The Perception of Clauses in 6- and 8-Month-Old German-Learning Infants: Influence of Pause Duration and the Natural Pause Hierarchy

Dissertation zur Erlangung des Doktorgrades der Philosophie

vorgelegt von

Michaela Schmitz, M.A.

Institut für Linguistik / Allgemeine Sprachwissenschaft
Humanwissenschaftliche Fakultät
Universität Potsdam

Eingereicht: 28. Mai 2008
Disputation: 10. Oktober 2008

Erstgutachterin: Prof. Dr. Barbara Höhle

Zweitgutachter: Prof. Dr. Jürgen Weissenborn

In Memoriam

Margot M. Schmitz

1937 - 2004

οἶδα οὐκ εἰδώς

Scio nescio

'I know that I don't know'

Sokrates (469 - 399 v. Chr.)

What we know is a drop.

What we don't know is an ocean.

Sir Isaac Newton (1643 - 1727)

Acknowledgements

Much of this study was conducted while I had the opportunity to work in the project C3 „Erwerb sprachlicher Mittel der Markierung von Informationsstruktur im Erstspracherwerb: Prosodische, syntaktische und lexikalische Aspekte“ of the DFG-funded Sonderforschungsbereich 632 „Informationsstruktur: Die sprachlichen Mittel der Gliederung von Äußerung, Satz und Text“.

How infants acquire their native language is a fascinating and intriguing topic. I have learned much about these issues during my time spent in Potsdam. I am very grateful for this time and this opportunity. My thanks go especially to Prof. Barbara Höhle and Prof. Jürgen Weissenborn for introducing me to the fascinating field of early language acquisition, for supervising and constantly supporting me during the long time it took to finish this work. Most of all, I'd like to thank the infants and their parents for participating in the experiments of this study. Without them, this work could never have been accomplished!

A special thank goes to Kerstin Leuckefeld and the Max-Planck-Institut für Kognitions- und Neurowissenschaften in Leipzig for providing the original auditory material which I used in my experiments, and to Claudia Männel for sharing and discussing the results of her work with me.

I would also like to thank my former and present colleagues: Frauke Berger, Heiner Drenhaus, Ines Fiedler, Jeannine Gies, Renate Kaldewey, Frank Kügler, Theo Marinis, Svetlana Petrova, Diana Pili, Anne Schwarz, Anke Sennema, Julia Siegmüller, Stavros Skopeteas, Nicole Stadie, Anja van Kampen, Ruben van de Vijver, and most of all Anja Müller, who shared the office with me for the last five years. Furthermore, I would like to thank all the colleagues and student assistants who have worked with me in the Babylab in Potsdam over the last 10 years: Sonja Bartels, Sylvia Dobler, Eva Gercke, who willingly undertook the laborious task of proofreading this thesis, Tom Fritzsche, Susan Ott, Lydia Pelzer, Anne Rosenthal, who was mainly responsible for the second coding of the data I present here, Caroline Schröder, Antje Schulz, and Katrin Skoruppa. It was a pleasure to have worked with you and I'd like to thank you all for the support I've received in various ways, and for the time we spent together discussing work and life.

A heartfelt thank goes to my friends who always believed in me, even when I myself didn't: Joanna Błaszczak, Birgit Herold, Ioulia Iouannidou, Anja Ischebeck, Christina Kauschke, Janette Timmer and Sabine Völker-Straub, some of you have been with me in some of the darkest hours of my life. I thank you all for your friendship and encouragement, and for answering that one telephone call in the middle of the night. And a special thank goes to Frank Wesel, for many a good conversation and many a good dinner – and for introducing me to Tango Argentino.

Last but not least, I'd like to thank my family: my mother, to whom this work is dedicated, my father, my sister and her family for their emotional support and confidence in me, and my aunt, Ute Schmitz, for keeping me company when I needed her.

Table of Contents

List of Abbreviations

ADS	Adult-Directed Speech
BP	Breathing Pause
CDS	Child-Directed Speech
CHT	Conditioned Headturn Paradigm
CPS	Closure Positive Shift
ERP	Event-Related Potentials
fMRI	Functional Magnetic Resonance Imaging
HAS	High-Amplitude Sucking Paradigm
HTP	Headturn Preference Paradigm
IPh	Intonational Phrase
ISI	Interstimulus Interval
MMN	Mismatch Negativity
MMR	Mismatch Response
ms	Milliseconds
MSS	Metrical Segmentation Strategy
NBP	Non-breathing Pause
NIRS	Near Infrared Spectroscopy
NP	Noun Phrase
OT	Orientation Time
PPh	Phonological Phrase
s	Second
S V O	Subject Verb Object
VP	Verb Phrase

List of Figures

1 Introduction

Language acquisition is a process which starts – at the latest – at the moment of the birth of an infant. When Mehler and colleagues (Mehler, Jusczyk, Lambertz, Halsted, Bertoncini & Amiel-Tison, 1988) found that newborns are able to discriminate their native language from another language on the basis of the differences in the global rhythmic properties of both languages, one can be inclined to think that this sensitivity to general rhythmic properties is based on previous experience. And indeed, several studies have shown that newborns prefer their mothers' voice to that of some other speaker (DeCasper & Fifer, 1980) and that they prefer speech which has a frequency range that matches the range which could be heard intra-uterine (cf., Lecanuet & Granier-Deferre, 1993; for an overview see Aslin, 1987; Lecanuet, 1998).

So in the strict sense there is no "pre-linguistic" stage in the developmental process of language acquisition. As soon as the auditory system is fully developed, the foetus starts to recognise differences between auditory stimuli. This was shown in several studies on the perception of sounds and speech in foetuses in the last trimester of the pregnancy (Lecanuet, Granier-Deferre, Jacquet & Busnel, 1992; Lecanuet, Granier-Deferre, Jacquet, Capponi & Ledru, 1993).

Despite the previous, and certainly limited, intra-uterine experience with the prosody of their native language, the infant is in a position comparable to an adult experiencing the first contact with a language totally foreign to her: the infant is confronted with an ongoing speech stream and has to identify relevant prosodic and / or syntactic units by attending to the prosodic characteristics provided by the input. This segmentation task is by no means trivial as there are few reliable cues, like pauses, that indicate word boundaries (Cutler, 1994). However, some cues are present at intra- and inter-sentence boundaries, like pauses, changes in fundamental frequency, and lengthening of the pre-boundary syllable (Cooper & Paccia-Cooper, 1980; Garnica, 1977). Thus, by discovering the prosodically marked units in the speech stream, like intonational phrases, the infant can ultimately discover syntactic units, like clauses, of her language.

The present study will focus on the question whether and under which conditions German-learning infants recognise clauses in fluent speech and the role a prosodic marker such as a pause may have in the segmentation process. This question is important since by being able to recognise clauses in fluent speech, the infant has access to the domain in which important grammatical rules apply. A clause minimally consists of a subject and a finite verb, as in *Mary sleeps*, or is more complex in that it consists of, for example, a subject, a finite verb and an object, as in *The boys play football*.[1, 2] These two examples show that subject-verb agreement is realised within the domain of the clause. Furthermore, clauses differ cross-linguistically in the ways the order of their constituents are realised, for instance, the difference between SVO (e.g., English) and SOV (e.g., Turkish) languages. On the clause level, but not on the phrase level, the infant can learn whether in her native language the subject of a clause must be overtly spelled out as in German or English, or can be left out, like in Italian, which is a so-called pro-drop language, compare: *Vedo un aero* 'see-1SG an airplane' vs. *I see an airplane* (example from Guasti, 2004, p. 153). Thus, these examples show that the clause is an important domain in the acquisition of the grammar of a language.

In the speech signal, clauses often coincide with intonational phrases (Nespor & Vogel, 1986, p. 190). The boundaries of these intonational phrases, and especially the right boundary, are marked by changes in the fundamental frequency, like a pitch fall or rise (e.g., Price, Ostendorf, Shattuck-Hufnagel & Fong, 1991), lengthening of the final syllable (e.g., Cooper & Paccia-Cooper, 1980; Wightman, Shattuck-Hufnagel, Ostendorf & Price, 1992) and the occurrence of a pause (Nespor & Vogel, 1986, p. 188). Thus, intonational phrases seem to be reliably marked by these cues in the speech stream, especially in child-directed speech (Fisher & Tokura, 1996a) and this may be used by infants to recognise them, but

> … it is unlikely that all languages employ the same set of prosodic cues to mark phrase [and clause] boundaries. In certain languages vowel lengthening is phonemic […], those languages do not use vowel lengthening as a cue […]. Tone languages […] use pitch pattern to make lexical distinctions and probably do not use such patterns to mark boundaries between phrases. (Morgan, 1986, p. 112)

---

[1] Clauses also may contain non-finite verbs, like the ones used in this study: *The aunt promises Katrin to swim and to explore the island.*

[2] In these examples the clause is also a complete sentence. However, a sentence may be composed of one or more clauses as, for instance, in *Mary sleeps and the boys play football*.

These considerations then might suggest that a pause could function as a reliable cue to indicate intonational phrase boundaries. Pauses occur in every language because of their physiological nature, that is, they are breaks in which the speaker can take breath during speech production, and these breaks often occur at the boundaries of syntactic / prosodic units. This is not the only function a pause can have in the speech production process. Often, hesitation pauses are produced to further plan the utterance. And (non-breathing) pauses are also used to demarcate syntactic / prosodic units in the speech (e.g., Grosjean, Grosjean & Lane, 1979; Grosjean & Collins, 1979, see also section 2.2.3.4).

Furthermore, corpus studies on the occurrence and distribution of pauses in several languages have revealed that there is a strong correlation between the duration of a pause and the type of boundary it marks (e.g., Goldman-Eisler, 1972, for English; Butcher, 1981, for German). Pauses between words are either non-existent or short, pauses between phrases are a bit longer, pauses between clauses are again longer than pauses between phrases, and pauses occurring at sentence boundaries have the longest duration. This suggests the existence of a natural pause hierarchy that complements the prosodic hierarchy described by Nespor and Vogel (1986). These hierarchies on the side of the speech signal correspond to the syntactic hierarchy of a language. Although there is not always a perfect one-to-one match, often a phonological utterance corresponds to a sentence, an intonational phrase corresponds to a clause, a phonological phrase corresponds to a syntactic phrase and a phonological word corresponds to a lexical word.

Thus, if infants, who mainly rely on bottom-up strategies during their language acquisition process, were able to recognise the prosodically marked units in the speech stream, they subsequently would have access to the grammatical units of their language. The mechanism that was suggested to be underlying this process is the *prosodic bootstrapping mechanism* (Gleitman & Wanner, 1982), which states that infants learn about one level of the language, the syntax, by exploiting regularities on another level of the language, in this case the prosodic level. The prosodic bootstrapping mechanism is compatible both with *inside-out* and *outside-in* theories of language acquisition (terminology by Hirsh-Pasek & Golinkoff, 1996).

Inside-out theories are defined as theories which assume that the child has innate domain-specific, that is, language-specific capacities and that the language acquisition process is driven by the linguistic structure itself. The most prominent and influential representative of the inside-out direction is the nativist theory inspired by Chomsky (1965). He was led by the so-called *logical problem of language acquisition*, that is, how children can acquire the complex linguistic system of their native language, given that they have only limited input (= positive evidence) and no negative evidence, which is the *poverty-of-the-stimulus argument* (cf., Guasti, 2004). Chomsky therefore postulated the existence of a language acquisition device, or what Pinker (1994) called an “instinct”, which allows any human being the acquisition of a language, either spoken or signed. In the framework of the inside-out theories, the child is innately predisposed to acquire the language and only has to uncover the underlying regularities appropriate for her language.

Outside-in theories are defined as theories which suppose that the child is equipped with innate but domain-general, that is, not language-specific learning mechanisms and that language acquisition takes place through the interaction with the environment, linguistic and non-linguistic. One theory within this framework suggests that the child’s language acquisition takes place by the interaction of the child’s domain-general learning capacities, like a distributional analysis of the language, and the cognitive abilities which allow the identification of, for instance, thematic roles like agent and patient. The categories then can be mapped on the output of the distributional analysis of the speech stream. Learning proceeds possibly by expanding the cognitive domain and re-analysis (cf., Bates & MacWhinney, 1989).

Thus, both theoretical approaches allow for innate knowledge, either being domain-specific or being domain-general. However, whether the parsing of speech and the perception of prosodically marked units in the speech stream, which leads to the identification of syntactic units, is driven by domain-specific knowledge or by domain-general learning mechanisms is still a matter of debate (Jusczyk, 1997, p. 38). Nevertheless, the idea that an infant makes use of the prosodic analysis of the speech stream to learn about the syntactic structure of her language was the starting point for an increasing number of studies in infants’ language acquisition research.

Many researchers studied infants’ ability to recognise words (2.2.3.1), phrases (2.2.3.2) and clauses (2.2.3.3) in fluent speech. The principal method of investigation which was used in many studies to address these topics was the Headturn Preference

(HTP) paradigm, which also will be used in the present study (3.1). Depending on the specific design of the experiment, this method can be used either to test infants' recognition of speech units and the learning mechanisms or to test what infants already know about their language. The second aspect is more important in the context of the present study. To test whether infants use pause as a cue to identify clause boundaries and whether they are sensitive to the natural pause hierarchy it is necessary to investigate infants' reactions based on their previous knowledge.

In a seminal study, Hirsh-Pasek and colleagues (Hirsh-Pasek, Kemler Nelson, Jusczyk, Cassidy, Druss & Kennedy, 1987) investigated whether English-learning 6- to 8-month-olds were able to recognise clauses in fluent speech. The results showed that infants preferred to listen to the speech passages in which artificial 1 s pauses coincided with the clause boundary compared to speech passages that had artificial pauses inserted within a clause. Thus, the authors concluded that infants are able to recognise well-formed clausal units in the speech stream. A further study on infants' recognition of clauses conducted by Nazzi and colleagues (Nazzi, Kemler Nelson, Jusczyk & Jusczyk, 2000b) also revealed that infants preferred to listen to clauses in the speech. These studies provide convergent evidence that infants as young as 6 months recognise clauses in the speech stream.

Subsequent studies showed that, among other strategies used to detect words in fluent speech, infants are also able to use the sensitivity to clause boundaries to begin segmenting words from the speech stream (e.g., Seidl & Johnson, 2006; Schmitz, Höhle, van de Vijver & Weissenborn, submitted). Thus, infants seem to be sensitive to the prosodic correlates of clause boundaries and are able to use these sensitivities to further analyse their input.

There are not many studies addressing the question of whether infants are sensitive to the occurrence of pauses, though some studies used to insert pauses in their test stimuli. The results of these studies indirectly provide evidence that infants seem to be sensitive to the occurrence of pauses at the clause boundary because infants preferred to listen to the clausal units where the boundary markers converged. Only recently the question whether infants use the occurrence of a pause as a cue to identify clause boundaries in fluent speech was addressed. In a series of studies, Seidl and colleagues (Seidl, 2007;

Johnson & Seidl, 2008; Seidl & Cristià, 2008) manipulated the occurrence of pauses at the clause boundary to investigate whether English- and Dutch-learning infants used pauses as a cue to identify clauses in fluent speech. They found that Dutch-learning 6-month-olds, but not English-learning 6-month-olds, seemed to rely on the pause as a necessary cue to identify clause boundaries. The English-learning 6-month-olds seemed to rely more on the pitch cue to recognise the clause boundaries in the speech.

These studies show that the pause is a decisive feature in the speech stream and that infants seem to use the pause as a cue to identify clause boundaries in fluent speech, though the weighting of the different cues to clause boundaries seems to be language-specific. However, none of the previously cited studies concentrated on the correlation between the type of boundary tested, that is, phrase boundaries vs. clause boundaries vs. sentence boundaries, and the duration of the pause that is associated with this boundary. More specifically, until now, the question of whether infants, who seem to rely on the pause as a cue to identify clause boundaries, are also sensitive to the natural pause hierarchy has not been posed. These questions will be addressed in the present study by first, testing infants' sensitivities in a previously not studied language, namely, German. And in a second step, the question is addressed whether the German-learning infants are sensitive to the natural pause hierarchy, reflected in different pause durations at the within-sentence clause and the sentence boundary, respectively.

The present dissertation is divided into two main parts: Part I provides the theoretical background for the empirical study presented in part II.

Following this introduction, chapter 2 reviews previous research on the main three questions regarding the recognition of clauses and pauses by infants, which are:

1. Are the cues indicating syntactic / prosodic boundaries, like changes in the fundamental frequency, final lengthening, and especially pauses, present in the input?
2. Are infants sensitive to units like, words, phrases, clauses and pauses in the input?

3. Do infants use these cues, and especially the pause cue, in their recognition of clauses in fluent speech?

With regard to the first question, studies are presented which examined whether the relevant cues marking prosodic / syntactic boundaries are available in the input (2.1). The focus is on studies which investigated the occurrence of pauses and their use and distribution in the speech input, more specifically, in the speech among adults, in the speech of children, and in child-directed speech. The results of these studies suggest the existence of a natural pause hierarchy that complements the prosodic hierarchy on the signal side. These hierarchies are closely related to the syntactic hierarchy of the language (Figure 2.1).

With regard to the second question, studies are presented which investigated whether infants are sensitive to the speech rhythm of their language and use this sensitivity by means of a prosodic bootstrapping mechanism to further analyse their input (2.2). Subsequently, studies on infants' sensitivities to words, phrases, clauses and pauses, are reviewed (2.2.3). Complementary, studies on adults' sensitivities to clauses and pauses will be discussed.

With regard to the third question, studies are presented which addressed the question whether adults and infants use the cues that are present in the input, especially the pause cue, to recognise clause boundaries in fluent speech (2.3). Part I concludes with a short summary (2.4).

In part II of the dissertation, chapter 3 introduces the principal method of investigation, the HTP paradigm (3.1). Theoretical considerations concerning the validity of the method and the direction of the effects found in the results of experiments using this method will be discussed. Furthermore, the sentence materials used in the subsequent experiments are presented, as well as the acoustic analyses of these sentences (3.2). Then an outline of the experimental study is given (3.3).

Chapter 4 presents five experiments which investigate whether German-learning infants use the pause cue to recognise clauses in fluent speech, and whether the infants are sensitive to the natural pause hierarchy. The first three experiments address the question whether and under which pause duration conditions German-learning 6-month-olds are able to recognise clause boundaries in fluent speech (4.1 – 4.3). The fourth experiment with 6-month-olds and the fifth experiment with 8-month-olds, directly address the question whether the recognition of within-sentence clause boundaries and sentence boundaries is guided by the infants' sensitivity to the natural pause hierarchy (4.4, 4.5).

Chapter 5 summarises the results of the present study and discusses the implications of the findings for the language acquisition process.

I Theoretical Background

2 Clauses and Pauses

How do infants manage to segment structures like words, phrases and clauses from fluent speech using bottom-up strategies? This question bears some importance, because usually infants are confronted with ongoing speech which contains no clear word boundary markings, such as pauses comparable to the blanks between words in written language (e.g., Cutler, 1994), and only some prosodic marking of larger units. In his study on child-directed speech (CDS) to a six-month-old infant, van de Weijer (1998) found that about 40% of the utterances were single word utterances and most of these single words were vocatives, social words or fillers. Without these vocatives, social words and fillers only 7% of the parental utterances were single word utterances. Moreover, even when parents were explicitly instructed to teach their children new words, the target word was produced in multi-word utterances in a majority of the cases (Aslin, Woodward, LaMendola & Bever, 1996). Thus, infants might use the occurrences of single words as a database on which they build early lexical representations as has been argued by Brent and Siskind (2001). However, this limited exposure to single words does not seem to suffice to acquire a language and especially its grammar. Infants also must be able to segment words and larger units like phrases and clauses, in which morpho-syntactic rules apply, from the far more often heard strings of concatenated speech. Pauses may play an important role in this segmentation process, because pauses, besides having other functions, often demarcate larger units like phrases, clauses and sentences in the speech stream. Furthermore, the occurrence of pauses to demarcate linguistic units seems to be universal across languages (Morgan, 1986) and thus might qualify as a good segmentation cue for the language-learning infants.

To pose the question of how infants might be able to segment fluent speech is to make three assumptions, namely, that the input provides cues to each of these structures, that infants are sensitive to these cues, and finally, that they use them (cf. e.g., Jusczyk, 1997). This chapter will present studies and research related to all three of these topics.

2.1 Prosodic cues in the input

2.1.1 Production of prosodic cues marking syntactic boundaries

Within the speech stream there are no reliable markers to word boundaries (e.g., Cutler, 1994). However, larger units in the speech stream, like phrases and clauses, are usually marked at their boundaries. These markings include the lengthening of the final syllable preceding the boundary (e.g., Cooper & Paccia-Cooper, 1980; Wightman et al., 1992) and a change in the fundamental frequency (e.g., Price et al., 1991). Furthermore, these larger units are often followed by pauses (see section 2.1.2). Most languages use a combination of these markers to signal a boundary, but the weighting of these cues seems to be language-specific (Gerken, 2001, p. 150f).

Kohler (1983) found that the duration of a bisyllabic word in German strongly correlates with the position of this word within an utterance. In utterance-medial position, the first syllable was about five times longer than the second one. When the same word occurred at a prosodic boundary, the first syllable was only about twice as long as the second syllable. This change is due to a proportionally much higher increase of the duration of the second syllable as compared to the first one. These data provide evidence of a considerable lengthening of the final pre-boundary syllable. In a recent study, Schmitz and colleagues (Schmitz et al., submitted) also found that whole words, and especially the final syllable produced at the utterance boundary had a significantly longer duration, than the same words produced within the utterance.

The same pattern emerged for changes in the fundamental frequency at a prosodic boundary compared to other locations within the sentence. In the analysis of their stimuli Schmitz et al. (submitted) found that words at the end of the utterance had a significantly lower pitch than the same words within the utterance. Furthermore, the second syllable of the pre-boundary word had a lower pitch than the first syllable of that word, whereas in utterance medial position the pitch of the second syllable was higher than the pitch of the first syllable, thus indicating ongoing speech. These data indicate that in German utterance boundaries are reliably marked by changes in the fundamental frequency and a longer duration of the final syllable preceding the boundary.

The prosodic hierarchy

The data on changes that mark a prosodic boundary together with the observation that certain phonological rules apply with phonological constituents but are blocked across these constituents led Nespor and Vogel (1986) to formulate their theory of *Prosodic Phonology*. They proposed a hierarchical prosodic organisation of the language which is interrelated but also sometimes independent from the also hierarchical syntactic organisation of the language. In the prosodic hierarchy, the largest unit, the *Phonological Utterance* is composed of one or more *Intonational Phrases* (IPh). These are composed of one or more *Phonological Phrases* (PPh), which are composed of one or more *Clitic Groups*. These *Clitic Groups* consist of one or more *Phonological Words*, which are composed of *Feet* and, as the smallest units, *Syllables*.

Nespor and Vogel (1986) also assumed that each of these levels is characterised by a relative prosodic prominence between the components, that is, one component is defined as strong while the other components are defined as weak elements (ibid., p. 7). The strong element receives primary stress. The weak elements are either unstressed or may receive secondary stress. On the word level for instance, bisyllabic words can be stressed on the first or on the second syllable. In a right-branching language like English or German, the rightmost element of the phonological phrase receives strong stress and all other components receive weak stress.[3] On the intonational phrase level, the position of the most prominent element may be influenced by information structural requirements, such as focus assignment. On each level in the prosodic hierarchy there are recurrent prosodic patterns.

Even if there are not always one-to-one mappings of the prosodic units to the syntactic units as Nespor and Vogel (1986) have shown, in many cases both hierarchies overlap: the phonological utterance coincides with a syntactic sentence, as well as the intonational phrase coincides with a syntactic clause, the phonological phrase with a syntactic phrase, and the phonological word with a lexical word (cf., Figure 2.1). Furthermore, the larger the syntactic / prosodic unit is, that is, the higher in the hierarchy it is ranked, the more reliably and consistently are its boundaries marked by changes in

[3] For example in the NP [the man], the determiner is unstressed whereas the noun is stressed. The same holds for VPs. In [buys the bread], the verb is unstressed and the complement receives phrasal stress.

fundamental frequency and longer durations of the pre-boundary element (Cooper & Paccia-Cooper, 1980) as well as by an increasing duration and a more reliable occurrence of pauses between the constituents, especially in CDS (Fisher & Tokura, 1996a, 1996b).

2.1.2 Production of pauses

Pauses, which are part of suprasegmental phonology, serve many functions in the language. One of these functions, which is of interest in the present study, is related to the demarcation of prosodic and / or syntactic units. The studies that will be presented in the next sections used different terms to describe the entities under investigation. Some of them investigated the occurrence and duration of pauses in relation to syntactically defined units, that is sentences and clauses, etc. (cf., Goldman-Eisler, 1972) and others investigated the occurrence and duration of pauses in relation to prosodically defined units, like intonational phrases, etc. (e.g., Ferreira, 1993) and even in relation to acoustically defined units (cf., Fernald, Taeschner, Dunn, Papousek, De Boysson-Bardies & Fukui, 1989). However, the overall picture that emerges from the different studies is cohesive.

Other functions of pauses in the language include pauses for breathtaking and hesitation pauses for language planning and emphasis, and their occurrence is not only determined by syntax, however, "… speakers undoubtedly combine grammatical pauses, BPs [breathing pauses] and hesitation pauses in their everyday speech." (Grosjean & Collins, 1979, p. 100).

2.1.2.1 Breathing pauses

One important function of a pause is of course to allow the speaker to take breath during the production of speech. Butcher (1981) cites several studies which investigated the occurrence of pauses used for breathing. For instance, it was found in a study on spoken Hungarian that all breathing pauses coincided with syntactic boundaries, that is, 88% occurred between sentences, 9% between clauses and 3% at other phrase boundaries (ibid., p. 13).

Grosjean and Collins (1979) investigated if and how breathing pauses were related to syntactic constituents in English. They studied pause occurrence by instructing six adult speakers to read a text passage at ten different speech rates, both faster and slower than the baseline rate for each speaker. They found that the number and duration of inserted pauses decreased as a function of speech rate, with about 23% non-breathing pauses (NBPs) with a mean duration of 0.5 s and 21% breathing pauses (BPs) with a mean duration of 10 s at the slowest speech rate to all pauses in fact being BPs with a mean duration of 0.4 s at the fastest speech rate. That is, the faster the speech rate, the less pauses were produced and only the need to breathe induced the occurrence of a pause. Over all speech rates, BPs tended to be twice as long as NBPs.

In the analysis of the correlation of the occurrence of pauses with the occurrence of major syntactic boundaries, Grosjean and Collins (1979) found that in slow reading both BPs and NBPs are found equally often at the sentence and at the clause boundary, in this case, before a coordinating conjunction. Both pause types had a mean duration of 1.3 s at the sentence boundary and 0.8 s at the clause boundary and together constituted 78% of all pause locations during slow reading. Again the occurrence and the duration of pauses decreased as a function of speech rate. The faster the speech rate, the less pauses were produced. At the fastest speech rate only BPs at the sentence boundaries were produced. The authors concluded that the occurrence of BPs did not only depend on the speech rate but also on the nature of the linguistic constituent, with more and longer BPs being produced at higher level syntactic boundaries and less and shorter BPs being produced between words within phrases. Thus, at normal speech rate, the speaker used both BPs and NBPs between constituents, and with increasing speech rate the number and duration of pauses decreased and the speaker used a pause mainly at major syntactic boundaries to take breath.

These findings were supported by a study on German conducted by Butcher (1981). He found that in spontaneous speech as well as in read texts the majority of BPs coincided with major prosodic boundaries (ibid., p. 112).

These studies provide evidence that speakers coordinate their need to breathe and their language planning, in that most breathing pauses coincide with major syntactic / prosodic boundaries, like the ends of clauses and sentences.

2.1.2.2 Hesitation pauses

The production of a second type of pauses is strongly related to the planning of speech or to word-finding during the speech production process. These pauses may be filled by filler words such as *ehm* [Em], *äh* [æ] in German and *uh* [u] or *um* [um] in English, or they may be silent pauses. Filled pauses may occur before larger syntactic units, as an indication of language planning, or be used to maintain the speech act, that is, to prevent the conversation partner from interrupting the flow of speech (Butcher, 1981, p. 17ff). Silent hesitation pauses also may be used to plan the language activity before larger syntactic units, but may also occur within syntactic units (cf., Reich, 1980). However, Ferreira (1993, footnote 1) pointed out that these hesitation pauses are distinct from what she called "timing based pauses" (ibid., p. 234), which are pauses following a prosodic boundary, like a phonological word, phonological phrase or intonational phrase (cf., Nespor & Vogel, 1986).

2.1.2.3 Pauses as cues to syntactic and prosodic boundaries

The third function of a speech pause, and the one which is most important for the present study, is to demarcate syntactic and / or prosodic boundaries in the speech stream. As has already been discussed in section 2.1.2.1, Grosjean and Collins (1979) found a correlation between pause occurrence and syntactic constituents.[4] In normal speech, sentence boundaries, clause boundaries and phrase boundaries (= NP / VP boundaries) were consistently marked by (NB-)pauses. Furthermore, shorter pauses were sometimes also found at locations between words within NP-, VP-, and PP-phrases. As a function of speech rate, the occurrence of NBPs decreased depending on their level within the syntactic hierarchy, that is, pauses between words within phrases were produced less than pauses between phrases than pauses between clauses. The next sections will present further studies investigating the occurrence and duration of pauses to indicate syntactic and prosodic boundaries in the speech stream.

[4] This correlation was also found in American Sign Language: although "breathing is not related to signing, signers also pause at important syntactic boundaries." (Grosjean & Collins, 1979, p. 111f)

Production of pauses in the speech of adults

In an earlier study, Goldman-Eisler (1972) found a comparable distribution of pauses as has been observed by Grosjean and Collins (1979). In her corpus study she included samples of English adult-directed speech (ADS) not only from read texts, but also from discussions, unscripted radio talks and natural conversations. She investigated the occurrence and duration of pauses in relation to the syntactic level, that is, words, different types of clauses and sentences of the speech in the corpus. The analyses of the natural speech samples revealed that more than 93% of transitions between words were fluent, that is, the pauses had a duration ranging from 0 to 250 ms. Between sentences, only 10% of transitions were fluent, about 35% of the pauses had a duration below 750 ms, and more than 50% of the pauses which occurred between sentences had a duration of more than 1 s.; 15% of these long pauses between sentences even had a duration of more than 2 s. The pause pattern that emerged for coordinated clauses, that is, sentences consisting of two clauses conjoined by a coordinating conjunction comparable to the stimuli used in the present study, is the following: about 33% of the transitions between the clauses were fluent. However, 43% of the pauses between clauses had a duration of more than 500 ms, with a proportion of more than 20% of the pauses between clauses having a duration between 750 ms and 1 s.

Goldmann-Eisler's (1972) findings on the pause pattern in natural speech show a hierarchical pause structure correlated with the syntactic level of constituents. Although there is no unequivocal correlation between the duration of a pause and the syntactic level, the pattern that emerges from the analyses shows that an overwhelming number of pauses with the shortest duration (0-250 ms) are produced between words. Clauses are in most cases separated by pauses of at least 500 ms duration and the duration of most pauses between sentences is more than 1 second. Thus, speakers tend to mark the syntactic levels of clauses and sentences quite reliable with pauses of different durations.

Grosjean and colleagues (Grosjean et al., 1979) also found a high correlation between the duration of pauses and the surface syntactic structure of a sentence (ibid., p. 69). The syntactic structure of a given sentence was a good predictor of pause durations in most cases. The longest pause durations were observed at clause boundaries, followed by shorter pause durations at phrase boundaries. However, in sentences in which the first clause consisted of a short Subject-NP, for instance a name, which was followed by a long VP, for instance a verb and a complex Object-NP, speakers also tended to divide the clause into chunks of equal length and produced a pause within the complex Object-NP. Thus, the production of a pause was seemingly influenced by some abstract timing requirement and not only by syntactic structure. The authors therefore concluded that the insertion of pauses in fluent speech is guided by "two (sometimes conflicting) demands […]: the need to respect linguistic structure […] and the need to balance the length of constituents in the output." (ibid., p. 75). Therefore, the data suggests that placement of pauses was not only guided by syntax, but sometimes also by the speaker's attempt to divide speech into chunks of equal length.

In a comprehensive corpus study Butcher (1981) investigated the occurrence and functions of speech pauses in German. The corpus included read texts and natural conversation. Across all conditions he found a total of 1151 pauses. The distribution of all pause durations showed that the pauses clustered around three main duration intervals: a short pause with a duration of 100 – 200 ms, a medium pause with a duration of 500 – 600 ms and a long pause with a duration of 1100 – 1200 ms. The same pattern emerged for the separate analyses of the spontaneous and the read speech samples, respectively. Furthermore, the majority of pauses over 600 ms occurred at junctures[5], that is, at phonological or intonational phrase boundaries or at an utterance boundary. A pause duration of 100 – 200 ms might indicate a non-juncture break, that is a pause between successive words, as well as a juncture break (ibid., p. 73f).

With respect to syntactic constituents, Butcher (1981) found the highest amount of juncture pauses before clause connectives, including coordinating and subordinating conjunctions, that is, most of the pauses occurred between clauses. Of these, the pauses before coordinating conjunctions, like *und* 'and' were the longest (average between 607 - 857 ms), followed by pauses before subordinating conjunctions, like *dass* 'that' (average between 433 – 575 ms) and shortest before a relative subordinating conjunction, like *wo* 'where' (average between 300 – 350 ms) depending on read or spoken modality.

[5] A juncture is defined by Butcher (1981, p. 61) as "a set of prosodic features marking the end of a tone group" or "breath-group". According to Lieberman (1996, p. 61) the boundaries of a breath-group are marked by a fall in the F0, lengthening of the final segment and "a sudden sharp inspiration" (= a breathing pause). A breath-group often spans the prosodic unit of a whole sentence, but may also correspond to a 'phonemic' phrase or 'phonemic' clause (Lieberman, 1966).

Furthermore, as has been observed in other studies as well, the most frequent and longest pauses occurred between sentences (average between 797 – 1120 ms), whereas pauses between words within phrases were less frequent and overall shorter (ibid., p. 168ff). Thus, the data found by Butcher (1981) for German speech corpora corresponds to the data found for English. This cross-linguistic evidence supports the concept of a pause hierarchy which is interrelated to the prosodic and the syntactic hierarchy in language.

However, the degree of influence that either the prosodic structure or the syntactic structure have on the realisation of pauses in speech is still not fully determined. In her paper, Ferreira (1993) therefore posed the question "whether the syntactic representation of a sentence directly influences phonological variables […] or whether syntax is used to construct an intermediate representation, a prosodic representation, which in turn influences variables such as word duration and stress" (ibid., p. 233f) and pausing. In her experiments she tested the influence of syntactic and prosodic structure by creating sentences in which each variable was manipulated separately. For instance, the syntactic structure remained unchanged, whereas the prosodic structure differed, like in *The crate contains the missing book*. vs. *The CRATE contains the missing book*. (emphasis by Ferreira, 1993, p. 238). If the duration of the word *crate* at the NP / VP boundary and a subsequent following pause depended on syntax, no differences in the realisation of both sentences were expected. However, if the realisation of the word was guided by prosody, a longer duration of the focused word *crate* and a longer pause were expected in the second sentence. That is, the first sentence was expected to be produced in a single IPh, whereas the second sentence was expected to be produced in two IPhs. Vice versa, Ferreira (1993) tested the duration of the target word (*cop*) and the occurrence of a pause in a condition in which the syntax of the test sentence was changed, but the prosody remained unchanged, like in *The friendliest cop infuriated the boyfriend of the girls* vs. *The friend of the cop infuriated the boyfriend of the girls* (ibid., p. 239).

Both experiments provided evidence that the realisation of the target word and the occurrence of a pause were guided by the speaker's prosodic realisation of the sentence and not simply by the syntactic structure of a sentence. Especially in the condition in which the target word occurred at different types of syntactic phrase boundaries, the duration of the pauses did not differ in the speech samples, but each differed from the neutral control condition, where the target was in a phrase medial position. Thus, Ferreira (1993) concluded that the syntactic structure of a sentence did not directly influence the realisation. Instead, in speech production the syntactic structure seemed to be transformed during the speech planning into prosodically relevant units which determined such variables as word and pause duration.[6] Thus, the production of pauses in speech seems to be determined by the interplay between the syntactic and the prosodic hierarchies.

Further support for these findings comes from a study by Krivokapić (2007). She showed that the duration of a pause is determined by factors such as the length and / or the complexity of an IPh, and not the syntactic structure per se.

Production of pauses in the speech of children

The previously presented studies on pauses in speech production all relied on adult informants. But do children also produce pauses in the speech in an adult-like and therefore target-like manner? This question was addressed in a study by Katz and colleagues (Katz, Beach, Jenouri & Verma, 1996), who compared utterances from adults and 5- and 7-year-old children. The participants were asked to describe a scene where three cubes, a pink one, a green one and a white one, were presented. In the scene, these cubes were either all separate or two of them were grouped together and one cube was separate. This elicited utterances like *pink and green and white* or, for instance, *[pink and green]*[7] *and white*. Katz et al. (1996) analysed the variables word duration, F0 change and pause duration at the phrase boundary. The data show that adults consistently used longer pauses to indicate phrase boundaries, that is, in the *pink and green and white* condition (all blocks separate), the pauses after pink and green had about equal duration, whereas in the *[pink and green] and white* condition, the pause after *green* was significantly longer than in the other condition. The children, however, did not reliably produce longer pauses to indicate phrase boundaries. Furthermore, the children also did not use final lengthening

---

[6] The theoretical model that Ferreira (1993) proposed in which she relates word duration and pause duration as instances of an overall timing of utterances will not be discussed here.

[7] The brackets indicate the grouping of the elements.

or changes in the F0 to indicate phrase boundaries in this task, whereas the adults consistently did.

These findings were further supported by a study from Dankovičová and colleagues (Dankovičová, Pigott, Wells & Peppé, 2004), who also found that 8-year-old children show great inter-individual differences in the marking of phrase boundaries. These children also did not show an adult-like usage of prosodic correlates to phrase boundaries.

Taken together, the studies provide evidence that children, at least up to schoolage, do not use pauses, and other prosodic cues, systematically to mark prosodic boundaries. There are no studies investigating the developmental path in this area, to find out when children first show adult-like performance.

Production of pauses in child-directed speech

Several studies have provided evidence that the prosodic correlates to phrase and clause boundaries are used even more reliably in speech to children than in speech among adults. In a cross-linguistic study including six languages, Fernald and colleagues (Fernald et al., 1989) investigated the prosodic features of CDS. They found that mothers as well as fathers spoke with a higher fundamental frequency and used shorter utterances and longer pauses between the utterances when speaking to their children than when speaking to an adult. An utterance was defined on an acoustic basis as “a section of speech bounded by pauses greater than 300 ms” (ibid., p. 485). Thus, an utterance in this study does not need to constitute a full-fledged syntactic unit, like a clause or a sentence, and also does not necessarily need to coincide with a prosodic phrase in Nespor and Vogel’s (1986) sense. Still, the data that Fernald et al. (1989) report is consistent with the data provided by studies using syntactic or prosodic units as their base of analysis.

The characteristic differences between the utterances in CDS and ADS were found in all of the analysed languages, which were French, German, Italian, Japanese, British and American English (all data are taken from the appendix of Fernald et al., 1989). The mean duration of pauses between utterances was consistently longer in CDS than in ADS in all languages. All pauses between utterances in this sample were long pauses with a duration of more than 1 s in all languages, thus in ADS and in CDS the utterances were demarcated with pauses associated with the sentence level in ADS. The mean difference between pause durations in ADS and CDS was about 200 ms in mothers' speech. Interestingly, the elongation of the pause duration was more pronounced in fathers speaking to their children. The pauses between utterances were about 500 ms longer when fathers spoke to their child than when speaking to another adult. Thus, parents provided input that consisted of shorter utterances which were demarcated by longer pauses when speaking to their children. Furthermore, the modifications used in CDS were more extreme in American English than in any of the other languages.

As Fernald et al. (1989) did not specify the internal structure of the analysed utterances their data provide no information about the correlation between the type of linguistic boundary and the prosodic modifications. How the different linguistic structures are marked in CDS was investigated in a study by Fisher and Tokura (1996a; for an overview see also Fisher & Tokura, 1996b). They analysed the speech of three American English-speaking and three Japanese-speaking mothers to their respective child and measured the prosodic changes of syllables within the utterances. An analysis of the types of utterances in the English corpus revealed that 52% of the utterances were full clauses, like declaratives, questions and imperatives. 18% of the utterances were well-formed phrase fragments, like verb phrases, noun phrases but also single nouns or adverbs, which then constitute an IPh. Furthermore, 29% of the utterances were exclamations or utterances of the child's name (Fisher & Tokura, 1996a, p. 3198). However, these type utterances also constitute single IPhs. Only 0.9% of utterances in the corpus were incomplete sentences or false starts. In the Japanese corpus the distribution of utterance types was more balanced: 30% full clauses, 37% fragments and 33% exclamations (ibid., p. 3206). There were no false starts in the Japanese corpus. Across all speech samples the largest changes in pause duration, F0 range and vowel duration were found around syllables at the utterance-final position compared to syllables in a non-final position (ibid., p. 3201). For instance, the mean duration of the pause was 859 ms in English and 838 ms in Japanese. Furthermore, in English, 59% of the utterances were followed by a pause and 96% of all pauses occurred at the boundaries of such utterances (ibid., p. 3202). In Japanese, even 69% of the utterances were followed by a pause and 96% of all pauses occurred at the boundaries of such utterances. Furthermore, the change in F0, the amplitude and the duration of the syllable preceding a pause was significantly different from syllables occurring within an utterance. The corpus analyses of Fisher and Tokura (1996a) provide evidence that the utterances are reliably marked by converging acoustic cues in CDS (see also Morgan, 1986).

The analyses of phrases in English, in this case the first major phrase within a full clause like the Subject-NP of an utterance or the fronted wh-phrases as in *What a nice girl you are* (ibid., p. 3199), revealed no significant longer pause durations at the phrase boundary compared to a non-final position. Both at within sentence locations and phrase boundaries the duration of the pause was 37 ms vs. 26 ms for English and 32 ms vs. 38 ms for Japanese, respectively. Thus, in English the mean pause duration at a phrase boundary was even shorter than the mean duration of a pause occurring elsewhere in the sentence. Fisher and Tokura (1996a) assumed that this was the case, because most of the Subject-NPs consisted of pronouns, which do not constitute a separate IPh but are integrated in the following IPh of the VP (cf., section 2.2.3.2). Thus, no explicit boundary marking was expected at such positions. In utterances containing a full Subject-NP a significant final lengthening was observed.

Fisher and Tokura (1996a) concluded that the observed acoustic differences, including pauses, between syllables at an utterance boundary and syllables within the utterances were large enough to be recognised by the infant, despite the variability of the acoustic features within an utterance. Especially the distribution of pauses within the CDS samples, that is, 96% of pauses occurred at utterance boundaries and 59% to 69% of utterances were followed by a long pause, might provide infants with reliable information about grammatical units in their native language (for a different view see Fernald & McRoberts, 1996).

In a recent study, Soderstrom and colleagues (Soderstrom, Blossom, Foygel & Morgan, 2008) investigated the maternal speech input of two infants between the age of six and ten months.[8] They analysed not only the boundaries of single clause utterances, they also examined the prosodic markings at clause boundaries within multiclausal utterances. Soderstrom et al. (2008, Table 7/8 and Appendix 2) found that both the

[8] There are also data from supplementary recordings at 12 and 18 months, the results of which will not be discussed here (cf., Soderstrom et al., 2008).

boundaries of single clauses and of within-utterance clauses were reliably marked in the input, including significant pause durations.[9]

The data presented on the production of pauses in ADS and CDS support the assumption that pauses are used systematically to demarcate syntactic and / or prosodic units in the speech stream. Even though pauses also have other functions in the speech, the converging cues of pauses and other acoustic features marking prosodic boundaries seem to be a reliable indicator of speech units in the input. Furthermore, in CDS nearly all full utterances and phrase segments were found to be grammatical units (Fisher & Tokura, 1996a), so that infants might be able learn about the grammar of their language by identifying IPhs in their caretakers' speech.

2.1.3 Correlation between prosody and syntax

The findings of a rather systematic use of different pause durations to demarcate different speech units suggest the existence of a *natural pause hierarchy*, which is interrelated to the prosodic hierarchy by Nespor and Vogel (1986). The following figure shows the way this pause hierarchy complements the prosodic hierarchy and how both are related to the syntactic hierarchy within a language.

[9] However, overall the pause durations in their speech corpus were considerably shorter than the ones found by Fisher and Tokura (1996a).

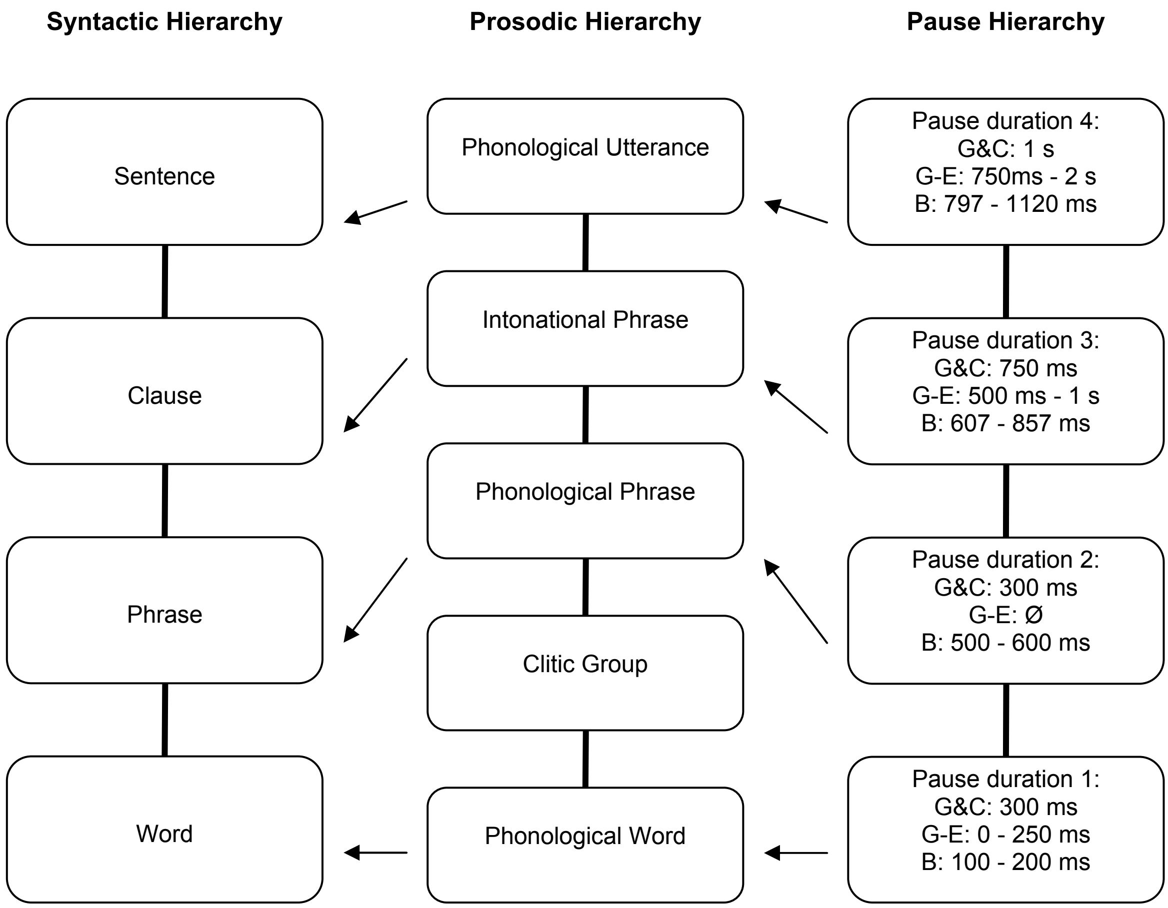

Figure 2.1: Outline of the correspondence between syntactic, prosodic and pause hierarchies.[10]

Figure 2.1 shows how the presented studies on the production of pauses suggest a hierarchical organisation of the different pause durations (right hand column). These pause durations are associated with the speech units organised in the prosodic hierarchy (middle column), in that pauses with a shorter duration are more likely to be inserted between phonological words or phrases (pause durations 1 & 2) and pauses with a longer duration tend to occur at intonational phrase boundaries (pause duration 3) or between phonological utterances (pause duration 4).

As the infant is mainly confronted with an ongoing speech stream, the sensitivity to the distribution of pauses adding to the organisation and marking of the prosodic units

[10] The syntactic hierarchy is simplified. The prosodic hierarchy follows Nespor & Vogel (1986). The mean pause durations given in the pause hierarchy graph are from: Grosjean and Collins (G&C) (1979, p. 110), Goldman-Eisler (G-E) (1972) for English, and Butcher (B) (1981) for German. Note: read the arrows as "corresponds to".

may then help her to learn more about the syntactic units of her language (left-hand column, cf., Morgan, 1986). So, if infants start out segmenting fluent speech with the working hypothesis that important syntactic units are prosodically marked at the boundaries and demarcated by pauses in the signal, they would be able to segment units like sentences and clauses correctly in the speech stream most of the time, although sometimes it would not result in the identification the correct units, for instance, in the cases where there is a mismatch between syntactic and prosodic units. It was suggested that in these cases the infant can rely on other strategies to infer the correct syntactic units of her language (cf., the study by Gerken, Jusczyk & Mandel, 1994, see also section 2.2.3.2).[11] That infants use their sensitivity to the signal side of the language to learn about the syntax was described in the *prosodic bootstrapping hypothesis* (cf., section 2.2.2).

Having identified the relevant domains in the speech stream, infants might be able to detect the recurrent prosodic pattern of strong and weak components within these units and thereby learn about the syntactic structure of their language. Thus, exploiting the prosodic cues the speech stream provides would help infants to identify syntactically relevant units in their language. Or as Gerken and colleagues (Gerken et al., 1994) phrase it: "… learners might use prosodic changes to infer the prosodic structure and in turn infer syntactic structure from prosodic structure." (ibid., p. 242).

One possible application of this type of inference from prosodic to syntactic structure was described by Nespor and colleagues (Nespor, Guasti & Christophe, 1996) in terms of a *rhythmic activation principle*. They suggested that the infant could set the head-direction parameter, the parameter that allocates the word order within a syntactic phrase, by detecting the relative prominence of constituents within the phonological / intonational phrase. The syntactic head-direction parameter can be set to the language-specific value, because the head of a phrase is always the weak (unstressed) element and the complement always the strong (stressed) element within a phrase (cf., also Guasti, Nespor, Christophe & Van Ooyen, 2001). Furthermore, Christophe and colleagues (Christophe, Nespor, Guasti & Van Ooyen, 2003) provided evidence that infants as young as 6 to 12 weeks are sensitive to the prosodic correlates of the head-direction parameter, that is, to the

[11] The idea that infants might start out with some working hypothesis that yields correct segmentation in many cases of her language also underlies the hypothesis of a metrical segmentation strategy for word segmentation proposed by Cutler (1994, see also section 2.2.2).

weak-strong distribution of constituents, in different languages like Turkish and French (see also Christophe, Guasti, Nespor, Dupoux & van Ooyen, 1997 for a comprehensive discussion of the phonological bootstrapping approach).

The present chapter presented evidence that the input provides cues to larger language units like phrases and clauses. These markers include changes in the fundamental frequency and duration in the vicinity of prosodic boundaries. Furthermore, the data on the production of pauses showed that there is a strong correlation between the duration of a pause and the type of boundary at which the pause occurred. This led to the postulation of a natural pause hierarchy which complements the prosodic hierarchy of the language as described by Nespor and Vogel (1986). It was suggested that the sensitivity to the prosodic entities of a language might help infants to learn, for instance, about the syntactical organisation of their language.

The next chapter will present studies that investigated infants' sensitivity to the prosodic cues provided by the input. Many of these studies were conducted within the framework of the *prosodic bootstrapping hypothesis*. The prosodic bootstrapping mechanism is assumed to allow infants, by exploiting their sensitivity to the prosodic markers of the speech signal, to learn about the syntactic organisation of their language (cf. e.g., Jusczyk, 1997).

2.2 *Infants' sensitivity to the prosodic cues provided by the input*

One of the prosodic cues that are part of the language-specific properties is the rhythm of a language. Rhythm is a suprasegmental feature and can be defined by language-specific timing units.[12] In stress-timed languages the interval between two stressed syllables is said to be of equal duration (cf., Lehiste, 1977), and is the reason for the perceived staccato rhythm of languages like German and English. One such interval includes one stressed syllable plus one or more unstressed syllables. Syllable-timed languages, like French, are characterised by having syllables of equal duration, irrespective of whether the syllables are stressed or unstressed. And mora-timed languages, like Japanese, rely on differences in the subsyllabic structure, that is, the mora. For example, a syllable consisting of a consonant and a short vowel is monomoraic, whereas a syllable consisting of a consonant and a long vowel is bimoraic (Mehler, Dupoux, Nazzi & Dehaene-Lambertz, 1996, footnote 1).

Lehiste (1977) was able to show that, although isochrony often could not be found in the measurements of speech samples, the adult listener tends to perceive the duration intervals between stressed syllables in English as isochronous. And an increase in the duration of an interstress interval, that is, a disruption in the perceived isochrony, was interpreted as a phrase boundary in a syntactic disambiguation task (cf., Lehiste, Olive & Streeter, 1976).

Thus, the sensitivity to the speech rhythm and especially to changes in that rhythm might provide infants with information about the speech units of their native language, because the boundaries of phonological and intonational phrases are marked, for instance, by changes in duration, that is, final lengthening (Cooper & Paccia-Cooper, 1980).

In the next section studies will be presented that addressed the question of infants' sensitivity to the rhythmic properties of languages (for an overview see also Guasti, 2004).

[12] For a different account for the classification of languages to each of these rhythm classes see, for instance, Ramus and colleagues (Ramus, Nespor & Mehler, 1999). They suggest that rhythm classes are defined by phonological differences in the syllable structure of the languages. So-called stress-timed languages, allow complex consonant cluster and vowel reduction, syllable-timed languages have less complex consonant clusters and allow most often no vowel reduction, and finally mora-timed languages do not allow consonant clusters.

2.2.1 Sensitivity to rhythmic properties of speech

Preference for the native language

From early on, infants are sensitive to and even prefer to listen to their native language. Using the High-Amplitude Sucking (HAS) paradigm, a habituation paradigm to test (speech) perception and discrimination abilities in newborns and young infants, Mehler and colleagues (Mehler et al., 1988) tested 4-day-old French infants on their ability to discriminate their native language (syllable-timed) from a language with different rhythmical properties, namely Russian (stress-timed). In a series of experiments they provided evidence that the infants were able to discriminate both languages, that is, the infants showed a dishabituation reaction in correlation to the change of language, even when the stimuli were low-pass filtered. In low-pass filtered speech, only prosodic information like rhythm and intonation is preserved, while most of the phonetic information is removed.

Furthermore, Mehler et al. (1988) found that the dishabituation reaction was stronger when the stimulus changed from Russian to French. Therefore, they concluded that the infants not only discriminated both languages but also preferred their native language over the non-native one. This preference for the native language, though it could not be replicated by the research group around Mehler themselves (cf., Mehler et al., 1996), was also found in a study by Moon and colleagues (Moon, Cooper & Fifer, 1993) in which they tested the preferences of Spanish and English 2-day-olds using a different methodology designed to test specific preferences. In their study Spanish and English newborns sucking behaviour was contingent upon hearing stimuli from either the native or the non-native language. Moon et al. (1993) found that the infants preferred to listen to their native language over the foreign language.

These studies were supplemented by findings from Bosch and Sebastián-Gallés (1997), who tested 4-month-old infants on their discrimination of two languages coming from the same rhythmic class, namely Spanish and Catalan, both syllable-timed languages (cf., Ramus, Dupoux & Mehler, 2003). They found that the infants, who were raised either in monolingual Spanish or in monolingual Catalan families, were able to distinguish between these two languages, again even when the stimuli were presented low-pass filtered. This result shows that 4-month-old infants are able to discriminate between their native and a second language, even when both languages are from the same rhythmical class. In a subsequent study, Bosch and Sebastián-Gallés (2001) also provided evidence that 4-month-old Spanish-Catalan bilinguals were able to discriminate between these two languages.

Thus, the studies on the early language perception of infants provide evidence of a sensitivity to and preference for the native language from birth on. Especially, the data from Bosch and Sebastián-Gallés (1997, 2001) seem to suggest that the ability to discriminate two languages found by Mehler et al. (1988) and Moon et al. (1993) was not initiated by the differences in the rhythmical class of these languages as such, but rather by the familiarity with the native language.

Discriminating two non-native languages

However, infants were also found to be able to discriminate between two non-native languages. In their first study, Mehler et al. (1988) also tested the ability of infants to discriminate between two non-native languages belonging to different rhythmical classes, for the French group between English (stress-timed) vs. Italian (syllable-timed) and for the American group French (syllable-timed) vs. Russian (stress-timed). Both groups did not seem to be able to discriminate between two non-native languages. This finding was later subject to reanalysis (Mehler & Christophe, 1995; Mehler et al., 1996). Looking at a dishabituation reaction across test conditions, that is, irrespective of which language was presented during the habituation phase, they found that the French 4-day-olds, but not the American 2-month-olds, showed a dishabituation reaction after the language change, indicating that they were able to discriminate languages from two different rhythmical classes. They assumed that a developmental change in infants' sensitivities was responsible for that pattern, in such a way that older infants already have more specific knowledge about their native language and treat both unknown languages as not fitting that model, and therefore fail to show discrimination between these languages (Mehler et al., 1996, p. 105f.)

Nazzi and colleagues (Nazzi, Bertonici & Mehler, 1998) also tested newborns on their abilities to discriminate between two languages belonging to the same rhythmical class, like English vs. Dutch. The French newborns tested with the HAS paradigm did not seem to be able to discriminate between English and Dutch, whereas they succeeded in discriminating English from Japanese (mora-timed). Furthermore, the French newborns were able to discriminate languages based on the rhythmic class, that is, they discriminated a test set consisting of English and Dutch utterances from another test set consisting of Spanish and Italian utterances. As all the stimuli presented in the study were low-pass filtered, Nazzi et al. (1998) concluded that newborns are sensitive to the differences in the rhythmic classes of languages (R-Hypothesis; ibid., p. 757), which then might help them to learn about the organisation of the rhythmic class their language belongs to.

A second study by Nazzi and colleagues (Nazzi, Jusczyk & Johnson, 2000a) tested the ability of older infants to discriminate languages across or within rhythmic classes. Using the HTP paradigm (for a detailed description of the paradigm see chapter 3), they found that 5-month-old American infants still were able to discriminate between languages from two different rhythmic classes, even when both were non-native (Italian vs. Japanese).[13] Furthermore, the 5-month-olds were not able to discriminate between languages that belong to the same rhythmic class when both languages were non-native (Italian vs. Spanish; Dutch vs. German). Only in the condition in which one of the languages from within the same class was the native language (British English vs. Dutch; and even American English vs. British English) discrimination could be seen. Nazzi et al. (2000a) concluded from the results of the series of experiments that infants did have a sensitivity to the rhythmic class their native language belongs to as opposed to other rhythmic classes. Furthermore, the familiarity with the native language seemed to be an important factor for the discrimination of languages belonging to the same rhythmic class, because the infants were able to discriminate between English and Dutch but not between Dutch and German (native-language acquisition hypothesis; ibid., p. 3f). However, Nazzi et al. (2000a) used natural speech and not low-pass filtered speech.

[13] This result partly contradicts the findings by Mehler et al. (1996), who found no discrimination of two non-native languages by 2-month-old infants. Nazzi et al. (2000a) assume that these different findings are due to the different experimental paradigms used in the two studies.

Therefore, the languages presented differed in more dimensions than just the differences or similarities in rhythm.

In a recent study, Ramus (2002) tried to investigate more closely the influence of rhythm and intonation on infants' discrimination of different languages. He tested newborns on their discrimination of Dutch and Japanese sentence stimuli, in a condition where the segmental information was manipulated[14] and also an identical intonation contour was superimposed on all sentences, leaving alone differences in the rhythm between the two "languages". In this experiment the newborns showed tentative discrimination between the two languages. As the effects found in this experiment were weaker than the effects found in studies using low-pass filtered speech, in which both rhythm and intonation were preserved, Ramus (2002) suggested that this might indicate that the integration of rhythmic and intonational cues might enhance infants' language discrimination abilities. The conclusion that not only rhythm but also intonation might play a role in infants' language discrimination could also be drawn from results of a study from Dehaene-Lambertz and Houston (1998). They found discrimination between English and French in 2-month-old English- or French-learning infants even when presenting rather short utterances (1.2 s), but not when the intonational organisation of the utterances was eliminated by scrambling the words within the utterances.

Taken together, the studies on the perception of rhythmic properties in speech provide evidence for an initial preference for the native language and also for infants' abilities to discriminate two non-native languages from different rhythmic classes. At the age of 4- to 5-months, the familiarity with prosodic characteristics of the native language seems to have evolved rather precise, because in addition to the ability to discriminate the native language from a different language from the same rhythmic class (Bosch & Sebastián-Gallés, 1997), infants were also able to discriminate between the American and British dialect of their native language (Nazzi et al., 2000a).

This sensitivity to the rhythmic, and also the intonational characteristics of the native language is a prerequisite to notice changes in the rhythmic and intonational pattern

[14] All consonants were replaced by a single consonant according to the manner of articulation. All vowels were replaced by /a/ (Ramus, 2002, p. 96).

that may indicate prosodic boundaries in the speech. As described in chapter 2.1 of the present study the boundaries of phrases and clauses are marked by changes in duration, pitch, and the occurrence of pauses. Furthermore, the sensitivity to the rhythmic units in a language, for instance, interstress intervals in stress-timed languages, may allow more sophisticated analysis of the language, as has been proposed by the *bootstrapping hypotheses*. These hypotheses claim that infants are able to use knowledge of one level of the language to infer about other levels of the language. In relation to the topic of the present study, an early sensitivity to the rhythmic properties of the native language might allow infants to establish the relations of larger prosodic units and pauses which then, once established, could be used to infer about the syntactic structure of the language (see Figure 2.1).

2.2.2 The Prosodic Bootstrapping Hypothesis

The described sensitivity to the rhythmic properties of the native language might also help infants to acquire other structures of the native language, like words, phrases and clauses (cf., Mehler et al., 1996). Gleitman and Wanner (1982) were the first to use the image of *bootstraps* that the infant might use to help her entering in the language acquisition process. The rationale behind the idea of *prosodic bootstrapping*[15] is that different units are marked in the speech stream (cf., chapter 2.1). For instance, as Nespor and Vogel (1986) have suggested at each level of the prosodic hierarchy only one element is defined as strong and the other elements are defined as being weak. Thus, each level in the hierarchy has its “own rhythm”: for words, the alternation of stressed and unstressed syllables, for phrases, the alternation of stressed and unstressed words and for utterances, the intonation assigning one main accent. Together with the sensitivity to the prosodic boundary markers, the sensitivity to rhythmic units in the language therefore might help infants to identify the relevant units on each level of the prosodic hierarchy.

---

[15] The *prosodic bootstrapping* is also sometimes called *phonological bootstrapping* (cf., Christophe et al., 1997). Morgan and Demuth (1996) claim that the term *prosodic bootstrapping* “is overly confining and, hence, misleading” (ibid, p. 2), because infants do not only make use of the prosody, but also use information on the phonetic, phonotactic and stochastic level and therefore, they propose the “general descriptor […] *phonological bootstrapping*” (ibid, p. 2). In this paper, however, the term *prosodic bootstrapping* will be used to cover both variants (cf., Jusczyk, 1997, p. 38).

Gleitman and Wanner (1982) proposed that the stressed syllable functions as the relevant unit in the identification of words. The idea is that the prosodic analysis of the input into stressed and unstressed syllables helps the infant in most cases to encode the domain in which morphological rules apply, namely the word. In a stress-timed language like English, about 90% of the bisyllabic content words start with a stressed syllable (Cutler & Carter, 1987), so that applying a segmentation based on the rhythm, a *metrical segmentation strategy* (Cutler, 1994, 1996), would yield correct segmentation results in most of the cases.[16] A corpus analysis of German, also a stress-timed language, provided a comparable picture: 85% of the bisyllabic content words also have stress on the first syllable (Höhle, 2002, p. 91). Thus, a segmentation strategy on one level, namely the prosodic level, may help the child to discover entities at another level, in this case the word on which morphological operations take place. Furthermore, Cutler (1994) convincingly showed that processing strategies of adults were influenced by their native and in the case of bilingual speakers, dominant language. Thus, English speakers used a stress-based word segmentation procedure, French speakers used a syllable-based segmentation procedure, and Japanese speakers used a mora-based segmentation procedure. Highly proficient bilingual speakers processed the speech according to their dominant language. Cutler (1994) claimed that the rhythmic segmentation procedures of adults are “simply traces which remain from a period when the segmentation problem dominated the infant’s language processing” (ibid., p. 96).

With respect to larger speech units like phrases and clauses, Gleitman and Wanner (1982) cited evidence that 7-year-olds rely on intonation properties rather than syntax in their analysis of complex utterances. Thus, they suggested that “the input signals are interpreted as ordered phonetic strings bracketed by stress into words and bracketed by intonation into phrases” (ibid., p. 26) and that an “infant who is innately biased to treat intonationally circumscribed utterance segments as potential syntactic constituents would be at considerable advantage in learning the syntactic rules of his language” (ibid., p. 26). That is, by paying attention to the prosodic characteristics of the input string, the child can learn much about different organisation units in her language.

[16] This identification of stressed syllables as relevant units in the sentence could also explain the frequently found pattern that unstressed function words, such as determiners, preposition, etc., are often not produced in the early utterances of children (cf., Gerken, 1991).

The reliance on prosody rather than syntax in speech processing was also demonstrated in adults. Dooling (1974) tested adults on their perception of sentences in white noise, which is a masking noise with constant amplitude over all frequencies. After being presented with a number of training sentences, the participants heard a test sentence which was identical to the training sentences or which either varied in rhythm or both in rhythm and syntax from the training sentences. All sentences were presented in monotone, so that there was no difference in intonation across conditions. The only differences occurred in the rhythm, which was realised by changing the amount of syllables of the words, and in the syntax of the sentences (ibid., p. 256f). The adults' task was to identify a keyword in each sentence. The data analyses showed that in the condition in which the syntax of the sentence remained unchanged and only the rhythm of the target sentence was changed the participants were less likely to identify the keyword than in the condition in which no change was present. The same pattern of results was found for the condition in which both rhythm and syntax changed as compared to the no-change condition. Dooling (1974) concluded that not the syntax but the rhythm was the relevant dimension which the participants used to process the sentences. Thus, adults can also rely on bottom-up mechanisms to process speech, and they have been shown to do so, for instance, in the processing of syntactic ambiguities (cf., Katz et al., 1996; Trueswell, Sekerina, Hill & Logrip, 1999).

Critical analysis of the prosodic bootstrapping hypothesis

In a critical analysis of the arguments in favour of a prosodic bootstrapping account for the acquisition of syntactic clauses, Fernald and McRoberts (1996) argued that the data in the input are not sufficient and not reliable enough, that is, the cue reliability is too low, to allow the proposed prosodic bootstrapping of syntactic units. Discussing each of the prosodic boundary markers of clauses in ADS, namely occurrence of pauses, changes in the F0, and final lengthening, Fernald and McRoberts (1996) stated that none of the cues is reliable enough to allow adults or infants to infer the existence of a syntactic clause boundary.

They argued that pauses also frequently occur at locations not related to syntactic clause boundaries and that syntactic clause boundaries, on the other hand, are not consistently marked by pauses; that changes in the F0-contour are not only related to clause boundaries, but are also influenced by the discourse structure, for instance the focus structure of a sentence, and also by the affective state of the speaker; and finally, that the lengthening of the pre-boundary element is merely a relative feature, which also can occur at other locations, depending on the speaker's mood or speech rate.[17] Therefore, they concluded that in "spontaneous ADS, prosodic features are only moderately reliable as cues to syntax" (ibid., p. 375).

Fernald and McRoberts (1996) further analysed CDS and also found that the boundary markers did not seem to be reliable enough to allow the bootstrapping of syntactic clauses, as most of the utterances in CDS did not consist of whole clauses but more often of sub-clausal elements like vocatives and clause fragments. Even the existing clauses were not simple SVO-structures but often non-canonical sentences like questions. Thus, even though the utterances in CDS are more reliably marked, an infant relying, for instance, on the pauses between utterances as a bootstrapping device would hardly be able to bootstrap a syntactic clause and more often come to wrong hypotheses about the structure of their language. The authors concluded that the cue reliability of the clause boundary markers is more probabilistic than actual and that therefore bootstrapping of clauses hardly seems possible or at least, that the "potential role [of prosody] has been oversimplified in the prosodic bootstrapping debate […]" (ibid., p. 384).

Even though Fernald and McRoberts (1996) argued that no single cue is sufficient to reliably mark phrase and clause boundaries in fluent speech, they themselves could not rule out that "prosodic cues in combination are probably more powerful than in isolation" (ibid., p. 373). And it is usually the combination of cues that marks a boundary.

A recent study has provided evidence that on the word level a combination of cues, even though each single one of them has a low reliability, allows for categorising

[17] Another factor that influences the duration of words is that in some languages like German, Dutch and English vowel duration is phonemic. For instance, in German long and short variants of a vowel are different phonemes. Thus, there are minimal pairs like like Miete [miːtə] 'rent' and Mitte [mItə] 'centre'. However, it is to be expected that both words undergo lengthening in the vicinity of a prosodic boundary, as lengthening is a relative feature in the speech stream (for a different view see Morgan, 1986, p. 112).

the input. Shi and colleagues (Shi, Morgan & Allopenna, 1998) analysed correlates of lexical vs. functional words in Mandarin Chinese and Turkish CDS. These correlates included, among others, word position in an utterance, number of syllables per word, syllable duration and pitch changes. They found that neither cue was reliable enough on its own to allow the attribution of a word to its word class. But when a set of cues was fed to a self-organised neural network, a network that does not need external teaching or feedback, thus resembling an infant learner, it led to the correct attribution of a word to the appropriate word class in about 90% of all cases. The authors hypothesised that these "results raise the possibility that infants may be able to sort segmented words or morphemes into appropriate superordinate grammatical categories before they have learned the meanings of any of these words [...]" (ibid., p. 195). Thus, by combining cues that are probabilistic, the infants may also achieve to bootstrap into their language.

Fisher and Tokura (1996a, 1996b) have shown that phrases and clauses, especially in CDS, are reliably marked in the input (cf., section 2.1.2.3). In contrast to the findings of Fernald and McRoberts (1996), they provided evidence that most of the utterances, though not always consisting of full-fledged clauses, were still grammatical utterances. Furthermore, the reliability of pause occurrence in relation to prosodic boundaries was high: 96% of pauses occurred at utterance boundaries and 59% to 69% of utterances were followed by a long pause. In combination with the changes in the fundamental frequency and the longer duration of the final syllable, these converging cues might provide infants with reliable information about grammatical units and ultimately about the grammar of their native language.

Fernald and McRoberts' (1996) criticism that the prosodic bootstrapping account does not seem to be applicable for data that has a probabilistic distribution in the input, like pauses occurring at the clause boundary but also at locations in which no clause boundary is marked, seems to be qualified by the results of these studies.

The prosodic bootstrapping into language indeed seems to be the first strategy that infants use in language acquisition (cf., Hirsh-Pasek & Golinkoff, 1996; Höhle, 2002). It is supposed to underlie the early perception of words, phrases and clauses. However, additional bootstrapping mechanisms like *distributional*, *semantic* and *syntactic bootstrapping* were suggested to support the language acquisition process (cf., Pinker, 1984, 1987).

The *distributional bootstrapping* hypothesis was based on the findings of Maratsos and Chalkley (1980) that types of words that tend to co-occur frequently with other types of words are prone to belong to the same category, for instance, a determiner frequently occurring before a noun. Infants have been shown to be sensitive to these distributions and to use them for language acquisition (cf. e.g., Höhle, Weissenborn, Kiefer, Schulz & Schmitz, 2004).

The *semantic bootstrapping* mechanism was proposed by Pinker (1984, 1987). He suggested that a child can learn the meaning of words from the input and can construct basic semantic representations of these words from the situational context and the meaning. These semantic entities include, for instance, "thing" or "causal agent" and the child infers from the highly structured CDS input that these entities map onto (innate) syntactic universals such as "noun", "subject", etc. (Pinker, 1987, p. 407). Furthermore, Pinker (1987) claims that these mappings are the input to the phrase structure rules of grammar. The level of abstraction is enhanced by learning more and more meanings of words and realising what each category has in common and vice versa, once a syntactic structure has been built, the next sentence that is encountered can be analysed according to that structure.

Gleitman (1990) proposed a *syntactic bootstrapping* account for the learning of verbs. Her main motivation was the question "How does the learner decide which particular phonological object corresponds to which particular verb concept?" because "[…] it is not clear at all that the required pairings are available to learners from their ambient experience of words and the world." (ibid., p. 4). Drawing from her research of language acquisition in blind children, who cannot observe the world in the sense Pinker (1984, 1987) intended, she proposes that the verb-argument structure allows the child to deduce the meaning of words. That is, by exploiting the innate syntactic knowledge, the child bootstraps into her language.

The *prosodic bootstrapping* hypothesis is compatible both with *inside-out*[18] (e.g., Chomsky, 1986; Pinker, 1994) and *outside-in* (e.g., Bates & MacWhinney, 1989) theories of language acquisition. The acoustic and prosodic analysis of the language is a prerequisite for the identification of relevant units in the language. However, the "issue of whether what is required for detecting cues to syntactic organization in the signal are innate linguistic capacities or more general perceptual capacities used in a language context is not settled […]" (Jusczyk, 1997, p. 38). Although Jusczyk made this remark about 10 years ago, the issue is still a matter of debate.

The hypothesis that infants use different bootstrapping mechanisms to acquire the native language was the starting point for an increasing number of studies on the segmentation and acquisition of language in infants. Some of these studies on the various levels of speech will be presented in the next sections, with special emphasis on the prosodic bootstrapping into language.

2.2.3 The perception of…

2.2.3.1 Words

When children start to utter their first words, an enormous segmentation work lies behind them: they had to find out what exactly in the speech stream constitutes a word. It was suggested that the supposed sensitivity to prosodic boundaries would help infants to segment words from fluent speech, as prosodic boundaries always coincide with word boundaries. Recent research has provided evidence that 8-month-old infants more readily segment words that occur at boundary positions as opposed to words that occur in utterance-medial positions. Seidl and Johnson (2006) found that American English-learning infants recognised monosyllabic words and non-words, like *cash* and *geff*, that had been previously presented in utterance-initial or utterance-final position easier than monosyllabic words that had been presented in utterance-medial position.

[18] This classification follows Hirsh-Pasek and Golinkoff (1996, p. 17ff). The rationale behind the use of this terminology is that both inside-out theories, like nativism, and outside-in theories, like cognitivism, assume innate capacities which allow the acquisition of the language. The difference is that inside-out theories claim these innate capacities to be domain-specific, that is, specifically and uniquely for language acquisition, whereas outside-in theories believe them to be domain-general.

In a second study, Seidl and Johnson (2008) provided further evidence for their *Edge Hypothesis*. In a series of experiments testing 11-month-olds on the recognition of monosyllabic vowel-initial words and non-words, like *ash* and *eeb*, they found differences in the facilitation effect of the positions: the segmentation and recognition of words presented in utterance-final position seemed to be easiest, followed by words presented in utterance-initial position and words that were presented in utterance-medial position seemed to be more difficult to segment. Thus, the occurrence of a word boundary that coincided with the prosodically marked right boundary of an utterance seemed to facilitate the segmentation of whole words in fluent speech.

The facilitation effect for segmenting words from the utterance-final position was also found for bisyllabic words in German-learning 9-month-olds. Schmitz et al. (submitted) found that infants more easily segmented and recognised bisyllabic words, like *Balken* ‘beam’ and *Pinsel* ‘paint-brush’, that had been presented in utterance-final position as opposed to words that had been presented in utterance-medial position. Thus, the sensitivity to prosodic boundaries not only helped infants to segment monosyllabic words which are identical with the prosodically marked final syllable of an utterance, but also allowed the segmentation of bisyllabic words from the utterance-boundary. That the sensitivity to a phrase boundary helped infants in the recognition of words was also found in the study by Gout and colleagues (Gout, Christophe & Morgan, 2004) which will be discussed in detail in the next section of the present paper (2.2.3.2).

The segmentation of monosyllabic words from fluent speech, in which the position of the target word was variable, was also found in earlier studies. Jusczyk and Aslin (1995) provided evidence that 7.5-month-old, but not 6-month-old American English-learning infants were able to recognise monosyllabic content words, like *cup* and *feet*, in fluent speech. Infants of this age were not only able to recognise stressed content words from fluent speech, but also unstressed closed-class elements, like the determiner *das* ‘the’ in German 7.5-month-olds (Höhle & Weissenborn, 2003). Shi and colleagues (Shi, Werker & Cutler, 2003) found that 13-month-olds, but not 8-month-olds, recognised real function words, like *its*, as opposed to nonsense function words, like *ots*, in an NP[19] (see also Shi, Cutler, Werker & Cruickshank, 2006). Furthermore, Shi and colleagues (Shi, Marquis &

[19] That 14-month-old infants used high frequent function words, like determiners, to categorise the subsequent word as a noun could be shown in a study by Höhle and colleagues (Höhle et al., 2004).

Gauthier, 2006) found that French-learning infants at the age of 6 months also recognised and segmented a determiner, like *la* 'the'$_{\text{(fem.)}}$ within an NP and are able to discriminate between the determiner and a nonsense function word like *ta*, at the age of 8 months. In all these studies, it was argued that the high frequency of occurrence of function words in the input may have allowed infants to recognise these words, despite the fact that they are prosodically less salient than the stressed content words. In addition, function words like determiners usually occur at the left boundary of an utterance in languages like German, English or French and therefore might also be easier to segment for the infants.

The sensitivity to the rhythm of their native language also seems to influence infants' ability to segment bisyllabic words from fluent speech. Jusczyk and colleagues (Jusczyk, Houston & Newsome, 1999) found that 7.5-month-old American English-learning infants were able to recognise trochaic words, like *kingdom* or *doctor*, from fluent speech, but not iambic words, like *guitar* or *surprise*. That is, the infants seemed to rely on the metrical segmentation strategy appropriate for stress-timed languages (cf., Cutler, 1994, 1996). The iambs were missegmented, that is, the infants seemed to perceive the strong syllable of the iamb as a word onset and integrate the next following syllable to form another trochee, that is, infants seemed to perceive a word combination like *'guitar is'* as *'taris'* (Jusczyk et al., 1999, Experiment 11). The experiments also suggested that the infants represented the whole trochee as a coherent unit rather than only the strong initial syllable, that is, *kingdom* rather than *king*. From 10.5 months on, the infants were also able to recognise iambic words in fluent speech and also represented these iambic words as coherent units. This was interpreted in a way that by the end of the first year the infants no longer rely only on the rhythmic segmentation strategy to acquire the language, but also use other information to integrate knowledge, like allophonic cues.

Recent studies on German infants' discrimination of trochaic vs. iambic lists of bisyllabic sequences, like '*gaba*[20] vs. *ga*'*ba*, provided evidence that 6-months-olds, but not 4-month-olds, preferred to listen to the trochaic over the iambic sequences (Höhle, Bijeljac-Babic, Herold, Weissenborn & Nazzi, submitted). However, Friederici and colleagues (Friederici, Friedrich & Christophe, 2007), who used a non-behavioural experimental technique, namely the Event-Related Potentials (ERP) paradigm, provided

[20] The symbol ' ' ' indicates that the following syllable is stressed in the examples.

evidence that infants as young as 4 months are able to discriminate trochaic and iambic stimuli. This was reflected in a mismatch response (MMR) of the brain, in which a positive wave in the EEG signal was found for the stimulus type that is less frequent in their language, that is, to the iambic stress pattern in German infants. Thus, it seems that infants are able to discriminate trochaic from iambic stimuli already by 4 to 6 months of age. But solving the more demanding task of recognising trochees in fluent speech can first be observed at about 7.5 months of age. Only by the age of 10.5 months, the infants seem to integrate other speech cues to allow them to discriminate iambic words from fluent speech.

But how do infants learn about the differences between a multi-syllabic word and a sequence of words, which is a prerequisite for identifying phrases and grammatical relations within phrases in the speech signal? Johnson and colleagues (Johnson, van de Weijer & Jusczyk, 2001) tested whether 7.5-month-old infants perceive a 3-word phrase as a coherent unit, that is, as a single word or not. Based on the finding by Santelmann and colleagues (Santelmann, Houston & Jusczyk, 1997) that 7.5-month-old infants were able to identify a three-syllabic word with a strong-weak-strong stress pattern like *parachute* as a coherent unit, Johnson et al. (2001) familiarised infants with sentence passages which contained a 3-word-phrase like *pair of mugs*, which also has a strong-weak-strong stress pattern. During the test phase, however, the 7.5-month-olds did not listen longer to the word lists containing the familiarised combination of words than to the lists containing previously unfamiliarised combinations of words. Johnson et al. (2001) concluded that the infants did not perceive and represent the 3-word-phrase as a coherent unit, that is, not as a single trisyllabic word, and therefore did not recognise the phrase in the test phase. They suggest that infants integrate speech cues like coarticulation cues as a means to perceive word boundaries and distinguish between multisyllabic words and multi-word phrases.

To summarise, infants are able to use a range of cues in their perception of words in fluent speech, including prosodic cues which occur at the boundaries of utterances. The studies cited above show that already by 7.5 months infants are able to discriminate monosyllabic, bisyllabic and even trisyllabic words from fluent speech and that they use their knowledge to further segment the speech stream and, during the second year of life, start to categorise words from the input. Furthermore, at the age of 7.5 months, the infants are able to perceive a 3-word-phrase as a non-coherent unit.

The next two sections will present studies on infants' perception of phrases and clauses in fluent speech. Many of these studies used the insertion of artificial pauses at boundaries or within utterances to investigate infants' recognition of phrases and clauses. That is, these studies used the insertion of a pause as a gap inserted between phrases or within phrases, without addressing the question of infants' sensitivity to pauses and the pause hierarchy. Therefore, studies on the perception of pauses will be presented in a separate section (2.2.3.4) in the present paper.

2.2.3.2 Phrases

The research on infants' perception of phrases was concerned with the question whether infants are able to recognise phrases, and what cues might help infants to identify phrases in the fluent speech, like concord morphology. The authors of the different studies sometimes used the term "phrase" to describe syntactically defined units (e.g., Jusczyk, Hirsh-Pasek, Kemler Nelson, Kennedy, Woodward & Piwoz, 1992), others referred to prosodically defined units, like phonological phrases in Nespor and Vogel's (1986) sense (e.g., Soderstrom, Seidl, Kemler Nelson & Jusczyk, 2003). However, this is only a difference in terminology, because "whenever there is a phonological phrase boundary, there is also a syntactic phrase boundary" (Gout et al., 2004, p. 550).

In a comprehensive study Jusczyk and colleagues (Jusczyk et al., 1992) investigated the ability of 6- and 9-month-old infants to recognise phrase boundaries in fluent speech. In a series of HTP experiments, they presented infants with speech passages in which an artificial pause of 1 s duration was either inserted at a phrase boundary (i.e., before the main verb; coincident version), like in *Did you ___ spill your ceral?*[21], or at a location within a phrase (i.e., after the main verb; non-coincident version), like in *Did you spill ___ your ceral?*. Jusczyk et al. (1992) found that 9-month-olds, but not 6-month-olds, preferred to listen to the speech samples in which the pause occurred at the

[21] The use of '___' indicates the position of the artificially inserted pause in the examples.

phrase boundary. This finding was confirmed in a longitudinal study, in which a group of infants was tested both at 6 and at 9 months of age.

Furthermore, Jusczyk et al. (1992) tested whether the pattern of results also held, when the non-coincident version had the artificial pause located within the subject-NP, like in *Many different kind ___ of animals live in the zoo* or within the VP, like in *The little boy at the piano is having ___ a birthday party*. They also tested the infants on low-pass filtered speech samples. All of these further experiments confirmed the findings of the first experiment. The 9-month-olds again preferred to listen to the coincident speech samples.

This preference for the coincident speech samples could be seen both in CDS and ADS. Thus, by 9 months of age, infants' perception of phrase boundaries was not dependent on the CDS register. In an acoustic analysis of the stimuli, Jusczyk et al. (1992) found that the phrase boundary between the NP and the VP was marked by a fall in the fundamental frequency and a longer duration of the syllable preceding the boundary, plus the additionally inserted pause. Thus, they concluded that the phrase boundary was reliably marked by converging cues in the signal, allowing infants to recognise phrases in speech. This series of experiments provided the first evidence that infants as young as 9 months are sensitive to the acoustic properties indicating a major phrase boundary (NP / VP), reflected in their longer listening times towards the coincident speech samples.

A subsequent study by Gerken and colleagues (Gerken et al., 1994) tested whether the preference for the coincident speech samples reported by Jusczyk et al. (1992) was driven by syntax or by prosody. As has been noted in section 2.1.1, there is not always a one-to-one mapping between the syntactic and the prosodic structure of a sentence or an utterance. Longer subject-NPs, or any stressed constituent, like a proper name or any focused constituent, form a separate phonological phrase. But weakly stressed constituents, like subject pronouns, tend to be prosodically incorporated in the following phonological phrase, that is, the pronoun forms a clitic group (see Figure 2.1) with the following verb. However, both lexical and pronoun subject-NPs have the same syntactic status. In their study, Gerken et al. (1994) presented 9-month-old infants in an HTP experiment with speech samples that either contained a lexical NP or a pronoun in subject position, as in *The caterpillar ate four strawberries* compared to *He ate four strawberries*.

Pauses, again with a 1 s duration, were inserted after the lexical NP or after the Pronoun, respectively, (coincident version) or after the main verb (non-coincident version). One group of infants was tested in the lexical-NP condition and the other group was tested in the pronoun-NP condition of the speech samples.

Gerken et al. (1994) found that infants preferred to listen to the coincident speech samples in the lexical-NP condition, thus replicating the findings of Jusczyk et al. (1992). However, in the pronoun-NP condition, the infants did not reveal listening preferences for the coincident over the non-coincident versions. The authors concluded that 9-month-old infants had access to the prosodic structure of the utterances but no immediate access to the syntactic structure of the utterances. They hypothesised that infants might ultimately learn that pronouns are of the syntactical category NP by cross-sentence comparison: utterances in which a pronoun was focused, and thereby received stress and formed its own phonological phrase might trigger the syntactic analyses of pronoun NPs. Thus again, by exploiting prosody infants might ultimately be able to acquire the syntax of their language.

That the phonological phrase boundary has a psychological reality already in infants and constrains the processing of strings of words was found in a recent study by Soderstrom and colleagues (Soderstrom et al., 2003). Using the HTP paradigm, they familiarised 6- and 9-month-old infants with two homophonous strings of words. One string consisted of a single NP, as in … *new watches for men are simple …*; the other string consisted of a noun plus a VP, thus including a subject-verb phrase boundary as in *…the old frightened gnu / watches for men and women ….*[22] During familiarisation, infants listened to both strings of words. During testing, infants were presented with two text passages that also included these familiarised strings, as well as a sentence consisting of a control string, unfamiliarised for that particular infant. A listening preference for the text passage that contained the string of words as a single NP would therefore indicate that infants, though being familiarised with both strings, recognise prosodically well-formed strings more readily in fluent speech than strings of words that contained a phrase boundary. The results of this study provided evidence that infants as young as 6 months,

[22] The slash ‘ / ’ indicates a phrase boundary in the examples. The underline ‘____’ marks the homophonous string.

as well as 9-month-olds, were able to discriminate between strings of words containing a phonological phrase boundary and single NPs.

This finding was extended in a further experiment which showed that 6-month-olds also prefer to listen to passages containing single well-formed verb phrases, as in *Inventive people design telephones in their homes*, than to passages containing strings of identical words containing a subject-verb phrase boundary, as in *The director of design / telephones her boss*. The acoustic analysis of the stimulus material revealed that this effect seemed to have been driven by differences like final lengthening and pitch change on the words before the prosodic boundaries compared to the same words within phrases. Pauses between the naturally produced phonological phrases were hardly found, thus playing no decisive role in the recognition of these phrases in this experiment. Soderstrom et al. (2003) concluded that infants as young as 6 months are sensitive to prosodic boundaries which coincide with syntactic boundaries. Furthermore, the results suggest that a phonological phrase boundary limited the processing of strings of words. Strings of words disrupted by a prosodic boundary did not seem to be recognised as a coherent unit, whereas strings of words that form a single phrase seemed to be identified by the infants as a coherent unit.

The psychological reality of a phonological phrase boundary in the processing of speech was further investigated in a study by Gout et al. (2004). In an HTP experiment, they familiarised 10- and 13-month-old infants on two trochaic words, like *paper* or *beacon*. In the test phase either the whole word was included in a sentence adjacent to a phonological phrase boundary, as in *The scandalous paper / sways him …*; or the first syllable of the trochee was the last syllable of one phonological phrase and the second syllable of the trochee was the first syllable of the subsequent phonological phrase, as in *The outstanding pay / persuades him…* (ibid., p. 552). Gout et al. (2004) hypothesised that if infants were aware of the phonological phrase boundary the recognition of the trochee in the second condition would be impeded. The acoustic analysis of the target words revealed the expected pattern of phrase-final vowel lengthening (*pay* vs. *paper*) and lengthening of the final syllable (*paper* vs. *per*). To ensure utmost comparability between the two conditions the naturally occurring pauses between the subsequent phonological phrases in the phrase condition, which had a mean duration of 133 ms, were eliminated. Thus, the infants were not able to distinguish between both conditions simply by the absence or presence of a pause. Gout et al. (2004) found that 13-month-olds, but not 10-month-olds, indeed preferred to listen to the speech passages in which the whole trochee was retained as opposed to the speech passages in which the two syllables straddled the phonological phrase boundary. They concluded from these results that between 10 and 13 months a development takes place in which infants are more and more able to integrate their sensitivities of existing phrase boundaries with the recognition of words in fluent speech.

In a further series of experiments, using the Conditioned Headturn (CHT) technique (for details see section 3.1.1), Gout et al. (2004) trained 10-month-olds and 13-month-olds to react to the occurrence of a target word within a stream of speech. Within each age group infants were divided into two test groups: one was trained on a bisyllabic word like *paper*, the other group was trained on a monosyllabic word being homophonous to the first syllable of the bisyllabic word, like *pay*. The texts presented during the test phase were the same as in the previous experiment. Thus, for the group trained on *paper*, the text containing the sequence *The scandalous paper / sways him …* retained the whole word followed by a phonological phrase boundary, whereas in the text containing the sequence *The outstanding pay / persuades him …* a phrase boundary occurred between the syllables of the trained word. It was expected that this group of infants would react more often to the occurrence of the whole trochee in the speech stream as opposed to the syllables straddling the boundary. The test conditions and result expectations for the group trained on the monosyllabic word were vice versa: in this case the infants were supposed to turn their head more often when they perceived the monosyllabic word at the phrase boundary than when the monosyllabic word occurred as the homophonous part of the trochee, that is, the initial syllable.

Using this experimental paradigm, Gout et al. (2004) found that also the 10-month-olds showed a recognition effect for the whole bisyllabic word in the speech stream compared to the trochee disrupted by the phrase boundary. However, in this experiment, as in the first experiments, the 13-month-olds showed a stronger effect than the 10-month-olds (80% vs. 48% correct reactions). For the group trained on the monosyllabic words the effect did not reach significance in the data of the 10-month-olds. The authors concluded that using a method which monitors immediate reaction to a single stimulus change may reveal infants' word detection earlier than the orientation-time based HTP technique. Thus, already at 10 months infants show an emerging ability to use prosodic cues of phonological phrase boundaries to recognise words in fluent speech. This sensitivity to phrase boundaries and the identification of words as belonging to a phrase is a prerequisite for attending to and acquiring grammatical regularities within phrases.

In a study with adults learning an artificial language, Morgan and colleagues (Morgan, Meier & Newport, 1987) systematically varied the cues indicating phrase structure, like prosody, occurrence of function words and concord morphology to evaluate the learning process. In the experiment on the influence of prosody on phrase structure learning, participants were tested on whether they were able to learn the artificial language under three test conditions. In one condition, no prosodic cues were present, that is, the words were presented in a monotone way. The second condition, prosodic cues were present in the input, but contradicting the syntactic phrasing. Only in the third condition, the prosody gave coherent information about the phrasing in the language. In all subsequent tests, the participants in the condition in which the prosody supported the information on syntax had better results than the other two groups. Morgan et al. (1987) concluded that the presence of prosody promotes the acquisition of syntax. The experiments in which the occurrence of function words and the occurrence of concord morphology were tested in an analogous way as in the prosody experiment yielded the same results. In each experiment, the group of learners that had information which attested the phrase structure of the artificial language had the best learning results. This suggests that not only prosody, but also the distribution of function words and concord morphology may help the learner to determine what constitutes a phrase in her language.

Pelzer and Höhle (2006b; see also Pelzer, 2008) investigated whether the occurrence of concord morphology within a syntactic phrase would also enhance infants' recognition of these phrases as opposed to phrases which are not marked by concord morphology. To address this question an experiment using the HTP paradigm with 10-month-old infants was conducted. During the familiarisation phase infants listened to two types of speech passages: in one passage each of the five subsequently presented sentences contained a phrase marked with concord morphology, like in *dies*<u>*en*</u> *faul*<u>*en*</u> *Henn*<u>*en*</u> 'these lazy hens' or *sein*<u>*en*</u> *groß*<u>*en*</u> *Katz*<u>*en*</u> 'his big cats'. In the other text passage the sentences contained phrases with non-concordant suffixes, as in *dieser faulen Henne* 'this lazy hen' or *seiner großen Katze* 'his big cat'. These (non-)concordant target phrases were embedded in otherwise identical carrier sentences. This construction of the experimental passages was possible, because in these phrases the dative case in the plural is marked by concordant morphemes, whereas the dative case in the singular is not. During familiarisation, a concordant *hen* passage was paired with a non-concordant *cat* passage and vice versa. During the test phase infants listened to four lists of isolated phrases: familiarised concordant (e.g., *diesen faulen Hennen* 'these lazy hens'), familiarised non-concordant (e.g., *seiner großen Katze* 'his big cat'), unfamiliarised concordant (e.g., *solchen dünnen Suppen* 'those watery soups') and unfamiliarised non-concordant (e.g., *ihrer jungen Meise* 'her young titmouse'; ibid., p. 429). Using this experimental design, it was possible to test the influences of familiarisation and concord morphology separately.

Pelzer and Höhle (2006b) hypothesised that if concord morphology helps identifying phrases in fluent speech, the infants would have longer listening times for the familiar concordant phrase than for the familiar non-concordant phrase. Furthermore, if familiarisation influences the recognition process, the infants were expected to have longer listening times for the familiar concordant phrase than to the unfamiliar concordant phrase. Both expectations were met in the data. These results suggest that infants are able to recognise phrases from fluent speech. The concord morphology used in this experiment seemed to facilitate the recognition, reflected in the longest listening times for the familiar concordant phrase. The results may have been caused by the infants' distributional analysis of the input, that is, distributional bootstrapping, in which the recurring structural pattern of concord morphology indicates the size and the internal structure of the relevant unit, in this case the phrase.

In further studies, Blenn and colleagues (Blenn, Seidl & Höhle, 2003) and Pelzer and Höhle (2006a) found that American English-learning infants were also sensitive to concord morphology, despite the fact that their native language does not make use of it. However, unlike the German-learning infants, the American infants only showed discrimination of concordant vs. non-concordant phrases in a language that marks concord at the beginning of words, like in Swahili. When tested on German concord morphology, the American infants displayed no discrimination abilities. The authors suggest that these findings may be attributed to language intrinsic differences, with German having a complex inflectional system, mostly realised as suffixes, so that attending to the end of words may help the infant in the language acquisition process. English, however, has a reduced system of inflectional morphology, so that infants might not be paying so much attention to word endings.

Concord morphology is a phenomenon that is restricted to syntactic phrases. However, it does not occur in all phrases of German, as has been shown, for instance, in the examples used by Pelzer and Höhle (2006b) and Pelzer (2008). Thus, the sensitivity to concord morphology can only be an additional feature in infants' recognition of phrases in fluent speech. As has been shown in this section, infants are also sensitive to the acoustic features like final lengthening to infer phrase boundaries in the speech.

Other grammatical relations are not that overtly marked and, furthermore, are realised in two phrases, for instance, subject-verb agreement which crosses the boundaries of the subject-NP and the VP in a clause. Therefore, infants must also be able to find larger structures like clauses in order to learn about the structure of their language.

2.2.3.3 Clauses / Sentences

The perception of clauses coinciding with intonational phrases, which also is the topic of the present study, has already been investigated in several languages so far. As sentences consist of one or more clauses, and the sentence boundaries are marked by the same set of prosodic features as clause boundaries, like final lengthening, changes in the pitch contour and occurrence of pauses, this section collapses findings on the perception of these boundaries. Thus, the terms clause and sentences will be used interchangeably most of the time, with clear identification of the term where a distinction is necessary. The present section will present studies on the perception of prosodic boundaries of clauses and on infants' sensitivities to changes within that domain, like word-order (violations). As in the studies on the perception of phrases, many of these studies also used the insertion of pauses with an invariant duration to investigate infants' recognition of the units without directly addressing the question of the function of the pause in relation to these units and also not posing the question of infants' sensitivity to the natural pause hierarchy.

Perception of clause boundaries in English-learning infants

In a seminal study, Hirsh-Pasek and colleagues (Hirsh-Pasek et al., 1987) examined the perception of clauses in American English-learning infants. Using the HTP paradigm, they tested whether infants discriminate between speech passages in which pauses were inserted either at clause boundaries or at a location within the clause. In a first experiment, 7- to 10- month-olds were presented with speech passages extracted from a recording of a mother interacting with her daughter. To generate their so-called natural speech samples[23], Hirsh-Pasek et al. (1987) inserted a 1 s pause at every clause boundary in the text. These locations included not only intra-sentence clause boundaries, but sometimes also sentence boundaries whose prosodic marking is even more pronounced.

An example of a part of a natural speech passage is: *Cinderella lived in a great big house, ___ but it was sort of dark ___ because she had this mean, mean, mean stepmother. ___ And [...]*. To generate their so-called unnatural speech samples, a 1 s pause was inserted at locations within the clauses. Furthermore, these speech samples started and ended in the middle of a clause, thus heightening the unnaturalness of the sample. An example of a part of the unnatural speech passage is: *in a great big house, but it was ___ sort of dark because she had ___ this mean, mean, mean stepmother. And [...]* (ibid., p. 273).

In this first experiment, the infants listened significantly longer to the natural speech samples than to the unnatural samples. The authors concluded that the infants were sensitive to the prosodic correlates of clause boundaries. However, in this first experiment, all other naturally occurring speech pauses were kept intact. As a result the unnatural speech passages had more frequent pauses than in the natural speech samples, that is,

[23] The different studies also used varying terms to describe the speech samples. Thus, *natural* and *coincident* always refer to speech samples with pauses located at the clause boundaries. The terms *unnatural* and *non-coincident* refer to speech samples with pauses at other locations within the clause.

in addition to the artificially inserted pauses all existing naturally pauses at the clause boundaries.

To test whether this discrepancy in the samples led the infants to prefer the natural speech samples, Hirsh-Pasek et al. (1987) conducted a second experiment. To ensure greater comparability of the stimuli all naturally occurring pauses longer than 450 ms were eliminated in the speech passages. Only a minimal pause was left (10-100 ms, depending on the phonemic context) so that the speech still sounded natural. Then again, artificial 1 s pauses were inserted either at the clause boundaries or at locations within the clause. In this experiment, 6- to 8-month-olds were tested. The infants again listened significantly longer to the natural speech samples, thus confirming the results of the first experiment. The authors concluded that infants in the second half of their first year of life are sensitive to the prosodic features, including pauses, indicating clauses in speech, so that they can start segmenting the speech stream into linguistically relevant units.

These findings were confirmed in a subsequent study by Kemler Nelson and colleagues (Kemler Nelson, Hirsh-Pasek, Jusczyk & Wright Cassidy, 1989), in which they tested infants on the perception of clauses in child-directed as compared to adult-directed speech. Using the same stimuli of child-directed speech as Hirsh-Pasek et al. (1987), Kemler Nelson et al. (1989) also found that 8-month-olds preferred to listen to the coincident speech samples. However, when presented with comparable speech samples in adult-directed speech, another group of 8-month-olds did not react differently to the coincident and non-coincident speech samples. The authors concluded that the sensitivity to the CDS register, in which clause boundaries are reliably marked in the speech stream (cf. e.g., Fisher & Tokura, 1996a, 1996b), helps the infants to segment syntactically relevant units from the speech stream.

Furthermore, the finding that infants are sensitive to clause boundaries in fluent speech held even when infants were tested on non-native speech samples. Using the HTP paradigm, Jusczyk (1997; see also Jusczyk, 2003) tested American English-learning 4.5- and 6-month-olds on their ability to discriminate coincident and non-coincident samples of CDS in English and Polish, both of which are stress-timed languages. He found that the 4.5-month-olds preferred to listen to the coincident speech samples in both English and Polish, whereas the 6-month-olds only preferred to listen to the coincident samples in their native language. The 6-month-olds' results persisted even when the speech samples were low-pass filtered to eliminate possibly disturbing phonemic and phonotactic idiosyncrasies of Polish. However, 4.5-month-old American English-learning infants who were tested on comparable speech stimuli in Japanese, which is mora-timed, displayed no preference for either the coincident or non-coincident speech samples. Thus, the perception of clause boundaries may be influenced by the prosodic properties of the native language from early on, so that by 4.5 months infants might still be able to discriminate coincident from non-coincident speech samples in a comparable stress-timed language, but not in a language from a different rhythmic class, that is, Japanese. By 6 months the ability to recognise clause boundaries seemed to be limited to the native language.

Fernald and McRoberts (1996) tried to replicate the studies by Hirsh-Pasek et al. (1987) and Jusczyk (1997; 2003). Using a different methodology[24], they tested 4-, 7- and 10-month-old American English-learning infants on their ability to discriminate coincident and non-coincident speech samples in English and German, both being stress-timed languages. The material consisted of samples of child-directed speech of the same story both in English and German. The stimuli were generated according to the previous studies by inserting a 1 s pause either at a clause boundary (coincident sample) or at a location within a clause (non-coincident sample).

The results of their experiments were twofold. First, they were not able to replicate the findings of Jusczyk (1997; 2003) that American English-learning 4-month-olds were able to discriminate coincident from non-coincident speech samples in a foreign language. Moreover, Fernald and McRoberts (1996) found that none of the tested age groups showed comparable discrimination abilities with the German speech materials. Furthermore, the 4-month-olds were also not able to discriminate the speech samples in their native language, though the 7- and 10-month-olds were.

The second finding was that by 7 months, the infants did not prefer the coincident versions of the samples, but showed a kind of novelty effect, preferring to listen to the non-coincident speech sample. Only by 10 months, the infants showed the same result

[24] The method was developed by Cooper and Aslin (1990). The presentation of the auditory stimulus is contingent on the infants' fixation of a checkerboard placed in front of the infant. This task may be even less demanding, because the infant does not have to turn her head to listen to the speech sounds presented.

pattern that was found by Hirsh-Pasek et al. (1987), namely the preference of the coincident over the non-coincident speech samples (for a discussion on the direction of preferences see sections 3.1.2 and 4.2.5). Fernald and McRoberts (1996) concluded that the sensitivity to clauses was not an innate capacity which can be found early in the infant regardless of the language presented, but that this sensitivity developed during the language acquisition process. And second, they claim that the infants, despite showing discrimination abilities in opposite directions at 7 and 10 months, do not discriminate clausal units, but that they have learned "to discriminate *continuous vocalisations* of their own language from those that are artificially interrupted" (ibid., p. 383).

The previously cited studies all investigated the perception of clauses / intonational phrases in American English-learning infants. Taken together, the studies provided evidence that from about 6 or 7 months on, despite the differences in the direction of the effect, infants seem to be able to recognise the prosodic correlates of clausal units in speech stimuli. However, the data on 4-month-olds and on the recognition of clauses in non-native languages seems controversial. The question remains, whether these findings can be generalised to infants learning other languages than English.

Perception of clause boundaries in Japanese-learning infants

This question was addressed in a study by Hayashi and colleagues (Hayashi, Tamekawa, Deguchi & Kiritani, 1996) who tested Japanese-learning infants on their perception of clauses in fluent speech. Using the same design as in the original Hirsh-Pasek et al. (1987) study, Hayashi et al. (1996) presented 6- and 10-month-old infants with speech passages in which a 1 s pause was either inserted at a clause boundary or at a location within the clause. All passages were in the CDS register. An example of a coincident speech passage is:

> *Ashi-kara haku mono-de-su ___ Kondo-wa kukku, o-kutsu-da-yo ___ O-kutsu-wa doo-suru-no-ka-naa. ___ O-kutsu-wa anyo-ni haku-n-de-shi-ta___ Yoisho yoisho-tte ___ Ashi-ni haku mono-de-shi-ta.*
>
> 'It's something you put on your foot ___ It's shoes this time, shoes ___ What do you do with shoes? ___ You wear your shoes on your feet ___ You do it like heave ho! heave ho! ___ It's something you put on your feet.'

An example of a part of the non-coincident speech passage is: *haku mono-de-su. Ashi-kara ___ haku mono-de-su. Kondo-wa ___ kukku, o-kutsu-da-yo. O-kutsu-wa ___* […] (ibid., p. 1566).[25] Again, the non-coincident sample started and ended in the middle of a clause.

In this study, the 10-month-olds, but not the 6-month-olds, had longer listening times towards the non-coincident speech samples than to the coincident speech samples. The 6-month-olds did not react differently to both speech passage types. The authors concluded that Japanese infants also are able to recognise clause boundaries in speech. Thus, the recognition of clauses in fluent speech is not a language-specific feature in language acquisition. But this ability is demonstrated at a later age in the Japanese-learning infants than in the American English-learning infants.

However, in a recent study Hayashi and colleagues (2002) presented data from a cross-sectional as well as from a longitudinal study in which Japanese-learning infants were tested on their perception of clauses in two-monthly intervals between the age of 4 to 5 months and 12 to 14 months. Using the same design as before, Hayashi et al. (2002) found that the recognition of clauses in speech was first detected in infants at the age of 8 months. Furthermore, the data provided evidence of three developmental stages. In a first stage, between the ages of 4 to 7 months, no preference for either a coincident or a non-coincident speech passage was found. Between the ages of 8 and 11 months, the infants preferred to listen to non-coincident speech samples and only at the ages of 12 to 14 months the preference for coincident speech samples was found. Despite the differences in the direction of the effect, the Japanese infants showed the ability to recognise the prosodic correlates of clauses in the speech passages about two month later than the American English-learning infants.

[25] Unfortunately, Hayashi et al (1996) did not provide an English translation of their speech stimuli. Many thanks to Anja Ischebeck and Shinichiro Ishihara for their help in translating the sentences.

Thus, the ability to recognise clause boundaries in fluent speech is a general rather than language specific feature and, at least for the languages tested so far, the sensitivity seems to be present in the infants at the beginning of the second half of their first year of life. In testing German-learning 6-month-olds on their recognition of clauses, the present study will provide further evidence for this claim.

Perception of clause boundaries in infants - Further evidence

Further studies investigating the perception of clauses and clause boundaries in infants did not make use of the pause insertion technique. A study by Nazzi and colleagues (Nazzi et al., 2000b) investigated whether 6-month-olds recognise sequences of words within speech passages better when these sequences were presented as coherent clausal units. Using the HTP paradigm, infants were familiarised with two single sequences of words. One sequence was a well-formed clause, which corresponded to a single intonational phrase like in *Rabbits eat leafy vegetables*. The other sequence was created of words straddling the prosodic boundary like in […] *rabbits eat. Leafy vegetables* […].

The acoustic analyses of the sequences revealed the patterns typical of intonational phrase boundaries compared to within clause locations. For instance, Nazzi et al. (2000b) found longer pauses when the sequence straddled the boundary, for instance the pause at the … *eat. Leafy* … boundary had a duration of 456 ms, whereas the duration of the pause between the same words within the clause was 27 ms. Furthermore, an increase in pitch and final lengthening was observed at the clause boundary. None of these parameters was manipulated. After the familiarisation phase, infants listened to speech passages in which the same sequence was embedded. With this experimental design Nazzi et al. (2000b) wanted to find out whether the presentation of a prosodically well-formed clause would help infants to encode the sequence during familiarisation and thus facilitate the recognition of that sequence during the test phase, even though the word content was identical to the condition that straddled the boundary.

The results of their experiment confirmed this hypothesis: the infants had significantly longer orientation times (OTs) towards the passages that contained the previously familiarised well-formed sequence, that is, the well-formed intonational phrase.

No such effect could be found for the familiarised sequences that straddled the boundary. This finding held even when the familiarisation sequences were not an exact acoustic match of the test sequences, that is, when they were extracted from the test passages, but also when the familiarisation sequences were produced as separate instances by the same speaker. The authors concluded that the intonational phrase is the organisational unit in which infants encode information during online processing of speech and that this ability is not dependent on an exact prosodic match between familiarisation and test. This capability is especially suited to the tasks which infants encounter in the natural language acquisition process as it is unlikely that infants will hear the same sentence spoken with exact the same intonation twice.

Further evidence that infants are able to recognise clauses in the speech stream was provided by a study from Soderstrom and colleagues (Soderstrom, Kemler Nelson & Jusczyk, 2005). Using a comparable experimental design, the authors familiarised 6-month-old infants with speech passages in which the same sequence of words used in the study by Nazzi et al. (2000b) was included as a well-formed intonational phrase or as a sequence straddling a clause boundary. Thus, one group of infants was familiarised with a passage containing […] what rabbits eat. *Leafy vegetables taste so good*. […]. A second group of infants listened to *Rabbits eat leafy vegetables*. Taste so good is […] (ibid., p.92; highlighting by the authors). This is a rather complex design in which each group of infants is familiarised both with a well-formed sequence (italics) and a sequence straddling the boundary (underlined). Furthermore, the well-formed unit for one group corresponds to the sequences straddling the boundary for the other group.

The acoustic analyses revealed that the intonational phrase boundary was reliably marked by a pause, a lengthening of the pre-boundary word and a pitch reset at the beginning of the next intonational phrase. After the familiarisation phase was completed, the infants were tested on four different speech passages. In these passages either the well-formed clauses (italics) were surrounded by different sentences compared to the familiarisation phase, or the sequences straddling the boundaries (underlined) were inserted in new carrier phrases, like in "[…] how rabbits eat. Leafy vegetables make […]". Thus the task for the infants was twofold: they were supposed to identify the target clausal unit within a speech passage during familiarisation and then transfer this knowledge to a different speech passage with different acoustic properties to recognise that same clause.

The analysis of the infants' listening times yielded two results. The infants showed a familiarisation effect, reflected in longer listening times for the familiarised sequences (both clausal and straddling) than to the unknown sequences (both clausal and straddling). However, the mean OT to the clausal sequences was longer than to the straddling sequences. Soderstrom et al. (2005) concluded that the familiarity with word sequences (clausal and non-clausal) subsequently helped the infants to recognise them in different speech passages and that the recognition was further enhanced when the sequence of words was presented as well-formed intonational phrase. These findings again provide evidence that infants are sensitive to clauses / intonational phrases during online processing of speech.

Whether infants are able to exploit the prosodic information related to clauses to learn about the inner organisation of sentences was addressed in two studies by Mandel and colleagues (Mandel, Jusczyk & Kemler Nelson, 1994; Mandel, Kemler Nelson & Jusczyk, 1996). They tested 2-month-old infants' memory for words in three different conditions. Either the words were presented within a prosodically well-formed unit, that is, a clause / sentence, or the same words presented as word lists or as a series of words belonging to two different phrases, that is, straddling a clause boundary. The experimental technique used was a variant of the HAS paradigm, in which a 2-minute delay between the habituation phase and the test phase was included. This 2-minute delay was inserted to test memory effects for changes in the acoustic stimuli.

In their first study, Mandel et al. (1994) presented infants with a short sentence like *The cat chased white mice* or the same words in the same order created of single spoken items, thus without sentential prosody *the, cat, chased, white, mice* until each infant habituated to the presented stimulus. Then a 2-minute silent pause was inserted in which the infants saw colourful slides. After that, the same stimulus was presented again (control group) or, for the experimental group, the stimulus changed to, for instance, *The rat chased white mice* in which only one segment differed, or to the respective word list. In this experimental group, only the infants in the condition with sentence prosody showed a significant dishabituation effect for the change stimulus, but not the children in the word list condition.

This finding was further extended in a second experiment in which infants were tested on a prosodically well-formed sentence as opposed to the same words straddling a clause boundary. An example of a stimulus for the habituation phase is *Cats like park benches* for the sentence condition vs. (*knows what) cats like. Park benches (are their favourite)* in the sentence-fragment condition. Again, after a 2-minute delay, only the infants in the sentence condition dishabituated on the change of the stimulus sentence to *Cats strike park benches*. The infants in the sentence-fragment condition did not seem to notice the segmental change in the words.

In a second study, Mandel et al. (1996) changed the test stimuli in a way that infants were no longer required to detect phonetic changes in the material but to detect word order changes within the sentence or sentence fragments, respectively. Using the same experimental technique, Mandel et al. (1996) tested 2-month-old infants on their perception of word order changes, like a change from *Cats would jump benches* to *Cats jump wood benches*. Comparable to the previous study, infants only dishabituated in the condition in which the words were part of a prosodically well-formed clause and not if the words were part of two adjoining clauses. This series of experiments led the authors to conclude that the sentential prosody helps infants to detect changes in the information encoded within the domain of a clause / sentence. That is, even 2-month-olds seem to perceive the prosodic correlate of a clause as an organisational unit in the speech stream and are sensitive to changes within that domain.

Perception of clause boundaries - Non-behavioural studies

In recent years a growing number of studies using different non-behavioural techniques have investigated the perception and processing of speech in the adult's, child's, and even in the infant's brain. A general lateralisation for the processing of speech has been found: in the left hemisphere syntactic, semantic and phonological information is being processed; in the right hemisphere the prosodic speech information is being processed. However, a dynamic interaction is assumed to fully process speech (Friederici &

Alter, 2004). There are several studies which have looked at the processing of the prosodic correlates of clause boundaries in sentences identical or compatible to the ones used in the present study. Therefore in this section mainly the results of these studies will be presented. In all of these, studies the intonational phrase boundary coincided with a syntactic clause boundary.

In an early study using the ERP technique, Steinhauer and colleagues (Steinhauer, Alter & Friederici, 1999) presented adults with sentences which either contained one IPh boundary like in *Peter verspricht Anna zu arbeiten / und das Büro zu putzen* (‘Peter promises Anna to work / and to clean the office’; translations by the authors) or which contained two IPh boundaries like in *Peter verspricht / Anna zu entlasten / und das Büro zu putzen* (‘Peter promises / to support Anna / and to clean the office’). The naturally produced pauses at the boundaries were not manipulated. The authors found an ERP component which was clearly related in the time-course to the processing of the IPh boundary, the so-called Closure Positive Shift (CPS). This positivity in the ERP signal was found to peak at about 500 ms after the occurrence of the prosodic boundary. For the sentences including one IPh boundary, only one CPS was found, for the sentences including two IPh boundaries correspondingly two CPS’ were found. The CPS component was found equally distributed over both hemispheres in adults, thus indicating the interaction of both hemispheres in the processing of speech.

A control condition was created in which the sentences were spliced by substituting the subject and the matrixverb, so that the sentence formerly containing one IPh boundary now had two IPh boundaries as in *Peter verspricht / Anna zu arbeiten / und das Büro zu putzen* (‘Peter promises / Anna to work / and to clean the office’), which lead to a prosody-syntax mismatch. In this condition also two CPS’ were found, however later components in the signal, like the N400 and the P600[26], indicated a syntactic and semantic reanalysis of the sentences after the occurrence of the verb *arbeiten* (‘work’). Steinhauer et al. (1999) concluded that the CPS was an automatic response driven by the prosodic processing of an utterance. Furthermore, the CPS also occurred when the existing pause at the IPh boundary was deleted (Steinhauer et al., 1999, second experiment).

[26] The N400 is associated with semantic processes and the P600 is associated with syntactic processes (cf., Friederici & Alter, 2004)

Thus, for German adults pause does not seem to be a necessary cue for the processing of IPh boundaries.

A subsequent study compared the processing of normal sentences and manipulated sentences. Using the ERP paradigm, Pannekamp and colleagues (Pannekamp, Toepel, Alter, Hahne & Friederici, 2005) tested adults on a comparable set of sentences containing either one or two IPh boundaries, like in *Kevin verspricht Mama zu schlafen / und ganz lange lieb zu sein* (‘Kevin promises mom to sleep / and to be a good boy for a while’; translations by the authors) vs. *Kevin verspricht / Mama zu küssen / und ganz lange lieb zu sein* (‘Kevin promises / to kiss mom / and to be a good boy for a while’). Again, the naturally occurring pauses were not manipulated. The processing of the natural sentences replicated the finding of a CPS distributed over both hemispheres, indicating the dynamic interaction of speech processing. In the condition where the sentences were hummed[27] a strong right lateralisation of the CPS was found. The results again support the notion that the occurrence of the CPS is prosodically driven.

Two additional studies investigated the processing of prosodic boundaries including pauses in children. Leuckefeld (2001) studied the occurrence of CPS in 9-year-old children, using sentences containing either one or two IPh boundaries like in *Die Tante verspricht Katrin zu schwimmen / und die Insel zu erkunden*. (‘The aunt promises Katrin to swim and to explore the island’) or *Die Tante verspricht / Katrin zu begleiten / und die Insel zu erkunden*. (‘The aunt promises to accompany Katrin and to explore the island’; for a thorough description of the stimuli see chapter 3.2 of the present study). These sentences were comparable to the ones in the original Steinhauer et al. (1999) study. Leuckefeld (2001) found that the children also displayed a bilaterally distributed CPS in relation to the processing of the prosodic boundary. That is, 9-year-old children already show an integrated processing of prosodic boundaries in an adult-like fashion.

In a recent study, Wartenburger and colleagues (Wartenburger, Steinbrink, Telkemeyer, Friedrich, Friederici & Obrig, 2007), using the same sentences as Pannekamp et al. (2005), investigated the processing of prosodic boundaries in 4-year-old children.

[27] It was taken care that even the hummed sentences conformed to the overall intonation contour of normal speech, thus clearly marking an intonational phrase boundary (Pannekamp et al., 2005, p. 418).

Wartenburger et al. (2007) presented the children with sentences either in a natural speech condition, a hummed speech condition or in a flat speech condition, which was created from the hummed sentences. In all sentences the naturally occurring pauses at the boundaries were left intact. Using Near Infrared Spectroscopy[28] (NIRS) as the experimental technique, they found that 4-year-olds already showed a lateralisation in the processing of speech. Natural speech elicited greater activation in the left-hemisphere than hummed speech, which in turn, elicited greater activation in the right hemisphere compared to natural speech. For the hummed sentences a greater activation in the right hemisphere was found compared to the flat sentences. Thus, the authors concluded that at the age of 4-years an adult-like processing of speech is in place.

As infants have been found to be sensitive to clause boundaries in the speech stream, the question remained whether infants would also display the CPS in the ERP pattern as a response to the processing of clause boundaries. This question was addressed in a recent study by Pannekamp and colleagues (Pannekamp, Weber & Friederici, 2006). Using the same sentences as in the study with adults (Pannekamp et al., 2005), the authors measured the reaction of 8-month-old German-learning infants to the critical difference in the first part of the sentence, that is, between *Kevin verspricht Mama zu schlafen ...* 'Kevin promises mom to sleep …' (no boundary) vs. *Kevin verspricht / Mama zu küssen ...* 'Kevin promises / mom to kiss…' (one boundary), leaving aside the other IPh boundary before the coordinating conjunction which both sentence types share. The results of the analysis of the ERP components showed that 8-month-olds reacted with a CPS in the ERP pattern after the processing of the prosodic boundary. However, as has also been found with other ERP components in children and infants, the occurrence of the positivity is delayed compared to the adults' data. Pannekamp et al. (2006) concluded that already 8-month-olds identify prosodic correlates to clause boundaries in the speech stream and use this ability to segment fluent speech, although it is not yet fully developed, reflected in the temporal delay in the pattern compared to adults.

[28] A method recently adapted for the investigation of speech in infants and children (cf., Aslin & Mehler, 2005). During the processing of speech the haemodynamics, more specifically, the changes in blood oxygen concentrations associated with neural activity are measured.

Taken together, the results from the behavioural and the non-behavioural studies provide evidence that adults, children and also infants are sensitive to the prosodic correlates of clause boundaries in the speech stream. However, none of these studies directly posed the question of the status of the pause in speech processing, though several studies used the insertion of pauses in speech materials as a means to test infants' sensitivity to language. The next section will focus on the perception of pauses in natural speech.

2.2.3.4 Pauses

Pause perception in adults - Behavioural studies

One interesting feature of pauses is that they are sometimes not perceived by the listener, even though there is a measurable break in the signal. And vice versa, sometimes a pause is perceived even if there is no measurable pause in the signal (cf., Butcher, 1981, p. 79f).

In a judgement task with adults, Butcher (1981, p. 83) found that 93% of actual pauses over 1 s duration were perceived as long pauses, 89% of pauses between 300 ms and 1 s are perceived as short pauses and 71% of pauses with a duration of less than 300 ms were not perceived at all. There was a high correlation between pauses perceived as long pauses and the occurrence of a major prosodic boundary, as well as the opposite correlation, that is, most of the unheard pauses occurred between words within a phonological phrase. There was also a correlation between the duration of the pause and the position it occurred in: between prosodic phrases pauses "were not heard by 75% of listeners until they are approximately 220 ms long", but within prosodic phrases pauses "are heard by the same proportion of listeners when only 80 ms long" (ibid., p. 205).

Duez (1993) showed that the manipulation of the duration of a pre-boundary syllable led to an increase in pause reports, though no pause was measurable in the signal. In a first experiment, Duez (1993) presented 10 naïve listeners with French speech samples either in a natural speech condition or in a spectrally inverted speech condition. The participants, all native speakers of French, were asked to press a button whenever they perceived a pause in the speech stream. In addition to the pauses that were measurable in the signal, many pauses were reported which were not measurable. Duez (1993) called these reported pauses *subjective pauses*. Subsequent analyses of the data revealed that most of these subjective pauses were perceived at syntactic phrase boundaries. Several of the pauses were reported at clause boundaries and only a few subjective pauses were reported at sentence boundaries. Some of the subjective pauses were perceived at within phrase locations. However, these pauses were due to the presence of a filled pause in the speech stream (ibid., p. 28). The acoustic analysis of the words preceding the subjective pauses revealed that in most cases there was a significant increase in vowel duration (final lengthening), a considerable F0 change between the vowel at the boundary relative to its preceding one, and also a change in the amplitude between these two vowels. However, only the factor vowel duration led to a significant increase in pause reporting, not the other two factors.

In a second experiment, Duez (1993) manipulated each of these factors in turn, leaving the respective other two unchanged. The results of this experiment provided further evidence that the manipulation of the duration of the vowel before a perceived pause was the most influential factor. The stepwise elongation of the vowel duration led to a corresponding increase in reported subjective pauses. This was not found in the condition in which F0 and intensity were likewise manipulated. Thus, the results of these two experiments provided evidence that pauses are expected in the speech stream following final lengthening, which in turn is most often found at syntactic boundaries. The distribution of the subjective pauses is inverse to the actual occurrence of pauses at these boundaries. Thus, even though in natural speech not each phrase boundary is marked by a pause, most of the subjective pauses were reported at such locations.

Cross-linguistic studies on the perception of pauses were presented by Stuckenberg and O'Connell (1988, comparing English and German) and Chiappetta and colleagues (Chiappetta, Monti & O'Connell, 1987, testing Italian and English speakers with Italian speech samples). Chiappetta et al. (1987) found no differences in the perception of pauses as a function of native or non-native language. Short pauses were more often correctly identified by the native speakers. However, pauses with a duration from 300 ms to 1 s were more often correctly identified by the non-native speakers. Long pauses with a duration of more than 1 s were identified by all listeners equally well (around 93%

accuracy). Both groups of listeners overestimated the actual duration of the pauses they perceived.

In the study by Stuckenberg and O'Connell (1988) it was found that the German speakers identified more pauses in their native language than did the American speakers. The German speakers also identified pauses in the native and the non-native speech samples equally well, whereas the American speaker were better in their native than in their non-native language. In their study, the short, medium and long pauses in the English samples were equally well identified by both groups, whereas these pauses in the German samples were better perceived by the native speakers than by the non-native speakers. The authors suggest that these findings are due to the German speakers' second language knowledge of English, whereas the American speakers had no second language knowledge of German. More subjective pauses were perceived in the English speech samples by both speaker groups than in the German speech samples. The reported subjective pauses could almost all be related to syntactic units, final lengthening effects or changes of amplitude or intonation (ibid., p. 26). Stuckenberg and O'Connell (1988) attributed the differences in the results compared to the study by Chiappetta et al. (1987), like the difference in the perception of subjective pauses, to the languages used in the studies, with German and English being prosodically more similar than Italian and English.

These cross-linguistic studies supported the findings by Butcher (1981) and Duez (1993), namely that the perception of a subjective pause seemed correlated with the occurrence of syntactic boundaries, especially in the presence of the lengthening of the word final vowel. That is, speakers tended to expect a pause at a location in speech which is characterised by a long duration of a word final vowel, a characteristic convergent with a syntactic boundary (cf., Kohler, 1983). Furthermore, knowledge of a language either as a native speaker or as a second language learner increased the accuracy in the perception of actual pauses.

Pause perception in infants - Behavioural studies

As already discussed in section 2.2.3, several studies employed a kind of pause insertion technique to investigate infants' recognition of clauses and phrases in the speech stream (e.g., Hirsh-Pasek et al., 1987; Kemler Nelson et al., 1989; Jusczyk et al., 1992; Gerken et al., 1994). These studies provided evidence that infants prefer to listen to speech passages that had pauses inserted at syntactic boundaries, like clause or phrase boundaries which coincided with major prosodic boundaries, rather than speech samples with pauses inserted elsewhere in the utterance. These results can be interpreted as evidence that infants as young as 6 months of age are sensitive to the occurrence of pauses at the boundaries of relevant speech units.

In a series of recent studies Seidl and colleagues (Seidl, 2007; Johnson & Seidl, 2008; Seidl & Cristià, 2008) tested infants on their perception of pauses, and other prosodic boundary markers, and the use infants make of these cues in their recognition of clauses and phrases in the speech stream. These studies will be presented in more detail in chapter 2.3. From these studies it could be seen that American English-learning 4-month-olds, but not 6-month-olds, and Dutch-learning 6-month-olds seemed to be sensitive to pauses which occurred at prosodic boundaries and seemed to rely on the occurrence of a pause as a necessary cue, because in experiments in which the duration of the pause was manipulated, the English 4-month-olds and the Dutch 6-month-olds did no longer recognise the clauses in the test passages.

Thus, pauses do seem to play an important role in the perception of phrases and clauses in adults and infants. It was found that adults expect pauses to occur at major syntactic boundaries, reflected in the fact that they reported subjective pauses for such locations even if there was no measurable break, especially in the environment of pre-boundary final lengthening. For infants, pauses also seem to be of some importance. The converging evidence that infants discriminate between speech stimuli in which a pause occurred at a boundary location as opposed to a location within a syntactic / prosodic unit shows that infants are sensitive to the occurrence and location of pauses in speech.

The previously presented studies all used behavioural methods to asses the sensitivity to pauses as a cue in the perception of boundaries. The next section will present studies using non-behavioural techniques.

Pause perception - Non-behavioural studies

There are not many studies which investigated the status of the pause in the perception of speech in non-behavioural studies. Most of the studies cited in section 2.2.3.3 did not manipulate the naturally produced pause at the IPh boundary. Thus, the pause was only one cue among other cues marking the prosodic boundary.

However, Steinhauer et al. (1999) tested whether the presence (Experiment 1, 2) or absence (Experiment 3) of a pause influenced the processing of IPh boundaries in adults. They found that even after the pause was removed in a sentence like *Peter verspricht* ___ *Anna zu entlasten und das Büro zu putzen*. ('Peter promises to support Anna and to clean the office.') a CPS in the ERP pattern could be seen. Thus, the prosodic processing of IPh boundaries in adults did not seem to rely on the pause, but on other prosodic cues like final lengthening and pitch change (cf., Aasland & Baum, 2003, for English). This finding can be related to the findings on the perception of a subjective pause in the vicinity of increased duration of the phrase-final vowel (cf., Duez, 1993) in adults.

This leaves open the question whether infants also are sensitive to the occurrence of a pause at the prosodic boundary. Ongoing research has provided first evidence that, contrary to the findings of Steinhauer et al. (1999) for adults, 5-month-old German-learning infants still seem to rely on the occurrence of the pause for the processing of a prosodic boundary.

Männel (p.c.) investigated whether 5-month-old infants display a CPS when the pause at the IPh boundary was deleted. Following the study of Pannekamp et al. (2006), she presented infants with sentences which either had a 500 ms pause at the IPh boundary or with sentences which had no pause[29] at the IPh boundary. In her study, Männel found that the infants did not display a CPS in the condition in which the pause was eliminated at

[29] A pause duration of 50 ms had to be retained because of the prosodic environment at the phrase boundaries and to ensure the naturalness of the utterance.

the clause boundary. However, even the alleged CPS in sentences containing the pause (cf., Pannekamp et al., 2006) might not be a genuine CPS after all, but rather a positivity related to the reset of the speech after the pause. This positivity is a so-called obligatory component in the ERP-signal related to the onset of speech. Thus, infants seemed to have perceived the pause within the sentence and reacted with a positivity to the reset of speech. However, these initial findings need to be more closely investigated in future research.

The combined studies on the perception of pauses in infants and adults might indicate that adults are more flexible in their perception of speech than infants. The findings that adults, on the one hand, still display a CPS in the ERP pattern, even when the pause at the prosodic boundary is removed, and, on the other hand, perceive a subjective pause in the vicinity of such a boundary, in addition to the identification of actually occurring pauses, indicates that not the physical presence of a pause in the speech signal triggers the perception of phrases and clauses in adults. Rather, pauses seem to have a more complementary status: if other cues indicate a major prosodic boundary, a pause is assumed to also be present. Furthermore, the finding that the estimated duration of a pause correlated with the type of prosodic boundaries further supports the notion of a pause hierarchy in the speech, like the one presented in Figure 2.1. Infants might need more cues and more consistent cues, like the actual occurrence of a pause in the speech signal, to identify major speech units in their language.

Thus, the third important question to be answered is which cues are used by adults and infants, respectively, to identify clause boundaries.

2.3 Use of prosodic cues to identify syntactic boundaries

This section addresses the question whether the prosodic cues to phrase and clause boundaries that are produced by the speaker and perceived by the listener are also used as cues to identify these boundaries.

Use of prosodic cues by adults

The studies presented in the previous sections on the perception of clauses and pauses in adults suggest that adults may rely on other cues than pauses to identify major syntactic / prosodic boundaries in their language. One possibility to test which cues are necessary to identify these units is to manipulate these cues at the potential boundaries of syntactically ambiguous sentences.

Lehiste and colleagues (Lehiste et al., 1976) manipulated the factor duration at the boundaries of ambiguous constructions like (1) *The hostess [[greeted][the girl]][with a smile]* meaning: the hostess smiled vs. (2) *The hostess [greeted][[the girl][with a smile]]*[30] meaning: the girl smiled. They found that when the critical word was expanded (e.g., *greeted*) and the following word(s) compressed (e.g., *the girl*) the adult listeners would reliably interpret the sentence as having a major prosodic boundary between *greeted* and *the girl*, thus, disambiguating the sentence as having the meaning denoted in (2). In the other case, when *greeted* was compressed and *the girl* expanded, the listeners disambiguated the sentences as having the meaning denoted in (1). To avoid influences of the pitch cue, the fundamental frequency in all the sentences was held constant and there were also no pauses anywhere between words. Lehiste et al. (1976) concluded that duration was a sufficient cue to disambiguate syntactically ambiguous sentences.

To test the individual influences of duration, pitch and amplitude, Streeter (1978) conducted a listening experiment using algebraic expressions like (1) *[A plus E] times O* vs. (2) *A plus [E times O]*. In an experiment where she imposed the values of duration, pitch and amplitude, either individually or in various combinations, of the utterance (1) onto the utterance (2), she found that cue 'duration' and the combinations of the cues

[30] The [ ] brackets show the two different readings of the syntactically ambiguous sentence.

‘duration and amplitude’, ‘duration and pitch’ and the combination of all three cues, led the listeners to interpret the heard utterance as having the meaning in (1). Amplitude and pitch alone did not lead the listener to infer the first reading, instead most of the listeners still disambiguated the utterance as having the reading (2). These results again support the notion that for adults ‘duration’ seems to be a good indicator for syntactic bracketing reflected in major prosodic boundaries. However, in her experiments Streeter (1978) found that the combination of cues ‘duration and pitch’ was more effective in disambiguating the utterance than the cue ‘duration’ alone.

The influence of pauses in the disambiguation of syntactically ambiguous sentences was investigated in a study by Scott (1982). She used utterances like (1) *Kate or [Pat and Tony] will come*. vs. (2) *[Kate or Pat] and Tony will come*. Depending on the bracketing of the sentence the speaker produced a pause either after the first name in utterance (1) or after the second name in utterance (2). In her experiments, Scott (1982) manipulated the utterances in that she inserted either a) artificial pauses up to a duration of 562 ms at the “wrong” position in the utterance, for instance, in utterance (1) an artificial pause was inserted after the second name. Or b) she lengthened the phrase-final segment and inserted a short pause at the “wrong” boundary; or c) she lengthened the duration of the whole foot to create a “wrong” boundary. To eliminate the pitch cue, the fundamental frequency in all utterances was held constant. The results of her experiments again showed that the lengthening of the pre-boundary element, irrespective of the type of lengthening used, led the listener to perceive a major boundary. Thus, as in condition a) in the absence of a segmental lengthening, the insertion of a pause led the listeners to interpret the utterance (1) as having the meaning of utterance (2). This was also true for condition b) and c). Therefore, Scott (1982) concluded that the increase in duration of the interstress interval (cf., Lehiste, 1977) was responsible for the disambiguation of syntactically ambiguous sentences.

These studies provide further evidence that for adults the pause does not seem to be a necessary cue to identify major boundaries, like phrase and clause boundaries, in the speech stream. One possible explanation for this might be that, as has been described in section 2.1.2, on the one hand, syntactic and / or prosodic units are often but not always demarcated by a pause in ADS and, on the other hand, pauses may occur in other positions in the speech stream and also have different functions. Therefore, adults, being competent language users, might no longer rely on pauses as a main indicator of syntactic and / or prosodic boundaries. However, infants might still exploit the occurrence of a pause in relation to a major prosodic boundary to infer the syntactic structure of their language. This question will be addressed in the following section.

Use of prosodic cues by infants

As has been mentioned before, the studies which investigated whether infants are sensitive to phrase and clause boundaries used the insertion of pauses more or less as a means to an end. The studies did not address the status of the pause in relation to the prosodic boundaries, for instance, in the studies investigating the perception of phrases (e.g., Jusczyk et al., 1992) and clauses (e.g., Hirsh-Pasek et al., 1987) the duration of the inserted pause was 1 s, though in the analyses of natural speech corpora it was found that the duration of a pause is correlated to the type of boundary present (cf., Grosjean & Collins, 1979). Only recently, the question of whether the presence or absence of a pause, and the manipulation of other boundary markers, influences infants' recognition of clause boundaries, was addressed.

In order to further specify which of the prosodic markers to clause boundaries, that is, final lengthening, changes in fundamental frequency and pauses, might be used by the infants to extract clauses from fluent speech, Seidl (2007) conducted a series of experiments with 6-month-old American English-learning infants. Using the same experimental design and the same word sequences as Nazzi et al. (2000b), that is, *Rabbits eat leafy vegetables* vs. […] *rabbits eat. Leafy vegetables* […], she manipulated each of the indicators to clause boundaries both individually and in different combinations in her experiments.

The acoustic analyses of the stimuli revealed that the stimuli differed significantly from each other depending on the clausal status of the sequence. The duration of the vowel, as in *eat*, was twice as long when followed by the clause boundary than when occurring within the clause. The duration of the pause was about eight to ten times longer at the clause boundary than between the same words within the clause, that is, 300 ms at the boundary and 30 ms within the phrase. Furthermore, a significant pitch change (resetting) was found between the words straddling the clause boundary as compared to the same words within the clause. Thus, the clause boundary was clearly marked by acoustic parameters.

The first experiment served as a baseline for the subsequent experiments in which the parameters were manipulated. Using the HTP paradigm, 6-month-old American English-learning infants were familiarised both with the clausal sequence *rabbits eat leafy vegetables* and with the sequence containing the same words but also including the prosodic boundary as in *rabbits eat / leafy vegetables*. The infants were tested on two text passages which contained the familiarised word sequences, that is, during testing infants again heard both conditions, the clausal and the non-clausal one. The results of this first experiment showed that infants significantly preferred to listen to the text passage that contained the previously familiarised clausal sequence as opposed to text passages that contained the also previously familiarised sequence of words straddling the prosodic boundary. Thus, Seidl (2007) was able to replicate the findings of the original study by Nazzi et al. (2000b).

In her second experiment Seidl (2007) manipulated the factor pause duration. The original duration of the pauses in the example stimuli were 30 ms between *eat* and *leafy* within the clause and 300 ms when straddling the clause. There were two conditions for the familiarisation phase in this experiment. In one condition, all existing pauses between the words straddling the clause boundary *eat / leafy* or the same words within the clause were removed. In the second condition, a pause of 300 ms was inserted at the relevant locations. During the test phase, the infants listened to the text passages of the original experiment. The results of this experiment suggested that the infants in both conditions again preferred to listen to the text passages that contained the familiarised clausal units. Seidl (2007) interpreted this result as an indication that the presence of a pause is not a necessary cue for the recognition of clauses in fluent speech: neither did the insertion of the 300 ms pause within the clause interrupt the identification of the clausal unit, nor did the deletion of the pause at the prosodic boundary lead the infants to interpret this unit as coherent. Note however, that the other prosodic cues indicating the clause boundary were still present in this condition.

In the next experiments either the pitch contour or the factor final lengthening were manipulated. In the experiment manipulating the pitch contour in which the familiarisation stimuli were presented with flattened pitch, the 6-month-olds did not show a preference for either the clausal or the non-clausal stimuli. Even when the pitch contour was reversed, that is, the pitch contour of the clause sequence was mapped onto the non-clausal sequence and vice versa, the infants again did not react differently to the text passages. Seidl (2007) interpreted these findings as an indication that pitch change is a necessary cue for the recognition of clauses. However, as the experiment in which the reversed pitch contour was used also did not lead 6-month-olds to prefer to listen to either of the speech samples, pitch alone could not be responsible for a reliable identification of a clause boundary. Rather, the author suggested, the convergence of cues seemed to be the best indicator of a clause boundary. The next two experiments that manipulated the factor final lengthening also provided evidence that infants preferred to listen to the text passages in which the previously familiarised clauses were included. Thus, as with the factor pause duration, the factor final lengthening did not seem to be a necessary cue for the recognition of clause boundaries.

To further test whether pitch is not only a necessary cue, but also a sufficient cue for the recognition of clause boundaries, Seidl (2007) conducted an experiment in which the factors pause and final lengthening were neutralised by again using a short or a long pause condition and by shortening or lengthening the final vowel of the critical word. Only the factor pitch remained unchanged. In this experiment the 6-month-olds did not react differently to the clausal and non-clausal passages. The author concluded that the factor pitch, though being a necessary cue, is not a sufficient cue for the recognition of clause boundaries in fluent speech.

The results of all experiments provided evidence that 6-month-old infants used some of the acoustic markers to identify clause boundaries in fluent speech, but that the cues were weighted differently. Pitch seemed to be a necessary cue, changes in the pitch contour resulted in a null effect in the experiment, pauses and final lengthening did not seem to be necessary cues. However, pitch alone was not sufficient for recognition, the infants rather seemed to rely on converging cues in their analysis of fluent speech.

In a follow-up study, Seidl and Cristià (2008), using the same stimuli as Seidl (2007), found that American English-learning 4-months-olds seem to rely on a holistic strategy in segmenting clauses. Only in the original experimental condition, 4-month-olds did discriminate the clauses from the non-clauses, but not in the condition where the cues had been manipulated individually. Thus, the 4-month-olds, but not the 6-month-olds tended to discriminate clauses from non-clauses only when all cues to clause boundaries converged.

To investigate infants' use of cues in the recognition of clause boundaries from a cross-linguistic perspective, Johnson and Seidl (2008) tested Dutch-learning 6-month-olds, using the same experimental design as in Seidl's (2007) original study. The Dutch-learning infants were familiarised with both the clause sequence *De jongens eten koude pizza* 'the boys eat cold pizza' and the non-clausal sequence *de jongens eten. Koude pizza* 'the boys eat. Cold pizza', which was extracted from the two sentences *Tante fragt zich af wat de jongen eten. Koude pizza smaakt niet so goed.* 'Aunt wonders what the boys eat / are eating. Cold pizza doesn't taste so good.' (translations by Johnson & Seidl, 2008). The two test passages also contained either the clausal sequence or the non-clausal sequence.

The acoustic analysis of the prosodic cues at the clause boundary revealed a lengthening of the pre-boundary full vowel, a pitch reset after the boundary and a pause with a duration of 820 ms between *eten* and *koude* (ibid., Table 1). Thus, the pause at the clause boundary was more than twice as long as the one found in the English stimuli of Seidl's original study. The results of the first experiment, using non-manipulated material, provided evidence that, like the English-learning infants, the Dutch-learning infants preferred to listen to the clausal sequence.

In a second experiment the pause cue was manipulated in the same way as in Seidl's original study, that is, either the pause at the clause boundary in the non-clausal unit was eliminated, or an artificial pause of 738 ms, which corresponds to the mean duration of pauses found at the clause boundary, was inserted between the words *eten* and *koude* within the clause. The test passages were the same as in the first experiment.

The collapsed analysis of the two test conditions (pause deletion and pause insertion) revealed that the Dutch-learning 6-month-olds did not react differently to the clause sequences than to the non-clausal sequences. Johnson and Seidl (2008) therefore concluded that for Dutch-learning infants the pause occurring at the clause boundary is a necessary cue for the identification of clauses in the speech stream.

When tested on the unmodified English stimuli from Seidl (2007), Dutch infants also failed to discriminate clauses from non-clauses. Johnson and Seidl (2008) assumed that this is because the Dutch 6-month-olds may already be attuned to the characteristics of Dutch prosodic boundary marking.

Thus, the importance of pauses for the recognition of clauses seems to be influenced by the target language weighting of cues, as English adults also have been shown to weight the pause cue less than other cues. Dutch adults, on the other hand, have been shown to rely more heavily on pauses as cues to syntactic boundaries (Seidl, 2007, p. 35).

Taken together, the results of these three studies demonstrate that infants are able to use prosodic cues, like changes in the pitch contour, final lengthening and pauses to identify clause boundaries in the fluent speech. These findings convincingly disprove Fernald and McRoberts' (1996) claim that the prosodic cues provided by the input are far too unreliable that infants could use them to bootstrap into their language. Specifically, they claimed that the cues signalling clause boundaries like pauses and pitch changes were merely probabilistic in their distribution in the speech signal and therefore not reliably exploitable. The studies by Seidl and colleagues (Seidl, 2007; Johnson & Seidl, 2008; Seidl & Cristià, 2008) demonstrated that even if a single cue is not sufficient to recognise clause boundaries, the cues like pause and pitch are necessary cues for the identification of these boundaries and infants are able to use these cues already at the age of 6 months.

The weighting of these cues seems to be both influenced by developmental factors and language-specific characteristics. English-learning 4-month-olds seemed to rely on the convergence of cues for the identification of clause boundaries, whereas English-learning 6-month-olds did not. For these infants pitch, in combination with lengthening or pause, was a necessary cue to identify the clause boundaries, whereas pause alone was not a necessary cue for them. For Dutch-learning 6-month-old infants the occurrence of a pause was a necessary cue to identify the clause boundaries. Johnson and Seidl (2008) discussed these findings in relation to target-language acoustic differences: in the English stimuli they found only a short pause of 300 ms at the clause boundary as compared to the Dutch stimuli which had a pause with a mean duration of 738 ms at the boundary. With respect to the changes in the pitch contour, they found that the pitch resets in their English stimuli were about two times greater than in their Dutch stimuli. Thus, the infants' weighting of the different cues may be driven by language-specific characteristics.

However, there may be an alternative explanation of why the English infants in Seidl's (2007) study preferred to listen to the clauses even in the pause neutralised conditions and the Dutch infants did not react correspondingly. This explanation takes into account the findings on different pause durations that suggested the existence of a natural pause hierarchy (cf., Figure 2.1). In both studies, Seidl (2007) and Johnson and Seidl (2008) manipulated the pause cue by inserting an artificial pause between words within the clause. For the English stimuli this resulted in the insertion of a pause of 300 ms duration, which was the longest pause duration that was found for the non-clausal stimuli[31] (Seidl, 2007, p. 32). Thus, in the original and the short pause condition, there was a pause of 30 ms between *eat* and *leafy* in the clause *Rabbits eat leafy vegetables* and in the long pause condition there was a pause of 300 ms between *eat* and *leafy*. Correspondingly, in the natural and long pause condition the pause between *eat. Leafy* at the clause boundary was 300 ms long and in the short pause condition only 30 ms long.[32]

In the Dutch stimuli there was a pause of 73 ms between the words *eten* and *koude* in the clause *De jongens eten koude pizza* in the natural and in the short pause condition, and in the long pause condition an artificial pause with a duration of 738 ms, which corresponded to the mean duration of pauses found in the non-clausal stimuli, was inserted

[31] This pause duration at the clause boundary is rather short compared to the findings on American English CDS (cf., Fernald et al., 1989).

[32] Therefore, the non-clausal sequences in the familiarisation string may constitute conflicting input for the infants. They hear an interstimulus pause of 1 s (Seidl, p.c.), associated with a pause between sentences, then the beginning of a sequence which is spliced out of a whole utterance, thus probably deviating from the prosodic characteristic of an utterance onset. Then a clause boundary follows, which is indicated by pitch and duration markers, but with a rather short pause. The offset of the sequence then again does not resemble the offset of an utterance.

between *eten* and *koude*. Furthermore, to create the familiarisation strings Seidl (p.c.) inserted an intertimulus pause of 1 s duration between the subsequent clauses and non-clausal sequences, respectively. This resulted in a familiarisation string for the clauses in English with the following pause distribution: […] *Rabbits eat* ___ $_{30/300\ ms}$ *leafy vegetables* ___ $_{1\ s}$ *Rabbits eat* […]. The familiarisation string for the Dutch clause stimuli had the following pause distribution: […] *Jongens eten* ___$_{73/738\ ms}$ *koude pizza* ___$_{1\ s}$ *Jongens eten* […].

If infants are sensitive to the natural pause hierarchy, the results of these experiments could be explained by the differences in the duration of the inserted pauses used. The natural pause hierarchy suggests that pauses between sentences have a duration of about 1 s and pauses between words within a phrase can have a duration between 0 ms and 300 ms (cf., Figure 2.1). Thus, the English infants listened in each condition of the familiarisation to well-formed strings of speech, allowing them to recognise them in the test phase. The Dutch infants were familiarised with strings that, only in the short pause condition correspond to the natural versions, and in the long pause condition violate the natural pause hierarchy, because the pause inserted between subsequent words in the phrase was too long, thus possibly impeding the recognition of these strings during the test phase[33], resulting in the null result found for this experiment.

This alternative explanation for the results of the studies by Seidl (2007) and Johnson and Seidl (2008) does not rule out that American English-learning infants also might use pause as a necessary cue to identify clause boundaries in fluent speech. A possible test of this hypothesis would be to test English-learning infants with stimuli that also violate the natural pause hierarchy in the clausal condition by inserting a pause with a longer duration, comparable to the manipulation of the Dutch clausal stimuli.

[33] Johnson and Seidl (2008) only present the collapsed results of both the short pause and long pause familiarisation condition and report no individual mean orientation times. It would be interesting to look at the individual data for each condition to verify or falsify the hypothesis presented here.

2.4 Conclusion of Part I

The question that was addressed in the present chapter was whether infants might be able to recognise syntactic clauses in the speech stream by relying on prosodic information which marks the boundaries of these clauses. The relevant prosodic cues like changes in the pitch contour and changes in segmental duration were briefly discussed while the main focus was on the role pauses might play in the segmentation process. To answer this question it was necessary to show, first, that the cues are present in the input, second, that infants are sensitive to these cues and third, that infants indeed make use of these cues.

Chapter 2.1 presented research related to the question whether the relevant cues are present in the input. It was shown that pauses, that also have other functions in the speech stream like allowing the speaker to breathe and to plan the utterance, quite frequently and reliably occurred at the boundaries of speech units, like clauses and / or intonational phrases. Furthermore, the analyses of German and English speech corpora (e.g., Goldman-Eisler, 1972; Butcher, 1981) revealed that the duration of a pause seemed to be closely related to the type of boundary it marks. Pauses which might occur between words in an utterance tended to have a short duration up to a maximum of 300 ms, the pauses between phrases were slightly longer with a duration between 300 ms and 600 ms, the pauses at the clause boundary were found to have a duration between a minimum of 500 ms and a maximum of 1 s, and the pauses between utterances were longest with a duration between a minimum of 750 ms and a maximum of 2 s. This distribution of pauses suggested the existence of a natural pause hierarchy (cf., Figure 2.1) which complements the prosodic hierarchy described by Nespor and Vogel (1986) on the signal side of the input which is correlated to the syntactic side of the language. If infants were sensitive to prosodic boundaries and especially to pauses marking these boundaries it might help them to learn about the syntactic units of their language.

Chapter 2.2 presented studies which investigated infants' sensitivity to the speech rhythm and the sensitivity to the building blocks of language, that is, words, phrases and clauses. Furthermore, the sensitivity of adults and infants to pauses in the speech stream was discussed. The mechanism that is assumed to underlie infants' ability to learn about syntactic units in their language by using their sensitivity to the prosodic units of their language is the prosodic bootstrapping mechanism (e.g., Gleitman & Wanner, 1982;

Jusczyk, 1997). More specifically, it claims that infants are using their sensitivities to the prosodic markings, like the prosodically marked boundaries of intonational phrases provided in the signal to identify relevant speech units, like clauses, in which grammatical rules apply. This sensitivity to prosodic boundaries has also been shown to help infants in segmenting words from fluent speech, as both the left and the right edge of an utterance also coincide with word boundaries (cf., Seidl & Johnson, 2006, 2008; Schmitz et al., submitted). Furthermore, various studies that used pause insertion as a means to test infants' perception of phrases and clauses in fluent speech provided evidence that infants not only are sensitive to phrases and clauses in the speech stream but that they also seem to be sensitive to the occurrence and distribution of pauses within the speech stream (e.g., Hirsh-Pasek et al., 1987; Jusczyk et al., 1992). However, none of these studies investigated whether infants use pause as a cue for the perception of clauses, nor did the studies consider the differences resulting from the type of boundary, that is, phrase vs. clause, and the duration of the pause associated with that boundary as reflected in the natural pause hierarchy.

Chapter 2.3 attempted to answer the question of whether infants use prosodic markers and especially the pauses to recognise clauses in the speech stream. In a series of studies, Seidl and colleagues (Seidl, 2007; Johnson & Seidl, 2008; Seidl & Cristià, 2008) investigated whether the pause was used as a cue in infants' perception of clauses. The results of these studies suggested a language-specific weighting of this cue: American English-learning 6-month-olds did not seem to rely on the pause as a cue in the recognition of clauses in fluent speech, whereas the Dutch-learning infants did.

However, one can discuss some methodological problems with respect to these studies. All three studies used the HTP paradigm with familiarisation phase, which is used to test infants' recognition of previously familiarised units during the test phase and does not test previous language knowledge of the infants (cf., chapter 3.1). Second, infants were familiarised both with clauses and with non-clausal units, spanning the same words, which had been spliced out of longer utterances. Therefore, these non-clausal units were also prosodically ill-formed, because they started and stopped in the middle of a natural utterance and contained a prosodic boundary in the middle. And third, there were considerable differences in the pause durations used in the experiments: in the English experiment the duration of the pause, which was artificially inserted in the clauses for the familiarisation, was 300 ms, whereas the pause duration used in the Dutch experiment was 738 ms. Furthermore, in each language the subsequent clauses and non-clausal units were divided by a 1 s interstimulus pause in the familiarisation string (Seidl, p.c.), which conforms to the pause which usually occurs between utterances. Therefore, the English-learning infants listening preference for the clauses during the experiment in which the pause cue was manipulated might have been due to the fact that the pause which was artificially inserted in the clause still was within the range of the possible pause durations found between words within a clause (cf., Figure 2.1). Furthermore, the duration of the pauses between subsequent clauses was consistent with the natural pause hierarchy. In the Dutch experiment, the artificially inserted pause was more than twice as long as in the English experiment, and therefore well outside the range of pause durations between words within a clause, which might have impeded the recognition of these clauses during the test phase. In sum, the results of these experiments do not rule out that English-learning infants also use the pause as a cue to recognise clause boundaries. Moreover, the results might also be a first indication that infants are sensitive to the natural pause hierarchy.

The present study will investigate German-learning infants' recognition of clauses and the sensitivity to pause durations and the natural pause hierarchy in a series of five experiments, using the HTP paradigm without familiarisation phase. This experimental method was chosen to test the knowledge infants already have acquired about their language. Furthermore, the test stimuli comprised sentences consisting of two coordinated clauses, therefore it is possible to test the sensitivity to both, clause and sentence boundaries. The results of these experiments will provide further evidence for the fact that infants use pauses as a cue to recognise clauses in fluent speech. The study will also provide evidence that infants as young as 8 months already are sensitive to distribution of pauses conforming to the values of the natural pause hierarchy of their native language.

II Empirical Study

3 Method and Stimuli

In this chapter the principal method of investigation used in this study and the auditory stimuli used in all five experiments will be presented. First, the developmental history of the method will be described and its reliability and adequacy for the use in the planned experiments compared to other methods used in infant language acquisition studies will be discussed. Second, the experimental setting and procedure of the method used in this study is described. Furthermore, the acoustic characteristics of the auditory stimuli used in all experiments of this study will be analysed. At last the outline of the present study is given.

3.1 Method – The Headturn Preference Paradigm

The Headturn Preference (HTP) paradigm is now being used for more than twenty years to study the early language acquisition processes in infants. An early version of the paradigm was used by Fernald (1985). In her experiment she tested 4-month-old infants on their preference for infant- versus adult-directed speech. During the experiment the child was sitting on her parent's lap in a three sided test booth. At the beginning of an experimental trial a green lamp started to flash in front of the child, when the child looked towards it, this lamp was switched off. In the four pre-test training trials the loudspeaker on one side was activated and presented the predetermined stimulus type. The types of stimuli (infant-directed vs. adult-directed) and the side of presentation were contingent, that is, the adult-directed speech always came from one side and the infant-directed speech from the other side. These training trials were used to acquaint the child with the experimental design and the types of stimuli presented. In the immediately following test phase, the first headturn of the infant, after the green lamp stopped flashing, initiated the presentation of the stimulus contingent on the side the infant looked at, simultaneously a red light on this side started to flash. The dependent variable in her experiment was the number of headturns the infant initiated to each stimulus type. Fernald (1985) found that during the experiment the infants more often turned their head to the side from which the infant-directed speech was presented. Thus, she concluded that 4-month-old infants preferred to listen to infant-directed speech compared to adult-directed speech. Using this procedure Fernald and colleagues (Fernald, 1985; Fernald & Kuhl, 1987) were able to assess the spontaneous preferences of the infant.

Hirsh-Pasek and colleagues (1987) also used the headturn-induced presentation of the stimuli in their experiment on the recognition of clause boundaries in infants (cf., section 2.2.3.3 for a detailed description of the experiment). As a second dependent variable they calculated the amount of time the infant spent looking at the side from which the auditory stimulus was presented (orientation time: OT). Comparing the validity of both measures they found that the direction of the initial headturn gave no significant indication of infants' preference, but that the duration of the headturn was a more reliable cue to infants' preference of one type of stimulus. Subsequently, the measurement of the duration of infants' orientation to the auditory stimuli became the standard analysis tool in HTP studies. The measurement of the OT as the dependent variable made it no longer necessary for the type-of-stimulus and side-of-presentation to be contingent. This allowed the development of more complex experimental designs. Furthermore, the possibility that the result of an experiment was due to an infant's preference of one side regardless of the type of stimulus presented from that side was diminished. With the availability of computer technology, it became easier to computer-control the experiments and record the OT, thus diminishing experimenter induced errors (cf., Hirsh-Pasek et al., 1987, p. 279). This variant of the method is used to assess the knowledge the infants already have acquired about (their) language.

Jusczyk and Aslin (1995) further developed the HTP procedure by including a so-called familiarisation phase. This familiarisation phase precedes the immediately following test phase and the experimental design is suitable for testing infants' learning mechanisms. During the familiarisation phase, the infant listens to speech stimuli that are different from the ones presented during the test phase. However, the procedure is the same during familiarisation and test phase to acquaint the infant with the course of events. In their study, Jusczyk and Aslin (1995) familiarised 7.5-month-old infants with lists of two monosyllabic content words for about 30 seconds each. In the immediately following test phase, the infants heard four different short text passages, two of which contained the familiarised words and two of which contained new, not-familiarised words. Jusczyk and

Aslin (1995) found that infants had longer OTs towards the text passages that included the previously familiarised words than to the text passages that included new words. They concluded that the infants were able to recognise the familiarised words in continuous speech.

3.1.1 Reliability of the HTP paradigm and range of application

In their evaluation of the reliability of the HTP paradigm, Kemler Nelson and her colleagues (Kemler Nelson, Jusczyk, Mandel, Myers, Turk & Gerken, 1995) compared the online coding of the OT during the experiment and data obtained from offline coding by a second experimenter. Both experimenters are blind with respect to the kinds of stimuli presented and to the randomised order of presentation of these stimuli in order to prevent influences on the data recording. During online coding, the experimenter controls the experiment with a response box which is connected to a computer. She starts the flashing of the centre lamp by pushing a button on the box, and when the infant orients towards this lamp, one of the side lamps is activated by a second push of the same button. The side on which the side lamp starts to flash is computer controlled. When the infant orients towards this lamp, a third push of the button starts the presentation of the auditory stimuli. The time the infant spends looking towards the side on which the stimulus is presented is recorded by the computer. If the infant turns away from the presentation side, the time recording is stopped by pushing another button on the box. If the infant returns her attention within a two second predefined cut-off window, the recording of the time is continued. If the infant looks away for more than two seconds, the presentation of the stimulus is stopped and a new trial is started. The experiment is videotaped to allow for offline coding by a second experimenter.

For this offline coding the second experimenter views the videotape of the experiment and also uses the push-button box in the same way as described above, as if online coding the experiment. The data obtained by this second coding is compared to the data obtained from the original coding. Kemler Nelson et al. (1995) report that the inter-coder reliability was very high (r = 0.94 to r = 0.99; ibid., p. 113f). Other studies confirmed the high inter-coder reliability (cf., Seidl, 2007). Thus, the HTP paradigm is a credible tool in the research of infants' language acquisition and development.

Compared to two other methods used in the research on language acquisition in infants, namely the High-Amplitude-Sucking procedure (HAS) (Eimas, 1974; Mehler et al., 1988) and the Conditioned Headturn technique (CHT) (Morgan, 1994; Werker & Tees, 1984), the HTP has the advantage of being applicable to a wider age range when testing infants' language acquisition development.

The HAS paradigm is the only one of these three methods which can be used with newborns, but it can only be applied up to the age of 4 to 6 months, the age when the natural sucking reflex in the baby starts to decay. The HAS is a habituation paradigm in which the natural sucking reflex is used to test infants' discrimination abilities. When presented with a stimulus, the infant usually starts to suck faster and / or more intense, that is, with a high amplitude, on a blind nipple connected to a transducer and computer. The longer the infant listens to the stimulus, the more habituated she becomes, reflected in a decrease in sucking rate. When the sucking rate falls below a predetermined criterion amplitude, the next high amplitude suck will result in the presentation of a new stimulus type (experimental group) or the same stimulus type (control group). If the infant is able to detect the stimulus change, an increase in sucking rate can bee seen compared to infants of the control group. This paradigm has been used to test infants' sensitivities, for instance, to rhythmic properties of language (cf., chapter 2.2).

The CHT technique has been successfully used with infants between 5.5 and 18 months, but it seems to be most suited to test infants between 6 and 10 months, because at these ages "the infant is a 'captive audience' and not easily bored with the reinforcer" (Werker, Polka & Pegg, 1997, p. 173).[34] In the first phase of an experimental session, the infant is trained to turn her head towards a loudspeaker on the detection of a stimulus change compared to the background speech or sound stream. If the detection was correct a reinforcer, usually a mechanical toy, starts to move. A second experimenter is needed in

[34] However, with modifications this method may be used with older children and even adults. These modifications include using a button which is pressed instead of the conditioned headturn, when a stimulus change is identified (Werker et al., 1997).

this kind of experiment to divert the infant from the reinforcer, so that a headturn can be uniquely interpreted as a response to the detection of a change in the auditory material. During the test phase the auditory material is changed and the recognition of the stimulus on which the infants was trained can be observed. The CHT technique "may provide a more sensitive index of infants' speech processing abilities [than the HTP]" as it is "time-locked to particular stimuli" (Gout et al., 2004, p. 555).[35] This technique was used in studies on word recognition (cf., chapter 2).

However, unlike CHT, which can only be used to test the discrimination and categorisation of relatively short speech items, like syllables, words or short melodic patterns (Werker et al., 1997, p. 176), the HTP paradigm also allows to study infants' knowledge about more complex structures in language like dependency structures (Santelmann & Jusczyk, 1998; Höhle, Schmitz, Santelmann & Weissenborn, 2006), or clausal structures, which are the main interest of this study (Hirsh-Pasek et al., 1987; Hayashi et al., 1996; Hayashi & Mazuka, 2002). In addition, the HTP has been successfully used to study infants' language acquisition development from 4 months[36] to about 20 months of age (Höhle, Weissenborn, Schmitz & Ischebeck, 2001; Höhle, 2002; Höhle et al., 2006; Santelmann & Jusczyk, 1998).

3.1.2 Direction of preference: Familiarity vs. Novelty

In studies using the HTP paradigm, the infants' preferences have been found to go in either of two directions: either the infants preferred to listen to the stimuli which are more familiar to them, the so-called *familiarity effect,* or the infants preferred to listen to the new / unknown stimuli, the so-called *novelty effect*. The familiarity effect might be expected in HTP studies which use the paradigm with the familiarisation phase. And indeed, in many studies using this design the infants showed a preference for the previously familiarised stimulus-type (e.g., Jusczyk, Cutler & Redanz, 1993; Jusczyk &

---

[35] For more detailed descriptions of the HAS and the CHT methods see, for instance, Jusczyk (1997) and Werker et al. (1998).

[36] There has to be a slight change in the experimental setting to test infants that young: the lamps which are usually placed at a 90° angle at the sides of the test booth are now placed to the left and right side of the centre lamp. The reaction of the 4-month-olds, who cannot easily turn their head, now can be identified by their eye-movement towards the blinking lamp.

Aslin, 1995; Jusczyk et al., 1999; Bortfeld, Morgan, Golinkoff & Rathbun, 2005). However, in other studies the researchers found that in this experimental setting the infants preferred to listen to the unfamiliar stimuli (e.g., Saffran, Aslin & Newport, 1996; Echols, Crowhurst & Childers, 1997; Johnson & Jusczyk, 2001; Höhle et al., 2004).

Likewise, in studies which used the HTP without the familiarisation phase the results of the experiments were going in both directions. Such opposite findings in the direction of preferences were also reported in previous studies on the recognition of clauses in fluent speech, which is of special interest for the present study. Hirsh-Pasek et al. (1987) found that American English-learning 6- to 10-month-old infants preferred to listen to the natural rather than to the unnatural speech samples. However, Hayashi et al. (1996; 2002), who used comparable materials and the same experimental design, found that Japanese-learning 8- to 11-month-old infants preferred to listen to the unnatural rather than to the natural speech samples. Furthermore, in a replication study of Hirsh-Pasek et al.'s (1987) experiment, Fernald and McRoberts (1996), though using a slightly different methodology (ibid., p. 382), also found a novelty effect in the experiment with 7-month-old infants.[37] So how can these differences be accounted for?

The model of Hunter and Ames (1988)

There are currently no models explaining the outcome of studies which use auditory stimuli and which do not include a familiarisation or habituation phase. Hunter and Ames (1988) discussed possible factors that influence the preference for familiar or novel stimuli in visual preference procedures containing a familiarisation phase. However, their model might be adapted to experiments using auditory stimuli (cf., Houston-Price & Nakai, 2004), and the explanatory power of the model for HTP experiments on the perception of clauses which do not include a familiarisation phase, will be discussed in the following sections.

[37] During the following discussion of the experiments on the recognition of clauses in fluent speech, the term 'familiarity' will be used for infants' preferences for the natural speech samples, and the term 'novelty' will be used to describe infants' preferences for the unnatural speech samples.

Hunter and Ames (1988) claimed that there are invariant stages the infant passes through during the experimental session. These are, first, an initial no-preference stage, as the information is started to be processed. The second stage is defined by a preference for the familiar stimulus. During this stage the stimulus is compared to an internal representation of the stimulus, or a mental representation of this stimulus is being built (cf., Sokolov, 1963, 1969). The third stage, which may be very brief, is again a no-preference stage in which the familiarised stimulus becomes "boring", because the comparison with or building of the mental representation is completed, and the preference starts to switch. The last stage is a novelty preference stage, in which the unfamiliar stimulus is explored. The model is dynamic with respect to the speed with which infants proceed through each stage during an experiment. This process is influenced by the interplay of three factors: familiarisation time, age of the infants, and task difficulty.

The first factor is the amount of familiarisation prior to the test phase. The model predicts that the longer the familiarisation phase, that is, the amount of time and the amount of stimulus repetition during familiarisation, the more likely is a novelty preference during testing. This conforms to the results of such different studies as the study of Jusczyk and Aslin (1995), who familiarised 7.5-month-old infants for 30 seconds with a list of words and found a familiarisation effect for these words during testing, and the study of Saffran et al. (1996), who familiarised 8-month-old infants for two minutes with a string of syllables and found a novelty effect during testing. In their experiment the infants preferred to listen not to the 'words'[38] included in the syllable string but to a combination of syllables created from different 'words' of the string. The amount of familiarity with the stimuli presented in an experiment without familiarisation phase cannot be reliably estimated. However, as the age of the infants in both Hirsh-Pasek et al.'s (1987, second experiment) and Fernald and McRoberts' (1996) studies was 7 months, the infants' pre-test experience with the native language, both being American English-learning infants, can be supposed to be similar. Therefore, the differences in the direction of the effect in these two studies cannot be explained by differences in prior familiarity with clauses in the input.

[38] A 'word' in the syllable string is defined as a combination of syllables which always co-occur, that is, having a transitional probability of 1 (ibid., p. 1927).

The second factor that Hunter and Ames (1988) discussed, namely differences in the preference depending on the age of the infants tested, also cannot be responsible for the different directions of the effect in these two studies on clause perception. Hunter and Ames (1988) claimed that the older the infants are, the faster they proceed through the different stages of the model, due to greater cognitive abilities and / or more pre-experimental knowledge. This holds if the familiarisation time and task difficulty are constant. Thus, the younger the infants are, the more likely is a familiarisation effect as the result of an experiment and the older the infants are the more likely is a novelty effect as the result of that same experiment. However, as already pointed out above, the infants tested in the studies by Hirsh-Pasek et al. (1987, second experiment) and Fernald and McRoberts (1996) were both 7 months old.

Furthermore, Fernald and McRoberts (1996) also tested 10-month-old infants, which showed a familiarity effect. In their series of experiments, the younger infants showed a novelty effect, whereas the older infants, using the same procedure and experimental materials, showed a familiarity effect. This distribution was also found in the studies by Hayashi et al. (1996; 2002). In their series of experiments the youngest infants (4-7 months) did not reveal any preference for the different stimuli, whereas the medium age group (8-11 months) showed a novelty preference, and the oldest infants (12-14 months) showed a familiarity preference. However, this distribution might be explained by the fact that infants at the end of their first year of life become more attuned to the properties of the native language and might therefore prefer to listen to stimuli that conform more to the input.

Empirical support for these two factors influencing the outcome of experiments including a familiarisation phase is presented in the study by Thiessen and colleagues (Thiessen, Hill & Saffran, 2005). In their first experiment, they familiarised 7-month-olds with sentences of a nonsense language, in which words could only be identified based on the transitional probabilities between the syllables. When tested on syllable lists which either coincided with a word (transitional probability: 1) or which were constructed of syllables of adjoining words (transitional probability: 0.25), the 7-month-olds preferred to listen to the lists of words. However, this preference was only found when the nonsense language was presented in CDS, but not when the language was presented in ADS.

In a second experiment, Thiessen et al. (2005) increased the duration of the familiarisation time and also tested older infants, namely 8-month-olds. The researchers predicted that if the CDS register facilitated the segmentation of the words from the language, the infants in this experimental condition should show a novelty effect. And indeed, the 8-month-olds which had been exposed to a longer familiarisation period preferred to listen to the lists constructed from adjoining syllables, thus showing a novelty effect. Thus, the experiments provided evidence that an increase in age and familiarisation time led to the preference for novel stimuli.

The third factor which can influence the direction of the preference in studies comes from the differences in task difficulty and / or the complexity of the stimuli. The degree of salience of two stimuli does not influence the direction of a possible effect, but the effect itself. That is, stimuli with equal salience might not be treated differently by the infants. Though they might be able to discriminate the stimuli, at the same time infants might find both stimuli equally interesting or boring (cf., Houston-Price & Nakai, 2004, p. 342), resulting in a null effect in the analysis of the OTs.

However, differences in the task difficulty and the complexity between the stimuli might lead to differences in the direction of an effect. Hunter and Ames (1988) claimed that the younger the infants and the more difficult or complex the task, the more likely is a familiarity preference. On the other end of the continuum: the older the infants and the easier the task, the more likely is the novelty preference. This concept might account for the finding of the familiarity preference in the study by Hirsh-Pasek et al. (1987). Given that the unnatural samples were perceived by the infants as more complex than the natural speech samples, the familiarity preference would reflect the infants' preference for the easier type of stimulus. But again, Fernald and McRoberts (1996) and Hayashi et al. (1996; 2002) found contrary preferences. The preference for the unnatural speech samples in these studies may be explained by differences in the salience between the stimuli. The argumentation here would be complementary to the one before. The unnatural speech samples may have been perceived as more salient by the infants because the pauses interrupt the natural flow of speech more noticeably within a clause than between clauses (cf., Fernald & McRoberts, 1996, p. 383). The novelty preference then would reflect the infants' preference for acoustically more salient stimuli. However, there seems to be no explanation which resolves the contrary findings on infants' preferences in the perception of clauses.

To sum up, the complex interaction of familiarity with the stimulus, age of the infants and task difficulty responsible for the direction of an effect in experiments using auditory stimuli is not yet fully understood. The model of Hunter and Ames (1988) does seem to be explanatory for the factor familiarisation time in HTP experiments which include a familiarisation phase (cf., Jusczyk & Aslin, 1995; Saffran et al., 1996). The age related preferences suggested in the model by Hunter and Ames (1988), due to a faster process through the suggested stages, do not seem to hold for experiments using auditory stimuli. In contrast, the studies by Fernald and McRoberts (1996) and Hayashi et al. (1996; 2002) show that the younger the infants are, the more likely it is that they show a novelty effect and the older and thus, the more attuned to the native language they are, the more likely it is that they show a familiarity effect. However, the result of the study by Hirsh-Pasek et al. (1987), who found a familiarity preference in 7-month-olds, though the procedure and the materials used were comparable to the study by Hayashi et al. (1996; 2002), defies the account of an age-related factor in the prediction of preferences in infants in these experiments. Hayashi et al. (1996, p. 1568) suggested that language-specific differences in speech processing might account for the contradictory results of the English and Japanese studies.[39] The differences in the complexity of the stimulus or differences in the salience of the stimuli also do not seem to be reliable indicators for the prediction of the infants' preferences in experiments with auditory stimuli.

In consequence, for the experiments presented in this study no prediction with respect to the possible direction of the infants' preferences will be made. Any statistically significant differences in the OTs towards the natural and unnatural speech samples will be taken as an indication that the infants discriminated between both types of stimuli, because the "direction of a looking preference is largely irrelevant […]; any deviation from random behaviour indicates that a difference between the stimuli has been detected" (Houston-Price & Nakai, 2004, p. 344).

[39] However, they did not discuss the nature of the language-specific processing differences that they think are responsible for the outcome of the studies.

The next section describes the experimental setting and the procedure of the HTP paradigm used in the present study.

3.1.3 Experimental setting and procedure used in the present study

The setup of the HTP paradigm and the procedure used in this study is as follows. During the experiment the infant sits on a caregiver's lap in the centre of a three-sided test booth. In front of the infant at eye-level a green lamp is fixed at the cardboard wall of the booth. Above this lamp there is a hole for the lens of a video camera. In a 90° angle from the infant a red lamp is fixed on each side of the booth, also at eye-level. Behind these lamps, out of sight of the infant, there are two loudspeakers mounted for the presentation of the auditory stimuli (see Figure 3.1). During the experiment, the caregiver is wearing headphones over which masking music is played, so that s/he cannot influence the spontaneous reaction of the infant. In addition, the caregivers are instructed not to point to the lamps or to otherwise interfere with the procedure. In an adjoining room, the experimenter monitors the behaviour of the infant. The presentation of the stimulus and the flashing of the lamps is computer controlled. The experimenter starts each trial of the experiment and codes the behaviour of the infant with a push-button box. She is blind (or better: deaf) to the stimulus played to the child on a given trial, so that she also cannot influence the experimental results.

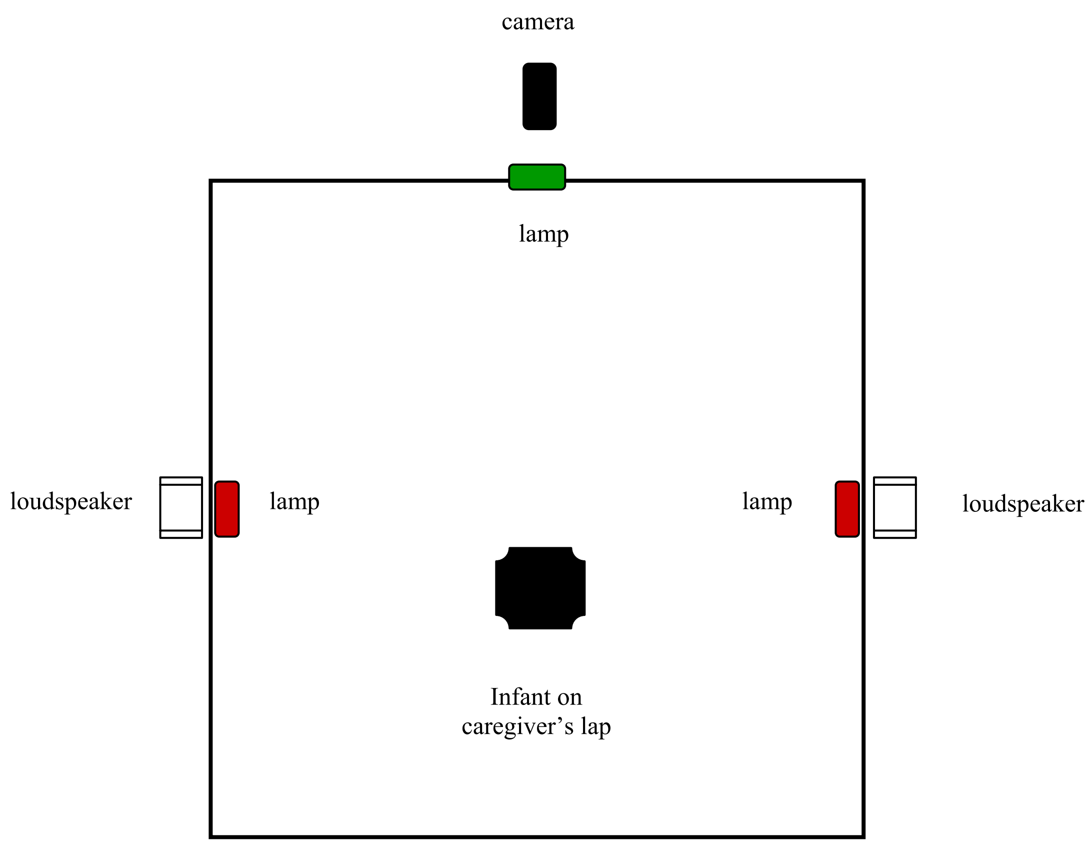

Figure 3.1: Sketch of the HTP test booth seen from above.

Each experimental trial starts with the flashing of the centre green lamp to attract the infant's attention. Then the lamp is switched off and one of the side lamps starts to flash. When the infant turns her head to the blinking lamp, the auditory stimulus is presented from the same side. The dependent variable is the time the infant spends looking in that direction (= OT). If the infant looks away for more than two seconds, the trial is stopped and the next one is started with the blinking of the centre lamp. If the infant looks away for less than two seconds, the amount of time she looks away is not included in the OT, though the auditory stimulus continues to play.

In the present study, the HTP paradigm without familiarisation phase is used, thus, the first four trials of the experiment are training trials in which the infant is acquainted with the procedure. The remaining trials, twelve in this study, are the basis of the data analyses in each experiment. Each experimental session is videotaped for offline coding.

The next chapter describes the auditory stimuli used in all five experiments of the present study. Specifics about the manipulation and presentation of the stimuli in each experiment will be presented in the respective sections in chapter 4.

3.2 *Stimuli for all experiments*

3.2.1 Sentence materials

The sentences used in the present study are a subset of the stimulus materials originally constructed for an ERP-study on the perception of intonation phrase boundaries in 9-year-old children (Leuckefeld, 2001).[40] Leuckefeld used sentences consisting of either two or three intonation phrases (IPhs) depending on the transitivity of the first embedded verb. The sentences were all of the type: NP MatrixV Name $V_{inf\ tr/intr}$ Coordinating Conjunction VP [ObjNP $V_{inf,trans}$]. An example of a sentence consisting of two IPhs, in which the first embedded verb is intransitive, is the following: *Die Tante verspricht Katrin zu schwimmen / und die Insel zu erkunden*.[41] ('The aunt promises Katrin to swim and to explore the island.'). The sentences consisting of three IPhs, in which the first embedded verb is transitive, is the following: *Die Tante verspricht / Katrin zu begleiten / und die Insel zu erkunden*. ('The aunt promises to accompany Katrin and to explore the island.').[42]

Both types of sentences showed the same acoustic characteristics for marking the IPh boundaries, that is, longer duration of the pre-boundary element, a longer pause at the boundary and a high pitch on the following nominal element, that is, a pitch reset after the boundary (Leuckefeld, 2001, p. 62ff). The experiments conducted with 9-year-old German-learning children revealed a closure positive shift (CPS) (cf., Steinhauer et al., 1999, and section 2.2.3.3) in the ERP pattern that was correlated to the occurrence of an IPh boundary. The CPS was found at each IPh boundary and seemed to be an automatic response of the auditory processing system. Thus, the results of this study provided evidence that the intonation phrase boundaries in these sentences were recognisable by the listener.

For the series of experiments presented in this study, only the sentences consisting of two intonation phrases were used to study the ability of 6-month-old German-learning infants to recognise clause boundaries in speech. This subset was chosen because the

[40] Many thanks to the MPI for Human Cognitive and Brain Sciences, Leipzig, and especially to Kerstin Leuckefeld for providing the original auditory material.

[41] The ' / ' indicates the position of an IPh boundary.

[42] These and all subsequent translations, which are more or less literal translations to illustrate the syntactic structure of the experimental stimuli, are my own.

intonation phrase boundary coincided with a syntactic clause boundary. The auditory material consisted of 48 co-ordinated sentences. 24 of the sentences were of the type: *Die Tante verspricht Katrin zu schwimmen und die Insel zu erkunden*. ('The aunt promises Katrin to swim and to explore the island'; sentence type 1), the other 24 sentences were of the type: *Der Lehrer erlaubt Judith zu laufen und das Brot zu kaufen*. ('The teacher allows Judith to run and to buy the (loaf of) bread'; sentence type 2). Each of the sentences contained the same number of words, but they differed in syntactic structure. In sentence type1, the unexpressed agent in the infinitival clause is controlled by the subject of the matrix sentence, thus having the syntactic structure 'The aunt$_i$ promises Katrin [[PRO$_i$ to swim] and [PRO$_i$ to explore the island]]'. The verbs used in these 24 sentences are: *versprechen* ('to promise'), *drohen* ('to threaten') and *schwören* ('to vow'), which are subject control verbs in German. In sentence type 2, the unexpressed agent of the infinitival clauses is controlled by the object of the matrix clause, thus having the syntactic structure 'The teacher allows Judith$_j$ [[PRO$_j$ to run] and [PRO$_j$ to buy the (loaf of) bread]]'. The verbs used in these 24 sentences are: *verbieten* ('to forbid'), *erlauben* ('to allow') and *helfen* ('to help'), which are object control verbs in German (cf. e.g., Haegeman, 1994). For a complete list of the sentences see Appendix A.

All sentences were produced in isolation by a trained female speaker of German. The speaker took care to pronounce all sentences as similar as possible to each other with regard to speed and prosodic contour. Though being instructed to speak in a "neutral" manner, the speaker was aware that the sentences were to be used in an experiment with children. The subsequent prosodic analysis revealed that the sentences had a significantly slower speaking rate but only a slightly greater pitch range than comparable adult-directed sentences of the same speaker (Leuckefeld, 2001, p. 64f). The mean fundamental frequency of the sentences ranged from 105 to 280 Hz. These values are more compatible with findings on adult-directed speech of women in German. In their analysis of child- and adult-directed speech Fernald et al. (1989) found a pitch range for adult-directed speech from 160 to 282 Hz, whereas the child-directed speech of women had a range from 178 to 367 Hz. Thus, the sentences used in this study display some but not all characteristics of child-directed speech.

3.2.2 Acoustic analyses of the stimuli

The duration of the original sentences used in this study ranged from 4346 ms to 5358 ms (mean duration: 4800 ms). There was a significant difference with regard to the sentence type: mean duration of the sentence type 1 was 4691 ms, whereas the mean duration of the sentence type 2 was longer, namely 4910 ms ($t_{(46)} = 3.534$; $p = 0.001$).

The mean fundamental frequency (F0) of each word was measured. Figure 3.2 shows the typical intonation contour of the two sentence types based on the F0 values averaged across sentences.

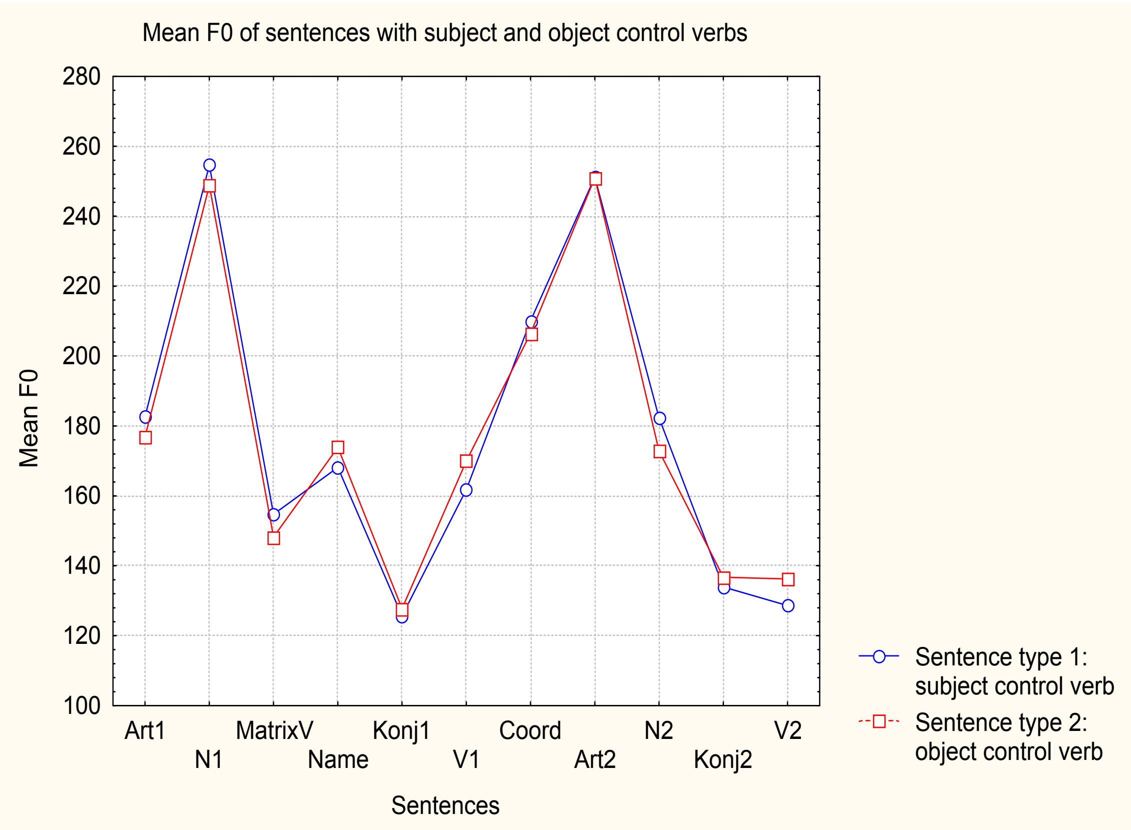

Figure 3.2: Plot of the mean fundamental frequency of the test sentences.

The graph shows that the intonation contour is nearly identical for both sentence types, tough they differ in syntactic structure and interpretation. Therefore, no distinction between the sentence types was made in the data analysis of the experiments in this study.

Also, in the analyses presented in the following paragraphs the data for both sentence types will be collapsed.

The mean duration of the first clause was 2585 ms (range: 2284 ms to 2963 ms) and the mean duration of the second clause was 1700 ms (range: 1428 ms to 2033 ms). The pause between the two clauses in the original sentences had a mean duration of 523 ms (range: 368 ms to 750 ms; cf., Appendix A). As clause and sentence boundaries are not only demarcated by pauses but are also characterised by prosodic features like lengthening of the pre-boundary element and rising or falling F0-contours (Cooper & Paccia-Cooper, 1980), the edges of the clauses have been examined with respect to their prosodic characteristics.

Boundary tones, which can be either high (H%[43]) or low (L%), mark the end of an IPh. They are realised on the syllables of the last word of the IPh (cf., Féry, 1993, p. 51). To evaluate whether the clauses used as stimuli are marked by IPh boundary tones, the fundamental frequency of the last syllables of the verbs at the end of the clauses was measured. For example, in the sentence *Der Doktor verspricht Daniel zu flüstern und das Fieber zu messen* ('The doctor promises Daniel to whisper and to take the (his) temperature.'), the F0 values of the accented syllable *flüs* and the coda *tern* as well as the F0 values of the accented syllable *mes* and the coda *sen* were measured. The measurement of the F0 values revealed that the prosodic contour of the first embedded verb shows a L-H% boundary, and the contour of the second embedded verb shows a H-L% boundary (cf., Féry, 1993; Müller, 1998). As L-H tones are associated with ongoing intonation, the prosodic contour at the end of the first IPh signals that the utterance is going on. The L% boundary of the second intonation phrase signals the end of the utterance. Thus, in addition to the pause between the clauses, the IPhs are also reliably marked by boundary tones.

The data within the sentences allowed for no reliable analysis of a syllables' pre-boundary lengthening compared to an identical syllable at a no-boundary location. So, no analysis of the supposed pre-boundary lengthening was carried out. However, this is not a major drawback for the present study as both the analyses of the two other major prosodic

[43] The annotation follows the GToBI (German Tone and Break Indices) labelling system (cf., Müller, 1998, p. 10ff)

characteristics of clause and sentence boundaries (IPhs) showed that the IPhs are reliably marked: the first clause boundary in these sentences is characterised by high boundary tone and the following pause. The second clause boundary has a low boundary tone, indicating the end of the utterance (cf., Figure 3.3).

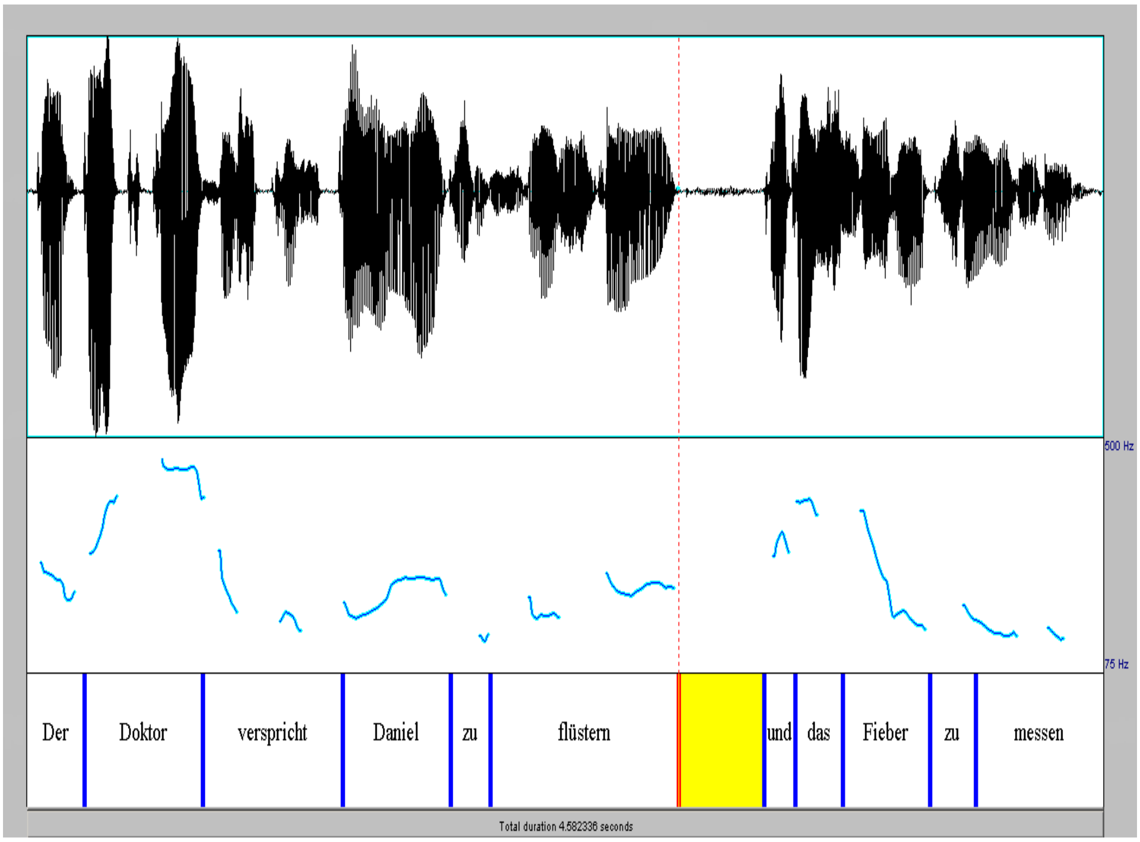

Figure 3.3: Screenshot of the waveform and pitch contour of the sentence 'The doctor promises Daniel to whisper and to take the (his) temperature'. The naturally occurring pause is marked in yellow.

Thus, as the analyses have shown and as also can be seen in the pitch contour depicted in Figure 3.3, the IPhs boundaries of the utterances are reliably marked and this might allow infants to recognise them in the speech stream. The next section will give an outline of the design of the present study.

3.3 Outline of the present study

In the first three experiments of this study the question was addressed whether, and under which pause duration conditions, 6-month-old German-learning infants recognise clause boundaries in fluent speech. The sentences were manipulated according to the material of the study of Hirsh-Pasek et al. (1987). Artificial pauses were inserted either between the two intonation phrases, that is, at a clause boundary (natural condition), or within the first intonation phrase, that is, within a syntactic clause (unnatural condition). The specifics concerning the duration of the inserted pauses will be given in the descriptions of each experiment in chapter 4. The artificial pause at the clause boundary was inserted before the coordinating conjunction *und* ('and'), resulting in: *Die Tante verspricht Katrin zu schwimmen ___ und die Insel zu erkunden*.[44] ('The aunt promises Katrin to swim ___ and to explore the island.'). The sentences for the unnatural condition were generated by inserting a pause after the infinitive particle *zu* ('to') within the first clause, resulting in the structure: *Die Tante verspricht Katrin zu ___ schwimmen und die Insel zu erkunden*. ('The aunt promises Katrin to ___ swim and to explore the island'). In this condition the pause is inserted at a location which also did not coincide with a syntactic / prosodic phrase boundary. The HTP design requires the sentences to be presented as sample passages usually consisting of six sentences rather than in isolation. Therefore additional pauses had to be inserted between subsequent sentences. The duration of these pauses is also specified in the description of each experiment in chapter 4.

In Experiments 4 and 5, 6- and 8-month-old German-learning infants were tested on their sensitivity to the natural pause hierarchy (see chapter 2.1 and Figure 2.1). For that, the artificial pauses were inserted only at the clause boundaries and between subsequent sentences. In the natural pause condition, the pauses between the co-ordinated clauses were shorter than the pauses between the sentences. To generate an unnatural version of the sentences, the pause durations were reversed, that is, the pause duration between the co-ordinated clauses was longer than the pause between the sentences (reversed pause condition).

[44] The use of '___' indicates the position of the artificially inserted pause in the examples.

Given the conflicting results of the study by Hirsh-Pasek et al. (1987) that infants preferred to listen to the natural samples during the experiment and the results of the experiments by Fernald and McRoberts (1996) and Hayashi et al. (1996) that infants preferred to listen to the unnatural samples during the experiment, the hypotheses in this study are two-tailed. This means, that if the infants' mean OTs towards the natural and unnatural samples show any statistical difference, despite the direction of the effect, this result will be interpreted as an indication that infants indeed discriminated these samples.

In the next chapter of this dissertation the five experiments conducted for the present study are presented and the results will be discussed.

4 The Study

4.1 Experiment 1

4.1.1 Participants

The participants of the first experiment were 30 6-month-old German infants (15 girls and 15 boys; mean age: 6 months and 7 days; with a range from 6 months 0 days to 6 months 24 days). All infants were from monolingual German-speaking families from the Potsdam area and were developing normally according to a questionnaire filled in by the parents at their visit to the lab. This questionnaire collected data on whether the children suffered from previous ear infections (none of them had), and whether the child was developing normally according to the last medical check by their paediatrician (all of them were). Fifteen additional infants were tested, but their data could not be included in the analysis due to the following reasons: problems with the technical equipment (2), infants did not finish the experimental session or were too fussy (10), or having a mean orientation time of less than 3 seconds (s) per condition (3).[45]

4.1.2 Stimuli

The stimuli for the experiment were generated from the set of 48 sentences described in the previous section. A complete list of all sentences used in the study is given in Appendix A. These sentences were used in all the experiments of this study (Experiments 1 – 5).

The stimuli were digitized (44.1 KHz) and edited for the presentation in the experiment. In the first experiment a pause with the duration of 1 s was inserted into the sentences (cf., Hirsh-Pasek et al., 1987). Two conditions were generated depending on the position of the pause. In the natural condition, the pause was inserted at the clause boundary before the coordinating conjunction *und* (‘and’). To create an unnatural version

[45] This exclusion criterion was pre-defined on the basis of the duration of the first clause in the original material. The durations of the first clause ranged from 2284 ms to 2963 ms (mean: 2585 ms). Thus, infants with a mean orientation time of less than 3 seconds in one condition were considered not to have heard enough of the sentences to recognise the critical difference between the natural and unnatural samples at all.

of each sentence, a pause was inserted between the infinitive particle *zu* ('to') and the verb of the first clause, a position which did not coincide with a phrase boundary.

In the natural condition, the naturally occurring pause at the clause boundary, that is, before the coordinating conjunction *und* ('and') was deleted and replaced by an artificially generated silent pause of 1 s duration. In the original sentence material, the naturally occurring pauses at the clause boundary had a mean duration of 523 ms, with a range from 368 ms to 750 ms.

In the unnatural condition a pause was inserted between the infinitive particle *zu* ('to') and the verb in the first clause. In the original sentences there was not always a measurable pause between the two words. In only 19 of the 48 sentences a pause occurred. The duration of these pauses was between 0.02 and 0.09 s. This indicates that usually no pause is inserted between the two words in naturally spoken sentences. In addition, the naturally occurring pause before *und* ('and') was shortened to 100 ms in the sentences in this condition. This additional manipulation was made to avoid too much variability in the duration of the sentences between the natural and the unnatural condition (cf., Hirsh-Pasek et al., 1987, second experiment; and Gerken et al., 1994, footnote 5). All other naturally occurring pause locations were the same in both conditions and therefore remained intact.

To test the possible range of a shortened pause duration that would not interfere with the overall sentence intonation, a pre-test with adults was conducted. Two sets of test sentences were prepared. The first test-set had the original pause duration at the clause boundary shortened to 200 ms before *und* ('and'). The second test-set had a shortened pause of 100 ms duration at the boundary. To verify whether the shorter duration did not impair the perception of the prosody of the whole sentence, a rating was conducted. Five adult listeners, all native speakers of German and naïve with regard to the purpose of the experiment, listened to the two sentence sets. The sentences were presented pairwise in randomised order. The listeners indicated on a sheet of paper whether the first or the second presentation of each sentence sounded more natural (two-alternative forced-choice decision task). The analysis of these ratings showed that the five adult listeners performed at chance level. They had a 50.3% preference for the test sentences with a 100 ms pause duration compared to a 49.7% preference for sentences with a 200 ms pause duration at the clause boundary. Based on the result of this pre-test, the naturally occurring pauses at the clause boundary were shortened to 100 ms in the sentences of the unnatural condition in the infant experiment. Other features, like final lengthening or pitch in the surroundings of the artificially inserted pauses were not manipulated.

As the original material consisted of isolated sentences as opposed to natural conversation, pauses also had to be inserted between the six subsequent sentences constituting a test passage in the experiment. Research on pause structures in natural speech indicated that pauses between sentences are longer than pauses within sentences (e.g., Butcher, 1981, for German). To preserve this natural pause hierarchy, a pause of 1.5 s duration was inserted between the sentences. The resulting auditory stimuli for the infant experiment consisted of 16 blocks of six sentences each. Half of the blocks (natural condition) had an artificial pause of 1 s duration inserted at the clause boundary and the other half (unnatural condition) had the pause inserted at a non-phrase boundary location within the first clause of the sentence, resulting in stimulus blocks like the following example, which is the first block of the training trials in the experiment (for the organisation of the 48 sentences into text blocks see Appendix B):

<u>Natural condition:</u>

Der Bäcker verspricht Anja zu lächeln ___$_{1\ s}$ *und die Sache zu vergessen.* ___$_{1.5\ s}$
Der Cousin droht Markus zu verschwinden ___$_{1\ s}$ *und das Fahrrad zu stehlen.* ___$_{1.5\ s}$
Die Mutter schwört Holger zu überlegen ___$_{1\ s}$ *und das Gedicht zu schreiben.* ___$_{1.5\ s}$
Der Onkel droht Sonja zu lärmen ___$_{1\ s}$ *und das Bild zu zerreißen.* ___$_{1.5\ s}$
Die Köchin verspricht Peter zu gehen ___$_{1\ s}$ *und den Einkauf zu erledigen.* ___$_{1.5\ s}$
Der Bruder schwört Ellen zu schleichen ___$_{1\ s}$ *und die Angst zu überwinden.*

‘The baker promises Anja to smile ___$_{1\ s}$ and to forget about it.’ ___$_{1.5\ s}$
‘The cousin threatens Marcus to disappear ___$_{1\ s}$ and to steal the bicycle.’ ___$_{1.5\ s}$
‘The mother vows (to) Holger to think ___$_{1\ s}$ and to write the poem.’ ___$_{1.5\ s}$
‘The uncle threatens Sonja to be noisy ____$_{1\ s}$ and to tear apart the picture.’ ___$_{1.5\ s}$
‘The cook promises Peter to go ___$_{1\ s}$ and to do the shopping.’ ___$_{1.5\ s}$
‘The brother vows (to) Ellen to tiptoe ___$_{1\ s}$ and to overcome the fear.’

Unnatural condition:

Der Bäcker verspricht Anja zu ___$_{\text{1 s}}$ *lächeln und die Sache zu vergessen.* ___$_{\text{1.5 s}}$
Der Cousin droht Markus zu ___$_{\text{1 s}}$ *verschwinden und das Fahrrad zu stehlen.* ___$_{\text{1.5 s}}$
Die Mutter schwört Holger zu ___$_{\text{1 s}}$ *überlegen und das Gedicht zu schreiben.* ___$_{\text{1.5 s}}$
Der Onkel droht Sonja zu ___$_{\text{1 s}}$ *lärmen und das Bild zu zerreißen.* ___$_{\text{1.5 s}}$
Die Köchin verspricht Peter zu ___$_{\text{1 s}}$ *gehen und den Einkauf zu erledigen.* ___$_{\text{1.5 s}}$
Der Bruder schwört Ellen zu ___$_{\text{1 s}}$ *schleichen und die Angst zu überwinden.*

‘The baker promises Anja to ___$_{\text{1 s}}$ smile and to forget about it.’ ___$_{\text{1.5 s}}$
‘The cousin threatens Marcus to ___$_{\text{1 s}}$ disappear and to steal the bicycle.’ ___$_{\text{1.5 s}}$
‘The mother vows (to) Holger to ___$_{\text{1 s}}$ think and to write the poem.’ ___$_{\text{1.5 s}}$
‘The uncle threatens Sonja to be ___$_{\text{1 s}}$ noisy and to tear apart the picture.’ ___$_{\text{1.5 s}}$
‘The cook promises Peter to ___$_{\text{1 s}}$ go and to do the shopping.’ ___$_{\text{1.5 s}}$
‘The brother vows (to) Ellen to ___$_{\text{1 s}}$ tiptoe and to overcome the fear.’

The mean duration of the sentences in the natural condition was 5.28 s (range: 4.90 s to 5.70 s) and the mean duration of a sentence in the unnatural condition was 5.35 s (range: 4.93 s to 5.80 s; see Appendix C).

The mean duration of the blocks for the natural condition was 39.17 s (ranging from 38.60 s to 40.43 s) and the mean duration of the blocks for the unnatural condition was 39.60 s (with a range from 39.05 s to 40.75 s; see Appendix D).

4.1.3 Method and Procedure

The method used in this and all subsequent conducted experiments was the Headturn Preference (HTP) paradigm without the familiarisation phase (see chapter 3 for a detailed description). This paradigm is designed to test the infants’ knowledge of their language at a certain level. For instance, the present experiment investigated whether infants are sensitive to the correlation between the occurrence of a pause and a syntactic boundary which coincides with an intonation phrase boundary.

For the experiment four different randomisations of block presentation were created. In the randomisations the natural and unnatural condition samples were not presented in strictly alternating trials. There was no contingency between the type of stimulus and the side of presentation. Two of the randomisations started with a block of the natural condition (natural first), the other two randomisations started with a block of the unnatural condition (unnatural first). There were the same number of boys and girls in each of these two randomisation groups. The first four trials in each experiment, two natural samples and the corresponding two unnatural samples, were training trials to familiarise the infant with the method itself and with the kind of stimuli presented. The orientation time data of these four trials were not included in the analysis. During the immediately following test phase, the remaining twelve passages were presented according to the randomisation.

4.1.4 Results

The data file for each infant was recoded by a second experienced experimenter[46] via video-analysis and only this recoded data will be presented here. For recoding, the second experimenter viewed the videotape and checked for differences in the infants' behaviour and the first experimenters' coding, such as missed looking-away events. The original result file was then corrected on the basis of the recoded data and used for the analysis. Corrections were made in 11 of the 30 datasets (36.7%). Comparable to other experiments using the HTP paradigm the inter-coder reliability was very high: $r = 0.99$ (Pearson, cf., Kemler Nelson et al., 1995).

The mean orientation times towards the natural and the unnatural passages were calculated for each infant. The mean orientation time towards the natural samples was 9049 ms (SD = 5413 ms) and the orientation time towards the unnatural samples was 9410 ms (SD = 4677 ms). This difference was not significant $t_{(29)} = 0.72$; $p = 0.47$ (paired t-test). Eighteen of the 30 infants had longer OTs towards the unnatural samples.

[46] Many thanks to Anne Rosenthal and Katrin Skoruppa who helped in the offline coding of the experiments in the present study.

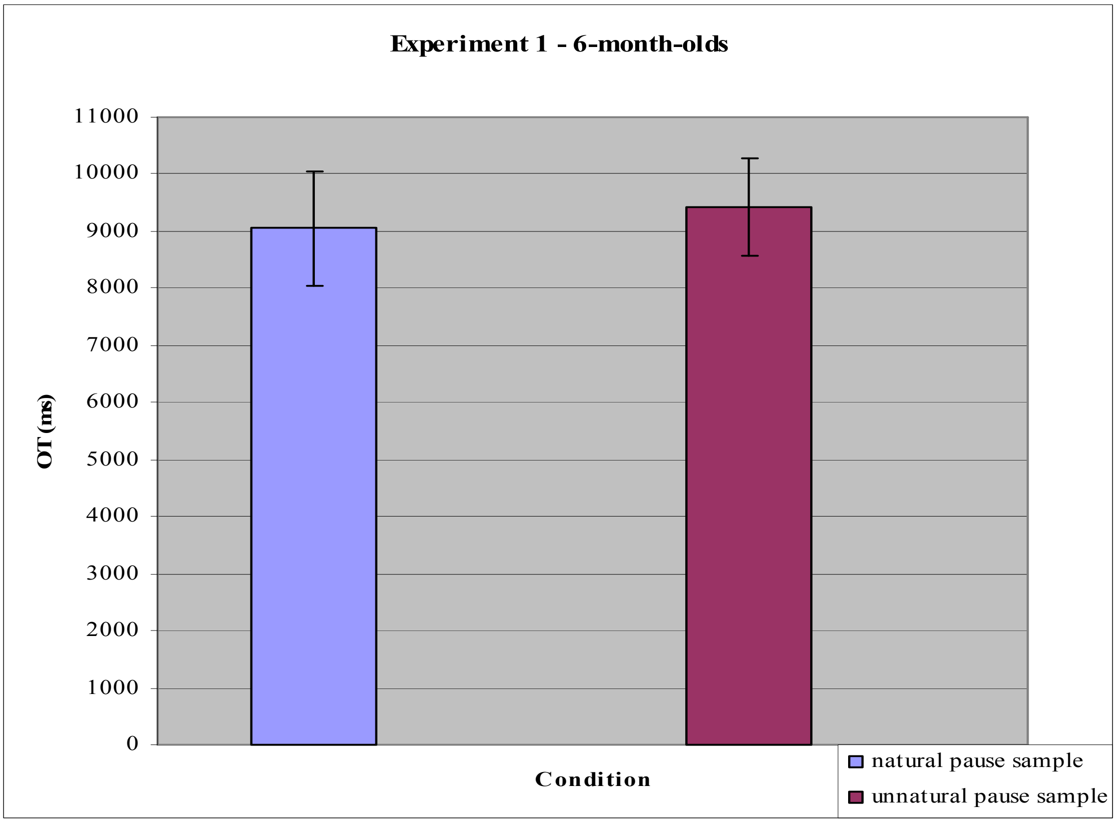

Figure 4.1: Experiment 1 - Results of 6-month-olds' discrimination of sentences containing pauses at a natural or unnatural position.

To evaluate whether the order of presentation or the sex of the infants influenced the result, additional statistical analyses were conducted. First, an ANOVA was conducted with the between-subjects factor Order of Presentation (natural first vs. unnatural first) and the within-subjects factor Condition (natural vs. unnatural). There was no significant main effect for order of presentation ($F_{(1,28)}$ = 0.85; $p < 1$) and no significant interaction ($F_{(1,28)}$ = 0.212; $p < 1$). A second ANOVA was conducted with Sex of the infant as the between-subjects variable and Condition as the within-subjects factor. There was also no main effect for sex of the infant ($F_{(1,28)}$ = 0.01; $p < 1$) and no significant interaction ($F_{(1,28)}$ = 0.35; $p < 1$).

In a third step, it was analysed whether the result of this first experiment was due to a change in the infants' preferences over time during the experiment. Figure 4.2 shows the mean orientation time for each condition and each presentation, including the pre-test training trials.

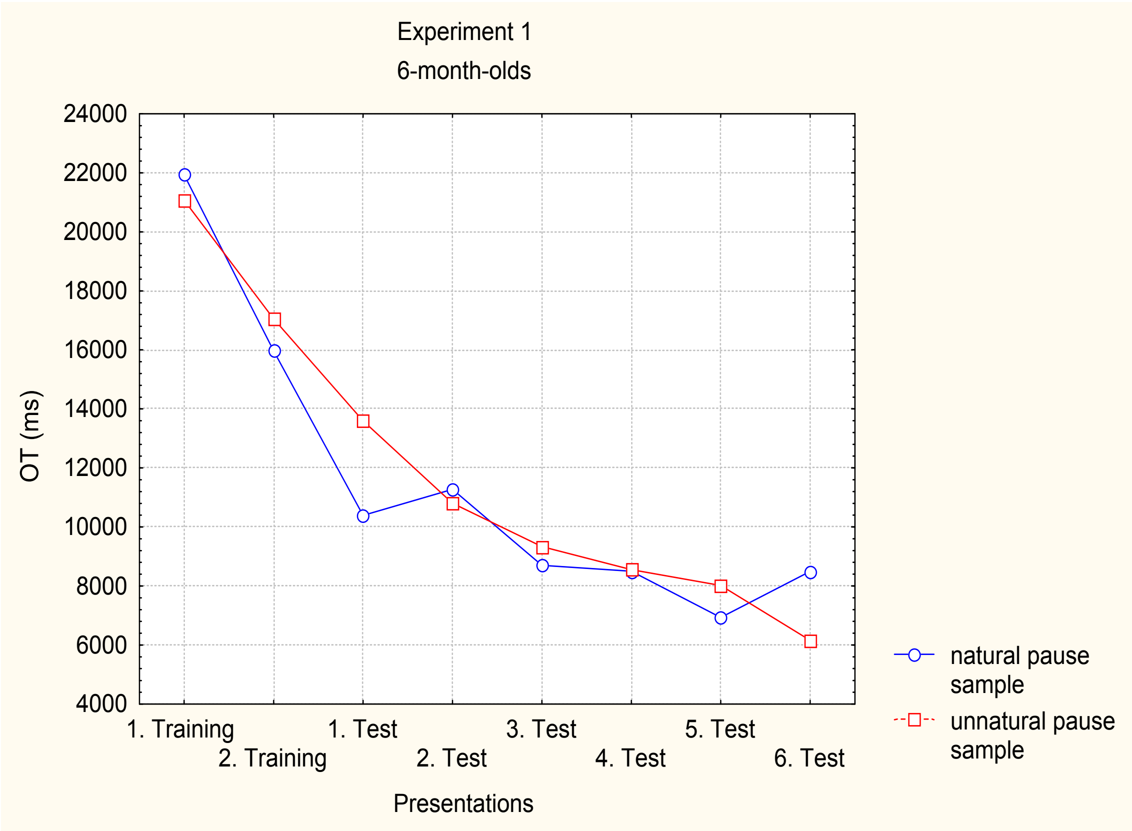

Figure 4.2: Experiment 1 - Mean orientation time for each condition over the presentations during the experiment.

The graph shows that there is a constant decrease in the mean orientation times for both conditions over the experiment, except in the last test trial. The first training trial shows a non-significant longer mean orientation time for the natural speech sample ($t_{(29)} = 0.30$; $p = 0.76$ (paired t-test)). Therefore, this cannot be interpreted as an initial (pre-test) preference for the natural stimuli. During some of the other presentations a longer mean orientation time towards the unnatural samples can be seen, or the orientation time towards the conditions is similar. Only in the last test trial the mean orientation time for the natural sample is longer than for the unnatural sample. However, this difference is only marginally significant ($t_{(29)} = 1.90$; $p = 0.07$ (paired t-test)). These analyses show that the null result of this first experiment is not due to a turning in the infants' preferences during the experiment. Furthermore, the similar orientation times towards both conditions in some of the presentations suggest that the infants did not seem to discriminate between both conditions, or at least, they did not prefer to listen to one of the conditions.

4.1.5 Discussion

The German-learning 6-month-old infants gave no indication of preferring either of the two conditions. In contrast to the findings of Hirsh-Pasek et al. (1987), no preference for the sentences in the natural condition could be observed, instead the German infants had non-significant longer mean orientation times for the unnatural condition. The result of this first experiment is striking: German 6-month-olds do not react differently to the presentation of sentences with inserted pauses in the natural and the unnatural condition, whereas studies in other languages with the same research question (Hirsh-Pasek et al., 1987; Hayashi et al., 1996; Hayashi & Mazuka, 2002) found that both American and Japanese infants of comparable age showed the ability to discriminate between the natural and unnatural condition.

There are two main differences in the auditory material, that is, differences in the register and differences in the kind of discourse, which could be responsible for the different results in the studies.

Use of register as an explanation of the result

The stimuli used by Hirsh-Pasek et al. (1987) were constructed using American English child-directed speech (CDS), whereas the stimuli of this experiment display only some characteristics of child-directed speech, like slower rate of speech, and some characteristics of adult-directed speech (ADS), like the F0 range (see chapter 3.2). Studies have documented that CDS has different characteristics than ADS, like shorter utterances, exaggerated pitch contours and more pronounced final lengthening of vowels at syntactic boundaries (cf., Fernald et al., 1989). Furthermore, the duration and appearance of pauses at syntactic boundaries is more reliable in CDS than in ADS (e.g., Fernald & Simon, 1984; Fisher & Tokura, 1996a; Morgan et al., 1987). It was also shown that infants are sensitive to CDS (e.g., Fernald, 1985; Cooper & Aslin, 1990) and use this register as a bootstrap into speech segmentation (Thiessen et al., 2005; for an overview see also Soderstrom, 2007).

Nevertheless, there are indications that the absence of a clearly child-directed register in this experiment may not be the reason for the infants' different behaviour in the present study compared to the original Hirsh-Pasek et al. (1987) study. Some of the features infants have been shown to be sensitive to are present in the stimuli used in this experiment. First, the analyses of the stimuli revealed that the syntactic boundaries, that is, the clause as well as the sentence boundaries are clearly marked by pitch change. Second, the insertion of the artificial pause led to the constant co-occurrence of pauses and syntactic boundaries, at least in the natural condition. What is not present in the auditory stimuli used in this experiment is the exaggerated pitch range that was found in CDS (cf., Fernald et al., 1989). However, recent studies have shown that infants at the age of 7 months do not distinguish between speech samples presented in CDS and speech samples presented in ADS (Hayashi, Tamekawa & Kiritani, 2001; Hayashi & Mazuka, 2002). These studies suggest that the sensitivity to child-directed vs. adult-directed speech may follow a U-shaped development, at least for the recognition of clauses in fluent speech: A preference for CDS in the first 6 months of life, a relative insensitivity to the differences in speech register at about 7-9 months of age and again a growing preference for CDS at about 10 months of age. A comparison of German and Japanese reveals that both languages are similar in most acoustic features in adult- as well as in child-directed speech, whereas American English CDS has much more exaggerated acoustic features (cf., appendix of Fernald et al., 1989). Therefore, infants at about the age tested in this study might not rely on the prosodic characteristics marking IPhs in CDS, like the more pronounced variations in fundamental frequency, to infer a clause boundary in fluent speech.

Recent studies on the preference of CDS over ADS in American English support the hypothesis that in the second half of their first year infants do not necessarily prefer CDS. Newman and Hussain (2006) found that only by 4.5 months the infants revealed a preference for speech passages presented in CDS over the same passages presented in ADS. By the age of 9- and 13-months no preference was found. Thus, Newman and Hussain (2006) were not able to find a comparable U-shaped preference curve for CDS that had been reported for the Japanese-learning infants. Furthermore, Zangl and Mills (2007) found in their ERP study that 6-month-olds had an increased brain activity (N600)[47] only to familiar words presented in CDS as opposed to familiar words presented in ADS. Non-familiar words did not elicit a greater brain activity even when presented in CDS. The positive influence of the CDS register on the segmentation of words was also found in the study by Thiessen et al. (2005). They found that 7- and 8-month-olds were able to segment words from a nonsense language on the basis of statistical cues when the language was familiarised in the CDS register, but not when the language was familiarised in the ADS register.

Therefore, the use infants make of the cues provided by the CDS around the middle of their first year of life seems to depend on the task that is to be solved. The processing of words seems to be facilitated by the prosodic cues of CDS whereas in the recognition of clauses CDS does not seem to be that relevant.[48] Thus, one can hypothesise that the absence of an explicit child-directed register in this study was not the reason for the 6-month-olds' indifferent reaction to the natural and the unnatural condition. However, as there is no comparable longitudinal data of a possible preference of CDS over ADS, and its use in the processing of speech in German-learning infants, the question whether the use of an explicit CDS register might have altered the results of this first experiment remains open.

[47] The N600 is a negativity in the ERP pattern, peaking around 600 ms to 800/900 ms after the onset of the target word. It is related to continued attention for familiar words in the speech stream.

[48] However, see the study of Kemler Nelson et al. (1989) on the recognition of clause boundaries by American English-learning infants. They found that the infants recognised the clause boundaries only in CDS but not in ADS speech samples. As has been mentioned before, American-English CDS shows more exaggerated prosodic characteristics than Japanese and German and therefore infants might benefit more from the use of this register in English.

Use of discourse type as an explanation of the result

There is a second difference between the material used in the present study and that of previous studies: the auditory stimuli used by Hirsh-Pasek et al. (1987) consisted of naturally produced text excerpts, in which a mother read a fairytale to her daughter. In contrast, the auditory stimuli used in this experiment consisted of 48 sentences produced in isolation. To create the natural speech samples for their experiment, Hirsh-Pasek et al. (1987) inserted the artificial pauses at several different boundaries in the texts, so that sometimes the artificial pause coincided with a clause boundary which was also a sentence boundary, thus being even more acoustically salient. Furthermore, the excerpts containing the artificial pauses in the middle of clauses, thus rendering artificial speech samples, also started and stopped in the middle of a clause. The presentation of such prosodically interrupted speech samples may also have influenced the infants in their listening preferences, that is, longer orientation times towards the natural samples. This interpretation is supported by recent findings of Nazzi and colleagues (2000b) and Seidl (2007). In both HTP-studies 6-month-old infants were familiarised with either a well-formed prosodic sentence or a sequence consisting of two sentence fragments. In the test phase, the infants significantly preferred to listen to the well-formed prosodic sentences.

In contrast, the stimuli used in this experiment consisted only of well-formed isolated sentences which were composed of two syntactic clauses each coinciding with an intonation phrase (cf., Nespor & Vogel, 1986). So there was a decision to be made concerning the insertion not only of the artificial pauses within the sentences, but in the absence of discourse prosody with naturally occurring pauses between the sentences, a second pause had to be inserted at the sentence boundaries as well. In this first experiment, following the natural pause hierarchy in which pauses between sentences usually have a longer duration than pauses within sentences, it was decided to insert a pause of one and a half seconds between subsequent sentences of a given block.

Taking into account the findings by Butcher (1981; cf., Goldman-Eisler, 1972) a possible reason for the result pattern found in the first experiment may be that, although corresponding to the natural pause hierarchy, the pauses were simply too long. Butcher (1981, p. 84ff) reported that in German pauses within sentences, that is, between clauses were about 600 ms to 850 ms long and pauses between sentences were about 800 ms to 1.2 s long. As the pauses within the sentence were 1 s long and the pauses between sentences had a duration of 1.5 s, infants might not have registered that the sentence was going on with the second clause, both in the case of the natural and the unnatural versions, but might just have rated the interruption of the pause higher than the other acoustic cues indicating an ongoing sentence.

This hypothesis is also supported by findings from Cheour and colleagues (Cheour, Ceponiené, Leppänen, Alho, Kujala, Renlund, Fellman & Näätänen, 2002). In their ERP study they tested the auditory memory of newborns. In an oddball design the infants were presented with standard stimulus pure tones of 1000 Hz and a deviant stimulus of 1100 Hz. They were able to elicit a mismatch negativity (MMN), indicating the detection of the stimulus change, in the newborns only in the condition in which the interstimulus interval (ISI) between the standard and the deviant was 700 ms long, but not in a condition in which this interval was 1.4 s long. Adults, and even 8-year-olds, detect the stimulus change even after an ISI of 10 s. Thus, the time span for the auditory memory in online processing of sounds, and consequently also speech sounds is considerably shorter in infants compared to older children. Even though Cheour et al. (2002) investigated the auditory memory span in newborns, one can presume that by six months the memory span may be a bit but not considerably longer.

To test the hypothesis that infants are able to discriminate between natural and unnatural passages, a second experiment was conducted, using the same sentence material but shortening the artificially inserted pause to a duration much more compatible with natural speech.

4.2 Experiment 2

4.2.1 Participants

In this experiment a new group of 30 infants (15 boys, 15 girls) was tested. The same selection criteria as in the previous study applied. The mean age of the infants was 6 months and 9 days (range: 6 months 1 day to 6 months 24 days). Twelve infants were tested additionally but their data could not be included in the analysis due to the following reasons: problems with the technical equipment (5), infants did not finish the experimental session or were too fussy (5), or having a mean orientation time of less than 3 s per condition (2).

4.2.2 Stimuli

The same sentences were used as in the previous experiment. To test a version which does reflect the natural speech more closely, the inserted pauses within the sentences, both in the natural and in the unnatural condition were shortened to 750 ms. The natural pause hierarchy was maintained with a pause duration of 1 s between the subsequent sentences of a given block. These pause durations were chosen on the basis of the mean pause duration correlated to these boundary types (cf., Figure 2.1).

<u>Natural condition:</u>

Der Bäcker verspricht Anja zu lächeln $\_\_\__{750\ ms}$ *und die Sache zu vergessen.* $\_\_\__{1\ s}$
Der Cousin droht Markus zu verschwinden $\_\_\__{750\ ms}$ *und das Fahrrad zu stehlen.* $\_\_\__{1\ s}$
.....

'The baker promises Anja to smile $\_\_\__{750\ ms}$ and to forget about it.' $\_\_\__{1\ s}$
'The cousin threatens Marcus to disappear $\_\_\__{750\ ms}$ and to steal the bicycle.' $\_\_\__{1\ s}$
…..

Unnatural condition:

Der Bäcker verspricht Anja zu ___$_{750\ ms}$ lächeln und die Sache zu vergessen. ___$_{1\ s}$
Der Cousin droht Markus zu ___$_{750\ ms}$ verschwinden und das Fahrrad zu stehlen. ___$_{1\ s}$
…..

'The baker promises Anja to ___$_{750\ ms}$ smile and to forget about it.' ___$_{1\ s}$
'The cousin threatens Marcus to ___$_{750\ ms}$ disappear and to steal the bicycle.' ___$_{1\ s}$
…..

The mean duration of the sentences in the natural condition was 5.03 s (range: 4.65 s to 5.46 s) and the mean duration of a sentence in the unnatural version was 5.10 s (range: 4.68 s to 5.56 s; see Appendix C).

The mean duration of the blocks for the natural condition was 35.17 s (ranging from 34.60 s to 36.43 s) and the mean duration of the blocks for the unnatural condition was 35.60 s (with a range from 35.05 s to 36.75 s; see Appendix D)

4.2.3 Method and Procedure

The method and procedure were identical to the one in the previous experiment.

4.2.4 Results

For each child the data were prepared for the analyses as described in the previous experiment. Also the same exclusion criterion applied. Though corrections were made in 17 of the 30 datasets (56.7%), the inter-coder reliability was r = 0.99 (Pearson). This suggests that there were only minor differences between the online and the offline coding.

The mean orientation times towards the natural and unnatural samples were calculated. In the experimental version with the overall shortened pauses, the mean orientation time towards the natural samples was 6711 ms (SD = 2310 ms). The mean orientation time towards the unnatural samples was 8546 ms (SD = 4399 ms). This difference was significant: $t_{(29)} = 2.97$; $p = 0.005$ (paired t-test). Twenty-one of the 30 infants had longer OTs towards the unnatural samples.

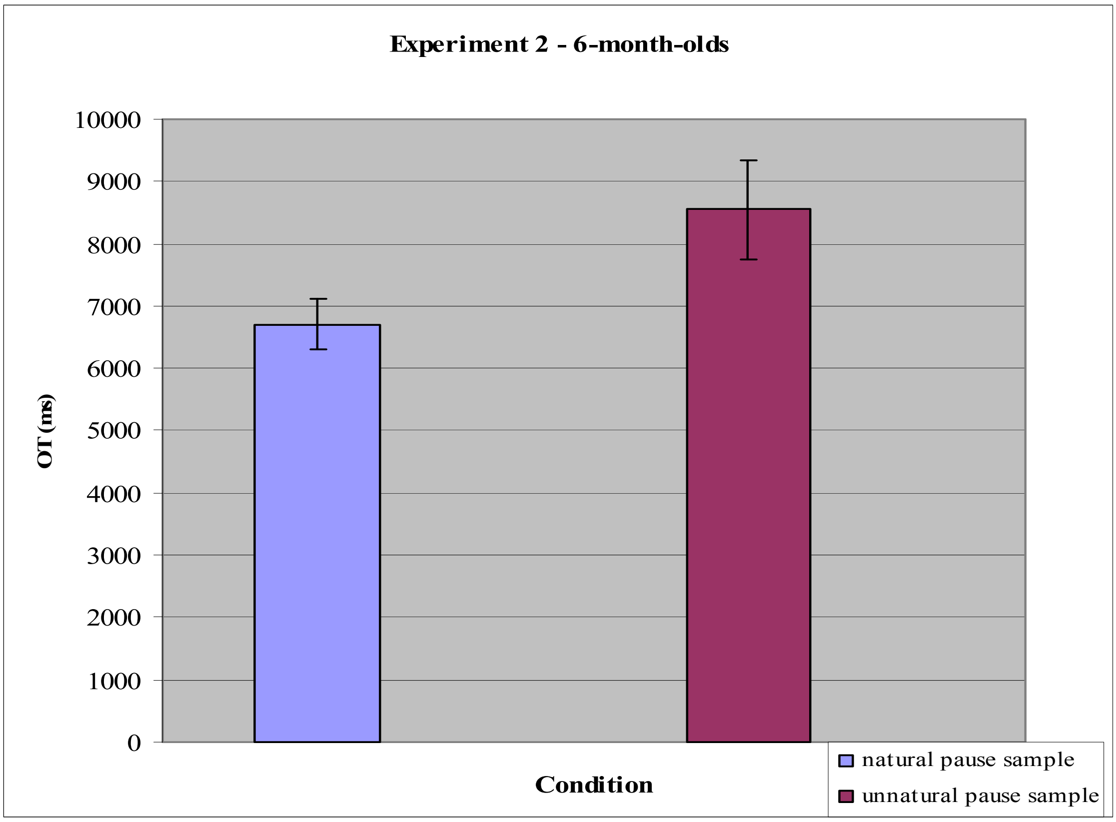

Figure 4.3: Experiment 2 - Results of 6-month-olds' discrimination of sentences with overall shortened pause durations.

To evaluate whether this result was influenced by the order of presentation (natural first vs. unnatural first) or the sex of the infant, again two statistical analyses using ANOVA were conducted. In the analysis with the between-subjects factor Order of Presentation (natural first vs. unnatural first) and the within-subjects factor Condition (natural vs. unnatural) there was no significant main effect for order of presentation ($F_{(1,28)} = 0.17$; $p < 1$) and no significant interaction ($F_{(1,28)} = 0.68$; $p < 1$). In the analysis with Sex of the infant as the between-subjects variable and Condition as the within-subjects factor there was also no main effect for sex of the infant ($F_{(1,28)} = 0.36$; $p < 1$) and no significant interaction ($F_{(1,28)} = 1.75$; $p < 1$).

In a third step a possible change in the infants' preferences over time was analysed. In Figure 4.4 the mean orientation time for each condition and each presentation, including the pre-test training trials is plotted.

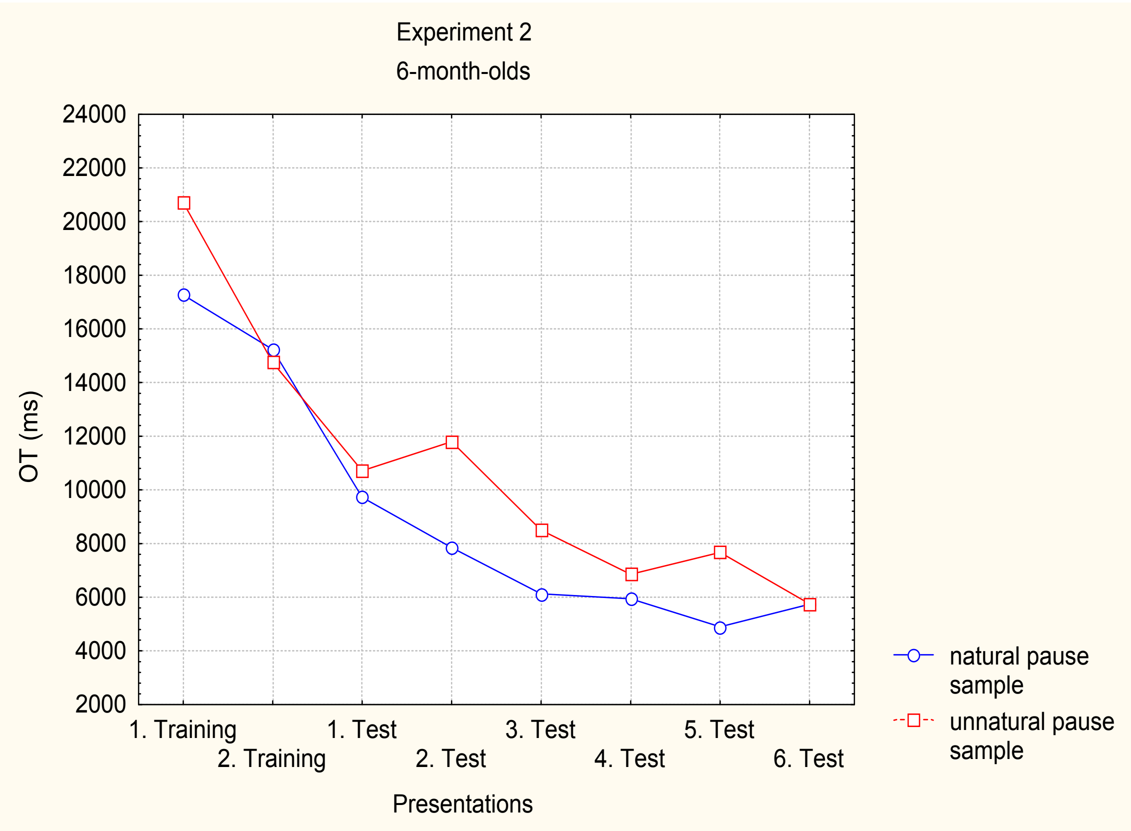

Figure 4.4: Experiment 2 - Mean orientation time for each condition over the presentations during the experiment.

The graph shows a general decrease in the infants' mean orientation times to both conditions during the experiment. Furthermore, it can be seen that the mean orientation time towards the unnatural samples is longer in most of the presentations, only in the second training trial and the last test trial the orientation times are nearly identical. The difference in the orientation time on the first training trial is non-significant: $t(29) = 1.33$; $p = 0.20$ (paired t-test). The constant longer mean orientation times towards the unnatural samples during the test phase of the experiment lead to the significant orientation time differences reported above. Thus, the infants start with a longer orientation time towards the unnatural sample during the training presentation and this preference is maintained during the test presentations. This suggests that the infants indeed discriminated between both conditions.

4.2.5 Discussion

In this experiment using sentences in which the inserted pauses were more closely reflecting natural speech, the German 6-month-olds are able to discriminate between the natural and the unnatural condition. The infants had longer mean orientation times towards the unnatural condition than towards the natural condition. These results on German complement the findings of studies on other languages that infants are able to recognise clauses and clause boundaries in fluent speech (Hirsh-Pasek et al., 1987; Hayashi & Mazuka, 2002; Hayashi et al., 1996; Nazzi et al., 2000b; Seidl, 2007). The emergence of this ability, at least in German, seems to be closely related to the pattern and duration of speech pauses in the experimental material.

This supports the hypothesis that the infants' inability to show a discrimination effect between the natural and the unnatural condition in the first experiment was caused by the overall longer duration of the inserted pauses. Since no other acoustic feature was manipulated, the result of the second experiment can be attributed to the shortened duration of the artificially inserted pauses. Furthermore, the hypothesis that the absence of an explicit child-directed register was not responsible for the 6-month-olds failure to discriminate the natural from the unnatural condition in the first experiment was corroborated.

The infants recognise the clause boundaries in fluent speech when the pause structure in the samples of the experimental conditions corresponds to the duration of naturally occurring speech pauses as reflected in the natural pause hierarchy (Butcher, 1981; Goldman-Eisler, 1972). Thus, for the German-learning infants pause seems to be a necessary cue for the recognition of clause boundaries in fluent speech.

An explanation for the novelty preference

The result pattern of the orientation times in this experiment is contrary to the one found in the original study by Hirsh-Pasek and colleagues (1987). In their study the 6- and 7-month-old American English-learning infants had longer mean orientation times to the natural passages. In the present experiment the infants have longer mean orientation times towards the unnatural samples than to the natural ones. This effect can be interpreted as a novelty effect, in which the infants prefer to listen to the more unusual / unknown acoustic stimulus. This reaction pattern has been observed in other HTP-studies as well (e.g., Höhle et al., 2004).

Furthermore, several studies, in which the original study by Hirsh-Pasek et al. (1987) was replicated or extended to other languages, also found this kind of novelty effect. In their replication of the original experiment, Fernald and McRoberts (1996) found that infants did not react differently to the natural and unnatural samples. In a subsequent study, using a different methodology, Fernald and McRoberts (1996) tested 4-, 7-, and 10-month-old American English-learning infants. In this study they found that the 4-month-olds did not react differently to the natural and the unnatural samples. At the age of 7 months, however, the infants preferred to listen to the unnatural samples, and only at the age of 10 months a preference for the natural samples was found. Hayashi and colleagues (2002) found a similar pattern. In their study 4- to 7-month-olds did not discriminate between the natural and the unnatural samples, whereas 8- to 11-month-old Japanese-learning infants were able to recognise clause boundaries in fluent speech, but the infants had longer mean orientation times to the unnatural condition than to the natural condition. Only the oldest age group from 12 months on showed a preference for the natural speech samples. Thus, in these two studies the age of the children seemed to be the decisive factor for the direction of the preference displayed by the infants. The results of these two studies contradict the predictions made by the model of Hunter and Ames (1988), who suggest that with increasing age a novelty preference is to be expected. One can speculate that at the age of about 12 months, when infants start to produce their first words, they prefer the natural speech model in order to improve their own communication strategies.

Taken together, the results of the experiments on clause perception are heterogeneous with respect to the direction of preference: American English-learning 6- and 7-month-olds displayed a familiarity reaction in the study by Hirsh-Pasek et al. (1987), whereas the 7-month-olds in the study by Fernald and McRoberts (1996) and the 8-month-olds in the study by Hayashi et al. (2002) showed a preference for the unnatural condition.

In the present study a preference for the unnatural condition was observed in 6-month-old German-learning infants, using the same methodology as the previous studies but different stimulus materials. Allowing for inter-individual and language-specific differences, this result conforms to the findings by Fernald and McRoberts (1996) and Hayashi et al. (2002). As discussed in section 3.1.2, the prediction of the direction of an effect is difficult, but regarding the stimuli in this experiment there might be a possible explanation for this novelty effect. On the whole, the isolated sentences used as stimuli in the present study all had the same structure and were pronounced in the same way (cf., the discussion in section 4.1.5). Therefore the stimuli of the present study are more homogeneous, lacking, for instance, discourse prosody, than the stimuli of the previous studies, that consisted of passages of child-directed speech. To create the natural samples, the artificial pause was inserted at the clause boundaries so that all acoustic cues marking the boundary coincided. Furthermore, the inserted pauses had a duration which more corresponds to the natural speech environment. To create the unnatural samples, the artificial pause was inserted after the acoustically short infinitive particle *zu* ('to') and thus interrupts the speech flow more noticeably than the pause inserted at the clause boundary. Furthermore, the pause after the infinitive particle contradicts other acoustic cues in the speech, that is, there are no other cues marking a boundary in that position.

The differences in the acoustic properties between the, perhaps more monotonous, natural and the unnatural samples may have caused the longer orientation times towards the unnatural samples. The model of Hunter and Ames (1988) would in this case predict a short phase of preference for the natural samples followed by a faster turning of the infants' attention to the unnatural samples. However, as can bee seen in Figure 4.4, the infants showed longer mean orientation times towards the unnatural samples in almost all experimental trials and no turning point could be discerned. This could be interpreted as a real novelty effect driven by the acoustic idiosyncrasies of the unnatural speech samples.

However, the main question in the first experiments of this study, as well as in other studies, concerned the recognition of clauses by the infants. Thus, the direction of the effect per se is not decisive for the interpretation of the results (two-tailed hypothesis).

The results of the first two experiments of the present study provide evidence that German-learning infants are able to recognise the prosodic correlates of clause boundaries in fluent speech when the duration of the inserted pauses closely reflects natural speech. In the first experiment, the infants did not prefer either of the conditions. It was hypothesised that this was due to the duration of the inserted pauses which were longer than those found in German speech corpora (cf., Butcher, 1981). The results of the second experiment support this hypothesis. As no other acoustic properties of the speech samples were manipulated the infants' ability to discriminate both conditions can only be attributed to the shorter pause durations. Thus, German-learning 6-month-olds seem to use the pause as a cue in their segmentation of clauses from fluent speech.

However, the stimuli used in the previous two experiments were manipulated in one further point: the naturally occurring pause at the clause boundary had been shortened to 100 ms in the unnatural versions of the sentences to ensure greater comparability of the sentences in both conditions. Therefore, the sentences in the unnatural condition deviated in two ways from the sentences in the natural condition: position of the inserted pause and a very short pause at the clause boundary. This also might have influenced the direction of the preference found in the second experiment. To test whether the infants show the same discrimination abilities when the originally occurring pause before the clause boundary remained intact, a third experiment was conducted.

4.3 Experiment 3

4.3.1 Participants

A new group of 30 infants from monolingual German-speaking families of the Potsdam area participated in this experiment (20 girls and 10 boys). The mean age of the infants was 6 months and 7 days, ranging from 6 months 0 days to 6 months and 20 days. The same criteria for participating in the study applied. Another eight infants were tested but their data could not be included in the analysis due to the following reasons: problems with the technical equipment (2), infants did not finish the experimental session or were too fussy (5), or parent intervention during testing (1).

4.3.2 Stimuli

The same sentences and pause durations as in the second experiment were used. Additionally, the naturally occurring pause at the clause boundary in the original sentences was kept in the unnatural condition. As described above, the naturally occurring pause before the coordinating conjunction *und* (‘and’) had a mean duration of 523 ms, with a range from 368 ms to 750 ms. Thus the sentences in the unnatural condition in this experiment had an additional pause of considerable duration.

<u>Natural condition:</u>

Der Bäcker verspricht Anja zu lächeln ___$_{750\ ms}$ *und die Sache zu vergessen.* ___$_{1\ s}$
Der Cousin droht Markus zu verschwinden ___$_{750\ ms}$ *und das Fahrrad zu stehlen.* ___$_{1\ s}$
.....

‘The baker promises Anja to smile ___$_{750\ ms}$ and to forget about it.’ ___$_{1\ s}$
‘The cousin threatens Marcus to disappear ___$_{750\ ms}$ and to steal the bicycle.’ ___$_{1\ s}$
…..

Unnatural condition:

Der Bäcker verspricht Anja zu ___$_{750\ ms}$ *lächeln* ___$_{PO}$ *und die Sache zu vergessen.* ___$_{1\ s}$
Der Cousin droht Markus zu ___$_{750\ ms}$ *verschwinden* ___$_{PO}$ *und das Fahrrad zu stehlen.* ___$_{1\ s}$[49]
…..

'The baker promises Anja to ___$_{750\ ms}$ smile ___ $_{PO}$ and to forget about it.' ___$_{1\ s}$
'The cousin threatens Marcus to ___$_{750\ ms}$ disappear ___ $_{PO}$ and to steal the bicycle.' ___$_{1\ s}$
…..

The mean duration of the sentences in the natural condition was 5.03 s (range: 4.65 s to 5.46 s). The mean duration of the sentences in the unnatural condition was 5.55 s (range: 5.10 s to 6.11. s; see Appendix C).

The mean duration of the blocks for the natural condition was 35.17 s (ranging from 34.60 s to 36.43 s) and the mean duration of the blocks for the unnatural condition had a mean duration of 38.34 s (range: 37.59 s to 39.53 s; see Appendix D).

4.3.3 Method and Procedure

The method and procedure were identical to the one in the previous experiments.

4.3.4 Results

For each child the data were recoded by a second experimenter as described in the first experiment. Also the same exclusion criterion applied. Though corrections were made in 17 of the 30 datasets (56.7.0%), the inter-coder reliability was r = 0.99 (Pearson). This suggests that there were only minor differences between the online and the offline coding. The mean orientation time towards the natural condition was 7751 ms (SD = 3619 ms) and the mean orientation time towards the unnatural condition was 8320 ms (SD = 3868 ms).

[49] Note: PO corresponds to the pause originally occurring at the clause boundary (368 ms – 750 ms).

This difference was not significant: $t_{(29)}$ = 0.93; p = 0.36 (paired t-test). Sixteen of the 30 infants had longer mean OTs towards the unnatural samples.

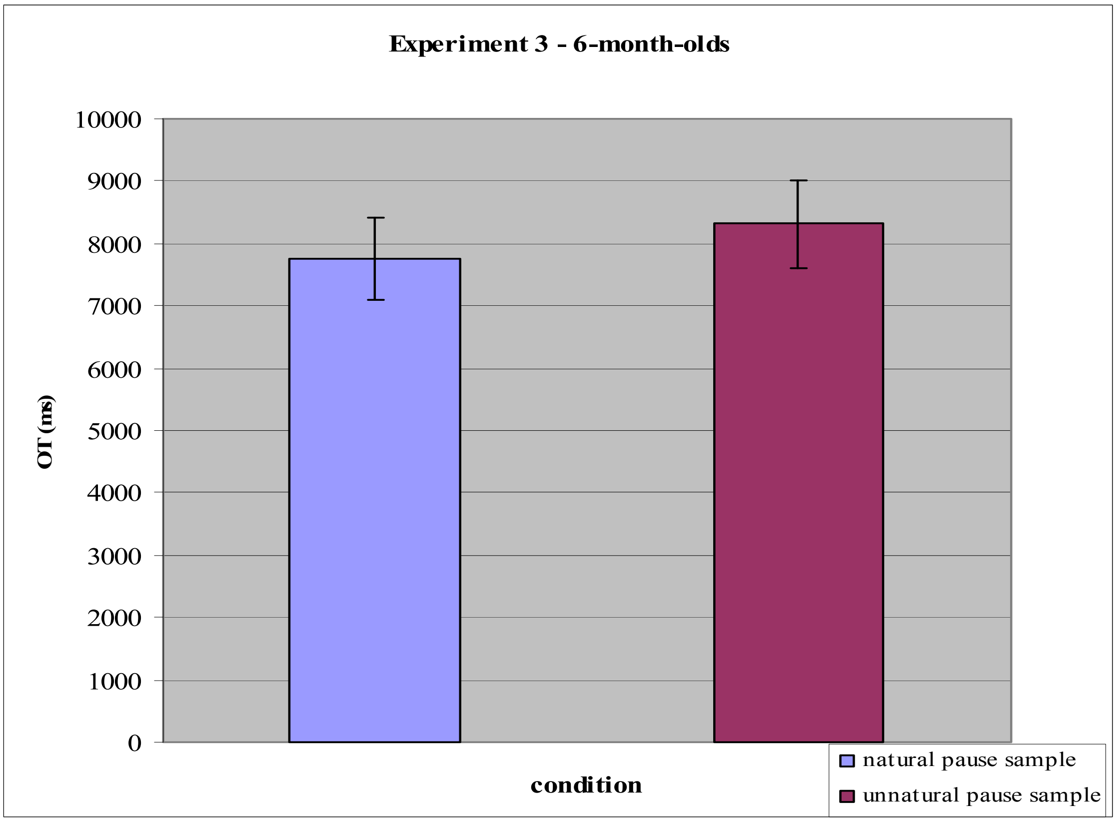

Figure 4.5: Experiment 3 - Results of 6-month-olds' discrimination of sentences with shortened unnatural pauses and original pause durations before the coordinating conjunction *und ('and')*.

To evaluate whether in this experiment the order of presentation or the sex of the infants influenced the result, again additional statistical analyses were conducted. First, an ANOVA was conducted with the between-subjects factor Order of Presentation (natural first vs. unnatural first) and the within-subjects factor Condition (natural vs. unnatural). There was no significant main effect for order of presentation ($F_{(1,28)}$ = 0.00; $p < 1$) and no significant interaction ($F_{(1,28)}$ = 0.30; $p < 1$). A second ANOVA was conducted with Sex of the infant as the between-subjects variable and Condition as the within-subjects factor. There was also no main effect for sex of the infant ($F_{(1,28)}$ = 0.04; $p < 1$) and no significant interaction ($F_{(1,28)}$ = 0.00; $p < 1$).

Furthermore, it was analysed whether this null result was due to a change in the infants' preferences over time during the experiment. Figure 4.6 shows the mean orientation time for each condition and each presentation, including the pre-test training trials.

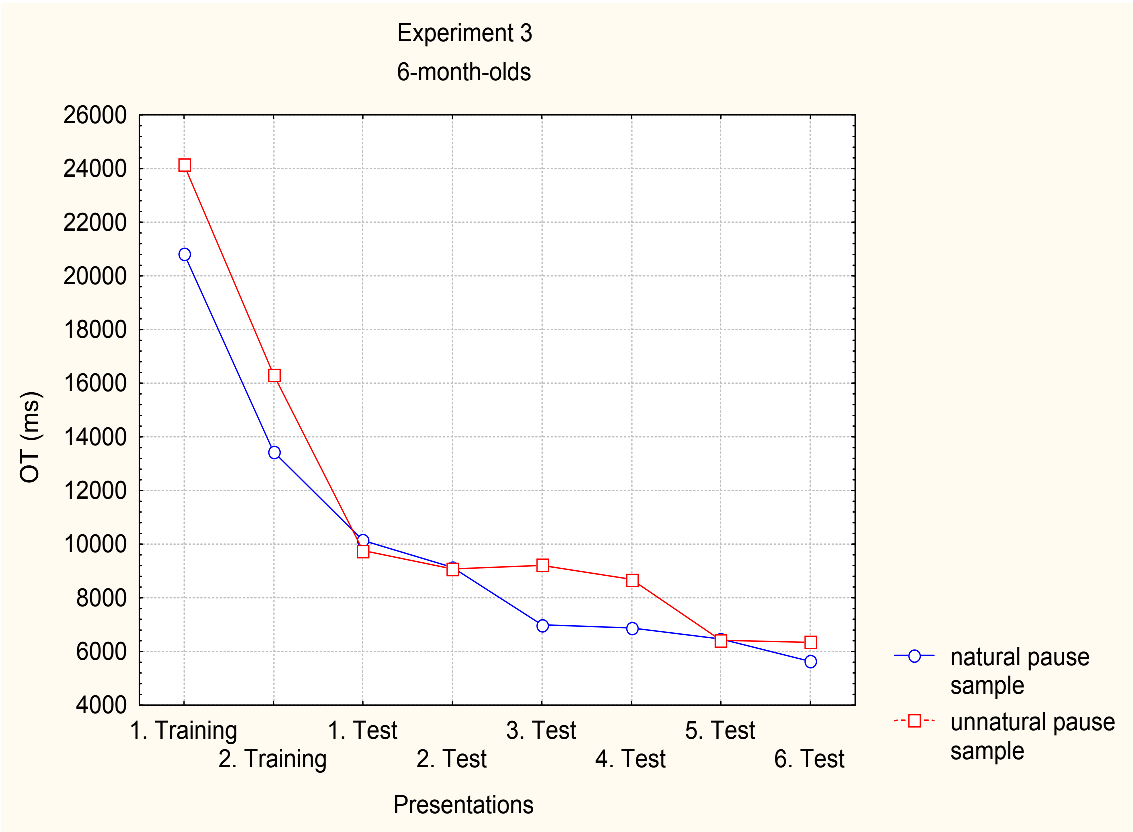

Figure 4.6: Experiment 3 - Mean orientation time for each condition over the presentations during the experiment.

The graph shows a pattern comparable to that of the first two experiments, that is, a general decrease in the mean orientation times for both conditions. Furthermore, as in the second experiment, a longer mean orientation time for the unnatural condition was found in most of the presentations. However, none of the differences in the orientation times between the natural and unnatural condition reached significance level (first training trial: $t_{(29)} = 1.09$; $p = 0.29$; second training trial: $t_{(29)} = 1.23$; $p = 0.23$; third test trial: $t_{(29)} = 1.50$; $p = 0.14$; fourth test trial: $t_{(29)} = 1.16$; p = 0.26 (paired t-tests)). No shift in the infants' preferences over the course of the experiment can be seen. The similar mean orientation times for both conditions on three of the six test presentations and the non-significant differences in the orientation times for the other three test presentations seem to indicate that the infants did not discriminate between the natural and the unnatural condition, or at least that they did not prefer either of the two conditions.

4.3.5 Discussion

In the present experiment, the German-learning infants again did not react differently to the two conditions. This result is contrary to the findings in Hirsh-Pasek et al.'s (1987) first experiment. In their experiment the infants preferred to listen to the natural samples even when the original pauses were preserved. Hirsh-Pasek et al. (1987) assumed that this familiarity effect was due to the differences in coherence in the samples. They argued that in the unnatural samples the original pauses plus the additionally inserted artificial pauses were disturbing the infants' perception and caused them to prefer the more coherent natural samples. In the experiment with the German infants this result was not replicated. Neither did the infants prefer the natural samples nor did they show a discrimination of the two conditions at all.

However, it is possible that exactly the more frequent pauses in the unnatural condition may have influenced the infants' reaction in the present experiment. The explanation for the diverging result pattern may be found in the overall direction of the effect the infants showed in the experiments. The American English-learning infants in the original study always showed a preference for the natural samples. Taking into account the results of all three experiments conducted in this study, it can be seen that in each experiment the German-learning infants tended to listen longer to the unnatural condition than to the natural condition, though the preference in the first and third experiment did not reach a significance level. Only in the second experiment, in which the pause durations more closely reflected the natural speech conditions, the infants had significantly longer mean OTs to the unnatural speech samples. This finding was validated by the comparison of the mean orientation times towards both conditions over the course of the experiment. In all three experiments the infants showed a trend to listen longer towards the unnatural condition in most of the presentations. It was argued above that the null result in the first experiment was due to the overall too long durations of the inserted pauses. Likewise, it seems possible that the result in this experiment might be due to the fact that a significant preference for the unnatural samples was impeded by the frequent interruptions in the stimuli, created by the insertion of the artificial pauses and the occurrence of a long natural pause before the coordinating conjunction.

To sum up, the results of these first three experiments provide evidence that German-learning 6-month-olds are able to recognise clause boundaries in fluent speech. They discriminated natural from unnatural speech samples in a condition in which pause durations reflect the pause durations found in natural German speech corpora. The nature of the stimuli used in the experiments made it necessary to insert two different types of pauses, shorter pauses between clauses and longer pauses between subsequent sentences, in an experimental text block. All previous studies used naturally spoken text passages as the basis of the stimuli and only one type of pause with a fixed duration was inserted within those passages, irrespective of whether the boundary was a within-sentence clause boundary or a between-sentence boundary. Therefore, none of the previous studies investigating infants' recognition of clause boundaries in fluent speech addressed the question of whether infants as young as 6 months already are sensitive to the pause hierarchy found in natural speech (cf., Figure 2.1).

In the present study, the three experiments reported so far all presupposed that infants are sensitive to the natural pause hierarchy in speech. The results of the first and the second experiment provide indirect evidence that infants are indeed sensitive to the natural pause hierarchy. Now, the question will be addressed whether infants do not only make use of pauses as a means to identify clause boundaries in fluent speech, but that they also are sensitive to the fact that there are correlations between pause durations and the type of boundary in the speech stream.

Based on the results of the second experiment, the following experiments are designed to explicitly test the hypothesis that German-learning infants are sensitive to the natural pause hierarchy. The experiments are designed to test the infants on the discrimination between speech samples either corresponding to the natural pause hierarchy (natural pause condition) and samples in which the pauses were reversed (reversed pause condition). Two cross-sectional experiments with 6-month-old and 8-month-old German-learning infants were conducted.

4.4 *Experiment 4*

4.4.1 Participants

A new group of 30 6-month-old German-learning infants (15 girls, 15 boys) participated in this experiment (mean age: 6 months 8 days, range: 5 months 30 days to 6 months 23 days). The same selection criteria as in the previous experiments applied. Twelve additional infants were tested, but their data was not included in the analysis due to the following reasons: infants did not finish the experimental session or were too fussy (11), or having a mean orientation time of less than 3 s per condition (1).

4.4.2 Stimuli

To test the hypothesis that infants are sensitive to the natural pause hierarchy, a new set of stimuli was generated from the original sentences. Artificial pauses within the sentences in both conditions were inserted only at the clause boundary before the coordinating conjunction *und* ('and'). All other naturally occurring pauses in the sentences remained intact. That is, in contrast to the previous experiments, in this experiment there is no longer a condition in which the pause is inserted at an unnatural position within the sentence. For the natural pause condition the manipulated sentences of the second experiment were used, that is, the pauses within the sentences at the clause boundary had a duration of 750 ms and the pauses between the subsequent sentences of a sample had a duration of 1 s. To create the stimuli for the reversed pause condition the duration of the pauses was inverted, that is, the artificial pauses at the clause boundary before the coordinating conjunction *und* ('and') had a duration of 1 s and the pauses between subsequent sentences in a sample had a duration of 750 ms.

Natural pause condition:

Der Bäcker verspricht Anja zu lächeln ___$_{750\ ms}$ *und die Sache zu vergessen.* ___$_{1\ s}$
Der Cousin droht Markus zu verschwinden ___$_{750\ ms}$ *und das Fahrrad zu stehlen.* ___$_{1\ s}$
…..

'The baker promises Anja to smile ___$_{750\ ms}$ and to forget about it.' ___$_{1\ s}$
'The cousin threatens Marcus to disappear ___$_{750\ ms}$ and to steal the bicycle.' ___$_{1\ s}$
…..

Reversed pause condition:

Der Bäcker verspricht Anja zu ___$_{1\ s}$ *lächeln und die Sache zu vergessen.* ___$_{750\ ms}$
Der Cousin droht Markus zu ___$_{1\ s}$ *verschwinden und das Fahrrad zu stehlen.* ___$_{750\ ms}$
…..

'The baker promises Anja to smile ___$_{1\ s}$ and to forget about it.' ___$_{750\ ms}$
'The cousin threatens Marcus to disappear ___$_{1\ s}$ and to steal the bicycle.' ___$_{750\ ms}$
…..

The mean duration of the sentences in the natural pause condition was 5.03 s (range: 4.65 s to 5.46 s) and the mean duration of a sentence in the reversed pause condition was 5.28 s (range: 4.90 s to 5.75 s; see Appendix C).

The mean duration of the blocks for the natural condition was 35.17 s (ranging from 34.60 s to 36.43 s) and the mean duration of for blocks for the reversed pause condition was 35.42 s (ranging from 34.85 s to 36.68 s; see Appendix D).

4.4.3 Method and Procedure

The method and procedure were identical to the one in the previous experiments. Two of the four randomisations for the presentation started with a block of the natural pause condition (natural pause first), the other two randomisations started with a block of the reversed pause condition (reversed pause first).

4.4.4 Results

For each child the data were recoded by a second experimenter as described in the first experiment. Also the same exclusion criterion applied. Corrections were made in nine of the 30 datasets (30.0%), the inter-coder reliability was $r = 0.99$ (Pearson).

The mean orientation time towards the natural pause condition was 8848 ms (SD = 3075) and the mean orientation time towards the reversed pause condition was 8393 ms (SD = 3277). This difference was not significant: $t_{(29)} = 0.91$; $p = 0.37$ (paired t-test). Fifteen of the 30 infants had longer OTs towards the natural pause condition.

Figure 4.7: Experiment 4 - Results of 6-month-olds' discrimination of sentences with natural and reversed pause durations.

To evaluate whether this result was influenced by the order of presentation (natural pause first vs. reversed pause first) or the sex of the infant, again two statistical analyses using ANOVA were conducted. In the analysis with the between-subjects factor Order of

Presentation (natural pause first vs. reversed pause first) and the within-subjects factor Condition (natural pause vs. reversed pause) there was no significant main effect for order of presentation ($F_{(1,28)} = 1.94$; $p < 1$) and no significant interaction ($F_{(1,28)} = 1.47$; $p < 1$). In the analysis with Sex of the infant as the between-subjects variable and Condition as the within-subjects factor there was also no main effect for sex of the infant ($F_{(1,28)} = 0.83$; $p < 1$) and no significant interaction ($F_{(1,28)} = 0.53$; $p < 1$).

Furthermore, it was analysed whether this result was due to a change in the infants' preferences over time during the experiment. Figure 4.8 shows the mean orientation time for each condition and each presentation, including the pre-test training trials.

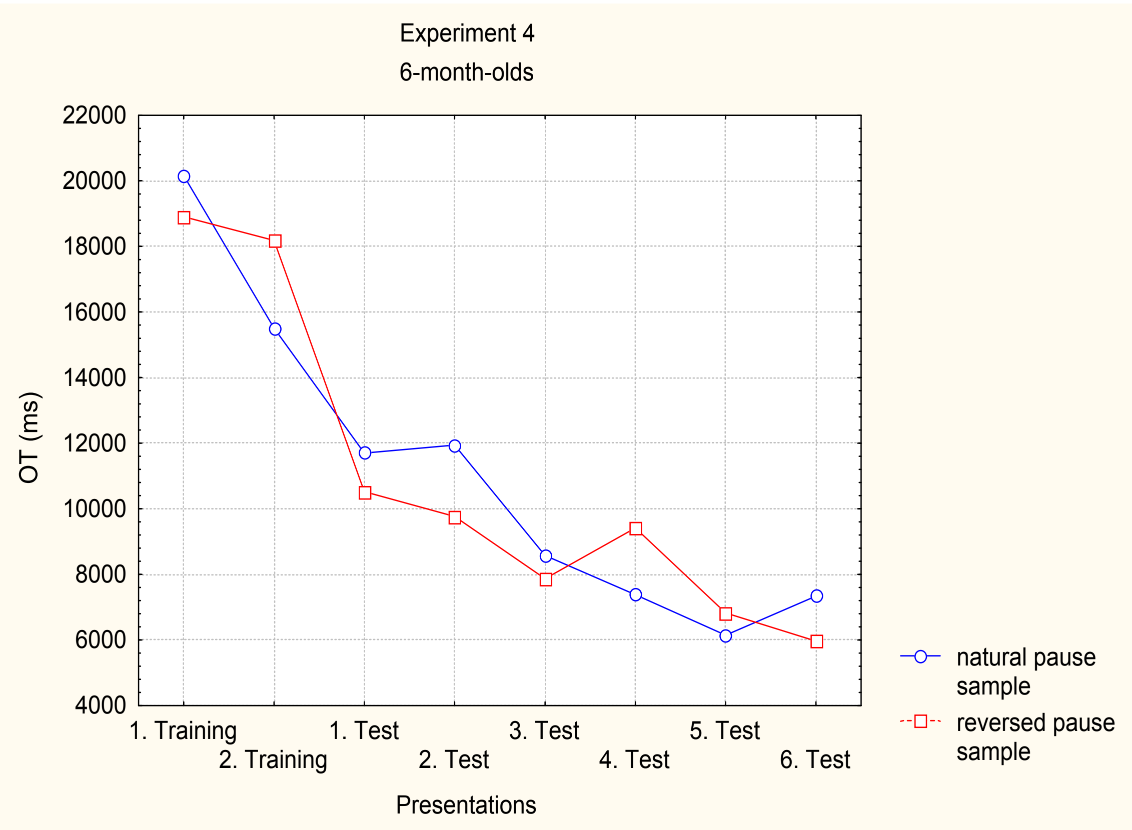

Figure 4.8: Experiment 4 - Mean orientation time for each condition over the presentations during the experiment.

The graph shows a permanent changing of infants' preferences over the course of the experiment. Again, there is an overall decrease in the mean orientation times towards both conditions. Looking at the distribution of orientation times for the six test trials a picture emerges which seems compatible with the model of preference changes over time: after the brief training in the first presentations, the 6-month-olds show an initial preference for the natural pause condition. Over the course of the experiment this initial preference changes to a preference for the reversed pause condition. In the last test presentation there is again a change in the preference. This looking behaviour on the last presentation was also seen in the first experiment. However, the orientation time differences on the second and fourth test presentation, in which the change is most pronounced, do not reach statistical significance, which might be due to the great inter-individual difference in the orientation times of the infants. The orientation times for the second test presentation are 1195 ms (SD = 761 ms) for the natural pause condition and 976 ms (SD = 854 ms) for the reversed pause condition ($t_{(29)}$ = 1.08; p = 0.29 (paired t-test)). The orientation times for the fourth test presentation are 738 ms (SD = 635 ms) for the natural pause condition and 941 ms (SD = 687 ms) for the reversed pause condition ($t_{(29)}$ = 1.40; p = 0.17 (paired t-test)). Thus, the null result of this experiment cannot be attributed to the change of the infants' preferences, though on a descriptive level the behaviour does seem to support such an interpretation.

Thus, the German 6-month-olds gave no indication of discriminating between the speech samples containing sentences corresponding to the natural pause hierarchy and speech samples containing sentences conflicting with this hierarchy. However, it should be noticed that for the first time in this series of experiments the listening preference tends towards the speech samples of the natural pause condition.

4.4.5 Discussion

A possible explanation of why the 6-month-olds did not react differently to the speech samples with natural pause condition and the reversed pause condition might originate from the minimal time difference in these conditions. The intra- and inter-sentence pauses in both conditions differed by 250 ms only, with the position of the pauses being constant.

However, a study by Werner and colleagues (Werner, Marean, Halpin, Spetner & Gillenwater, 1992), using a gap detection technique in a behavioural study, provided evidence that 6-month-olds have a gap detection threshold of only 45 ms to 80 ms, depending on the frequency cut-off, to detect silent intervals in broadband noise. This threshold was still about ten times higher than for adults in the same experimental conditions.

In another study, Trainor and colleagues (Trainor, Samuel, Desjardins & Sonnadara, 2001), using mismatch negativity (MMN) as an automatic response in the ERP paradigm, were able to elicit an MMN in infants at the age of 6- to 7-months with gaps as short as 4 ms in sinewave tones (deviant) compared to a tone without a gap (standard). Thus, it seems that the detection of short pauses on the basic level of auditory processing, that is, "relatively peripherally in the auditory system" (Trainor et al., 2001, p. 2447) is possible for infants from early on. However, this kind of processing only gives indications of the minimal duration of pauses that still can be discriminated in tone sequences, but does not include memory processes and the integration of cues that are necessary in speech processing.

In their ERP study on the memory trace in newborns, Cheour et al. (2002) were able to elicit an MMN only in a condition in which the interstimulus interval (ISI) was 700 ms, but not in a condition with 350 ms or 1400 ms. That is, the change in stimulus was detected when the ISI was 700 ms, but not when it was considerably shorter or longer. Thus, Cheour et al. (2002) suggested that the memory trace of the previous stimulus needed to compare this to the following stimulus can last for a duration of 700 ms. The finding with the 350 ms ISI suggested that infants may need a longer stimulus encoding time to extract auditory information, rendering the representation of the stimuli incomplete, or that the trace of the previous deviant stimulus was still present, so no reaction to the following deviant stimulus was obtained (ibid., p. 37). Thus, the duration of the pause was too short and interfered with the encoding and comparison of one stimulus with another.

The previously cited studies all used tone signals and pauses to determine the shortest duration of pauses that still can be discriminated by infants or the longest duration a pause may have to still identify the following tone as either being equal to the last one or deviate from it. The design of the present experiment required 6-months-olds to recognise differences in the pause duration of 250 ms between the intra- and inter-sentence pauses within fluent speech in the two conditions. Thus, the infants had to compute the relative duration of pauses in both conditions and, assuming their sensitivity to the natural pause hierarchy, to compare the distribution of the different pause durations with the model. This task appears to be more demanding than the identification of tones. Note that in the pre-test for the generation of the stimuli for this study (cf., section 4.1.2), adults performed at chance level when asked to rate the “naturalness” of sentences containing a 100 ms pause at the clause boundary compared to sentences that contained a 200 ms pause at the clause boundary. Thus, one can conclude that adults also were not sensitive to a difference of 100 ms in pause duration when processing fluent speech.

An alternative account for the result pattern of this experiment would be that infants at 6 months of age simply are not sensitive to the natural pause hierarchy. However, the 6-month-olds showed a tendency to listen to the natural speech samples longer than to the reversed pause samples. This makes it plausible to assume that the result of the present experiment is not due to a general lack in the sensitivity to the pause hierarchy but to a developmental process in the recognition of subtle time differences in speech pauses. Therefore it was decided to examine a group of 8-month-old infants to test the hypothesis that older infants are sensitive to the natural pause hierarchy.

4.5 Experiment 5

4.5.1 Participants

Thirty 8-month-old German infants participated in this experiment (15 girls, 15 boys; mean age: 8 months 13 days; range: 8 months 0 days to 8 months 29 days). The same selection criteria as in the previous experiments applied. None of the infants had participated in one of the previous experiments of this study conducted with 6-month-olds. Thirteen additional infants were tested but their data could not be included in the analysis due to the following reasons: problems with the technical equipment (2), infants did not finish the experimental session or were too fussy (6), or having a mean orientation time of less than 3 s per condition (5).

4.5.2 Stimuli

The stimuli were identical to the ones in the previous experiment.

4.5.3 Method and Procedure

The method and procedure were identical to the one in the previous experiment.

4.5.4 Results

For each child the data were recoded by a second experimenter as described in the first experiment. Also the same exclusion criterion applied. Corrections were made in eight of the 30 datasets (26.7%), the inter-coder reliability was $r = 0.99$ (Pearson).

The mean orientation time towards the natural pause condition was 7208 ms (SD = 4007 ms) and the mean orientation time towards the reversed pause condition was 6090 ms (SD = 2439 ms). This difference was significant: $t_{(29)} = 2.22$; $p = 0.03$ (paired t-test). Nineteen of the 30 infants had longer OTs towards the natural pause condition.

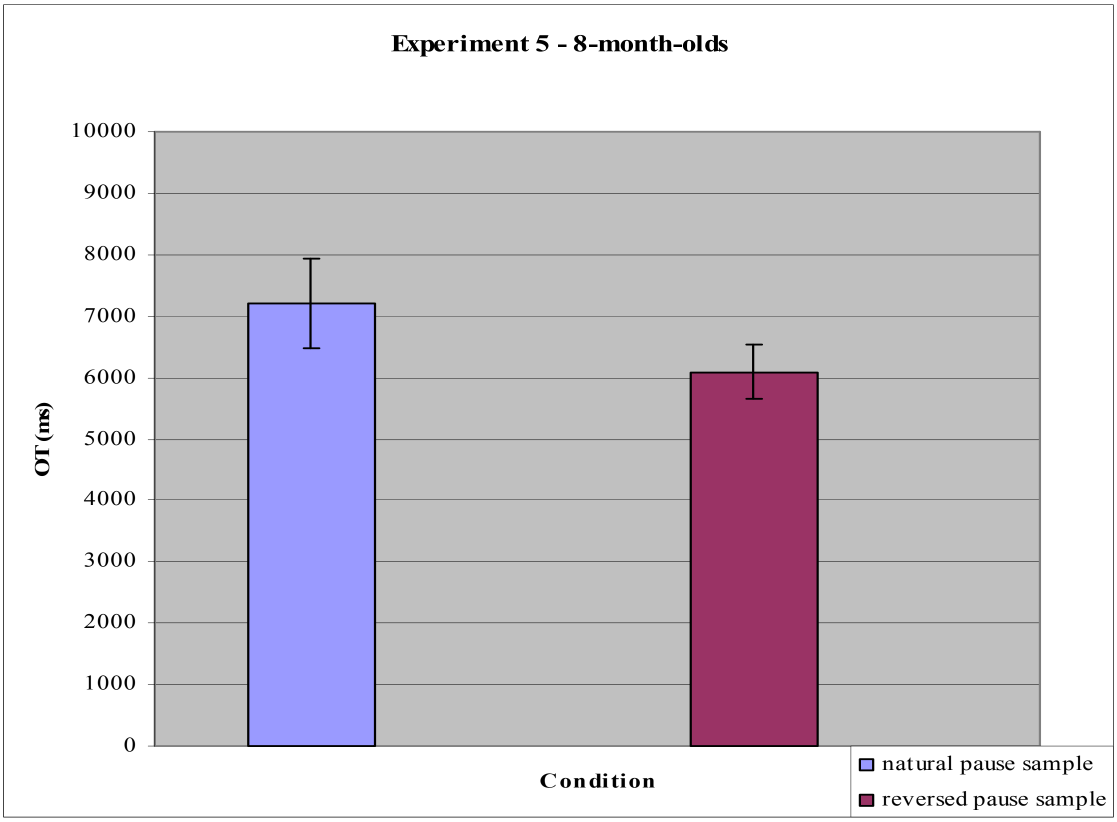

Figure 4.9: Experiment 5 - Results of 8-month-olds' discrimination of sentences with natural and reversed pause durations.

To evaluate whether this result was influenced by the order of presentation (natural pause first vs. reversed pause first) or the sex of the infant, again two statistical analyses using ANOVA were conducted. In the analysis with the between-subjects factor Order of Presentation (natural pause first vs. reversed pause first) and the within-subjects factor Condition (natural pause vs. reversed pause) there was no significant main effect for order of presentation ($F_{(1,28)} = 0.01$; $p < 1$) and no significant interaction ($F_{(1,28)} = 0.49$; $p < 1$). In the analysis with Sex of the infant as the between-subjects variable and Condition as the within-subjects factor there was also no main effect for sex of the infant ($F_{(1,28)} = 0.08$; $p < 1$) and no significant interaction ($F_{(1,28)} = 0.38$; $p < 1$).

For this experiment also a possible change in the infants' preferences over time was analysed. In Figure 4.10 the mean orientation time for each condition and each presentation, including the pre-test training trials is plotted.

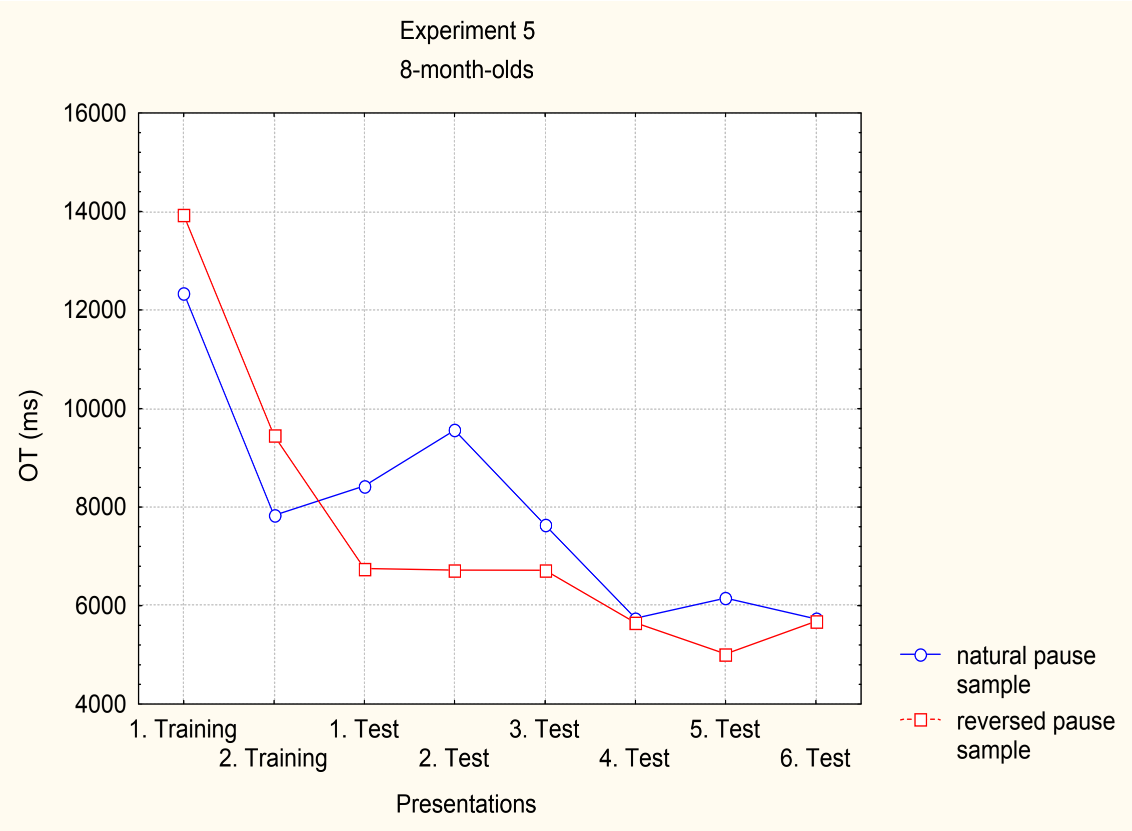

Figure 4.10: Experiment 5 - Mean orientation time for each condition over the presentations during the experiment.

For both training trials the mean orientation times of the infants are longer for the reversed pause condition. These differences, however, do not reach significance level. For the first training trial: $t_{(29)} = 0.86$; $p = 0.40$, and for the second training trial: $t_{(29)} = 0.89$; $p = 0.38$ (paired t-tests). During most of the six test presentations a constant preference for the natural pause condition can be seen. The constant longer mean orientation times towards the natural pause samples during the test phase of the experiment lead to the significant orientation time differences reported above.[50] This constant behaviour of the 8-month-olds leads to the interpretation that the preference for the natural pause condition is a genuine discrimination and preference effect. The model of preference changes (see section 3.1.2)

[50] Interestingly, the pattern of orientation times during the test phase mirrors the one found in the second experiment, in which the infants showed a significant preference for the unnatural condition: the most pronounced orientation time difference can be seen in the second test presentation, whereas the orientation times to both conditions on the fourth and sixth presentation are nearly identical. Furthermore, the preference for one type of stimulus remains constant throughout the test phase.

would predict a change in the preference of the older infants to the novel condition, that is, in this experiment for the reversed pause condition, early in the course of the experiment.

To evaluate whether the difference in the age of the infants is indeed the decisive factor for the results of the previous two experiments, a statistical analyses using ANOVA was conducted. The analysis with the between-subjects factor Age (6 months vs. 8 months) and the within-subjects factor Condition (natural pause vs. reversed pause) revealed a significant main effect for Age ($F_{(1,58)} = 6.73$; $p < 0.01$) and a significant main effect for Condition ($F_{(1,58)} = 4.89$; $p < 0.03$). There is no significant interaction between the two factors ($F_{(1,58)} = 0.87$; $p < 1$).

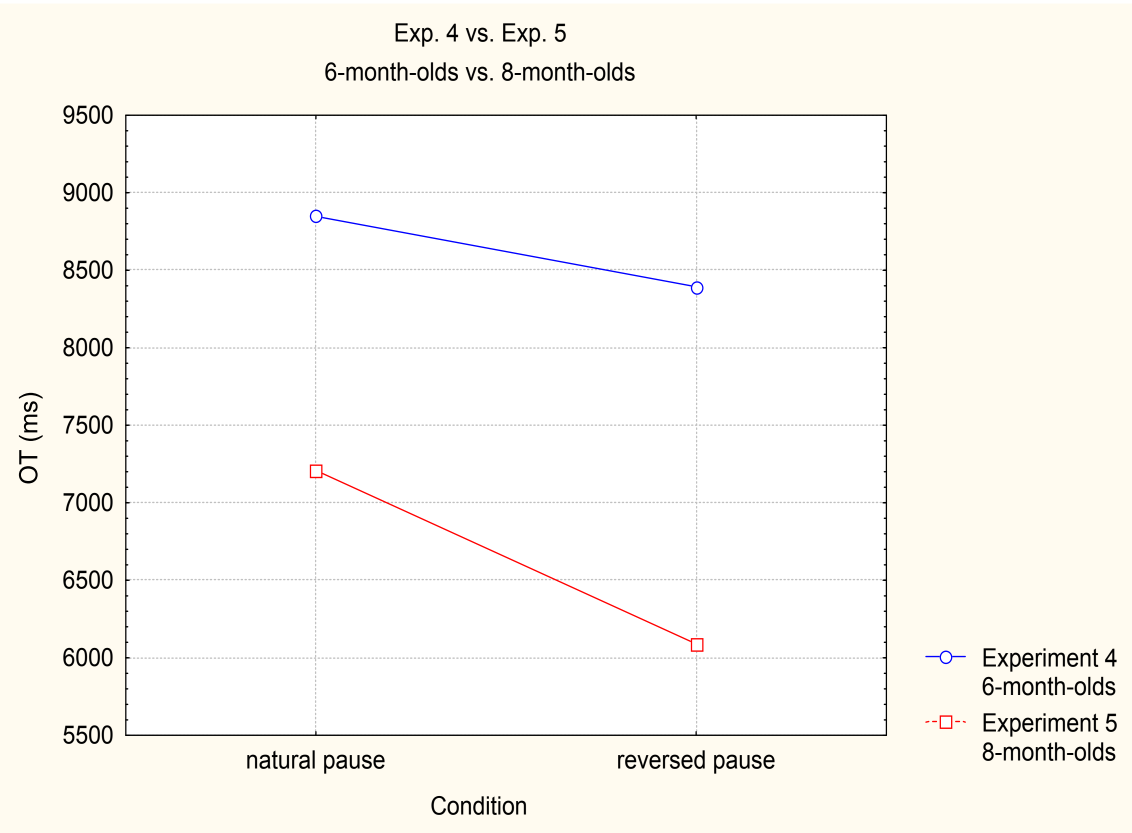

Figure 4.11: Comparison of the results of Experiment 4 and Experiment 5.

The significant main effect for age shows that, overall, the 6-month-old German infants have longer mean OTs to both conditions than the 8-month-olds. The significant main effect for Condition shows that both age groups have longer OTs to the natural pause condition than to the reversed pause condition. Furthermore, as also can be seen in the graph of Figure 4.11, the lack of an interaction suggests that both age groups tend to react similarly to the stimuli.

4.5.5 Discussion

The findings of these last two experiments support the interpretation that there is a development in the sensitivity to the natural pause hierarchy, which can be seen already in 6-month-olds but is verifiable first in the 8-month-olds. The results of the 8-month-olds also seems indicate that they have a sophisticated structural knowledge of their native language, because the development of the ability to perceive subtle time changes alone would not necessarily have resulted in the preference for the natural pause condition.

In this experiment, in which the pauses were inserted only at the location coinciding with the clause boundary and the difference in the samples was caused by a timing difference, the infants preferred to listen to the natural samples rather than to the reversed pause samples. This supports the notion that the novelty effect found in the second experiment was due to the highly irregular prosody caused by the insertion of the pause after the infinitive particle.

The question whether infants are sensitive to the natural pause hierarchy in their native language has never before been addressed. The results of the present experiments suggest that by 6 and 8months of age infants have a sophisticated knowledge not only of the location in which a pause might occur in correlation to a syntactic boundary, but also of the timing in their native language.

5 General Discussion and Conclusion

The present study investigated whether and under which conditions German-learning infants recognise clauses in fluent speech and the role a prosodic marker such as a pause may have in the segmentation process.

The clause is an important domain in the grammar of a language, because it is the domain in which grammatical rules such as subject-verb agreement apply. Furthermore, clauses differ cross-linguistically in the ways the order of their constituents are realised, for instance, the difference between SVO (e.g., English) and SOV (e.g., Turkish) languages. On the clause level, but not on the phrase level, the infant can learn whether her native language is a pro-drop language or not.

In the speech stream, a syntactic clause often coincides with an intonational phrase (Nespor & Vogel, 1986, p. 190) and in the stimuli used in the experiments of the present study always did. The boundaries of these intonational phrases are marked by changes in the fundamental frequency, that is, a fall or rise in pitch (e.g., Price et al., 1991), and a lengthening of the syllable adjacent to this boundary (e.g., Cooper & Paccia-Cooper, 1980; Wightman et al., 1992).

Furthermore, previous studies investigating the occurrence and distribution of pauses in speech corpora of different languages have revealed that pauses often demarcate linguistically relevant units, like clauses / intonational phrases and sentences / phonological utterances (e.g., Butcher, 1981, for German). The findings of these studies suggest the existence of a natural pause hierarchy (cf., Figure 2.1) that complements the prosodic hierarchy described by Nespor and Vogel (1986). These hierarchies on the signal side correspond to the syntactic hierarchy of a language. Although there is not always a perfect one-to-one match, often a phonological utterance corresponds to a sentence, an intonational phrase corresponds to a clause, a phonological phrase corresponds to a syntactic phrase and a phonological word corresponds to a lexical word.

Thus, if infants start out with the working hypothesis that important syntactic units are prosodically marked at the boundaries and demarcated by pauses in the signal, they would be able to segment units like sentences and clauses correctly in the speech stream most of the time. The mechanism that was suggested to allow infants to exploit this correlation between the signal side and the syntactic side is the prosodic bootstrapping mechanism (e.g., Gleitman & Wanner, 1982; Jusczyk, 1997).

Previous research (e.g., Hirsh-Pasek et al., 1987; Nazzi et al., 2000b) provided evidence that infants as young as 6 months are able to recognise clauses in fluent speech. Although some of the studies used the insertion of pauses as a means to test infants' recognition of clauses, only recently the question has been addressed whether infants use the pause as a cue to identify clause boundaries in the speech stream (Seidl, 2007; Johnson & Seidl, 2008; Seidl & Cristià, 2008). The results of these studies indicated that cross-linguistically the pause cue is weighted differently: English-learning 6-month-olds did not seem to rely on the pause as a cue to recognise clause boundaries, whereas Dutch-learning infants did. However, none of these studies investigated whether infants not only use the pause as a cue to recognise clauses in fluent speech but whether they are also sensitive to the natural pause hierarchy in their language.

Therefore, the open research questions that have been addressed in this dissertation were: Do German-learning 6-month-olds use the pause as a cue to recognise clause boundaries in fluent speech? And are the infants sensitive to the correlation of different pause durations and the type of boundaries they demarcate?

Five experiments were conducted to investigate these issues. The first three experiments addressed the question whether German-learning 6-month-olds use the pause cue to recognise clause boundaries in fluent speech. These experiments were designed according to the original study by Hirsh-Pasek et al. (1987). The stimulus material used in their study consisted of speech excerpts of a natural mother-child interaction. The speech passages were manipulated in a way that a pause of 1 s duration was inserted at every clause boundary in the sample in the natural condition. This means that sometimes the pause was inserted at a within-sentence clause boundary and sometimes at a sentence boundary, which might be marked even in a more pronounced way. To create the unnatural condition, pauses of 1 s duration were inserted at different locations within the clauses. Furthermore, these samples started and ended in the middle of a clause. Thus, the infants were confronted with rather heterogeneous speech materials.

To avoid this heterogeneity and to control for the difference between the within-sentence clause boundary and the between-sentence boundary, the stimuli in the present experiments were composed using isolated sentences consisting of two coordinated clauses which coincided with intonational phrases as in *Die Tante verspricht Katrin zu schwimmen und die Insel zu erkunden*. (‘The aunt promises Katrin to swim and to explore the island’). The choice of isolated sentences rather than speech passages had the advantage that the pause durations of the artificially inserted pauses could be matched to reflect the natural pause hierarchy. A speech sample for the natural condition consisted of six sentences in which the pauses at the within-sentence clause boundaries were shorter than the pauses between subsequent sentences. In the six sentences of a speech sample for the unnatural condition, a pause was inserted within the first clause after the infinitive particle *zu* (‘to’), and the originally occurring pause before the coordinating conjunction *und* (‘and’) was shortened to 100 ms. This additional manipulation was made to avoid too much variability between the duration of sentences in the natural and the unnatural condition.

In the first experiment, the duration of the inserted pauses at the clause boundary, or within the clause, respectively, was 1 s. This duration corresponded to the one used by Hirsh-Pasek et al. (1987). The pauses between subsequent sentences were 1.5 s long. The results of this experiment showed that in this pause duration condition, the infants did not seem to discriminate between the natural and the unnatural samples. This result was rather striking given that 6- to 8-month-old English-learning infants in the experiment by Hirsh-Pasek et al. (1987) already showed a preference for the natural samples. However, one reason for differences in the outcome of the two studies might have been the fact that the duration of the inserted pauses in the present experiment was longer than the mean duration of pauses usually occurring at the clause and sentence boundaries in German (cf., Butcher, 1981). A second experiment was conducted to test the hypothesis that German-learning 6-month-olds recognise clause boundaries when the duration of the inserted pauses corresponded to the mean duration of pauses in natural speech.

In the second experiment, the pauses that were inserted between the clauses within the sentence, or within the first clause, had a duration of 750 ms and the duration of pauses between subsequent sentences was 1 s. These durations were chosen because they corresponded to the mean duration of pauses at the clause and the sentence level, respectively, in German (cf., Butcher, 1981). In this experiment, the infants had significant longer mean orientation times to the unnatural speech samples than to the natural speech samples. This was taken as an indication that the infants recognised the clause boundaries in the speech samples which were marked by the converging cues of pitch change, final lengthening and the inserted pauses. The direction of the effect, that is, the longer orientation times to the unnatural speech samples has been attributed to the prosodic idiosyncrasies in this condition: the pause was inserted after the short infinitive marker *zu* ('to') and in this condition the other prosodic markers indicating a boundary were not present at this location. Thus, the natural clause prosody was disturbed in this condition which might have heightened the salience of this stimulus type for the infants.

A first conclusion that can be drawn on the basis of the results of the two experiments with German-learning infants is that they seem to be using the pause as a cue to recognise clause and sentence boundaries. These results on German complement the findings in studies on other languages that infants are able to recognise clauses and clause boundaries in fluent speech (Hirsh-Pasek et al., 1987, for English; Hayashi & Mazuka, 2002; Hayashi et al., 1996, for Japanese). However, the studies on English and Japanese did not investigate the correlation between pause duration and the type of boundary it marks. In these studies a pause of 1 s duration was inserted both at clause and at sentence boundaries.

The finding that the recognition of clause boundaries seems to be closely related to the duration of pauses, that is, infants discriminated the natural from the unnatural condition only when the pauses had a duration that conforms to the duration of the respective pauses in natural German speech, can be interpreted in two ways. Either the working memory of the 6-month-olds is not developed enough to allow them to track the ongoing sentence over a pause duration of 1 s (cf., Cheour et al., 2002), or the infants are already sensitive to the pause durations found in the pause hierarchy in German. These alternatives were addressed in Experiment 4.

The third experiment in this series examined whether infants also discriminate between the natural and the unnatural speech samples when the originally occurring pause before the coordinating conjunction, which had been shortened in the first two experiments, was retained. This resulted for the unnatural condition in sentences which included three pauses: the artificial pause inserted within the clause, the original pause (mean duration: 523 ms, with a range from 368 ms to 750 ms) before the conjunction, and the pause at the sentence boundary. In the natural condition, the sentences had only the artificially inserted pause at the clause and at the sentence boundary. The pause duration of the artificial pauses was kept identical to the second experiment. The results of Experiment 3 show that the infants did not seem to discriminate between the natural and the unnatural condition. Thus, in contrast to the second experiment in which a listening preference for the unnatural condition was found, the infants participating in the third experiment did not prefer to listen to either condition. It was argued that this might have been due to the fact that the more frequent pauses in the sentences of the unnatural condition impeded the development of the preference that was seen in the second experiment.

Experiments 4 and 5 directly addressed the question whether German-learning infants are sensitive to the natural pause hierarchy of their native language. To test this hypothesis a new set of test stimuli was created. Artificial pauses within the sentences in both conditions were inserted only at the clause boundary before the coordinating conjunction. All other naturally occurring pauses in the sentences remained intact. That is, in contrast to the previous experiments, in this experiment there was no longer a condition in which the pause was inserted at an unnatural position within the sentence. To create the natural pause condition, the pauses between the clauses within the sentence had a duration of 750 ms and the pauses between subsequent sentences had a duration of 1 s. To create the reversed pause condition, the duration of the pauses was inverted: the pause at the clause boundary was 1 s long and the pause at the sentence boundary was 750 ms long. Thus, to discriminate between the conditions, infants must be able notice the subtle time changes. In the fourth Experiment 6-month-old infants were tested. The results of this experiment showed that the infants did not significantly prefer to listen to either condition, although they had a tendency to listen longer to the natural pause samples. Thus, even though it was only a tendency, already the 6-month-olds had longer OTs to the speech samples that conformed to the natural pause hierarchy. This could be interpreted as tentative evidence that the result of the second experiment was also influenced by infants' sensitivity to the pause hierarchy. However, the differences in the pause duration between the clauses and between the sentences in the natural pause and the reversed pause conditions were rather short: only 250 ms. Therefore, the 6-month-olds might not have been able to compute such a subtle timing difference.

The fifth experiment was conducted to test the hypothesis that the sensitivity to the natural pause hierarchy was subject to development. And indeed, the 8-month-old infants, tested on the same stimuli as the 6-month-olds in the previous experiment, listened significantly longer to the speech samples of the natural pause condition. This might be interpreted in a way that the 8-month-olds already have sophisticated structural knowledge of their language as the development of the ability to perceive subtle time changes alone would not necessarily have resulted in the preference for the natural pause condition.

As no other acoustic cue to prosodic boundaries like final lengthening and / or the F0 was manipulated in the present experiments, the results of the experiments can only be attributed to the infants' use of pauses to identify clause / IPh boundaries and their sensitivity to the natural pause hierarchy.

With respect to the debate of the direction of the effect in experiments with infants, the experiments of the present study seem to suggest that the preference for the unnatural condition in the first three experiments may have been driven by the prosodic idiosyncrasies of the unnatural sentences. In this condition the pause was inserted after the short infinitive particle and in this position did not coincide with other prosodic markers indicating a boundary. This interrupts the natural clause prosody and thereby might have attracted infants' attention, reflected in the longer orientation time towards these samples. Thus, with respect to the model of Hunter and Ames (1988) the factor complexity seems to have led to the kind of novelty effect found in these experiments. The factor age does not seem to be of consequence, because in the fourth experiment already the 6-month-olds showed a tendency to listen to the natural pause samples. In this experiment the artificial pauses were inserted at the clause and the sentence boundary only and their respective duration was varied. Furthermore, the fact that the 8-month-olds significantly preferred to listen to the natural pause condition also suggests that the factor age was not decisive in the outcome of the experiments, because according to Hunter and Ames (1988) the increasing age of an infant should potentially lead to a novelty effect.

What do the results of the present study contribute to the recently initiated line of research on the language-specific use and weighting of cues to recognise clause boundaries? First of all, in the cross-linguistic context, German-learning 6-month-olds seem to rely on the pause as a cue to recognise clause boundaries, as do their Dutch-learning peers (Johnson & Seidl, 2008). For English-learning 6-month-olds, on the other hand, the pause did not seem to be a necessary cue to recognise clause boundaries in fluent speech (Seidl, 2007). These results are interesting because all three languages are Germanic languages. So, what might be the reason for the different weighting of cues in German and Dutch, on the one hand and English, on the other hand? One difference between the languages is that in English vowel duration is not phonemic, whereas it is in German and Dutch. Following the idea of Morgan (1986, p. 112) that languages probably do not use inherent features to mark boundaries, English infants might rely more on the duration cue than Dutch and German infants. English adults have been shown to rely on the duration cue to disambiguate syntactically ambiguous sentences (e.g., Lehiste et al., 1976). However, Seidl (2007) found that English-learning infants seemed to rely on the pitch cue to recognise clause boundaries. This difference between the infants and adults might have a developmental explanation: the English-learning 6-month-olds may not yet be sensitive to changes in duration. This would be parallel to the German-learning 6-months-olds' reaction in Experiment 4, in which they were not able to discriminate between the natural pause condition and the reversed pause condition. English-learning infants might rely more on the pitch cue because of their experience with the child-directed register, in which the pitch is a rather exaggerated feature (c.f., Fernald et al., 1989). In contrast, Dutch- and German-learning infants, in whose language vowel duration is phonemic, might rely more on the pause cue to recognise clause boundaries. Dutch adults also seemed to rely more on the pause cue than English adults did (cf., Johnson & Seidl, 2008; Scott, 1982).

To further investigate the language-specific weighting of cues to recognise clause boundaries, it would be necessary to test German-learning infants on their use of the pitch cue and the duration cue to gather more cross-linguistic evidence. Furthermore, it would be interesting to test German adults and to examine whether they, like the Dutch adults, also rely on the pause cue. Although the study of Steinhauer et al. (1999) suggests the German adults might not rely on the pause cue, because a CPS at the IPh boundary was also found in the absence of a pause.

With respect to the sensitivity to the natural pause hierarchy, this dissertation is the first study which addresses this question directly. However, one aspect in the experiments conducted by Seidl and colleagues (Seidl, 2007; Johnson & Seidl, 2008) might be reinterpreted to indicate that also the Dutch- and English-learning infants might be sensitive to the natural pause hierarchy. In their experiments testing the use of the pause cue, the infants were familiarised either with a well-formed clause or with a sequence of the same words spanning a prosodic boundary. In both studies, Seidl (2007) and Johnson and Seidl (2008) manipulated the pause cue by inserting an artificial pause between words within the clause, or by deleting the pause between the words at the boundary. The present line of argument is based on the manipulations of the well-formed clauses in the long pause condition.

In the English stimuli a pause of 300 ms duration was inserted within the clause, which was the longest pause duration that was found for the non-clausal stimuli (Seidl, 2007, p. 32). Thus, in the clause *Rabbits eat leafy vegetables* there was a pause of 300 ms between *eat* and *leafy*. In the Dutch stimuli there was a pause of 738 ms between the words *eten* and *koude* in the clause *De jongens eten koude pizza,* which corresponded to the mean duration of pauses found in the non-clausal stimuli. Furthermore, to create the familiarisation strings, Seidl (p.c.) inserted a pause of 1 s duration between the subsequent clauses. This resulted in a familiarisation string for the clauses in English with the following pause distribution: *Rabbits eat* $\_\_\__{300\ ms}$ *leafy vegetables* $\_\_\__{1\ s}$ *Rabbits eat* […]. The familiarisation string for the Dutch clause stimuli had the following pause distribution: *Jongens eten* $\_\_\__{738\ ms}$ *koude pizza* $\_\_\__{1\ s}$ *Jongens eten* […].

If infants are sensitive to the natural pause hierarchy, the results of these experiments, that is, the English-learning infants preferred to listen to the well-formed clause even with the pause insertion, whereas the Dutch-learning infants did not, could be explained by the differences in the duration of the inserted pauses used. The natural pause hierarchy suggests that pauses between sentences have a duration of about 1 s and pauses between words within a phrase can have a duration between 0 ms and 300 ms (cf., Figure 2.1). Thus, the English-learning infants listened to strings of speech which were well-formed even with the inserted pause, allowing them to recognise the clauses in the test phase. The Dutch-learning infants were familiarised with strings that violated the natural pause hierarchy, because the pause inserted between subsequent words in the phrase was too long, thus possibly impeding the recognition of these clauses during the test phase. Thus, the reaction of the infants to the well-formed clauses in these experiments might be reinterpreted in a way that also learners of other languages might be sensitive to the natural pause hierarchy. Further research is clearly necessary to validate or falsify this hypothesis.

Another interesting aspect would be to investigate how the weighting of cues changes over time. As has been mentioned above, English-learning infants seem to rely on other cues than English adults do. In Dutch, the pause seems to be an important cue for both infants and adults. However, given that pauses have other functions in the speech stream as well, and given that adults detect a stimulus change over a 10 s delay, whereas newborns were not able to do so over a 1.4 s delay (Cheour et al., 2002), the question arises: When does the speaker / listener become more flexible? Besides their function to demarcate linguistically relevant units, pauses also have pragmatic functions. For instance, during a pause in the speech a turn-taking, that is, a speaker change, may take place. Furthermore, a long pause may also indicate the change in the topic the speaker is talking about. However, one can assume that changes in that direction will be a later achievement in the language acquisition process.

An immediate advantage of the sensitivity to the natural pause hierarchy is that it seems to help infants to find the relevant linguistic units in their language, starting with the larger ones, like clauses, at 6 months of age. Reviewing the literature on the perception of words, phrases and clauses in infants (e.g., Jusczyk, 1997), an interesting time scale emerges: Clauses seem to be recognised first at about 6 months of age, and about one and a half months later, at about 7.5 months, the first words are recognised in the speech stream. As clause boundaries are also word boundaries, the recognition of the clause boundary may have helped infants to segment the first words from fluent speech (cf. e.g., Seidl & Johnson, 2006, 2008; Schmitz et al., submitted). These recognised words included content words (Jusczyk & Aslin, 1995) as well as function words (Höhle & Weissenborn, 2003), and also nonsense words like ‘taris’ from the sequence ‘guitar is’ (Jusczyk et al., 1999). Therefore, one can assume that infants recognise the sound pattern of the words, but do not yet attribute a function to these words. Again, and one a half months later, at the age of 9 months, the recognition of phrases can be seen (Jusczyk et al., 1992). This might be due to the fact that phrase boundaries are not so well marked in the speech stream. The findings of the present study may suggest that the sensitivity to the natural pause hierarchy developed in 8-month-olds, among other factors, also may play a role in the recognition of phrases. Again, one and a half months later, at the age of 10.5 months, infants reacted differently to speech passages that included real function words as opposed to speech passages that included nonsense function words (Shady, 1996; cited by Jusczyk, 1998). Thus, the sensitivity to phrases in the speech may have helped infants to learn about function words, which usually occur at the boundary of such phrases. Thus, the sensitivity to the pause hierarchy may also help infants in “dividing and conquering” the linguistic input (Jusczyk, 1998).

To conclude, the results of the experiments presented in this dissertation provide evidence that German-learning 6-month-olds are able to recognise clause boundaries in fluent speech. This finding complements the results of previous studies on languages like English and Japanese. Furthermore, the experiments show that, at least for German, the sensitivity to clause boundaries is closely related to the pause duration in the speech samples. The results of the present study suggest that by 6 and 8 months of age the infants have a sophisticated knowledge not only of the location in which a pause might occur in relation to a syntactic boundary, but also of the timing in their native language. That is, in interaction with the other prosodic features indicating a boundary, infants preferred to listen to speech samples in which the pauses at the sentence boundary were longer than at the clause boundary. This is in accordance with the natural pause hierarchy.

Because all clause boundaries in the experimental stimuli coincided with intonational phrase boundaries, it is not possible to tear these features apart. However, previous research indicated that the recognition of clauses most likely seems to be mediated by the perception of prosodic features occurring at these boundaries, because infants were able to recognise NP / VP boundaries only in a condition in which the NP was fully fledged and not in a condition in which the NP consisted of a pronoun, whose prosodic realisation is usually incorporated in the intonational phrase of the VP.

Further research is necessary to verify the findings of infants' sensitivity to the natural pause hierarchy. This, for instance, includes the testing of infants' perception of minor boundaries, that is, phonological phrase boundaries, in relation to appropriate and inappropriate duration of inserted pauses.

6 References

Aasland, W. A. & Baum, S. R. (2003). Temporal parameters as cues to phrasal boundaries: A comparison of processing by left- and right-hemisphere brain-damaged individuals. *Brain and Language, 87*, 385-399.

Aslin, R. N. (1987). Visual and Auditory Development in Infancy. In Osofsky, J. D. (Ed.), *Handbook of Infant Development* (2nd ed., pp. 5-97). New York: Wiley.

Aslin, R. N. & Mehler, J. (2005). Near-infrared spectroscopy for functional studies of brain activity in human infants: promise, prospects, and challenges. *Journal of Biomedical Optics, 10*(1), 0110091-0110093.

Aslin, R. N., Woodward, J., LaMendola, N. & Bever, T. (1996). Models of word segmentation in fluent maternal speech to infants. In Morgan, J. L. & Demuth, K. (Eds.), *Signal to Syntax: Bootstrapping from Speech to Grammar in Early Acquisition* (pp. 117-134). Mahwah, NJ: Lawrence Erlbaum.

Bates, E. & MacWhinney, B. (1989). Functionalism and the competition model. In MacWhinney, B. & Bates, E. (Eds.), *The cross-linguistic study of sentence processing*. Cambridge: Cambridge University Press.

Blenn, L., Seidl, A. & Höhle, B. (2003). Recognition of Phrases in Early Language Acquisition: The Role of Morphological Markers. In Beachley, B. (Ed.), *Proceedings of the 27th Annual Boston University Conference on Language Development* (Vol. 1, pp. 138-149). Somerville, MA.: Cascadilla Press.

Bortfeld, H., Morgan, J. L., Golinkoff, R. M. & Rathbun, K. (2005). Mommy and Me - Familiar Names Help Launch Babies Into Speech-Stream Segmentation. *Psychological Science, 16*(4), 298-304.

Bosch, L. & Sebastián-Gallés, N. (1997). Native-language recognition abilities in 4-moth-old infants from monolingual and bilingual environments. *Cognition, 65*, 33-69.

Bosch, L. & Sebastián-Gallés, N. (2001). Evidence of early language discrimination abilities in infants from bilingual environments. *Infancy, 2*(1), 29-49.

Brent, M. R. & Siskind, J. M. (2001). The role of exposure to isolated words in early vocabulary development. *Cognition, 81*, B33-B44.

Butcher, A. (1981). *Aspects of the Speech Pause: Phonetic Correlates and Communicative Functions*. Kiel: Institut für Phonetik, Universität Kiel.

Cheour, M., Ceponiené, R., Leppänen, P., Alho, K., Kujala, T., Renlund, M., Fellman, V. & Näätänen, R. (2002). The auditory sensory memory trace decays rapidly in newborns. *Scandinavian Journal of Psychology, 43*, 33-39.

Chiappetta, J., Monti, L. A. & O'Connell, D. (1987). Pause Perception: Some cross-linguistic comparisions. *Bulletin of the Psychonomic Society, 25*(2), 103-105.

Chomsky, N. (1965). *Aspects of the theory of syntax*. Cambridge, MA.: The MIT Press.

Chomsky, N. (1986). *Knowledge of Language. Its Nature, Origin and Use*. New York: Praeger Publishers.

Christophe, A., Guasti, M. T., Nespor, M., Dupoux, E. & van Ooyen, B. (1997). Reflections on Phonological Bootstrapping: Its Role for Lexical and Syntactic Acquisition. *Language and Cognitive Processes, 12*(5/6), 585-612.

Christophe, A., Nespor, M., Guasti, M. T. & Van Ooyen, B. (2003). Prosodic structure and syntactic acquisition: the case of the head-direction parameter. *Developmental Science, 6*(2), 211-220.

Cooper, R. P. & Aslin, R. N. (1990). Preference for Infant-directed Speech in the First Month after Birth. *Child Development, 61*(5), 1584-1595.

Cooper, W. E. & Paccia-Cooper, J. (1980). *Syntax and Speech*. Cambridge, MA: Harvard University Press.

Cutler, A. (1994). Segmentation problems, rhythmic solutions. *Lingua, 92*, 81-104.

Cutler, A. (1996). Prosody and the Word Boundary Problem. In Morgan, J. L. & Demtuh, K. (Eds.), *Signal to Syntax: Bootstrapping from Speech to Grammar in Early Acquisition* (pp. 87-99). Mahwah, N.J.: Lawrence Erlbaum.

Cutler, A. & Carter, D. M. (1987). The predominance of strong initial syllables in the English vocabulary. *Computer Speech and Language, 2*, 133-142.

Dankovičová, J., Pigott, K., Wells, B. & Peppé, S. (2004). Temporal markers of prosodic boundaries in children's speech production. *Journal of the International Phonetic Association, 34*(1), 17-36.

DeCasper, A. J. & Fifer, W. P. (1980). On human bonding: Newborns prefer their mothers' voices *Science, 208*, 1174-1176.

Dehaene-Lambertz, G. & Houston, D. M. (1998). Language Discrimination Response Latencies In Two-Month-Old Infants. *Language And Speech, 41*(1), 21-43.

Dooling, D. J. (1974). Rhythm and Syntax in Sentence Perception. *Journal of Verbal Learning and Verbal Behavior, 13*, 255-264.

Duez, D. (1993). Acoustic Correlates of Subjective Pauses. *Journal of Psycholinguistic Research, 22*(1), 21-39.

Echols, C. H., Crowhurst, M. J. & Childers, J. B. (1997). The Perception of Rhythmic Units in Speech by Infants and Adults. *Journal of Memory and Language, 36*, 202-225.

Eimas, P. D. (1974). Auditory and linguistic processing of cues for place of articulation by infants. *Perception & Psychophysics, 16*(3), 513-521.

Fernald, A. (1985). Four-Month-Old Infants Prefer to Listen to Motherese. *Infant Behavior and Development, 8*, 181-195.

Fernald, A. & Kuhl, P. K. (1987). Acousitc Determinants of Infant Preference for Motherese Speech. *Infant Behavior and Development, 10*, 279-293.

Fernald, A. & McRoberts, G. (1996). Prosodic Bootstrapping: A Critical Analysis of the Argument and the Evidence. In Morgan, J. L. & Demuth, K. (Eds.), *Signal to Syntax: Bootstrapping from Speech to Grammar in Early Acquisition* (pp. 365-388). Mahwah, N.J.: Lawrence Erlbaum.

Fernald, A. & Simon, T. (1984). Expanded Intonation Contours in Mother's Speech to Newborns. *Developmental Psychology, 20*(1), 104-113.

Fernald, A., Taeschner, T., Dunn, J., Papousek, M., De Boysson-Bardies, B. & Fukui, I. (1989). A cross-language study of prosodic modifications in mothers' and fathers' speech to preverbal infants. *Journal of Child Language, 16*, 477-501.

Ferreira, F. (1993). Creation of Prosody during Sentence Production. *Psychological Review, 100*(2), 233-253.

Féry, C. (1993). *German Intonational Patterns*. Tübingen: Max Niemeyer Verlag.

Fisher, C. & Tokura, H. (1996a). Acoustic Cues to Grammatical Structure in Infant-Directed Speech: Cross-Linguistic Evidence. *Child Development, 67*, 3192-3218.

Fisher, C. & Tokura, H. (1996b). Prosody in speech to infants: Direct and indirect cues to syntactic structure. In Morgan, J. L. & Demuth, K. (Eds.), *Signal to Syntax: Bootstrapping from Speech to Grammar in Early Acquisition* (pp. 343-364). Mahwah, N.J.: Lawrence Erlbaum.

Friederici, A. D. & Alter, K. (2004). Lateralization of auditory language functions: A dynamic dual pathway model. *Brain and Language, 89*, 267-276.

Friederici, A. D., Friedrich, M. & Christophe, A. (2007). Brain Responses in 4-Month-Old Infants Are Already Language Specific. *Current Biology, 17*, 1208-1211.

Garnica, O. K. (1977). Some prosodic paralinguistic features of speech to young children. In Snow, C. E. & Ferguson, C. A. (Eds.), *Talking to Children: Language Input and Acquisition* (pp. 63-88). Cambridge: Cambridge University Press.

Gerken, L. (1991). The Metrical Basis for Children's Subjectless Sentences. *Journal of Memory and Language, 30*(4), 431-451.

Gerken, L. (2001). Signal to Syntax: Building a Bridge. In Weissenborn, J. & Höhle, B. (Eds.), *Approaches to Bootstrapping: Phonological, Syntactic and Neurophysiological Aspects of Early Language Acquisition* (Vol. 1, pp. 147-166). Amsterdam: Benjamins.

Gerken, L., Jusczyk, P. W. & Mandel, D. R. (1994). When prosody fails to cue syntactic structure: 9-month-olds' sensitivity to phonological versus syntactic phrases. *Cognition, 51*, 237-265.

Gleitman, L. R. (1990). The Structural Sources of Verb Meanings. *Language Acquisition, 1*(1), 3-55.

Gleitman, L. R. & Wanner, E. (1982). Language Acquisition: the state of the state of the art. In Gleitman, L. R. & Wanner, E. (Eds.), *Language Acquisition - The State of the Art* (pp. 3-48). Cambridge: Cambridge University Press.

Goldman-Eisler, F. (1972). Pauses, Clauses, Sentences. *Language And Speech, 15*, 103-113.

Gout, A., Christophe, A. & Morgan, J. L. (2004). Phonological phrase boundaries constrain lexical access. II. Infant data. *Journal of Memory and Language, 51*, 548-567.

Grosjean, F. & Collins, M. (1979). Breathing, Pausing and Reading. *Phonetica, 36*, 98-114.

Grosjean, F., Grosjean, L. & Lane, H. (1979). The Patterns of Silence: Performance Structures in Sentence Production. *Cognitive Psychology, 11*, 58-81.

Guasti, M. T. (2004). *Language Acquisition. The Growth of Grammar* (PB ed.). Cambridge, MA.: The MIT Press.

Guasti, M. T., Nespor, M., Christophe, A. & Van Ooyen, B. (2001). Pre-lexical Setting of the Head Complement Parameter through Prosody. In Weissenborn, J. & Höhle, B. (Eds.), *Approaches to Bootstrapping: Phonological, Lexical, Syntactic and Neurophysiological Aspects of Early Language Acquisition* (Vol. 1, pp. 231-248). Amsterdam: Benjamins.

Haegeman, L. (1994). *Introduction to Government and Binding Theory* (2nd ed.). Oxford: Blackwell.

Hayashi, A. & Mazuka, R. (2002). *Developmental Change in Infants' Preferences for Continuousness in Japanese Speech*. Paper presented at the 13th Biennial International Conference on Infant Studies.

Hayashi, A., Tamekawa, Y., Deguchi, T. & Kiritani, S. (1996). *Developmental Change in Perception of Clause Boundaries by 6- and 10-Month-Old Japanese Infants.* Paper presented at the ICSLP 1996.

Hayashi, A., Tamekawa, Y. & Kiritani, S. (2001). Developmental Change in Auditory Preferences for Speech Stimuli in Japanese Infants. *Journal of Speech, Language and Hearing Research, 44*, 1189-1200.

Hirsh-Pasek, K. & Golinkoff, R. M. (1996). *The Origins of Grammar: Evidence from Early Language Comprehension* (PB ed.). Cambridge, MA.: The MIT Press.

Hirsh-Pasek, K., Kemler Nelson, D. G., Jusczyk, P. W., Cassidy, K., Druss, B. & Kennedy, L. (1987). Clauses are perceptual units for young infants. *Cognition, 26*, 269-286.

Höhle, B. (2002). *Der Einstieg in die Grammatik: Die Rolle der Phonologie/Syntax-Schnittstelle für Sprachverarbeitung und Spracherwerb.* Habilitationsschrift, Freie Universität, Berlin.

Höhle, B., Bijeljac-Babic, R., Herold, B., Weissenborn, J. & Nazzi, T. (submitted). Language specific prosodic preferences during the first half year of life: Evidence from German and French infants. *Cognitive Psychology*.

Höhle, B., Schmitz, M., Santelmann, L. M. & Weissenborn, J. (2006). The Recognition of Discontinuous Verbal Dependencies by German 19-Month-Olds: Evidence for Lexical and Structural Influences on Children's Early Processing Capacities. *Language Learning and Development, 2*(4), 277-300.

Höhle, B. & Weissenborn, J. (2003). German-learning infants' ability to detect unstressed closed-class elements in continuous speech. *Developmental Science, 6*(2), 122-127.

Höhle, B., Weissenborn, J., Kiefer, D., Schulz, A. & Schmitz, M. (2004). Functional Elements in Infants' Speech Processing: The role of Determiners in the Syntactic Categorization of Lexical Elements. *Infancy, 5*(3), 341-353.

Höhle, B., Weissenborn, J., Schmitz, M. & Ischebeck, A. (2001). Discovering Word-Order Regularities: The Role of Prosodic Information for Early Parameter Setting. In Weissenborn, J. & Höhle, B. (Eds.), *Approaches to Bootstrapping: Phonological, Lexical, Syntactic and Neurophysiological Aspects of Early Language Acquisition* (Vol. 1, pp. 249-265). Amsterdam: Benjamins.

Houston-Price, C. & Nakai, S. (2004). Distinguishing Novelty and Familiarity Effects in Infant Preference Procedures. *Infant and Child Development, 13*, 341-348.

Hunter, M. A. & Ames, E. W. (1988). A Multifactoral Model of Infant Preferences for Novel and Familiar Stimuli. In Lipsitt, L. P. & Rovee-Collier, C. (Eds.), *Advances in Infancy Research* (Vol. 5, pp. 69-95). Norwood, N.J.: Ablex.

Johnson, E. K. & Jusczyk, P. W. (2001). Word Segmentation by 8-Month-Olds: When Speech Cues Count More Than Statistics. *Journal of Memory and Language, 44*(4), 548-567.

Johnson, E. K. & Seidl, A. (2008). Clause Segmentation by 6-Month-Old Infants: A Crosslinguistic Perspective. *Infancy, 13*(5), 440-455.

Johnson, E. K., van de Weijer, J. & Jusczyk, P. W. (2001). Word Segmentation by 7.5-month-olds: Three Words do not Equal One. In Do, A. H.-J., Domínguez, L. & Johansen, A. (Eds.), *Proceedings of the 25th Annual Boston University Conference on Language Development* (Vol. 1, pp. 389-400). Somerville, MA.: Cascadilla Press.

Jusczyk, P. W. (1997). *The Discovery of Spoken Language*. Cambridge, MA.: The MIT Press.

Jusczyk, P. W. (1998). Dividing and Conquering Linguistic Input. In Gruber, M. C., Higgins, D., Olsen, K. S. & Wysocki, T. (Eds.), *CLS 34* (Vol. II: The Panels, pp. 293-310). Chicago, IL.: The Chicago Linguistic Society.

Jusczyk, P. W. (2003). American Infants' Perception of Cues to Grammatical Units in Non-native Languages and Music: Evidence from Polish and Japanese. In Houston, D. M., Seidl, A., Hollich, G. J., Johnson, E. K. & Jusczyk, A. M. (Eds.), *Jusczyk Lab Final Report.*: Retrieved from http://hincapie.psych.purdue.edu/Jusczyk.

Jusczyk, P. W. & Aslin, R. N. (1995). Infants' Detection of the Sound Patterns of Words in Fluent Speech. *Cognitive Psychology, 29*, 1-23.

Jusczyk, P. W., Cutler, A. & Redanz, N. J. (1993). Infants' Preference for the Predominant Stress Patterns of English Words. *Child Development, 64*, 675-687.

Jusczyk, P. W., Hirsh-Pasek, K., Kemler Nelson, D. G., Kennedy, L. J., Woodward, A. & Piwoz, J. (1992). Perception of Acoustic Correlates of Major Phrasal Units by Young Infants. *Cognitive Psychology, 24*, 252-293.

Jusczyk, P. W., Houston, D. M. & Newsome, M. (1999). The Beginnings of Word Segmentation in English-Learning Infants. *Cognitive Psychology, 39*, 159-207.

Katz, W. F., Beach, C. M., Jenouri, K. & Verma, S. (1996). Duration and fundamental frequency correlates of phrase boundaries in productions by children and adults. *Journal of the Acoustical Society of America, 99*(5), 3179-3191.

Kemler Nelson, D. G., Hirsh-Pasek, K., Jusczyk, P. W. & Wright Cassidy, K. (1989). How the prosodic cues in motherese might assist language learning. *Journal of Child Language, 16*, 55-68.

Kemler Nelson, D. G., Jusczyk, P. W., Mandel, D. R., Myers, J., Turk, A. & Gerken, L. (1995). The Head-Turn Preference Procedure for Testing Auditory Perception. *Infant Behavior and Development, 18*, 111-116.

Kohler, K. J. (1983). Prosodic Boundary Signals in German. *Phonetica, 40*, 89-134.

Krivokapić, J. (2007). Prosodic Planning: Effects of phrasal length and complexity on pause duration. *Journal of Phonetics, 35*, 162-179.

Lecanuet, J.-P. (1998). Foetal responses to auditory and speech stimuli. In Slater, A. (Ed.), *Perceptual development: Visual, auditory, and speech perception in infancy* (pp. 317-355). Hove: Psychology Press.

Lecanuet, J.-P. & Granier-Deferre, C. (1993). Speech Stimuli in the Fetal Environment. In de Boysson-Bardies, B., de Schonen, S., Jusczyk, P. W., McNeilage, P. & Morton, J. (Eds.), *Developmental Neurocognition: Speech and Face Processing in the First Year of Life* (pp. 237-248). Dordrecht, Boston, London: Kluwer Academic Publishers.

Lecanuet, J.-P., Granier-Deferre, C., Jacquet, A.-Y. & Busnel, M.-C. (1992). Decelerative Cardiac Responsiveness to Acoustical Stimulation in the Near Term Fetus. *The Quarterly Journal of Experimental Psychology, 44B*(3/4), 279-303.

Lecanuet, J.-P., Granier-Deferre, C., Jacquet, A.-Y., Capponi, I. & Ledru, L. (1993). Prenatal discrimination of a male and female voice uttering the same sentence. *Early Development and Parenting, 2*, 217-228.

Lehiste, I. (1977). Isochrony reconsidered. *Journal of Phonetics, 5*, 253-263.

Lehiste, I., Olive, J. P. & Streeter, L. A. (1976). Role of duration in disambiguating syntactically ambiguous sentences. *Journal of the Acoustical Society of America, 60*(5), 1199-1202.

Leuckefeld, K. (2001). *Prosodische Information und syntaktische Strukturierung: eine Studie zu ereigniskorrelierten Potentialen bei Kindern.* Diplomarbeit, Universität Potsdam, Potsdam.

Lieberman, P. (1966). *Intonation, Perception, and Language.* PhD dissertation, MIT, Cambridge, MA.

Lieberman, P. (1996). Some Bilogical Constraints on the Analysis of Prosody. In Morgan, J. L. & Demuth, K. (Eds.), *Signal to Syntax: Bootstrapping from Speech to Grammar in Early Acquisition* (pp. 55-65). Mahwah, N.J.: Lawrence Erlbaum.

Mandel, D. R., Jusczyk, P. W. & Kemler Nelson, D. G. (1994). Does sentential prosody help infants to organize and remember speech information? *Cognition, 53*, 155-180.

Mandel, D. R., Kemler Nelson, D. G. & Jusczyk, P. W. (1996). Infants Remember the Order of Words in a Spoken Sentence. *Cognitive Development, 11*, 181-196.

Männel, C. (p.c.). Telephone calls concerning her study investigating the occurrence of CPS in infants in the presence or absence of a pause at the prosodic boundary. December 2007; April 2008.

Maratsos, M. P. & Chalkley, M. A. (1980). The internal language of children's syntax: The ontogenesis and representation of syntactic categories. In Nelson, K. E. (Ed.), *Children's language* (Vol. 2, pp. 127-214). New York: Gardner.

Mehler, J. & Christophe, A. (1995). Maturation and Learning of Language in the First Year of Life. In Gazzaniga, M. S. (Ed.), *The Cognitive Neurosciences* (pp. 943-954). Cambridge, MA.: The MIT Press.

Mehler, J., Dupoux, E., Nazzi, T. & Dehaene-Lambertz, G. (1996). Coping with linguistic diversity: the infants' viewpoint. In Morgan, J. L. & Demuth, K. (Eds.), *Signal to Syntax: Bootstrapping from Speech to Grammar in Early Acquisition* (pp. 101-116). Mahwah, N.J.: Lawrence Erlbaum.

Mehler, J., Jusczyk, P. W., Lambertz, G., Halsted, N., Bertoncini, J. & Amiel-Tison, C. (1988). A precursor of language acquisition in young infants. *Cognition, 29*, 143-178.

Moon, C., Cooper, R. P. & Fifer, W. P. (1993). Two-day old infants prefer native language. *Infant Behavior and Development, 16*(4), 495-500.

Morgan, J. L. (1986). *From Simple Input to Complex Grammar*. Cambridge, MA: The MIT Press.

Morgan, J. L. (1994). Converging Measures of Speech Segmentation in Preverbal Infants. *Infant Behavior and Development, 17*, 389-403.

Morgan, J. L. & Demuth, K. (1996). Signal to Syntax: An Overview. In Morgan, J. L. & Demuth, K. (Eds.), *Signal to Syntax: Bootstrapping from Speech to Grammar in Early Acquisition* (pp. 1-22). Mahwah, N.J.: Lawrence Erlbaum.

Morgan, J. L., Meier, R. P. & Newport, E. L. (1987). Structural packaging in the input to language learning: Contributions of prosodic and morphological marking of phrases in the acquisition of language. *Cognitive Psychology, 19*, 498-550.

Müller, K. E. (1998). *German Focus Particles and their Influence on Intonation.* Diplomarbeit, Universität Stuttgart, Stuttgart.

Nazzi, T., Bertonici, J. & Mehler, J. (1998). Language Discrimination by Newborns: Toward an Understanding of the Role of Rhythm. *Journal of Experimental Psychology: Human Perception and Performance, 24*(3), 756-766.

Nazzi, T., Jusczyk, P. W. & Johnson, E. K. (2000a). Language Discrimination by English-Learning 5-Month-Olds: Effects of Rhythm and Familiarity. *Journal of Memory and Language, 43*(1), 1-19.

Nazzi, T., Kemler Nelson, D. G., Jusczyk, P. W. & Jusczyk, A. M. (2000b). Six-month-olds' Detection of Clauses Embedded in Continuous Speech: Effects of Prosodic Well-Formedness. *Infancy, 1*(1), 123-147.

Nespor, M., Guasti, M. T. & Christophe, A. (1996). Selecting Word Order: The Rhythmic Activation Principle. In Kleinhenz, U. (Ed.), *Interfaces in Phonology* (pp. 1-26). Berlin: Akademie Verlag.

Nespor, M. & Vogel, I. (1986). *Prosodic Phonology*. Dordrecht: Foris.

Newman, R. S. & Hussain, I. (2006). Changes in the Preference for Infant-Directed Speech in Low and Moderate Noise by 4.5- to 13-Month-Olds. *Infancy, 10*(1), 61-76.

Pannekamp, A., Toepel, U., Alter, K., Hahne, A. & Friederici, A. D. (2005). Prosody-driven Sentence Processing: An Event-related Brain Potential Study. *Journal of Cognitive Neuroscience, 17*(3), 407-421.

Pannekamp, A., Weber, C. & Friederici, A. D. (2006). Prosodic processing at the sentence level in infants. *NeuroReport, 17*(6), 675-678.

Pelzer, L. (2008). *Die Rolle von Konkordanzmarkierungen für die Segmentierung von Phrasen aus dem Sprachstrom - Untersuchungen bei Säuglingen und Erwachsenen* PhD Thesis. University of Potsdam, Potsdam: urn:nbn:de:kobv:517-opus-16687.

Pelzer, L. & Höhle, B. (2006a). The Impact of Morphological Markers in Infants' and Adults' Speech Processing. In Bamman, D., Magnitskaia, T. & Zaller, C. (Eds.), *BUCLD 30: Proceedings of the 30th annual Boston University Conference on Language Development* (Vol. 2, pp. 452-463). Somerville, MA: Cascadilla Press.

Pelzer, L. & Höhle, B. (2006b). Processing of morphological markers as a cue to syntactic phrases by 10-month-old German-learning infants. In Belletti, A., Bennati, E., Chesi, C., Di Domenico, E. & Ferrari, I. (Eds.), *Language Acquisition and Development. Proceedings of the GALA 2005 Conference* (pp. 427-439). Cambridge: Cambridge Scholars Press.

Pinker, S. (1984). *Language Learnability and Language Development*. Cambridge, MA.: Harvard University Press.

Pinker, S. (1987). The Bootstrapping Problem in Language Acquisition. In MacWhinney, B. (Ed.), *Mechanisms of Language Acquisition* (pp. 399-441). Hillsdale, N.J.: Lawrence Erlbaum.

Pinker, S. (1994). *The Language Instinct*. New York: HarperCollins Publishers.

Price, P. J., Ostendorf, M., Shattuck-Hufnagel, S. & Fong, C. (1991). The use of prosody in syntactic disambiguation. *Journal of the Acoustical Society of America, 90*(6), 2956-2970.

Ramus, F. (2002). Language discrimination by newborns: Teasing apart phonotactic, rhytmic and intonational cues. *Annual Review of Language Acquisition, 2*, 85-115.

Ramus, F., Dupoux, E. & Mehler, J. (2003). *The psychological reality of rhythm classes: perceptual studies.* Paper presented at the 15th International Congress of Phonetic Sciences, Barcelona.

Ramus, F., Nespor, M. & Mehler, J. (1999). Correlates of linguistic rhythm in the speech signal. *Cognition, 73*, 265-292.

Reich, S. S. (1980). Significance of Pauses for Speech Perception. *Journal of Psycholinguistic Research, 9*(4), 379-389.

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

Santelmann, L. M., Houston, D. M. & Jusczyk, P. W. (1997). 7.5-month-old Infants' Segmentation of Multisyllabic Words in Fluent Speech. In Hughes, E., Hughes, M. & Greenhill, A. (Eds.), *Proceedings of the 21st Annual Boston University Conference on Language Development* (Vol. 2, pp. 495-505). Somerville, MA.: Cascadilla Press.

Santelmann, L. M. & Jusczyk, P. W. (1998). Sensitivity to discontinuous dependencies in language learners: evidence for limitations in processing space. *Cognition, 69*, 105-134.

Schmitz, M., Höhle, B., van de Vijver, R. & Weissenborn, J. (submitted). The Edge Hypothesis Revisited: The Effect of Position on the Word Detection Skills in 9-Month-Old German-Learning Infants. *Journal of Child Language*.

Scott, D. R. (1982). Duration as a cue to the perception of a phrase boundary. *Journal of the Acoustical Society of America, 71*(4), 996-1007.

Seidl, A. (2007). Infants' use and weighting of prosodic cues in clause segmentation. *Journal of Memory and Language, 57*(1), 24-48.

Seidl, A. (p.c.). E-mail answer to my question concerning the ISI used in the familiarisation phase. December 6, 2007.

Seidl, A. & Cristià, A. (2008). Developmental changes in the weighting of prosodic cues. *Developmental Science, 11*(4), 596-606.

Seidl, A. & Johnson, E. K. (2006). Infant word segmentation revisited: edge alignment facilitates target extraction. *Developmental Science, 9*(6), 565-573.

Seidl, A. & Johnson, E. K. (2008). Boundary alignment enables 11-month-olds to segment vowel initial words from speech. *Journal of Child Language, 35*, 1-24.

Shady, M. (1996). *Infants' senisivity to function morphemes.* Doctoral Dissertation, State University of New York at Buffalo, Buffalo, N.Y.

Shi, R., Cutler, A., Werker, J. F. & Cruickshank, M. (2006). Frequency and form as determinants of functor sesitivity in English-acquiring infants. *Journal of the Acoustical Society of America, 119*(6), EL61-EL67.

Shi, R., Marquis, A. & Gauthier, B. (2006). Segmentation and representation of function words in preverbal French-learning infants. In Bamman, D., Magnitskaia, T. & Zaller, C. (Eds.), *Proceedings of the 30th Annual Boston University Conference on Language Development* (Vol. 2, pp. 549-560). Somerville, MA: Cascadilla Press.

Shi, R., Morgan, J. L. & Allopenna, P. (1998). Phonological and acoustic bases for earliest grammatical category assignment: a cross-linguistic perspective. *Journal of Child Language, 25*, 169-201.

Shi, R., Werker, J. F. & Cutler, A. (2003). Function Words in Early Speech Perception. In Solé, M. J., Recasens, D. & Romero, J. (Eds.), *Proceedings of the 15th International Congress of Phonetic Scienes* (CD-ROM ed., pp. 3009-3012). Barcelona, Spain: Casual Productions.

Soderstrom, M. (2007). Beyond babytalk: Re-evaluating the nature and content of speech input to preverbal infants. *Developmental Review, 27*, 501–532.

Soderstrom, M., Blossom, M., Foygel, R. & Morgan, J. L. (2008). Acoustical cues and grammatical units in speech to two preverbal infants. *Journal of Child Language, 35*, 869-902.

Soderstrom, M., Kemler Nelson, D. G. & Jusczyk, P. W. (2005). Six-month-olds recognize clauses embedded in different passages of fluent speech. *Infant Behavior and Development, 28*, 87-94.

Soderstrom, M., Seidl, A., Kemler Nelson, D. G. & Jusczyk, P. W. (2003). The prosodic bootstrapping of phrases: Evidence from prelinguistic infants. *Journal of Memory and Language, 49*(2), 249-267.

Sokolov, E. N. (1963). *Perception and the conditioned reflex*. New York: Pergamon Press.

Sokolov, E. N. (1969). The modeling properties of the nervous system. In Coles, M. & Maltzman, I. (Eds.), *A handbook of contemporary Soviet psychology*. New York: Basic Books.

Steinhauer, K., Alter, K. & Friederici, A. D. (1999). Brian potentials indicate immediate use of prosodic cues in natural speech processing. *Nature Neuroscience, 2*(2), 191-196.

Streeter, L. A. (1978). Acoustic determinants of phrase boundary perception. *Journal of the Acoustical Society of America, 64*(4), 1582-1592.

Stuckenberg, A. & O'Connell, D. C. (1988). The Long and Short of It: Reports of Pause Occurrence and Duration in Speech. *Journal of Psycholinguistic Research, 17*(1), 19-28.

Thiessen, E. D., Hill, E. A. & Saffran, J. R. (2005). Infant-Directed Speech Facilitates Word Segmentation. *Infancy, 7*(1), 53-71.

Trainor, L. J., Samuel, S. S., Desjardins, R. N. & Sonnadara, R. R. (2001). Measuring temporal resolution in infants using mismatch negativity. *NeuroReport, 12*(11), 2443-2448.

Trueswell, J. C., Sekerina, I., Hill, N. M. & Logrip, M. L. (1999). The kindergarten-path effect: studying on-line sentence processing in young children. *Cognition, 73*, 89-134.

van de Weijer, J. (1998). *Language Input for Word Discovery*. Wageningen: Ponsen & Looijen.

Wartenburger, I., Steinbrink, J., Telkemeyer, S., Friedrich, M., Friederici, A. D. & Obrig, H. (2007). The processing of prosody: Evidence for interhemsipheric specialization at the age of four. *NeuroImage, 34*, 416-425.

Werker, J. F., Polka, L. & Pegg, J. E. (1997). The Conditioned Head Turn Procedure as a Method for Testing Infant Speech Perception. *Early Development and Parenting, 6*, 171-178.

Werker, J. F., Shi, R., Desjardins, R., Pegg, J. E., Polka, L. & Patterson, M. (1998). Three methods for testing infant speech perception. In Slater, A. (Ed.), *Perceptual Development: Visual, auditory, and speech perception in infancy* (pp. 389-420). Hove: Psychology Press.

Werker, J. F. & Tees, R. C. (1984). Cross-Language Speech Perception: Evidence for Perceptual Reorganization During the First Year of Life. *Infant Behavior and Development, 7*, 49-63.

Werner, L. A., Marean, G. C., Halpin, C. F., Spetner, N. B. & Gillenwater, J. M. (1992). Infant Auditory Temporal Acuity: Gap Detection. *Child Development, 63*, 260-272.

Wightman, C. W., Shattuck-Hufnagel, S., Ostendorf, M. & Price, P. J. (1992). Segmental Duration in the vicinity of prosodic phrase boundaries. *Journal of the Acoustical Society of America, 91*(3), 1707-1717.

Zangl, R. & Mills, D. L. (2007). Increased Brain Activity to Infant-Directed Speech in 6- and 13-Month-Old Infants. *Infancy, 11*(1), 31-62.

Appendices

Appendix A: The sentences

A.1 Subject Control Sentences

Sentence Number	Sentence	Duration of the original sentences (ms)	Pause duration between the clauses (ms)
S1Katrin	*Die Tante verspricht Katrin zu schwimmen und die Insel zu erkunden.* 'The aunt promises Katrin to swim and to explore the island.'[51]	4857	454
S2Jochen	*Der Sportler verspricht Jochen zu tauchen und die Uhr zu besorgen.* 'The athlete promises Jochen to dive and to get the watch.'	4643	540
S3Anja	*Der Bäcker verspricht Anja zu lächeln und die Sache zu vergessen.* 'The baker promises Anja to smile and to forget about it.'	4454	406
S4Jana	*Die Freundin verspricht Jana zu lachen und den Ärger zu verbergen.* 'The friend promises Jana to laugh and to hide the annoyance.'	4720	482
S5Lena	*Die Schwester verspricht Lena zu schlafen und die Tür zu schließen.* 'The sister promises Lena to sleep and to close the door.'	4668	451
S6Daniel	*Der Doktor verspricht Daniel zu flüstern und das Fieber zu messen.* 'The doctor promises Daniel to whisper and to take the (his) temperature.'	4582	368

[51] The translation of the sentences is a more or less literal one to illustrate the syntactic structure of the German examples.

Sentence Number	Sentence	Duration of the original sentences (ms)	Pause duration between the clauses (ms)
S7Peter	*Die Köchin verspricht Peter zu gehen und den Einkauf zu erledigen.* ‘The cook promises Peter to go and to do the hopping.’	4771	512
S8Lukas	*Die Putzfrau verspricht Lukas zu warten und den Bus zu nehmen.* The cleaning woman promises Lucas to wait and to take the bus.’	4503	430
S9Anke	*Der Nachbar droht Anke zu schreien und die Trompete zu blasen.* ‘The neighbour threatens Anke to shout and to play the trumpet.’	4759	559
S10Felix	*Die Oma droht Felix zu jammern und die Stimmung zu verderben.* ‘The grandmother threatens Felix to whimper and to spoil the atmosphere.’	4477	458
S11Sonja	*Der Onkel droht Sonja zu lärmen und das Bild zu zerreißen.* ‘The uncle threatens Sonja to be noisy and to tear apart the picture.’	4472	457
S12Katja	*Die Tante droht Katja zu schummeln und die Regeln zu missachten.* ‘The aunt threatens Katja to cheat and to ignore the rules.’	4641	455
S13Markus	*Der Cousin droht Markus zu verschwinden und das Fahrrad zu stehlen.* ‘The cousin threatens Markus to disappear and to steal the bicycle.’	4931	505
S14Stefan	*Der Sportler droht Stefan zu plaudern und das Geheimnis zu erzählen.* ‘The athlete threatens Stefan to chat and to relate the secret.’	5139	467

Sentence Number	Sentence	Duration of the original sentences (ms)	Pause duration between the clauses (ms)
S15Robert	*Der Opa droht Robert zu klingeln und den Brief zu übergeben.* 'The grandfather threatens Robert to ring and to hand over the letter.'	4680	556
S16Anna	*Der Gärtner droht Anna zu hupen und die Ruhe zu stören.* 'The gardener threatens Anna to honk and to disrupt the silence.'	4346	445
S17Carsten	*Der Nachbar schwört Carsten zu schweigen und den Namen zu verheimlichen.* 'The neighbour vows (to) Carsten to remain silent and to keep the name secret.'	4940	432
S18Boris	*Der Vater schwört Boris zu kommen und den Sieg zu feiern.* 'The father vows (to) Boris to come and to celebrate the victory.'	4427	517
S19Holger	*Die Mutter schwört Holger zu überlegen und das Gedicht zu schreiben.* 'The mother vows (to) Holger to think about it and to write the poem.'	4718	443
S20Dagmar	*Der Cousin schwört Dagmar zu kämpfen und die Brücke zu überqueren.* 'The cousin vows (to) Dagmar to fight and to cross the bridge.'	4749	543
S21Ellen	*Der Bruder schwört Ellen zu schleichen und die Angst zu überwinden.* 'The brother vows (to) Ellen to sneak and to overcome the fear.'	4835	490
S22Lisa	*Der Trainer schwört Lisa zu jubeln und die Fahne zu schwenken.* 'The trainer vows (to) Lisa to cheer and to wave the flag.'	4751	524

Sentence Number	Sentence	Duration of the original sentences (ms)	Pause duration between the clauses (ms)
S23Dennis	*Die Oma schwört Dennis zu tanzen und den Abend zu genießen.* ‘The grandmother vows (to) Dennis to dance and to enjoy the evening.’	4926	567
S24Hanna	*Der Sheriff schwört Hanna zu rennen und das Feuer zu löschen.* ‘The sheriff vows /to) Hanna to run and to extinguish the fire.’	4598	630

A.2 Object Control Sentences

Sentence Number	Sentence	Duration of the original sentences (ms)	Pause duration between the clauses (ms)
S25Carmen	*Der Lehrer verbietet Carmen zu verschlafen und die Schule zu schwänzen.* 'The teacher forbids Carmen to oversleep and to play truant.'	5134	569
S26Christoph	*Die Ärztin verbietet Christoph zu quengeln und die Schokolade zu essen.* 'The doctor forbids Christoph to whine and to eat the chocolate.'	5145	553
S27Nico	*Die Köchin verbietet Nico zu schwindeln und den Keks zu nehmen.* 'The cook forbids Nico to swindle and to take the cookie.'	4840	421
S28Kerstin	*Die Putzfrau verbietet Kerstin zu nörgeln und den Rock zu zerknittern.* 'The cleaning woman forbids Kerstin to nag and to crumple the skirt.'	5181	537
S29Petra	*Der Sheriff verbietet Petra zu kriechen und die Höhle zu betreten.* 'The sheriff forbids Petra to crawl and to enter the cave.'	5075	595
S30Martin	*Der Kellner verbietet Martin zu kleckern und den Saft zu verschütten.* 'The waiter forbids Martin to make a mess and to spill the juice.'	4988	552
S31Eva	*Der Gärtner verbietet Eva zu krabbeln und die Erdbeeren zu pflücken.* 'The gardener forbids Eva to crawl and to pick the strawberries.'	5158	582
S32Philipp	*Der Cowboy verbietet Philipp zu zelten und das Boot zu mieten.* 'The cowboy forbids Philipp to camp and to hire the boat.'	4933	495

Sentence Number	Sentence	Duration of the original sentences (ms)	Pause duration between the clauses (ms)
S33Elke	*Der Vater erlaubt Elke zu trödeln und die Cola zu trinken.* 'The father allows Elke to dawdle and to drink the coke.'	4714	498
S34Ingo	*Der Onkel erlaubt Ingo zu faulenzen und das Buch zu lesen.* 'The uncle allows Ingo to be idle and to read the book.'	4799	478
S35Heike	*Der Trainer erlaubt Heike zu turnen und das Training zu beginnen.* 'The trainer allows Heike to do gymnastics and to start the training.'	4699	477
S36Judith	*Der Lehrer erlaubt Judith zu laufen und das Brot zu kaufen.* 'The teacher allows Judith to run and to buy the (loaf of) bread.'	4610	474
S37Christian	*Der Torwart erlaubt Christian zu hüpfen und den Ball zu werfen.* 'The keeper allows Christian to hop and to throw the ball.'	4897	635
S38Moritz	*Die Ärztin erlaubt Moritz zu flitzen und das Medikament zu besorgen.* 'The doctor allows Moritz to scamper and to procure the medicine.'	5170	461
S39Sandra	*Der Doktor erlaubt Sandra zu reden und das Geschenk zu überreichen.* 'The doctor allows Sandra to talk and to hand over the present.'	5086	612
S40Ulrich	*Der Opa erlaubt Ulrich zu bleiben und den Film zu gucken.* 'The grandfather allows Ulrich to stay and to watch the movie.'	4849	701

Sentence Number	Sentence	Duration of the original sentences (ms)	Pause duration between the clauses (ms)
S41Heiko	*Der Cowboy hilft Heiko zu klettern und den Berg zu besteigen.* ‘The cowboy helps Heiko to scramble and to climb the mountain.’	4851	555
S42Torsten	*Der Torwart hilft Torsten zu siegen und die Fahne zu hissen.* ‘The keeper helps Thorsten to win and to fly the flag.’	4574	594
S43Tanja	*Die Schwester hilft Tanja zu mogeln und das Spiel zu gewinnen.* ‘The sister helps Tanja to cheat and to win the game.’	4681	583
S44Sascha	*Der Kellner hilft Sascha zu rechnen und das Geld zu zählen.* ‘The waiter helps Sascha to calculate and to count the money.’	4649	630
S45Silke	*Die Mutter hilft Silke zu stehen und den Zopf zu flechten.* ‘The mother helps Silke to stand and to braid the hair.’	4477	489
S46Thomas	*Die Freundin hilft Thomas zu springen und das Baumhaus zu erreichen.* ‘The friend helps Thomas to jump and to reach the tree house.’	5358	750
S47Saskia	*Der Bruder hilft Saskia zu balancieren und das Gleichgewicht zu halten.* ‘The brother helps Saskia to poise and to keep her balance.’	5184	518
S48Tina	*Der Bäcker hilft Tina zu arbeiten und den Kuchen zu backen.* ‘The baker helps Tina to work and to bake the cake.’	4796	643

Appendix B: Grouping of the sentences for the experiments

Block	**Sentence Number**	**Sentence**
Block 1 (Training)	S3Anja	Der Bäcker verspricht Anja zu lächeln und die Sache zu vergessen.
	S13Markus	Der Cousin droht Markus zu verschwinden und das Fahrrad zu stehlen.
	S19Holger	Die Mutter schwört Holger zu überlegen und das Gedicht zu schreiben.
	S11Sonja	Der Onkel droht Sonja zu lärmen und das Bild zu zerreißen.
	S7Peter	Die Köchin verspricht Peter zu gehen und den Einkauf zu erledigen.
	S21Ellen	Der Bruder schwört Ellen zu schleichen und die Angst zu überwinden.
Block 2 (Training)	S28Kerstin	Die Putzfrau verbietet Kerstin zu nörgeln und den Rock zu zerknittern.
	S34Ingo	Der Onkel erlaubt Ingo zu faulenzen und das Buch zu lesen.
	S45Silke	Die Mutter hilft Silke zu stehen und den Zopf zu flechten.
	S38Moritz	Die Ärztin erlaubt Moritz zu flitzen und das Medikament zu besorgen.
	S31Eva	Der Gärtner verbietet Eva zu krabbeln und die Erdbeeren zu pflücken.
	S42Torsten	Der Torwart hilft Torsten zu siegen und die Fahne zu hissen.
Block 3 (Test)	S1Katrin	Die Tante verspricht Katrin zu schwimmen und die Insel zu erkunden.
	S2Jochen	Der Sportler verspricht Jochen zu tauchen und die Uhr zu besorgen.
	S4Jana	Die Freundin verspricht Jana zu lachen und den Ärger zu verbergen.
	S6Daniel	Der Doktor verspricht Daniel zu flüstern und das Fieber zu messen.
	S5Lena	Die Schwester verspricht Lena zu schlafen und die Tür zu schließen.
	S8Lukas	Die Putzfrau verspricht Lukas zu warten und den Bus zu nehmen.

Block	Sentence Number	Sentence
Block 4 (Test)	S9Anke	Der Nachbar droht Anke zu schreien und die Trompete zu blasen.
	S10Felix	Die Oma droht Felix zu jammern und die Stimmung zu verderben.
	S12Katja	Die Tante droht Katja zu schummeln und die Regeln zu missachten.
	S14Stefan	Der Sportler droht Stefan zu plaudern und das Geheimnis zu erzählen.
	S15Robert	Der Opa droht Robert zu klingeln und den Brief zu übergeben.
	S16Anna	Der Gärtner droht Anna zu hupen und die Ruhe zu stören.
Block 5 (Test)	S17Carsten	Der Nachbar schwört Carsten zu schweigen und den Namen zu verheimlichen.
	S18Boris	Der Vater schwört Boris zu kommen und den Sieg zu feiern.
	S20Dagmar	Der Cousin schwört Dagmar zu kämpfen und die Brücke zu überqueren.
	S22Lisa	Der Trainer schwört Lisa zu jubeln und die Fahne zu schwenken.
	S23Dennis	Die Oma schwört Dennis zu tanzen und den Abend zu genießen.
	S24Hanna	Der Sheriff schwört Hanna zu rennen und das Feuer zu löschen.
Block 6 (Test)	S25Carmen	Der Lehrer verbietet Carmen zu verschlafen und die Schule zu schwänzen.
	S26Christoph	Die Ärztin verbietet Christoph zu quengeln und die Schokolade zu essen.
	S27Nico	Die Köchin verbietet Nico zu schwindeln und den Keks zu nehmen.
	S29Petra	Der Sheriff verbietet Petra zu kriechen und die Höhle zu betreten.
	S30Martin	Der Kellner verbietet Martin zu kleckern und den Saft zu verschütten.
	S32Philipp	Der Cowboy verbietet Philipp zu zelten und das Boot zu mieten.

Block	Sentence Number	Sentence
Block 7 (Test)	S33Elke	Der Vater erlaubt Elke zu trödeln und die Cola zu trinken.
	S40Ulrich	Der Opa erlaubt Ulrich zu bleiben und den Film zu gucken.
	S35Heike	Der Trainer erlaubt Heike zu turnen und das Training zu beginnen.
	S36Judith	Der Lehrer erlaubt Judith zu laufen und das Brot zu kaufen.
	S37Christian	Der Torwart erlaubt Christian zu hüpfen und den Ball zu werfen.
	S39Sandra	Der Doktor erlaubt Sandra zu reden und das Geschenk zu überreichen.
Block 8 (Test)	S41Heiko	Der Cowboy hilft Heiko zu klettern und den Berg zu besteigen.
	S43Tanja	Die Schwester hilft Tanja zu mogeln und das Spiel zu gewinnen.
	S44Sascha	Der Kellner hilft Sascha zu rechnen und das Geld zu zählen.
	S46Thomas	Die Freundin hilft Thomas zu springen und das Baumhaus zu erreichen.
	S47Saskia	Der Bruder hilft Saskia zu balancieren und das Gleichgewicht zu halten.
	S48Tina	Der Bäcker hilft Tina zu arbeiten und den Kuchen zu backen.

Appendix C: Duration (ms) of the sentences in each experiment

	Exp.1 - natural condition	Exp. 1 - unnatural condition	Exp. 2 - natural condition	Exp. 2 - artifiical condition	Exp. 3 - natural condition	Exp. 3 - unnatural condition	Exp. 4 & 5 - natural pause condition	Exp. 4 & 5 - reversed pause condition
S1Katrin	5,403	5,503	5,153	5,253	5,153	5,607	5,153	5,403
S2Jochen	5,102	5,125	4,852	4,875	4,852	5,393	4,852	5,102
S3Anja	5,048	5,148	4,798	4,898	4,798	5,204	4,798	5,048
S4Jana	5,237	5,338	4,987	5,088	4,987	5,470	4,987	5,237
S5Lena	5,216	5,317	4,966	5,067	4,966	5,418	4,966	5,216
S6Daniel	5,214	5,314	4,964	5,064	4,964	5,332	4,964	5,214
S7Peter	5,259	5,302	5,009	5,052	5,009	5,521	5,009	5,259
S8Lukas	5,073	5,173	4,823	4,923	4,823	5,253	4,823	5,073
S9Anke	5,199	5,300	4,949	5,050	4,949	5,509	4,949	5,199
S10Felix	5,019	5,116	4,769	4,866	4,769	5,227	4,769	5,019
S11Sonja	5,015	5,115	4,765	4,865	4,765	5,222	4,765	5,015
S12Katja	5,185	5,286	4,935	5,036	4,935	5,391	4,935	5,185
S13Markus	5,426	5,526	5,176	5,276	5,176	5,681	5,176	5,426
S14Stefan	5,671	5,684	5,421	5,434	5,421	5,889	5,421	5,671
S15Robert	5,124	5,162	4,874	4,912	4,874	5,430	4,874	5,124
S16Anna	4,900	5,001	4,650	4,751	4,650	5,096	4,650	4,900
S17Carsten	5,508	5,608	5,258	5,358	5,258	5,690	5,258	5,508
S18Boris	4,910	4,931	4,660	4,681	4,660	5,177	4,660	4,910
S19Holger	5,274	5,375	5,024	5,125	5,024	5,468	5,024	5,274
S20Dagmar	5,205	5,239	4,955	4,989	4,955	5,499	4,955	5,205
S21Ellen	5,345	5,445	5,095	5,195	5,095	5,585	5,095	5,345
S22Lisa	5,227	5,327	4,977	5,077	4,977	5,501	4,977	5,227
S23Dennis	5,359	5,386	5,109	5,136	5,109	5,676	5,109	5,359
S24Hanna	4,968	5,068	4,718	4,818	4,718	5,348	4,718	4,968
S25Carmen	5,565	5,665	5,315	5,415	5,315	5,884	5,315	5,565
S26Christoph	5,592	5,599	5,342	5,349	5,342	5,895	5,342	5,592
S27Nico	5,418	5,519	5,168	5,269	5,168	5,590	5,168	5,418
S28Kerstin	5,644	5,744	5,394	5,494	5,394	5,931	5,394	5,644
S29Petra	5,479	5,507	5,229	5,257	5,229	5,825	5,229	5,479
S30Martin	5,436	5,475	5,186	5,225	5,186	5,738	5,186	5,436
S31Eva	5,575	5,603	5,325	5,353	5,325	5,908	5,325	5,575
S32Philipp	5,438	5,481	5,188	5,231	5,188	5,683	5,188	5,438
S33Elke	5,215	5,245	4,965	4,995	4,965	5,464	4,965	5,215
S34Ingo	5,320	5,421	5,070	5,171	5,070	5,549	5,070	5,320
S35Heike	5,222	5,233	4,972	4,983	4,972	5,449	4,972	5,222
S36Judith	5,135	5,236	4,885	4,986	4,885	5,360	4,885	5,135
S37Christian	5,262	5,362	5,012	5,112	5,012	5,647	5,012	5,262
S38Moritz	5,708	5,809	5,458	5,559	5,458	5,920	5,458	5,708
S39Sandra	5,474	5,550	5,224	5,300	5,224	5,836	5,224	5,474
S40Ulrich	5,148	5,173	4,898	4,923	4,898	5,599	4,898	5,148
S41Heiko	5,296	5,336	5,046	5,086	5,046	5,601	5,046	5,296

	Exp.1 - natural condition	Exp. 1 - unnatural condition	Exp. 2 - natural condition	Exp. 2 - artifiical condition	Exp. 3 - natural condition	Exp. 3 - unnatural condition	Exp. 4 & 5 - natural pause condition	Exp. 4 & 5 - reversed pause condition
S42Torsten	4,979	5,079	4,729	4,829	4,729	5,324	4,729	4,979
S43Tanja	5,098	5,198	4,848	4,948	4,848	5,431	4,848	5,098
S44Sascha	5,018	5,119	4,768	4,869	4,768	5,399	4,768	5,018
S45Silke	4,987	5,088	4,737	4,838	4,737	5,227	4,737	4,987
S46Thomas	5,608	5,708	5,358	5,458	5,358	6,108	5,358	5,608
S47Saskia	5,665	5,737	5,415	5,487	5,415	5,934	5,415	5,665
S48Tina	5,153	5,173	4,903	4,923	4,903	5,546	4,903	5,153

Appendix D: Mean duration (ms) of the sample passages in both conditions and order of presentation of the blocks in the experiments

			natural condition		unnatural/reversed pause condition	
			Order of Presentation 1	Order of Presentation 2	Order of Presentation 1	Order of Presentation 2
Trial 1	Name Wave		**Korr1**	**Korr1**	**Inkorr1**	**Inkorr1**
	Duration (ms)	Experiment 1	38,871	38,871	39,411	39,411
		Experiment 2	34,871	34,871	35,411	35,411
		Experiment 3	35,070	35,070	37,824	37,824
			Ko75011	**Ko75011**	**Ko17501**	**Ko17501**
		Experiment 4 & 5	34,871	34,871	35,121	35,121
Trial 2	Name Wave		**Korr2**	**Korr2**	**Inkorr2**	**Inkorr2**
	Duration (ms)	Experiment 1	39,717	39,717	40,244	40,244
		Experiment 2	35,717	35,717	36,244	36,244
		Experiment 3	35,916	35,916	38,986	38,986
			Ko75012	**Ko75012**	**Ko17502**	**Ko17502**
		Experiment 4 & 5	35,717	35,717	35,967	35,967
Trial 3	Name Wave		**Korr6**	**Korr4**	**Inkorr4**	**Inkorr6**
	Duration (ms)	Experiment 1	40,433	38,602	39,049	40,746
		Experiment 2	36,433	34,602	35,049	36,745
		Experiment 3	36,632	34,801	37,589	39,531
			Ko75016	**Ko75014**	**Ko17504**	**Ko17506**
		Experiment 4 & 5	36,433	34,602	34,852	36,683

			natural condition		unnatural/reversed pause condition	
			Order of Presentation 1	Order of Presentation 2	Order of Presentation 1	Order of Presentation 2
Trial 4	Name Wave		**Korr8**	**Korr6**	**Inkorr6**	**Inkorr8**
	Duration (ms)	Experiment 1	39,342	40,433	40,746	39,772
		Experiment 2	35,342	36,433	36,746	35,772
		Experiment 3	35,542	36,632	39,531	39,051
			Ko75018	**Ko75016**	**Ko17506**	**Ko17508**
		Experiment 4 & 5	35,342	36,433	36,683	35,592
Trial 5	Name Wave		**Korr3**	**Korr8**	**Inkorr8**	**Inkorr3**
	Duration (ms)	Experiment 1	38,749	39,342	39,772	39,272
		Experiment 2	34,749	35,342	35,772	35,272
		Experiment 3	34,948	35,542	39,051	37,596
			Ko75013	**Ko75018**	**Ko17058**	**Ko17503**
		Experiment 4 & 5	34,749	39,342	35,592	34,999
Trial 6	Name Wave		**Korr5**	**Korr3**	**Inkorr3**	**Inkorr5**
	Duration (ms)	Experiment 1	38,678	38,749	39,272	39,060
		Experiment 2	34,678	34,749	35,272	35,060
		Experiment 3	34,878	34,948	37,596	37,873
			Ko75015	**Ko75013**	**Ko17503**	**Ko17505**
		Experiment 4 & 5	34,678	34,749	34,999	34,928
Trial 7	Name Wave		**Korr4**	**Korr7**	**Inkorr7**	**Inkorr4**
	Duration (ms)	Experiment 1	38,602	38,959	39,299	39,049
		Experiment 2	34,602	34,959	35,299	35,049
		Experiment 3	34,801	35,159	38,295	37,589
			Ko75014	**Ko75017**	**Ko17507**	**Ko17504**
		Experiment 4 & 5	34,602	34,960	35,210	34,852

			natural condition		unnatural/reversed pause condition	
			Order of Presentation 1	Order of Presentation 2	Order of Presentation 1	Order of Presentation 2
Trial 8	Name Wave		**Korr7**	**Korr5**	**Inkorr5**	**Inkorr7**
	Duration (ms)	Experiment 1	38,959	38,678	39,060	39,299
		Experiment 2	34,959	34,678	35,060	35,299
		Experiment 3	35,159	34,878	37,873	38,295
			Ko75017	**Ko75015**	**Ko1055**	**Ko1057**
		Experiment 4 & 5	34,960	34,678	33,678	33,960