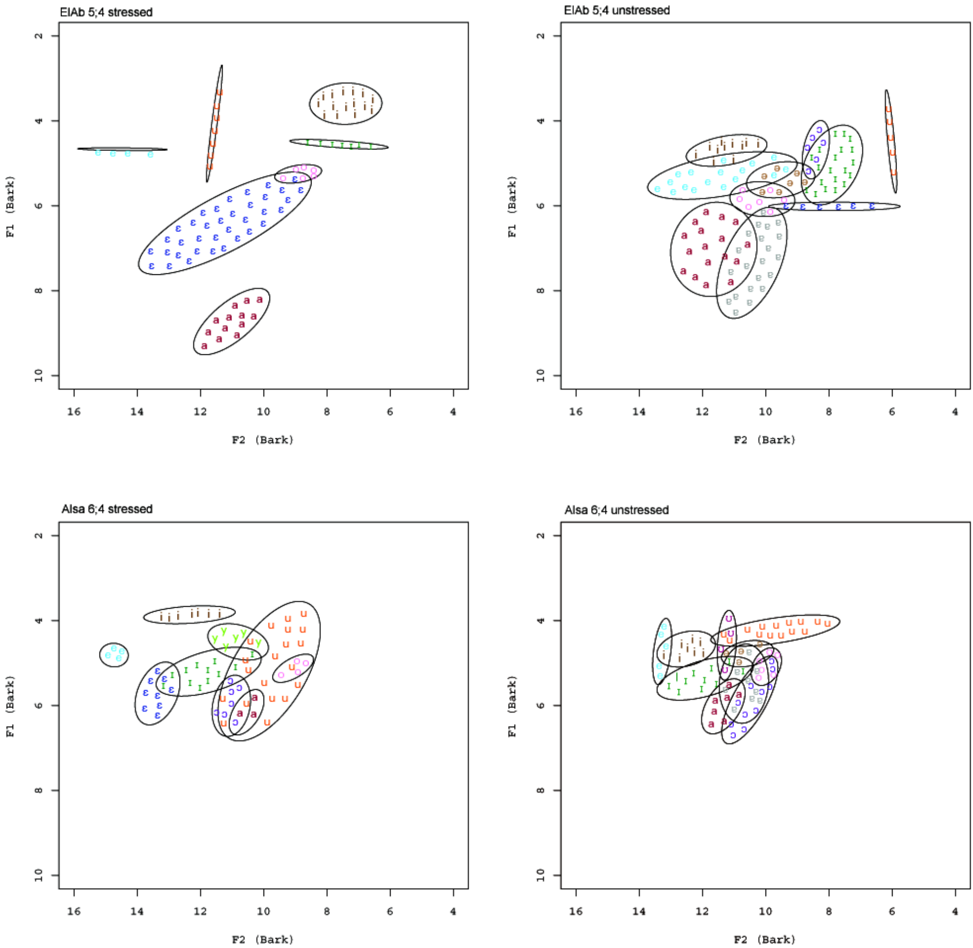


Figure 6.3: Dispersion ellipses (± 1.5 standard errors from the mean) in the F1 vs. F2 space in Bark for **stressed** (left) and **unstressed** (right) vowels of German. Data are presented for one speaker representative for the age group 5, and 6 years.

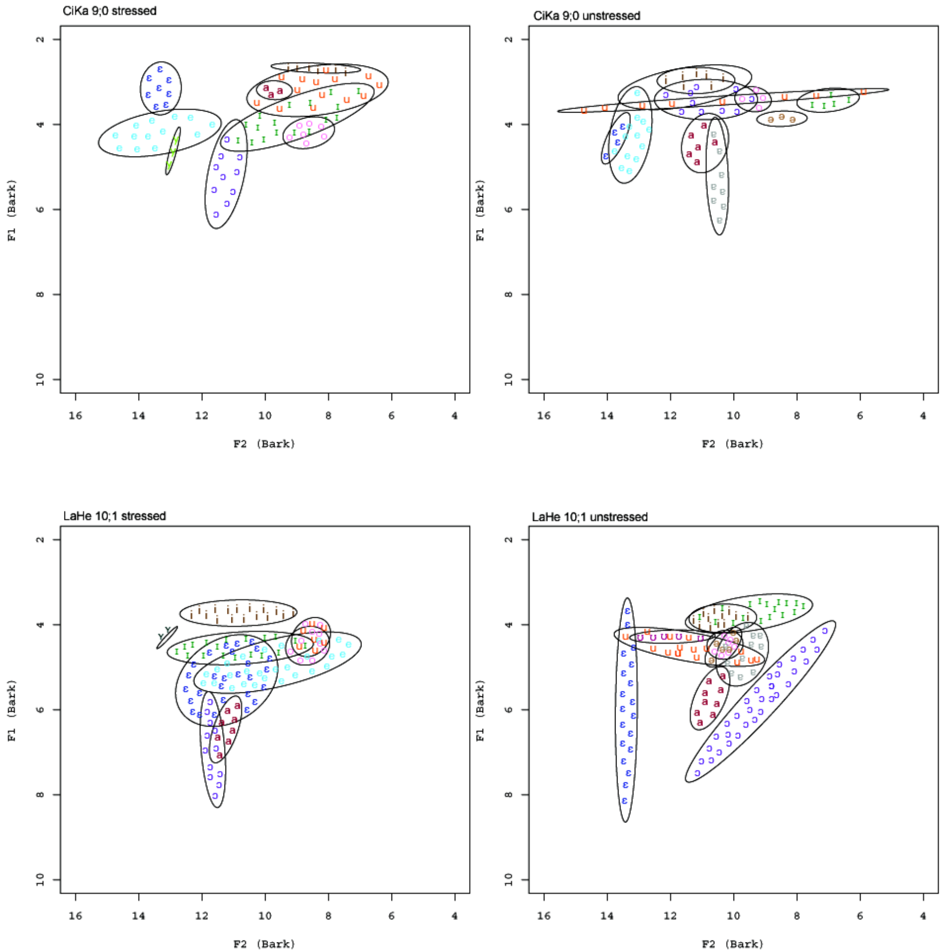


Figure 6.4: Dispersion ellipses ($\pm$ 1.5 standard errors from the mean) in the F1 vs. F2 space in Bark for **stressed** (left) and **unstressed** (right) vowels of German. Data are presented for one speaker representative for the age group 9 and 10 years.

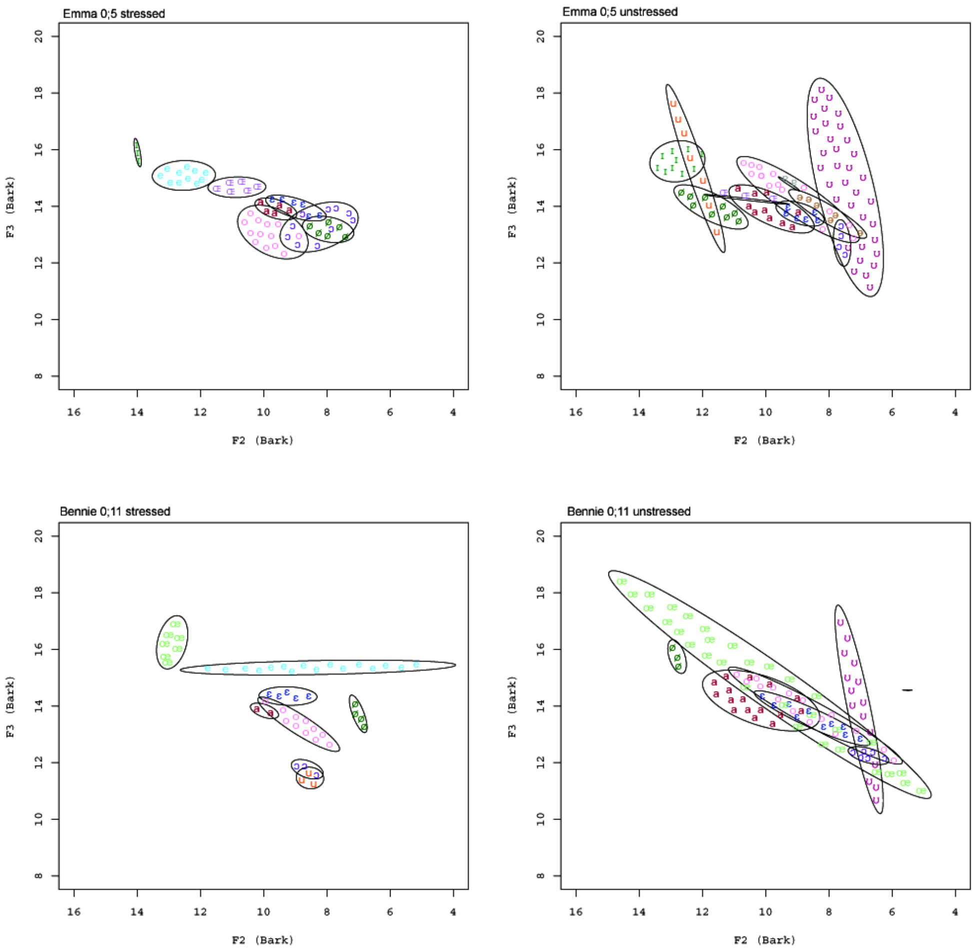


Figure 6.5: Dispersion ellipses (± 1.5 standard errors from the mean) in the F2 vs. F3 space in Bark for **stressed** (left) and **unstressed** (right) vowels of German. Data are presented for one speaker representative for the age group 0 and 1 year.

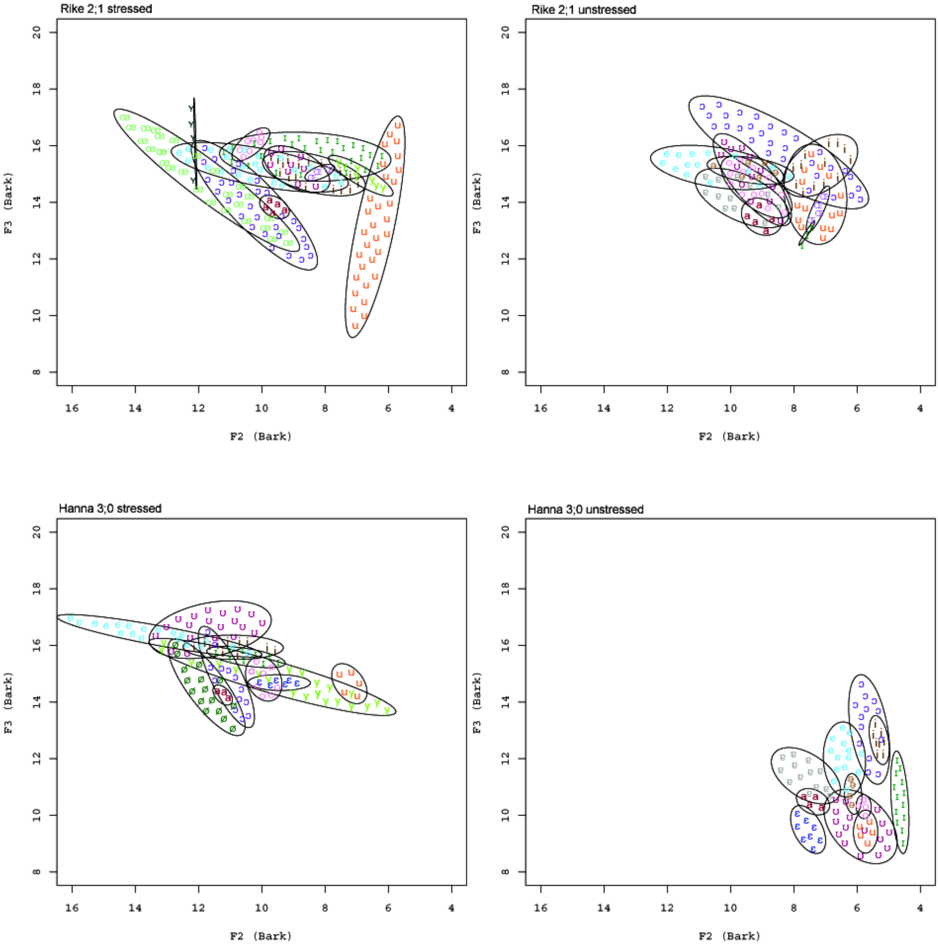


Figure 6.6: Dispersion ellipses (± 1.5 standard errors from the mean) in the F2 vs. F3 vowel space in Bark **stressed** (left) and **unstressed** (right) vowels of German. Data are presented for one speaker representative for the age group 2 and 3 years.

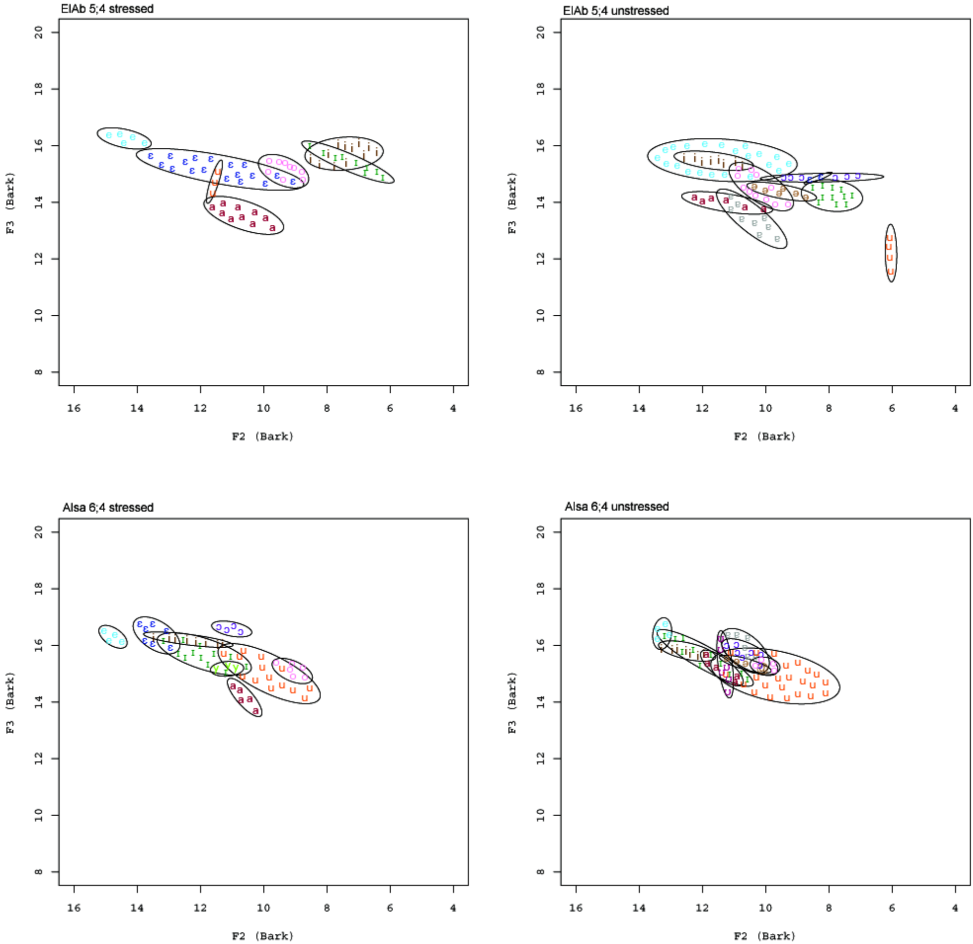


Figure 6.7: Dispersion ellipses (± 1.5 standard errors from the mean) in the F2 vs. F3 vowel space in Bark **stressed** (left) and **unstressed** (right) vowels of German. Data are presented for one speaker representative for the age group 5 and 6 years.

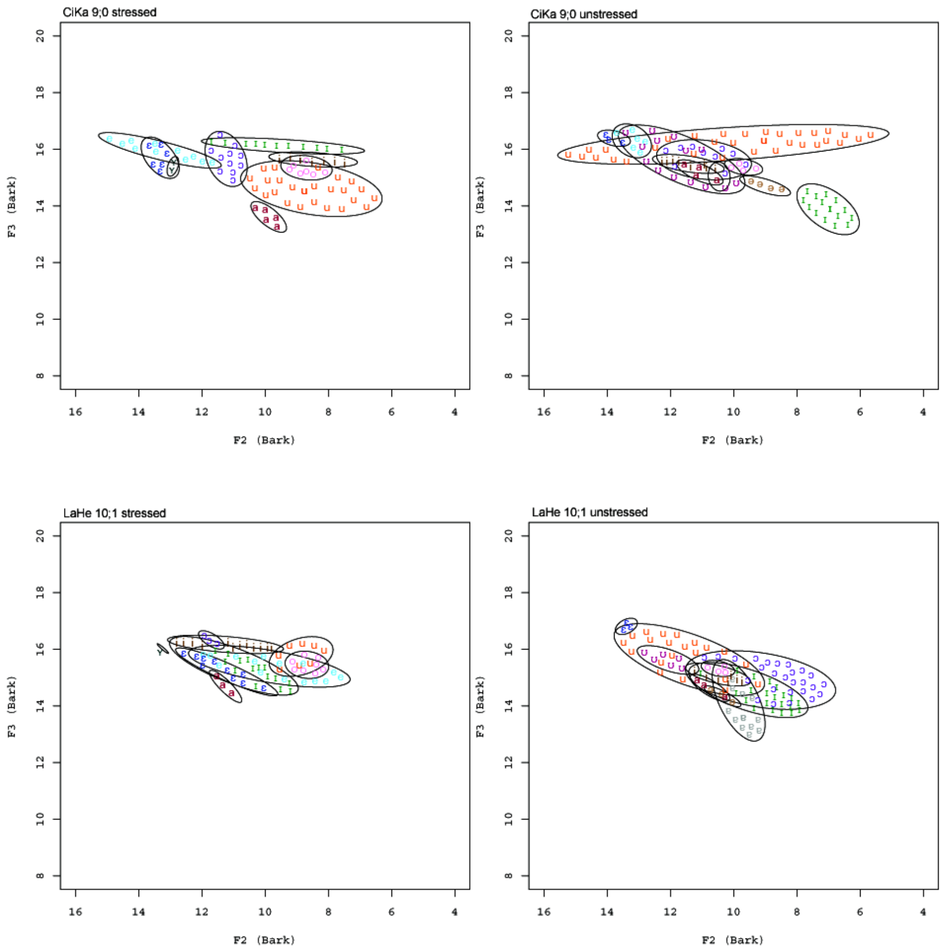


Figure 6.8: Dispersion ellipses (± 1.5 standard errors from the mean) in the F2 vs. F3 vowel space in Bark of **stressed** (left) and **unstressed** (right) vowels of German. Data are presented for one speaker representative for the age group 9 and 10 years.

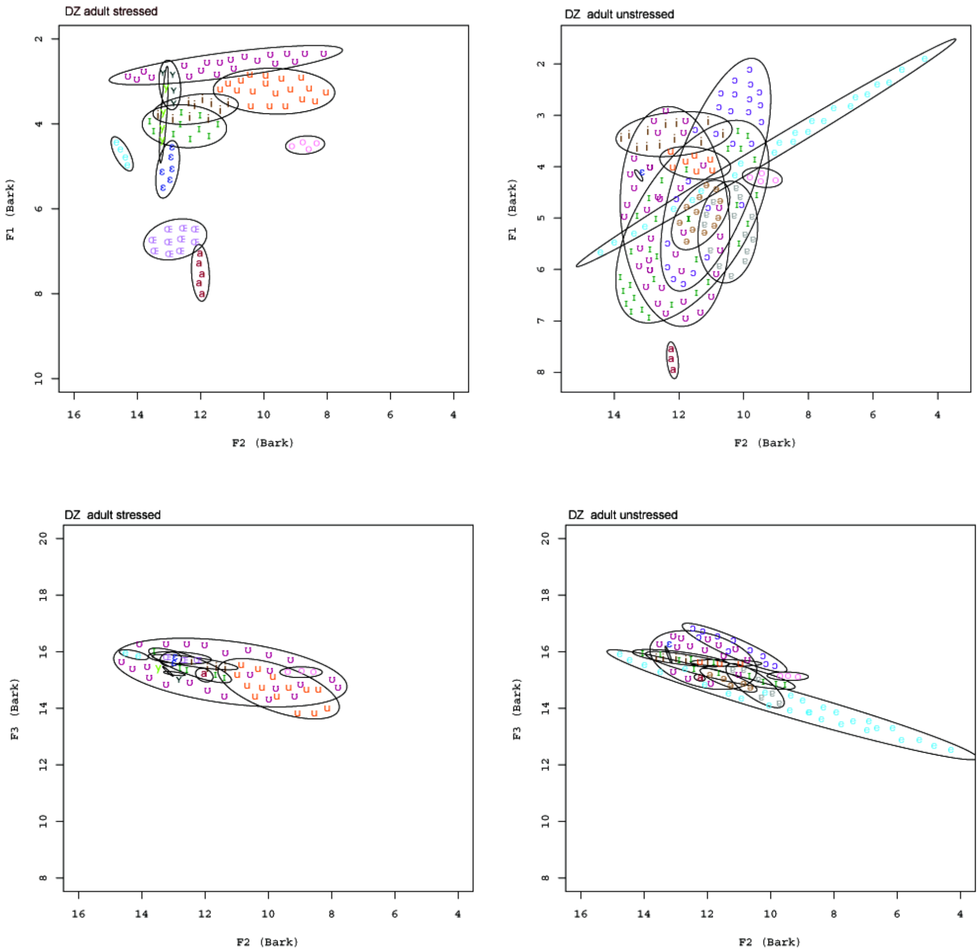


Figure 6.9: Dispersion ellipses ($\pm$ 1.5 standard errors from the mean) in the F1 vs. F2 (top) and F2 vs. F3 (bottom) vowel space in Bark of **stressed** (left) and **unstressed** (right) vowels of German. Data are presented for one speaker representative for adult speakers.

6.3.2 Development of vowels

To analyse, if the developmental pattern of vowels differed for stress, stressed and unstressed vowels were grouped according to their height, front–backness, and roundness. Five groups of vowels were chosen according to the height dimension.

- high: /i/, /ɪ/, /y/, /ʏ/, /u/, /ʊ/
- mid–high: /e/,/ø/,/o/
- mid–low: /ɛ/,/œ/,/ɔ/
- central: (only unstressed vowels): /ə/, /ɐ/
- low: /a/

Two groups of vowels were chosen according to the front–backness dimension, whereas the front vowels were also grouped according to their lip rounding.

- front, unrounded : /i/, /ɪ/, /e/, /ɛ/
- front, rounded : /y/, /ʏ/, /ø/,/œ/
- back: /u/, /ʊ/, /o/,/ɔ/

The area of each ellipsis [2] in the F1 vs. F2 vs. F3 space, in Bark, was calculated for each age and each speaker. The mean dispersion values of the vowels grouped in different heights as well as for front, back, rounded and unrounded vowels are given for each age group in Figure 6.10 for stressed and in Figure 6.11 for unstressed vowels in the dimension of F1 vs. F2.

For stressed and unstressed vowels in the F1 vs. F2 and F2 vs. F3 vowel acoustic space, univariate ANOVAs were carried out for each feature group. Area values was the dependent variable and feature and age group were the factors. In Table 6.3.2 the results of the ANOVAs are given. A developmental pattern were observed as the effect of age was indeed significant on the level $p<0.05$ for all features for stressed as well as for unstressed vowel space in the F1–F2 and F2–F3 vowel space dimension. For all three features the dispersion ellipses showed an increase of the areas until the age of about five years and then a slight decrease were observed. But the area of the dispersion ellipses up to the age of 10 years were larger and showed more variability than for the adult speakers. In the F1–F2 vowel space, height had a significant effect on the area of dispersion ellipses for stressed vowels. In the F2–F3 vowel space, height was significant for stressed as well as for unstressed vowels. The height factor showed no significant interaction between stress and age. Therefore, the differences in area depended more on

[2] If a and b are, respectively, the lengths in Bark of the major and minor axes of the dispersion ellipsis, provided by the diagonalisation of the covariance matrix, the area is provided by: Area = πab.

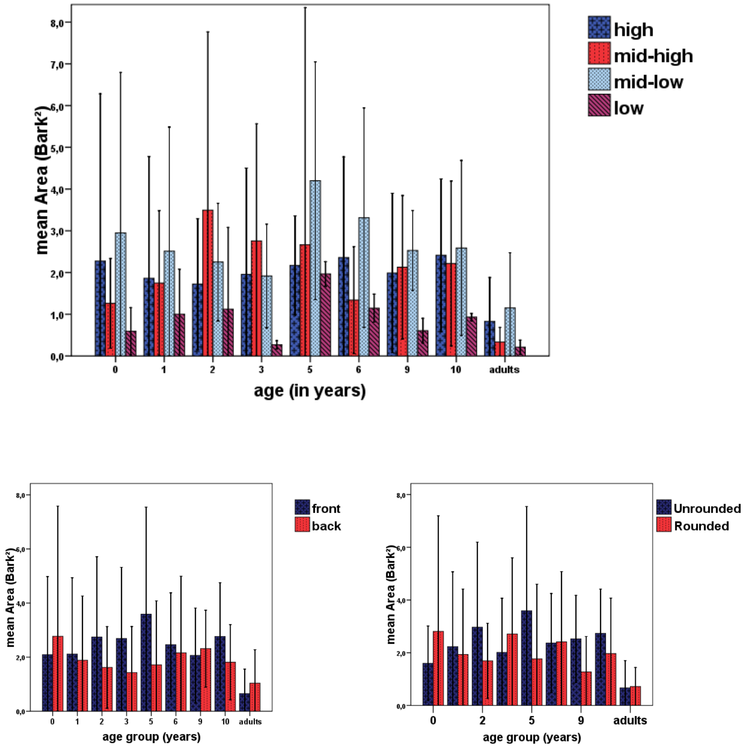


Figure 6.10: Area of the dispersion ellipses related to the produced **stressed** vowels, in the F1 vs F2 dimension (in $Bark^2$), grouped according to their height, front–backness, and roundness. The x–axis represents the nine speaker groups: 0–, 1–, 2–, 3–, 5–, 6–, 9–, 10– years old, and adults.

inter–speakern variability than on age. Low vowels were produced with the lowest area of dispersion ellipses thus with less variability than the other height groups. High vowels were produced with the largest area of dispersion ellipses but only in the F2–F3 vowel space, which gives evidence for a greater variability of F3. Front–backness showed a significant effect on unstressed vowels in the dimension of F1–F2. Until the age of 10 years, the children produced front vowels with greater area of dispersion ellipses than back vowels. Adult speakers in contrast produced back vowels with greater area. Round-

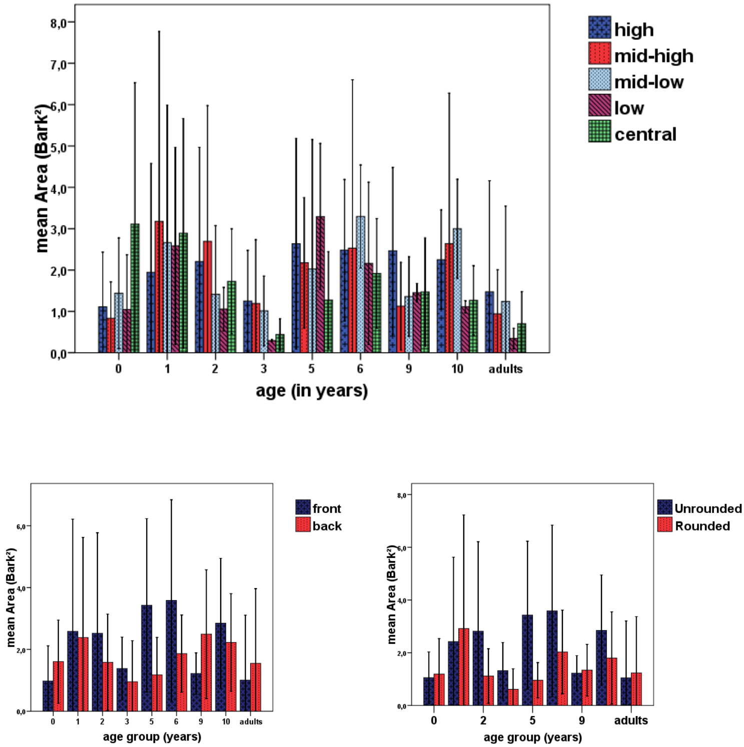


Figure 6.11: Area of the dispersion ellipses related to the produced **unstressed** vowels, in the F1 vs. F2 dimension (in $Bark^2$), grouped according to their height, front–backness, and roundness. The x–axis represents the nine speaker groups: 0–, 1–, 2–, 3–, 5–, 6–, 9–, 10–years old, and adults.

ing was significantly affected by age. The main articulators of rounding are the lips and children gained better motor control of their lips with age. In comparison to the vowel space for stressed and unstressed vowels, unstressed vowels were produced with lower formant frequencies and more centralised. To sum up, different vowel spaces for stressed and unstressed vowels have to be developed. With the beginning of word production, the vowel space was expanded to the expected adult vowel space. The cardinal vowels were produced at the extreme limits of the F1 vs. F2 vowel space.

group	stress	vowel space	source	degree of freedom	F Statistic	p Value
height	unstressed	F21	age	8,587	4.148	**0.000***
			height	4,587	0.374	0.827
			age*height	32,587	0.646	0.935
	stressed	F21	age	8,615	2.959	**0.003***
			height	3,615	4.053	**0.007***
			age*height	24,615	0.776	0.769
	unstressed	F23	age	8,596	5.181	**0.000***
			height	4,596	4.047	**0.003***
			age*height	32,596	0.763	0.825
	stressed	F23	age	8,606	2.705	**0.006***
			height	3,606	6.004	**0.000***
			age*height	24,606	0.581	0.946
front–backness	unstressed	F21	age	8,447	1.993	0.046
			front–backness	1,447	1.338	0.239
			age*front–backness	8,447	1.325	0.229
	stressed	F21	age	8,561	2.624	**0.008***
			front–backness	1,561	2.427	0.120
			age*front–backness	8,561	1,124	0.345
	unstressed	F23	age	8,455	4.746	**0.000***
			front–backness	1,455	9.332	**0.002***
			age*front–backness	8,455	0.694	0.697
	stressed	F23	age	8,554	3.505	**0.001***
			front–backness	1,554	2.906	0.089
			age*front–backness	8,554	0.469	0.878
rounding	unstressed	F21	age	8,360	2.422	**0.015***
			rounding	1,360	2.904	0.089
			age*rounding	8,360	1.422	0.186
	stressed	F21	age	8,469	3.061	**0.002***
			rounding	1,469	1.298	0.255
			age*rounding	8,469	1.267	0.259
	unstressed	F23	age	8,368	4.513	**0.000***
			rounding	1,368	0.399	0.528
			age*rounding	8,368	0.925	0.496
	stressed	F23	age	8,462	4.746	**0.000***
			rounding	1,462	1.638	0.201
			age*rounding	8,462	0.915	0.504

Table 6.1: Results of the univariate ANOVAs for each feature group (height, front–backness, rounding) for stressed and unstressed vowels in the F1–F2 (F21)and F2–F3 (F23) vowel space.

6.3.3 Individual vowel development

The individual development were described using the longitudinal data. The same statistical analyses were done individually for each child. In order to evaluate the development of different vowel spaces for stressed and unstressed vowels, univariate ANOVAs were carried out on the area of the dispersion ellipses as dependent variable and stress and child's age as factors. If stressed and unstressed vowels were produced in the same manner, no significant differences between the area should occurred as the variance should be the same. If two different vowel spaces have to be created, the frequency of occurrence as well as the area the observed feature groups should be dependent on stress and should show significant different developmental patterns. Therefore, the frequency of

Table 6.2: Frequency of occurrence(%) according to vowel height, front–backness, and roundness for **Hanna** depending on her age (in months).

stress	feature	5,0	7,0	15,0	18,0	20,0	27,0	30,0	34,0	36,0
unstressed	central	2,3	7,5		4,6	67,7	42,1	30,7	36,7	24,2
	high	11,6	12,2		22,7		5,3	30,7	15,6	28,7
	low	32,6	30,8	61,1	29,6	22,6	26,3	4,4	18,9	13,5
	mid–high	9,3	16,8	5,6	13,6	9,7	26,3	30,7	25,6	29,2
	mid–low	44,2	32,7	33,3	29,6			3,5	3,3	4,5
	back	9,3	17,8	5,6	13,6	6,5	15,8	21,1	21,1	29,8
	front	55,8	43,9	33,3	52,3	3,2	15,8	43,9	23,3	32,6
	rounded	18,6	11,2	11,1	4,6			0,89		1,1
	unrounded	37,2	32,7	22,2	47,7	3,2	15,8	44,0	23,3	31,5
stressed	high	10,8	6,2	14,6	19,0	26,9	42,9	26,6	37,9	33,3
	low	35,1	39,2	53,7	44,0	38,5	21,4	34,4	39,7	29,4
	mid–high	2,7	12,4	12,2	18,0	26,9	21,4	23,4	5,2	22,6
	mid–low	51,4	41,2	19,5	19,0	7,7	14,3	15,6	17,2	14,7
	back	24,5		6,2	26,8	13,0	26,9	28,6	26,6	12,1
	front	64,9	53,6	19,5	43,0	34,6	50,0	39,1	48,3	46,1
	rounded	8,1	9,3	2,4	5,0	3,9	21,4	1,6	1,7	3,9
	unrounded	56,8	44,3	17,1	38,0	30,8	28,6	37,5	46,6	42,2

occurrence depending on the observed feature groups as well as their area were analysed for each child. Additionally, the development of the cardinal vowels /i/, /y/, /a/ and /u/ at the extreme limits of the F1 vs. F2 vs. F3 vowel space were described. The upper left corner of the vowel space (lowest F1, highest F3) is represented by /i/, /u/ corresponds to the upper right corner (lowest F1, lowest F2). The lower limit of the vowel space with highest F1, is represented by /a/. The lower left corner of the F2 vs. F3 vowel space is represented by /y/.

Hanna

For Hanna, the univariate ANOVA of the area of the dispersion ellipses showed a significant main effect of stress in the F1–F2 space ($F(1,125)=15.382$; $p<0.05$). The factor age and the interaction between stress and age were not significant. For the F2 vs. F3 vowel space no main effects were significant. Hanna produced stressed and unstressed vowels with differences in the variability, independent of her age. She mainly produced stressed vowels with larger area values of the dispersion ellipses and therefore with greater variability than unstressed vowels. With increasing age the size of the dispersion ellipses increased, but not significantly. In the F2–F3 vowel space no developmental pattern could be observed. F2 and F3 were produced with too much variability for significant results. Hanna produced unstressed vowels with lower frequencies. No developmental pattern could be observed for the unstressed vowels. They showed a different dispersion in the vowel space than stressed vowels. F1 as well as F2 showed lower values, unstressed vowels tend to be produced at the upper right edge whereas stressed vowels were produced with different F1 and F2 values depending on the vowel quality. A stress dependency could only be found for the factor roundness. The other factor groups showed nearly the same occurrences for stressed and unstressed vowels. During babbling, Hanna produced mainly low, mid–low vowels, and high vowels with a preference for front, unrounded vowels (Table 6.2). Once she produced meaningful speech, the production of more adult–like targets was necessary. In this stage more and more high vowels as well as mid–high vowels occurred and even back vowels were produced by Hanna. Until the age of 15 months, Hanna produced more unstressed rounded vowels than stressed rounded vowels. At 18 months of age, with the beginning of word production, an opposite pattern was produced. More stressed than unstressed rounded vowels occurred now. The vowel inventory of Hanna increased with age. The production of the cardinal vowels /i y a u/ began during the period of observation. During babbling only /a/ and /i/ from the cardinal vowels were produced, but not at the extreme limits of the F1 vs F2 vs F3 acoustic space. With increasing age also the other cardinal vowels occurred. But until the age of 36 months, Hanna did not always produce them at the extreme limits of the vowel space. While front vowels were preferred in the early babbling phase, more and more back vowels were produced with increasing age. The internal organisation of the stressed vowel system was not finished by 36 months of age.

With 36 months of age, Hanna produced all German vowels and the cardinal vowels were produced at the extreme limits of her vowel space. She mastered this production for /a/ at the lower limit of this space (highest F1) and located /u/ to the upper right corner (lowest F1, lowest F2). She produced /i/ and /y/ with high F2 but not with the expected lowest F1. In Appendix D.2 the area of the dispersion ellipses related to the vowels produced by Hanna is given. Neither the height factor nor the front–backness factor showed a significant effect on the area values of the dispersion ellipses of stressed or unstressed vowels, either as a main effect or in relation to age, until the age of 36 months. However the analysis revealed that the rounding factor had a significant effect on the area in the F1–F2 dimension for unstressed vowels ($F(1,20)=6.534$; $p<0.05$). The interaction between age and rounding in the F2–F3 dimension was significant for stressed vowels ($F(7,24)=4.404$; $p<0.05$). However, no age–dependent pattern were observed.

The area value of the dispersion ellipses was not dependent on the height dimension. Hanna produced high, mid–high, mid–low and low vowels with nearly the same variability, as the near–identical area values showed. It could be concluded that stressed vowel space differs significantly from the unstressed vowel space in the F1–F2 vowel space but not in the F2–F3 vowel space. Until the age of 36 months, Hanna was not able to produce vowels with stable F3–pattern vowel identity.

Nils

For Nils, the univariate ANOVA for the area of dispersion ellipses as variable showed no significant effect of stress, neither in the F1–F2 nor in the F2–F3 dimension. A developmental pattern could be observed as age showed a significant effect on both dimensions (F1–F2: $F(8,122)=2.597$; $p<0.05$; F2–F3: $F(8,124)=2.164$, $p<0.05$). The interaction between stress and age was significant in the F1–F2 vowel space

Table 6.3: Frequency of occurrence(%) according to vowel height, front–backness, and roundness for **Nils** depending on his age (in months).

stress	**feature**	**7,0**	**9,0**	**12,0**	**15,0**	**18,0**	**20,0**	**30,0**	**34,0**	**36,0**
unstressed	central	8,5	2,6		7,9	1,6	37,5	64,3	30,3	34,4
	high	20,0	13,8	4,8	37,1	17,9	25,0	14,3	22,6	20,0
	low	57,1	32,9	42,9	42,7	47,4		14,3	20,0	23,2
	mid–high	5,7	10,5		4,5	28,0	31,3	7,1	18,1	19,2
	mid–low	8,6	40,1	52,4	7,9	5,2	6,3		9,0	3,2
	back	8,6	2,0		2,3	33,3	37,5	3,6	25,8	24,8
	front	25,7	62,5	57,1	47,2	17,7	25,0	17,9	23,9	17,6
	rounded	2,9	2,0			1,2	6,3			
	unrounded	22,9	60,5	57,1	47,2	16,5	18,8	17,9	23,9	17,6
stressed	high	9,9	5,5	10,3	40,0	16,6	20,0	39,1	31,3	39,1
	low	58,0	40,9	51,3	39,4	65,9	33,3	26,1	35,8	28,7
	mid–high	2,5	10,0	10,3	11,9	12,6	13,3	26,1	19,4	17,2
	mid–low	24,7	39,1	28,2	6,3	4,9	26,7	8,7	13,4	14,9
	back	11,1		7,7	8,8	16,1	6,7	26,1	23,9	28,7
	front	25,9	54,5	41,0	49,4	18,0	53,3	47,9	40,3	42,5
	rounded	1,2	1,8		5,6	2,8	13,3	4,4	1,5	3,5
	unrounded	24,7	52,7	41,0	43,8	15,2	40,0	43,5	38,8	39,1

($F(7,122)=2.597$; $p<0.05$) but not in the F2–F3 vowel space. In Table 6.3, the frequency of occurrences of the vowels is given. With the beginning of babbling, Nils produced mainly front unrounded vowels and back vowels (in low frequency) until the age of 18 months. Stress influenced the vowel height, as Nils produced more high vowels in unstressed syllables than in stressed and more mid–low vowels in stressed than unstressed syllables. Nearly the same pattern occurred during the whole babbling phase. With the beginning of word production, a higher occurrence of high and mid–low vowels in stressed syllables was observed and increasingly more back vowels were produced. Nils hardly ever produced unstressed rounded vowels but he was able to produce rounded vowels. Stressed rounded vowels occurred nearly in all recordings. During babbling, Nils prpduced all possible German vowels. The frequency of occurrence differed with age as well as the dispersion of the ellipses in the vowel space. Nils' vowel inventory increased with age, as expected, and was completed at the beginning of meaningful speech. All possible German vowels occurred. However, even at the end of the observation period (32 months of age), no consistent patterns were produced. High variability in the area of dispersion ellipses as well as in their distribution in the vowel space were found. Nils mastered the production of the cardinal vowels /i y a u/ during the babbling phase. Until the age of 32 months, Nils did not always produce them at the extreme limits of the vowel space. While front vowels were preferred in the early babbling phase, more and more back vowels were produced with increasing age. An equal distribution of front and back vowels from the age of 20 months could be observed. The phonemic distribution in the vowel space was not yet finished at 32 months as no stable use of F1 nor F2 could be observed.

A developmental pattern for unstressed vowels can be observed (Appendix D.3).

Table 6.4: Frequency of occurrence (%) according to vowel height, front–backness, and roundness for **Bennie** depending on his age (in months).

stress	**feature**	**12,0**	**15,0**	**18,0**	**20,0**	**22,0**	**26,0**	**30,0**	**36,0**
unstressed	central	4,4		2,9	4,0	40,3	44,4	39,2	34,1
	high	8,7	12,5	17,7	32,0	8,2	10,4	17,7	22,5
	low	30,4	54,2	26,5	32,0	20,2	15,6	19,0	16,8
	mid–high	17,4	4,2	8,8		22,4	24,4	19,0	22,0
	mid–low	39,1	29,2	44,1	32,0	9,0	5,2	5,1	4,5
	back	26,1		23,5	16,0	16,4	17,8	21,5	28,4
	front	39,1	45,8	47,1	48,0	23,1	22,2	20,3	20,6
	rounded	13,0	8,3	5,9		0,8	0,7		0,2
	unrounded	26,1	37,5	41,2	48,0	22,4	21,5	20,3	20,4
stressed	high	4,8	3,8	5,0	7,5	23,0	40,6	35,4	30,6
	low	41,9	72,3	63,9	51,1	28,7	35,4	25,0	31,1
	mid–high	24,2	3,3	7,3	20,3	30,9	18,8	25,0	22,6
	mid–low	29,0	20,7	23,7	19,6	15,7	5,2	14,6	15,3
	back	24,2	8,2	17,8	20,3	36,5	22,9	31,3	26,0
	front	33,9	19,6	18,3	27,1	33,2	41,7	43,8	42,6
	rounded	11,3	4,9	2,7	0,8	2,3	1,0	2,1	0,9
	unrounded	22,6	14,7	15,5	26,3	30,9	40,6	41,7	41,7

An univariate ANOVA was significant for age in the F1–F2 dimension (height: $F(7,32)=2.640$; $p<0.05$; front–backness: $F(7,32)=2.568$; $p<0.05$; roundness: $F(7,19)=10.251$; $p<0.05$) as well as in the F2–F3 dimension (front–backness: $F(7,40)=2.768$; $p<0.05$; the interaction between age and roundness was significant $F(3,24)=3.967$; $p<0.05$). Stressed vowels did not show a developmental pattern. A significant effect of age on the factor front–backness of the F2–F3 dimension were observed ($F(8,41)=2.440$; $p<0.05$). The height factor showed a significant effect on the area value only in the F1–F2 dimension for stressed vowels ($F(3,40)=3.660$; $p<0.05$). The height factor did not have a significant effect on the area of the dispersion ellipses, neither as a main factor nor in interaction with the age factor. Similar results were found for the front–backness factor. However, the analyses revealed for unstressed vowels a significant effect of roundness on the area in the F1–F2 dimension ($F(1,19)=10.251$; $p<0.05$) as well as in the interaction between roundness and age ($F(2,19)=8.583$); $p<0.05$). In the F2–F3 dimension for unstressed vowels, age showed a significant influence on roundness ($F(3,24)=3.967$; $p<0.05$). Nils produced low vowels with the smallest area values, but no differences between high, mid–high and mid–low vowels according to their area values were observed.

Bennie

For Bennie, no significant influence either from stress, age or of the interaction of stress and age could be observed. He showed no preference for low and mid–low front vowels (Table 6.4). The first recordings were done at 12 months of age. He mastered the production of all German vowels with the beginning of meaningful speech with a preference for unstressed high vowels over stressed high vowels. For stressed vowels an early stability

Table 6.5: Frequency of occurrence (%) according to vowel height, front–backness, and roundness for **Rike** depending on her age.

stress	**feature**	**12,0**	**15,0**	**18,0**	**19**	**20,0**	**22,0**	**23,0**	**24,0**	**30,0**
unstressed	central	10,4	11,6	27,5	47,6	38,1	35,0	30,5	42,4	24,8
	high	20,6	28,4	14,5	14,3	19,1	20,0	21,0	18,6	28,1
	low	18,9	8,4	20,3	14,3	23,8	10,0	16,2	10,2	14,8
	mid–high	12,3	31,6	11,6	14,3	19,1	30,0	31,4	23,7	25,2
	mid–low	37,7	20,0	26,1	9,5		5,0	1,0	5,1	7,1
	back	7,6	7,4	2,9	9,5	14,3	15,0	25,7	28,8	34,3
	front	63,2	72,6	49,3	28,6	23,8	40,0	27,6	18,6	26,2
	rounded	12,3	4,2							1,0
	unrounded	50,9	68,4	49,3	28,6	23,8	40,0	27,6	18,6	25,2
stressed	high	29,0	19,9	52,1	38,1	63,2	53,3	42,9	25,3	39,7
	low	28,0	42,9	33,3	28,6	15,8	13,3	31,8	48,1	29,4
	mid–high	11,5	20,9	6,3	19,1	21,1	20,0	15,9	6,3	20,6
	mid–low	30,5	16,2	7,3	14,3		13,3	9,5	17,7	10,3
	back	16,5	14,7	26,0	38,1	26,3	13,3	22,2	15,2	27,8
	front	54,5	42,4	39,6	33,3	57,9	73,3	46,0	34,2	42,9
	rounded	9,0	5,2			10,5		4,8	3,8	4,8
	unrounded	45,5	37,2	39,6	33,3	47,4	73,3	41,3	30,4	38,1

for F1 could be observed, F2 was variable until the age of 22 months. With 36 months of age a more consistent production of F2 was observed. The produced formant values of F1 and F2 were nearly in the expected way. At the end of analysis, Bennie produced the cardinal vowels /i y a u/ at the extreme limits of the F1 vs. F2 vowel space. Unstressed vowels were produced with more variability in their phonemic distribution and more centralised.

A developmental pattern only for stressed vowels were observed (Appendix D.4). Age had a significant effect on the area of dispersion ellipses in the F1–F2 dimension ($F(7,25)=3.314$; $p<0.05$) as well as in the F2–F3 dimension ($F(7,24)=6.547$, $p<0.05$) on the factor roundness. The height factor did not have a significant effect on the area of the dispersion ellipses, neither as a main factor nor in interaction with the age. Similar results were found for the front–backness factor. For stressed vowels the roundness factor had a significant effect on the area in the F2–F3 dimension ($F(1,24)=9.913$; $p<0.05$) as well as in interaction with age ($F(7,24)=3.871$); $p<0.05$).

Rike

In Rike' case, no significant effect of stress was found. A significant main effect of age on the area for the F1–F2 vowel space ($F(8,157)=4.844$; $p<0.05$) as well as for the F2–F3 vowel space ($F(8,157)=2.732$; $p<0.05$) was found. A developmental pattern were observed as the area of the dispersion ellipses was larger at the beginning of the observation and tended to get smaller with increasing age. Rike showed no preference for low or mid–low front vowels. She produced all German vowels with nearly no differences in the frequency of occurrence for stressed and unstressed vowels (Table 6.5). The production of F1 for stressed vowels had been mastered at the beginning of observation period. But

Table 6.6: Frequency of occurrence (%) according to vowel height, front–backness, and roundness for **Emma** depending on her age (in months).

stress	feature	5,0	9,0	12,0	15,0	18,0	22,0	24,0	26,0
unstressed	central	19,4	24,8		11,8	6,5	23,6	31,4	24,7
	high	11,1	8,3	63,6	25,0	32,3	17,0	18,1	8,8
	low	19,4	27,8	9,1	22,1	16,1	30,9	26,7	33,0
	mid–high	13,9	7,5	9,1	8,8	22,6	9,7	11,4	12,8
	mid–low	36,1	31,6	18,2	32,4	22,6	18,8	12,4	20,7
	back	25,0	5,3	72,7	39,7	25,8	14,6	13,3	18,1
	front	36,1	42,1	18,2	26,5	51,6	30,9	28,6	24,2
	round	5,6	1,5		10,3	9,7	2,4	3,8	2,6
	unrounded	30,6	40,6	18,2	16,2	41,9	28,5	24,8	21,6
stressed	high	2,5	5,4	16,7	11,4	23,1	20,1	23,7	14,5
	low	37,5	49,4	23,3	37,7	31,3	44,4	53,2	53,9
	mid–high	17,5	10,7	16,7	16,6	20,9	9,2	7,7	13,2
	mid–low	42,5	34,5	43,3	34,3	24,6	26,4	15,4	18,4
	back	12,5	2,4	16,7	9,7	14,9	14,6	17,3	15,8
	front	50,0	48,2	60,0	52,6	53,7	41,0	29,5	30,3
	rounded	5,0	4,8	6,7	8,6	10,5	3,4	1,3	2,9
	unrounded	45,0	43,5	53,3	44,0	43,3	37,7	28,2	27,4

until the age of 23 months a great variability in the area of dispersion ellipses could be seen. She always tried to produce the cardinal vowels at the extreme limits of the vowel space and had mastered this for F1 as well as for F2 by 30 months of age. Unstressed vowels tended to be produced more centralised than stressed vowels. With increasing age the distribution of stressed and unstressed vowels showed separate distributions in the vowel space.

A developmental pattern were observed for unstressed as well as for stressed vowels (Appendix D.5). Age generally had a significant influence, although this did not hold true in all cases. The area of dispersion ellipses of unstressed vowels depended on the F1–F2 vowel space (height:$F(8,40)=2.575$; $p<0.05$) but also in the F2–F3 vowel space (front–backness: $F(8,42)=2.421$; $p<0.05$; roundness: $F(8,29)=2.446$; $p<0.05$). For stressed vowels age showed a significant effect in the F1–F2 dimension (height: $F(8,56)=2.353$; $p<0.05$; front: $F(8,64)=2.439$; $p<0.05$). For the height factor no significant effect was found. Similar for the factor rounding. For Rike, the front–backness factor had a significant effect on the area of stressed vowels in the F2–F3 vowel space ($F(1,64)=8.923$; $p<0.05$), back vowels being mainly associated with larger ellipses than front vowels.

Emma

In Emma's case, no significant effect of stress or age on area was found. With the beginning of babbling, Emma preferred to produce low and mid–low front stressed vowels (Table 6.6). In the unstressed syllables, front and back vowels were produced with nearly the same frequency. When she began producing meaningful speech, the production scheme changed and no preference could be observed any longer. Even though

Table 6.7: Frequency of occurrence (%) according to vowel height, front–backness, and roundness for **Ollie** depending on his age (in months).

stress	feature	9,0	12,0	15,0	18,0	33,0
unstressed	central	4,7	21,6	18,4	7,7	33,3
	high	4,7	2,0	21,1	19,2	6,7
	low	18,6	15,7	15,8	19,2	22,2
	mid–high	20,9	19,6	23,7	26,9	28,9
	mid–low	51,2	41,2	21,1	26,9	8,9
	back			18,4	38,5	26,7
	front	76,7	62,8	47,5	34,6	17,8
	rounded	2,3	2,0	7,9	3,9	
	unrounded	74,4	60,8	39,5	30,8	17,8
stressed	high	11,4	6,6	7,1	10,3	24,3
	low	48,6	30,2	28,6	37,9	27,0
	mid–high	17,1	29,0	27,4	13,8	32,4
	mid–low	22,9	34,2	36,9	37,9	16,2
	back		5,3	13,1	12,1	29,7
	front	51,4	64,5	58,3	50,0	43,2
		48,6	35,5	41,7	50,0	56,8
	rounded	2,9	4,0	3,6	6,9	2,7
	unrounded	48,6	60,5	54,8	43,1	40,5

all German vowels were produced already from the age of 15 months, no stable use for either F1 or F2 could be found up to the age of 26 months. Emma still produced the different vowels with great variability as could be seen in different areas of the dispersion ellipses as well as in different phonemic distributions of the vowel space. Stressed and unstressed ellipses differ in their dispersion in the vowel space but unstressed vowels showed no centralisation as expected. Only in interaction with age, height showed a significant effect on the ellipses of stressed vowels in the F1–F2 dimension ($F(21,60)=1.802$; $p<0.05$). The area of high and low vowels decrease with increasing age (Appendix D.6). For front–backness, no significant influence could be found, neither as main factor nor in interaction with age. The rounding factor showed no significant effect as a single factor, however, in interaction with age, a significant effect on the ellipses could be found for stressed vowels in the F2–F3 dimension ($F(7,40)=3.308$; $p<0.05$). Rounded vowels were mainly associated with larger ellipses in the F2–F3 vowel space than unrounded vowels.

Ollie

For Ollie, the univariate ANOVA for the area value showed a significant effect of stress in the F1–F2 vowel space ($F(1,78)=6.085$; $p<0.05$). Age and the interaction between stress and age were not significant. For the F2–F3 dimension no main effect was significant. Ollie produced stressed and unstressed vowels with differences in the variability, independent of age. He produced stressed vowels with larger area values of the dispersion ellipses and therefore with greater variability than unstressed vowels. The variability in the production of stressed and unstressed vowels increases with age as could be seen by larger ellipses, but not significantly. In the F2–F3 vowel space, no developmental pattern could be observed as they were produced with too much variability for significant results. During babbling, Ollie produced mainly low and mid–low front vowels in stressed as well as in unstressed syllables (Table 6.7). When first words occurred, more back vowels were produced. He produced rounded vowels and showed a clear preference for unrounded vowels during the whole observation period. The vowel inventory have been acquired at about 15 months of age. At this age, F1 was produced more stable. With 33 months of age, also F2 was produced stable. The cardinal vowels /i y a u/ were attempted at the extreme limits of the F1 vs. F2 vowel space. Univariate ANOVAs were carried out on the area values with the features and age as factors. The height feature did not had a significant effect on the area of the dispersion ellipses, neither as a main effect nor in interaction with age. The front–backness factor showed a significant effect on the ellipses for unstressed vowels in the F1–F2 dimension ($F(1,25)=6.271$; $p<0.05$) (Appendix D.7). The front vowels were associated with larger ellipses than back vowels. The roundness factor showed no significant effect neither as a main effect nor in interaction with age in the F2–F3 dimension. In the F1–F2 dimension a significant effect of the roundness factor in interaction with age on unstressed vowels could be found ($F(2,15)=6.000$; $p< 0.05$).

6.4 Summary

The first aim of this study was to analyse whether stressed and unstressed vowels differ in their distribution in the vowel space. Therefore, stressed and unstressed vowels of different speakers from 5 months up to adulthood were plotted as dispersion ellipses and their distributions were compared throughout their aligned vowel space. Stressed and unstressed vowels show different dimensions of the dispersion ellipses as well as a different expansion in the vowel space. During the babbling stage, no clear pattern in the vowel distribution can be observed. At the beginning of word production, the stressed vowels expanded in the vowel space whereas unstressed vowel production became more and more centralised in the F1 vs. F2 vowel space. The observed acoustic values vary among the speakers as the distribution of the dispersion ellipses differ. These findings show that the internal organisation of the vowel spaces produced by different speakers between 5 months and adulthood can be characterised by great variability. Even at the beginning of babbling, stressed and unstressed vowels have different distributions in the vowel space. Different acoustic representations of stressed and unstressed vowels have to be built up. During development two different vowel spaces, one for stressed and one for unstressed vowels have to be developed. The stressed vowels shared a vowel space different from that of unstressed vowels, which tend to be more centralised in vowel space.

A second aim was to describe the development of the produced vowel spaces. The infant's vowel space expands towards the edges of the adults vowel space with age. Depending on the shorter vocal tract length (Vorperian and Kent, 2007), an infant's vowel distribution shows higher F1, F2, and F3 values and a gradual decrease in formant frequencies with age. The formant–frequency variability decreases with age and stable productions of F1 can mainly be observed at the beginning of word production. The differences between stressed and unstressed vowel space can be described best by comparison to the cardinal vowels. With stressed syllables, all speakers drift to produce the cardinal vowels /i/, /y/, /u/ and /a/ at the extreme limits of the F1 vs. F2 vs. F3 vowel acoustic space with increasing age. Mainly, /i/ takes the upper left corner of the vowel space (lowest F1, highest F3), /u/ corresponds to the upper right corner (lowest F1, lowest F2), /a/ is produced at the lower limit of the vowel space with highest F1, and /y/ is located at the lower left corner of the F2 vs. F3. The sizes of the dispersion ellipses decrease with age. For all three features (height, front–backness, roundness), the dispersion ellipses were significantly broader for the children up to the age of 10 years. The adults showed smaller area values and therefore less variability in the vowel space.

For the longitudinal data, a high variability in the area of the dispersion ellipses could be observed up to the age of three years due to the immaturity of motor control. Detailed analyses of the longitudinal data of the six children from 5 up to 36 months of age showed individual differences in the developmental process after 24 months of age, particularly in the developmental speed in the expansion of the vowel space. The infants had gone through the babbling period and now produced mainly meaningful words. The developmental speed seemed to be influenced by the parents' production of the vowel space (Kemler Nelson et al., 1989). In the parents' data, the dispersion in the vowel spaces also shows inter–individual variability. Children develop their own vowel space according to their individual environment. This can be confirmed by our data of parents compared to the children. A correlation in production performance between parents and their children were observed. Children can produce clearly separated vowel classes earlier if their parents have a well–defined vowel space themselves. These findings are supported by results that the mother's vowel space area is significantly correlated with an infant's speech discrimination performance (Liu et al., 2003).

6.5 Discussion

Despite the inter–speaker and intra–speaker variability, some patterns in the development of vowel space can be observed. Apart from the increase of the vowel inventory with age, the characteristics of vowels changed over time. We observed a gradual change in the infants' formant values towards values that are typical for an adult speaker. This process started during the babbling phase and had not yet been finished by the age of 36 months. These results, as expected, are caused by developmental changes in the infants' abilities and anatomy. The large overlaps between the ellipses become smaller with age due to the spread of the vowels and reduced variability. Thus the observed changes in the infants' vowel production may originate from the infants' increasing motor abilities as well as from the perception–production loop and the necessity to develop more adult–like representation of vowels.

Based on the non–uniform vocal tract growth (Fant, 1960; Vorperian et al., 2005) and the fact that the vocal tract morphology does not prevent the infant from achieving the perceptual targets representing the native vowels (Ménard et al., 2004), the children had to improve their articulatory abilities. During babbling, no template has to be followed and the produced vowels are randomly produced and distributed. Their articulatory abilities have to be developed. While listening to their own articulation, they learn to move their articulatory organs in a certain way and to thereby produce different sound changes. Through outward perception and self–perception, the produced sounds are linked with each other (Vihman, 1996). At the beginning of word production, these linked production schemes are compared with the target production via self–feedback as well as compared with the underlying representation. Therefore in order to reach the target, different modifications of the articulatory system have to be done by the children to establish more language–appropriate acoustic representations.

When babbling begins, the poor control of the jaw, tongue, and lip articulators causes the production of mainly open front and central vowels (Davis and MacNeilage, 1995). The jaw opening, which constricts the vocal tract toward the glottal end and expands it toward the lip end, is the deciding factor for the first formant. This formant frequency rises as the jaw is opened wider. The second formant is most sensitive to the shape of the body of the tongue, and the third formant is most sensitive to the tip of the tongue. Lip rounding, also influence F3, exaggerates the acoustic effect of backness, and helps make the back vowels more easily distinguishable from front vowels.

With increase control over jaw and tongue infants are able to add articulatory strategies to their acoustic vowel representations. They learn to produce a larger vowel space with greater variability in the acoustic space as can be seen in different extensions of the dispersion ellipses. The most frequently produced vowels during babbling were perceived as mid–low, low, front and central vowels. With increasing motor–control, more high and back vowels were produced. For the production of rounded vowels, the motor control of the lip articulators is imperative. Lip rounding is achieved by the children very individually as they increase in age.

These results suggest a developmental expansion of the vowel space. During babbling, the child produces the most frequent syllable patterns of the native language, constrained by its motor skills (MacNeilage, 1998). An increase of the acoustic variability is associated with the ability to retract the tongue. With the production of first words the articulation improves and the control over the tongue as an articulator for vowel production (Fikkert and Levelt, 2008) increases and therefore more adult–like productions occur depending on the infant's discrimination performance, as a well–defined vowel space in the input of the mother or father seems to assist infants in discovering the appropriate acoustic cues (Kemler Nelson et al., 1989; Liu et al., 2003). Based on self–feedback as well as feedback from the environment a more detailed representation can be built up.

Based on the production–perception loop a cognitive map that codes this information is developed. During the emergence of speech, children have to compare their own signals by self–feedback with the adult target to establish appropriate articulatory settings for the production of a good vowel exemplar depending on the acoustic input. Each production is by self–perception also mapped to the child's representation of the vowel. During the non–uniform vocal tract growth, in order to reach a common perceptual goal, children have to recalibrate the articulatory–to–acoustic relations of their vocal tract by improving motor control.

6.6 Conclusion

Babbling is one key mechanism that permits infants to discover and to produce native language patterns. At the beginning of meaningful speech, the production patterns practised during the babbling phase are checked against the perceived target and the production improves with the growing number and quality of input elements. The development of a categorically separated vowel space requires the development of articulatory skills and is not limited by the vocal tract morphology (Ménard et al., 2004) but constrained by the infant's motor abilities. The development of articulatory skills begins in the early stages of an infant's speech development and the developmental progression in the F1–F2 and F2–F3 pattern are results of the anatomical growth of the vocal tract (Vorperian et al., 2005), the refinement of speech motor control and the establishment of internal representations for the vocal tract configurations for the vowels. When speech begins, infants tend to imitate the vowels they hear by using their articulatory skills (Kuhl and Melztzoff, 1996).

With increased speech feedback from the auditory system, more examples can be stored in the cognitive map (Goldinger, 1997; Pierrehumbert, 2002). Speech feedback as well as self–monitoring are sufficient for the development of language–specific representations of vowels. Depending on the input mainly given by the parents an inter–personal influence on the development of representations and on the developmental speed can be assumed and be reflected in the variable developmental patterns of the six longitudinal observed infants.

7 General discussion

In this general discussion, the main findings of the various experiments are summarised and the results are integrated in a neural model of speech production and speech acquisition following the ideas of Guenther's neural model of sensorimotor control of speech production (Guenther, 1994, 2003). The main focus lies on the differences between babbling and word production and how the neural network is developed.

7.1 Introduction

The various experiments, described in detail in the preceding chapters, were meant to describe the development of stress and of a mental syllabary for children learning German. The knowledge of this process as a whole still remains incomplete, but the various parts of this study provide important information about the development of word stress in children's speech production. In this concluding chapter, mainly the differences in the production of word stress in babbling and words are discussed. Then, some implications for a developmental model during the first three years of life are described.

Long before children are able to produce first linguistic sounds, they are already able to perceive language-specific patterns (Jusczyk et al., 1994; Christophe et al., 2003b; Echols and Marti, 2004; Maye et al., 2008; Nazzi et al., 2008). With these perceived patterns, they build up prosody-based rules to bootstrap the speech signal (Jusczyk, 1996; Demuth, 1996b; Ramus et al., 1999; Demuth, 2001; Echols, 2001; Fisher and Church, 2001; Höhle et al., 2001). These prosody-based rules build up the phonological encoding in the mental lexicon (Levelt et al., 1999b). Long time it was fairly accepted in the acquisition literature that the underlying representation of a word is quite close to the adult/target form (Smith, 1973; Kehoe and Stoel-Gammon, 1997). But this view is less and less accepted and clearly challenged by Fikkert and Levelt (2008). Lexical representations are not adult-like from the start but rather depending on the child's ability of segmentation (Goad and Rose, 2004; Fikkert and Levelt, 2008).

Growing evidence suggests that context-sensitive acoustic-phonetic details are part of these representations (cf. Curtin, 2002; Pierrehumbert, 2003a; Swingley, 2003; Burns et al., 2007; Shukla et al., 2007; Thiessen and Saffran, 2007; Zamuner, 2009). It has been shown that stress information in the native language not only shapes the representational landscape used by infants in segmentation, but that this information is also encoded in the representations of parsed sequences (Lindfield et al., 1999; Saffran et al., 1999; Curtin, 2002; Thiessen and Saffran, 2004; Curtin et al., 2005). Based on this assumption, the representations are not maximally underspecified as it is assumed in an abstractionist view (cf. Chomsky and Halle, 1968; Werker, 2000). Context-sensitive fine phonetic-acoustic details are part of the representations. In order to learn to speak, children have to develop this enriched representation from observed examples. Furthermore, a lag between perception and production during acquisition can be observed. Children do not immediately produce everything they perceive, their production of speech sounds depends on their articulatory abilities, which are practised during the babbling phase. While babbling, children learn to control their articulatory organs in a certain way to produce different sounds. When word production begins, these trained articulatory patterns have to be connected to phonological encoding to produce words.

And what role does prosody play in production? It is assumed that the perceived representations are enriched and stressed and unstressed syllables are represented in different ways in the infants' perceptual space (Curtin et al., 2005). Therefore, acoustic cues of stress are also stored in this perceptual space. However, with the beginning of first pre-linguistic utterances the stored information cannot be implemented as the articulatory abilities of the infants first have to be established. The question arises as to how prosodic structure is developed in parallel to the phonemes and at which age the prosodic and syllabic structure of the ambient language can be produced with constant acoustic parameters. The phonological development has been described in detail by, e.g. Smith (1973); Vihman (1996); Stoel-Gammon (1998); de Boysson-Bardies (1999) and others. There are also several works describing the development of prosodic structure either focused on word development (Allen and Hawkins, 1980; Fikkert, 1994; Demuth and Fee, 1995; Kehoe and Stoel-Gammon, 1997; Grimm, 2008) or babbling (Vihman et al., 1998; Davis et al., 2000; DePaolis et al., 2008). However, the development of the production of the different possible prosodic correlates of stress in babbling and words in a longitudinal study has not yet been examined.

Therefore, the focus of this thesis is on comparing the development of prosodic structure as well as syllable structure depending on word stress in babbling and words and to describe the phonological development as well as phonetic development of stress. In addition, to describe the development of language specific probability distributions over the phonetic space, the development of vowels is also described. The theoretical focus of the thesis thus ultimately lies on the development of an example-based model of the acquisition of prosodic structure. In the next sections, the results are represented in a neural model of speech production. The conducted model is divided in the production process during babbling and during word production and is based on a comprehensive neural model of speech production using self-organising neural network maps (Kröger et al., 2008, 2009). A central property in the presented model is the co-activation of the phonemic, sensory, and motor plan of a syllable with the phonetic map. As this model comprises a production and a perception part, the switch from prosodic productions between babbling and the word stage can be discussed in light of this model.

7.2 Neural model of speech production

It is assumed that in order to produce meaningful speech, a mental lexicon including concepts, lemmas, and the phonological codes of lexical items is a central component for successful production (Levelt and Wheeldon, 1994; Levelt et al., 1999b; Schiller et al., 2006). Furthermore, it is assumed that the phonological codes, the underlying motor plans as well as the sensory representations of frequent syllables are stored within the mental syllabary. In speech production the mental lexicon selects the necessary syllabification. Frequent syllables can be produced directly via the stored representations in the mental syllabary, infrequent syllables have to be produced with stored representations of each unique phoneme.

Based on this model of speech production and following the ideas of Guenther and colleagues (e.g. Perkell et al., 2000; Guenther, 2001, 2003; Guenther and Perkell, 2004) that the production part is divided into feed-forward and feedback control, Kröger et al. (2008, 2009) developed a comprehensive neural model of speech production using self-organising neural network maps (Figure 7.1). This model can be used to describe the different speech acquisition stages in a logical way. The principal aspect of this model is self-organising maps (so called SOM's) which model the relations between the phonemic, sensory, and motor levels. A central property of this model is the co-activation of the phonemic, sensory, and motor plans of a syllable with the phonetic map.

When produced, two different paths can be selected. The phonetic map in this model corresponds to the idea of a mental syllabary by Levelt and Wheeldon (1994); Levelt et al. (1999b); Schiller et al. (2006). For frequent syllables an activation of the phonetic map including the neural activation of the appropriate sensory state as well as the appropriate motor plan state is assumed. Therefore, the auditory, somatosensory, phonemic, and motor planning information for each frequent speech item is brought together in one map. For infrequent syllables, another production path using the motor planning module and sub-syllabic units (Varley and Whiteside, 2001) is activated.

During speech acquisition, the mental lexicon as well as the mental syllabary has to be built up. With the beginning of perception infants first only store the perceived exemplars in the memory. In an exemplar-based theory of speech acquisition each stored phonological representation is not abstract, but includes context-sensitive fine phonetic-acoustic details and is acquired bottom-up from the acoustic signal by means of statistical learning procedures (Pierrehumbert, 2003a,b). These details are part of the representations, which form the basis of the mental lexicon necessary for word production. During first production of pre-linguistic sounds a *Protosyllabary* is built up (Levelt et al., 1999b) containing proto-word forms sharing four sequential sound patterns (three CV and one CVC sequences) which may form the basic to the emergence of words (MacNeilage and Davis, 2000).

The production of different babbling phrases activates a mapping of auditory-phonetic processing to articulatory-acoustic gestures through self-feedback. The main task during babbling is to associate the auditive and the somatosensory outcome of a random

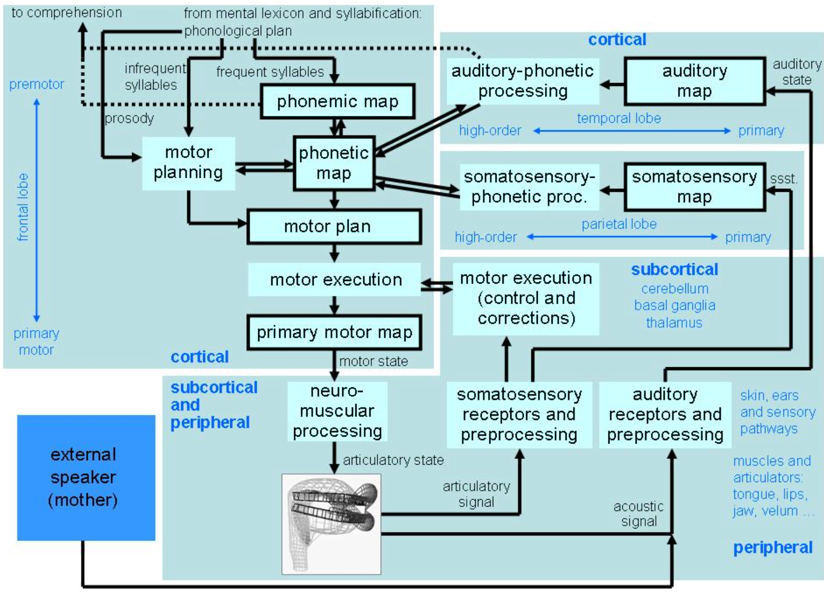


Figure 7.1: Organisation of the neurocomputational model after Kröger et al. (2008, 2009). Boxes with black outline represent neural maps. Arrows indicate processing paths or neural mappings. Boxes without outline indicate processing modules.

vocal tract motor activity and to form the sensorimotor mappings from the phonetic map (Figure 7.2). At the beginning of the imitation of words or production of words, the basic sensorimotor knowledge gained during babbling is activated but a knowledge of phonological categories of sounds and of the meaning of words is also needed. This knowledge is stored in the mental lexicon where neural connections between acoustic realisation of a word and its meaning are built up during interaction with care-givers. For a successful acquisition, human-human interaction is important to associate the meaning and phonetic word forms (Kröger et al., 2009). In the next section the results of the phonological as well as the phonetic development of stress are integrated in this neural model of speech production and perception

7.3 Model of the development of word stress

Even though children at the beginning of the babbling stage and therefore about 5 months old are able to produce perceivable different stress patterns, no restrictions either of the metrical pattern (see Section 4.3) nor of the syllable pattern (Section 4.4) can be observed for German children. Furthermore, the implementation and usage of the acoustic parameters to mark word stress (Appendix C) are used inconsistently as each measurable parameter is produced variably by the children. It appears that during babbling, the produced syllable types do not depend on stress and that the different metrical patterns are produced at random. Babble phrases show variable stress pattern with nearly the same dispersion during the whole babbling phase. The production of stressed syllables seems to be the unmarked case as only about 30% of all produced babble syllables are perceived as unstressed. Stressed syllables are produced most frequently, followed by syllables perceived as less prominent but not as weak (level stress).

This can also be observed in the production of the vowel space (Section 6.3.1). During babbling, the vowels are randomly produced as no target has to be reached. Although stressed and unstressed vowels show different dimensions of the dispersion ellipses as well as a different expansion in the vowel space, no clear pattern in the vowel distribution, for either stressed or unstressed vowels can be observed for babbled vowels at the beginning of the babbling phase. As neither vowel height, front-backness or roundness show significant differences at this stage, the production of vowels are based on the corresponding consonants and their tongue movement and the biomechanical properties of the vocal apparatus. Based on gradual convergence towards the target system babbling permits infants to discover and to produce more and more language-specific patterns and phonemes. The different motor speech patterns are stored as examples in the protosyllabary. Therefore, babbling is training for the articulation and their underlying articulatory movements. At this stage, external input (red arrows in Figure 7.2) activates the auditory-phonetic processing through the auditory processing and auditory map (right of mental syllabary in Figure 7.2). As part of this process, the phonological encoding already built up in the mental lexicon helps to bootstrap the speech signal and enriched representations of the perceived language-specific structures are stored in the mental syllabary (Figure 7.2, red area). The production of babbles also activates the auditory processing and the connected auditory map by self-feedback (blue arrows in Figure 7.2). Each self-produced babble is therefore also stored as an enriched representation in the mental syllabary. In the case of production, not only the perceived representation is stored but the corresponding somato-sensory map and articulatory-acoustic gestures are also part of the enriched representation (blue area in Figure 7.2). The connection of enriched representations and associated articulatory-acoustic gestures are the basis for the development of a mental syllabary. Even during babbling no word template has to be followed, the children are edging towards more and more language-specific phonotactics. External feedback only serves to activate the auditory path and build up enriched representations and phonological encoding necessary for bootstrapping the sig-

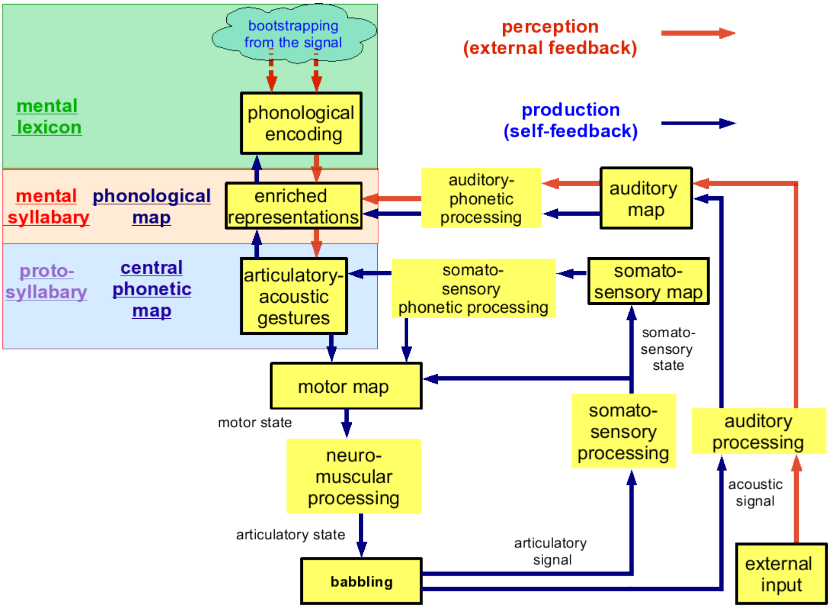


Figure 7.2: Neural speech production model during **babbling** (based on Kröger et al. (2008)). Boxes with blue outliers represent activated neural maps for production, blue arrows indicate processing paths or neural mappings during production. Boxes with red outliers represent activated neural maps for perception, red arrows indicated processing paths or neural mappings for perception. Boxes without outline indicate processing modules.

nal (red arrows in Figure 7.2) and to discover the language-specific phonotactic patterns. However, the representations built up from external input have no connection to corresponding articulatory-acoustic gestures as these connections can only be built up in connection with the auditory and somatosensory paths. In the early stages of babbling, the syllables produced depend more on the rhythmic mandibular cycle and the resulting consonant-vowel alternations than on underlying targets. According to MacNeilage and Davis (2000), there are four underlying units of the protosyllabary, where central vowels mainly co-occur with labial consonants, front vowels with coronal and back vowels with dorsal consonants. The fourth pattern describes the labial consonant-vowel-coronal consonant sequence.

In this phase, a neural network of auditory-articulatory gestures has to be built and clusters of different gestural positions are stored in the protosyllabary. With increasing age, reinforcement from the environment and more experience in the production of prelinguistic sounds, the influence of the phonological knowledge stored in the phonological encoding is integrated in the babbles produced. The inventory of babbled vowels and consonants assimilates to the speech inventory of the ambient language and depends no longer on the rhythmic mandibular cycle (characterised with dashed blue arrows from phonological encoding to enriched representations in Figure 7.2).

Between 12 and 20 months of age, babbles as well as an increasing quantity of words are produced in parallel. At this stage, the differences in the organisation and control in the production of babbles and words can be observed most clearly. Utterances marked as babbles still show no restrictions in metrical pattern. In word production, a bias can be observed for stress on the first syllable for German learning children, as mainly trochaic stressed disyllabic and antepenultimate stressed trisyllabic words are produced by our children at the first word stage (see Section 4.3). The other word stress patterns still have to be developed for words even if a great variety of metrical stress patterns can be observed for babbled phrases at the same time. However, the production of different syllable structures still shows no restriction on the production of babbles. In the production of words, a developmental order for the different syllable structures can be drawn out (Section 4.4). The development of syllable structures depends on stress. Complex codas are relevant for stress placement in German. Complex onsets are universally irrelevant for stress (Goedemans, 1998). The stressed syllable structure has to be acquired first before the unstressed counterpart can be produced.

The developmental order of the different syllable structures corresponds relatively well to the frequency of syllable structures for German. The first words which are produced show the most frequently occurring syllable structures CV and CVC. Then, two different developmental steps can be observed for the acquisition of V, VC, CCV, and CCVC. This could be due to the fact that the four syllable structures have nearly the same frequency of occurrence. The main developmental order for German is that first V and VC is acquired, followed by complex onset clusters. Only one boy acquired complex onset cluster before V and VC. Complex coda, which is less frequent in German, is acquired last in the developmental order.

When words begin to be produced, different production schemes can be observed as the phonological encoding has to be established to manage the various syllable structures and metrical patterns (green area in Figure 7.3). Therefore, when the child begins to actually produce words, an infant has to connect the underlying articulatory-acoustic gestures trained in the babbling stage and stored in the proto-syllabary to the meaning of an object and to build up a mental syllabary with a phonological and central phonetic map (red area in Figure 7.3). To improve the output, self-feedback (blue arrows in Figure 7.3) as well as feedback from the environment (red arrows in Figure 7.3) are now necessary. At this stage, the control over the articulators also increases and the produced vowel space expands (Chapter 6). A more detailed representation of the vowel

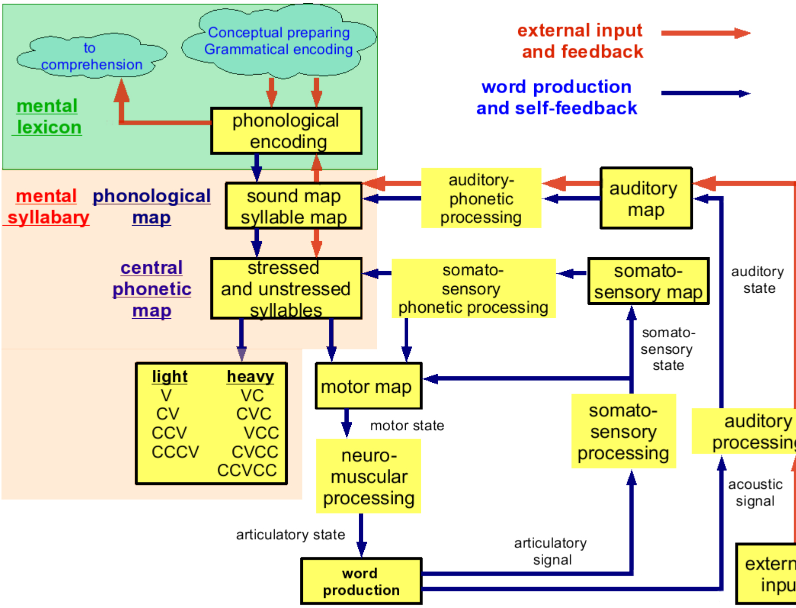


Figure 7.3: Neural speech production model during **word production**. Boxes with blue outliers represent activated neural maps for production, blue arrows indicate processing paths or neural mappings during production. Boxes with red outliers represent activated neural maps for perception, red arrows indicated processing paths or neural mappings for perception. Boxes without outline indicate processing modules.

space can be built up. Stressed vowels expand in the vowel space and production of the cardinal vowels /i y a u/ is attempted at the extreme limits of the F1 vs. F2 vowel space. In contrast, unstressed vowels tend to be produced more centralised in the vowel space and no extreme limits in the production of unstressed vowels can be observed. With increasing control over jaw and tongue, children are able to add articulatory strategies to their acoustic vowel representations. This can be seen by the production of larger vowel spaces as they get older. The older the child, the greater the flexibility in the acoustic space. The dispersion ellipses of the produced vowels show different extensions and reflect the parental vowel spaces increasingly. However, the vowel space continues expanding until the age of ten.

With increasing age the children improve the production of perceivable differences in the acoustic parameters to mark stress. A greater variety of metrical stress patterns occur and the production of complex onsets and codas increases. The prosodic structure of the ambient language in production is developed at about three years of age, the acoustic realisation of stress is still inconsistent, although duration is the most reliable parameter to mark stress for the older children as well as for the parents. Furthermore, RMS Intensity, and the voice quality parameters *Glottal Opening*, *Skewness*, *Rate of Closure*, and *Completeness of Closure* demonstrate parent-like usage patterns as the child grows in age.

The acoustic realisation of word stress is still inconsistent and variable up to the age of 36 months. The children still have to gain motor control over their articulation for a more adult-like use of the acoustic parameters (Appendix C).

The production of their first words is mainly the result of imitation of some targets given by the external input. At this stage, not only self-feedback like during babbling but also feedback from the environment plays an increasingly important role. With the beginning of word production, the language-specific syllabic structure plays an important role in production as a target has to be reached (phonological map, red area in Figure 7.3). Now the phonological encoding (green area in Figure 7.3), based on grammatical encoding and conceptual preparation, has to be fully integrated into the selection of specific syllable structures and acoustic realisations stored in the mental syllabary (red area in Figure 7.3) to reach a target word. When word production begins, infants have to connect the underlying articulatory-acoustic gestures trained during babbling and stored in the proto-syllabary to the meaning of an object. To improve their output, self-feedback as well as feedback from the environment is now necessary. The language-specific syllabic structure is now integrated in word productions and has to be built up from the one foot template. In the first word production the target words are always reduced to one foot and the stressed syllable remains preserved in the truncations.

With increasing age, the initial stage of one foot is expanded. The optimal stage is reached at about two years of age. At this stage, target-like prosodic shapes are produced with complex onsets and codas, stored in the central phonetic map of the mental syllabary. Therefore, with more and more experience, self-feedback and feedback from the environment the production gets better and more adult-like. The right syllable structures as well as metrical patterns have to be selected from the central phonetic map and the acoustic production improved. First the stressed syllable structure is produced. For the unstressed version, the reduction of the acoustic correlates of stress has to be managed in such a way as to produce a perceivable difference between stressed and unstressed syllables.

As most of the syllables during babbling are perceived as stressed and not as weak, the unmarked case is the production of stressed syllables. The development of the acoustic parameters of stress can therefore be characterised as involving increased control over the articulatory movements to derive perceptually identifiable stress contrasts. The most important cue to mark perceivable stress patterns is to learn to reduce the acoustic parameters. During acquisition, children have to learn to fine tune the different acoustic parameters to mark word stress. The unique phonetic target region for each phoneme in the stressed and unstressed condition has to be established.

This is in line with the assumption that one of the requirements of learning to talk is learning to produce *un*stressed syllables. Based on perceptual constraints (Allen and Hawkins, 1980), the metrical account of the ambient language influences the development of prosodic structure. In this case, a trochaic bias is assumed in the production of German learning children. In a motor-based perspective, babbling could be described as a sequence of stressed syllables, and suggests that getting the unstressed syllables into the sequence is what may be difficult (MacNeilage and Davis, 1990; MacNeilage, 1998).

Although these specific arguments differ, both give strong evidence for perceptual contributions to the form of early words. Perceptual salience would serve only to indicate which elements are most important for inclusion in a production. The children's earliest representations would include all of the segmental material of the adult target (Echols, 1993) and at the beginning of word production they would try to reach this adult target.

Prosody permits the infants to detect the phonetic and phonotactic structures of the ambient language. The perception of language-specific prosodic cues begins during maturation as the amniotic fluid preserves prosodic characteristics like the rhythmic properties of the ambient language. The infants build up language-specific representations and store them as enriched representations in the mental syllabary. At the beginning of production, no influence of the ambient language can be observed at first. However, with increasing age, a preference for a more language-specific production can be observed. The produced articulatory-acoustic gestures – stored in the proto-syllabary– are more and more connected to the perceived enriched representations of the mental syllabary and match to the perceived representations regulated by improved motor abilities. At the beginning of word production, the articulatory-acoustic gestures formerly stored in the proto-syllabary, are now connected to the language-specific central phonetic map and integrated into the motor map. A mental syllabary and mental lexicon is developed (Levelt and Wheeldon, 1994; Levelt et al., 1999b; Cholin et al., 2006). Therefore, learning to speak requires the development of a well-distributed language-specific neural network. Once this system is established, the network is supposed be the hard-wired part of phonology (Dogil, 2010) as it is preserved in pathological speech. On the other hand, when this system cannot be created, or problems occur during the development of this neural network, phonological deficits like developmental dyslexia emerge (van Alphen et al., 2007; de Bree et al., 2008).

The introduced neural model of the acquisition of stress in German is based on the implemented computational model of Kröger et al. (2008). Nevertheless, a computational model, based on the presented neural model, which interprets the variability in the acquisition of stress as a function of the temporal development was not the focus of this thesis but should be implemented in future research.

8 Summary

In this thesis several aspects of the phonological as well as the phonetic development of stress of German children are described. The most important research question is: how does the production of word stress develop in the speech of children, and, above all, can it be described with an exemplar-based model?

First, based on the assumption that even infants are able to encode the salient properties of the speech signal and that stress information does not only shape the representational landscape but that this information is also encoded in the representations of parsed sequences, an exemplar-based model of speech production is introduced in Chapter 2. The development of exemplars is embedded in a neural network of auditory-articulatory mapping as well as in a neural model of speech production and perception.

Chapter 3 describes how the speech samples of the present study were recorded and annotated. In order to process the large number of speech samples, different automatic analysis tools are introduced and evaluated for child speech. As the main focus of this thesis is to describe the development of stress between 5 and 36 months of age, mainly longitudinal data from six children are used to describe the phonological as well as phonetic development of stress. To be able to compare the child's use of the acoustic parameters, the parental use of the acoustic parameter of stress was also annotated and analysed. Furthermore, in order to describe the vowel development, cross-sectional data for children between 4 and 10 years were evaluated.

In Chapter 4 the phonological development of stress between 5 and 36 months of age is described. Therefore, both the development of metrical pattern and the development of different syllable types depending on stress are analysed. The results show that during babbling, no restriction exists on either the production of different metrical stress patterns or the different syllable structures. With increasing age, when babbles as well as words are produced, the picture changes for babbles and words. For babble phrases, a great variety of metrical stress patterns and syllable structures can still be observed. The production of first words is restricted to a metrical pattern with stress on the first syllable. Furthermore, the production of different syllable structures is restricted to CV and CVC syllables. The developmental order of stressed and unstressed syllable structure corresponds well with the frequency of the different syllable structures for German.

In Chapter 5, the phonetic development of stress and the acoustic cues of German word stress (vowel length, loudness, pitch, formants, voice quality) are analysed for children between 5 and 36 months of age and their parents. The children use different strategies to produce stress. All acoustic parameters are used to mark stress. During babbling, the most reliable correlate of word stress were intensity and f_0. With increasing age, the children show tendencies to use all the stress cues reported for adults, as there are vowel duration, intensity, and the voice quality parameters *Glottal Opening*, *Skewness*, *Rate of Closure*, and *Completeness of Closure* in the parent-like usage pattern, but not consistently.

For the adult use of the parameters, it can be concluded that vowel duration as well as fundamental frequency and the spectral tilt parameters *Skewness* and *Rate of Closure* are used to differentiate between stressed and unstressed syllables. Stressed vowels are produced with longer vowel duration, higher fundamental frequency and with faster and more abruptly closing vocal folds than unstressed vowels. No clear tendencies can be observed for secondary stressed vowels.

In German secondary stress is used mainly in compounds, where the initial constituent obtains main stress and the other constituent obtains secondary stress, at least when the compounds are derived from monosyllabic or trochaic constituents. However, no clear use of vowel duration, fundamental frequency or spectral tilt parameters to mark secondary stress can be found. To differentiate secondary stressed vowels from unstressed and stressed vowels, *Open Quotient*, *Glottal Opening* as well as *Completeness of Closure* and *Incompleteness of Closure* are used by the parents. This gives evidence that these parameters are used to emphasise stress.

In Chapter 6 the development of German vowels and their language-specific probability distribution over the phonetic space for children between 5 months and 10 years of age and for adult speakers is analysed in order to describe the development of a cognitive map of representations in the perceptual space.

Despite the inter-speaker and intra-speaker variability, some regularities in the development of the vowel space can be observed. Apart from the increase of the vowel inventory with age, the characteristics of vowels change over time. A gradual change in the infants' formant values towards values that are typical for an adult speaker can be observed. This process started during the babbling phase and had not been finished by the age of 36 months. These results, as expected, are caused by developmental changes in the infants' abilities and anatomy. The large overlaps between the typical dispersion ellipses for the F1–F2 as well as F2–F3 vowel space become smaller with age due to the spread of the vowels and reduced variability. Thus the observed changes in the infants' vowel production may originate from the infants' increasing motor abilities as well as from the perception-production-loop and the necessity to build up more adult-like representations of vowels.

In Chapter 7 the results of the different analyses are integrated in a comprehensive neural model of speech production using a self-organising neural network. This model is very appropriate to describe the differences in the production of babbling and words as there are possibilities to develop different neural maps depending on the stage of development as well as depending on the articulatory abilities. Self-feedback as well as feedback from the environment is also integrated in this model of production and perception.

9 Zusammenfassung

9.1 Motivation

Kinder sind, schon bevor sie überhaupt ihre ersten Wörter produzieren, sensitiv für die Auftretenshäufigkeiten bestimmter Lautmuster in ihrer Muttersprache (Jusczyk et al., 1994). Von Geburt an speichert die kindliche Wahrnehmung kontinuierlich alle Sprachwahrnehmungen (Vihman et al., 2004). Schon Kleinkinder können die salienten Eigenschaften des Sprachsignals wahrnehmen und verarbeiten. Sie bilden aufgrund ihrer Erfahrung mit einer Anzahl wahrgenommener Exemplare sprachspezifische Verteilungen interner Modelle aus und generalisieren, um so mittels der abgelegten internen Lautmuster Bedeutungen oder syntaktische Anforderungen einzelner Wörter aus dem sie umgebenden Sprachstrom extrahieren zu können (cf. Saffran et al., 1996; ten Bosch and Cranen, 2007).

Untersuchungen sowohl zur Erwachsenensprache (Pierrehumbert, 2001) als auch zum Spracherwerbsprozess (Curtin, 2002; Beckman, 2003; Fisher et al., 2004; Zamuner et al., 2004) legen die Annahme nahe, dass die im mentalen Lexikon zugrundeliegenden phonologischen Repräsentationen nicht völlig abstrakt sind, sondern dass feine phonetisch-akustische Details Teil der Repräsentation sind. Jede wahrgenommene lexikalische Repräsentation enthält Echospuren von z.B. Betonung und Stimmqualität, und wird mit ihrer Merkmals- und Kontextbeschreibung im mentalen Lexikon abgespeichert. In ihrer Untersuchung zur mentalen Repräsentationen der Betonung zeigen Curtin et al. (2005), dass auch Betonung in der Repräsentation einer analysierten Sequenz kodiert wird. Betonte und unbetonte Silben scheinen unterschiedlich organisiert zu werden und einen unterschiedlichen perzeptuellen Raum einzunehmen. Schon bei 5–7 Monate alten Kindern ist Betonung offenbar Teil der proto-lexikalischen Repräsentation. Doch inwiefern folgt die Produktion diesen unterschiedlichen proto-lexikalischen Repräsentationen? Es scheint, dass mittels Perzeption unterschiedliche Exemplare zusammen mit einer Merkmals- und Kontextbeschreibung abgelegt werden. Auf Basis eines exemplarbasierten Sprachproduktionsmodells sollte das Kind bei der Produktion Zugriff auf diese Repräsentationen haben und diese zur Produktion verwenden. Es scheint jedoch zuerst eine kritische Anzahl wahrgenommener Exemplare nötig zu sein, um eine Kategorie auch produzieren zu können. Im Spracherwerbsprozess kommt es daher zu einer Diskrepanz zwischen Perzeption und Produktion. Längst nicht alle Repräsentationen und Kategorien, die mittels Perzeptetionstests nachgewiesen wurden (Jusczyk et al., 1994; Vihman et al., 1998; Beckman, 2003; Christophe et al., 2003b; Echols and Marti, 2004; Nazzi et al., 2006; Burns et al., 2007; Thiessen and Saffran, 2007; Maye et al., 2008;

Nazzi et al., 2008), können mit Beginn der Produktion auch produziert werden. Die ersten prälinguistischen Äußerungen eine Kleinkindes basieren eher auf rhythmischem Öffnen und Schließen der Stimmbänder als auf gezielter Produktion.

Bezüglich der Wortbetonung hat sich gezeigt, dass unterschiedliche Repräsentationen für betonte und unbetonte Silben mit Beginn der Lallphase vorhanden sind. Doch wie äußern sich diese unterschiedlichen Repräsentationen in der Produktion? Darüber gibt es, vor allem für das Deutsche, noch keine aussagekräftigen Arbeiten. Daher soll das Hauptaugenmerk in dieser Dissertation darauf liegen, zu untersuchen, wie Kinder die deutsche Wortbetonung erwerben und welchen Einfluß die Lallphase auf die spätere Wortproduktion hat. Ein weiteres Ziel dieser Dissertation ist es, auch die Variabilität der Entwicklungschritte, die beim Erwerb der Betonung beobachtet werden können, zu erklären und ein exemplar-basiertes Spracherwerbsmodell von Betonung im Deutschen zu entwickeln. Dafür sollen sowohl phonologische wie auch phonetische Aspekte der Wortbetonung im Deutschen analysiert werden.

Dazu werden in dieser Dissertation folgende Schwerpunkte untersucht:

1. Für die Beschreibung der phonologischen Entwicklung wird sowohl der Erwerb der prosodischen Struktur als auch der Erwerb der Silbenstruktur in Abhängigkeit von der Betonung untersucht. Dabei soll die Rolle der Proto-Silben der Lallphase für die Wortproduktion erklärt werden.

2. Mit Hilfe unterschiedlicher akustischer Parameter der Betonung wird die phonetische Entwicklung der Wortbetonung beschrieben. Es wird analysiert, wie Betonung im Deutschen von Kindern abhängig von ihrem Alter realisiert wird und ob ihre Verwendung von den untersuchten akustischen Parametern vergleichbar ist mit dem elterlichen Gebrauch der Parameter.

3. Die Entwicklung eines perzeptuellen Raums wird dann beispielhaft durch die Entwicklung der unterschiedlichen Vokalräume für betonte und unbetonte Vokale beschrieben.

4. Zum Schluss wird ein exemplar-basiertes Sprachmodell für den Erwerb der Betonung im Deutschen vorgestellt.

9.2 Korpus und Methode

Für die Analysen wurde ein Kindersprachkorpus erstellt, der die sprachliche Entwicklung von 15 deutschsprachigen Kindern im Alter zwischen 5 Monaten und 8 Jahren dokumentiert. Eine genaue Übersicht von allen Kinder findet sich in Appendix 3.1. Alle Aufnahmen fanden in der häuslichen Umgebung der jeweiligen Kinder statt. Über ein an der Kleidung befestigtes Mikrofon wurden die sprachlichen Äußerungen der Kinder auf entsprechende Medien (DAT) aufgenommen. Die Eltern wurden in die Aufnahmetechnik eingewiesen und ebenfalls aufgenommen.

Hauptaugenmerk in dieser Arbeit liegt in der Analyse der Daten zwischen 5 Monaten und 3 Jahren. Daher werden im weiteren nur die Aufnahmebedingungen und Analysen dieser Altersgruppen besprochen. Insgesamt wurden von 6 Kindern (3 Mädchen und 3 Jungen) beginnend von einem Alter von 5 bis 12 Monaten über einen Zeitraum von 2 Jahren regelmäßig Aufnahmen gemacht. Eine genaue Übersicht der einzelnen Aufnahmen findet sich in Appendix A.1.

Den Kinder wurden, abhängig vom jeweiligen Alter und der damit verbundenen Phase des Spracherwerbs, unterschiedliche Aufnahmeszenarien geboten. Zwischen 5 und 18 Monaten wurden alle 6–8 Wochen Sprachaufnahmen der Lallphase von insgesamt 6 Kindern gemacht. Die Interaktion mit ihren Eltern wurde auch aufgenommen. Mit Beginn der Produktion erster Wörter wurde bei 3 Kindern (Nils, Rike und Bennie) Bildkarten mit unterschiedlichen Betonungsmustern eingeführt, um so den Erwerb der Silbe und der Betonungsmuster dokumentieren zu können (siehe Abbildung 3.1). Die gesamte Wortliste ist in Appendix A.4) gegeben.

Die erhobenen Sprachdaten wurden manuell auf acht verschiedenen Labelspuren für die automatische Analyse annotiert. Alle störungsfreien Äußerungen von Kind und Elternteil wurden in die Analysen mit einbezogen.

Die Wortbetonung wurde auditiv über die wahrgenommene Prominenz der einzelnen Silbe innerhalb einer Äußerung markiert. Die prominenteste Silbe wurde als betonte Silbe (1), die nicht prominenten Silben als unbetonte Silben (0) markiert. Des weiteren wurden Silben, die prominent waren, jedoch nicht als die prominenteste Silbe in der Äußerung schien, als sekundärbetont (level stress, 2) markiert. Bei Wortproduktionen wurde auf die korrekte lexikalische Betonung geachtet, ansonsten wurde mittels Prominenz der einzelnen Silben die Betonung markiert.

9.3 Phonologische Entwicklung von Wortbetonung

Als prosodische Domäne von Wortbetonung ist die Silbe Teil der universalen prosodischen Hierarchie (Blevins, 1995). Eine Aufgabe der Betonung ist die hierarchische Organisation von rhythmischen Einheiten. Die Silben werden in Füße und die Füße in prosodische Wörter gegliedert. Diese Gliederung ist verantwortlich für die mögliche Positionierung der Hauptbetonung in einem Wort. Es hat sich gezeigt, dass die prosodische Entwicklung einhergeht mit der Entwicklung von metrischen Füßen und abhängt von der Betonung (Fikkert (1994) für das Niederländische, Demuth and Fee (1995); Kehoe and Stoel-Gammon (1997) für das Englische, Grimm (2008) für das Deutsche).

Für das Deutsche können drei mögliche metrische Strukturen angenommen werden (Tabelle 4.2). Im Deutschen hängt die Wortbetoung davon ab, wie die Silben in die verschiedenen Füße gegliedert werden. Daher ist es für die Entwicklung der Wortbetonung wichtig, sowohl die prosodische Struktur als auch die Silbenstruktur von Äußerungen in der Lallphase wie auch Wortproduktionen in Abhängigkeit von der Betonung zu beschreiben. Im folgenden Abschnitt soll daher zuerst die Entwicklung der prosodischen

Struktur, vor allem von zwei- und dreisilbigen Äußerungen, sowie die Entwicklung verschiedener Silbenstrukturen beschrieben werden. Für die Entwicklung der prosodischen Struktur werden die produzierten metrischen Muster auf ihre Häufigkeiten abhängig von der Produktionsphase untersucht.

Basierend auf einer Studie von Allen and Hawkins (1980) zur frühen Produktion von englischen Kleinkindern wurde häufig angenommen, dass Kleinkinder eine angeborene Tendenz besitzen, trochäischen (stark-schwach) Füßen den Vorzug in ihrer phonologischen Entwicklung zu geben (Fikkert, 1994; Archibald, 1995; Gerken, 1996). Vihman et al. (1998) hingegen fand für Englisch und Französisch lernende Kinder keine solchen trochäische Tendenz. Die Entwicklung der prosodischen Struktur scheint sich eher durch eine Expansion der prosodischen Domäne von einem einzelnen Fuß hin zu mehrsilbigen Wörten auszuzeichnen. Während dieser Entwicklung zeigen die Kinder eine Vielfalt an produzierten prosodischen Mustern, da erst einmal die zugrunde liegende Syllabifikation in der Phase der Wortproduktion erlernt werden muss. Während der prälinguistischen Phase scheinen keine Restriktionen bezüglich der prosodischen Strukturen vorhanden zu sein, da sowohl trochäische (stark-schwach), iambische (schwach-stark) wie auch spondeische (stark-stark) Muster produziert werden können (Vihman et al., 1998; Davis et al., 2000; DePaolis et al., 2008).

Zum Beschreiben der Silbenstruktur wurde das von Levelt et al. (2000) beschriebene Verfahren angewandt. Aufgrund möglicher Extrasilbizität wurden nur Silbenstrukturen, die dem Sonoritätsprinzip folgen, untersucht. Im Deutschen kommt dazu noch eine systematisch distinktive Unterscheidung von langen (V:) und kurzen (V) Vokalen da auch während des Spracherwerbs zu keiner Phase von einer rein zufälligen Produktion von Vokallänge ausgegangen werden kann (Kehoe and Lleó, 2003). Folgende unterschiedliche Silbenstrukturen, bezeichnet mit C für Konsonanten und V für Vokale (unabhängig von Vokallänge) wurden abhängig von der Betonung analysiert: CV, CVC, V, CCV, VC, CCVC, CVCC, CCVCC, VCC.

Um die prosodische Entwicklung zu beschreiben, wurden von den beschriebenen sechs Kindern in einer Zeitspanne von 5 Monaten bis 3 Jahren alle gelabelten Äußerungen untersucht. Abhängig vom Alter und der Phase der Produktion (Lallen oder Wort) wurden Kreuztabellen für alle produzierten metrischen Strukturen über alle Kinder berechnet und ausgewertet. Für die Entwicklung der Silbenstruktur wurden zuerst für jedes Kind abhängig vom Alter, der Phase der Produktion und der Betonung Kreuztabellen für jede produzierte Silbenstruktur, sowohl mit langen wie auch mit kurzen Vokalen, berechnet und ausgewertet. Es zeigt sich, dass innerhalb des Lallens eine große Variabilität der produzierten Silbenstrukturen, unabhängig von der Betonung, zu beobachten ist. In der prälinguistischen Phase sind keine Restriktionen hinsichtlich der Silbenstruktur zu erkennen. Dies ändert sich jedoch mit der Produktion von ersten Wörtern. Diese sind sowohl hinsichtlich ihrer Silbenstruktur beschränkt wie auch abhängig von der Betonung. Es zeigt sich, dass im Deutschen zuerst die "Kernsilben" CV und CVC erworben werden. Es scheint, dass zuerst die Silbenstruktur an sich erworben wird und diese standardmäßig zuerst betont produziert wird. Erst wenn die Struktur erworben ist, kann der Betonungsstatus geändert werden und durch eine Reduzierung der akustischen Parameter auch unbetont produziert werden. Desweiteren lässt sich erkennen, dass es im Deutschen zu einer bestimmten Abfolge im Erwerb der untersuchten Silbenstrukturen kommt. Nur für ein Kind kann eine etwas unterschiedliche Abfolge mittels Skalierung ermittelt werden. Zuerst werden von allen untersuchten Kindern die im Deutschen auch am häufigsten auftretenden Silbenstrukturen CV und CVC erworben. Die Abfolgen V, VC, CCV und CCVC zeigen Variabilität in ihrer Erwerbsabfolge. Die meisten untersuchten Kinder folgen der Abfolge, dass zuerst V und VC erworben wird und dann erst komplexere Onsetkluster. Ein Junge (Bennie) jedoch produziert zuerst komplexere Onsetkluster, bevor auch V und VC in seiner Produktion auftauchen. Es zeigt sich, dass diese Silbenstrukturen im Deutschen fast die gleichen relativen Häufigkeiten zeigen. Es könnte daher sein, dass Bennie mit mehr Wörtern mit komplexen Onsetklustern konfrontiert wurde als die anderen Kinder und daher diese aufgrund der größeren absoluten Häufigkeit etwas eher erworben hat. Komplexe Codas, die seltensten Silbenstrukturen im Deutschen, wurden bei beiden Entwicklungsgruppen als Letztes erworben.

9.4 Phonetische Entwicklung von Wortbetonung

Betonung bezieht sich auf die Prominenz einer Silbe (Wortbetonung) und wird mit einer Anzahl daran beteiligter akustischer Parameter realisiert. Akustische und perzeptuelle Studien von Erwachsenensprache zeigen, dass im Deutschen die betonte Silbe durch längere Dauer, ansteigende Grundfrequenz, ansteigende Intensität, meist gemessen als spektrale Neigung, und Änderungen in der Stimmqualität markiert ist (Pompino-Marschall, 1990; Heuft and Portele, 1994; Heuft et al., 1995; Jessen et al., 1995; Dogil and Williams, 1999). Die beschriebenen Kindersprachdaten wurden mit Hilfe eines am Institut für Maschinelle Sprachverarbeitung der Universität Stuttgart (IMS) entwickelten Programmpakets zur automatischen Extraktion und Berechnung der akustischen und Stimmqualitätsparameter analysiert. Dieses Tool wurde von Pützer and Wokurek (2006) entwickelt und zuvor nur für die Analysen männlicher, deutscher Sprecher getestet. Es verwendet Amplituden- und Frequenzmessungen an den spektralen Gipfeln der Harmonischen als Basis für die akustischen Analysen. Dies ermöglicht die Bearbeitung umfangreicher Sprachdaten. Das Analysetool misst automatisch zwischen den gesetzten Labels mit einer vorgegebenen Fensterlänge und Schrittweite an 5 Zeitpunkten (bei 10%, 30%, 50%, 70% und 90% der Vokaldauer) (Shih et al., 1999; Möbius, 2004) innerhalb eines jeden Vokals verschiedene akustische Parameter wie die Grundfrequenz (F_0), Intensität (RMS), Formanten, Harmonische und Amplituden. Über die so erhaltenen Messwerte werden dann mittels mathematischer Formeln verschiedene Stimmqualitätsparameter berechnet. Die Vokaldauer ergibt sich trivial aus den markierten Vokalintervallen.

Da alle untersuchten Parameter vokalabhängig sind, wurden normalisierte (z-transformierte) Werte der Vokaldauer, Intensität und Stimmqualitätsparameter für die statistische Analyse verwendet. Die Formantwerte wurden in die Bark-Skala überführt, um eine sprechernormalisierte Ausdehnung der Formanten im Vokalraum zu erreichen.

Es zeigt sich, dass schon in der prälinguistischen Phase von den Kleinkindern unterschiedliche Betonungsmuster produziert werden können. Aber die Realisierung und der Gebrauch der verschiedenen akustischen Parameter ist noch sehr inkonsistent. Jeder gemessene Parameter wird auch mehr oder weniger, abhängig vom Alter des Kindes, verwendet, um unterschiedliche Prominenz der Silben zu produzieren. Es zeigt sich aber, dass für die untersuchten Kinder Intensität und Grundfrequenz die am zuverlässigsten verwendeten Parameter sind. Mit ansteigendem Alter sind die Kinder immer besser in der Lage, auch Dauerunterschiede zum Markieren von Prominenz einzusetzen. Die untersuchten Stimmqualitätsparameter werden auch abhängig vom Alter verwendet, doch auch hier zeigt sich mit ansteigendem Alter eine immer bessere Kontrolle über die Stimmlippen und damit einhergehend eine immer konstantere Verwendung der Parameter. Die Verwendung der möglichen Parameter im Deutschen, um Wortbetonung zu markieren, nähert sich mit ansteigendem Alter immer mehr dem elterlichen Gebrauch dieser Parameter an. Es zeigt sich, dass außer Dauer, die von allen untersuchten Eltern als Hauptkorrelat der Betonung eingesetzt wurde, diejenigen Korrelate von den Kinder übernommen werden, mit denen sie durch ihre Eltern am häufigsten konfrontiert wurden. Dies zeigt sich insbesondere dann, wenn die Eltern, außer Dauer, unterschiedliche Korrelate der Betonung bevorzugen und das Kind die zu übernehmen scheint, mit denen es je nach Betreuungszeiten vorwiegend konfrontiert wurde.

9.5 Entwicklung des Vokalraums

Die Entwicklung des Vokalraumes für englisch- und französisch-lernende Kinder wurde von einer Anzahl von Querschnittsstudien beschrieben (Kent and Murray, 1982; de Boysson-Bardies et al., 1989; Lee et al., 1999; Davis et al., 2000; Rvachew et al., 2006; Ménard et al., 2007b).

In der ersten Phase des kanonischen Lallens zeigen Kleinkinder eine starke Präferenz für die offenen und zentralen Vokale /a ɛ ə/. Mit ansteigendem Alter kann eine graduelle Erweiterung des Vokalinventars beobachtet werden. Diese ist abhängig von der immer besser werdenden Kontrolle über die Artikulatoren (MacNeilage, 1998; MacNeilage and Davis, 2000; Lacerda, 2003). Um nun sprachspezifische perzeptuelle Ziele zu erreichen (Ménard et al., 2004), müssen die Kinder im Laufe des Erwerbsprozesses ihre artikulatorisch-akustischen Verbindungen, abhängig von ihrer verbesserten Motorkontrolle sowie ihren anatomischen Veränderungen anpassen. Die Kinder müssen, um ein bestimmtest Ziel zu erreichen, entsprechende artikulatorische Routinen aufbauen und diese Routinen wiederum mit den entsprechend wahrgenommenen akustischen Realisierungen verbinden (Dogil, 2010). Kinder müssen somit erst einmal die physiologischen

Einschränkungen ihrer noch nicht ausgebildeten Motorkontrolle überwinden (Ménard et al., 2004), um einen sprachspezifischen Vokalraum aufzubauen.

Abhängig von der sukzessiven Verbesserung der Motorkontrolle wird der Vokalraum immer weiter erweitert und immer unterschiedlichere Vokalkategorien werden gebildet (cf. Peterson and Barney, 1952; Hillenbrand et al., 1995; Busby and Plant, 1995; Palethorpe et al., 1996; Smith and Kenney, 1998, 1999; Lee et al., 1999; Whiteside, 2001; Yildrim et al., 2003). Die relative Position eines Vokals im akustischen Raum ist somit abhängig sowohl vom Alter wie auch von der immer besseren Motorkontrolle (Ménard et al., 2007b) und der darauf nötigen Rekalibrierung der artikulatorisch-akustischen Verbindungen, um ein bestimmtes perzeptuelles Ziel zu erreichen.

Die bisher vorgestellten Studien beschreiben hauptsächlich den Vokalraum betonter Vokale, da die verwendeten Daten auf Einsilblern beruhen, und diese als betont gelten. Es kann aber davon ausgegangen werden, dass betonte und unbetonte Vokale unterschiedliche Kategorien im perzeptuellen Raum ausbilden. Im Deutschen zeigt sich, dass betonte und unbetonte Vokale unterschiedlichen artikulatorischen Routinen unterliegen (Dogil, 1995). Daher wurde in dieser Arbeit untersucht, inwieweit sich unterschiedliche Vokalräume abhängig von der Betonung ausbilden.

Um diese Entwicklung aufzuzeigen, wurden Daten von allen 15 deutschen Vokalen in Spontansprache von 27 Sprechern ausgewertet. Das Alter der Sprecher liegt zwischen 5 Monaten und 10 Jahren für die Kinderdaten, des weiteren wurden auch die Erwachsenendaten des beschriebenen Korpus ausgewertet. Dabei soll vor allem die Frage beantwortet werden, welche artikulatorischen Strategien die Kinder innerhalb ihres Erwerbsprozesses verwenden, um sprachspezifische Vokale zu produzieren, und wie sich der Vokalraum entsprechend dem Alter und der Betonung individuell entwickelt. Um die Entwicklung des Vokalraumes zu beschreiben, wurden für alle betonten und unbetonten Vokale Verteilungsellipsen (mit Radius korrespondierend zum $\pm$ 1,5 fachen Standardfehler der individuellen Daten um den Mittelpunkt) im Vokalraum gebildet und deren Fläche berechnet. Es zeigt sich, dass schon in der Lallphase unterschiedliche Vokalräume für betont und unbetont wahrgenommene Vokale produziert werden. Mit Beginn der Wortproduktion expandiert der Vokalraum für betonte Vokale, wohingegen unbetonte Vokale eine Zentralisierung in F1-F2 Raum zeigen. Die interne Organisation des Vokalraums ist sprecherspezifisch, welches durch eine breite Variabilität der untersuchten Vokalräume zu beobachten ist. Um die Entwicklung der Vokale zu beschreiben, wurden diese dann aufgrund ihrer Artikulationsstelle, Öffnungsgrad und Rundung in unterschiedliche Gruppen eingeteilt. Die Verteilung der unterschiedlichen Ellipsenflächen wurde für jeden Parameter in Abhängigkeit vom Alter untersucht und es zeigt sich bis zu einem Alter von 10 Jahren eine sehr große Variabilität in den Ellipsenflächen, die für alle untersuchten Parameter signifikant größer sind als für die untersuchten Erwachsenen. Die Ellipsenflächen der Erwachsenendaten sind kleiner und zeigen somit eine geringere Variabiliät und exakte Kategorisierung.

Desweiteren wird die individuelle Entwicklung der Vokale zwischen 5 Monaten und 3 Jahren anhand der beschriebenen Langzeitdaten für sechs Kinder analysiert. Der Einfluß der Umgebung auf den produzierten Vokalraum zeigt sich bei den sechs Kindern der Langzeitstudie ab einem Alter von etwa 24 Monaten. Bis zu diesem Zeitraum kommt es bei allen Kindern zu einer großen Variabilität sowohl des Vokalraumes wie auch innerhalb einer Vokalkategorie. Ab etwa 2 Jahren dagegen entwickeln sich die Vokalräume der Kinder dahingehend, dass die extremen Kardinalvokale immer mehr die Eckpunkte des produzierten Vokalraums bilden. Es zeigt sich aber ab diesem Zeitpunkt auch ein Einfluss der Eltern dahingehend, dass diejenigen Kinder, deren Eltern klar definierte Vokalkategorien zeigen, eher unterschiedliche Kategorien ausbilden und weniger überlappende Vokalräume zeigen als Kinder, deren Eltern keine so klare Trennung der Vokalkategorien produzieren.

9.6 Diskussion der Ergebnisse

Die gefundenen Ergebnisse der verschiedenen Analysen zur Silbenstruktur, metrischen Mustern, akustischer Realisation von Betonung sowie der Entwicklung des Vokalraums von betonten und unbetonten Silben wurde dann in ein neuronales Sprachproduktionsmodell eingebunden. Das entwickelte Modell zeigt zwei unterschiedliche Produktionsprozesse für die Lallphase und Wortproduktion auf. Es basiert auf dem neuronalem Sprachproduktionsmodell von Kröger und Kollegen (2008, 2009), und verwendet sich selbst-organisierende neuronale Netzwerke. Ein zentraler Punkt dieses Modells ist die zeitgleiche Aktivierung der phonemischen, sensorischen, und motorischen Planung zusammen mit der phonetischen Speicherung.

In dem vorgestellten Sprachproduktionsmodell, basierend auf den Daten zum Erwerb von Betonung im Deutschen, aktiviert externer Input während der Lallphase mittels der auditorischen Verarbeitung und auditorischen Speicherung das auditorisch-phonetische Verarbeitungsmodul. Als Teil dieses Prozesses helfen die schon vorhandenen phonologischen Reprsentationen im mentalen Lexikon das Sprachsignal zu selektieren. Diese wahrgenommenen sprachspezifischen Strukturen werden wiederum im mentalen Silbenspeicher gespeichert. Bei der Produktion dagegen werden nicht nur die bei der Perzeption aktivierten Verarbeitungsprozesse aktiviert sondern zusätzlich auch die somatosensorische sowie artikulatorisch-akustische Bewegung werden Teil der abgespeicherten Repräsentationen und somit Teil des Proto-Silbenspeichers.

Externer Input scheint somit während der Lallphase nur zu einer Aktivierung der auditorischen Verarbeitung und dem Aufbau sprachspezifischer Repräsentationen für die Zerlegung des Sprachsignals zu führen. Diese Repräsentationen haben keine Verbindung zu einer artikulatorisch-akustischen Bewegung, da diese nur in Verbindung mit der somato-sensorischen Verarbeitung stattfinden kann.

Diese Verarbeitungsprozesse ändern sich mit Beginn der Wortproduktion. Nun muss ein Abgleich von Produktion und Zielwort stattfinden. um zu immer besseren Ergebnissen zu kommen. Dafür muß nun auch die phonologische Kodierung, welche auf der grammatischen Kodierung sowie konzeptueller Vorbereitung basiert, bei der Produktion mit einbezogen werden. Ein sprachspezifischer Plan der verwendeten Laut für betonte und unbetonte Silben wird im mentalen Silbenspeicher, basierend auf dem während der Lallphase aufgebauten Proto-Silbenspeicher, aufgebaut und mittels externem Feedback kontinuierlich verbessert.

Es scheint, dass die Kinder während der Lallphase die sprachspezifischen Muster und Phoneme erkennen und zu produzieren versuchen. Es zeigt sich, dass in der prälinguistischen Phase die ersten Sprachproduktionen eher durch biomechanische Beschränkungen beeinflusst werden. Im Spracherwerbsprozess müssen die Kleinkinder erst einmal lernen, basierend auf ihren perzeptuellen Erfahrungen, die unterschiedlichen akustischen Realisierungsmöglichkeiten und die unterschiedlichen artikulatorischen Abläufe für die Realisierung einer Lautabfolge zu kontrollieren und vor allem zu koordinieren. Es zeigt sich, dass der größte Teil der gelallten Äußerungen eher als prominent wahrgenommen wird und nur etwa 30% als wirklich nicht prominent gelten können. Daher scheint die Produktion von prominenten Silben als der unmarkierte Fall zu gelten, Kinder müssen daher durch eine Reduzierung der unterschiedlichen akustischen Parameter erst einmal lernen, nicht prominente Silben zu produzieren. Dafür ist jedoch eine Feinjustierung der unterschiedlichen akustischen Realisierungsmöglichkeiten nötig. Um dieses Ziel zu erreichen, muss innerhalb der Lallphase die Kontrolle über die verschiedenen Artikulatoren erlangt werden. Die zufällig produzierten unterschiedlichen Prominenzen werden dann in der Wortproduktion abhängig von der Wortbetonung und der Silbenstruktur mit Hilfe des phonologischen Wissens ausgewählt. Dafür müssen allerdings zuerst die zugrundeliegenden artikulatorisch-akustischen Abläufe, die in der Lallphase antrainiert wurden und im Proto-Silbenspeicher abgelegt sind, mit der Bedeutung eines Objektes verbunden werden. Um ihre Äußerungen zu verbessern, ist nun sowohl Selbstkontrolle nötig als auch Resonanz auf die Äußerung aus der unmittelbaren Umgebung. Altersspezifisch werden nun kontinuierlich mit wachsender Erfahrung immer besser die verschiedenen akustischen Parameter für die Realisierung unterschiedlicher Betonungsstadien verwendet, und immer unterschiedlichere Betonungsmuster können produziert werden. Die kindliche Produktion nähert sich somit in diesem Aspekt immer mehr der Erwachsenensprache an.

Durch die Prosodie der Muttersprache erkennen schon Babys die phonetische und phonotaktische Struktur der zu erlernenden Sprache. Sprachspezifische Erwartungen an die sie umgebende Sprache werden aufgebaut und als angereicherte Repräsentationen im mentalen Silbenspeicher abgelegt. Um eine Sprache zu lernen, muss somit zuerst einmal ein feinverteiltes, sprachspezifisches neuronales Netzwerk aufgebaut werden. Dieses neuronale Netzwerk wird zum festen Bestandteil des sprachlichen Netzwerkes, da dieses scheinbar auch bei pathologischer Sprache erhalten bleibt (Dogil, 2010). Bei einer Stöung im Aufbau dieses sprachlichen Netzwerkes scheint eine Sprache nicht fehlerfrei erlernt werden zu können. In diesem Fall kann es zu phonologischen Defiziten, wie sie zum Beispiel bei Dyslexie auftreten (van Alphen et al., 2007; de Bree et al., 2008), kommen.

A Method

Participant information

Word lists

Explorative data analysis

Babbling phrases

Table A.1: Participant information of analysed speech recordings.

Subject	**Age** (years;months:days)	**recording time**	**total of utterances**	**syllable tokens**	**vowel tokens**
Emma	0;5:4	15 min	34	92	100
	0;9:7	40 min	176	414	448
	1;1:2	10 min	25	59	57
	1;2:22	63 min	157	399	426
	1;5:14	66 min	132	260	277
	1;10:8	59 min	269	483	532
	2;0:15	22 min	174	320	338
	2;2:8	36 min	295	571	649
Hanna	0;5:23(-28)	16 min	29	150	158
	0;7:0(-23)	28 min	93	285	288
	1;3:11(-13)	10 min	47	75	68
	1;6:0	13 min	84	179	183
	1;8:11	15 min	30	70	70
	2;3:2	7 min	18	39	39
	2;6:7	29 min	63	190	191
	2;10:8	52 min	65	170	170
	3;0:19(-22)	35 min	108	314	314
Ollie	0;9:1(-12)	20 min	31	99	115
	1;0:12(-15)	17 min	62	126	165
	1;2:7(-10)	12 min	72	128	156
	1;6:0	7 min	53	86	103
	2;9:0	11 min	38	100	102
Bennie	0;11:0(-2)	27 min	65	130	134
	1;2:17(-21)	30 min	185	247	254
	1;5:15(-20)	45 min	228	304	317
	1;7:14	24 min	140	206	219
	1;9:20(21)	42 min	190	335	342
	2;2:11(12)	27 min	102	261	263
	2;4:27	15 min	48	141	141
	2;11:12(14))	45 min	255	748	750
Nils	0;7:17(-19)	20 min	71	167	179
	0;8:22(-28)	47 min	91	238	302
	0;11:20(-23)	28min	37	74	76
	1;2:0	36 min	138	215	292
	1;5:27(-1;6:2)	51 min	531	1101	1385
	1;8:4(-6)	21 min	17	35	35
	1;9:25	15 min	25	55	55
	2;7:27	22 min	84	245	245
	2;8:12	18 min	90	245	248
Rike	1;0:21(-25)	39 min	198	344	371
	1;4:3	28 min	188	297	338
	1;5:28(-1;6:7)	24 min	84	168	183
	1;7:17	9 min	25	46	46
	1;8:15	15 min	20	42	42
	1;9:24	10 min	17	42	42
	1;11:13	41 min	72	198	198
	2;1:15	35 min	84	150	170
	2;6:6(-7)	64 min	128	370	376

Table A.2: Participant information of the parents recordings of the Stuttgart Child Language Chorpus, cds means child-directed speech, ads means adult-directed speech

Subject	parent of child	speech task	total of utterances	syllables	vowels
BL	mother of Emma	cds	164	250	259
(the author)	evaluation	cds	49	146	144
CH	mother of Hanna	cds	103	218	230
DB	father of Nils	cds	60	167	171
		ads	58	166	170
DZ	mother of Ollie	cds	104	206	210
vH	father of Hanna	cds	64	122	132
MB	mother of Nils	cds	391	737	977
		ads	54	157	159
NW	mother of Bennie	cds	413	1052	1076
		ads	48	145	145
PW	father of Bennie	cds	164	479	482
		ads	49	139	139
VZ	mother of Rike	cds	801	1284	1339

Table A.3: Participant information of the evaluation data

participant	date of recording	age	gender	place of recording
ElAb	2006/05/05	5;2:9	female	Kindergarten
LuPa	2006/05/03	5;2:22	male	Kindergarten
LeWu	2006/05/03	5;3:18	female	Kindergarten
LeVo	2006/05/03	5;10:25	male	Kindergarten
ToRi	2006/05/03	6;6:14	male	Kindergarten
AlSa	2005/11/11	6;4:10	female	primary school
LuMa	2005/11/11	6;9:14	male	primary school
NiPe	2005/11/11	7;0:3	male	primary school
DaWo	2005/11/12	9;4:8	male	primary school
CiKa	2005/11/12	9;11:7	female	primary school
PhVo	2005/11/12	10;0:3	male	primary school
LaHe	2005/11/12	10;1:25	female	primary school

Table A.4: Word list for Memory task

orthography	english word	phones	CV	stress
Akrobat	acrobat	a k r oː ˈb aː t	v-ccv-CVvC	0-0-1
Anorak	anorak	ˈa n oː r a k	V-cvv-cvc	1-0-0
Banane	banana	b a ˈn aː n ə	cv-CVv-cv	0-1-0
Bikini	bikini	b iː ˈk iː n iː	cvv-CVv-cvv	0-1-0
Buchstabe	letter	ˈb uː x ˌʃ t aː b ə	CVC-ccvv-cv	1-0-2
Eisenbahn	train	ˈaɪ z ə n b aː n	VV-cvc-cvc	1-0-0
Elefant	elephant	eː l eː ˈf a n t	vv-cvv-CVCC	0-0-1
Eskimo	eskimo	ˈɛ s k i m o	VC-cv-cv	1-0-0
Fliege	fly	ˈf l iː g ə	CCV-cv	1-0
Flughafen	airport	ˈf l uː k ˌh aː f ə n	CCVC-cv-cvc	1-2-0
Fotograf	fotograph	f o t o ˈg r aː f	cv-cv-CCVC	0-0-1
Gardine	curtain	g a r ˈ d iː n ə	cvc-CVv-cv	0-1-0
Giraffe	giraffe	g i ˈr a f ə	cv-CVC-cv	0-1-0
Gitarre	guitar	g i ˈt a r ə	cv-CVC-cv	0-1-0
Gorilla	gorilla	g o ˈr ɪ l a	cv-CVC-cv	0-1-0
Kamel	camel	k a ˈm eː l	cv-CVC	0-1
Kanone	cannon	k aˈ n oː n ə	cv-CVv-cv	0-1-0
Kobolde	hobgoblins	ˈk oː b ɔ l d ə	CVv-cvc-cv	1-0-0
Kokosnuss	coconut	ˈ k oː k ɔ sˌ n ʊ s	CVv-cvc-cvc	1-0-2
Krokodil	crocodile	k r o k o ˈd iː l	ccv-cv-CVvC	0-0-1
Lastwagen	truck	ˈl a s t ˌv aː g ə n	CVCC-cv-cvc	1-2-0
Lawine	snowslide	l a ˈv iː n ə	cv-CVv-cv	0-1-0
Malerin	paintress	ˈ m aː l ə r ɪ n	CVv-cv-cvc	1-0-0
Matrose	seaman	m a ˈt r oː z ə	cv-CCVv-cv	0-1-0
Mikrofon	microphone	ˈm i: k r o f o: n	CV-ccv-cvc	1-0-0
Mikrofon	microhpone	m i k r o ˈf oː n	cv-ccv-CVvC	0-0-1
Muellwagen	garbage truck	ˈm ʏ l ˌv aː g ə n	CVC-cv-cvc	1-2-0
Paket	package	p a k ˈeː t	cv-CVC	0-1
Papagei	parrot	p a p a ˈg aɪ	cv-cv-CVV	0-0-1
Papagei	parrot	ˈp a p a g aɪ	CV-cv-cvv	1-0-0
Pinguin	penguin	ˈp ɪ ŋ g uː iː n	CVC-cv-vc	1-0-0
Pistole	pistol	p ɪ sˈ t oː l ə	cvc-CVv-cv	0-1-0
Polizei	police	p o l i ˈt s aɪ	cv-cv-CCVV	0-0-1
Polizist	policeman	p o l i ˈt s ɪ s t	cv-cv-CCVCC	0-0-1
Postauto	mail car	ˈp ɔ s tˌ aʊ t oː	CVCC-vv-cv	1-2-0
Postbote	mailman	ˈp ɔ s t ˌb oː t ə	CVCC-cv-cv	1-2-0
Prinzessin	princess	p r ɪ n ˈt s ɛ s ɪ n	ccvc-CCVC-cv	0-1-0
Pullover	jumper	p ʊ ˈl oː v ɒ	cvc-CVv-cv	0-1-0
Radfahrer	cyclist	ˈr aː t ˌf aː r ɒ	CVC-cv-cvc	1-2-0
Rennfahrer	racer	ˈr ɛ n ˌf aː r ɒ	CVC-cv-cvc	1-2-0

orthography	english word	phones	CV	stress
Sandale	sandal	z a n' d aː l ə	cvc-CVv-cv	0-1-0
Saxophon	saxophone	z a k s o 'f oː n	cvc-cv-CVC	0-0-1
Saxophon	saxophone	' z a k s o f oː n	CVC-cv-cvc	1-0-0
Schmetterling	butterfly	'ʃ m ɛ t ɒ l ɪ ŋ	CCVC-cv-cvc	1-0-0
Skifahrer	skier	'ʃ iː ˌf aː r ɒ	CVv-cv-cv	1-2-0
Spiegel	mirror	' ʃ p iː g ə l	CVv-cvc	1-0
Stadion	stage	'ʃ t aː d (i) ɔ n	CCVv-cv-vc	1-0-0
Teddybaer	teddy bear	't ɛ d iː ˌb ɛː ɒ	CVC-cv-cvv	1-0-2
Tomate	tomato	t o' m aː t ə	cv-CV-cv	0-1-0
Trampolin	trampoline	t r a m p o' l iː n	ccvc-cv-CVvC	0-0-1
Trampolin	trampoline	' t r a m p o l iː n	CCVC-cv-cvvc	1-0-0
Trompete	trumpet	t r ɔ m' p eː t ə	ccvc-CVv-cv	0-1-0
Vulkan	volcano	v ʊ l' k aː n	cvc-CVvc	0-1
Vulkane	volcanos	v ʊ l 'k aː n ə	cvc-CVv-cv	0-1-0
Wohnwagen	caravan	' v oː n ˌv aː g ə n	CVvC-cv-cvc	1-2-0
Zitrone	citron	t s i 't r oː n ə	ccv-CCVv-cv	0-1-0

Table A.5: Contrastive stress names and toys used in the TAKi task

Animal toy 1	**Animal toy 2**
brown bear ['bimo]]	polar bear [bi'mo]
big zebra ['nami]	small zebra [na'mi]
otter ['doba]	badger [do'ba]
big tiger ['midano]	small tiger [mida'no]
flying eagle ['badoni]	standing eagle [bado'ni]

Table A.6: Types, tokens and discarded values (in %) for each child.

		Hanna	Nils	Bennie	Rike	Emma	Olli
duration	**types**	1366	2606	2398	1753	2606	518
	tokens	1324	2498	2308	1665	2511	498
	% discarded	3.07%	4.14%	3.75%	5.02%	3.65%	3.86%
F_0	**types**	5369	11493	10733	7381	10506	2181
	tokens	5208	11179	10366	7159	10220	2128
	% discarded	3.00%	2.73%	3.42%	3.00%	2.72%	2.43%
F1	**types**	5706	12107	11273	7706	10700	2223
	tokens	5478	11559	11001	7500	10477	2162
	% discarded	4.00%	4.5%	2.41%	2.67%	2.08%	2.74%
F2	**types**	5708	12107	11274	7706	10700	2223
	tokens	5454	11611	10915	7269	10411	2149
	% discarded	4.45%	4.10%	3.18%	5.57%	2.70%	3.33%
F3	**types**	5708	12107	11274	7706	10700	2223
	tokens	5438	11759	10997	7472	10418	2125
	% discarded	4.73%	2.87%	2.46%	3.04%	2.64%	4.41%
F4	**types**	5708	12107	11274	7706	10700	2223
	tokens	5496	11807	10989	7530	10407	2131
	% discarded	3.71%	2.48%	2.53%	2.28%	2.74%	4.14%
RMS	**types**	5710	12108	11274	7739	10700	2223
	tokens	5477	11410	10704	7291	10073	2110
	% discarded	4.08%	5.76%	5.01%	5.79%	5.86%	5.08%

Table A.7: Utterances longer than 6 syllables classified as babbles

child (years;months:days)	**syllable structure**	**word stress pattern**	**ipa-transkription**
Emma(0;5:4)	v-cv-V-v-v-V-V	s-W-S-W-W-S-S	a-hə-a-ʊ-œ-a-ø-ɛ
Emma (0;5:4)	V-v-v-v-cv-CV-cv	S-W-W-W-W-S-W	æ-ɛ-u-ʊ-lɪ-ja-və
Emma (0;5:4)	V-CVC-v-vc-v-cv-v-V-v	S-S-W-s-s-s-W-S-W	ø-vɛl-ɛ-ɛl-ɛ-lø-e-ʉ-ə
Hanna (0;5:23)	cv-CV-cv-CV-cv-cv-cv-cv	s-S-s-S-s-s-s-S	hɛ-dɛ-dɛ-dɛ-dɛ-dɛ-de-dɛ
Hanna (0;5:23)	vvc-cv-CVC-vc-vvcc-CVC-cccvc	W-s-S-W-W-W-S-s	oin-dɛ-dɛn-œd-ealn-dɔn-bla-lnbœ
Hanna (0;5:23)	vc-CVC-cvc-cvc-CVC-cvc-cvc-cvc	W-S-s-s-S-s-s-s	ad-dan-dɛn-jan-jain-jain-jan-jœng
Hanna (0;5:23)	CVV-ccvc-cvc-cv-vvcc-cvvc-ccv-VVC-ccvv	S-s-s-W-W-W-s-S-s	hœi-gjɛn-bin-de-aind-join-djɛ-ɛin-djɛu
Hanna (0;5:23)	VC-VC-cvc-ccvc-cvcc-cvc-cvc-CCVVC-cvvc-cvc	S-S-s-s-s-W-s-S-W-s	ɪl-ɛn-den-djɛn-bɛln-dɛn-dɛn-djoin-guig-jah
Hanna (0;5:27)	cvc-CV-cv-cv-cv-cv-cvc	s-S-W-W-s-s-W	han-na-an-na-an-na-nɛh
Hanna (0;5:27)	v-cv-vc-cv-vv-CV-cv-cvvc	W-W-W-s-W-S-s-s	a-la-an-la-œa-la-ja-rœah
Hanna (0;7:0)	cvc-cvc-CVV-cvv-vv-CV-vv	W-s-S-s-s-S-W	hɛl-dɪn-deɪ-vaɪ-ɛ-je-aɪ
Hanna (0;7:0)	vc-CVV-cv-v-v-v-v	W-S-s-W-W-W-W	ɛn-jaɪ-ja-a-a-a
Hanna (0;7:3)	cv-cvc-vc-v-cvc-CVC-cvc-cvc-cvc-cv	W-W-s-s-S-W-W-W-W-W	ha-jɛn-in-a-naŋ-ŋɛn-non-nan-nan-nə
Hanna (0;7:23)	v-cv-cv-CV-cvc-cv-cv	W-W-W-S-W-s-W	æ-jæ-lœ-dɛ-dœŋ-a-də
Nils (0;7:17)	CV-cv-cv-cv-cv-CV-cv-cv	S-s-W-W-W-S-s-W	va-dɛ-ja-da-da-ja-da-da
Emma (0;8:9)	cv-cv-cv-cv-cv-cv-CV	s-W-W-s-W-s-S	ja-ja-ja-jal-ja-jal-ja
Emma (0;8:9)	VVC-cv-cvvc-cv-cv-cv-CV-cv	S-W-s-W-s-s-S-W	aiŋ-jɛ-cɛ-ja-ha-ha-ja-la
Emma (0;8:9)	v-cv-cv-cvc-vc-v-v-v-CV	W-s-s-W-s-W-s-W-S	e-ja-ja-gɛg-aŋ-a-a-a-ga
Emma (0;8:9)	v-cv-vv-v-v-cvc-cvv-v-cv-cv-v-CV	s-W-s-W-W-s-Ws-W-W-s-W-S	a-hɛɪ-ɛɪ-æ-ɪ-jal-e-lɛɪ-val-a-ja
Nils (0;8:23)	vv-CV-cvv-cvv-cvv-cvv-cvv-v	s-S-W-W-s-W-W-s	ai-je-dæ-dai-ai-æ
Nils (0;8:27)	v-V-cv-CV-cv-cv-cv-cv	W-S-W-S-W-W-W-W	ɛ-jæ-jæ-dæ-jæ-dæ-dæ-dæ
Nils (0;8:22)	v-cv-CV-cvv-v-v-v-vv	W-W-S-s-W-W-W-W	æ-gæ-ga-gaɪ-a-æ-ɛ-aɪ
Nils (0;8:23)	CV-cv-CV-v-cv-cv-cv-cv-CV	S-W-S-W-s-W-W-s-S	ja-gæ-da-æ-ja-ja-ja-da-ja
Nils (0;8:23)	V-cv-cvv-cv-cv-CVV-cvv-cv-v	S-W-W-W-W-S-s-W-W	œ-jɛ-daɔ-ʔœ-pæ-dai-dai-dæ-a
Nils (0;8:27)	cv-CV-cv-cv-cv-cv-cv-cvc-cvc-cv-CV	W-S-W-W-W-W-W-W-W-W-S	dæ-dæ-nɛ-næ-nɛ-nɛ-dæ-lɛn-nɛn-da-dɛ

child (years;months:days)	syllable structure	word stress pattern	ipa-transkription
Ollie (0;9:1)	cv-cv-cv-cv-cv-cv-cv-CV	s-W-W-s-s-s-s-S	ne-ne-ne-ne-ne-ne-ne-ne
Ollie (0;9:11)	CV-cv-cv-cvc-cv-cv-CV	S-W-W-W-W-s-S	ŋe-dɛ-dɛj-jæ-jɛ-dɛ-ji
Ollie (0;9:1)	v-cv-cv-CV-cv-cv-cv-CV	W-s-W-S-W-s-s-S	ɛe-ne-ne-na-na-næ-næ-na
Ollie (0;9:11)	cv-cv-cvc-cv-cv-cv-cv-cv-cv-cvv-cv-V-cv	W-W-W-W-W-W-W-W-W-W-s-S-s	va-vɛ-væl-va-vɛ-va-vɛ-væ-vɛ-vaː-va-ɛ-vhɛ
Nils (0;11:23)	v-cv-cv-cv-CVC-cv-CVC	W-s-W-W-S-s-S	a-la-la-lɛ-lal-lal-lɛl
Ollie(1;0:12)	cv-v-cv-cv-cv-cv-cv-cv-CVV-cv-cv-v-cvv-cvv-CVC-cvc-cvc-cvc-cv-cvc-CV-cvc	W-s-W-s-W-W-W-W-S-W-W-W-W-s-S-s-s-s-W-s-S-W	nɛ-e-ne-ne-nɛ-nɛ-nɛ-nɛ-mel-mɛ-me-ɛ-nel-nal-nɛn-nɛn-nɛn-nɛn-nɛ-mɛn-nɛ-nɛn
Rike (1;0:24)	vv-CVC-cv-cv-cv-v-cv	W-S-s-W-W-W-s	ai-gad-dø-lø-lœ-œ-lœ
Emma (1;2:22)	cv-v-CV-cvc-cv-vc-cvc	s-s-S-s-W-s-s	ga-ɛ-ja-bɛm-vœ-av-dah
Emma (1;2:22)	v-CVV-CV-v-cvv-cv-cv	s-S-S-W-s-W	ɛ-ol-jo-a-laʊ-ve-lœ
Emma (1;2:22)	cvv-CVC-vc-v-V-cv-cv	s-W-S-s-W-S-W	daʊ-mɛm-ɛn-œ-a-la-ra
Emma (1;2:22)	CVC-cvc-cvc-cv-cvc-cvc-cv	S-s-W-s-s-W-W	kak-gak-dæt-da-gag-gag-ga
Emma (1;2:22)	CV-cv-cvv-v-cvc-ccv-cvc	S-s-W-W-s-s-W	bʏ-hɛ-laʊ-o-vɛn-djo-vax
Emma (1;2:22)	CV-CCV-vv-cv-ccv-ccv-cv	S-S-s-s-W-s	po-pvi-aʊ-pɪ-njo-bja-la
Emma (1;2:22)	v-CV-cv-V-cv-cvc-cvcc	W-S-W-S-W-W-W	ɛ-la-bɛ-a-pɛ-lab-ev-vɛlh
Hanna (1;3:11)	v-cv-CV-cv-cv-cv-v-cv	W-W-S-W-s-W-W-W	a-dœ-la-dɛ-la-lœ-a-la
Ollie (1;2:7)	CV-cv-CV-CV-CV-cv-CV	S-W-S-S-S-W-S	hæ-hæ-he-he-he-he-he
Ollie (1;2:7)	CV-cv-cv-cv-cv-cv-cv-cv	S-s-s-s-s-s-s-W	he-he-he-he-he-he-he
Ollie (1;2:7)	cv-CV-cv-cv-CV-CV-CV-cv	W-S-W-W-S-S-S-W	he-he-he-hæ-hæ-hæ-hæ-he
Emma (1;5:14)	CV-v-v-ccv-v-cv-cv-cv-V-cvc	S-W-W-W-W-W-W-S-W	nø-a-a-njo-ʏ-na-no-ja-ɛ-nɔn
Nils (1;5:28)	VV-v-v-VV-cv-VV-cv	S-W-W-S-W-S-W	aʊ-ə-o-aɪ-do-aɪ-do
Nils (1;5:28)	VV-cv-VV-cv-VV-cv-VV-cv	S-W-S-W-S-W-S-W	aɪ-to-aɪ-to-aɪ-to-aɪ-to
Nils (1;5:28)	V-cv-VV-cv-vv-cv-vv-cv-vc-cv	S-W-S-W-W-W-W-W-W-W	a-to-æɪ-to-aæ-to-aɪ-to-al-do
Emma (1;10:8)	CV-cv-CV-cv-cv-cv-cv-cvc-cv-cvc-cv-cvc-cvc-cv-cv	S-W-S-W-W-W-W-W-W-W-W-W-W-W-W	ba-ba-ma-ba-ma-ma-ma-mam-ma-bɔm-ma-mam-mam-mæ-ma

B Phonological development

Cross tabulations

Gutman scalings

Frequency count of syllable structures

Table B.1: Crosstabs for disyllabic utterances containing words (W) and babbles (B) between 12 and 22 months of age.

age	Stage	Statistics	WS	sS	SW	Ss	Spondee	comp_Ss	Observed
12	B	Count	18	9	45	32			104
		Row %	17.3%	8.7%	43.3%	30.8%			100.0%
		stand.Res.	0.6	0.5	-0.7	0.2			
	W	Count	0	0	13	4			17
		Row %	0.0%	0.0%	76.5%	23.5%			100.0%
		stand.Res.	-1.6	-1.1	1.7	-0.5			
15	B	Count	19	20	66	37	3		145
		Row %	13.1%	13.8%	45.5%	25.5%	2.1%		100.0%
		stand.Res.	1.3	0.7	-1.2	0.6	-0.4		
	W	Count	3	7	54	16	3		83
		Row %	3.6%	8.4%	65.1%	19.3%	3.6%		100.0%
		stand.Res.	-1.8	-0.9	1.6	-0.7	0.6		
18	B	Count	47	17	101	38	5	0	208
		Row %	22.6%	8.2%	48.6%	18.3%	2.4%	0.0%	100.0%
		stand.Res.	1.7	0.7	-2.3	2.3	0.9	-0.6	
	W	Count	49	21	234	31	4	1	340
		Row %	14.4%	6.2%	68.8%	9.1%	1.2%	0.3%	100.0%
		stand.Res.	-1.4	-0.5	1.8	-1.8	-0.7	0.5	
20	B	Count	2	7	2	3		0	14
		Row %	14.3%	50.0%	14.3%	21.4%		0.0%	100.0%
		stand.Res.	0.4	2.7	-2.0	0.6		-0.3	
	W	Count	11	15	61	15		1	103
		Row %	10.7%	14.6%	59.2%	14.6%		1.0%	100.0%
		stand.Res.	-0.1	-1.0	0.7	-0.2		0.1	
22	B	Count	7	0	18	2	0	0	27
		Row %	25.9%	0.0%	66.7%	7.4%	0.0%	0.0%	100.0%
		stand.Res.	0.4	-0.7	-0.2	0.3	-0.3	-0.4	
	W	Count	60	5	193	16	1	2	277
		Row %	21.7%	1.8%	69.7%	5.8%	0.4%	0.7%	100.0%
		stand.Res.	0.1	0.1				-0.1	

Table B.2: Crosstabs for trisyllabic utterances containing words (w) and babbles (b) for 12 and 15 months of age.

age	stage	statistics	SWW	WSW	sWS	Sss	SsS	SsW	SSs	SSS	SSW	SWs	SWS	sSs	sSS	sSW	ssS	WsS	observed
12	B	Count	2	3	3	1	1	6	1			6	3				6	3	35
		Row %	5.7%	8.6%	8.6%	2.9%	2.9%	17.1%	2.9%		17.1%	8.6%				17.1%		8.6%	100.0%
		stand.Res.	0.1	0.1	0.1	0.1	0.1	-0.2	0.1			-0.2	0.1			0.1		0.1	
	W	Count	0	0	0	0	0	1	0			1	0			0		0	2
		Row %	0.0%	0.0%	0.0%	0.0%	0.0%	50.0%	0.0%			50.0%	0.0%			0.0%		0.0%	100.0%
		stand.Res.	-0.3	-0.4	-0.4	-0.2	-0.2	1.0	-0.2			1.0	-0.4			-0.6		-0.4	
15	B	Count	13	5	3	4	1	7	1	1	1	3	2	4	2	9	0	1	57
		Row %	22.8%	8.8%	5.3%	7.0%	1.8%	12.3%	1.8%	1.8%	1.8%	5.3%	3.5%	7.0%	3.5%	15.8%	0.0%	1.8%	100.0%
		stand.Res.	0.0	0.2	0.1	0.1	0.1	-0.5	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.2	-1.0	0.1	
	W	Count	1	0	0	0	0	2	0	0	0	0	0	0	0	0	1	0	4
		Row %	25.0%	0.0%	0.0%	0.0%	0.0%	50.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	25.0%	0.0%	100.0%
		stand.Res.	0.1	-0.6	-0.4	-0.5	-0.3	1.8	-0.3	-0.3	-0.3	-0.4	-0.4	-0.5	-0.4	-0.8	3.6	-0.3	

Table B.3: Crosstabs for trisyllabic utterances containing words (w) and babbles (b) between 12 and 22 months of age.

age	stage	statistics	SWW	WSW	sWS	SsW	SSs	SSW	SWs	SWS	SWW	WSs	sSW	WsS	observed
18	B	Count	9	3	3	2	1	3	0	1	1	3	5	2	33
		Row %	27.3%	9.1%	9.1%	6.1%	3.0%	9.1%	0.0%	3.0%	3.0%	9.1%	15.2%	6.1%	100.0%
		stand.Res.	0.1	-1.0	-0.4	-0.1	0.6	0.2	-0.7	-0.8	0.6	0.6	1.0	0.9	
	W	Count	7	7	4	2	0	2	1	3	0	1	1	0	28
		Row %	25.0%	25.0%	14.3%	7.1%	0.0%	7.1%	3.6%	10.7%	0.0%	3.6%	3.6%	0.0%	100.0%
		stand.Res.	-0.1	1.1	0.4	0.1	-0.7	-0.2	0.8	0.9	-0.7	-0.6	-1.1	-1.0	
22	B	Count	1	1	1	1			0	1		0	0	0	5
		Row %	20.0%	20.0%	20.0%	20.0%			0.0%	20.0%		0.0%	0.0%	0.0%	100.0%
		stand.Res.	0.9	-0.6	1.8	0.0			-0.8	1.8		-0.4	-0.6	-0.4	
	W	Count	1	8	0	4			3	0		1	2	1	20
		Row %	5.0%	40.0%	0.0%	20.0%			15.0%	0.0%		5.0%	10.0%	5.0%	100.0%
		stand.Res.	-0.5	0.3	-0.9	0.0			0.4	-0.9		0.2	0.3	0.2	

Table B.4: Guttman scaling for syllable types of words depending on stress (stressed (1), unstressed (0), or secondary stress (2)) until 18 months of age

child	stress	cv	cv:	cvc	cv:c	v	v:	vc	v:c	ccv	ccv:	ccvc	ccv:c	cvcc	cv:cc	ccvcc	ccv:cc	vcc	v:cc
Bennie	1	+		+						+		+							
Bennie	0	+																	
Bennie	2	+		+															
Emma	1	+		+		+		+		+		+							
Emma	0	+		+		+		+											
Emma	2	+		+		+		+		+									
Hanna	1	+		+		+													
Hanna	0	+		+															
Hanna	2	+		+															
Nils	1	+	+	+		+		+		+									
Nils	0	+		+		+		+		+									
Nils	2	+						+											
Ollie	1	+		+															
Ollie	0	+																	
Ollie	2	+																	
Rike	1	+		+		+		+		+		+							
Rike	0	+		+		+													
Rike	2	+								+									

Table B.5: Guttman scaling for syllable types of words depending on stress (stressed (1), unstressed (0), or secondary stress (2)) until 24 months of age

child	stress	cv	cv:	cvc	cv:c	v	v:	vc	v:c	ccv	ccv:	ccvc	ccv:c	cvcc	cv:cc	ccvcc	ccv:cc	vcc	v:cc
Bennie	1	+	+	+	+	+				+		+							
Bennie	0	+	+	+	+	+				+									
Bennie	2	+	+		+														
Emma	1	+		+		+		+		+		+							
Emma	0	+		+		+		+											
Emma	2	+		+		+		+		+									
Hanna	1	+	+	+	+	+		+											
Hanna	0	+		+	+														
Hanna	2	+		+															
Nils	1	+	+	+		+		+		+									
Nils	0	+	+	+		+		+		+									
Nils	2	+						+											
Rike	1	+	+	+	+	+	+	+		+	+	+		+					
Rike	0	+	+	+	+	+		+		+		+		+					
Rike	2	+	+	+	+	+				+									

Table B.6: Guttman scaling for syllable types of words depending on stress (stressed (1), unstressed (0), or secondary stress (2)) until 30 months of age

child	stress	cv	cv:	cvc	cv:c	v	v:	vc	v:c	ccv	ccv:	ccvc	ccv:c	cvcc	cv:cc	ccvcc	ccv:cc	vcc	v:cc
Bennie	1	+	+	+	+	+	+	+		+		+		+					
Bennie	0	+	+	+	+	+	+	+		+		+		+					
Bennie	2	+	+	+		+		+		+	+			+					
Emma	1	+		+		+		+		+		+				+		+	
Emma	0	+		+		+		+											
Emma	2	+		+		+		+		+									
Hanna	1	+	+	+	+	+		+		+	+	+		+					
Hanna	0	+	+	+	+	+	+			+									
Hanna	2	+	+	+		+					+								
Nils	1	+	+	+	+	+		+		+	+	+		+					
Nils	0	+	+	+	+	+		+		+		+		+					
Nils	2	+	+	+	+			+											
Rike	1	+	+	+	+	+	+	+		+	+	+	+	+		+			
Rike	0	+	+	+	+	+	+	+	+	+	+	+		+	+				
Rike	2	+	+	+	+	+				+					+				

Table B.7: Guttman scaling for syllable types of words depending on stress (stressed (1), unstressed (0), or secondary stress (2)) until 36 months of age

child	stress	cv	cv:	cvc	cv:c	v	v:	vc	v:c	ccv	ccv:	ccvc	ccv:c	cvcc	cv:cc	ccvcc	ccv:cc	vcc	v:cc
Bennie	1	+	+	+	+	+	+	+	+	+	+	+	+	+	+			+	+
Bennie	0	+	+	+	+	+	+	+	+	+	+	+	+	+					
Bennie	2	+	+	+		+	+	+		+	+			+	+				
Hanna	1	+	+	+	+	+		+		+	+	+	+	+		+			
Hanna	0	+	+	+	+	+	+	+	+	+	+	+		+					
Hanna	2	+	+	+		+				+	+						+		
Nils	1	+	+	+	+	+		+		+	+	+	+	+	+				
Nils	0	+	+	+	+	+	+	+		+	+	+	+	+	+				
Nils	2	+	+	+	+	+		+						+	+				
Ollie	1	+	+	+	+			+											
Ollie	0	+	+	+	+	+													
Ollie	2	+	+	+															

Table B.8: Syllable structure and corresponding proportion of syllable tokens seperated by vowel length and pooled together.

Syllable (with short vowel)	**Tokens**	**Frequency** (in %)	**Syllable** (with long vowel)	**Tokens**	**Frequency** (in %)	**Total**	**Frequency** total (in %)
CCCVVCC	1	0,0037				1	0,0037
CVVCCCC	1	0,0037				1	0,0037
CCVCCCC	2	0,0074				2	0,0074
VVCCC	2	0,0074				2	0,0074
CCVVCCC	3	0,0111				3	0,0111
CVVCCC	5	0,0185				5	0,0185
CCCVCC	6	0,0222	CCCV:CC	1	0,0037	7	0,0259
VVCC	7	0,0259				7	0,0259
CCCVVC	9	0,0333				9	0,0333
CVCCCC	10	0,0370				10	0,0370
VCCC	15	0,0555	V:CCC	2	0,0074	17	0,0629
CCCVV	27	0,0999				27	0,0999
CCVCCC	26	0,0962	CCV:CCC	1	0,0037	27	0,0999
CCCVC	21	0,0777	CCCV:C	13	0,0481	34	0,1258
CCVVCC	36	0,1332				36	0,1332
CCCV	29	0,1073	CCCV:	27	0,0999	56	0,2072
CVVCC	72	0,2664				72	0,2664
CCVVC	121	0,4477				121	0,4477
CCVCC	171	0,6327	CCV:CC	35	0,1295	206	0,7622
CVCCC	240	0,8880	CV:CCC	5	0,0185	245	0,9065
CCVV	305	1,1285				305	1,1285
VV	330	1,2210				330	1,2210
CVVC	419	1,5504				419	1,5504
CVV	684	2,5309				684	2,5309
VVC	735	2,7196				735	2,7196
VCC	761	2,8158	V:CC	19	0,0703	780	2,8861
CCVC	598	2,2127	CCV:C	235	0,8695	833	3,0822
CVCC	824	3,0489	CV:CC	133	0,4921	957	3,5410
CCV	282	1,0434	CCV:	680	2,5161	962	3,5595
V	364	1,3469	V:	881	3,2598	1245	4,6067
VC	1576	5,8314	V:C	175	0,6475	1751	6,4789
CVC	6296	23,2961	CV:C	1314	4,8620	7610	28,1581
CV	5477	20,2657	CV:	4049	14,9819	9526	35,2475
	27026	100,0000				27026	100,0000

C Phonetic Development

Formant values

Children's use of voice quality

Adults' use of voice quality

Overview of parameters

Table C.1: Results of the ANOVAs for formant values of Emma, Bennie, and Rike.

child	**source**	**variable**	**degree of freedom**	**F Statistic**	**p Value**
Emma	stress	F1 (Bark)	2,7749	77.682	**0.000***
		F2 (Bark)	2,7749	3.487	**0.031***
		F3 (Bark)	2,7749	10.996	**0.000***
		F4 (Bark)	2,749	2.529	0.080
	age	F1 (Bark)	7,7749	33.650	**0.000***
		F2 (Bark)	7,7749	44.365	**0.000***
		F3 (Bark)	7,749	24.279	**0.000***
		F4 (Bark)	7,7749	15.890	**0.000***
	stress * age	F1 (Bark)	14,7749	2.717	**0.001***
		F2 (Bark)	14,.7749	3.228	**0.000***
		F3 (Bark)	14,749	2.757	**0.000***
		F4 (Bark)	14,7749	3.839	**0.000***
Bennie	stress	F1 (Bark)	2,804	56.388	**0.000***
		F2 (Bark)	2,04	9.301	**0.000***
		F3 (Bark)	2,804	11.658	**0.000***
		F4 (Bark)	2,804	2.690	0.068
	age	F1 (Bark)	7,8040	18.021	**0.000***
		F2 (Bark)	7,8040	17.822	**0.000***
		F3 (Bark)	7,8040	18.693	**0.000***
		F4 (Bark)	7,8040	21.152	**0.000***
	stress * age	F1 (Bark)	14,8040	9.595	**0.000***
		F2 (Bark)	14,8040	4.681	**0.000***
		F3 (Bark)	14,8040	12.240	**0.000***
		F4 (Bark)	14,8040	4.095	**0.000***
Rike	stress	F1 (Bark)	2,5402	22.772	**0.000***
		F2 (Bark)	2,.5402	16.372	**0.000***
		F3 (Bark)	2,402	45.472	**0.000***
		F4 (Bark)	2,5402	0.809	0.446
	age	F1 (Bark)	8,5402	26.509	**0.000***
		F2 (Bark)	8,5402	26.572	**0.000***
		F3 (Bark)	8,5402	12.569	**0.000***
		F4 (Bark)	8,5402	3.624	**0.000***
	stress * age	F1 (Bark)	14.5402	7.687	**0.000***
		F2 (Bark)	14,5402	5.756	**0.000***
		F3 (Bark)	14,5402	7.231	**0.000***
		F4 (Bark)	14,5402	4.208	**0.000***

Table C.2: Results of the ANOVAs for formant values of Hanna, Nils, and Ollie.

child	**source**	**variable**	**degree of freedom**	**F Statistic**	**p Value**
Hanna	stress	F1 (Bark)	2,4478	50.065	**0.000***
		F2 (Bark)	2,4478	19.685	**0.000***
		F3 (Bark)	2,4478	14.856	**0.000***
		F4 (Bark)	2,4478	6.530	**0.001***
	age	F1 (Bark)	8,4478	65.478	**0.000***
		F2 (Bark)	8,4478	40.942	**0.000***
		F3 (Bark)	8,.4478	13.899	**0.000***
		F4 (Bark)	8,478	28.822	**0.000***
	stress * age	F1 (Bark)	16,4478	8.585	**0.000***
		F2 (Bark)	16,4478	5.832	**0.000***
		F3 (Bark)	16,4478	4.445	**0.000***
		F4 (Bark)	16,4478	2.876	**0.000***
Nils	stress	F1 (Bark)	2,9110	33.250	**0.000***
		F2 (Bark)	2,9110	1.394	0.248
		F3 (Bark)	2,9110	12.802	**0.000***
		F4 (Bark)	2,9110	1.962	0.141
	age	F1 (Bark)	8,9110	294.492	**0.000***
		F2 (Bark)	8,9110	180.054	**0.000***
		F3 (Bark)	8,.9110	122.497	**0.000***
		F4 (Bark)	8,110	56.328	**0.000***
	stress * age	F1 (Bark)	14,9110	20.761	**0.000***
		F2 (Bark)	14,9110	11.304	**0.000***
		F3 (Bark)	14,9110	8.759	**0.000***
		F4 (Bark)	14,9110	6.070	**0.000***
Ollie	stress	F1 (Bark)	2,1742	19.939	**0.000***
		F2 (Bark)	2,1742	1.244	0.289
		F3 (Bark)	2,1742	6.331	**0.002***
		F4 (Bark)	2,1742	3.801	**0.023***
	age	F1 (Bark)	4,1742	45.126	**0.000***
		F2 (Bark)	4,1742	28.899	**0.000***
		F3 (Bark)	4,1742	37.823	**0.000***
		F4 (Bark)	4,1742	5.583	**0.000***
	stress * age	F1 (Bark)	8,1742	4.687	**0.000***
		F2 (Bark)	8,1742	5.049	**0.000***
		F3 (Bark)	8,1742	7.589	**0.000***
		F4 (Bark)	8,1742	0.667	0.721

Table C.3: Results of the ANOVAs for formant values for Emma's, Rike's, and Ollie's mother.

Parents	**source**	**variable**	**degree of freedom**	**F Statistic**	**p Value**
mother of Emma	stress	F1(Bark)	2,1095	0.308	0.735
		F2(Bark)	2,1095	8.653	**0.000***
		F3(Bark)	2,1095	0.840	0.432
		F4(Bark)	2,1095	2.113	0.121
	age of child	F1(Bark)	4,1095	12.778	**0.000***
		F2(Bark)	4,1095	1.849	0.117
		F3(Bark)	4,1095	5.459	**0.000***
		F4(Bark)	4,1095	13.243	**0.000***
	stress * age of child	F1(Bark)	7,1095	5.831	**0.000***
		F2(Bark)	7,1095	3.350	**0.002***
		F3(Bark)	7,.1095	1.396	0.203
		F4(Bark)	7,095	0.718	0.657
mother of Rike	stress	F1(Bark)	2,3123	1.692	0.184
		F2(Bark)	2,123	3.450	**0.032***
		F3(Bark)	2,3123	3.417	**0.033***
		F4(Bark)	2,3123	0.817	0.442
	age of child	F1(Bark)	7,3123	11.319	**0.000***
		F2(Bark)	7,3123	9.044	**0.000***
		F3(Bark)	7,3123	6.622	**0.000***
		F4(Bark)	7,3123	4.776	**0.000***
	stress * age of child	F1(Bark)	14,3123	8.485	**0.000***
		F2(Bark)	14,3123	4.832	**0.000***
		F3(Bark)	14,3123	3.937	**0.000***
		F4(Bark)	14,3123	2.293	**0.004***
mother of Ollie	stress	F1(Bark)	2,633	0.010	0.990
		F2(Bark)	2,633	13.409	**0.000***
		F3(Bark)	2,633	30.156	**0.000***
		F4(Bark)	2,633	2.984	0.051
	age of child	F1(Bark)	2,633	10.655	**0.000***
		F2(Bark)	2,633	0.092	0.912
		F3(Bark)	2,633	8.818	**0.000***
		F4(Bark)	2,633	5.964	**0.003***
	stress * age of child	F1(Bark)	4,633	7.429	**0.000***
		F2(Bark)	4,633	7.414	**0.000***
		F3(Bark)	4,633	17.770	**0.000***
		F4(Bark)	4,633	0.919	0.452

Table C.4: Results of the ANOVA for formant values for Hanna's and Nils's parents.

Parents	**source**	**variable**	**degree of freedom**	**F Statistic**	**p Value**
mother of Hanna	stress	F1(Bark)	2,563	24.579	**0.000***
		F2(Bark)	2,563	3.185	**0.042***
		F3(Bark)	2,563	8.338	**0.000***
		F4(Bark)	2,563	2.663	0.071
	age of child	F1(Bark)	3,563	22.548	**0.000***
		F2(Bark)	3,563	5.061	**0.002***
		F3(Bark)	3,563	15.990	**0.000***
		F4(Bark)	3,563	2.512	0.058
	stress * age of child	F1(Bark)	5,563	3.494	**0.004***
		F2(Bark)	5,563	3.782	**0.002***
		F3(Bark)	5,563	3.028	**0.010***
		F4(Bark)	5,563	1.264	0.278
father of Hanna	stress	F1(Bark)	2,215	10.581	**0.000***
		F2(Bark)	2,215	1.991	0.139
		F3(Bark)	2,215	6.005	**0.003***
		F4(Bark)	2,215	2.612	0.076
mother of Nils	stress	F1(Bark)	2,3974	3.822	**0.022***
		F2(Bark)	2,3974	24.188	**0.000***
		F3(Bark)	2,3974	8.356	**0.000***
		F4(Bark)	2,3974	7.130	**0.001***
	age of child	F1(Bark)	6,3974	28.275	**0.000***
		F2(Bark)	6,3974	13.858	**0.000***
		F3(Bark)	6,3974	4.634	**0.000***
		F4(Bark)	6,3974	7.585	**0.000***
	stress * age of child	F1(Bark)	12,3974	7.246	**0.000***
		F2(Bark)	12,3974	5.551	**0.000***
		F3(Bark)	12,3974	2.681	**0.001***
		F4(Bark)	12,974	2.706	**0.001***
father of Nils	stress	F1(Bark)	2,1247	23.850	**0.000***
		F2(Bark)	2,1247	2.855	0.058
		F3(Bark)	2,1247	1.985	0.138
		F4(Bark)	2,247	1.877	0.153
	age of child	F1(Bark)	2,1247	3.194	**0.041***
		F2(Bark)	2,.1247	11.338	**0.000***
		F3(Bark)	2,247	35.940	**0.000***
		F4(Bark)	2,1247	0.041	0.960
	stress * age of child	F1(Bark)	3,1247	0.351	0.788
		F2(Bark)	3,1247	4.336	**0.005***
		F3(Bark)	3,1247	9.991	**0.000***
		F4(Bark)	3,1247	1.459	0.224

Table C.5: Results of the ANOVA for formant values for Bennie's parents.

Parents	**source**	**variable**	**degree of freedom**	**F Statistic**	**p Value**
mother of Bennie	stress	F1(Bark)	2,4547	33.065	**0.000***
		F2(Bark)	2,4547	3.522	**0.030***
		F3(Bark)	2,4547	8.162	**0.000***
		F4(Bark)	2,547	4.385	**0.013***
	age of child	F1(Bark)	7,4547	28.051	**0.000***
		F2(Bark)	7,.4547	12.830	**0.000***
		F3(Bark)	7,547	11.508	**0.000***
		F4(Bark)	7,4547	5.023	**0.000***
	stress * age of child	F1(Bark)	13,4547	3.421	**0.000***
		F2(Bark)	13,547	2.613	**0.001***
		F3(Bark)	13,4547	7.781	**0.000***
		F4(Bark)	13,4547	7.164	**0.000***
father of Bennie	stress	F1(Bark)	2,2479	24.236	**0.000***
		F2(Bark)	2,2479	15.120	**0.000***
		F3(Bark)	2,2479	13.836	**0.000***
		F4(Bark)	2,2479	1.147	0.318
	age of child	F1(Bark)	4,2479	15.663	**0.000***
		F2(Bark)	4,2479	22.339	**0.000***
		F3(Bark)	4,2479	5.159	**0.000***
		F4(Bark)	4,2479	18.132	**0.000***
	stress * age of child	F1(Bark)	8,2479	1.765	0.079
		F2(Bark)	8,2479	2.170	**0.027***
		F3(Bark)	8,2479	3.920	**0.000***
		F4(Bark)	8,2479	6.753	**0.000***

Table C.6: ANOVA results for **Bennie** for the voice quality parameters for raw and normalised data (Z).

source	**variable**	**degree of freedom**	**F Statistic**	**p Value**
stress	OQG	2,5316	2.835	0.059
	ZOQG	2,5316	2.140	0.118
age	OQG	7,5316	16.804	**0.000***
	ZOQG	7,5316	0.426	0.886
stress*age	OQG	14,5316	7.808	**0.000***
	ZOQG	14,5316	1.445	0.123
stress	GOG	2,3879	0.301	0.740
	ZGOG	2,3879	7.415	**0.001***
age	GOG	7,3879	28.608	**0.000***
	ZGOG	7,3879	28.962	**0.000***
stress*age	GOG	14,3879	5.053	**0.000***
	ZGOG	14,3879	6.535	**0.000***
stress	SKG	2,4591	3.008	0.050
	ZSKG	2,4591	0.837	0.433
age	SKG	7,4591	15.387	**0.000***
	ZSKG	7,4591	1.165	0.319
stress*age	SKG	14,4591	6.411	**0.000***
	ZSKG	14,4591	3.008	**0.000***
stress	RCG	2,4591	4.971	**0.007***
	ZRCG	2,4591	1.141	0.320
age	RCG	7,4591	9.418	**0.000***
	ZRCG	7,4591	1.541	0.148
stress*age	RCG	14,4591	6.734	**0.000***
	ZRCG	14,4591	2.102	**0.009***
stress	T4G	2,4591	2.740	0.065
	ZT4G	2,4591	0.330	0.719
age	T4G	7,4591	25.027	**0.000***
	ZT4G	7,4591	1.689	0.107
stress*age	T4G	14,4591	5.959	**0.000***
	ZT4G	14,4591	3.581	**0.000***
stress	CC	2,4591	9.755	**0.000***
	ZCC	2,4591	2.898	0.055
age	CC	7,4591	5.469	**0.000***
	ZCC	7,4591	0.888	0.515
stress*age	CC	14,4591	7.102	**0.000***
	ZCC	14,4591	3.407	**0.000***
stress	IC	2,4591	8.469	**0.000***
	ZIC	2,4591	3.446	**0.032***
age	IC	7,4591	7.418	**0.000***
	ZIC	7,4591	0.372	0.919
stress*age	IC	14,4591	3.519	**0.000***
	ZIC	14,4591	2.530	**0.001***

Table C.7: ANOVA results for **Emma** for the voice quality parameters for raw and normalised data (Z).

source	variable	degree of freedom	F Statistic	p Value
stress	OQG	2,3898	1.478	0.228
	ZOQG	2,3898	0.970	0.379
age	OQG	4,3898	6.594	**0.000***
	ZOQG	4,3898	1.082	0.364
stress*age	OQG	8,3898	2.694	**0.006***
	ZOQG	8,3898	2.494	**0.011***
stress	GOG	2,2960	0.575	0.563
	ZGOG	2,2960	10.656	**0.000***
age	GOG	4,2960	1.280	0.275
	ZGOG	4,2960	5.056	**0.000***
stress*age	GOG	8,2960	2.328	**0.017***
	ZGOG	8,960	11.035	**0.000***
stress	SKG	2,3955	0.294	0.745
	ZSKG	2,3955	0.663	0.515
age	SKG	4,3955	0.418	0.796
	ZSKG	4,3955	1.874	0.112
stress*age	SKG	8,3955	0.938	0.484
	ZSKG	8,3955	3.601	**0.000***
stress	RCG	2,3955	0.856	0.425
	ZRCG	2,3955	0.226	0.798
age	RCG	4.3955	3.851	**0.004***
	ZRCG	4,955	1.181	0.317
stress*age	RCG	8,3955	6.703	**0.000***
	ZRCG	8,3955	3.052	**0.002***
stress	T4G	2,3955	1.717	0.180
	ZT4G	2,3955	0.203	0.817
age	T4G	4,3955	10.197	**0.000***
	ZT4G	4,3955	2.028	0.088
stress*age	T4G	8,3955	6.855	**0.000***
	ZT4G	8,3955	3.352	**0.001***
stress	CC	2,3955	0.404	0.667
	ZCC	2,3955	0.961	0.383
age	CC	4,3955	10.738	**0.000***
	ZCC	4,3955	0.438	0.781
stress*age	CC	8,3955	2.006	**0.042***
	ZCC	8,3955	1.014	0.422
stress	IC	2,3955	0.045	0.956
	ZIC	2,3955	1.077	0.341
age	IC	4,3955	21.491	**0.000***
	ZIC	4,3955	0.751	0.557
stress*age	IC	8,3955	2.575	**0.008***
	ZIC	8,3955	1.091	0.366

Table C.8: ANOVA results for **Rike** for the voice quality parameters for raw and normalised data (Z).

source	**variable**	**degree of freedom**	**F Statistic**	**p Value**
stress	OQG	2,4377	11.417	**0.000***
	ZOQG	2,4377	4.967	**0.007***
age	OQG	8,4377	24.745	**0.000***
	ZOQG	8,4377	2.112	**0.031***
stress*age	OQG	13,4377	12.608	**0.000***
	ZOQG	13,4377	4.115	**0.000***
stress	GOG	2,2685	0.987	0.373
	ZGOG	2,2685	15.399	**0.000***
age	GOG	8,2685	10.478	**0.000***
	ZGOG	8,2685	9.478	**0.000***
stress*age	GOG	13,2685	2.811	**0.001***
	ZGOG	12,2685	5.186	**0.000***
stress	SKG	2,3555	0.686	0.503
	ZSKG	2,3555	5.414	**0.004***
age	SKG	8,3555	0.567	0.806
	ZSKG	8,3555	0.445	0.895
stress*age	SKG	13,3555	1.148	0.313
	ZSKG	13,3555	4.525	**0.000***
stress	RCG	2,3555	21.411	**0.000***
	ZRCG	2,3555	9.862	**0.000***
age	RCG	8,3555	20.383	**0.000***
	ZRCG	8,3555	0.481	0.871
stress*age	RCG	13,3555	7.025	**0.000***
	ZRCG	13,3555	4.036	**0.000***
stress	T4G	2,3555	24.909	**0.000***
	ZT4G	2,3555	10.143	**0.000***
age	T4G	8,3555	17.074	**0.000***
	ZT4G	8,3555	0.520	0.843
stress*age	T4G	13,3555	5.145	**0.000***
	ZT4G	13,3555	4.320	**0.000***
stress	CC	2,3555	1.716	0.180
	ZCC	2,3555	2.174	0.114
age	CC	8,3555	11.687	**0.000***
	ZCC	8,3555	0.441	0.897
stress*age	CC	13,3555	3.082	**0.000***
	ZCC	13,3555	1.685	0.057
stress	IC	2,3555	2.717	0.066
	ZIC	2,3555	2.751	0.064
age	IC	8,3555	8.306	**0.000***
	ZIC	8,3555	0.572	0.801
stress*age	IC	13,3555	2.596	**0.001***
	ZIC	13,3555	1.948	**0.021***

Table C.9: ANOVA results for **Hanna** for the voice quality parameters for raw and normalised data (Z).

source	**variable**	**degree of freedom**	**F Statistic**	**p Value**
stress	OQG	2.2539	2.570	0.077
	ZOG	2,2538	2.133	0.119
age	OQG	6,2538	9.909	**0.000***
	ZOG	6,2538	1.645	0.131
stress*age	OQG	12,2538	2.373	**0.005***
	ZOG	12,2538	1.741	0.053
stress	GOG	2,1905	3.725	**0.024***
	ZGOG	2,1905	3.199	**0.041***
age	GOG	6,1905	9.855	**0.000***
	ZGOG	6,1905	3.505	**0.002***
stress*age	GOG	12,1905	2.927	**0.000***
	ZGOG	23,2905	2.867	**0.034***
stress	SKG	2,2794	1.276	0.279
	ZSKG	2,2794	0.843	0.430
age	SKG	6,2794	0.712	0.640
	ZSKG	6,2794	1.528	0.165
stress*age	SKG	12,1794	1.455	0.134
	ZSKG	12,1794	1.710	0.058
stress	RC	2,2794	1.057	0.348
	ZRC	2,2794	0.605	0.546
age	RCG	6,2794	32.463	**0.000***
	ZRCG	6,2794	1.221	0.292
stress*age	RCG	12,1794	3.3337	**0.000***
	ZRCG	12,1794	1.107	0.350
stress	T4G	2,2794	34.139	**0.000***
	ZT4G	2,2794	1.937	0.144
age	T4G	6,2794	34.139	**0.000***
	ZT4G	6,2794	1.939	0.071
stress*age	T4G	12,1794	2.984	**0.000***
	ZT4G	12,1794	1.900	**0.030***
stress	CC	2,2794	11.781	**0.000***
	ZCC	2,2794	5.879	**0.003***
age	CC	6,2794	0.846	0.535
	ZCC	6,2794	1.165	0.322
stress*age	CC	12,2794	1.883	**0.032***
	ZCC	12,2794	1.958	**0.024***
stress	IC	2,2794	17.499	**0.000***
	ZIC	2,2794	6.328	**0.002***
age	IC	6,2794	19.166	**0.000***
	ZIC	6,2794	1.598	0.144
stress*age	IC	12,2794	5.344	**0.000***
	ZIC	12,2794	3.267	**0.000***

Table C.10: ANOVA results for **Nils** for the voice quality parameters for raw and normalised data (Z).

source	**variable**	**degree of freedom**	**F Statistic**	**p Value**
stress	OQG	2,3496	5.808	**0.003***
	ZOQG	2,3496	1.813	0.163
age	OQG	6,3496	3.006	**0.006***
	ZOQG	6,3496	1.247	0.279
stress*age	OQG	10,3496	9.467	**0.000***
	ZOQG	10,3496	3.450	**0.000***
stress	GOG	2,1785	2.531	0.080
	ZGOG	2,1785	3.562	**0.029***
age	GOG	6,1785	7.349	**0.000***
	ZGOG	6,1785	12.385	**0.000***
stress*age	GOG	10,1785	5.551	**0.000***
	ZGOG	10,1785	0.328	0.974
stress	SKG	2,4040	0.291	0.748
	ZSKG	2,4040	11.660	**0.000***
age	SKG	6,4040	1.886	0.079
	ZSKG	6,4040	0.846	0.534
stress*age	SKG	10,4040	0.813	0.617
	ZSKG	10,4040	6.436	**0.000***
stress	RCG	2,4040	29.648	**0.000***
	ZRCG	2,4040	21.917	**0.000***
age	RCG	6,4040	81.801	**0.000***
	ZRCG	6,4040	0.622	0.713
stress*age	RCG	10,4040	8.078	**0.000***
	ZRCG	10,4040	3.034	**0.001***
stress	T4G	2,4040	52.326	**0.000***
	ZT4G	2,4040	25.449	**0.000***
age	T4G	6,4040	65.332	**0.000***
	ZT4G	6,4040	0.391	0.885
stress*age	T4G	10,4040	6.719	**0.000***
	ZT4G	10,4040	2.607	**0.004***
stress	CC	2,4040	12.822	**0.000***
	ZCC	2,4040	1.704	0.182
age	CC	6,4040	55.162	**0.000***
	ZCC	6,4040	0.140	0.991
stress*age	CC	10,4040	27.233	**0.000***
	ZCC	10,4040	5.295	**0.000***
stress	IC	2,4040	0.300	0.741
	ZIC	2,4040	3.622	**0.027***
age	IC	6,4040	2.206	**0.040***
	ZIC	6,4040	0.246	0.961
stress*age	IC	10,4040	0.627	0.792
	ZIC	10,4040	1.257	0.249

Table C.11: ANOVA results for **Ollie** for the voice quality parameters for raw and normalised data (Z).

source	variable	degree of freedom	F Statistic	p Value
stress	OQG	2,997	3.187	**0.042***
	ZOQG	2,997	6.276	**0.002***
age	OQG	2,997	8.178	**0.000***
	ZOQG	2,997	3.047	**0.048***
stress*age	OQG	4,997	0.429	0.788
	ZOQG	4,997	1.632	0.164
stress	GOG	2,796	4.874	**0.008***
	ZGOG	2,796	1.774	0.170
age	GOG	2,796	8.144	**0.000***
	ZGOG	2,796	9.805	**0.000***
stress*age	GOG	4,796	0.471	0.757
	ZGOG	4,796	8.387	**0.000***
stress	SKG	2,1122	0.416	0.660
	ZSKG	2,1122	13.253	**0.000***
age	SKG	2,1122	0.441	0.644
	ZSKG	2,1122	0.292	0.747
stress*age	SKG	4,1122	0.521	0.720
	ZSKG	4,1122	3.804	**0.004***
stress	RCG	2,1122	15.146	**0.000***
	ZRCG	2,1122	14.982	**0.000***
age	RCG	2,1122	8.293	**0.000***
	ZRCG	2,1122	0.448	0.639
stress*age	RCG	4,1122	1.814	0.124
	ZRCG	4,1122	3.188	**0.013***
stress	T4G	2,1122	17.583	**0.000***
	ZT4G	2,1122	17.209	**0.000***
age	T4G	2,1122	3.872	**0.021***
	ZT4G	2,1122	0.695	0.499
stress*age	T4G	4,1122	2.389	**0.049***
	ZT4G	4,1122	2.026	0.089
stress	CC	2,1122	0.096	0.909
	ZCC	2,1122	0.973	0.378
age	CC	2,1122	1.202	0.301
	ZCC	2,1122	0.225	0.799
stress*age	CC	4,1122	4.168	**0.002***
	ZCC	4,1122	6.900	**0.000***
stress	IC	2,1122	1.877	0.154
	ZIC	2,1122	0.698	0.498
age	IC	2,1122	0.848	0.428
	ZIC	2,1122	0.227	0.797
stress*age	IC	4,1122	2.807	**0.025***
	ZIC	4,1122	5.551	**0.000***

Table C.12: ANOVA results for Emma's, Ollie's and Rike's mother for the normalised voice quality parameters.

Parents	source	variable	degree of freedom	F Statistic	p Value
mother of Emma	stress	ZOQG	2,423	1.176	0.310
	age of child	ZOQG	2,423	0.320	0.726
	stress*age of child	ZOQG	3,423	1.077	0.358
	stress	ZGOG	2,515	1.116	0.328
	age of child	ZGOG	2,515	1.276	0.280
	stress*age of child	ZGOG	3,515	1.277	0.282
	stress	ZSKG	2,684	0.060	0.942
	age of child	ZSKG	2,684	0.020	0.980
	stress*age of child	ZSKG	3,684	1.832	0.140
	stress	ZRCG	2,684	5.997	**0.003***
	age of child	ZRCG	2,684	4.867	**0.008***
	stress*age of child	ZRCG	3,685	7.072	**0.000***
	stress	ZT4G	2,684	0.861	0.423
	age of child	ZT4G	2,684	3.218	**0.041***
	stress*age of child	ZT4G	3,688	7.165	**0.000***
	stress	ZCC	2,684	6.177	**0.002***
	age of child	ZCC	2,684	0.570	0.566
	stress*age of child	ZCC	3,686	0.709	0.547
	stress	ZIC	2,684	2.441	0.088
	age of child	ZIC	2,684	4.698	**0.009***
	stress*age of child	ZIC	3,687	5.631	**0.001***
mother of Ollie	stress	ZOQG	2,480	1.477	0.229
	stress	ZGOG	2,512	11.906	**0.000***
	stress	ZSKG	2.677	5.550	**0.004***
	stress	ZRCG	2.677	4.377	**0.013***
	stress	ZCC	2.677	5.239	**0.006***
	stress	ZIC	2.677	7.508	**0.001***
	stress	ZT4G	2.677	4.817	**0.008***
mother of Rike	stress	ZOQG	2,1134	0.395	0.674
	age of child	ZOQG	7,1134	2.589	**0.012***
	stress*age of child	ZOQG	14,1134	2.313	**0.004***
	stress	ZGOG	2,813	1.808	0.165
	age of child	ZGOG	7,813	2.983	**0.004***
	stress*age of child	ZGOG	14,813	3.460	**0.000***
	stress	ZSKG	2,1418	3.660	**0.026***
	age of child	ZSKG	7,1418	0.436	0.880
	stress*age of child	ZSKG	14,1418	1.682	0.053
	stress	ZRCG	2,1418	3.523	**0.030***
	age of child	ZRCG	7,1418	0.359	0.926
	stress*age of child	ZRCG	14,1418	1.304	0.197
	stress	ZT4G	2,1418	1.009	0.365
	age of child	ZT4G	7,1418	0.195	0.987
	stress*age of child	ZT4G	14,1418	1.023	0.427
	stress	ZCC	2,1418	1.700	0.183
	age of child	ZCC	7,1418	1.654	0.116
	stress*age of child	ZCC	14,1418	1.164	0.297
	stress	ZIC	2,1418	2.542	0.079
	age of child	ZIC	7,1418	1.655	0.116
	stress*age of child	ZIC	14,1418	1.412	0.139

Table C.13: ANOVA results of Hanna's parents for the normalised voice quality parameters.

Parents	**source**	**variable**	**degree of freedom**	**F Statistic**	**p Value**
mother of Hanna	stress	ZOQG	2,513	0.307	0.736
	age of child	ZOQG	3,513	3.347	**0.019***
	stress*age of child	ZOQG	4,513	2.943	**0.020***
	stress	ZGOG	2,429	12.838	**0.000***
	age of child	ZGIG	3,429	11.374	**0.000***
	stress*age of child	ZGOG	4,429	6.754	**0.000***
	stress	ZSKG	2,570	4.360	**0.013***
	age of child	ZSKG	3,570	10.395	**0.000***
	stress*age of child	ZSKG	4,570	1.218	0.302
	stress	ZRCG	2,570	3.921	**0.020***
	age of child	ZRCG	3,570	22.484	**0.000***
	stress*age of child	ZRCG	4,570	5.602	**0.000***
	stress	ZT4G	2,570	1.970	0.140
	age of child	ZT4G	3,570	11.159	**0.000***
	stress*age of child	ZT4G	4,570	2.523	**0.040***
	stress	ZCC	2,570	11.628	**0.000***
	age of child	ZCC	3,570	17.619	**0.000***
	stress*age of child	ZCC	4,570	4.509	**0.001***
	stress	ZIC	2,570	10.818	**0.000***
	age of child	ZIC	3,570	20.928	**0.000***
	stress*age of child	ZIC	4,570	7.230	**0.000***
father of Hanna	stress	ZOQG	2,82	4.316	**0.017***
	stress	ZGOG	2,155	2.108	0.125
	stress	ZSKG	2,162	1.999	0.139
	stress	ZRCG	2,162	2.844	0.061
	stress	ZT4G	2,162	1.455	0.236
	stress	ZCC	2,162	6.611	**0.002***
	stress	ZIC	2,162	6.105	**0.003***

Table C.14: ANOVA results of Nils's parents for the normalised voice quality parameters.

Parents	source	variable	degree of freedom	F Statistic	p Value
father of Nils	stress	ZOQG	2,799	5.256	**0.005***
	age of child	ZOQG	2,799	2.765	0.064
	stress*age of child	ZOQG	3,799	0.342	0.795
	stress	ZGOG	2,967	0.180	0.835
	age of child	ZGOG	2,967	5.885	**0.003***
	stress*age of child	ZGOG	2,967	3.325	**0.036***
	stress	ZSKG	2,1095	2.031	0.132
	age of child	ZSKG	2,1095	2.163	0.115
	stress*age of child	ZSKG	3,1095	2.234	0.083
	stress	ZRCG	2,1095	1.352	0.259
	age of child	ZRCG	2,1095	2.904	0.055
	stress*age of child	ZRCG	3,1095	3.892	**0.009***
	stress	ZT4G	2,1095	1.674	0.188
	age of child	ZT4G	2,1095	1.764	0.172
	stress*age of child	ZT4G	3,1095	2.176	0.089
	stress	ZCC	2,1095	0.495	0.610
	age of child	ZCC	2,1095	0.058	0.944
	stress*age of child	ZCC	3,1095	0.451	0.717
	stress	ZIC	2,1095	2.046	0.130
	age of child	ZIC	2,1095	0.084	0.919
	stress*age of child	ZIC	3,1095	0.482	0.695
mother of Nils	stress	ZOQG	2,930	7.822	**0.000***
	age of child	ZOQG	5,930	0.766	0.574
	stress*age of child	ZOQG	10,930	1.239	0.262
	stress	ZGOG	2,753	12.339	**0.000***
	age of child	ZGOG	5,753	1.367	0.234
	stress*age of child	ZGOG	9,753	6.378	**0.000***
	stress	ZSKG	2,1817	1.801	0.165
	age of child	ZSKG	6,1817	0.822	0.553
	stress*age of child	ZSKG	12,1817	2.320	**0.006***
	stress	ZRCG	2,1817	0.521	0.594
	age of child	ZRCG	6,1817	0.552	0.769
	stress*age of child	ZRCG	12,1817	2.398	**0.004***
	stress	ZT4G	2,1817	1.546	0.213
	age of child	ZT4G	6,1817	0.858	0.525
	stress*age of child	ZT4G	12,1817	2.205	**0.010***
	stress	ZCC	2,1817	3.061	**0.047***
	age of child	ZCC	6,1817	0.729	0.626
	stress*age of child	ZCC	12,1817	2.087	**0.015***
	stress	ZIC	2,1817	2.096	0.123
	age of child	ZIC	6,1817	0.417	0.868
	stress*age of child	ZIC	12,1817	1.497	0.118

Table C.15: ANOVA results of Bennie's parents for the normalised voice quality parameters.

Parents	source	variable	degree of freedom	F Statistic	p Value
mother of Bennie	stress	ZOQG	2,1218	3.956	**0.019***
	age of child	ZOQG	6,1218	3.526	**0.002***
	stress*age of child	ZOQG	11,1218	2.257	**0.010***
	stress	ZGOG	2,1409	0.786	0.456
	age of child	ZGOG	6,1409	0.614	0.719
	stress*age of child	ZGOG	9,1409	2.138	**0.024***
	stress	ZSKG	2,1884	1.189	0.305
	age of child	ZSKG	7,1884	0.477	0.852
	stress*age of child	ZSKG	13,1884	1.435	0.135
	stres	ZRCG	2,1884	1.688	0.185
	age of child	ZRCG	7,1884	0.700	0.672
	stress*age of child	ZRCG	13,1884	1.700	0.055
	stres	ZT4G	2,1884	3.047	**0.048***
	age of child	ZT4G	7,1884	0.902	0.504
	stress*age of child	ZT4G	13,1884	1.484	0.116
	stres	ZCC	2,1884	0.262	0.770
	age of child	ZCC	7,1884	0.855	0.542
	stress*age of child	ZCC	13,1884	1.790	0.039
	stres	ZIC	2,1884	1.100	0.333
	age of child	ZIC	7,1884	0.688	0.682
	stress*age of child	ZIC	13,1884	1.421	0.142
father of Bennie	stress	ZOQG	2,172	1.399	0.250
	age of child	ZOQG	2,172	4.163	**0.017***
	stress*age of child	ZOQG	4,172	2.051	0.089
	stress	ZGOG	2,778	5.036	**0.007***
	age of child	ZGOG	3,778	0.544	0.652
	stress*age of child	ZGOG	4,778	2.417	**0.047***
	stress	ZSKG	2,995	0.083	0.921
	age of child	ZSKG	4,995	0.425	0.791
	stress*age of child	ZSKG	5,995	0.939	0.454
	stress	ZRCG	2,995	5.711	**0.003***
	age of child	ZRCG	4,995	0.351	0.843
	stress*age of child	ZRCG	5,995	1.055	0.384
	stress	ZT4G	2,995	0.948	0.388
	age of child	ZT4G	4,995	0.190	0.944
	stress*age of child	ZT4G	5,995	1.200	0.307
	stress	ZCC	2,995	1.269	0.282
	age of child	ZCC	4,995	0.117	0.977
	stress*age of child	ZCC	5,995	1.146	0.334
	stress	ZIC	2,995	5.776	**0.003***
	age of child	ZIC	4,995	0.163	0.957
	stress*age of child	ZIC	5,995	0.735	0.597

Table C.16: Overview of the MANOVA results of the different parameters to mark stress for Hanna and her parents. Significance is marked with (*), no significance with (–). If a clear tendency were observed a main use of the parameters to differ stress is given (Tamhane post hoc). It is described wether higher (>) or lower (<) values were observed.

vp	source	duration	RMS Intensity	F0	OQ	GO	SK	RC	T4	CC	IC
Hanna	stress	*	*	*	-	*	-	-	-	*	*
	age	-	*	*	-	*	-	-	-	-	-
	position	-	-								
	stress*age	-	*	*	-	*	-	-	*	*	*
	stress*position	-	-								
	stress*age*position	-	*								
	Tamhane	variable	0<1,2	2<0<1	1<0	0,1<2	1,2<0	1,2<0	1,2<0	1,0<2	1,0<2
mother of Hanna	stress	*	*	*	-	*	*	*	-	*	*
	age	-	*	*	*	*	*	*	*	*	*
	position	-	*								
	stress*age	-	*	*	*	*	*	*	*	*	*
	stress*position	-	*								
	stress*age*position	-	-								
	Tamhane	0<1	0<1<2	0,2<1		2<0,1	2,1<0	2,1<0	2,1<0	2<=1<=0	2<=1<=0
father of Hanna	stress	*	*	-	*	-	-	-	-	*	*
	age	-	-	-	-	-	-	-	-	-	-
	position	-	*								
	stress*age	-	na	-	-	*	-	*	*	*	-
	stress*position	*	-								
	stress*age*position	-	na								
	Tamhane	0<1	0<1,2		2<=0<=1					2,1<0	2,1<0

Table C.17: Overview of the MANOVA results of the different parameters to mark stress for Nils and his parents. Significance is marked with (*), no significance with (–). If a clear tendency were observed a main use of the parameters to differ stress is given (Tamhane post hoc). It is described wether higher (>) or lower (<) values were observed.

vp	source	duration	RMS Intensity	F0	OQ	GO	SK	RC	T4	CC	IC
Nils	stress	*	*	*	-	*	*	*	*	-	-
	age	*	*	*	-	*	-	-	-	-	-
	position	*	*								
	stress*age	*	*	*	*	*	*	*	*	-	-
	stress*position	-	*								
	stress*age*position	*	*								
	Tamhane	variable	0<2<1	0<1<2	1,2<0	1<0	1<0,2	1<0,2	1<0,2	2<0,1	0,1<2
mother of Nils	stress	*	*	*	*	*	-	-	-	*	-
	age	-	*	*	-	-	-	-	-	-	-
	position	*	-								
	stress*age-	*	*	-	*	-	*	*	*	-	
	stress*position	*	-								
	stress*age*position	-	*								
	Tamhane	0<1,2	2<0<1	2<0<1	2<1,0	1,0<2	CDS:2<1	CDS:2<1	CDS:2<1	1,0<2	1,0<2
father of Nils	stress	*	*	*	*	-	-	-	-	-	-
	age	-	-	*	-	*	-	-	-	-	-
	position	*	-								
	stress*age	-	*	*	-	*	-	*	-	-	-
	stress*position	*	-								
	stress*age*position	-	-								
	Tamhane	0<1,2	0<1	0,2<1	2<1,0	ADS:2<0,1		CDS:1,0<2			

Table C.18: Overview of the MANOVA results of the different parameters to mark stress for Bennie and his parents. Significance is marked with (*), no significance with (–). If a clear tendency were observed a main use of the parameters to differ stress is given (Tamhane post hoc). It is described wether higher (>) or lower (<) values were observed.

vp	source	duration	RMS Intensity	F0	OQ	GO	SK	RC	T4	CC	IC
Bennie	stress	*	*	*	-	*	-	-	-	-	*
	age	-	*	*	-	*	-	-	-	-	-
	position	*	-								
	stress*age	*	*	*	-	*	*	*	*	*	*
	stress*position	-	*								
	stress*age*position	*	*								
	Tamhane	0<1	0,2<1	0<2<1	1<0,2	1,0<=2	1,2<0	1,2<0	1,2<0	1,0<2	2<0,1
mother of Bennie	stress	*	-	*	*	-	-	-	-	-	-
	age	-	-	*	*	-	-	-	-	-	-
	position	-	-								
	stress*age	-	-	*	*	*	-	-	-	*	-
	stress*position	-	-								
	stress*age*position	-	-								
	Tamhane	0<1,2		0,2<1	2<1,0	1,0<2					
father of Bennie	stress	*	*	*	-	*	-	*	-	-	*
	age	-	-	*	*	-	-	-	-	-	-
	position	-	-								
	stress*age	-	*	*	-	*	-	-	-	-	-
	stress*position	-	-								
	stress*age*position	-	-								
	Tamhane	0<2<1	0,2<1	0<2<1		1,0<2		1,2<0			

Table C.19: Overview of the MANOVA results for Rike and her mother

vp	source	duration	RMS Intensity	F0	OQ	GO	SK	RC	T4	CC	IC
Rike	stress	*	*	*	*	*	*	*	*	-	-
	age	-	*	*	*	*	-	-	-	-	-
	position	*	*								
	stress*age	*	*	*	*	*	*	*	*	-	*
	stress*position	-	*								
stress*age*position	*	*									
	Tamhane	variable	0<2<1	0<2<1		2<0,1	1<=2<=0	1<=2<=0	1<=2<=0	2<0,1	
mother of Rike	stress	*	-	*	-	-	*	*	-	-	-
	age	-	*	*	*	*	-	-	-	-	-
	position	-	*								
	stress*age	-	*	*	*	*	-	-	-	-	-
	stress*position	*	-								
	stress*age*position	-	*								
	Tamhane	0<1	1,0<2	2<0<1		1<0,2	1<=2<=0		1,0<2	1,0<2	

Table C.20: Overview of the MANOVA results for Emma and her mother

vp	source	duration	RMS Intensity	F0	OQ	GO	SK	RC	T4	CC	IC
Emma	stress	*	*	-	-	*	-	-	-	-	-
	age	-	*	*	-	*	-	-	-	-	-
	position	*	*								
	stress*age	-	*	*	*	*	*	*	*	-	-
	stress*position	-	*								
	stress*age*position	-	*								
	Tamhane	0<1,2	0<1<2	0<1,2		1,0<2	1<0	1<0	1<0		
mother of Emma	stress	*	*	*	-	-	-	*	-	*	-
	age	-	-	*	-	-	-	*	*	-	*
	position	-	-								
	stress*age	-	*	*	-	-	-	*	*	-	*
	stress*position	*	-								
	stress*age*position	-	*								
	Tamhane	0<1	0<1	0,2<1		1,0<2		1<0,2	1<0,2	1,0<2	2<0,1

Table C.21: Overview of the MANOVA results for Ollie and his mother

vp	source	duration	RMS Intensity	F0	OQ	GO	SK	RC	T4	CC	IC
Ollie	stress	*	*	*	*	-	*	*	*	-	-
	age	-	*	*	*	*	-	-	-	-	-
	position	*	-								
	stress*age	*	*	*	-	*	*	*	-	*	*
	stress*position	-	-								
	stress*age*position	-	*								
	Tamhane	variable	0,2<1	0<2<1		1,0<2					
mother of Ollie	stress	*	-	*	-	*	*	*	*	*	*
	age	*	*	*	-	-	-	-	-	-	-
	position	-	*								
	stress*age	-	*	*	-	-	-	-	-	-	-
	stress*position	-	*								
	stress*age*position	-	-								
	Tamhane	0<1	0<2,1	0<1,2		1,0<2	1,0<2	1,0<2	1<=0<=2	1,2<0	1,2<0

D Development of vowel space

Dispersion ellipses

Figure D.1: Colour code for the different vowels.

Individual vowel development

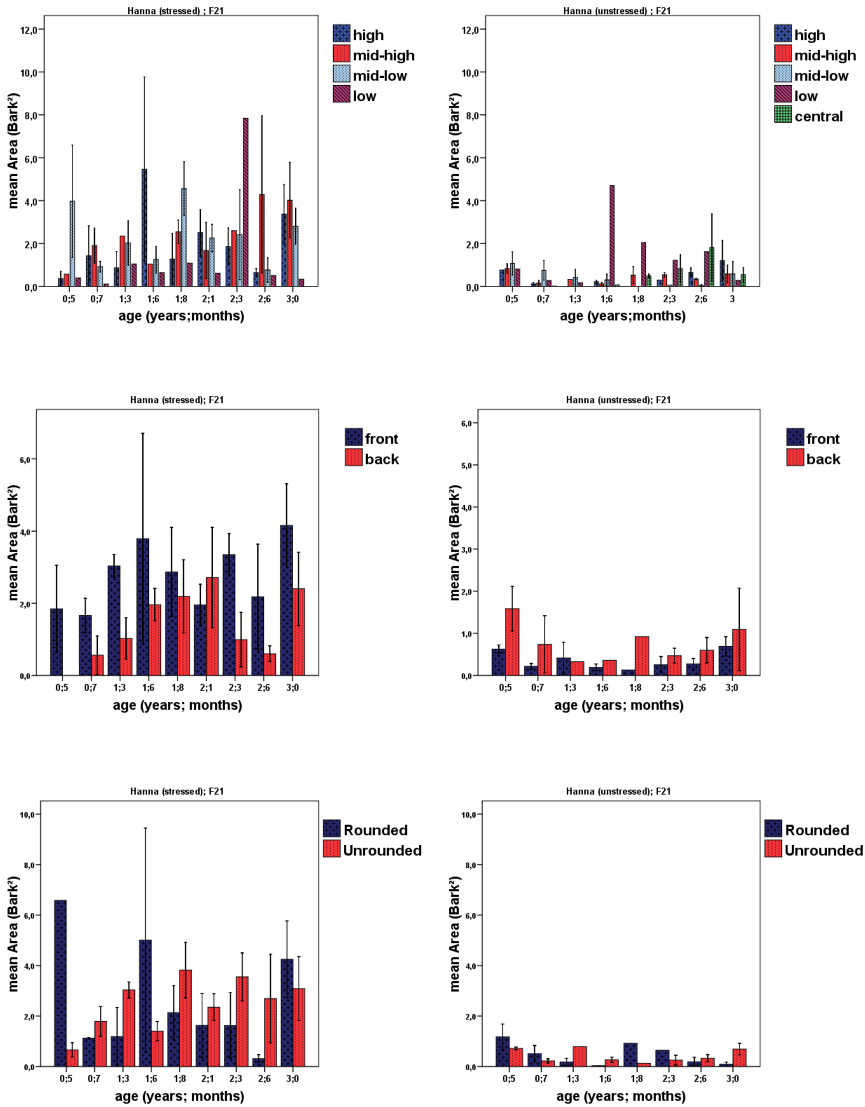


Figure D.2: Area of the dispersion ellipses related to the produced **stressed** (left) and **unstressed** (right) vowels, in the F1 vs. F2 dimension (in $Bark^2$), grouped according to their height, front-backness, and roundness depending on age for **Hanna**.

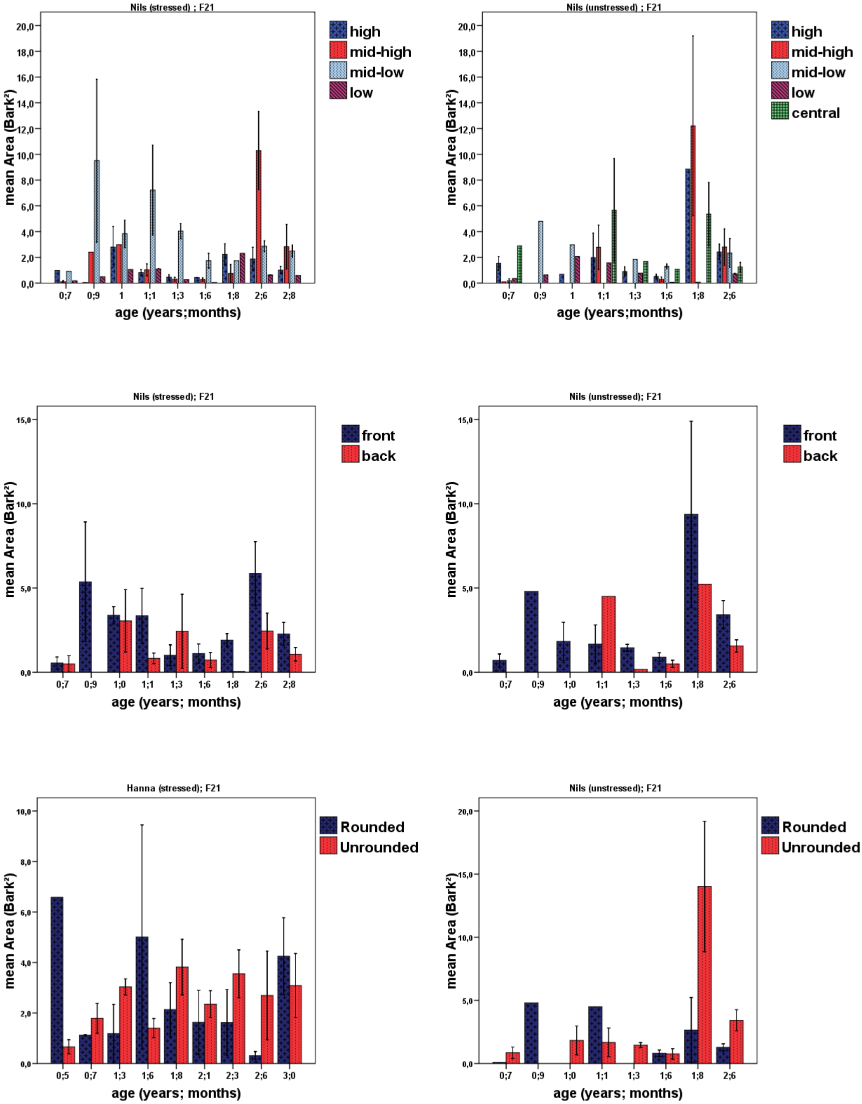


Figure D.3: Area of the dispersion ellipses related to the produced **stressed** (left) and **unstressed** (right) vowels, in the F1 vs. F2 dimension (in $Bark^2$), grouped according to their height, front-backness, and roundness depending on age for **Nils**.

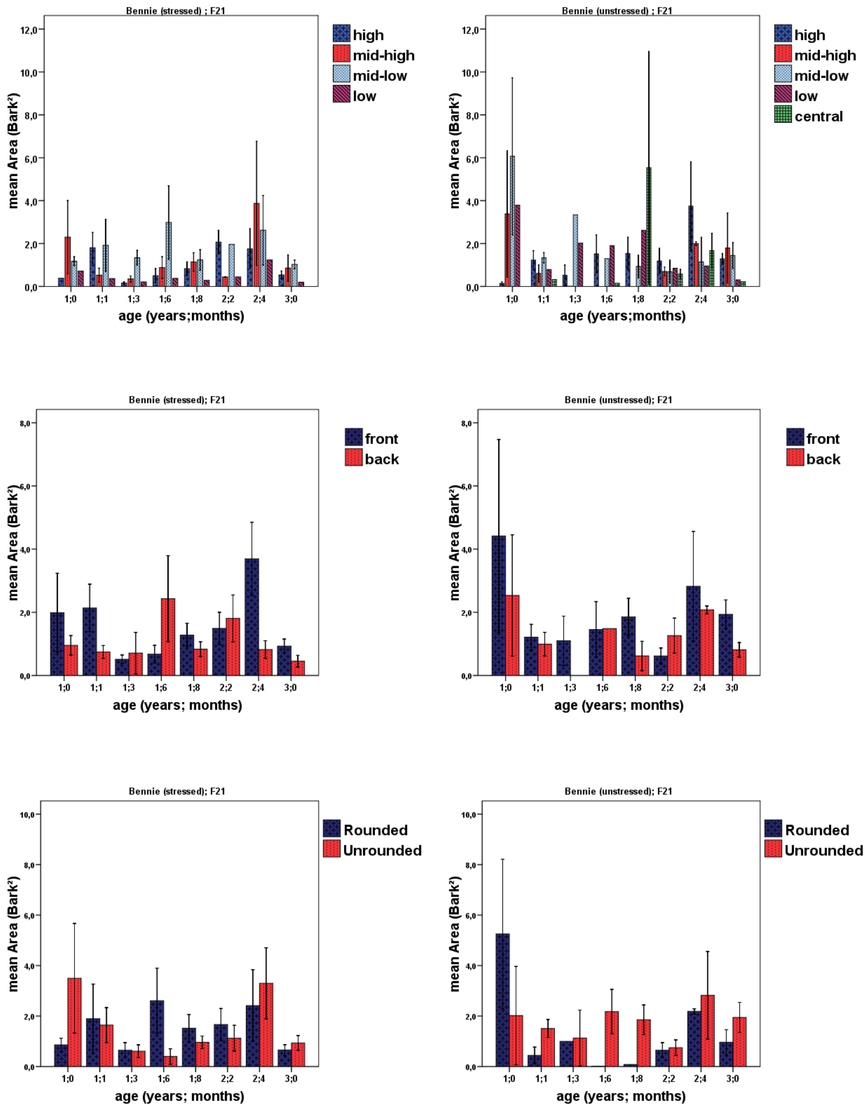


Figure D.4: Area of the dispersion ellipses related to the produced **stressed** (left) and **unstressed** (right) vowels, in the F1 vs. F2 dimension (in $Bark^2$), grouped according to their height, front-backness, and roundness depending on age for **Bennie**.

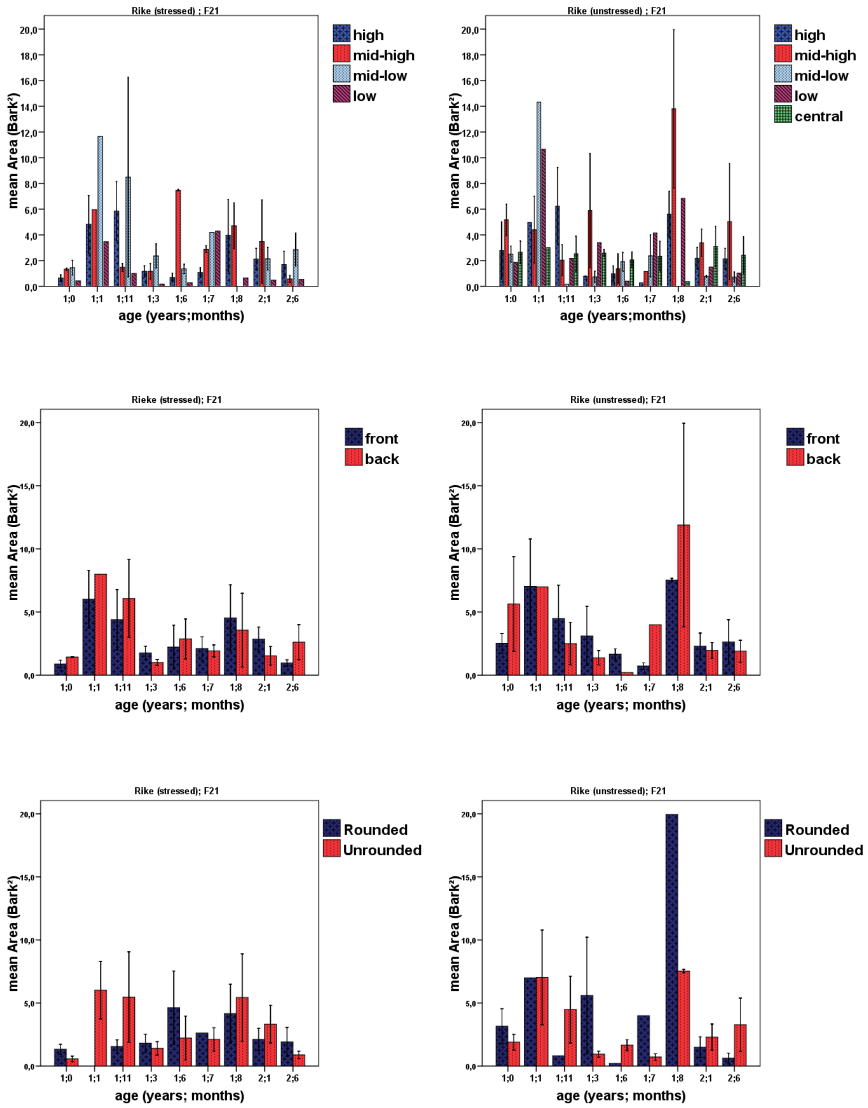


Figure D.5: Area of the dispersion ellipses related to the produced **stressed** (left) and **unstressed** (right) vowels, in the F1 vs. F2 dimension (in $Bark^2$), grouped according to their height, front-backness, and roundness depending on age for **Rike**.

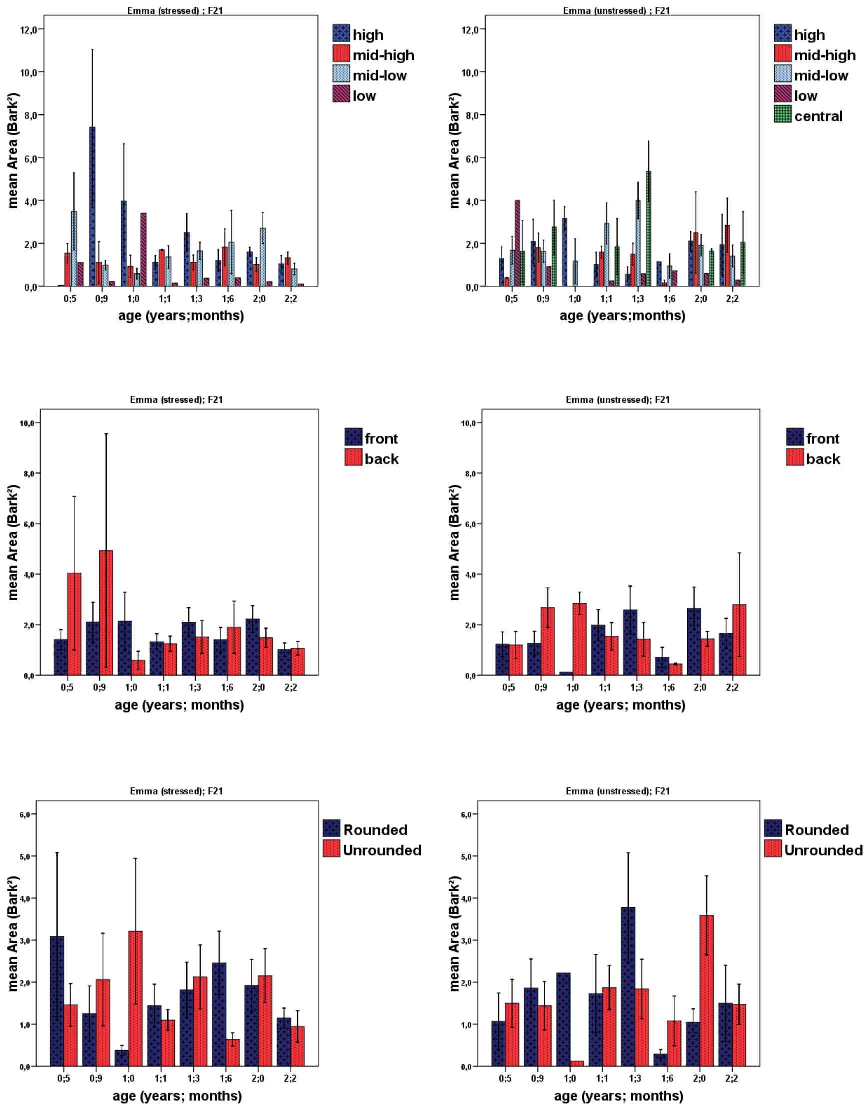


Figure D.6: Area of the dispersion ellipses related to the produced **stressed** (left) and **unstressed** (right) vowels, in the F1 vs. F2 dimension (in $Bark^2$), grouped according to their height, front-backness and roundness depending on age for **Emma**.

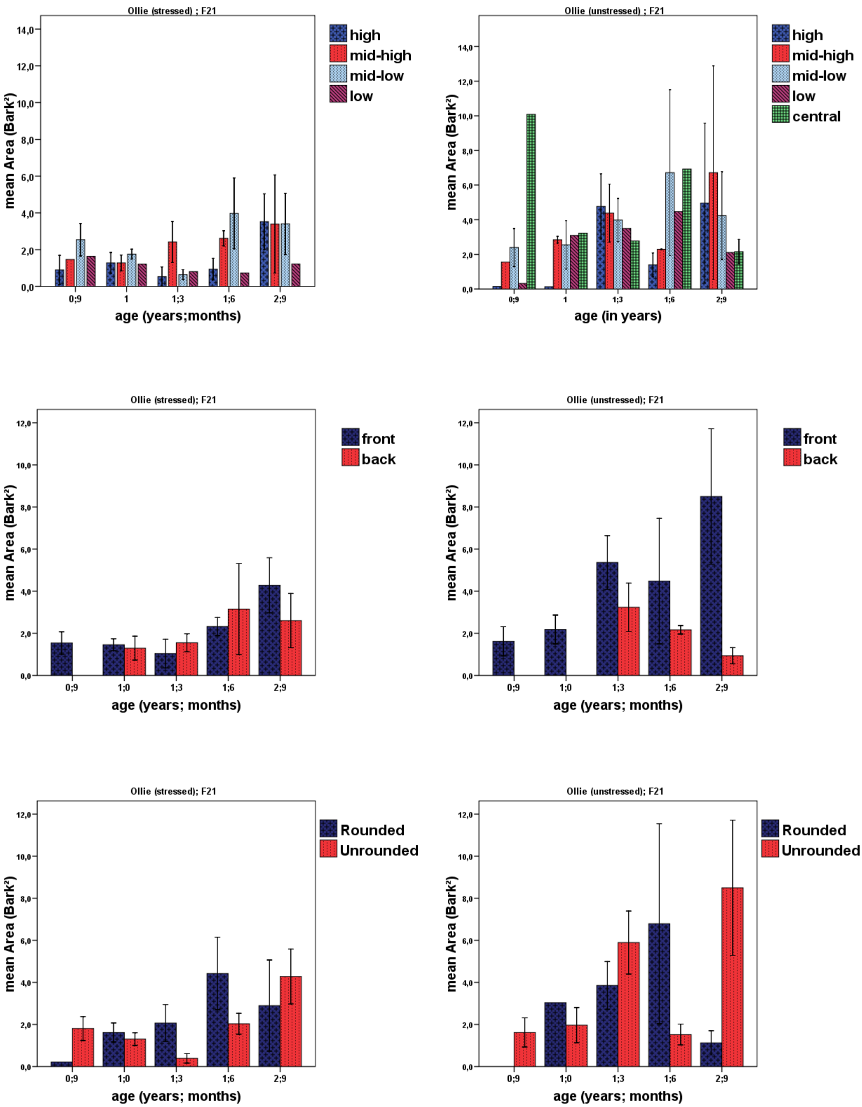


Figure D.7: Area of the dispersion ellipses related to the produced **stressed** (left) and **unstressed** (right) vowels, in the F1 vs. F2 dimension (in $Bark^2$), grouped according to their height, front-backness, and roundness depending on age for **Ollie**.

Bibliography

Adank, P., R. Smits, and R. van Hout (2004): A comparision of vowel normalization procedures for language variation research. Journal of the Acoustical Society of America 116(5), 3099–3107.

Allen, G. D. (1981): Development of Prosodic Phonology in Children's Speech: Further Evidence from the TAKI Task. In: Dressler, W. U., O. E. Pfeiffer, and J. R. Rennison (eds.), Phonologica 1980, *Innsbrucker Beiträge zur Sprachwissenschaft*, vol. 36, 9–14.

Allen, G. D. and S. Hawkins (1980): Phonological rhythm: Definition and development. In: Yeni-Komshian, G., J. Kavanagh, and C. Ferguson (eds.), Child Phonology: Volume 1. Production, New York: Academic Press, 227–256.

Altmann, H. and B. Kabak (2000): The use of prosodic information for disambiguation by German children: An experimental investigation. Master's thesis, University of Delaware.

Anderson, J. L., J. L. Morgan, and K. S. White (2003): A Statistical Basis for Speech Sound Discrimination. Language and Speech 46(2–3), 155–182.

Archibald, J. (1995): The acquisition of stress. In: Archibald, J. (ed.), Phonological acquisition and phonological theory, Hillsdale, NJ: Lawrence Erlbaum Associates, 81–109.

Atkinson-King, K. (1973): Children's acquisition of phonological stress contrasts. Ph.D. thesis, University of California.

Baayen, R. H., R. Piepenbrock, and L. Gulikers (1995): The CELEX lexical database. Philadelphia: Linguistic Data Consortium, University of Pennsylvania, release 2 [cd-rom] ed.

Baddeley, A. (1996): Working Memory. Oxford, University Press.

Baddeley, A. D., S. E. Gathercole, and C. Papagno (1998): The phonological loop as a language learning device. Psychological Review 105, 158–173.

Barton, D. (1976): The Role of Perception in the Acquisition of Phonology. Ph.D. thesis, University of London.

Beckman, M. E. (1986): Stress and Non-Stress Accent. Dordrecht: Foris.

Beckman, M. E. (2003): Input Representations (Inside the Mind and Out). In: Garding, G. and M. Tsujimura (eds.), WCCFL 22 Proceedings, Somerville, MA: Cascadilla Press, 70–94.

Beckman, M. E. and J. Edwards (2000): The Ontogeny of Phonological Categories and the Primacy of Lexical Learning in Linguistic Development. Child Development 71(1), 240–249.

Behrens, H. and U. Gut (2005): The relationship between prosodic and syntactic organization in early multiword speech. Journal of Child Language 32, 1–34.

Blevins, J. (1995): The Syllable in Phonological Theory. In: Goldsmith, J. A. (ed.), The Handbook of Phonological Theory, Oxford : Blackwell, 207–244.

Borden, G. J., K. S. Harris, and L. J. Raphael (1994): Speech science primer. Baltimore: Williams & Wilkins.

Bosch, L. and N. Sebastián-Gallés (2003): Simultaneous Bilingualism and the Perception of a Language–Specific Vowel Contrast in the First Year of Life. Language and Speech 46(2–3), 217–243.

Brunner, M., B. Pfeiffer, C. Heinrich, and U. Pröschel (2005): Entwicklung und Erprobung des Heidelberger Vorschulscreenings zur auditiven Wahrnehmung und Sprachverarbeitung HVS. Folia Phoniatrica et Logopaedica 57, 48–58.

Buhr, R. D. (1980): The emergence of vowels in an infant. Journal of Speech and Hearing Research 23, 73–94.

Burns, T. C., K. A. Yoshida, K. Hill, and J. F. Werker (2007): The development of phonetic representation in bilingual and monolingual infants. Applied Psycholinguistics 28, 455–474.

Busby, P. A. and G. L. Plant (1995): Formant frequency values of vowels produced by preadolescent boys and girls. Journal of the Acoustical Society of America 97, 2603–2606.

Carter, P. (2002): Structured Variation in British English Liquids: The Role of Resonance. Ph.D. thesis, University of York.

Cho, T. and P. Keating (2009): Effects of initial position versus prominence in English. Journal of Phonetics 37(4), 466–485.

Cholin, J., W. J. M. Levelt, and N. O. Schiller (2006): Effects of syllable frequency in speech production. Cognition 99(2), 205–235.

Chomsky, N. and M. Halle (1968): The Sound Pattern of English. New York: Harper & Row.

Christiansen, M. H., J. Allen, and M. S. Seidenberg (1998): Learning to segment speech using multiple cues: A connectionist model. Language and Cognitive Processess 13, 221–268.

Christophe, A., A. Gout, S. Peperkamp, and J. Morgan (2003a): Discovering words in the continuous speech stream: The role of prosody. Journal of Phonetics , 585–598.

Christophe, A., M. Nespor, M. T. Guasti, and B. van Ooyen (2003b): Prosodic structure and syntactic acquisition: The case of the head–direction parameter. Developmental Science 6(2), 211–220.

Claßen, K., G. Dogil, M. Jessen, K. Marasek, and W. Wokurek (1998): Stimmqualität und Wortbetonung im Deutschen. In: Linguistische Berichte, Westdeutscher Verlag, vol. 174, 202–245.

Curtin, S. L. (2002): Representational richness in phonological development. Ph.D. thesis, University of Southern California.

Curtin, S. L., T. H. Mintz, and M. H. Christiansen (2005): Stress changes the representational landscape: Evidence from word segmentation. Cognition 96(3), 233–262.

Daelemans, W., G. Durieux, and S. Gillis (1994): The acquisition of stress: a data–oriented approach. Computational Linguistics 20(3), 421–451.

Davis, B. L. and P. F. MacNeilage (1995): The Articulatory Basis of Babbling. Journal of Speech and Hearing Research 38, 1199–1211.

Davis, B. L., P. F. MacNeilage, and C. Matyear (2002): Aquisition of Serial Complexity in Speech Production: A Comparison of Phonetic and Phonological Approaches to First Word Production. Phonetica 59, 75–107.

Davis, B. L., P. F. MacNeilage, C. Matyear, and J. Powell (2000): Prosodic Correlates of Stress in Babbling: An Acoustical Study. Child Development 71(5), 1258–1270.

de Boysson-Bardies, B. (1999): How language comes to children: from birth to two years. Cambridge, MA: MIT Press.

de Boysson-Bardies, B., P. Halle, L. Sagart, and C. Durand (1989): A crosslingusitic investigation of vowel formants in babbling. Journal of Child Language 16, 1–17.

de Boysson-Bardies, B., L. Sagart, and C. Durand (1984): Discernible differences in the babbling of infants according to target language. Journal of Child Language 11(1), 1–15.

de Boysson-Bardies, B., L. Sagart, P. Halle, and C. Durand (1986): Acoustic investigation of crosslinguistic variability in babbling. In: Lindblom and Zetterström (eds.), Precursors of early speech, New York: Stockton Press., 113–126.

de Boysson-Bardies, B. and M. M. Vihman (1991): Adaptation to language: Evidence from babbling and first words in four languages. Language 67, 297–319.

de Bree, E., P. van Alphen, P. Fikkert, and F. Wijnen (2008): Metrical stress in comprehension and production of Dutch chidlren at-risk of dyslexia. In: Chan, H., H. Jacob, and E. Kapia (eds.), BUCLD 32: Proceedings of the 32nd annual conference on language development, 60–71.

Demuth, K. (1996a): Alignment, stress, and parsing in early phonological words. In: Bernhardt, B., J. Gilbert, and D. Ingram (eds.), Proceedings of the UBC International Conference on Phonological Acquisition, Cascadilla Press, 113–125.

Demuth, K. (1996b): The Prosodic Structure of Early Words. In: Morgan, J. and K. Demuth (eds.), Signal to Syntax: Bootstrapping from Speech to Grammer in Early Acquisition, Mahwah, N.J.: Lawrence Erlbaum Associates, 171–183.

Demuth, K. (2001): Prosodic Constraints on Morphological Development. In: Weissenborn, J. and B. Höhle (eds.), Approaches to bootstrapping: phonological, lexical, syntactic and neurophysiological aspects of early language acquisition, John Benjamins, Amsterdam/Philadelphia, vol. 2, 3–21.

Demuth, K. and J. E. Fee (1995): Minimal words in early phonological development. Ms. Brown University and Dalhousie University.

Demuth, K. and A. Tremblay (2008): Prosodically–conditioned variability in children's production of French determiners. Journal of Child Language 35(1), 99–127.

DePaolis, R. A., M. M. Vihman, and S. Kunnari (2008): Prosody in production at the onset of word use: A cross-linguistic study. Journal of Phonetics 36, 406–422.

Dogil, G. (1995): The Phonetic Manifestation of Stress. Arbeitspapiere des Instituts für Maschinelle Sprachverarbeitung (Univ. Stuttgart), AIMS 2 (2), University of Stuttgart.

Dogil, G. (2003): Understanding Prosody. In: Rickheit, G., T. Herman, and W. Deutsch (eds.), Psycholinguistik– Ein internationales Handbuch / Psycholinguistics– An International Handbook, Berlin: de Gruyter, 544–566.

Dogil, G. (2010): Hard–wired phonology: Limitations and latitude of phonological variation in pathological speech. In: Fougeron, C., B. Kühnert, M. D'Imperio, and N. Valleé (eds.), Laboratory Phonology 10, Berlin: de Gruyter.

Dogil, G. and B. Williams (1999): The phonetic manifestation of word stress. In: van der Hulst, H. (ed.), Word Prosodic Systems in the Languages of Europe, Berlin: De Gruyter, 273–334.

Domahs, U., R. Wiese, I. Bornkessel-Schlesewsky, and M. Schlesewsky (2008): The processing of German word stress: Evidence for the prosodic hierarchy. Phonology 25(1).

Dresher, B. E. and J. D. Kaye (1990): A computational learning model for metrical theory. Cognition 34, 137–195.

Dupoux, E. and S. Peperkamp (2001): A robust method to study stress "deafness". Journal of the Acoustical Society of America 110(3), 1606–1618.

Echols, C. H. (1993): A perceptually–based model of children's earliest productions. Cognition 46, 245–296.

Echols, C. H. (2001): Contributions of Prosody to Infants' Segmentation and Representation of Speech. In: Weissenborn, J. and B. Höhle (eds.), Approaches to bootstrapping: phonological, lexical, syntactic and neurophysiological aspects of early language acquisition, John Benjamins, Amsterdam/Philadelphia, vol. 1, 25–46.

Echols, C. H. and N. C. Marti (2004): The Identification of Words and Their Meanings: From Perceptual Biases to Language–Specific Cues. In: Hall, D. G. and S. R. Waxman (eds.), Weaving a lexicon, Cambridge, MA: MIT Press, 41–78.

Eisenberg, P. (1991): Syllabische Struktur und Wortakzent: Prinzipien der Prosodik deutscher Wörter. Zeitschrift für Sprachwissenschaft 10, 37–64.

Engstrand, O., K. Williams, and F. Lacerda (2003): Does Babbling Sound Native? Listener Responses to Vocalizations Produced by Swedish and American 12-and 18-Month-Olds. Phonetica 60, 17–44.

Fant, G. (1960): Acoustic theory of speech production. The Hague: Mouton Press.

Féry, C. (1998): German Word Stress in Optimality Theory. Journal of comparative Germanic Linguistics 2, 101–142.

Féry, C. (2008): Phonologie des Deutschen : eine optimalitätstheoretische Einführung, vol. 2. URL: http://opus.kobv.de/ubp/volltexte/2006/1091/: Potsdam: University of Potsdam.

Fikkert, P. (1994): On the Acquisition of Prosodic Structure. Ph.D. thesis, University of Leiden. Spracherwerb.

Fikkert, P. (2005): Getting sound structures in mind. In: Cutler, A. (ed.), Twenty-First Century Psycholinguistics: Four Cornerstones, Lawrence Erlbaum, Mahwah, New Jersey, 43–56.

Fikkert, P. and M. J. Freitas (1997): Acquisition of syllable structure constraints: Evidence from Dutch and Portuguese. In: Language acquisition: Knowledge representation and processing. Proceedings of GALA '97, Edinburgh: Edinburgh University Press, 217–222.

Fikkert, P. and C. C. Levelt (2008): How does place fall into place? The lexicon and emergent constraints in the developing phonological grammar. In: Avery, P., B. E. Dresher, and K. Rice (eds.), Contrast in phonology: Theory, Perception, Acquisition, Berlin, New York: Mouton de Gruyter, 231–268.

Fikkert, P., M. van Heugten, P. Offermans, and T. S. Zamuner (2005): Rhymes as a Window into Grammar. In: Brugos, A., M. R. Clark-Cotton, and S. Ha (eds.), Proceedings of the 29th annual BUCLD, Sommerville, MA, Cascadilla Press, vol. 1, 204–215.

Fisher, C. and B. A. Church (2001): Implicit Memory Support for Language Acquisition. In: Weissenborn, J. and B. Höhle (eds.), Approaches to bootstrapping: phonological, lexical, syntactic and neurophysiological aspects of early language acquisition, John Benjamins B.V., Amsterdam/Philadelphia, vol. 1, 47–71.

Fisher, C., B. A. Church, and K. E. Chambers (2004): Learning to identify spoken words. In: Hall, D. and S. Waxman (eds.), Weaving a lexicon, Cambridge, MA: MIT Press, 3–40.

Fougeron, C. and P. A. Keating (1997): Articulatory strengthening at edges of prosodic domains. Journal of the Acoustical Society of America 101, 3728–3740.

Friederici, A. D. and J. Wessels (1993): Phonotactic knowledge of word boundaries and its use in infant speech perception. Perception and Psychophysics 43, 287–295.

Friedrich, C. K., L. Aditi, and C. Eulitz (2008): Neurophysiological Evidence for Underspecified Lexical Representations: Asymmetries With Word Initial Variations. Journal of Experimental Psychology: Human Perception adn Performance 34(6), 1545–1559.

Gerken, L. A. (1996): Prosodic structure in young children's language production. Language 72, 683–712.

Giegerich, H. J. (1985): Metrical Phonology and Phonological Structure: German and English. Cambridge: Cambridge University Press.

Gilbert, H. R., M. P. Robb, and Y. Chen (1997): Formant Frequency Development: 15 to 36 Months. Journal of Voice 11(3), 260–266.

Gillis, S., W. Daelemans, and G. Durieux (2000): A comparision of natural and machine learning of word stress. In: Broeder, P. and J. Murre (eds.), Models in Language Acquisition, Oxford University Press, 76–99.

Goad, H. and Y. Rose (2004): Input Elaboration, Head Faithfulness and Evidence for Representation in the Acquisition of Left-edge Clusters in West Germanic. In: Kager, R., J. Pater, and W. Zonneveld (eds.), Constraints in Phonological Acquisition, Cambridge: Cambridge University Press, 109–157.

Goedemans, R. (1998): Weightless Segments. A Phonetic and Phonological Study Concerning the Metrical Irrelevance of Syllable Onsets. Ph.D. thesis, University of Leiden.

Goldinger, S. D. (1997): Words and Voices – Perception and Production in an Episodic Lexicon. In: Johnson, K. and J. W. Mullennix (eds.), Talker Variability in Speech Processing, San Diego: Academic Press, 33–66.

Goldinger, S. D. (1998): Echoes of Echoes? An Episodic Theory of Lexical Access. Psychological Review 105(2), 251–279.

Goldstein, U. G. (1980): An articulatory model for the vocal tracts of growing children. Ph.D. thesis, MIT.

Grimm, A. (2008): The development of early prosodic word structure in child German: Simplex words and compounds. Ph.D. thesis, University of Potsdam.

Grimm, H. and H. Doil (2004): ELFRA, Elternfragebögen für die Früherke nnung von Risikokindern. Hogrefe, Verlag für Psychologie.

Guenther, F. H. (1994): A Neural Network Model of Speech Acquisition and Motor Equivalent Speech Production. Biological Cybernetics 72, 43–53.

Guenther, F. H. (1995): A modeling framework for speech motor development and kinematic articulator control. In: Proceedings of the International Congress of Phonetic Science 1995, Stockholm, 92–99.

Guenther, F. H. (2001): Neural Modeling of Speech Production. In: Proceedings of the 4th International Nijmegen Speech Motor Conference, 12–15.

Guenther, F. H. (2003): Neural control of speech movements. In: Meyer, A. and N. Schiller (eds.), Phonetics and Phonology in Language Comprehension and Production: Differences and Similarities, Berlin: Mouton de Gruyter, 209–239.

Guenther, F. H., M. Hampson, and D. Johnson (1998): A Theoretical Investigation of Reference Frames for the Planning of Speech Movements. Psychological Review 105(4), 611–633.

Guenther, F. H. and J. S. Perkell (2004): A neural model of speech production and its application to studies of the role of auditory feedback in speech. In: Maassen, B., R. Kent, H. Peters, P. Van Lieshout, and W. Hulstijn (eds.), Speech Motor Control in Normal and Disordered Speech, Oxford: Oxford University Press, 29–49.

Halle, P. A., B. de Boysson-Bardies, and M. M. Vihman (1991): Beginnings of Prosodic Organization: Intonation and Duration patternd of Disyllables produced by Japanese andFrench Infants. Language and Speech 34(4), 299–318.

Hanson, H. M. (1995): Glottal Characteristics of Female Speakers. Ph.D. thesis, Harvard University.

Hanson, H. M. (1997): Glottal characeristics of female speakers: Acoustic correlates. Journal of the Acoustical Society of America 101, 466–481.

Hayes, B. (1995): Metrical stress theory: principles and case studies. Chicago: University of Chicago Press.

Heuft, B. and T. Portele (1994): Zur akustischen Realisierung des Wortakzents. In: 5. Gemeinsame Konferenz Elektronische Sprachsignalverarbeitung, 197–204.

Heuft, B., T. Portele, F. Höfer, H. Meyer, and M. Rauth (1995): Betonungsstufen von Silben und ihre Beziehung zum Sprachsignal. In: Fortschritte der Akustik – DAGA ’95, 999–1002.

Hillenbrand, J., L. A. Getty, M. J. Clark, and K. Wheeler (1995): Acoustic characteristics of American English vowels. Journal of the Acoustical Society of America 97(5), 3099–3111.

Hochberg, J. G. (1988): Learning Spanish stress: Developmental and theoretical perspectives. Language 64(4), 683–706.

Höhle, B., J. Weissenborn, M. Schmitz, and A. Ischebeck (2001): Discovering Word Order Regularities: The role of prosodic information for early parameter setting. In: Weissenborn, J. and B. Höhle (eds.), Approaches to bootstrapping: phonological, lexical, syntactic and neurophysiological aspects of early language acquisition, John Benjamins, Amsterdam/Philadelphia, vol. 1, 251–267.

House, D. (1996): Differential perception of tonal contours through the syllable. In: Proceedings of the International Conference on Spoken Language Processing (Philadelphia, PA), vol. 1, 2048–2051.

Howard, I. S. and P. Messum (2007): A Computational Model of Infant Speech Development. In: Proceedings of Specom, Moscow, Russia, 756–764.

Huber, J. E., E. T. Stathopoulos, G. M. Curione, T. A. Ash, and K. Johnson (1999): Formant of children, women, and men: The effects of vocal intensity variation. Journal of the Acoustical Society of America 106(3), 1532–1542.

Isačenko, A. V. and H. J. Schädlich (1966): Untersuchungen über die deutsche Satzintonation. Studia Grammatica 7, 7–64.

Ishizuka, K., R. Mugitani, H. Kato, and S. Amano (2007): Longitudinal developmental changes in spectral peaks of vowels produced by Japanese infants. Journal of the Acoustical Society of America 121(4), 2272–2282.

Janßen, U. (2003): Untersuchungen zum Wortakzent im Deutschen und Niederländischen. Ph.D. thesis, University of Düsseldorf.

Jessen, M. (1997): Phonetic and phonology of the tens and lax obstruents in German. Ph.D. thesis, University of Stuttgart.

Jessen, M. (1999): German. In: van der Hulst, H. (ed.), Word Prosodic Systems in the Languages of Europe, Berlin, New York: Mouton de Gruyter, 515–545.

Jessen, M., K. Marasek, K. Schneider, and K. Claßen (1995): Acoustic correlates of word stress and the tense/lax opposition in the vowel system of German. In: Proceedings of the International Congress of Phonetic Sciences 13 (Stockholm), vol. 4, 428–431.

Johnson, K. (1997): Speech perception without speaker normalization. In: Johnson, K. and J. W. Mullennix (eds.), Talker variability in speech processing, San Diego: Academic Press, 145–165.

Jusczyk, P. W. (1996): Developmental Speech Perception. In: Lass, N. J. (ed.), Principles of Experimental Phonetics, Mosby–Year Book, chap. 9, 328–361. Prosodic bootstrapping, mental lexicon, perception experiments, talker variability.

Jusczyk, P. W. (1997): The discovery of spoken language. The MIT Press, Cambridge, MA, London, England.

Jusczyk, P. W., A. Cutler, and N. Redanz (1993a): Infant's preference for the predominant stress patterns of English words. Child Development 64, 675–687.

Jusczyk, P. W., A. D. Friederici, J. Wessels, V. Y. Svenkerud, and A. M. Jusczyk (1993b): Infants' sensitivity to the sound patterns of native language words. Journal of Memory and Language 32, 402–420.

Jusczyk, P. W., K. Hirsh-Pasek, D. G. Kemler Nelson, L. J. Kennedy, A. Woodward, and J. Piwoz (1992): Perception of acoustic correlates of major phrasal units by young infants. Cognitive Psychology 24, 252–293.

Jusczyk, P. W., P. A. Luce, and J. Charles-Luce (1994): Infants Sensitivity to Phonotactic Patterns in the Native Language. Journal of Memory and Language 33, 630–645.

Kehoe, M. M. (2000): Truncation Without Shape Constraints: The Latter Stages of Prosodic Acquisition. Language Acquisition 8(1), 23–67.

Kehoe, M. M. and C. Lleó (2003): The acquisition of nuclei: a longitudinal analysis pf phonological vowel length in three German-speaking children. Journal of Child Language 30, 527–556.

Kehoe, M. M. and C. Stoel-Gammon (1997): The acquisition of prosodic structure: An investigation of currents accounts of children's prosodic development. Language: Journal of the Linguistic Society of America 73(1), 113–145.

Kehoe, M. M., C. Stoel-Gammon, and E. H. Buder (1995): Acoustic Correlates of Stress in Young Children's Speech. Journal of Speech and Hearing Research 38, 338–350.

Kemler Nelson, D. G., K. Hirsh-Pasek, P. W. Jusczyk, and C. K. Wright (1989): How the prosodiec cues in motherese might assist language learning. Journal of Child Language 16, 55–68.

Kent, R. D. and G. Miolo (1995): Phonetik Abilities in the First Year of Life. In: Fletcher, P. and B. MacWhinney (eds.), The Handbook of Child Language, Oxford & Cambridge, MA: Blackwell, 303–334.

Kent, R. D. and A. D. Murray (1982): Acoustic features of infant vocalic utterances at 3,6, and 9 months. Journal of the Acoustical Society of America 72(2), 353–365.

Kirchner, R. (1999): Preliminary thoughts on phonologization within an exemplar–based speech processing system. In: Gordon, M. K. (ed.), Papers in Phonology 2, UCLA, *UCLA Working Papers in Linguistics*, vol. 1, 207–231.

Kröger, B. J., J. Kannampuzha, and C. Neuschaefer-Rube (2009): Towards a neurocomputational model of speech production and perception. Speech Communication 51(9), 793–809.

Kröger, B. J., A. Lowit, and R. Schnitker (2008): The Organization of a Neurocomputational Control Model for Articulatory Speech Synthesis. In: Esposito, A., N. Bourbakis, N. Avouris, and I. Hatzilygeroudis (eds.), Verbal and Nonverbal Features of Human-Human and Human-Machine Interaction. Selected papers from COST Action 2102 International Conference, Patras, Greece, Springer Verlag, New York, 121–135.

Kuhl, P. K. and A. N. Melztzoff (1995): Vocal learning in infants: Development of perceptual–motor links for speech. In: Proceedings from the 13th International Congress of Phonetic Science, Stockholm, vol. 1, 146–149.

Kuhl, P. K. and A. N. Melztzoff (1996): Infant vocalizations in response to speech: Vocal imitation. Journal of the Acoustical Society of America 100, 2425–2438.

Kuhl, P. K., E. Stevens, A. Hayashi, T. Deguchi, S. Kiritani, and P. Iverson (2006): Infanst show a faciliation effect for native lagnuage phonetic perception between 6 and 12 months. Developmental Science 9(2), F13–F21.

Kuhl, P. K., K. A. Williams, F. Lacerda, K. N. Stevens, and B. Lindblom (1992): Lingusitic experience alters phonetic perception in infants by 6 months of age. Science 255, 606–608.

Lacerda, F. (1995): The perceptual–magnet effect: An emergent consequence of exemplar–based phonetic memory. In: Elenius, K. and P. Branderud (eds.), Proceedings ICPhS 1995, Stockholm, vol. 2, 140–147.

Lacerda, F. (2003): Phonology: An emergent consequence of memory constraints and sensory input. Reading and Writing: An Interdisciplinary Journal 16, 41–59.

Ladefoged, P. and D. E. Broadbent (1957): Information conveyed by vowels. Journal of the Acoustical Society of America 29, 88–104.

Lane, H., M. Denny, F. H. Guenther, M. L. Matthies, L. Ménard, J. S. Perkell, E. Stockmann, M. Tiede, J. Vick, and M. Zandipour (2005): Effects of bite blocks and hearing status on vowel production. Journal of the Acoustical Society of America 118(3), 1636–1646.

Lecanuet, J.-P. and B. Schaal (2002): Sensory performance in the human foetus: a brief summary of research. Intellectica 34(1), 29–56.

Lee, S., A. Potamianos, and N. Shrikanth (1999): Acoustics of children's speech: Developmental changes of temporal and spectral parameters. Journal of the Acoustical Society of America 105(3), 1455–1468.

Levelt, C. C., N. O. Schiller, and P. Fikkert (2003): Metrical Priming in Speech Production. In: M., S., D. Recasens, and J. Romero (eds.), Proceedings of the 15th International Conference of Phonetic Sciences, 2481–2484.

Levelt, C. C., N. O. Schiller, and W. J. M. Levelt (1999a): A Developmental Grammar for Syllable Structure in the Production of Child Language. Brain and Language 68, 291–299.

Levelt, C. C., N. O. Schiller, and W. J. M. Levelt (2000): The Acquisition of Syllable Types. Language Acquisition 8(3), 237–264.

Levelt, W. J. M. (1989): Speaking: From Intonation to Articulation. MIT Press, Cambridge, MA.

Levelt, W. J. M., A. Roelofs, and A. S. Meyer (1999b): A theory of lexical access in speech production. Behavioral and Brain Sciences 22, 1–75.

Levelt, W. J. M. and L. Wheeldon (1994): Do speakers have access to a mental syllabary? Cognition 50, 239–269.

Liberman, M. Y. (1975): The Intonational System of English. Ph.D. thesis, Cambridge, MA: MIT.

Lieberman, P. (1980): On the development of vowel production in young children. In: Yeni-Komshian, G., J. Kavanagh, and C. Ferguson (eds.), Child Phonology, Volume 1, Production, London: Academic Press, 113–142.

Lindfield, K. C., A. Wingfield, and H. Goodglass (1999): The Role of Prosody in the Mental Lexicon. Brain and Language 68, 312–317.

Lintfert, B. and K. Schneider (2006): Prosodic correlates of stress in babbling and first words: Acoustic analysis of German infants. LabPhon10, Paris.

Lintfert, B. and U. Vollmer (2007): Acoustic analysis of stress in babbling and first words in German. Child Language Seminar, Reading.

Lintfert, B. and W. Wokurek (2005): Voice quality dimensions of pitch accents. In: Proceedings of Interspeech/Eurospeech 2005 (Lisbon), 2409–2412.

Liu, H. M., P. K. Kuhl, and F. M. Tsao (2003): An association between mothers'speech clarity and infants' speech discrimination skills. Developmental Science 6(3), F1–F10.

Lobanov, B. M. (1971): Classification of Russioan Vowels Spoken by Different Speakers. Journal of the Acoustical Society of America , 606–607.

Locke, J. L. (1993): The Child's Path to Spoken Language. Cambridge, MA & London: Harvard University Press.

MacNeilage, P. F. (1997): Acquisition of Speech. In: Hardcastle, W. and J. Laver (eds.), The Handbook of Phonetic Sciences, Oxford & Malden: Blackwell, 301–332.

MacNeilage, P. F. (1998): The frame/content theory of evolution of speech production. Behavioral and Brain Sciences 21, 499–546.

MacNeilage, P. F. and B. L. Davis (1990): Vowel–Consonant Relations in Babbling. In: Proceedings of the XIIth ICPhS, Aix–en–Provence, France, vol. 1, 338–343. Babbling.

MacNeilage, P. F. and B. L. Davis (2000): On the origin of internal structure of word forms. Science 288, 527–531.

MacNeilage, P. F., B. L. Davis, A. Kinney, and C. L. Matyear (2000): The Motor Core of Speech: A Comparision of Serial Organization Patterns in Infants and Languages. Child Development 71(1), 153–163.

Mattys, S. L., L. White, and J. F. Melhorn (2005): Integration of Multiple Speech SegmentationCues: A Hierarchical Framework. Journal of Experimental Psychology: General 134(4), 477–500.

Maye, J., D. J. Weiss, and R. N. Aslin (2008): Statistical phonetic learning in infants: Facilitation and feature generalization. Developmental Science 11(1), 122–134.

Maye, J., J. F. Werker, and L. A. Gerken (2002): Infant sensitivity to distributional information can affect phonetic discrimination. Cognition 82, B101–B111.

McCarthy, J. J. and A. S. Prince (1996): Prosodic Morphology 1986. http://ruccs.rutgers.edu/pub/papers/pm86all.pdf.

Mehler, J., A. Christophe, and F. Ramus (2000): How infants acquire language: Some preliminary observations. In: Marantz, A., Y. Miyashita, and W. O'Neil (eds.), Image, Language, Brain: Papers from the first Mind-Brain Articulation Project symposium, Cambridge, MA: MIT Press, 51–75.

Ménard, L., B. Davis, and L.-J. Boë (2007a): Perceptual categorization of synthesized English vowels from birth to adulthood. In: Trouvain, J. and W. J. Barry (eds.), Proceedings of the 16th International Congress of Phonetic Sciences, 315–320.

Ménard, L., J.-L. Schwartz, and L.-J. Boë (2004): Role of Vocal Tract Morphology in Speech Development: Perceptual Targets and Sensorimotor Maps for Synthesized French Vowels From Birth to Adulthood. Journal of Speech, Language, and Hearing Research 47(5), 1059–1080.

Ménard, L., J.-L. Schwartz, L.-J. Boë, and J. Aubin (2007b): Articulatory–acoustic relationships during vocal tract growth for French vowels: Analysis of real data and simulations with an articulatory model. Journal of Phonetics 35, 1–19.

Menn, L. and C. Stoel-Gammon (1995): Phonological Development. In: Fletcher, P. and B. MacWhinney (eds.), The Handbook of Child Language, Oxford & Cambridge, MA: Blackwell, 335–359.

Messum, P. R. (2007): The Role of Imitation in Learning to Pronounce. Ph.D. thesis, London University.

Miller, J. D. (1989): Auditory–perceptual interpretation of the vowel. Journal of the Acoustical Society of America 85, 2114–2134.

Möbius, B. (2001): German and Multilingual Speech Synthesis. Arbeitspapiere des Instituts für Maschinelle Sprachverarbeitung (Univ. Stuttgart), AIMS 7 (4), University of Stuttgart.

Möbius, B. (2004): Corpus-based investigations on the phonetics of consonant voicing. Folia Linguistica 38(1-2), 5–26.

Nazzi, T., G. Iakimova, J. Bertoncini, S. Frédonie, and C. Alcantara (2006): Early segmentation of fluent speech by infants acquiring French: Emerging evidence for crosslinguistic differences. Journal of Memory and Language 54, 283–299.

Nazzi, T., G. Iakimova, J. Bertoncini, S. Mottet, J. Serres, and S. de Schonen (2008): Behavioral and electrophysiological exploration of early word segmentation in French: Distinguishing the syllabic and lexical levels. In: Friederici, A. D. and G. Thierry (eds.), Early language development: bridging brain and behaviour, John Benjamins B.V., *Trends in Language Acquisition Reearch*, vol. 5, 65–90.

Nazzi, T., D. G. Kemler Nelson, P. W. Jusczyk, and A. M. Jusczyk (2000): Six-month-olds' Detection of Clauses Embedded in Continouous Speech: Effects of Prosodic Well–Formedness. Infancy 1, 123–147. Antrag.

Nespor, M. and I. Vogel (1986): Prosodic phonology. Dordrecht: Foris.

Nosofsky, R. M. (1986): Attention, similarity, and the identification–categorization relationship. Journal of Experimental Psychology: General 115(1), 39–57.

Packman, A., C. Codde, and M. Onslow (2007): On the cause of stuttering: Integrating theory with brain and behavioral research. Journal of Neurolinguistics 20, 353–362.

Palethorpe, S., R. Wales, J. E. Clark, and T. Senserrick (1996): Vowel classification in children. Journal of the acoustical Society of America 100, 3843–3851.

Pater, J. (2004): Bridging the gap between receptive and productive development with minimally violable constraints. In: Kager, R., J. Pater, and W. Zonneveld (eds.), Contraints in Phonological Acquisition, Cambridge University Press, 219–244.

Payne, E. (2006): Phonetic motifs and the formation of sound structure. In: UCL Working Paper in Linguistics, vol. 18, 321–343.

Peperkamp, S. (2003): Phonological Acquisition: Recent Attainments and New Challenges. Language and Speech 46(2–3), 87–113.

Peperkamp, S. (2004): Lexical exceptions in stress systems: Arguments from early language acquisition and adult speech perception. Language 80, 98–126.

Peperkamp, S. and E. Dupoux (2002): A typological study of stress 'deafness'. In: Gussenhoven, C. and N. Warner (eds.), Laboratory Phonology 7, Mouton de Gruyter: Berlin, 203–240.

Perkell, J. S., F. H. Guenther, H. Lane, M. L. Matthies, P. Perrier, J. Vick, R. Wilhelms-Tricarico, and M. Zandipour (2000): A theory of speech motor control and supporting data from speakers with normal hearing and with profound hearing loss. Journal of Phonetics 28(3), 233–272.

Perrachione, T. K., J. B. Pierrehumbert, and P. C. M. Wong (2010): Differential neural contributions to native– and foreign–language talker identification. Journal of Experimental Psychology – Human Perception and Performance .

Peterson, G. E. and H. L. Barney (1952): Control Methods Used in a Study of the Vowels. Journal of the Acoustical Society of America 24(2), 175–184.

Pierrehumbert, J. B. (2000): The phonetic grounding of phonology. Bulletin de la Communication Parlee 5 , 7–23.

Pierrehumbert, J. B. (2001): Exemplar dynamics: Word frequency, lenition, and contrast. In: Bybee, J. and P. Hopper (eds.), Frequency effects and the emergence of linguistic structure, John Benjamins, Amsterdam, 137–157.

Pierrehumbert, J. B. (2002): Word-specific phonetics. In: Gussenhoven, C. and N. Warner (eds.), Laboratory Phonology VII, Mouton de Gruyter, Berlin, vol. 7, 101–140.

Pierrehumbert, J. B. (2003a): Phonetic Diversity, Statistical Learning and Acquisition of Phonology. Language and Speech 46(2–3), 115–154.

Pierrehumbert, J. B. (2003b): Probabilistic Phonology: Discrimation and Robustness. In: Bod, R., J. Hay, and S. Jannedy (eds.), Probabilistic linguistics, Cambridge, MA: MIT Press, 175–228.

Pierrehumbert, J. B., M. E. Beckman, and D. R. Ladd (2000): Conceptual Foundations of Phonology as a Laboratory Science. In: Burton-Roberts, N., P. Carr, and G. Docherty (eds.), Phonological Knowledge, Oxford University Press, 273–303.

Pollock, K. E., D. M. Brammer, and C. F. Hageman (1993): An acoustic analysis of young children's productions of word stress. Journal of Phonetics 21, 183–203.

Pols, L. C. W., H. R. C. Tromp, and R. Plomp (1973): Frequency analysis of Dutch vowels from 50 male speakers. Journal of the Acoustical Society of America 53, 1093–1101.

Pompino-Marschall, B. (1990): Die Silbenprosodie. Ein elementarer Aspekt der Wahrnehmung von Sprechrhythmus und Sprechtempo. Niemeyer, Tübingen.

Prince, A. and P. Smolensky (2004): Optimality Theory: Constraint Interaction in Generative Grammar. Blachwell Publishing.

Pützer, M. and W. Wokurek (2006): Multiparametrische Stimmprofildifferenzierung zu männlichen und weiblichen Normalstimmen auf der Grundlage akustischer Analysen. Laryngo–Rhino–Otologie 85, 1–8.

Ramus, F. (2002): Language discirmination by newborns: Teasing apart phonotactic, rhythmic, and intonational cues. Annual Review of Language Acquisition 2, 85–115.

Ramus, F., M. Nespor, and J. Mehler (1999): Correlates of linguistic rhythm in the speech signal. Cognition 73, 265–292. Spracherwerb, Prosodie, phonological bootstrapping.

Robb, M. P., Y. Chen, and G. H. R. (1997): Developmental aspects of formant frequency and bandwiths in infants and toddlers. Folia Phoniatrica et Logopaedica 49, 88–95.

Robb, M. P. and A. A. Tyler (1995): Durations of young children's word and nonword vocalizations. Journal of the Acoustical Society of America 98, 1348–1354.

Rose, Y. (2000): Headedness and Prosodic Licensing in the L1 Acquisition of Phonology. Ph.D. thesis, McGill University.

Rose, Y. (in press): Internal and External Influences on Child Language Production. In: Chitoran, I., F. Pellegrino, and E. Marsico (eds.), Approaches to Phonological Complexity, Berlin: Mouton de Gruyter.

Rose, Y. and C. Champdoizeau (in press): There is no Innate Trochaic Bias: Acoustic Evidence in Favour of the Neutral Start Hypothesis. In: Gavarró, A. and M. J. Freitas (eds.), Proceedings of the Generative Approaches to Language Acquisition Conference 2007, Cambridge Scholars Publishing.

Rvachew, S., K. Mattock, and L. Polka (2006): Developmental and cross–linguistic variation in the infant vowel space: The case of Canadian English and Canadian French. Journal of the Acoustical Society of America 120(4), 2250–2259.

Rvachew, S., E. B. Slawinski, M. Williams, and C. L. Green (1996): Formant frequencies of vowels produced by infants with and without early onset otitis media. Canadian Acoustics 24, 19–28.

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

Saffran, J. R., E. K. Johnson, R. N. Aslin, and E. L. Newport (1999): Statistical learning of tone sequences by human infants and adults. Cognition 70, 27–52.

Sawashima, M. and H. Hirose (1983): Laryngeal Gestures in Speech Production. In: MacNeilage, P. (ed.), The Production of Speech, New York, Berlin: Springer Verlag, 11–38.

Schiller, N. O. (2006): Lexical stress encoding in single word production estimated by event–related brain potentials. Brain Research 1112, 201–212.

Schiller, N. O., B. M. Jansma, J. Peter, and W. J. M. Levelt (2006): Monitoring metrical stress in polysyllabic words. Language and Congitive Processes 21(1-3), 112–140.

Schneider, K. (2007): Acquisition of word stress in German: Vowel duration and Incompleteness of Closure. In: Proceedings of the 16th ICPhS (Saarbrücken), 1565–1568.

Schneider, K. and B. Lintfert (2005): Acquisition of contrastive stress in German children. Posterpresentation at the 10th International Congress for the Study of Child Language (IASCL), Berlin.

Schneider, K., B. Lintfert, G. Dogil, and B. Möbius (2006): Phonetic grounding of prosodic categories. In: Sudhoff, S., D. Lenertov, R. Meyer, S. Pappert, P. Augurzky, I. Mleinek, N. Richter, and J. Schlieer (eds.), Methods in Empirical Prosody Research, De Gruyter, Berlin, 335–361.

Schneider, K. and B. Möbius (2006): Production of word stress in German: Children and adults. In: Proceedings of Speech Prosody 2006 (Dresden), 333–336.

Schneider, K. and B. Möbius (2007): Word stress correlates in spontaneous child-directed speech in German. In: Proceedings of Interspeech 2007 (Antwerpen), 1394–1397.

Schwartz, R. G., K. Petinou, L. Goffman, G. Lazowski, and C. Cartusciello (1996): Young children's production of syllable stress: An acoustic analysis. Journal of the Acoustical Society of America 99(5), 3192–3200.

Schweitzer, A. and B. Möbius (2004): Exemplar-based production of prosody: Evidence from segment and syllable duration. In: Proceedings of the Speech Prosodz 2004 Conference (Nara, Japan), 459–462.

Selby, J. C., M. P. Robb, and H. R. Gilbert (2000): Normal vowel articulations between 15 and 36 months of age. Clinical Linguistics & Phonetics 14(4), 255–265.

Selkirk, E. O. (1984): Phonology and syntax: The relation between sound and structure. Cambridge, MA: MIT Press.

Serkhane, J. E., J.-L. Schwartz, L.-J. Boë, B. L. Davis, and C. L. Matyear (2007): Infants' vocalizations analyzed with an articulatory model: A preliminary report. Journal of Phonetics 35(3), 321–340.

Shih, C., B. Möbius, and B. Narasimhan (1999): Contextual effects on consonant voicing profiles: A cross–linguistic study. In: Proceedings of the 14th International Congress of Phonetic Science (San Francisco, CA), vol. 2, 989–992.

Shukla, M., M. Nespor, and J. Mehler (2007): An interaction between prosody and statistics in the segmentation of fluent speech. Cogitiv Psychology 54(1), 1–32.

Sluijter, A. M. C. (1995): Phonetic Correlates of Stress and Accent. Ph.D. thesis, University of Leiden.

Smith, B. L. and M. K. Kenney (1998): An assessment of several acoustic parameters in children's speech production development: longitudinal data. Journal of Phonetics 26, 95–108.

Smith, B. L. and M. K. Kenney (1999): A Longitudinal Study of the Development of Temporal Properties of Speech Production: Data from 4 Children. Phonetica 56, 73–102.

Smith, N. V. (1973): The acquisition of phonology. Cambridge, University Press.

Snow, D. (2006): Regression and Reorganization of Intonation Between 6 and 23 Months. Child Development 77(2), 281–296.

Stevens, K. N. and H. M. Hanson (1994): Classification of glottal vibration from acoustic measurements. In: Fujimura, O. and M. Hirano (eds.), Vocal Fold Physiology: Voice Quality Control, Singular, San Diego, 147–170.

Stoel-Gammon, C. (1989): Prespeech and early speech development of two late talkers. First Language 9, 207–224.

Stoel-Gammon, C. (1998): Sound and Words im Early Language Acquisition: The Relationship Between Lexical and Phonological Development. In: Paul, R. (ed.), Exploring the speech–language connection, Paul H. Brooks Publishing, Baltimore, Maryland, *Communication and Language Intervention Series*, vol. 8, 25–52.

Stoel-Gammon, C. (2001): Transcribing the Speech of Young Children. Topics in Language Disproders 21(4), 12–21.

Swingley, D. (2003): Phonetic Detail in the Developing Lexicon. Language and Speech 46(2–3), 265–294.

Swingley, D. (2005): Statistical clustering and the contents of the infant vocabulary. Cognitive Psychology 50, 86–132.

Syrdal, A. K. and H. S. Gopal (1986): A perceptual model of vowel recognition based on the auditory representation of American English vowels. Journal of the Acoustical Society of America 79(4), 1086–1100.

ten Bosch, L. and B. Cranen (2007): A computational model for unsupervised word discovery. In: Proceedings of Interspeech 2007 (Antwerpen), 1481–1484.

Thiessen, E. D. and J. R. Saffran (2003): When cues collide: use of stress and statistical cues to word boundaries by 7– to 9–month–old infants. Developmental Psychology 39(4), 506–716.

Thiessen, E. D. and J. R. Saffran (2004): Infants' Acquisition of Stress–based Word Segmentation Strategies. In: Brugos, A., L. Micciulla, and C. Smith (eds.), BUCLD 28: Proceedings of the 28th annual Boston University Conference on Language Development, vol. 2, 608–619.

Thiessen, E. D. and J. R. Saffran (2007): Learning to Learn: Infant's Acquisition of Stress–Based Strategies for Word Segmentation. Language Learning and Development 3(1), 73–100.

Traunmüller, H. (1990): Analytical expressions for the tonotopic sensory scale. Journal of the Acoustic Society of America 88(1), 97–100.

Trommelen, M. and W. Zonneveld (1999): Dutch. In: Word prosodic systems in the languages of Europe, Berlin, New York: Mouton de Gruyter, 492–515.

Vallabha, G. K., J. L. McClelland, F. Pons, J. F. Werker, and S. Amano (2007): Unsupervised learning of vowel categories from infant–directed speech. Proceedings of the National Academy of Sciences of the United States of America 104(33), 13273–13278.

Vallabha, G. K. and B. Tuller (2002): Systematic errors in the formant analysis of steady–state vowels. Speech Communication 38, 141–160.

van Alphen, P., E. de Bree, P. Fikkert, and F. Wijnen (2007): The role of metrical stress in comprehension and production in Dutch children at-risk of dyslexia. In: Proceedings of Interspeech 2007 (Antwerpen), 2313–2316.

van de Weijer, J. C. (1998): Language Input for Word Discovery. Ph.D. thesis, University of Nijmegen. Spracherwerb, Input.

van der Hulst, H. (ed.) (1999): Word Prosodic Systems in the Languages of Europe. Emprical approaches to language typology; 20-4, Berlin; New York: Mouton de Gruyter.

Varley, R. and S. Whiteside (2001): What is the underlying impairment in acquired apraxia of speech. Aphasiology 15, 39–49.

Vennemann, T. (1986): Neuere Entwicklungen in der Phonologie. Berlin: Mouton de Gruyter.

Vihman, M. M. (1996): Phonological development: The origins of language in the child. Blackwell.

Vihman, M. M. (2002): Getting started without a system: From phonetics to phonology in bilingual development. The International Journal of Bilingualism 6(3), 239–254.

Vihman, M. M., R. A. DePaolis, and B. L. Davis (1998): Is there a "Trochaic Bias" in Early Word Learning? Evidence from Infant Production in English and French. Child Development 69(4), 935–949.

Vihman, M. M., M. A. Macken, R. Miller, H. Simmons, and J. Miller (1985): From babbling to speech: A re–assessment of the continuity issue. Language 61(2), 397–445.

Vihman, M. M., S. Nakai, and R. A. DePaolis (2006): Getting the rhythm right: A cross-linguistic study of segmental duration in babbling and first words. In: Goldstein, L., D. Whalen, and C. Best (eds.), Papers in Laboratory Phonology 8: Varieties of Phonological Competence, Berlin; New York: Mouton de Gruyter, 341–366.

Vihman, M. M., S. Nakai, R. A. DePaolis, and P. Hallé (2004): The role of accentual pattern in early lexical representation. Journal of Memory and Language 50, 336–353.

Vihman, M. M. and S. L. Velleman (2000): The construction of a first phonology. Phonetica 57, 255–266.

Vogel, I. and E. Raimy (2002): The acquisition of compound vs. phrasal stress: The role of prosodic constituents. Journal of Child Language 29, 225–250.

Vorperian, H. K. and R. D. Kent (2007): Vowel acoustic space development in children: a synthesis of acoustic and anatomic data. Journal of Speech, Language and Heraing Research 50(6), 1510–1545.

Vorperian, H. K., R. D. Kent, M. J. Lindstrom, C. M. Kalina, L. R. Gentry, and B. S. Yandell (2005): Development of vocal tract length during early childhood: A magnetic resonance imaging study. Journal of the Acoustical Society of America 117(1), 338–350.

Wade, T., G. Dogil, H. Schütze, M. Walsh, and B. Möbius (2010): Syllable frequency effects in a context–sensitive segment production model. Journal of Phonetics 38, 227–239.

Walsh, M., B. Möbius, T. Wade, and H. Schütze (2010): Multi-level Exemplar Theory. Cognitive Science 34, 537–582.

Werker, C. L., Janet F.and Stager (2000): Developmental changes in infant speech perception and early word learning: Is there a link? In: Broe, M. B. and J. B. Pierrehumbert (eds.), Papers in Laboratory Phonology V; Acquisition and the lexicon, New York: Cambridge University Press, 181–193.

Werker, J. F. and C. T. Fennell (2004): From listening to sounds to listening to words: Early steps in word learning. In: Hall, G. and S. Waxman (eds.), Weaving a Lexicon, Cambridge, Mass.: MIT Press, 79–109.

Werker, J. F. and L. Polka (1993): Developmental Changes in speech perception: new challenges and new directions. Journal of Phonetics 21, 83–101.

Werker, J. F. and R. C. Tees (1984): Cross–language speech perception: evidence for perceptual reorganization during the first year of life. Infant Behavior and Development 7, 49–63.

Whalen, D. H., A. G. Levitt, and Q. Wang (1991): Intonational differences between the reduplicative babbling of French- and English-learning infants. Journal of Child Language 18, 501–516.

Whiteside, S. P. (2001): Sex–specific fundamental and formant frequency patterns in a cross–sectional study. Journal of the Acoustical Society of America 110, 464–478.

Wiese, R. (1996): The Phonology of German. Oxford University Press.

Wokurek, W. and M. Pützer (2003): Automated Corpus Based Spectral Measurement of Voice Quality Parameters. In: Sole, M., D. Recasens, and J. Romero (eds.), Proceedings of the 15th International Congress of Phonetic Sciences, Barcelona, 2173–2176.

Yildrim, S., D. B. Narayanan, and S. Khurana (2003): Acoustic Analysis of Preschool Children's Speech. In: Proceedings of the 15th International Congress of Phonetic Sciences, Barcelona, 949–952.

Zamuner, T. S. (2009): Phonotactic probabilities at the onset of language development: Speech production and word position. Journal of Speech, Language, and Hearing Research. .

Zamuner, T. S., L. Gerken, and M. Hammond (2004): Phonotactic probalities in young children's speech production. Journal of Child Language 31, 515–536.