Claudia Friedrich: Prosody and Spoken Word Recognition: Behavioral and ERP Correlates. Leipzig: Max Planck Institute of Cognitive Neuroscience, 2003 (MPI Series in Cognitive Neuroscience; 36)

Prosody and spoken word recognition:

Behavioral and ERP correlates

Von der Fakultät für Biowissenschaften, Pharmazie und Psychologie
der Universität Leipzig
genehmigte
DISSERTATION

zur Erlangung des akademischen Grades
DOCTOR RERUM NATURALIUM
(Dr. rer. nat.)

vorgelegt
von Diplom-Psychologin Claudia Katrin Friedrich
geboren am 25. Mai 1974 in Berlin

Dekan:
Prof. Dr. Martin Schlegel

Gutachter
Prof. Dr. Angela D. Friederici
Prof. Dr. Anne Cutler
Prof. Dr. Erich Schröger

Tag der Verteidigung: 27. Februar 2003

Vorwort

Mein Interesse an der Aufzeichung subtiler elektrischer Aktivität des Gehirns führte mich an das Max-Planck-Institut für neuropsychologische Forschung in Leipzig. Die Anwendung dieser Methode zur Erforschung von Prozessen, die dem Verstehen gesprochener Sprache zu Grunde liegen, brachte mir Prof. Angela D. Friederici durch begeisternde Vorträge und Diskussionen nahe. Ihre inhaltlichen Anregungen und ihre Unterstützung haben zu der vorliegenden Arbeit wesentlich beigetragen. Meine Faszination für frühe Mechanismen der auditiven Worterkennung wurde v.a. durch Befunde der Arbeitsgruppe von Prof. Anne Cutler vom Max-Planck-Institut für Psycholinguistik in Nijmegen geweckt. Ich freue mich, dass ich Prof. Anne Cutler und Prof. Angela D. Friederici, ebenso wie Prof. Erich Schröger von der Universität Leipzig als Gutachter für die Dissertation gewinnen konnte.

Motiviert und begleitet wurden die vorliegenden Experimente von Dr. Sonja A. Kotz und Dr. Kai Alter. Gemeinsam versuchten wir verschiedene Probleme zu lösen, die bei der Untersuchung von Prozessen der auditiven Worterkennung mit der elektrophysiologischen Methode auftraten. Die hieraus resultierenden Erfahrungen und Erkenntnisse haben nicht nur meine Arbeit, sondern auch meine persönliche Entwicklung bereichert.

Die vorliegende Dissertation wäre in dieser Form nicht ohne den methodisch-kritischen Blick entstanden, den ich mir während meines Studiums an der Technischen Universität Berlin bei Prof. Gisela Erdmann angeeignet habe. Ebenso waren die Diskussionen mit Dr. Monica de Filippis, Dr. Thomas C. Gunter und Dipl. Psych. Annett Schirmer am Max-Planck-Institut in Leipzig wegweisend für die Konzeption und Interpretation der von mir durchgeführten Experimente. Schließlich gilt mein Dank Cornelia Schmidt und Sylvia Stasch für ihre Unterstützung bei der Aufnahme der Daten für die EKP-Studien II, III und IV. Die Dissertation wurde von der Deutschen Forschungsgemeinschaft im Rahmen des Projektes FR 519/17-2 gefördert.

Ich danke meiner Mutter für die unverzichtbare emotionale Unterstützung, die sie mir voraussetzungsfrei bietet, meinen Großeltern für das unerschütterliche Vertrauen in meine Fähigkeiten, allen Freunden in Berlin und Leipzig für die Erdung zum normalen Leben und Frank, dafür dass er mich nach Leipzig begleitet hat und für sein Augenzwinkern im richtigen Moment.

Contents

Chapter 1

Introduction

An autopilot with no prosody implemented might confuse a German driver with the statement *’Das Denkmal müssen Sie umfahren.’*. There are two alternatives a human speaker can pronounce the last word of such an utterance: If the second syllable of *umfahren* is stressed it tells the driver to go around a monument. If, however, the first syllable of *umfahren* is stressed, it would advise the driver to run over the monument. Recent empirical evidence suggests that word prosody is an important aspect of spoken word recognition. The first part of the introduction provides an overview of what is currently known about how prosodic information is used during spoken word recognition. Thereafter, the recording of Event-Related brain Potentials (ERPs) and the correlates of spoken word processing in the ERP signal are addressed, as one major goal of the current thesis is to relate the processing of word prosody to the electrophysiological response of the brain. Electrophysiological correlates of the processing of word prosody might allow to draw a more detailed picture on the time course of spoken word processing as it is provided so far by behavioral methods.

1.1 Word prosody

The term *prosody* comes from ancient Greek where it was used for a *song sung with instrumental music*. The term is now used to refer to those properties of speech that cannot be derived from the sequence of phonemes underlying human utterances. Phonemes are referred to as a words segmental information, whereas prosody is alternatively labeled as the *suprasegmental* information of a word. Typically, features of pitch, loudness and duration are relevant here (e.g. Clark & Yallop, 1995). They are used to mark the organization of syllables within words. However, prosodic features are not only relevant at the word level. They also serve different functions on sentence or discourse level, where they are used, for instance, to structure phrases and to encode focus within sentences (see Cutler, Dahan & van Donselaar, 1997, for a recent review) or to convey emotions (see Schirmer, Kotz & Friederici, 2001, for recent empirical work on this issue).

The function of prosodic information in spoken word recognition has been investigated in a variety of languages including English (Coninne, Clifton & Cutler, 1987; Cooper, Cutler & Wales, 2002; Cutler, 1986), Dutch (Cutler & van Donselaar, 2001; van Donselaar, Koster & Cutler, submitted), Spanish (Soto-Faraco, Sebastian-Galles & Cutler, 2001), Japanese (Cutler & Otake, 1999; Sekiguchi & Nakajima, 1999; Otake & Cutler, 1999), Cantonese (Cutler & Chen, 1997) and Mandarin (Ye & Connine, 1999). Prosodic features are exploited differently across these languages. With respect to prosody, languages are commonly classified as stress languages, tone languages or pitch accent languages.

In *stress languages* one syllable within each word is marked by a longer duration and a greater intensity as compared to other syllables of the word. Furthermore, stressed syllables are characterized by a higher pitch and a stronger pitch movement than unstressed ones. Examples of stress languages are German, English, Dutch and Spanish. Pitch appears to be the most important parameter for the perception of stress in English (Clark & Yallop, 1995, ch. 9). In German the emphasis of the single parameters, that is pitch, duration and amplitude, on the perception of stress is less clear (Dogil, 1999). Some stress languages have stress fixed on a specific syllable of all words. In Finish, for example, stress is invariably located on the first syllable of a word, in Polish stress is fixed on the penultimate syllable. In other stress languages such as German, English or Dutch the position of the stressed syllable varies across words. For instance, in the previously mentioned example, *UMfahren* is stressed on the first syllable, whereas, *umFAHren* is stressed on the second syllable (upper case henceforth denotes a stressed syllable). However, words that differ only in stress information, so called minimal stress pairs, as illustrated in the string *umfahren* are rarely found in stress languages.

The relevance of word prosody fundamentally differs for *tone languages*. The term tone refers to pitch or, more specifically, to a pitch pattern. Tones are not only used in *tone languages*. The English word *no* or the German equivalent *nein*, for instance, might be uttered with falling tone to express refusal or with a rising tone to express a query (see Halliday, 1967, for examples how tones serve communicative function in non-tone languages). However, in contrast to stress languages, tone is associated with lexical items in tone languages. There is only a restricted amount of syllable structures in tone languages which is combined with a large tone repertoire (e.g., six tones for Cantonese or five tones for Mandarin). In other words, tone in these languages is used to distinguish the meaning of words. For instance, the Cantonese consonant-vowel structure /si/ illustrates such a minimal tone group. It means *poem* with high falling Tone 1, *history* with middle rising Tone 2, *time* with low-level Tone 6 and so on (Cutler & Chen, 1997). Commonly, the lexical relevance of tone is considered to be an important characteristic of a tone language.

Pitch accent languages, in contrast, are somewhere in between stress and tone languages. Comparable to tone languages, word prosody is established via pitch. However, pitch accent languages have only restricted tonal options in contrast to tone languages that have a large tone repertoire. For instance, there are only two tones (high versus low) in Japanese. The variable is the point at which the high pitch falls. Take the Japanese string *kakiga* in two spoken forms as an example of a minimal pitch accent pair. It means *oyster* if high pitch falls on the first syllable or *fence* if high pitch falls on the second syllable (Clark & Yallop, 1995, ch. 9). However, as stress in stress languages, pitch accent information rarely distinguishes the meaning of words in Japanese. Taken together, there are differences in the exploitation of prosodic features across different languages. Whereas pitch appears to have an important role to establish stress across all languages, stress languages such as German additionally make use of duration and amplitude to mark a specific syllable within a word. Furthermore, prosodic information has critical lexical relevance in tone languages, but not in stress and pitch accent languages.

1.2 Spoken word access processes

1.2.1 Overview

Research in the field of spoken word recognition suggests that the mapping of sound onto meaning is a highly complex cognitive device (see Cutler & Clifton, 1999; Frauenfelder & Floccia, 1999, for a recent review). Processes by which the listener derives words from the acoustic speech signal are referred to as spoken word access processes (McQueen & Cutler, 2001). Different models have been developed in order to explain how and based on which information such processes operate. Commonly they assume that spoken word recognition involves the activation of a set of lexical competitors and the selection of the target word from this activated set. However, influential models diverge in how they specify the more detailed processing mechanisms and representations. In the present section basic concepts of influential models of spoken word recognition are briefly reviewed. What is known about the role of prosodic information during spoken word access processes is discussed in section 1.2.2.

As the acoustic input that represents a word is highly variable for instance due to the talkers age, sex, dialect, speaking stile, speech rate or due to background noise, most models of spoken word recognition such as TRACE (McClelland & Elman, 1986), the Neighborhood Activation Model (NAM, Luce & Piosini, 1998) or Shortlist (Norris, 1994) assume that the speech stream is prelexically coded as a normalized, language-specific phonological representation. This abstract representation may consist of so called *units of perception* such as phonemes or phonetic features. Note, that phonetic features are articulatory units such as the voiced versus voiceless distinction or the place of articulation of a phoneme. That is, that several phonetic features constitute a phoneme. Models which assume a prelexical level of representation are distinguished from *direct* access models. The latter ones claim that the acoustic input is rather directly mapped onto stored representations than initially transformed into an abstract prelexical code (e.g., Gaskell & Marslen-Wilson, 1997; Goldinger, 1998; Pisoni, 1997). So far empirical evidence (e.g. Pallier, Colome & Sebastian-Galles, 2001; Davis, Marslen-Wilson & Gaskell, 2002) confirms both assumptions, suggesting that the applied paradigms do not distinguish clearly between direct or mediated activation models.

Stored entries in the mental lexicon are referred to as *lexical representations*. These representations are assumed to contain a variety of information about the phonological orthographic form or about syntactic and semantic characteristics (see McQueen & Cutler, 1997). Whereas some models assume that the lexicon consists of abstract prelexical representations (e.g. McClelland & Elman, 1986; Norris, 1994; Norris, McQueen & Cutler, 2000), others propose that the lexicon consists of detailed acoustic traces (Goldinger, 1998; Pisoni, 1997). Again, empirical evidence exists for abstract lexical representations (e.g., Pallier, Colome & Sebastian-Galles, 2001; Norris, McQueen & Cutler, 2002) as well as for detailed ones (e.g., Goldinger, 1996). The form of lexical representations is related to the question whether prelexical representations are needed to recognize a word or not (see McQueen & Cutler, 2001). If segments are represented lexically, abstract prelexical representations which map specific features are a necessary assumption. In contrast, there is no need to extract phonological features on a prelexical level if feature-based lexical representations are suggested (e.g. Gaskell & Marslen-Wilson, 1997; Lahiri & Marslen-Wilson, 1991).

There is considerable agreement among current models of spoken word recognition that the acoustic input activates - directly or via prelexical phonological representations - multiple matching representations in the lexicon. This assumption was first introduced in the COHORT-model of spoken word recognition (Marslen-Wilson & Welsh, 1978). According to this theory a *cohort* of lexical representations is activated on the basis of the initial acoustic-phonetic input of a word. The selection of the first cohort is purely data driven. For instance the perception of the sequence of the phonemes /g/ and /aː/, which forms the beginning of the German word *Galerie* (Engl.: gallery), activates words such as *Garten* (Engl.: garden), *Garage* (Engl.: garage), *Gala* (Engl.: gala) or *Galaxie* (Engl.: galaxy) and *Galerie*. The COHORT model claims that all activated representations in the cohort monitor the subsequent sensory input and remove themselves from the cohort as soon as mismatching information is encountered. That is, the perception of the phoneme /l/ folowing /gaː/ will cause the lexical entries of *Garten* and *Garage* to drop out from the cohort. If features are the basic units of activation and competition, even the place of articulation of the phoneme /aː/, which is more anterior for /l/ than for /r/ would resolve the ambiguity for the fragment /gaː/. Thus, lexical entries for *Galerie* and *Gala* and *Galaxie* would be stronger activated by this phoneme sequence than *Garten* and *Garage* (see Marslen-Wilson & Warren, 1994; McQueen, Norris & Cutler, 1999, for empirical evidence for feature based activation).

Subsequent models of spoken word recognition, including the revised COHORT model (Marslen-Wilson, Moss & van Halen, 1996), TRACE (McClelland & Elman, 1986), NAM (Luce & Piosini, 1998), and Shortlist (Norris, 1994), accounted for the reduction of the initial cohort with the assumption of inhibitory connections at the lexical level. According to these models activated lexical candidates actively compete with each other for selection. Lexical items that completely match the acoustic input are stronger competitors than partially mismatching ones. As a candidate becomes incompatible with the input it looses in activation and eventually drops out of the competition. That is, with the perception of the phoneme /l/ that follows /aː/, or even with the place of articulation of /aː in the word *Galerie*, entries such as *Gala* and *Galerie* will become stronger competitors and, thus, effectively inhibit *Garage* and *Garten*.

In a selection process the element is chosen that best matches the available input to the system. In contrast to the initial activation, selection is not only based on phoneme or feature information of the acoustic signal. The selection of the final item is assisted by contextual information, such as the sentence or the situation in which the word is spoken (Zwitserlood, 1989).

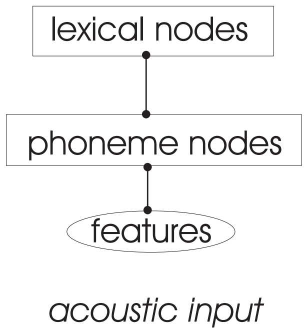

Figure 1.1: *The basic architecture of models of spoken word recognition that assume prelexical processing of the acoustic input. Activation spreads from the prelexical to the lexical level. Some model propose feedback conncetions between the lexical and the prelexical level, others not (see text for details).*

To summarize, important concepts underlying spoken word access in cognitive research are: multiple parallel activation of lexical representations in a mental lexicon, competition among lexical items for recognition and selection of the finally recognized word. Activation is assumed to spread from a prelexical level of representation - or directly from features - to the lexical level. Figure 1.1 illustrates the assumption of a prelexical level, which codes phoneme information. Single phoneme nodes are connected with lexical entries which contain these phonemes. For instance, the phoneme /g/ is connected with representations for *Garten*, *Gala* and so on. The activation of a phoneme node influences the activation of connected lexical nodes. Interactive-activation models such as COHORT and TRACE propose bidirectional excitatory connections between the prelexical and the lexical level. That is, the activity of prelexical nodes might potentially be influenced by lexical nodes. However, recently it has been questioned whether feedback connections from the lexical to the prelexical level are really necessary (Norris, McQueen & Cutler, 2000). Examples for models which assume only forward activation flow from prelexical to lexical nodes are Shortlist (Norris, 1994) and NAM (Luce & Cluff, 1998).

Models of spoken word recognition concentrated so far on the phonemic structure of the acoustic input. Suprasegmental information is commonly not implemented in these models. However, empirical evidence, which is discussed in the following sections, suggests that prosodic information is exploited by the listeners during spoken word access processes. In other words, it appears to make a difference for initial processes of spoken word recognition whether the string /ga/

comes from *GAla* or from *galaXIE*. Both strings differ only in their prosody. The first syllable of *GAla* is stressed, whereas that of *galaXIE* is unstressed. In the following three sections, selected paradigms that investigate spoken word access processes are shortly introduced and effects of suprasegmental information found with these paradigms are discussed. The first section focuses on whether prelexical and or lexical representations code prosodic information. Of course, this issue is related to the question, whether suprasegmental information is used for lexical activation and competition which is discussed thereafter.

1.2.2 The role of prosody

1.2.2.1 Prelexical and lexical representations

Studies that investigated a possible prosodic structure of lexical representations used phoneme or tone judgment tasks (Coninne, Clifton & Cutler, 1987; Ye & Connine, 1999) and delayed repetition priming (Cutler & van Donselaar, 2001; Cutler & Otake, 1999). Both paradigms are reported to show lexicality effects. That is, different results were observed for words as compared to nonwords. In phoneme judgments, listeners are required to make decisions about categories of speech sounds. Phoneme judgments are thought to be influenced indirectly by prelexical and lexical representations which supply information to a specific phoneme decision level (Norris, McQueen & Cutler, 2000). In a phonetic categorization task ambiguous auditory strings are presented and subjects have to guess the phoneme underlying the ambiguous information (see McQueen, 1996, for a review). For instance, the English word *dash* and the nonword *tash* as well as the nonword *dask* and the English word *task* differ only in their voice onset times of the first phoneme. When this parameter is varied along a continuum, subjects show a tendency to categorize a phoneme so that the resulting string would be a word. That is, an ambiguous /d/ in *dask* is more likely to be judged as /t/ than an ambiguous /d/ in *dash*. Vice versa, an ambiguous /t/ is more likely to be classified as /d/ when it occurs in *tash* than when it occurs in *task* (Ganong, 1980). A similar effect for stress is reported by Connine, Clifton and Cutler (1987). In their study, subjects were asked to categorize an ambiguous word onset consonant varying along a continuum between the English word *TIgress* and the nonword *DIgress* or between the English word *diGRESS* and the nonword *tiGRESS*. The responses show effects of lexical status: /t/ was reported more often for the initially stressed items, whereas /d/ was reported more often for the initially unstressed items.

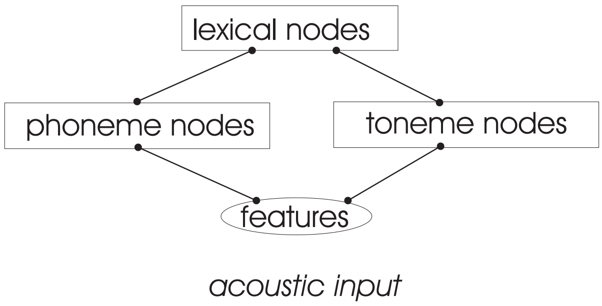

Figure 1.2: *Model of spoken word recognition with a prelexical level that codes not only phoneme but also tone information (see Ye and Connine, 1999)*

In a phoneme monitoring task (see Connine & Titone, 1996, for a review) subjects are required to detect a target phoneme, for instance /d/. This decision is faster in a word such as *date* than in a nonword such as *dac* (Cutler, Mehler, Norris & Segui, 1987). Similar to effects of phonemic categorization, this effect has been interpreted as being influenced by lexical representations. Recently, a lexicality effect has been observed for tone monitoring in Mandarin idioms (Ye & Connine, 1999). Subjects were required to detect a specific tone in a idiomatic context as fast as possible. The idiomatic German expression *das Eis brechen* which equals the English string *breaking the ice* is a combination of words for which meaning is not derived from the meaning of the individual words. One hypothesis of idiomatic processing is that they are lexically represented by their own entries (Connine, Blasko, Brandt & Layer, 1992; Swinney & Cutler, 1979; Titone & Connine, 1994). Ye and Connine (1999) showed that tone monitoring of meaningful Mandarin syllables is faster in idioms than in a neutral context. In line with the lexical representation hypothesis for idioms, this result can be taken as evidence that tone is coded at the lexical level. Furthermore, in an idiom context, tone monitoring was faster than the monitoring of a single phoneme. This led the authors to the conclusion that tone processing is highly lexically involved and that there might be a separate prelexical nodes representing tone information, which are called *toneme nodes*. These nodes are thought to code tone information in Chinese and to coexist with phoneme nodes on the prelexical level (see Figure 1.2).

However, the results of Ye and Connine (1999) are not fully in accordance with an assumed prelexical level in spoken word recognition. In a third task they showed that tone monitoring in idiom context depends on whether the tone of the target word is close to the expected word or highly different from the expected word. If, for instance, the appropriate word has Tone 2, a word with Tone 4, which is close to Tone 2, is responded to faster than Tone 3, which is more different from Tone 2. Thus, smaller units than labeled by the tone distinction might be processed during spoken word recognition in tone languages. This parallels the assumption of feature based processing on the segmental level, an account which, as argued in the previous section, not necessarily need a prelexical level of representation.

The results of Ye and Connine (1999) contrast to only weak effects of tone information in a Cantonese study by Cutler and Chen (1997). Nonwords that mismatch words in tone or in phoneme information were presented in a lexical decision task and in similarity judgments. Subjects more often accepted a nonword as a word when the difference was in tone than when it was in phoneme information. Furthermore, same-different judgments were slower and less accurate when two strings differed in tone than when they differed in a phoneme. However, the results might rather reflect a perceptual disadvantage for tone, which appears to become available later than phoneme information, than the involvement of the lexicon for phoneme information, but not for tone information. In contrast, the paradigm by Ye and Connine (1999) clearly taps lexical representations and, the results can be taken as a strong hint that tone information is an important part of prelexical and/or lexical representations in tone languages.

A further tool to study the format of lexical representations is to present lists of spoken words or nonwords in which items are repeated after several trials. Subjects have to perform a lexical decision, that is they must simply decide whether an item is a word or not. After lags of more than four intervening trials there is a clear lexicality effect: When a word appears twice, subjects respond more rapidly. However, when a nonword is encountered for a second time, no difference is recorded in the participants' decision times (Mimura, Verfaellie & Milberg, 1997). As already pointed out, different responses for words and nonwords can be taken to argue that a paradigm taps the lexical level. It has been shown with this paradigm that word pairs which differ only in a single phonetic feature are represented and activated separately. Pallier, Colome and Sebastian-Galles (2001) showed that, for instance, the Catalan word /*netə*/ (Engl.: granddaughter) does not facilitate responses to the subsequent presentation of /*nɛtə*/ (Engl.: clean) in Catalan subjects. Note, that both words differ only in one feature in that /e/ is the closed counterpart of /ɛ/. Delayed repetition priming has also been used to investigate the processing of minimal stress pairs (Cutler & van Donselaar, 2001) and minimal pitch accent pairs (Cutler & Otake, 1999). Both studies report that lexical decisions to one version of a minimal stress pair are not facilitated by the prior presentation of the alternatively stressed version. Only the repetition of identically stressed words speeds up lexical decisions. For example, reactions to the Dutch word *VORnaam* (English: first name) are facilitated when *VORnaam* is presented some items earlier, but not when *vorNAAM* (English: respectable) is presented previously. The Dutch and the Japanese study indicate that prosodic information in form of stress and pitch accent is coded at the lexical level. Each member of a prosodic minimal pair is represented separately in the lexicon. To summarize, results of phoneme judgment and repetition priming tasks reveal that prelexical representations are not purely phonemic in nature. Tone, stress and pitch accent information is coded at the lexical level. The results of repetition priming not only reflect that prelexical and lexical representations code prosodic information, they are also a clear hint that prosody is exploited by the listeners during lexical activation. More evidence regarding this issue is presented in the following section.

1.2.2.2 Lexical activation

Whether the activation of lexical representations is a function of suprasegmental information has been investigated with cross-modal form priming. This paradigm investigates effects of spoken primes on visual targets (Zwitserlood, 1996). Reaction times (RTs) to the targets, usually recorded while subjects make lexical decisions, are assumed to reflect the lexical activation of the target's lexical entry. This interpretation is based on the assumption that two independent access routes, one for the auditory and one for the visual domain activate a modality independent lexical entry (Marslen-Wilson, 1990). As already pointed out, minimal pairs are a promising tool to investigate whether suprasegmental information is used during lexical activation. Positive evidence regarding this issue comes from delayed repetition priming in Dutch (Cutler & van Donselaar, 2001) and Japanese (Cutler & Otake, 1999). However, negative evidence is reported for the presentation of minimal stress pairs in mediated form priming. An English (Cutler, 1986) and a Dutch (Jongenburger, 1996) study that used mediated priming with minimal stress pairs were not able to show that lexical activation is restricted by word prosody. In both studies the activation of semantically related words to both versions of minimal stress pairs were tested. For example, consider the string *forbear*. Related to the initially stressed version *FORbear* is the word *ancestor*, whereas, *tolerate* is related to the initially unstressed version *forBEAR*. In the English study (Cutler, 1986) associated words to both members of the minimal stress pairs were facilitated. That is, *forBEAR* speeded up responses to *tolerate* as well as to *ancestor*. In the Dutch study no priming effect was observed (Jongenburger, 1996).

Apparently, both results are in conflict not only with each other, but also with the results of the above mentioned effects for prosodic minimal pairs in delayed repetition priming (Cutler & Otake, 1999; Cutler & van Donselaar, 2001), and cross-modal repetition priming in Japanese (Sekiguchi & Nakajima, 1999). In contrast to mediated priming, Sekiguchi and Nakajima tested the activation of each member of a minimal pitch accent pair directly and not via semantically related words. This is possible in Japanese as each word is represented by a specific visual symbol that contains no phonological information. As for delayed repetition priming in Japanese (Cutler & Otake, 1999) it was found that one version of a minimal pitch accent pair speeded up reactions only to the logographic symbol of that word, but not to the symbol of the alternatively accented word. The heterogeneous empirical results regarding the use of suprasegmental information for minimal pairs in repetition and semantic priming might be caused by differences in the applied paradigms. Mediated priming may well test other aspects of spoken word recognition than repetition priming. Thus, Cutler, McQueen, Norris and Butterfield (2000) failed to show effects of cross-modal mediated priming in sentence context and claimed that this type of priming might not be automatic in nature and that it thus might not be a direct reflection of lexical access or activation. Note, that Cutler, 1986, as well as Jongenburger, 1996 used carrier sentences for their prime words. In line with this, Gaskell and Marslen-Wilson (2001) argue that there are more processing stages underlying effects observed in mediated than in purely form priming. This may well lead to more variance in the observed data for the former than for the latter paradigm. Accordingly, results of repetition priming with minimal stress pairs (Cutler & Otake, 1999; Cutler & van Donselaar, 2001; Sekiguchi & Nakajima, 1999) appear to be more valid indicators whether prosodic information of minimal stress pairs is used during lexical activation than the results of mediated priming (Cutler, 1986; Jongenburger, 1996). The former studies clearly confirm the assumption that stress and pitch accent information is coded in prelexical representations which activate matching lexical representations with corresponding prosodic information.

Additional and unequivocal evidence regarding the use of suprasegmental information during lexical activation comes from cross-modal word fragment priming in English, Dutch and Spanish (Cooper, Cutler & Wales, 2002; Cutler & van Donselaar, 2001; Soto-Faraco, Sebastian-Galles & Cutler, 2001; van Donselaar, Koster & Cutler, submitted). This paradigm investigates the initial activation during spoken word recognition by means of word onset fragments. After presenting a prime fragment, responses to multiple matching candidates and even semantically related words to these candidates are facilitated. For example the spoken fragment *fee* taken from the English word *feel*, facilitates responses to *feel* as well as to *feed* (Marslen-Wilson, 1990). Similarly, the fragment *kapi* taken from the Dutch word *kapitain* (Engl.: captain) is reported to facilitate responses to *schip* (Engl.:ship) which is related to captain as well as responses to *geld* (Engl.: money) which is related to *kapitaal* (Engl.: capital). This can be taken to conclude that *kapi* activates not only *kapitain*, but also the initially identical word *kapitaal* (Zwitserlood, 1989). To investigate effects of the fragments stress information form related fragment priming has been used successfully. Using this paradigm it has been shown that target words such as the word *muSEUM* are responded to faster when they are preceded by a segmentally and suprasegmentally matching fragment such as *muS*, than when preceded by a segmentally matching but suprasegmentally mismatching fragment like *MUs*. This can be taken as clear evidence that lexical representations are differently activated due to prosodic information. Recently, the function of tone information for lexical activation has been addressed (Yip, 2001). However, in this study effects of tone were confounded with effects of segmental information, which was also modulated. Thus, no clear conclusions about the role of tone for lexical activation can be drawn so far.

1.2.2.3 Competition at the lexical level

Both preceding subsections show that word prosody appears to be coded at the lexical level and that prosodic information of a spoken word modulates activation at the lexical level. The following section concerns the question whether prosodic information influences lexical competition. Direct evidence for competition between activated candidates comes from word-spotting experiments (McQueen, Norris & Cutler, 1994). In those experiments listeners have to detect real words, for instance *mess*, embedded in nonsense strings. When a nonsense string is the onsets of a longer word it takes longer to detect an embedded word than when the nonsense string is not the onset of a longer word. Take the nonsense strings *domes*, which is the beginning of *domestic*, and *nemess*, which is not the beginning of an English word, as examples. Subjects need more time to detect the embedded *mess* in *domes* as compared to *nemess*. This result suggests that *domes* activates the entry for *domestic*, which competes with the entry for *mess*.

Recently, this word-spotting has been used to explore whether a word onset with mismatching prosodic information activates a competitor word in a word spotting experiment (Cutler & van Donselaar, 2001). For instance, the spoken nonsense string *muZEE* is the correct beginning of the Dutch word *muSEUM* (Engl.: museum), whereas *MUzee* does not start a Dutch word. When both strings are presented and subjects are engaged in detecting the Dutch word *zee*, this takes longer for the second string (Cutler & van Donselaar, 2001). That is, the prosodic information of the segmentally identical strings appears to be used for activation in that *muSEUM* is activated to a lesser degree by *MUzee* than by *muZEE*. Word spotting is easier for the former string because there is no or less competition from the activated entry for *muSEUM*. The data strengthen the hypothesis that prosodic information is exploited by listeners during lexical activation. Furthermore, they reveal that entries which are prosodically incongruent with the acoustic stream are weaker competitors than prosodically matching entries.

More recent results of cross-modal word fragment priming provide additional evidence for competition at the lexical level. In a Spanish study, a prime fragment such as *abon* was either followed by a completely matching word such as *abonado* (Engl.: subscriber), by a word mismatching the prime in a single vowel like *abanico* (Engl.: fan) or by a completely mismatching word such as *osado* (Engl.: daring Soto-Faraco, Sebastian-Galles & Cutler, 2001). Subjects reactions were faster for the complete match than for the complete mismatch condition. However, the partial mismatch condition produced even longer reactions than the complete mismatch condition. These results are interpreted in terms of competition. Lexical representations of the partially mismatching targets are thought to be activated by the prelexical access code. Simultaneously, however, the activation of partially matching representations is inhibited by their strongly activated competitors which fully match the acoustic information of the prime. Accordingly, competition due to suprasegmental information can be investigated with word onset fragments which initiate words with different prosody. For example the words *music* and *museum* share the same word onset segments, but differ in their suprasegmental word onset information. The initial syllable of *MUsic* is stressed, whereas that of *muSEUM* is unstressed. Cooper (2002) showed that the prime fragment *MUs* facilitates reaction times (RTs) for *MUsic* stronger than for *muSEUM*. Vice versa, *muSEUM* received greater activation than *MUsic* from the prime fragment *muS*. However, in that study suprasegmental mismatch did not slow down reactions. That is, *muSEUM* was responded to faster when it was preceded by *MUs* than when it was preceded by *mons* taken from *monster*. Thus, there is no hint for competition in the results of the study reported by Cooper (2002).

A different result is reported by Soto-Faraco et al. (2001). Bisyllabic prime fragments such as *PRINci* were presented. Segmentally and stress matching targets such as *PRINcipe* (Engl.: prince) were responded to faster than unrelated targets such as *mosQUIto* (Engl. mosquito). However, prolonged RTs in comparison to the unrelated condition were found for segmentally matching but stress mismatching targets like *prinCIpio* (Engl.: beginning). Taken together, a competition effect as a result of stress information was observed in the Spanish, but not in the English study. This difference might be caused by different fragment lengths in both studies, an assumption confirmed in a recent Dutch study (van Donselaar, Koster & Cutler, submitted). Fragments longer than the first syllable of a word such as *OCto* taken from the Dutch word *OCtopus* (Engl.: octopus), speeded up reactions to *OCtopus* but slowed down responses to *okTOber* as compared to an unrelated control word. In contrast, monosyllabic fragments such as *OC* only facilitated reactions to a stress matching word but did not enhance RTs for a stress mismatching word. Taken together, suprasegmental information of fragments no longer than the first syllable of a word activates all segmentally matching representations, but even more strongly the ones that match its suprasegmental information. As soon as the suprasegmental information of the second syllables is available, competition comes into play. Lexical representations that match the prosody of the incoming speech signal effectively inhibit activated representations that do not match the prosody of the speech signal.

To summarize, the results of different paradigms clearly reveal that prosodic information modulates spoken access processes. They indicate that different kinds of suprasegmental information such as tone (Ye & Connine, 1999), pitch accent (Cutler & Otake, 1999; Sekiguchi & Nakajima, 1999) as well as stress (Cooper, Cutler & Wales, 2002; Cutler & van Donselaar, 2001; Soto-Faraco, Sebastian-Galles & Cutler, 2001; van Donselaar, Koster & Cutler, submitted) are coded at the lexical level. Thus, although, stress and pitch accent information rarely have lexically distinctive function (see section 1.1), they appear to play a critical role during spoken word access processes. This might result from the large amount of words like the German words *PRObe* (Engl.: probe) and *proBLEM* (Engl. problem) that equal each other in their segmental onset, but can be distinguished by their prosodic information. This information clearly speeds up word recognition for words with segmentally ambiguous word onsets. It has been argued, that tone might be represented in separate prelexical representations (Ye & Connine, 1999) and this would of course by functionally equally of even more important than in stress and pitch accent languages. However, there is no comparable study reported for a stress or a pitch accent language to conclude that this information might be presented differently than tone.

Behavioral investigations provide only indirect hints about the timing of spoken word processing as they reflect only the final result of decision processes. The recording of ERPs reflects online processing more directly. It is the aim of the present thesis to bring behavioral and electrophysiological indicators more closely together. On the one hand, this approach may well provide more detailed insights into the time course of cognitive processes underlying the behavioral effects for word prosody. On the other hand, the combination of theoretically and empirically based behavioral techniques with the recording of ERPs might supply further consolidated findings about cognitive processes that underlay the ERP signal elicited by spoken words. In the following section the ERP method is briefly introduced. Thereafter, experiments are described that investigated spoken word processing by means of ERPs.

1.3 Event-related Potentials (ERPs)

1.3.1 Overview

Reaction times inform us about the end product of cognitive processes including stages of perception, evaluation and eliciting a response. In contrast, ERPs provide us with millisecond accuracy information about the time course of brain responses as they take place in real time. Thus, ERPs have been used when temporal aspects of cognitive processes have been the main focus of research. Scalp-recorded ERPs measure primarily summed post-synaptic potentials of simultaneously activated pyramidal cells in the neocortex time-locked to specific events, such as the presentation of spoken or written words. In order to derive ERPs, segments of the electrophysiological activity, time-locked to the event of interest are extracted and averaged. After averaging a sufficient number of epochs, the resulting ERP is thought of as reflecting the brains response to an event of interest free from any other unrelated brain activity (see Fabiani, Gratton & Coles, 2000, for an overview of the ERP method).

The ERP waveform consists of a specific set of temporally consecutive negative and positive peaks which are commonly referred to as *components*, *deflections*, *peaks* or *waves*. ERP components are described in terms of their characteristic scalp distribution, their polarity, their latency and their experimental variability. The earlier components in an ERP waveform - approximately up to 100 to 200 ms of the signal - are thought to be exclusively modulated by physical properties of an external eliciting event. Accordingly, these early peaks are called *exogenous components*. They are distinguished from later peaks which are more closely related to cognitive events and which are influenced by intentions and actions of a participant. Consequently, the later peaks in the ERP waveform are referred to as *endogenous components*.

Traditionally, psycholinguistic ERP research focused on the processing of written language, typically on the comprehension of written words in sentence context. Only recently, spoken language processing has received more attention (see Hagoort & Brown, 2000, for a review). Much of the ERP research on language processes compared ERPs elicited by correct and by incorrect linguistic stimuli. A major finding of this research is that sentences that violate syntactic constrains elicit ERP effects that are clearly distinct from such elicited by semantic violations (see Friederici, 2002; Kutas & Federmeier, 2000; Osterhout, McLaughlin & Bersick, 1997, for reviews). Semantic violations such as 'The cats won't *bake*' elicit a centroparietal negative wave labeled as N400, which will be discussed in more detail. In contrast to semantic violations, syntactic violations such as 'The cats won't *eating*' are correlated with two ERP components, a left-anterior negativity (LAN) in an early time-window and a late centro-parietal positivity, variously labeled as P600 or syntactic positive shift (SPS). The fast detection of word category errors such as the last word in the German sentence 'Die Bluse wurde am *gebügelt*' (Engl.: The blouse was on *ironed*), is associated with a left-anterior negativity that peaks even earlier than the LAN and is thus referred to as ELAN (early left anterior negativity). Recently, the processing of prosodic information in sentence context has been related to a large positive wave which is named closure positive shift (CPS) as it is assumed to reflect the processing of the closure of an intonational phrase (Steinhauer, Alter & Friederici, 1999). Taken together, the use of different kinds of linguistic information can be distinguished in the ERP waveform by means of several different components.

The finding of language-related ERP effects does not allow the inference that any of these effects are language specific. Presumably, the generators of these effects are sensitive to other cognitive processes as well. For example, faces (e.g., Eimer, 2000; Bentin & Deouell, 2000) and pictures (e.g., Federmeier & Kutas, 2001, 2002; West & Holcomb, 2002) elicit negative ERP deflections similar to the N400 observed for language. This suggests that the N400 reflects general processes related to semantic memory. Even the P600 effect appears to be not exclusively elicited by language related syntactic incongruities. An enhanced P600 is also observed for structural violations in music (Patel, Gibson, Ratner, Besson & Holcomb, 1998). Thus at least the N400 and the P600 might reflect general cognitive processes that are not language-specific. However, ERPs elicited in psycholinguistic experiments at the sentence level show that the brain distinguishes between semantic, syntactic and prosodic information. Thus, the ERP method provides a useful tool to investigate the processing of different types of linguistic information as it occurs in time. The following section summarizes ERP research related to the processing of spoken words. As already pointed out, this research utilizes the high temporal resolution of this method. The spatial resolution of ERPs can only be considered with caution, as the skull acts as a spatial filter that smears brain electrical activity (Davidson, Jackson & Larson, 2000). A review of the neural architecture underlying the processing of spoken word forms integrating ERP and neuroimaging findings is provided by Price, Indefrey and van Turenout (1999).

1.3.2 ERPs related to spoken word processing

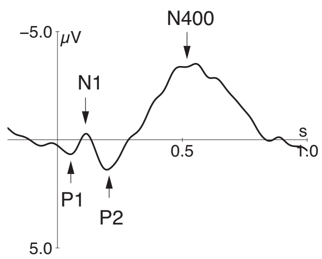

Figure 1.3: *Typical ERP waveform elicited by spoken words*

Similar to written words, spoken words evoke characteristic ERP deflections. The electrophysiological activity elicited by a spoken word is illustrated in Figure 1.3. This waveform is based on the averaging of the EEG signal elicited by several spoken words (compare previous section). The vertical bar marks the beginning of the presentation. As usual in ERP research, negativity is plotted upwards. To date, roughly three temporally consecutive sets of peaks and troughs in the ERP waveform can be distinguished in terms of electrophysiological properties and functional correlates. Typically, a P1-N1-P2 complex is elicited within the first 200 ms post stimulus onset (compare Figure 1.3). These three peaks are considered to be exogenous ERP components that reflect early modality specific input processes, that is automatic auditory processes that analyze a word. To date, the three components are reported to be modulated simultaneously (Hillyard, Luck, & Mangun, 1994). For words presented in isolation the P1-N1-P2 complex is followed by a substantial scalp negativity between 200 and 800 ms post-stimulus-onset (Cobianchi & Giaquinto, 1997; Soares, Collet & Duclaux, 1991; Woodward, Owens

& Thompson, 1990). This component is commonly referred to as N400. As described in the following sections, the N400 varies as a function of lexicality of the spoken strings as well as as function of different characteristics of the input that precedes a spoken word. Thus, it appears that this component reflects a family of different negative ERPs rather than a single wave.

Depending on task constrains, the N400 is followed by a late positive component (LPC). This ERP deflection may be considered as a member of the so-called P300 family reflecting controlled processing related to subjective expectancy, task relevance, decision making, or contextual updating (Donchin & Coles, 1988). In P300 studies auditory and visual P300 amplitudes are enlarged and latencies shortened when stimuli are easier to detect (e.g. Polich, Ellerson & Cohen, 1996). Furthermore, a positive relationship between P300 latency and reaction times has been obtained in a number of information processing tasks including lexical decision (e.g., Bentin, McCarthy & Wood, 1985). These effects and the late occurrence of the LPC exclude an interpretation of this component in terms of lexical processing. Thus, this component is not in focus of the present theoretical discussion of ERP effects related to lexical access and in the data analysis of the subsequent experiments.

1.3.2.1 Semantic and phonological processing

Similar to the classical N400 findings in the visual modality (Kutas & Hillyard, 1980), the N400 for spoken words can be modulated by manipulating the semantic correspondence between a word and the context in which it is embedded. A variety of studies have indicated that spoken words presented in sentence context elicit a higher amplitude in the N400 when they do not fit the preceding sentence than when they are appropriate in the sentence context (Besson, Faita, Czternasty & Kutas, 1997; Connolly, Stewart & Phillips, 1990; Connolly, Phillips, Stewart & Brake, 1992; Connolly & Phillips, 1994; Friederici, Pfeifer & Hahne, 1993; Hagoort & Brown, 2000; Herning, Jones & Hunt, 1987; Holcomb & Neville, 1991; McCallum, Farmer & Pocock, 1984; Van Petten, Coulson, Rubin, Plante & Parks, 1999). Similarly, when spoken words are preceded by semantically related words (Holcomb & Neville, 1990; Holcomb & Anderson, 1993) or semantically related sounds (Van Petten & Rheinfelder, 1995) they elicit a lower N400 amplitude than spoken words which are unrelated to preceding stimuli. These findings fit well with the general interpretation of the N400 as reflecting processes dealing with stored information in semantic memory (Kutas & Federmeier, 2000). Spoken non-meaningful strings generate an N400 with longer latency than words (Attias & Pratt, 1992; Soares, Collet & Duclaux, 1991). The same effect is reported for written pronounceable pseudowords (e.g., Picton & Hillyard, 1988) and it might be related to semantic integration effort for meaningless strings.

Most of the N400 literature is related to semantic processing during visual and auditory word recognition. It has been shown that a lower amplitude of the N400 occurs for words that fit a semantic context, as compared to unrelated words. The nature of the processes underlying the N400 effect has been investigated with priming paradigms. One of the most compelling findings, that the N400 does not reflect lexical activation is reported by Brown and Hagoort (1993). These authors showed that an N400 effect is not elicited when primes are masked, even though behavioral priming occurs for masked primes. In contrast, recent studies showed that N400 effects can be elicited by masked prime words (Deacon, Hewitt, Yang & Nagata, 2000; Kiefer, 2002; Brendel & Kiefer, 2002). However, that degraded visually presented stimuli elicit a behavioral priming effect, but no N400 effect (Holcomb, 1993). Even for the auditory domain there is a finding that calls into question that the N400 reflects lexical activation. In a task used by Bentin, Kutas and Hillyard (1995) two different word lists were presented dichotically. Subjects were instructed to memorize words presented to one ear while ignoring words presented to the other ear. When words were preceded by semantically related words on the attended site, the N400 component was significantly reduced as compared to words preceded by unrelated words. No such N400 effect was observed for words presented to the unattended ear. However, there was clear evidence that the unattended words activated their lexical representations. In a subsequent lexical decision task attended as well as unattended words were responded to faster than new words. This can be taken as evidence, that the N400 is not sensitive to semantic activation during spoken word recognition. Thus, the results are far from conclusive that the N400 is related to semantic spreading activation.

N400 effects have not only been observed for semantic relations of prime-target pairs, they also occur for phonologic relations. N400 effects are reported for rhyme priming in the visual and in the auditory domain. For visually presented items, orthographically similar targets that rhyme (e.g., *rung* and *sung*) as well as orthographically dissimilar targets that rhyme (e.g., *make* and *ache*) elicit reduced N400 amplitudes as compared to nonrhyming targets (Barrett & Rugg, 1989; Rugg, 1984a,b; Rugg & Barrett, 1987). For the auditory domain it is reported that the amplitude of the N400 is reduced for spoken words preceded by spoken rhyming words as compared to spoken non-rhyming words (Dumay, Benraiss, Barriol, Colin, Radeau & Besson, 2001; Radeau, Besson & Castro, 1998; Praamstra & Stegemann, 1993; Praamstra, Meyer & Levelt, 1994). These results show that the N400 is sensitive to phonological variables.

Praamstra et al. (1994) tested not only a rhyming condition such as the Dutch words *graaf-staaf* (Engl.: duke-stick). They also presented an onset overlap condition, such as the Dutch words *beeld-beest* (Engl.: statue-animal), referred to as alliterating condition. Spoken words preceded by spoken rhyming words started to differ from words preceded by non-rhyming words at 450 ms. There was no ERP effect for nonwords in the rhyming condition. In contrast, spoken words that share the initial consonantal and vowel onset with the prime words elicited a reduced amplitude in the ERP starting at already 250 ms. This effect was also found for alliterating nonwords. The results can be interpreted in a way that the alliterating condition taps prelexical representations or prelexical processing, whereas, the rhyming condition is related to later integrative processes that need the mental lexicon. Thus, the overlap of initial phonemic information between prime and target is likely to reflect priming at the prelexical level. The temporal delay of the rhyming effect as compared to the alliterating effect suggests that the N400 component might not be a unique electrophysiological wave. The results reported in the next section support this assumption. It has been found that early processes which are affected by both phonological as well as semantic information are correlated with negative-going ERP correlates starting at 200 ms.

1.3.2.2 Interaction of semantic and phonological information

If semantically inappropriate words which are presented in sentence context match the expected word in their initial phonemes they elicit a reduced negativity in the ERP than completely mismatching words (Connolly, Phillips, Stewart & Brake, 1992; Connolly & Phillips, 1994; Hagoort & Brown, 2000; van den Brink, Brown & Hagoort, 2001; Van Petten, Coulson, Rubin, Plante & Parks, 1999). For instance, the first phoneme of the Dutch word *pension* (Engl. pension) matches the expected terminal word *penseel* (Engl. paint brush) of the Dutch sentence *De schilder kleudre de details with a small* ... (Engl. The painter colored the details with a small ...). In contrast, the word *doolhof* (Engl. labyrinth) is completely unrelated to the sentence. *pension* was found to elicit a reduced negativity between 200 to 300 ms as compared to *doolhof*. This ERP deflection has been variously labeled as *phonological mismatch negativity* (PMN Connolly, Phillips, Stewart & Brake, 1992; Connolly & Phillips, 1994), N2b (D'Arcy, Connolly & Crocker, 2000), early N400 (Van Petten, Coulson, Rubin, Plante & Parks, 1999), N250 (Hagoort & Brown, 2000) or N200 (van den Brink, Brown & Hagoort, 2001). Slight differences in the latency of the effects across the studies might be attributed to the use of word-onset plosives in some studies (van den Brink, Brown & Hagoort, 2001). The effect can be clearly distinguished from the N400 by comparing both conditions semantically incorrect conditions with the correct one. In the above example, comparing the semantically correct word *penseel* with the semantically incorrect but phonologically fitting word *pension* as well as with the completely unrelated words *dollhof* results in a reduced N400 in the correct condition starting at 400 ms.

Originally, the effect preceding the N400 has been attributed to an intermediate processing stage in which both phonological and semantic information contribute to spoken word recognition. More recently, this assumption has been specified by van den Brink et al. (2001) who argued that the ERP for semantically mismatching but partially phonologically matching words reflects the fact that sentence final word are assessed early on with respect to their goodness of fit relative to initial acoustic information within the sentence context. The proposed process underlying these effects is the already mentioned *selection function* (see section 1.2.2.2). Although, the word *pension* does not fit the sentence context, it's initial phonemes activate not only *pension* but also *penseel*. This interpretation is in accordance with evidence from behavioral studies that semantic information is activated by only partial acoustic information of a word (Zwitserlood, 1989). Thus, semantic features for the appropriate sentence ending are within the activated cohort and early semantically based selection processes can act. In contrast, the word *doolhof* does not activate any entry related to the sentence appropriate ending. In this condition there is no semantic information fitting the sentence context as reflected in an enlarged early negativity (van den Brink, Brown & Hagoort, 2001).

To sum up, early negative ERP deflections might be interpreted as a reflection of spoken word access processes. They might indicate whether the initial acoustic information of a string such as *penseel* activates the lexical representation such as of *pension*. That is, early negative effects allow to test what kind of acoustic information activates a word which is the appropriate sentence ending. With this respect, ERPs for sentence final words can be taken to investigate whether the prosodic information modulates lexical activation. For instance, one member of a minimal stress pair can be presented as an inappropriate ending of sentence in which the alternatively stressed version of the minimal stress pair is expected. ERPs would than indicate, whether the incorrectly stressed version activates the alternatively stressed member of the minimal stress pair. Recently, such studies have been conducted. Their results are discussed in the following section.

1.3.2.3 Interaction of semantic and prosodic information

The question whether stress and context interact in auditory word recognition has already been investigated in a behavioral study (Slowiaczek, 1991). Subjects were presented with sentences in which the last word was replaced by enveloped shape noise. This manipulation replaces the segmental information by maintaining prosody. Following a sentence that ended in noise, subjects heard a word and should indicate whether the word fits both the sentence meaning and the stress pattern of the noise. Four conditions were tested: *A* words matching the sentence context and the prosody of the shape noise, *B* words matching the context but not the noise, *C* words matching only the noise, *D* words matching neither the context nor the noise. Responses were faster and more accurate in condition *A* than in condition *B*, indicating that prosody might effect responses for contextually appropriate words. Note, however, that only in condition A yes-responses were required. Thus, reaction times in this condition might profit from a behavioral facilitation for *yes* in comparison to *no* responses. For words inappropriate to the sentence (conditions C and D) it made no difference whether they matched the prosody of the noise or not. The results favor the conclusion of a highly dominant function of sentence context. However, sentences with more than one appropriate ending were used. That is, appropriate words could either have one or an alternative stress pattern. Thus, the results are not informative about whether the stress pattern of an highly expected word, modulates the processing of an inappropriate word with the same or a different stress pattern.

Effects of recent electrophysiological studies more clearly indicate a crucial rule of prosodic information during word processing in sentence context. They report effects related to pitch-accent information (Hayashi, Imaizumi, Koichi, Niimi, Ueno & Kiritani, 2001) and to tonal information (Wang & Connolly, 2001a,b). Moreover, the results can be taken as an indirect account that prosodic information might constrain lexical activation. Hayashi et al. presented Japanese minimal stress pairs in a question-answer paradigm while recording the magnetic brain responses of subjects. A quiz sentence, for example *What color is the stop light indicating stop?* was followed either by the correct answer word (e.g., *Aka* [Engl.: red]) or by its alternatively accented version (e.g., *aKA* [Engl. dirt]). The incorrect prosody evoked an enhanced magnetic brain response starting at 250 ms. The recording of event-related magnetic responses of the brain is, as well as the recording of event-related electrical responses of the brain, an neurophysiological method with high temporal resolution. Thus, the time windows of effects observed with one or the other method can be directly compared. The magnetic brain response observed by Hayashi et al. (2001) is in the same time window as the N200 reported by van den Brink et al. (2001). In parallel to the N200, the magneto-encephalographic response might be related to lexical activation. Following the interpretation of the N200, the stronger magnetic response for words that do not match the pitch-accent pattern of the expected answer indicates, that these words do not as strong activate the lexical entry with the correct pitch pattern. That is, the lexical entry for *aKA* is stronger activated by the correctly accented version *aKA* than by the alternatively accented version *Aka*. The magneto-encephalographic response might thus reflect enhanced selection effort for incorrectly accented words in sentence context. The interpretation of modulated activation due to pitch accent is in line with recent behavioral results (Cutler & Otake, 1999; Otake & Cutler, 1999; Sekiguchi & Nakajima, 1999, see section 1.2).

In line with the magneto-encephalographic finding for pitch accent information in Japanese are recent ERP findings for the processing of tone information in Chinese. In an auditory (Wang & Connolly, 2001a) and in a visual (Wang & Connolly, 2001b) experiment, sentences were presented whose terminal words were either expected due to sentence constrains or differed from the expected words in a single phoneme or in tone information. In both studies, words that were identical in phoneme or in tone to the correct ending elicited a smaller negativity in the ERP than completely incongruous words. However, both conditions elicited a higher amplitude in this ERP deflection than the completely matching condition. Although, the authors interpret the effect in terms of the N400 only, there is a clear distinction between the N400 and an earlier negativity in their data. According to the interpretation of van den Brink et al. (2001) these results might be taken as a first hint that tone information restricts lexical activation. A tonal mismatch does not prevent the appropriate sentence ending from being activated, as the reduced amplitude as compared to the incongruent condition reveals. However, the appropriate sentence ending is activated more strongly by the correct version than by the tonal mismatch version. Again, the electrophysiological findings are in line with recent behavioral results which suggest that tone information is used during spoken word access processes (Ye & Connine, 1999, see section 1.2.2.1). The ERP waves suggest that the restriction due to tone information might be equally powerful as the restriction due to phoneme information. Moreover, the results of the visual experiment (Wang & Connolly, 2001b) indicate that tonal information has a critical role in reading.

Taken together, brain responses as well as behavioral findings suggest that prosodic information has a crucial role in word processing. However, electrophysiological studies so far addressed the processing of word prosodic information by means of sentence constrains (Hayashi, Imaizumi, Koichi, Niimi, Ueno & Kiritani, 2001; Wang & Connolly, 2001a,b). Their findings can be interpreted as reflecting restricted lexical activation due to prosodic information. This interpretation is only indirect via an electrophysiological correlate of selection effort in sentence context. The ERP-studies presented in the following sections directly address the on-line processing of prosodic parameters for words presented in isolation.

1.4 The current experiments

A growing body of evidence originating from behavioral studies indicates that prosodic information plays a crucial role during spoken word recognition. The the behavioral results reviewed in section 1.2.2 point to a use of prosody during spoken word access processes. They suggest, that prosodic parameters are used for the activation of lexical entries and for the modulation of the activation status during competition. The recording of ERPs might provide additional evidence concerning the temporal structure of prosodic processing during spoken word recognition.

One main goal of the current ERP experiments was to investigate correlates of neural processes which are engaged with the extraction of prosodic parameters from the acoustic signal. The behavioral evidence reviewed in section 1.2.2 implies that prosodic parameters are available early to the recognition system.

Consequently, one might assume an early ERP component differentiating initially stressed and initially unstressed words. **ERP-study I** addresses this issue. Bisyllabic German words, either stressed on the first syllable, such as *AMboss* (Engl.: anvil), or stressed on the second syllable, such as *abTEI* (Engl.: abbey), were presented. However, there was no ERP-deflection that differentiates both types of words. In **ERP-study II** the question was asked, whether the failure to show any difference for initially stressed and initially unstressed words might be caused by latency jitter. Indeed, the results of that study suggest an influence of syllable duration on the electrophysiological signal. Consequently, duration of was controlled for **ERP-study III**. In that study a single prosodic parameter, namely pitch, was manipulated. This artificial modulation of the auditory signal leaves it's remaining properties such as duration and amplitude envelope constant. The results of ERP-study III indicate, that pitch is extracted early by the brain. An early ERP deflection peaking at 200 ms differed in amplitude for words with initially stressed pitch contour and the same words with initially unstressed pitch contour. Thus the present results suggest that at least pitch is available early for the processing of spoken words.

The second focus of the present thesis is more closely related to the role of prosody during spoken word access processes and their ERP correlates. As reported in section 1.3.1, much of the ERP findings in psycholinguistic research have been detected by a comparison between the processing of correct and incorrect language conditions. So far no attempt has been made to investigate electrophysiological effects of incorrect prosodic information leading to an irregular word. **ERP-study I** aimed to answer the question whether incorrect prosody is reflected in the electrophysiological signal. This was realized by presenting words once with correct stress (e.g. *AMboss* or *abTEI*) and once with incorrect stress (e.g., *amBOSS* or *ABtei*). An influence of the stress violation on behavioral responses was established in a **gating task**. Surprisingly, only a stress violation of initially stressed words elicited a reliable ERP effect, starting at 400 ms. Also behavioral results were more robust for initially stressed words in that study. This component might reflect selection effort for incorrectly stressed words as they are not as strongly activated and had more competitors co-activated than correctly stressed words. The different results for initially stressed and initially unstressed words were interpreted as biased by a different realization of the stress manipulation for both types of words. The speaker shifted all prosodic parameters to produce the incorrect versions of initially stressed words, whereas the pitch parameter remained invariant for the correct and incorrect versions of initially unstressed words.

The consequence of incorrect pitch was directly addressed in **ERP-study III**. When an initially unstressed pitch contour was applied to an initially stressed word and, vice versa, when an initially stressed pitch contour was applied to an initially unstressed word, reactions of the subjects were slower than for the same words with correctly applied pitch contour. However, there was no corresponding ERP effect. Again, latency jitter effects in the ERPs might have canceled out subtle effects of incorrect pitch on lexical access.

ERP-study IV attempts to clarify the role of pitch information during spoken word access processes. In this experiment word-onset fragments (e.g. *AM* taken from *AMboss*) were presented with stressed or with unstressed pitch contours. In a **word completion task** it was ensured, that the pitch contour when the fragments is used for spoken word recognition. In ERP-study IV the fragments were immediately followed by visual target words with different stress patterns. Reaction times were faster for visual words with a stress pattern as indexed by the pitch of the prime. Furthermore, an ERP potential named P350 was found to be sensitive to the pitch information. Together with the results of ERP-study I, these results imply that the word initial pitch contour is a crucial parameter for lexical activation.

Chapter 2

Correctly vs. incorrectly stressed words

2.1 Overview

The main objective of the present chapter is the examination of the processing of correct and incorrect word stress. In the first part a gating experiment is described which should reveal whether incorrect word stress modulates spoken word recognition in German. This task was chosen, as a considerable amount of evidence has been obtained which suggests that responses in gating tasks are modulated by suprasegmental information. In the second part an ERP study is introduced in which the online processing of incorrect stress was investigated. In that study the same words as in the gating study were presented. The ERP method provides, in contrast to the behavioral gating technique, a tool to investigate neural correlates of the processing of incorrect stress on the word level. The comparison of ERPs for correctly and incorrectly stressed words might reveal a detailed insight into the timing of use of prosodic information during spoken word recognition.

A number of behavioral studies investigated the effects of incorrect word stress on spoken word recognition (Bond & Small, 1983; Cutler & Clifton, 1984; Koster & Cutler, 1997; Slowiaczek, 1990). Some of these studies used words that did not only differ in their prosody, but also in their segmental information

(Bond & Small, 1983; Cutler & Clifton, 1984; van Heuven, 1985). In these studies responses to initially unstressed words with reduced vowels were compared to responses to initially stressed words with full vowels. Vowel reduction accounts for the fact that certain vowels might undergo changes in unstressed positions. Crucially vowel reduction is the change of a segment which can only occur in unstressed syllables. One vowel quality that shows up quite commonly in vowel reduction is *schwa* (/ə/). Schwa is in a variety of languages, such as German, the only reduced vowel that is used. Consider, for instance, the vowels of the first syllables of the German words *Besen* (Engl.: broom) and *Besitz* (Engl.: holdings). The vowel in *Besen*, which is an initially stressed word, is articulated in its full form /e/. The vowel in *Besitz*, which is an initially unstressed word, is shortened to /ə/.

In a metrical distinction, syllables bearing unreduced vowels are referred to as *strong*, whereas, syllables with reduced vowels are referred to as *weak*. Thus, an *unstressed syllable* can either be *weak*, if it contains a reduced vowel, or *strong*, if it contains a full vowel. A *stressed syllable*, in contrast, is always *strong* (see Chapter 3 for a further discussion of metrical stress). Reduced and full vowels differ in their feature representations. The /e/ in *Besen* is articulated with a fontal place of the tongue, whereas, the /ə/ in *Besitz* is articulated with a central place of the tongue. That is, when a stressed strong syllable is compared to a weak syllable there are next to suprasegmental differences also segmental ones. Thus, although all studies that contrasted strong- and weak initial words report delayed recognition for incorrectly stressed words, no unequivocal conclusion about the role of pure suprasegmental information can be drawn.

An English (Slowiaczek, 1990) and a Dutch study (Koster & Cutler, 1997) controlled for vowel reduction. Whereas in the English study only words with full vowels were presented, a carefully balanced design of correctly and incorrectly stressed words with full and reduced vowels was tested in the Dutch study. Results of both studies indicate that suprasegmental information alone affects word recognition in a way that correctly stressed words are responded to faster than incorrectly stressed words across a number of different paradigms. Moreover, the results of the Dutch study reveal that incorrect suprasegmental information as well as incorrect segmental information quite effectively delays responses.

Taken together, there is evidence from behavioral studies that incorrect stress information hampers the recognition of a spoken word. However, reaction time studies do not reveal the time course of such stress mismatch effects. Empirical evidence form paradigms which are considered to be closer *on-line*, such as cross-modal word fragment priming, suggests that stress information is already used during spoken word access processes (see section 1.2.2.2). Accordingly, an on-line measure of the brain's activity, such as the recording of ERPs, should reveal early presumably access-related differences for the processing of correctly and incorrectly stressed words.

2.2 Gating

2.2.1 Introduction

The first experiment clarifies with the help of the gating technique whether incorrect prosodic information harms spoken recognition in German. In a gating experiment spoken words are presented in segments of successively increasing duration. The first segment usually starts at the word beginning and the last one corresponds to the entire word (Grosjean, 1980, 1996). After listening to each word fragment subjects have to guess the correct word together with a rating of how confident they are about their guess. The data not only provide evidence about the amount of acoustic information that listeners need to correctly name the stimulus, they also reflect lexical hypotheses which listeners entertain when they are presented with fragments of word onset information. Accordingly, gating data can be analyzed for *quantitative* as well as for *qualitative* aspects.

Quantitative measures in gating experiments are the *isolation point* (IP) and the *recognition point* (RP). The IP is defined by the fragment size subjects need to correctly identify a word without changing their choice thereafter (Grosjean, 1980). The RP is defined by the fragment size subjects need after the IP to reach a particular confidence level without a change in their response thereafter (Tyler &

Wessels, 1983). It is not clear whether the RP reflects a word's actual recognition (Grosjean, 1996) and so far effects for the RP equal those of the IP. It is assumed that the IP and the RP may well reflect different processing stages of spoken word recognition (Tyler & Wessels, 1983), however, this issue is to date not answered satisfactorily (Grosjean, 1996). Words which are erroneously produced before the IP can be analyzed qualitatively in terms of frequency or segmental and suprasegmental structure. The results suggest to which extend fragment information guides the formation of hypotheses entertained by the subjects.

Several gating studies investigated effects of prosodic information. Most of these studies analyzed qualitative aspects of gating data. Error responses in Dutch (Jongenburger, 1996), English (Wingfield, Goodglass & Lindfield, 1997) and Japanese (Cutler & van Donselaar, 2001) reveal an influence of the fragments' prosody. In the Dutch and the Japanese study subjects' guesses were more frequently initially stressed for fragments from initially stressed words than for fragments from initially unstressed words. Vice versa, word completion was more frequently initially unstressed for fragments from initially unstressed words. In line with this, Japanese listeners are able to limit their word guesses or alternatives which match the initial pitch accent pattern of the input. Closely related to gating studies are the results of cued word completion tasks (Cooper, Cutler & Wales, 2002; Cutler & van Donselaar, 2001; Cutler & Otake, 1999; Mattys, 2000). In cued word completion tasks a guess has to be made towards the origin of a segmentally ambiguous word fragment. For example, *MUs* from *MUsic* is presented and subjects have to guess whether the origin of the fragment was *music* or *museum*. It has been shown that guesses of English (Cooper, Cutler & Wales, 2002; Mattys, 2000), Japanese (Cutler & Otake, 1999) and Dutch listeners (Cutler & van Donselaar, 2001) exceed chance levels, indicating that it is possible to make fine stress distinctions.

A quantitative approach to gating data was chosen by Lindfield, Wingfield and Goodglass (Lindfield, Wingfield & Goodglass, 1999). Three different types of word fragments were tested: *A* Isolated fragments as in the previous studies, *B* Fragments in which the correct word duration, as indicated by white noise was added, and *C* Fragments followed by the correct prosody of the stimulus word, as indicated by a low-pass filtered signal. Subjects needed less segmental information to correctly identify a word of condition *C* as compared to *A* and *B*. That is, the amount of segmental information needed to correctly identify a word can be reduced when extra prosodic information is given to the subjects. It should be mentioned, however, that in a low-pass filtered signal not only prosodic information, but also information about the number of syllables was given, and, thus, the facilitation effect for condition *C* might be increased due to syllabic information. To summarize, suprasegmental information is reported to modulate word completion in gating and in cued word completion tasks.

The present gating experiment investigated whether more segmental information is needed to identify an incorrectly stressed word in relation to its correctly stressed counterpart. If the fragments' prosody is used for word completion, IP and RP should be delayed for words with incorrect prosody as compared to the same words with correct prosody. Moreover, if prosody guides the formation of hypotheses for the correct word completion, fragments taken from incorrectly stressed words might favor lexical hypotheses with a stress pattern that differs from the stress pattern of the original word. That is, fragments of incorrect versions of initially stressed words should bias word completion by favoring initially unstressed words. Vice versa, fragments of incorrect versions of initially unstressed fragments should more frequently elicit initially stressed word completion than the correct fragments of these words.

2.2.2 Method

Subjects 20 subjects with no reported speech or hearing deficits participated in the gating experiment. All subjects were native speaker of German. As in all following experiments, subjects were paid a small fee for their participation (13 DM which equals 6.50 EUR per hour).

Material 80 initially stressed German words (e.g., BALsam [Engl.: balm]) and 80 initially unstressed German words (e.g., akTEUR [Engl.: participant]) were used. All words were bisyllabic concrete German nouns which contained only full vowels. A complete list of the words is presented in Appendix A. Initially stressed and unstressed words did not differ in their frequency in spoken language according to the CELEX lexical data base for German (Baayen, Piepenbrock & van Rijn, 1995).

A trained female native German actress spoke all words once with correct stress and once with incorrect stress. The recording was done in a sound-proven room. The manipulation of stress was practiced before the recording session. For an incorrectly stressed word the speaker was instructed to reverse the stress pattern of that word. For example, to produce the incorrectly stressed version of the initial stressed German word *AMboss* (Engl.: anvil) the speaker was required to stress the second syllable stronger than the first, such as in the German word *abTEI* (Engl.: abbey). Vice versa, to produce the incorrect version of the initially unstressed German word *abTEI* the speaker should stress the first syllable stronger than the second, such as in the word *AMboss* (see Figure 2.1 for an example of the stress manipulation for the word GUru [Engl.: guru]). All words were recorded onto Digital Audio tape and digitized at 44.1 KHz.

Acoustic analyses The digitized words were analyzed with regard to their main prosodic properties. Duration, pitch and amplitude features were analyzed with a sound editor (WinPitch v.1.6d ©Pitch Instruments Inc.).

Incorrectly stressed words produced by the speakers were significantly longer (M = 853 ms, SD = 77) than correctly stressed words (M = 827 ms, SD = 82, $t[79] = 3.54$, $p < .001$). This is a critical result for time sensitive measures such as the IP and the RP. The only comparable time interval with identical phoneme information for correctly and incorrectly stressed words is the length of the first syllable. The duration of the first syllables of the correctly stressed words (M = 332 ms, SD = 54) did not significantly differ from the duration of the first syllables of the incorrectly stressed words (M = 327 ms, SD = 49, $t[79] = 1.17$, $p = .25$).

	first syllable		second syllable	
	M	(SD)	M	(SD)
initially stressed				
correct stress				
duration (ms)	367	(84)	438	(118)
F_0 (Hz)	177	(49)	149	(40)
amplitude (dB)	38.0	(2.1)	35.1	(3.0)
incorrect stress				
duration (ms)	375	(85)	558	(96)
F_0 (Hz)	157	(42)	162	(53)
amplitude (dB)	36.4	(4.6)	36.9	(2.2)
initially unstressed				
correct stress				
duration (ms)	8297	(64)	552	(84)
F_0 (Hz)	170	(41)	154	(38)
amplitude (dB)	36.6	(2.0)	36.8	(2.1)
incorrect stress				
duration (ms)	278	(59)	494	(91)
F_0 (Hz)	169	(30)	155	(38)
amplitude (dB)	37.5	(1.4)	34.9	(2.6)

Table 2.1: *Prosodic realization of the first and the second syllable for correct and incorrect versions of initially stressed and initially unstressed words presented in the gating Experiment and in ERP-study I*

Prosodic relation of the stress pattern: Table 2.1 summarizes acoustic properties that are discussed to mark stress for the first and the second syllable of initially stressed and initially unstressed words with correct and incorrect stress. The stress pattern of a word can be analyzed in terms of the relation between the core prosodic parameters duration, pitch and amplitude of the single syllables. The acoustic data indicate, that the second syllables of all words were longer than the first syllables. This phenomenon might be caused by the tendency to mark the end of a word with final lengthening. As intended, the second syllables of initially stressed words were lengthened by the speaker to produce the incorrect stress pattern. Vice versa, second syllables of initially unstressed words were shortened for an incorrect stress pattern. Unfortunately, a variation of the fundamental frequency (F_0), which is perceived as pitch, was only realized for the incorrect version of the initially stressed words, but not for that of the initially unstressed words. F_0 was higher for the first syllables than for the second syllable of correctly stressed words with initial stress. The reversed duration pattern was realized for their incorrect counterparts. In contrast, the pitch of the first and second syllables of initially unstressed words with correct stress equaled that of the same words with incorrect stress. As can be seen in Table 2.1, the first syllables of initially stressed words with correct stress were spoken louder than their second syllables and this pattern was reversed for the same words with incorrect stress. Similarly, the intended variation of duration was observed for initially unstressed words. The second syllables of such with correct stress were spoken louder than the first syllables, whereas the same words with incorrect stress were characterized by the reversed pattern. Taken together, for all words duration and amplitude patterns across syllables are reversed for the incorrect as compared to the correct version. However, the F_0 pattern was only modulated for the stress violation of initially stressed words.

Prosodic realization of the first syllables: If one considers only the first syllables, which already provide important information for spoken word access processes (see section 1.2), it is clear that the realization of the stress manipulation again slightly differed for initially stressed and initially unstressed words. There was no difference in the length of the first syllables of initially stressed words with correct and incorrect stress. Moreover, the first syllables of initially unstressed words with incorrect stress which should have been longer than that of the correct version were in fact shorter. Again, a variation of pitch between correctly and incorrectly stressed words was only observed for initially stressed words. Such with correct stress had higher F_0 for the first syllables as compared to such with incorrect stress. There was no difference in F_0 for the first syllables of initially unstressed words in the correct or incorrect version. As intended the first syllables of initially stressed words with correct stress were spoken louder than the first syllables of the same words with incorrect stress. Vice versa, the first syllables of initially unstressed words with correct stress were spoken softer than the first syllables of the same words with incorrect stress.

To summarize, the speaker varied only the amplitude of the first syllable in the intended way. The expected pitch contrast was only established for the first syllables of initially stressed words in the correct and incorrect version, whereas, duration did not vary for the first syllables of words with correct and incorrect stress. Thus it may well be that the prosodic information of the incorrectly stressed words was more ambiguous than that of the correctly stressed words. Furthermore, the realization of the stress manipulation differs for initially stressed and initially unstressed words. Whereas the pitch of the initially stressed words was modulated for the incorrectly stressed version, the pitch pattern of the initially unstressed words was not. Accordingly, effects of the stress manipulation might differ for initially stressed and unstressed words. This is accounted for in the statistical analyses with a separate factor which contrasts effects of the stress manipulation for initially stressed and initially unstressed words.

Procedure Gating of all words began with fragments of 50 ms. Each successive fragment was 50 ms longer. The longest fragment represented the entire word. Each subject heard fragments of one word either in its correct or in its incorrect version. This was done to avoid repetition effects of segmentally identical words. Half of the words a subject heard were correctly stressed, the other words were incorrectly stressed. The presentation of correctly and incorrectly stressed words was randomized across subjects.

Subjects were run individually. They were instructed to listen to each presentation, to name the word they thought was being presented, and to indicate on a scale of 0 to 100 percent (very unsure/very sure) how confident they were about their guess. They were asked to give a response after each presentation, even though at first they might feel very unsure about it. The responses were recorded on audio tape and written down by the experimenter.

Statistical analysis Due to the significant differences in duration for correctly and incorrectly stressed words (see section 2.2.2) the isolation and recognition points could not be directly compared. Thus, the IP and the RP were related to phonemes. For each word the mean IP and the mean RP across all 20 subjects was located in the waveform and the corresponding phoneme was classified. This procedure resulted in a decimal variable. The nominal position of this variable coded the sequential position the phoneme were the IP or RP fell into. The tenth position coded the location of the IP or RP within that phoneme, *0* coded the first half of the phoneme, *5* the second half. For an example see Figure 2.1.

The mean IP of the correctly stressed word *Guru* falls in the second half of the second phoneme /u:/. This IP was recorded as *2.5*. In contrast, the mean IP of the incorrectly stressed version of *Guru* fell into the second half of the /R/ and is, therefore, coded as *3.5*. The calculated decimal variable was analyzed in a two-way ANOVA including the factors *original stress of the gated word* (initially stressed vs. initially unstressed) and *stress manipulation* (correctly stressed vs. incorrectly stressed).

For the analysis of the stress pattern of erroneously produced words, the stress pattern of polysyllabic words that were produced before the isolation point was determined. Words stressed on the first syllables were classified as initially stressed other polysyllabic words were classified as initially unstressed. The number of initially stressed error responses was added across all subjects. The same was done for the number of initially unstressed error responses. A three-way ANOVA including the factors *stress of the produced word* (initially stressed vs. initially unstressed), *original stress of the gated word* (initially stressed vs. initially un-

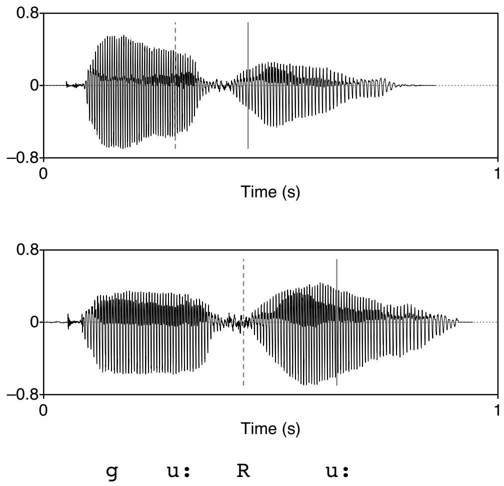

Figure 2.1: *Example of the stress manipulation for the German word 'Guru' (Engl.: guru). The upper waveform illustrates the correct version of 'Guru' which is initially stressed. The incorrectly stressed version is illustrated below. The mean IP for each version is marked by a vertical dashed line, the RP is marked by a vertical solid line.*

stressed) and *stress manipulation* (correctly vs. incorrectly stressed) was conducted. All statistical analyses of the gating experiment were conducted across items. This was done as each subject only heard half of the items (each word either in it's correct or in it's incorrect version, see section 2.2.2).

2.2.3 Results

IP and RP The mean IP was at 2.9 phonemes (SD = 0.55), the mean RP at 3.8 phonemes (SD = 0.51). Significant effects of the factors *stress manipulation* ($F[1,79] = 27.36$, $p < .001$) and *original stress of the gated word* ($F[1,79] = 18.83$, $p < .001$) were found for the two-way ANOVA of the IP. As expected, words with incorrect stress were isolated later, i.e. after 3.0 phonemes (SD = 0.6), than the same words with correct stress, which were isolated after 2.8 phonemes (SD = 0.6). Initially stressed words were isolated earlier (M = 2.7 phonemes, SD = 0.6]) than initially unstressed words (M = 3.2 phonemes, SD = 0.8) regardless of the stress manipulation. There were no significant interactions of the factors *stress manipulation* and *original stress of the gated word*.

The ANOVA of the RPs revealed significant effects of the factor *stress manipulation* ($F[1,79] = 24.93$, $p < .001$), of the factor *original stress of the gated word* ($F[1,79] = 19.16$, $p < .001$) and a significant interaction of both factors ($F[1,79] = 4.70$, $p = .03$). Initially stressed words with correct stress were recognized after 3.5 phonemes (SD = 0.7), such with incorrect stress were recognized after 3.6 phonemes (SD = 0.7). Initially unstressed words with correct stress were recognized after 3.9 phonemes (SD = 0.7). The recognition of initially unstressed words with incorrect stress needed 4.1 phonemes (SD = 0.6). The stress manipulation was effective for initially stressed words ($t[79] = 2.02$, $p = .05$) as well as for initially unstressed words ($t[79] = 5.10$, $p < .001$). Taken together, isolation and recognition of the correct word was delayed by incorrect stress. Isolation took approximately a fifth of a phoneme longer for words with incorrect stress as compared to the same words with correct stress. Recognition took a tenth of a phoneme longer for incorrectly as compared to correctly stressed words.

Stress pattern of error responses In this section the analysis of the stress of the words that were produced before the IP will be described. There was a significant effect of the factor *stress manipulation* ($F[1,79] = 11.91$, $p < .001$) revealing that more words (M = 4.6 words, SD = 3.7) were produced for word fragments with incorrect stress as compared to fragments with correct stress (M = 4.1 words, SD = 3.7). Note that this result includes initially stressed as well as initially unstressed words. That is, more word-alternatives of all stress patterns were produced for the incorrectly stressed fragments as compared to the correctly stressed fragments.

All in all, there was a main effect of the factor *stress of the produced words* ($F[1,79] = 282.02$, $p < .001$) indicating a predominance of initially stressed words. On average 13.7 different initially stressed words (SD = 6.3) and only 3.7 initially unstressed words (SD = 4.2) were produced before the IP of the correct word. An interaction of the factors *stress of the produced words* and *stress of the gated* *word* (F[1,79] = 5.44, p = .02) revealed that the production of initially unstressed words, but not the production of initially stressed words differed for words that were originally stressed or unstressed on their first syllables. More initially unstressed words were produced for fragments from initially unstressed words (M = 4.7 words, SD = 4.8) than for fragments from initially stressed words (M = 2.7 words, SD = 3.3, t[79] = 11.91, p $<$ .001) regardless of the stress manipulation.

Error responses for the first syllables Although there was a clear effect of the stress manipulation it could be argued that this effect is confounded by the longer duration of the incorrectly stressed words. As less lexical information is provided in equal time intervals for the incorrectly as compared to the correctly stressed words, fewer fragments were presented up to the IP of the correctly stressed words. Consequently, subjects might have guessed more word alternatives for the incorrectly stressed words as there were more responses possible. To rule out this possibility the same analysis was conducted for the stress pattern of words that were produced for gates up to the first syllable of the words. As already pointed out (see section 2.2.2), the duration of the first syllables did not differ for correctly and incorrectly stressed words. Thus, the number of gates and the amount of segmental information is identical up to the syllable boundary.

Again, there were significant effects of the factors *stress manipulation* (F[1,79] = 5.95, p = .02) and *stress pattern of the produced words* (F[1,79] = 384.48, p $<$ 001) and a significant interaction of both factors (F[1,79] = 5.24, p = .02). The former effect reveals that more words (M = 7.5 words, SD = 8.1) were produced for fragments from words with incorrect stress as compared to the same fragments with correct stress (M = 8.1 words, SD = 7.2). As this effect is not confounded with duration it clearly confirms greater variability of word completion for fragments with incorrect stress as compared to the same fragments with correct stress. Again, the second main effect indicates a predominant production of initially stressed words. On average 12.9 initially stressed words (SD = 6.0) and only 2.7 initially unstressed words (SD = 3.1) were produced erroneously, whereas the interaction of both factors revealed that more initially unstressed words were produced for the first syllables of initially unstressed words (M = 3.2 words, SD = 3.2) than for the first syllables of initially stressed words (M = 2.2 words, SD = 2.9, t[79] = 5.63, p = .02) regardless of the stress manipulation. No difference was found for the production of initially stressed words. To summarize, there was more variability of the produced words for incorrectly stressed word fragments than for correctly stressed word fragments.

2.2.4 Discussion

A novel aspect of the present experiment was the presentation of incorrectly stressed words in a gating task. The quantitative gating data clearly show the intended effect of the stress manipulation: Incorrectly stressed words are isolated and recognized later than correctly stressed words. Delayed IPs and RPs reveal that the acoustic information of incorrectly stressed fragments is perceptually longer ambiguous for the listener than the acoustic information of correctly stressed words. Following the present data, it can be generalized that incorrect prosody harms spoken word recognition not only in English (Slowiaczek, 1990) and Dutch (Koster & Cutler, 1997), but also in German. Thus, an implicit presumption of the experiments presented in the following experiments has been proofed by the results of the present gating task.

The quantitative data of the gating task are in the same line as the finding of Lindfiled, Wingfield and Goodglass (1999). These authors report that listeners need less information to identify a word in a gating task when the fragment and the prosody of the rest of the word is presented than when the fragment alone is presented. The data of Lindfield et al. reveal that prosodic information facilitates word recognition, whereas the present data indicate that incorrect prosody hampers this cognitive operation. Thus, the present results reveal, that the presentation of incorrectly stressed words in a gating task is a useful tool to investigate the processing of prosodic information during spoken word recognition. However, by presenting incorrectly stressed words one have to take duration effects of the single segments into account. In the present study this problem was solved by relating subjects responses to the phoneme information of the word and not on its temporal characteristics.

Although expected, the qualitative results of the gating task do neither reveal that fragments from incorrectly stressed weak-initial words invoke the production of initially stressed words, nor that fragments from incorrectly stressed strong-initial words invoke the production of initially unstressed words. However, they indicate that incorrectly stressed words are perceptually more ambiguous than correctly stressed words. There was more variance in the word alternatives that were produced for incorrectly stressed fragments than for the same fragments with correct stress. Unfortunately, these result can not be taken to draw conclusions about the phase of spoken word recognition that is effected by the stress manipulation. It has been doubt that the gating task is a clear on-line paradigm (see Grosjean, 1996). Thus, the greater variability of answers for incorrectly than for correctly stressed words might - as well as the delayed recognition time - reflect modulated initial activation and competition as well as modulated post-access processing.

In general, subjects produced more initially stressed than initially unstressed words. Jongenburger (1996) reports an equal result for a Dutch gating experiment. The preference for initially stressed word completion might be caused by the fact that although, the position of the stressed syllable can vary in German and Dutch (see section 1.1) it falls more often on word initial syllables than on any other position (Dogil, 1999; van Heuven & Hagman, 1988; Schreuder & Baayen, 1994). The same holds true for English (Cutler & Carter, 1987). It has been shown that listeners have information about the likelihood of a particular stress pattern available as English participants judge initially stressed nonsense words more English-like and process them more quickly than initially unstressed nonsense words (Vitevitch, Luce, Charles-Luce & Kemmerer, 1997). Accordingly, subjects in the current gating task as well as subjects in the Dutch study (Jongenburger, 1996) may well have biased their responses toward the more likely initially stressed words. Furthermore, the predominance of initially stressed words in German is apparently reflected in the competitors which begin with the same segments as the presented words. According to CELEX, each word (e.g., AMboss

[Engl.: anvil]) presented in the current gating experiment shares its initial syllable with on average 110 words (e.g., AMsel [Engl.: blackbird] or ambiENte [Engl.: ambience]). Most of these words (on average 61 words) are initially stressed, fewer (on average 49 words) are initially unstressed. Both, the predominance of initially stressed words in German and the initial segmental information of the words clearly favored the production of initially stressed words in the present experiment.

Even the predominance of initially unstressed word completion for fragments from initially unstressed words could be clearly attributed to the segmental information. Sixty-two competitors of the initially unstressed words were not stressed on the first syllable, whereas, only 48 segmental competitors of the initially stressed words were not stressed on the first syllable. That is, the prosodic information is clearly confounded with segmental information and, thus, word completion for fragments from initially stressed words cannot be meaningfully compared to word completion for fragments from initially unstressed words. It should be noted that a confounding like that is hard to avoid in German, which as already pointed out, has different amounts of initially stressed and initially unstressed words, and which has correlation between segmental and suprasegmental characteristics (e.g. Fery, 1998).

To summarize, the aim of the present study was to investigate whether incorrect stress harms spoken word recognition. Both, the qualitative and the quantitative gating data support this assumption. In line with results of previous gating tasks (Cutler & Otake, 1999; Jongenburger, 1996) and word completion tasks (Cooper, Cutler & Wales, 2002; Cutler & van Donselaar, 2001; Mattys, 2000) the current results indicate that subjects' guesses in word completion tasks are sensitive to prosodic information. This can be taken as a hint that word prosody plays a role during word recognition in German. Thus, the materials presented in the present gating study may well allow to investigate neural correlates of prosodic processing as recorded with ERPs (see ERP-study I). In that study the neural activity underlying the correlated to . The ERP is a clear online method which records neural activity as it unfolds over time.

The ERPs show as well as the gating data effects of the stress manipulation. However, these effects were separable in earlier and later ones. The later effect paralleled that of the gating data: a negative ERP deflection for all words is enhanced for incorrectly stressed words as compared to correctly stressed words. Furthermore, however, the ERP data show earlier effects, which show up only for the stress manipulation of initially stressed words. This is in line with the slightly different realization of the stress manipulation of initially stressed and initially unstressed words (see section 2.2.2). Taken the results of the gating task and ERP-study I together, it appears that responses in a gating task might not reflect the initial processing of spoken words.

2.3 ERP-study I

2.3.1 Introduction

ERP-study I investigates the processing of correctly and incorrectly stressed words with an on-line method, that is with the recording of ERPs. Words from the gating experiment were presented in two lexical tasks. In a prosodic judgment task, subjects had to decided whether a word was correctly stressed or not. In a semantic judgment task, subjects were asked to indicate whether a target word could be categorized as living or non-living. Both tasks clearly require successful word recognition.

The present ERP-experiment was designed to investigate the two main research questions of the present thesis (see section 1.4). A main issue is concerned with the time point of the differentiation of words with different stress patterns. The early extraction of prosodic information appears to be a necessary presumption for its use during spoken word access processes. ERPs might be powerful to discriminate neural processing concerned with the identification of the stress pattern of a word. Indeed, previous ERP research revealed promising results in that respect. Böcker, Baastiansen, Vroomen, Brunia and de Gelder (1999) report a negative ERP response peaking at around 325 ms which is higher in amplitude for weak-initial words as compared to strong-initial words with initial stress. This result was interpreted as reflecting prosodic processing of the words. However, as already pointed out in the introduction of the present chapter (see section 2.1), strong stressed and weak (necessarily unstressed) syllables are not only distinctive on a prosodic level. They also differ in terms of vowel quality. Strong syllables bear full vowels, whereas weak syllables are characterized by reduced vowels. Thus, vowel quality and prosody are confounded in the design of Böcker et al.. This drawback is prevented in the presented study. Only words with full vowels are presented. The results can be related to the findings of Böcker et al.. If there is again an enhanced N325 for initially unstressed words, this effect might well reflect prosodic processing. If there is no N325 effect in the present experiment, it can be argued that this ERP deflection is rather related to vowel quality than to prosodic parameters.

The second main research question of the present experiment is related to the processing of incorrect prosody. Following behavioral research, incorrectly stressed words should be harmed in lexical activation and/or competition (see section 1.2). Consequently, the ERP might be modulated in an access related component by incorrectly stressed words. The design of the present experiment allows to explore this issue. As in the gating experiment, all words were presented twice, once with correct stress and once with incorrect stress. The gating experiment revealed that incorrectly stressed words are recognized later than the same words with correct stress. However, these results were not clear with respect to the exact location of this effect during spoken word recognition.

In the present experiment subjects were engaged in a task that explicitly focuses on prosodic processing (the prosodic judgment task) and in a task that does not require conscious prosodic analysis (the semantic judgment). Effects that are only observed for the former task might not reflect automatic processing of prosody. Thus, the differentiation of ERP effects evoked by both tasks might be taken to relate them to processes commonly involved in natural spoken word recognition and such that are modulated by strategies induced by the experimental manipulation.

2.3.2 Method

Subjects Thirty-six right-handed subjects (16 women) aged 19-30 years (M = 23.3 years) participated in the experiment. All were native German speakers with no reported hearing or neurological deficits.

Materials All words presented in the Gating experiment were used (see section 2.2.2 for a detailed description of the material and Appendix A for a complete list). It should be noted that the stress manipulation slightly differed for initially stressed and initially unstressed words (see section 2.2.2). Again, two lists were created. One contained half of the words with correct stress and the others with incorrect stress. The other contained the words that were incorrectly stressed in the former list in their correct version and the others in their incorrect version.

Procedure Subjects were seated in an electrically and acoustically shielded chamber in front of a computer screen. Loudspeakers were placed to the left and to the right side of the screen. An experimental trial began with the presentation of a fixation cross in the center of the screen. Participants were instructed to fixate this cross at any time it appeared on the screen. After 500 ms a word was presented via loudspeaker while the fixation cross remained on the screen. Subjects were instructed to listen to the word and to indicate as fast as possible whether the word can be characterized as living or non-living (semantic judgment) or whether the word was correctly stressed or not (prosodic judgment). The semantic judgment was required in one task, the prosodic judgment in another task. Half of the subjects performed the semantic judgment first. The other half performed the prosodic judgment first. The fixation cross vanished 2500 ms post-stimulus onset. RTs were measured from the beginning up to 2500 ms after stimulus onset. Subjects responded with the thumbs of both hands. One hand indicated a yes-response the other indicated a no-response. The response hand was counterbalanced across subjects. After a variable inter-stimulus interval (1000-2000 ms) the next trial began.

Electrophysiological Recordings All ERP studies reported in this thesis used an electrode setup that followed the upgraded 10-20 system (Pivik, Broughton, Coppola, Davidson, Fox & Nuwer, 1993). This standardized system specifies electrode places in terms of their proximity to particular regions of the brain (see Figure 2.2). Frontal placements are labeled with *F* (FP indicates the frontal pole). Temporal placements are labeled with *T*, central ones with *C*, parietal ones with *P* and occipital ones with *O*. Odd numbers indicate left electrode placements, even numbers indicate electrodes placed at the right hemisphere and *Z* is used for electrodes places in the midline. For data reduction and power enhancement purpose electrode sites were summarized to *regions of interest* (ROIs) in the statistical analyses. As the scalp distribution of effects varied, especially depending on the applied paradigms and the ERP deflections of interest, ROIs were determined individually for the ERP-studies.

Electrophysiological recordings were standardized across all ERP-studies. The EEG was continuously recorded (250Hz; DC amplifier by Twente Medical Systems) from 58 Ag AgCl cap-mounted electrodes (Electro Cap International, Inc.) and from two separate Ag AgCl electrodes placed at the right and the left mastoid (referred to as A1 and A2 in Figure 2.2). All electrodes were referenced against the nose tip (referred to as NZ in Figure 2.2). To control for eye movements, four further electrodes provided bipolar recordings of the horizontal and vertical electro-occulogram (EOG). The vertical EOG (referred to as EOGV in Figure 2.2) was recorded from two Ag AgCL electrodes located above and below the right eye. The horizontal EOG (referred to as EOGH in Figure 2.2) was recorded from two electrodes positioned at the outer canthus of each eye. The ground electrode was placed at the sternum. Impedance was kept below 5 kΩ. Artifacts caused by facial and eye movements and by drifts were rejected off-line. Only correctly responded artifact-free trials were made were included in the ERP average. ERPs were averaged time-locked to the presentation of the stimulus of interest, relative to a 200-ms baseline. For illustration purpose only, the ERPs were smoothed off-line using a 10 Hz lowpass filter.

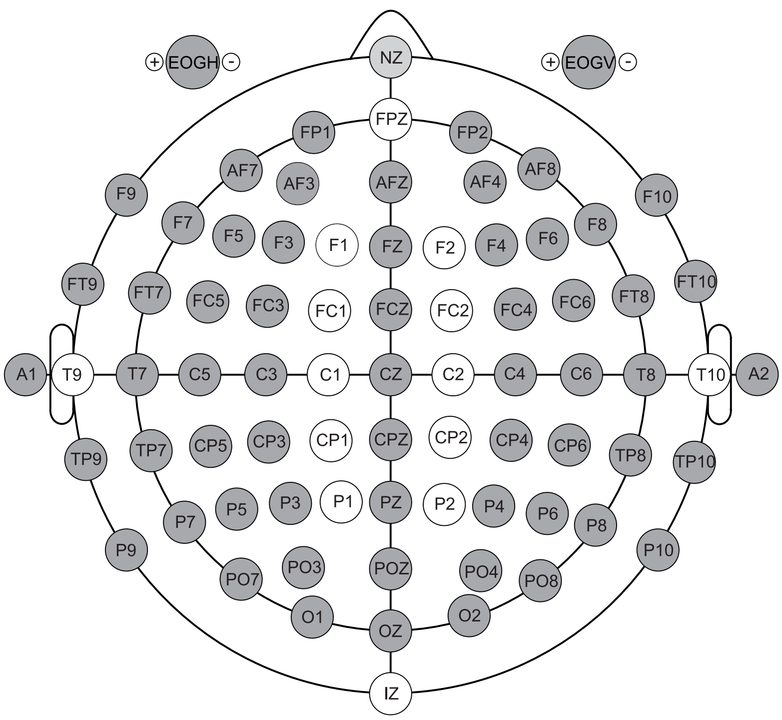

Figure 2.2: *10-20 system for electrode placement. Electrodes recorded in the present ERP studies are marked in gray. See text for a detailed explanation.*

Statistical analysis RTs were corrected for word length by subtracting this variable from each individual response. The percentage of correctly responded trials and the corresponding RTs were analyzed with repeated-measures analyses of variance (ANOVAs). Within-factors were *stress manipulation* (correctly stressed vs. incorrectly stressed), *original stress* (initially stressed vs. initially unstressed) and *task* (semantic decision vs. prosodic decision).

Five within-factors were analyzed in a repeated-measures ANOVA. As for the behavioral data the factors *stress manipulation*, *original stress* and *task* were analyzed. Two further factors served to explore the spatial distribution possible

ERP effects. Mean amplitudes of 36 electrode positions were included (see Figure 2.2). The factor *hemisphere* differentiates left and right electrode positions, the factor *region* anterior and posterior electrode positions. Both factors were based on the following ROIs: anterior left (FP1, AF7, AF3, F9, F7, F5, F3, FT7, FT9, FC5, FC3), anterior right (FP2, AF4, AF8, F4, F6, F10, F8, FC4, FC6, FT8, FT10), posterior left (TP9, TP7, CP5, CP3, P9, P7, P5, P3, PO7, PO3, O1) and posterior right (CP4, CP6, TP8, TP10, P4, P6, P8, P10, PO4, PO8, O2). All analyses were conducted across subjects.

2.3.3 Results

2.3.3.1 Behavioral data

Percent correct Across both tasks subjects' accuracy was high (M = 94.3 %, SD = 4.0). More errors were made for the prosodic judgment (M = 90.4 % correct, SD = 11.4) than for the semantic judgment (M=94.3 % correct, SD = 4.3). The effect of the factor *task* is significant at $p < .001$ level (F[1,31] = 19.11). Furthermore significant main effects of the factors *stress manipulation* (F[1,30] = 26.33, $p < .001$) and of the factor stress of the original word (F[1,31] = 15.15, $p < .001$) were found. However, significant two-way interactions of the factors *task* and *stress manipulation* (F[1,31] = 15.12, $p < .001$) and of the factors *original stress* and *task* (F[1,31] = 9.55, $p < .001$), and a significant three-way interaction of all three experimental factors (F[1,31] = 4.72, $p < .04$) indicate a complex scenario of the observed effects. The description of these effects focuses on the stress manipulation, which was apparently more effective in the prosodic decision task. The percentage of correct responses was greater for correctly stressed words (M = 94.8 %, SD = 3.8) than for incorrectly stressed words (M = 86.0 %, SD = 10.8, t[31] = 21.35, $p < .001$) in the prosodic decision task. Although there was the same tendency in the semantic decision task, incorrectly stressed words were not responded to significantly more erroneously (M = 93.9 % correct, SD = 2.6) than correctly stressed words (M = 94.7 % correct, SD = 3.4, t[31] = 2.55, p = .12). The three-way interaction is illustrated in Table 2.2.

	initially stressed		initially unstressed	
	M	(SD)	M	(SD)
semantic decision				
correctly stressed	95.5	(3.9)	93.8	(5.0)
incorrectly stressed	94.0	(3.9)	93.9	(4.5)
prosodic decision				
correctly stressed	95.6	(4.9)	94.0	(5.0)
incorrectly stressed	90.5	(9.2)	81.5	(16.6)

Table 2.2: *Correct responses (in percent) for correctly and incorrectly stressed words in both task split up for initially stressed and initially unstressed words*

The stress manipulation succeeded for the initially stressed words in both tasks (semantic judgment: t[31] = 3.33, p = .07, prosodic judgment: t[31] = 7.54, p = .01). However, no effect of the stress manipulation for initially unstressed words in the semantic judgment (t[31] = 0.01) contrasts a large effect in the prosodic judgment (t[31] = 17.00, $p < .001$). If there are effects of the pitch manipulation they are in the direction that, incorrectly stressed words were responded to more erroneously than correctly stressed words (compare Table 2.2).

RTs Subjects reacted on average 180 ms after a word ended. Prosodic judgments took longer (M = 438 ms, SD = 178) than semantic judgments (M = 259 ms, SD = 129 ms, F[1,31] = 88.28 $p < .001$). Again, there were significant main effects of the factors *stress manipulation* (F[1,31] = 12.26, p = .001) and *original stress* (F[1,31] = 73.14, $p < .001$) and significant interactions of the factors *task* and *stress manipulation* (F[1,31] = 6.06, p = .02), of the factors *original stress* and *stress manipulation* (F[1,30] = 60.35, $p < .001$) and of the factors *task* and *original stress* (F[1,31] = 20.83, $p < .001$), but there was no significant interaction of all three factors (F[1,31] = 0.72). Nevertheless, the single cells of the three-way analysis of the reaction times are summarized in Table 2.3 to enable a comparison between RTs effects on one hand and response accuracy on the other (see Table

2.2).

	initially stressed M	(SD)	initially unstressed M	(SD)
semantic decision				
correctly stressed	259	(131)	252	(128)
incorrectly stressed	234	(137)	292	(119)
prosodic decision				
correctly stressed	387	(131)	423	(162)
incorrectly stressed	411	(190)	533	(194)

Table 2.3: *RTs corrected for word length for correctly and incorrectly stressed words in both task split up for initially stressed and initially unstressed words*

The RTs as well as the correct responses reveal that the stress manipulation influenced the prosodic decision ($t[31] = 9.48$, $p < .01$) stronger than the semantic decision ($t[31] = 1.17$, $p = .28$). Incorrectly stressed words were responded to slower (prosodic decision: M = 472 ms, SD = 187, semantic decision: M = 265 ms, SD = 125) than correctly stressed words (prosodic decision: M = 405 ms, SD = 143, semantic decision: M = 258 ms, SD = 126). Across both tasks, the stress manipulation affected reaction times only for initially unstressed words ($t[31] = 36.76$, $p < .001$), but not for initially stressed words ($t[31] = 0.00$). Initially unstressed words with correct stress were responded to faster (M = 337 ms, SD = 131) than those with incorrect stress (M = 412 ms, SD = 143).

2.3.3.2 ERPs

The ERP waveform elicited in the present experiment is characterized by an P1-N1-P2 complex followed by an N400 component (see Figures 2.3 - 2.7). The waveform elicited in the present experiment is a typical one found for the presentation of spoken words (see section 1.3.2 for a detailed description).

There were no effects of stress or the stress manipulation in the early exogenous components. Furthermore, there was no ERP effect differentiating initially stressed and initially unstressed words spoken in their correct versions (see Figure 2.3). The N400 component reflected the stress violation (see Figure 2.4, 2.5, 2.6 and 2.7). However, there appear to be two different N400 effects. An earlier one, which ranges from 400 to 600 ms, is only observable for initially stressed words. The later effect occurred for both, initially stressed and initially unstressed words and ranges from 600 to 1000 ms. Both time windows were analyzed. Only significant main effects of the factor *original stress* or of the factor *stress manipulation*, and significant interactions with these factors are reported.

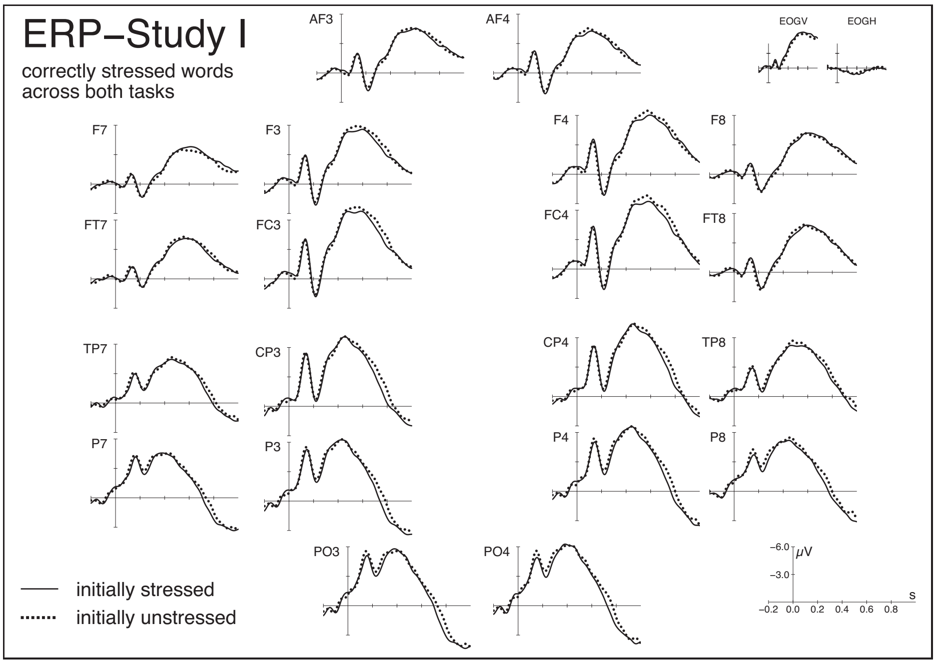

Figure 2.3: *ERPs for selected electrode sites. Solid lines indicate ERPs for initially stressed words in their correct version and dotted lines indicate ERPs for initially unstressed words in their correct version across both tasks.*

400 to 600 ms: N400a There were significant interactions of the factors *stress manipulation*, *original stress* and *hemisphere* ($F[1,31] = 7.08$, $p = .01$) and of the factors *stress manipulation*, *original stress*, *task* and *region* ($F[1,31] = 9.96$, $p < .01$). The first interaction reveals a significant effect of the stress manipulation for initially stressed words over the right hemisphere ($t[31] = 6.36$, $p = .02$). Initially stressed words with incorrect stress across both tasks elicited more negative ERPs ($M = -5.2\,\mu V$, $SD = 4.0$) than the same words with correct stress ($M = -4.4\,\mu V$, $SD = 4.0$, compare Figures 2.4 and 2.5). There was no significant effect for the stress manipulation of initially stressed words for the left hemisphere ($t[31] = 2.71$, $p = .11$), and no effects of the stress manipulation for initially unstressed words for either the right ($t[31] < 1$) or the left hemisphere ($t[31] < 1$).

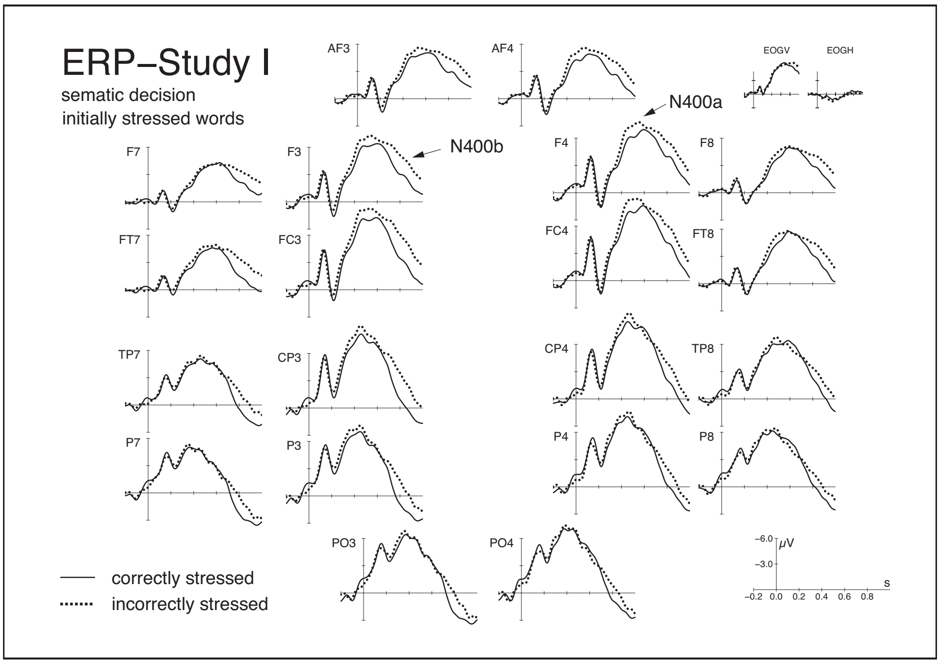

Figure 2.4: *ERPs for selected electrode sites for initially stressed words in the semantic decision. Solid lines: correct stress, dotted lines: incorrect stress.*

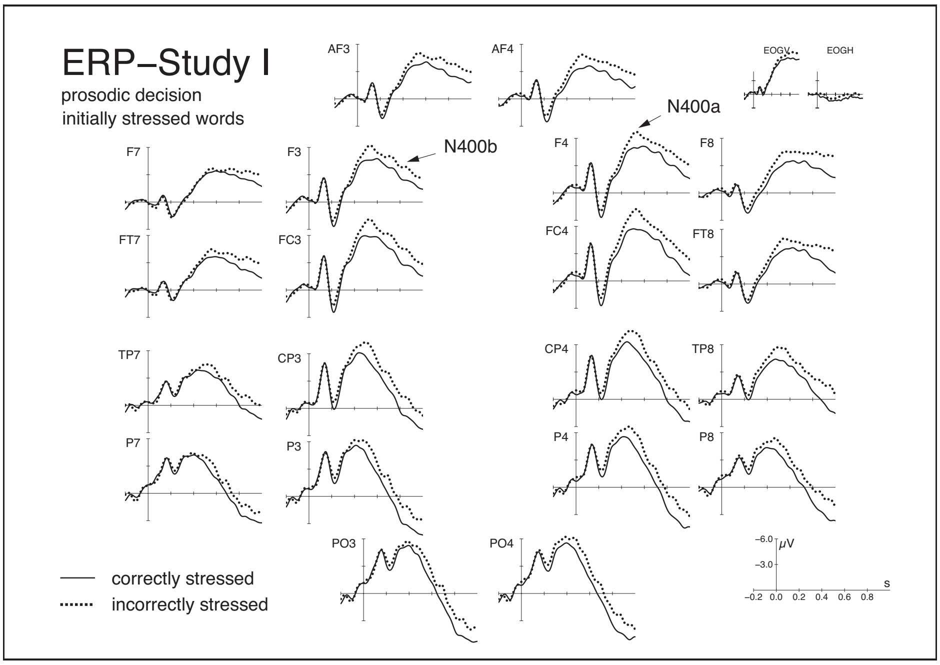

Figure 2.5: *ERPs for selected electrode sites for initially stressed words in the prosodic decision. Solid lines: correct stress, dotted lines: incorrect stress.*

		correct stress		incorrect stress			
		M	(SD)	M	(SD)	t[31]	p
semantic decision							
initially stressed	ANT	-4.4 μV	(3.9)	-5.0 μV	(4.3)	2.17	.15
	POS	-5.3 μV	(4.1)	-5.4 μV	(3.9)	0.01	
initially unstressed	ANT	-3.4 μV	(3.6)	-4.2 μV	(3.1)	6.17	.01
	POS	-3.7 μV	(3.5)	-4.6 μV	(3.4)	4.49	.04
prosodic decision							
initially stressed	ANT	-4.3 μV	(4.0)	-3.9 μV	(3.0)	0.77	
	POS	-5.0 μV	(4.9)	-4.8 μV	(4.2)	0.12	
initially unstressed	ANT	-3.7 μV	(3.7)	-4.2 μV	(3.7)	1.85	.18
	POS	-4.3 μV	(3.3)	-4.4 μV	(3.4)	0.02	

Table 2.4: *Mean amplitudes of ERPs elicited between 400 and 600 ms.Furthermore, t-values and their probabilities are given for each line. ANT = anterior scalp region; POS = posterior scalp region.*

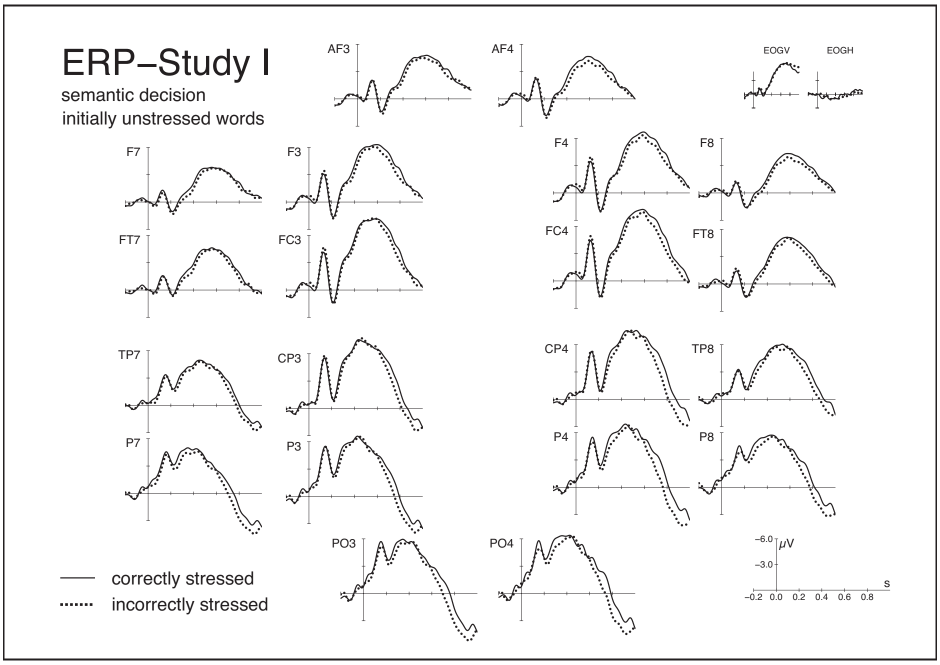

Figure 2.6: *ERPs for selected electrode sites for initially unstressed words in the semantic decision. Solid lines: correct stress, dotted lines: incorrect stress.*

Pairwise comparisons for the interaction of the stress manipulation with the factors original *stress*, *task* and *region* are summarized in Table 2.4. They indicate effects of the stress manipulation for initially stressed words (see Figure 2.4 and 2.5), but not for initially unstressed words (see Figure 2.6 and 2.7). However, even the effects for initially stressed words for the anterior and posterior scalp region were only significant for the prosodic judgment.
Taken together, in an early time window of the N400 there were only effects of the stress manipulation for initially stressed words. These effects were pronounced for the frontal-right scalp region and are stronger for the prosodic than for the semantic judgment.

600-1000 ms: N400b There is was main effect of the factor *stress manipulation* in the second time window of the N400 ($F[1,31] = 5.70$, $p = .02$). Furthermore, there were interactions of the factor *stress manipulation* with the factor *region* ($F[1,31] = 6.22$, $p = .02$), with the factors *original stress* and *hemisphere* ($F[1,31] = 4.54$, $p = .04$), and with the factors *original stress*, *task* and *hemisphere* ($F[1,31] = 10.78$, $p < .01$). The two-way interaction with region indicates that incorrectly stressed words elicited a more negative amplitude of the N400b (M = -3.5 μV, SD = 3.5) than the same words with correct stress (M = -2.8 μV, SD = 3.5, $t[31] = 17.24$, $p < .001$) for anterior electrodes. This effect was found regardless of the original stress of the word and the type of the judgment. There was no such effect for posterior electrodes ($t[31] = 1.36$, $p = .25$).

Results of t-tests that compare the ERPs elicited by correctly and incorrectly stressed words in the third time window split up for the factors *original stress of the words*, *task* and *hemisphere* are summarized in Table 2.5. As can be seen, there were ERP effects of the stress manipulation for initially stressed words, whereas, effects for initially unstressed words were not significant. Furthermore, as in the previous time window there were stronger effects for the prosodic judgment than for the semantic judgment.

To summarize, mean amplitudes in an early time window of the N400, named N400a, were only sensitive to the stress violation of initially stressed words, whereas ERP amplitudes in a later time window, named N400b, were effected by the stress violation of initially stressed as well as initially unstressed words. However, even the latter effects were more robust across both tasks for initially stressed as compared to initially unstressed words.

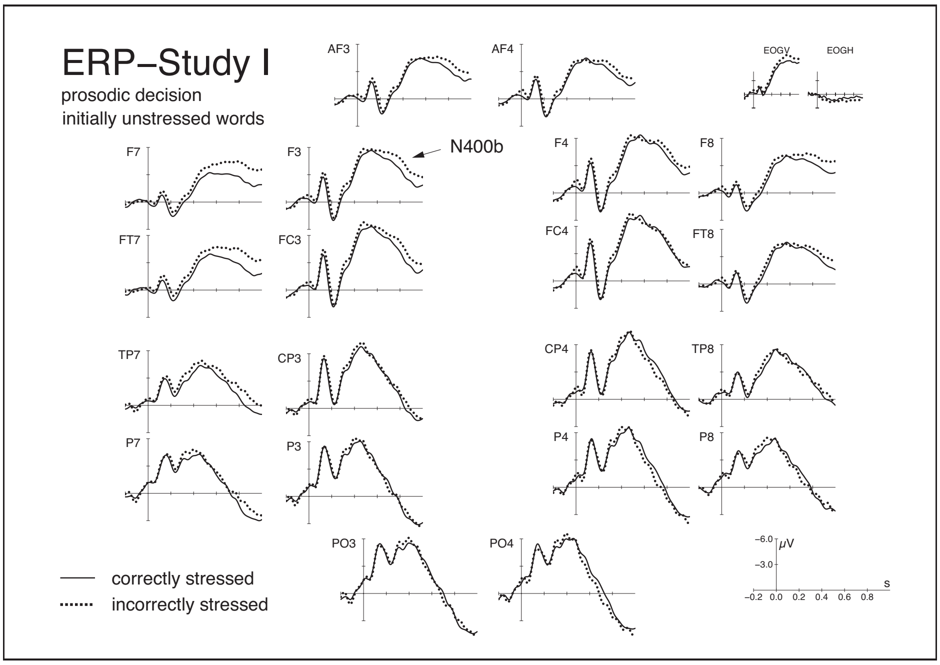

Figure 2.7: *ERPs for selected electrode sites for initially unstressed words in the prosodic decision. Solid lines: correct stress, dotted lines: incorrect stress.*

		correct stress		incorrect stress			
		M	(SD)	M	(SD)	t[31]	p
semantic decision							
initially stressed	LH	-1.1 μV	(4.7)	-2.2 μV	(4.5)	6.03	.02
	RH	-2.1 μV	(4.7)	-3.0 μV	(4.8)	3.85	.06
initially unstressed	LH	-0.7 μV	(4.2)	-1.7 μV	(3.8)	5.64	.02
	RH	-0.9 μV	(4.0)	-2.3 μV	(3.8)	12.86	.01
prosodic decision							
initially stressed	LH	-1.1 μV	(4.9)	-0.5 μV	(4.1)	1.60	.22
	RH	-1.8 μV	(4.7)	-1.1 μV	(3.9)	3.85	.10
initially unstressed	LH	-1.1 μV	(3.7)	-1.9 μV	(4.1)	3.67	.06
	RH	-1.6 μV	(3.3)	-1.8 μV	(4.2)	0.26	

Table 2.5: *Mean amplitudes (and standard deviation) of ERPs elicited between 600 and 1000 ms. Furthermore, t-values and their probabilities are given for each line. LH = left hemisphere; RH = right hemisphere.*

2.3.4 Discussion

The main research question of the present ERP experiment regarded the neural correlates of the processing of word stress. Firstly, ERP-study I was aimed to explore the time point of the extraction of prosodic information. If stress is used during spoken word access processes, different stress patterns should be differentiated early on using online measures. However, there was apparently no difference in the ERPs for initially stressed versus initially unstressed words pronounced with correct stress (see Figure 2.3). It is possible that *latency jitter* might smear ERP differences for initially stressed and initially unstressed words.

Latency jitter accounts to differences in the temporal availability of information between conditions. Availability of information can be altered by the duration of auditory material. Latency jitter can artificially decrease the amplitude of ERP-effects (e.g., Spencer, Vila Abad & Donchin, 2000). If there would be, for instance, an enhanced amplitude in the ERP related to the processing of initially stressed words, the enhanced duration and the greater variability of the first syllable of these words as compared to initially unstressed words (see Table 2.1) might prevent such an effect from being apparent in the ERP waveform. Furthermore, the onset of voiced parts that signal F_0 varies between the single words, and might smear ERP effects of the extraction of the pitch pattern of the first syllables. According to these considerations, the presentation of naturally spoken words is confounded by duration and might not be a powerful means to investigate neural processes engaged with the early discrimination of word stress.

The failure to show an ERP difference for initially stressed and initially unstressed words is in contrast to the results of Böcker, Baastiansen, Vroomen, Brunia and de Gelder (1999). They reported an enhanced negative ongoing component at 325 ms for weak-initial (unstressed) words as compared to strong-initial (stressed) words. This ERP deflection was interpreted as a correlate of prosodic processing. However, as already pointed out in the introduction of the present experiment (see 2.3.1), there are a number of drawbacks of that interpretation. On the basis of the present failure to find an equal ERP correlate with material that was held constant for vowel quality, the interpretation of Böcker et al. can be excluded. Accordingly, alternative views of their data should be considered.

As discussed in the introduction of the present chapter (see section 2.1), metrical stress is not only a prosodic phenomenon. Basically it is a function of the sequence of strong and weak syllables. Weak syllables bear different segments than strong syllables. Whereas vowels are always reduced in weak syllables, they are always full in strong syllables. With that respect, the N325 might be a correlate of processing different vowels or different vowel qualities. However, strong and weak syllables are not only established via different vowels, they also differ dramatically in duration. Thus, the N325 might be alternatively considered as an artifact of syllable duration. Weak syllables might evoke less latency jitter than strong syllables, because they are shorter and show less variation than strong syllables. As a consequence words with weak syllables on initial position would evoke a stronger peak in the ERP than words with strong syllables in initial position.

The previous considerations strengthens the crucial role of duration for the interpretation of ERP effects of spoken word recognition. ERP-study II regards this aspect. In that study, spoken words which bear full or reduced vowels were tested. The results of that study demonstrate, that it is probably the very short duration of the first syllable of weak-initial words, which accounts for the N325 effect. In terms of the present results it might be concluded that ERP correlates of processes that extract prosodic information might be smeared with naturally spoken material. Consequently, words, that were manipulated in a single prosodic parameter, namely pitch, were presented in ERP-study III. This manipulation leaves the duration of the words constant. In ERP-study III a difference in the amplitude of the P2 for words with different pitch contours was found. This result confirms the assumption that the ERP is sensitive to early prosodic processing. Furthermore, it can be taken as additional evidence that the N325, which shows a later onset and a reversed polarity is not likely to reflect prosodic processing.

The second major aim of the present experiment regarded the effects of incorrect stress. The behavioral results of the present study revealed only effects for the stress manipulation of initially stressed words. Across both tasks, subjects made more errors for initially stressed words with incorrect prosody than for the same words with correct prosody. However, the reaction times were only enhanced for the prosodic decision but not for the semantic decision. This might be attributable to the correction of the reaction time data. Initially unstressed words with incorrect stress show the longest duration (see Table 2.1). However, this was not attributable to the duration of the first syllables, which was comparable across both versions. The difference in duration between initially stressed words with correct and incorrect stress is mainly caused by the second syllables. Thus, the segmental information of the word onset might provide sufficient information for the semantic judgment. For the prosodic information, however, listeners might have taken the duration of the second syllable into account and thus, their responses were delayed. Taken together, correct responses and partially RTs for initially stressed words confirm the assumption that incorrect stress harms spoken word recognition. However, there were no behavioral effects related to the stress violation of initially unstressed words.

The ERPs extend the behavioral evidence regarding the use of stress during spoken word recognition. Effects of the stress manipulation were observable in the N400 component. Although there was an earlier and a more robust N400 effect in the ERPs for the stress violation of initially stressed words, there were also effects for initially unstressed words. Words with incorrect stress consistently elicited a more negative N400 than the same words with correct stress. This effect started at 400 ms for initially stressed words and at 600 ms for initially unstressed words. Incorrect versions of initially unstressed words, only elicited effects in the prosodic judgment, incorrect versions of initially unstressed words effected ERPs for both tasks. The ERP results indicate that the N400 reflects prosodic processing. However, in contrast to N400 effects related to the processing of semantic incongruity, which show a centro-parietal distribution (see Kutas, 1997), the N400 effect for a prosodic violation is largest over left-frontal scalp regions. This finding suggests different neural generators underlying prosodically and semantically related N400 effects. Furthermore, it confirms the assumptions that more than a single cognitive operation is reflected in the negative ERP wave, named N400, which occurs approximately 400 ms after stimulus onset.

Furthermore, the prosodically related N400 effects suggest that prosodic information might be used during different processing stages. An early N400 effect, referred to as N400a, can be differentiated from a later one, named N400b. The N400a, was only elicited by the stress violation of initially stressed words but it appeared consistently across both tasks. The N400b, in contrast, was elicited by a prosodic violation of initially stressed as well as initially unstressed words. However, for initially unstressed words it was restricted to the prosodic judgment. The N400a appears to reflect more 'general' or 'natural' processing of prosody during spoken word recognition, because it is elicited across both tasks. The N400b, in contrast, might be related to strategic effects induced by the prosodic judgment task.

A possible explanation of the N400a effect might be that it reflects enhanced effort for the selection of an appropriate candidate among activated entries. The initial acoustic signal of words with incorrect stress might activate wrong lexical or more lexical entries than the same words with correct stress. The greater number of competitors or the wrong activation of lexical entries might enhance the effort to select the appropriate word. That is, the N400a effect might be closely related to spoken word access processes. This interpretation fits well with the interpretation of early negative ERP effects in spoken sentence comprehension. An enhanced early negativity has been found for words that do semantically and phonologically not match the preceding sentence context as compared to words that fit the initially phonemes of semantically appropriate words (see section 1.3.2.2, Hagoort & Brown, 2000; van den Brink, Brown & Hagoort, 2001). It has bee argued that this effect, referred to as N200, is related to the selection of the appropriate word among activated candidates. The later N400 effect for all semantically inappropriate words is interpreted as a correlate of lexical integration (Hagoort & Brown, 2000; van den Brink, Brown & Hagoort, 2001). As the N400a in the present experiment, the N200 differs in distribution to the classical N400 found for semantic violations or for nonwords. The N400 is distributed across the posterior scalp region, whereas the N200 has a flat distribution across the scalp and is most clearly visible over frontal sites (van den Brink, Brown & Hagoort, 2001). The N400a effect is also pronounced over frontal sites. Thus, the present finding strengthens the assumption, that the N400 for auditory speech material is a conglomerate of different electrophysiological deflections related to different processes of language comprehension.

The question remains, whether the N400a is only modulated by incorrectly stressed words which were originally stressed on their first syllables. This result parallels the behavioral effects of the stress manipulation in the present experiment, which were also only observed for initially stressed words. There are two possible explanations for the heterogeneous effects for both types of words. First, they might reflect that lexical access and/or post-lexical processing is differently realized for initially stressed and initially unstressed words (see the following Chapter for a related theoretical approach). Similarly, however, the different results for initially stressed and initially unstressed words might be caused by the different realization of the stress violation. As discussed in section 2.2.2, the speaker used duration, amplitude and pitch to establish incorrect stress for initially stressed words. For the initially unstressed words, in contrast, the speaker only varied duration and amplitude, but not pitch. Thus, the absent N400a effect for initially unstressed words might be caused by the fact that fewer prosodic parameters were incorrectly applied in this condition as compared to the stress violation of initially stressed words. According to this interpretation lexical access should only be harmed if *enough* prosodic information is incorrect. Alternatively, however, incorrect pitch might be the most important prosodic parameter to elicit the N400a and the behavioral effect for incorrect stress.

ERP-study III (see section 4.2) concerns the clarification of the role of pitch. In that experiments words which were only modulated in their pitch contours were presented. The reaction times were delayed for initially stressed words with initially unstressed pitch contour as well as for initially unstressed words with initially stressed pitch contour. In ERP-study III the delay in response times was larger for a stress violation of the initially unstressed words as compared to the initially stressed words. Thus, there is conclusive hint from the present study and ERP-study III, that initially stressed and initially unstressed words are differently processed on the lexical level. However, these results do also not confirm, that pitch is the critical prosodic parameter which evokes an N400a effect when it is violated. It appears that a complex interaction of all prosodic parameters as realized in the present experiment elicits this ERP deflection. A complex interaction of all prosodic parameters is also suggested by the results of ERP-study IV. ERPs obtained for targets in cross-modal word fragment priming were earlier modulated by the primes pitch than by their remaining prosodic parameters. Furthermore, the ERPs and the reaction times revealed that pitch is, in contrast to the remaining prosody, used independently of the segments of the primes. These results suggest a use of pitch for the activation of lexical entries. The remaining prosody might, in contrast be used to constrain the initial activation in competition processes. Thus only the activation pattern that results from a modified initial activation as well as from modified competition processes might harm the selection of the final candidate from the activated that, the process that is probably reflected in the enlarged N400a in the present experiment.

The N400a effect in the ERP, which may well be related to lexical processing, is followed by an N400b effect. The N400b differed in it's time window and in it's sensitivity to the experimental manipulation from the N400a. The N400a may reflect the stress manipulation for initially stressed words only, whereas the N400b resulted from the stress violation of both, initially stressed and initially unstressed words. Crucially, the gating data for the words presented in ERP-study I are in line with the N400b effects. That is, in the gating task the stress manipulation affected subjects responses to incorrectly stressed words regardless of their original stress pattern. This might indicate that there are processes active during spoken word recognition, as reflected in the N400a effect, which do not affect subjects' responses in a gating task. Moreover, these processes might operate earlier and might be of more general importance during spoken word recognition than the effects reflected in the subjects' responses.

Taken together, the gating experiment and the present ERP-study clearly reveal effects of the stress manipulation in German. However, the ERPs point to a more complex picture of results. Different effects for initially stressed and initially unstressed words might be caused by the different prosodic realization for both types of words. The following experiments rely on the origin of these differences. In Chapter 3 the relation between syllable duration and ERPs observed for spoken words is investigated. Chapters 4 and 5 selectively focus on the function of pitch during spoken word recognition.

Chapter 3

Full versus reduced vowels

3.1 Introduction

The present chapter mainly concerns an methodical issue of ERP research with spoken language material, namely a confounding of the on-line signal and duration. A second aspect on which the present chapter focuses is the role of metrical stress, an issue which was already discussed in the previous chapter. Metrical stress is considered to be an important aspect of spoken language as it appears to be a marker for speech segmentation.

In connected speech a listener has to identify a single word within the acoustic stream. This acoustic stream provides - unlike written language - only few cues for the boundaries of words. Such cues are for instance the lengthening of word onset and word final syllables in connected speech (Gow & Gordon, 1995). Another possibility the listener has to extract single words, is the exploitation of probabilistic aspects of the utterance. That is word boundaries are located where they are most likely. There are, for instance, statistical regularities of lexical items such as distributional regularity and phonotactics (see Brent & Cartwright, 1996).

In some language stress position is a further statistic probability for speech segmentation. As already pointed out (see section 2.2.4), stress more often falls on the first than on any other syllable of a word in German, English and Dutch. Moreover, weak syllables are extremely unlikely to form a word beginning (Cutler

& Carter, 1987). Thus, at least weak syllables are assumed to guide segmentation. The metrical segmentation strategy (MSS Cutler & Norris, 1988) proposes that listeners expect a word boundary preceding a strong syllable, but not preceding a weak syllable. Research in English (Cutler & Norris, 1988; Cutler & Butterfield, 1992) and Dutch (Vroomen, Zon & de Gelder, 1996; Quene & Koster, 1998) has confirmed this idea. The original demonstration of the MSS by Cutler and Norris (1988) showed that a monosyllabic English word, for instance *mint*, is more difficult to recognize when it is embedded in a string with two strong syllables such as /mɪntef/ as opposed to a string with a strong and a weak syllables such as /mɪntəf/. The explanation of this effect is that a segmentation point falls in-between both strong syllables in /mɪntef/, whereas no segmentation is induced by /mɪntəf/.

An attempt to integrate the MSS into models of spoken word recognition was provided by Mattys and colleagues (Mattys, 1997; Mattys & Samuel, 1997; Mattys, 2000; Mattys & Samuel, 2000). Following this approach, the listener is assumed to wait until a strong syllable is present in the speech stream to initiate lexical access. For weak-initial words a *backtracking* mechanism is proposed. That is, each initially unstressed word should be delayed in its lexical access until a strong syllable occurs in the speech signal. According to Mattys and colleagues, backtracking is induced by prosodic parameters. Thus, they generalized this effect to initially stressed and initially unstressed words. Indeed, they showed, that initially stressed words are responded to slightly faster than initially unstressed words (Mattys & Samuel, 2000). However, the results of Zwitserlood (1989) show that it might not be the case that lexical access waits until a stressed syllable is present in the speech stream. Zwitserlood reports that in neutral sentence contexts fragments such as *kapi* taken from the initially unstressed Dutch word *kapiTAIN* (Engl. captain) or *kapiTAAL* (Engl.: capital) reliably activate lexical entries of their target words (e.g., ship and money). This contradicts the assumption that lexical access for both words only takes place if the stressed syllable *TAAL* is heard. As described in section 1.2 it appears to be rather the case that prosody restricts the activation to matching entries, but not that it prevents lexical activation.

Recently, effects of metrical segmentation have been linked to competition processes of spoken word recognition (Norris, McQueen, Cutler & Butterfield, 1997). Following this account, MSS does not exert an influence on lexical activation. It is rather used to modify the activation status of activated entries. Activation is reduced for entries that do not lead to an optimal segmentation of the utterance. Metrical stress is assumed to be taken into account in the calculation of the optimal segmentation of the input. That is, a weak syllable is assumed to activate lexical entries which begin with that syllable. Thereafter however, the activation of that entry might be reduced if the weak syllable can be attached to a strong syllable that precedes it. According to this proposal, weak initial words are penalized in the competition phase.

As already mentioned in section 2.1, a fundamental difference between weak and strong syllables is vowel quality. For instance, /ə/ in /mɪntəf/ differs on a phonetic domain, namely on the closed-unclosed continuum, from /e/ in /mɪntef/. As Fear, Cutler and Butterfield (1995) showed for English, the strong versus weak syllable distinction in English is mainly based on vowel quality and not on the prosodic information of the syllables.

In light of the previous considerations, the results of the ERP study by Böcker et al. (1999) should be discussed again. As already mentioned did these authors present strong- and weak-initial words. Therefore, the observed ERP effects are critical to interpret in two aspects. Firstly, weak and strong initial words fundamentally differ in their duration of the first syllable. Reduced vowels are extremely short as compared to full syllables. As the ERP continuously records ongoing brain activity, the signal might be extremely sensitive to timing differences of stimuli. There might be an ERP deflection in the time window of the 'N325', which is unrelated to the processing of prosody. For instance, the ERP might be sensitive in that time window to syllable boundaries. Note, that there is evidence, that the detection of phrase boundaries is reflected in the ERP (Steinhauer, Alter & Friederici, 1999). Such an ERP correlate for syllable boundaries would be stronger affected by latency jitter for strong-initial words than for weak-initial words, because the strong-initial words have longer first syllables and a greater variability in duration of the first syllables than weak-initial words. Thus, it is not clear whether the ’N325’ is an artifact of syllable length or whether it is a correlate of the processing of metrical stress.

Secondly, as already pointed out, strong and weak syllables do not only differ in terms of their prosodic information, they also vary in their segmental information. Especially in the design of Böcker et al. (1999) the ’N325’ might rather reflect segmental processing. They presented weak initial words which always bear /ə/ in the first syllable. In contrast, the strong initial words they presented are characterized by different full vowels (/ɛ/, /e/, /a/ or /o/). Thus, words with different vowels were contrasted to words with a similar vowel. That is, an alternative explanation of the ’N325’ would be that it is related to segmental processing rather than to the differentiation of stress patterns by the brain.

Finally, the interpretation of the ’N325’ as an electrophysiological marker of prosodic differentiation is not only critical in terms of theoretical aspects, it also conflicts with the results ERP-study I which did not reveal an ’N325’ deflection in the ERP related to the processing of initially stressed and initially unstressed words (see section 2.3). Therefore, it appears, that one of the two previously introduced alternative hypotheses might be a better account for this ERP deflection than the interpretation of Böcker et al. (1999). ERP-study II aimed at clarifying whether the ’N325’ is caused by syllable length or by segmental characteristics of the first syllable. If the ’N325’ reflects the processing of vowel quality, it should vary for groups of words with different vowels in the first syllable. If the ’N325’ is somehow related to syllable length, the mean ERP amplitude in the ’N325’ time window should correlate with this parameter.

3.2 ERP-study II

In the present study bisyllabic German words were selected in a way that half of them were stressed on the first syllable and half were stressed on the second syllable. Within each stress group two vowel qualities were tested: Half of the initially stressed words contained /a/, the other half /ɛ/ in the first syllable. Half of the initially unstressed words contained /a/, the other half contained the reduced vowel /ə/ in the first syllable. If the 'N325' is correlated with different vowels in the first syllable, it would be expected to vary across the different vowel conditions, that is for words with /a/, /ɛ/ and /e/ in the first syllable. If there is a correlation between the duration of the first syllable and the 'N325' amplitude, this would be a clear hint that the 'N325' is an artifactual consequence of the stimulus material.

3.2.1 Method

Subjects Twenty-four right-handed subjects (12 women) aged 19-30 years (M = 23.3) participated in the experiment. All were native German speakers with no reported hearing or neurological deficits.

Material 168 bisyllabic German words were selected (see Appendix B). All words started with a consonant or a consonant cluster. The first syllable was stressed for half of the words, the remaining 84 words were initially unstressed. 42 initially-stressed words contained the full vowel /a/ in the first syllable, 42 initially-stressed words contained the full vowel /ɛ/ in the first syllable. 42 initially-unstressed words contained the full vowel /a/ in the first syllable, 42 initially-unstressed words contained the reduced vowel /ə/ in the first syllable. The four subgroups of words were controlled for word frequency according to the CELEX lexical database for German (Baayen, Piepenbrock & van Rijn, 1995). The words, were spoken by a female native speaker of German, recorded onto Digital Audio tape, and digitized at 44.1 KHz. Each word was extracted with a sound editor (CoolEdit v. 1.52, ©Syntrillium Software Corp.). Prosodic features of the experimental words are summarized in Table 3.1.

Twenty-eight bisyllabic German words (14 initially stressed) different from the experimental words were spoken with incorrect stress (see Appendix B). Fourteen of the words were initially stressed in their original versions and contained an /a/ in the first syllable, 14 words were initially unstressed in their original versions and contained an /ɛ/ in the first syllable. Initially stressed words (e.g., *BAsis*

	first syllable		second syllable	
	M	(SD)	M	(SD)
initially stressed /a/				
duration (ms)	283	(74)	373	(75)
F_0 (Hz)	205	(25)	179	(34)
amplitude (dB)	37.4	(1.5)	35.0	(2.1)
initially stressed /ɛ/				
duration (ms)	293	(90)	337	(107)
F_0 (Hz)	206	(38)	190	(73)
amplitude (dB)	37.0	(1.3)	34.3	(2.5)
initially unstressed /a/				
duration (ms)	209	(56)	447	(107)
F_0 (Hz)	187	(24)	206	(65)
amplitude (dB)	35.6	(1.3)	37.0	(1.4)
initially unstressed /ə/				
duration (ms)	96	(38)	533	(137)
F_0 (Hz)	190	(10)	216	(84)
amplitude (dB)	35.7	(2.0)	37.3	(1.1)

Table 3.1: *Prosodic realization of the first and the second syllable for the words presented in ERP study II. Initially stressed words with /a/ and /ɛ/ in the first syllable are presented in left columns, and initially unstressed words with /a/ or /ə/ in the first syllable are presented in right columns.*

[Engl.: base]) were presented with stress on the second syllable (e.g., baSIS). Vice versa, initially unstressed word (e.g., *terMIN* [Engl.: date]) were presented with stress on the first syllable (e.g., *TERmin*. Words with incorrect stress were spoken by the same female speaker who also spoke the correctly stressed words.

Procedure Subjects were seated in an electrically and acoustically shielded chamber in front of a computer screen. Loudspeakers were placed to the left and to the right side of the screen. An experimental trial began with the presentation of a fixation cross in the center of the screen. Participants were instructed to fixate this cross at any time it appeared on the screen. After 300 ms a word was presented via loudspeaker while the fixation cross remained on the screen. As in the study of Böcker et al. (1999) the task of the subjects focused on prosodic processing. Subjects had to respond only when they thought that the word was incorrectly stressed. No responses were required for correctly stressed words. Half of the subjects made responses with the thumb of their left hand, half of the subjects made responses with the thumb of their right hand. Reaction times were measured from stimulus onset. After a variable inter-stimulus interval (1000-2000 ms) the next trial began.

Electrophysiological Recordings. See section 2.3.2.

Statistical analyses. The mean hit rate and the corresponding reaction times were analyzed to describe the performance of the subjects. Incorrectly stressed words were not further analyzed. This is based on the low number of these words. Correct rejections, that is correct non-responses, were analyzed with a repeated-measures analysis of variance (ANOVA). Factors were *stress* (initially stressed vs. initially unstressed) and *vowel* (/a/ vs. /ɛ/ for initially stressed words and /a/ vs. /ə/ for initially unstressed words).

Only correctly rejected experimental words were averaged for the calculation of the mean ERPs. Visual inspection of the ERPs revealed an 'N325' for initially unstressed words with /ə/ in the first syllable at anterior electrode sites. The time window at which this component occurred (250 to 370 ms) was explored statistically. As in ERP-study I an N400a and an N400b time window was analyzed. Four factors were analyzed per time window in a repeated-measures ANOVA. As for the behavioral data the factors *stress* and *vowel* were analyzed. Two further factors served to explore the spatial distribution of the 'N325'. Mean amplitudes of 36 electrode positions were included. The factor hemisphere differentiates left and right electrode positions, the factor region anterior and posterior electrode positions. Both factors were based on the following regions of interest: anterior left (AF7, AF3, F7, F5, F3, FT7, FC5, FC3, C3), anterior right (AF4, AF8, F4, F6, F8, FC4, FC6, FT8, C4), posterior left (TP7, CP5, CP3, P7, P5, P3, PO7, PO3, O1) and posterior right (CP4, CP6, TP8, P4, P6, P8, PO4, PO8, O2; see Figure 2.2 for electrode locations). For the correlation analysis the mean amplitude within the time window of the 'N325' was computed for each word and correlated with the duration of the first syllable of that word.

3.2.2 Results

3.2.2.1 Behavioral data

On average 24.1 (SD = 2.7) of the 28 incorrectly stressed words were recognized by the subjects. The mean reaction time for hits was 1045 ms (SD = 54). The analysis of the correct rejections revealed a significant main effect of the factor *stress* ($F[1,23] = 16.83$, $p < .001$), of the factor *vowel* ($F[1,23] = 13.96$, $p = .001$), and a significant interaction of both factors ($F[1,23] = 6.73$, $p = .02$). Single comparisons revealed the lowest number of correct rejections for initially unstressed words with /a/ in the first syllable (M = 40.2, SD = 1.6). The number of correct rejections for initially stressed words with /a/ in the first syllable (M = 41.5, SD = 0.7), for initially stressed words with /ɛ/ in the first syllable (M = 41.8, SD = 0.5) and for initially unstressed words with /ə/ in the first syllable (M = 41.5, SD = .09) were significantly higher than that for initially unstressed words with /a/ in the first syllable ($t[23] = 13.75$, $p = .001$, $t[23] = 19.89$, $p < .001$, $t[23] = 11.6$, $p < .01$). No other single comparison was significant ($p < .10$). In other words, subjects categorized initially unstressed words with /a/ in the first syllable more frequently as incorrectly stressed than words of the other three correctly stressed conditions.

3.2.2.2 ERPs

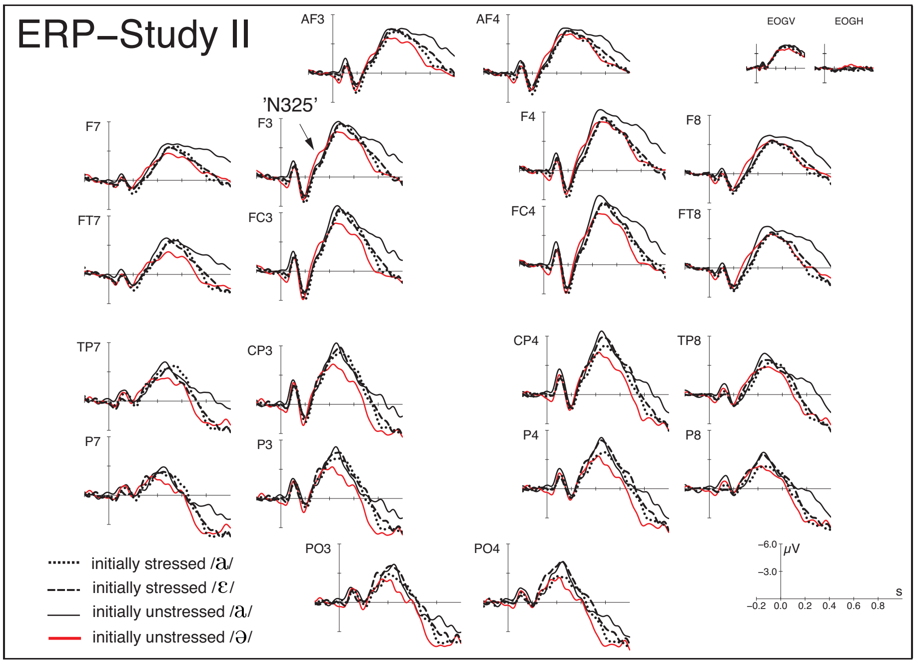

Figure 3.1: *ERPs for selected electrode sites. Dashed black lines indicate ERPs for initially stressed words with /a/ in the first syllable, dotted black lines indicate ERPs for initially stressed words with /ɛ/ in the first syllable, solid black lines indicate ERPs for initially unstressed words with /a/ in the first syllable, solid red lines indicate ERPs for initially unstressed words with /ə/ in the first syllable.*

250 to 370 ms: 'N325' The analysis of the ERPs for the correctly stressed words revealed a significant interaction of the factors *stress* and *region* (F[1,23] = 10.22, p = .004) and of the factors *stress*, *vowel* and *region* (F[1,23] = 4.35, p = .048). Post-hoc comparisons for the anterior region revealed a significantly more negative mean ERP amplitude elicited by initially unstressed words with /ə/ in the first syllable (M = -1.55 μV, SD = 2.4) than for initially stressed words with either an /a/ (M = -0.56 μV, SD 1.9, t[23] = 6.56 p = .017) or an /ɛ/ in the first syllable (M = -0.76 μV, SD = 2.4, t[23] = 5.17 p = .033, see Figure 3.1 for an illustration of the

ERPs). There was no significant difference between the 'N325' amplitude elicited by initially unstressed words with /a/ (M = -1.08 μV, SD = 2.1) or with /ə/ in the first syllable (t[23] = 1.51 $p > .10$), and for initially stressed words with /a/ or /ɛ/ in the first syllable ($t > 1$). In contrast to initially unstressed words with /ə/, words with /a/ did not differ from the initially stressed words with either /a/ (t[23] = 2.75, $p > .10$), or with /ɛ/ in the first syllable ($t > 1$). The latter result might be related to the intermediate duration of the first syllables of initially unstressed words with /a/. They were even longer than the first syllables of initially unstressed with /ə/, but slightly shorter than the first syllables of initially stressed words (see Table 3.1). This was confirmed by a significant correlation of the mean N325 amplitude over the anterior scalp region and the duration of the first syllable of the words (Pearsons Correlation Coefficient = 0.24, p = .002). That is, words with shorter syllables elicited more negative ongoing waveforms in the time window of the 'N325' than words with longer syllables and vice versa.

400 to 600 ms: N400a There was a significant interaction of the factors *vowel* and *stress* (F[1,23] = 8.85, $p < .01$). Post-hoc comparisons revealed, that the mean ERP-amplitudes elicited in this time window were more negative for initially unstressed words with /a/ (M = -3.84 μV, SD = 2.55) than for initially unstressed words with /ə/ (M = -2.79 μV, SD = 2.76, t[23] = 5.34, p = .03). There was no difference for initially stressed words with either /a/ (M = -3.16 μV, SD = 2.28) or /e/ (M = -3.52 μV, SD = 2.77) in the first syllable ($t < 1$). However, mean amplitudes of the N400a elicited by words with /ə/ in the first syllable did not differ significantly from initially stressed words ($t < 1$). That for initially unstressed words with /a/ differed from initially stressed words with /a/ (t[23] = 5.22, p = .03), but not from that with /e/ ($t > 1$). Taken together, it appears that words with /ə/ elicit a small N400a amplitude and initially unstressed words with /a/ elicit a high N400a amplitude (see Figure 3.1).

600 to 1000 ms: N400b There was again a significant interaction of the factors *vowel* and *stress* ($F[1,23] = 14.12$, $p = .001$). Again, post-hoc comparisons revealed, that the mean ERP-amplitudes elicited in this time window were more negative for initially unstressed words with /a/ (M = -1.60 μV, SD = 4.01) than for initially unstressed words with /ə/ (M = 0.36 μV, SD = 3.52, t[23] = 12.29, p = .002). There was no difference for initially stressed words with either /a/ (M = -0.27 μV, SD = 3.44) or /e/ (M = -0.53 μV, SD = 3.21) in the first syllable ($t < 1$). Mean amplitudes of the N400a elicited by initially unstressed words with /a/ in the first syllable differed significantly from initially stressed words with /a/ (t[1,23] = 9.36, p = .01) but not significantly from initially stressed words with /e/ (t[1,23] = 2.67, $p > .10$). There were no other significant differences between conditions (all $p > .10$). Taken together, initially stressed words with /a/ in the first syllable elicit an N400b with high amplitude.

3.2.3 Discussion

The results of the present experiment as well as of the results of ERP-study I (see section 2.3) do not confirm that the 'N325' reflects prosodic processing. There is no 'N325' effect for initially stressed versus initial unstressed words that contain full vowels in the first syllable. The present results strongly point to an interpretation of this ERP deflection as an artifactual consequence of syllable length. The assumption that the 'N325' might be caused by the shorter duration of syllables with reduced vowels was confirmed by a significant correlation between the length of the first syllable of a word and the 'N325' amplitude.

The present data clearly rule out this possibility that the 'N325' is related to vowel processing. There are no differences in the 'N325' amplitude for conditions that contain the different full vowels /e/ and /a/. Furthermore, it is not likely that the 'N325' reflects the differentiation of full and reduced vowels. ERP amplitudes in the time window of the 'N325' do not differ for initially unstressed words with the full vowel /a/ in the first syllable from words with a reduced vowel in the first syllable. A unifying explanation of the results is based on the duration of the first syllable. The shorter the first syllable of word is, the higher is the amplitude elicited in the time window of the 'N325'. It appears that the first syllable or the syllable boundary triggers linguistic processing. The ERP correlate of such processing is stronger time-locked for words with short first syllables than for words with long first syllables. Thus, latency jitter acts weaker on words with short first syllables, leading to an enhanced ERP amplitude for such words.

One might argue that the present effects do not finally rule out the possibility that the 'N325' is associated with the processing of prosody. If weak syllables are one end of a stressed-unstressed continuum and stressed strong syllables are the other end, unstressed strong syllables might range in-between both ends. Thus, the intermediate amplitude of the 'N325' for strong initial words with an unstressed first syllable might be caused by their intermediate prosodic status. However, these prosodic effects would be very small, as there was neither in the present study nor in ERP-study I a significant difference for initially stressed and initially unstressed words with full vowels. Thus, the present results should stimulate a critical view of ERP effects in spoken word recognition.

Artifacts caused by duration differences are hardly avoidable if effects of prosody are investigated with naturally spoken materials. A perfectly time-locked approach appears to be necessary for a meaningful interpretation of effects of subtle acoustic differences for spoken words. The experiments presented in the following chapters of this thesis account for this necessity. In these experiments pitch manipulated words were presented. This manipulation leaves temporal characteristics of the stimuli constant. The main aim of this manipulation was to test the effects of pitch on spoken word recognition. This attempt is based on the results of ERP-study I. They suggest an important role of pitch during spoken word recognition. Moreover, the pitch manipulation allows to directly compare ERPs between different experimental groups without potential artifacts of syllable. This appears a necessary precondition for a meaningful interpretation of the ERPs.

The results for the N400a do not confirm the assumption of *backtracking* for weak initial words (Mattys, 1997; Mattys & Samuel, 1997; Mattys, 2000; Mattys & Samuel, 2000, see section 3.1). In contrast, the reduced N400a amplitude (see Figure 3.1) might suggest weaker processing effort for these words as compared to words with full vowels in the first syllable. This effect might result from the small group of weak initial words in German. The limited number of weak-initial lexical entries results in the activation of fewer competitors. In contrast, the strong initial words activate more competitors. Even the initially unstressed words with the full vowel /a/ activate the large group of initially stressed words with that vowel. Thus, the selection of the appropriate word might be facilitated for weak-initial words as compared the selection for for strong-initial words. This interpretation of the N400a parallels that of the N400a in ERP-study I (see section 2.3.4). Both relate this ERP deflection to the activation pattern available to the selection process.

One might argue, that due to the metrical segmentation strategy (MSS) weak-initial words should be penalized in the competition process. However, there is no reason to assume, that the competition is modulated by MSS for words presented in isolation. There is a clear word boundary in this kind of presentation, namely an interval of silence before the onset of the word. The silence cue makes it unnecessary to use metrical hints for word boundaries. Taken together, the present data suggest that the initial activation is sensitive to vowel reduction. If a word begins with a reduced syllable, only weak-initial words are activated in the lexicon. In contrast, the results can not be taken to argue, that lexical activation waits until a stressed syllable is in the speech stream. This advantage for weak-initial words is not modulated in the competition process. Before their ending 620 ms after word onset, the reduced N400a indicates that selection for weak-initial words is facilitated.

The results for the N400b indicate differences between initially unstressed words with the full vowel (/a/) in the first syllable and the remaining three conditions. Subjects also produced more errors for words with the full vowel (/a/) in an unstressed first syllable. The behavioral result suggests that it was more difficult for the subjects to decide whether a word is correctly stressed or not, if it is initially unstressed but contains a full vowel. This interpretation confirms the assumption of a stressed-unstressed continuum in which initially unstressed words with a strong first syllable are more ambiguous than weak initial words or initially stressed words with a full vowel in the first syllable. The question remains why there was no difference between initially stressed and initially unstressed words in ERP-study I. It is possible, that subjects had different response criteria in both experiments. Half of the words were incorrectly stressed in the former experiment, whereas there was only one out of six words incorrectly stressed in the present experiment. Furthermore, subjects were required to respond after each presentation in ERP-study I. In contrast subjects should only respond to incorrectly stressed words in the present experiment. Additionally, the presentation rate was faster in the present than in the former experiment. These factors might have caused the disadvantage for initially unstressed words with full vowels in relation to all other words of the present experiment. It is interesting to note, that this discrimination disadvantage in the prosodic decision is paralleled by an enlarged N400a amplitude and an N400b amplitude for initially unstressed words with full vowels as compared to the other words. These results support the interpretation of ERP-study I in that N400 effects can be related to prosodic processing at the word level.

Chapter 4

Pitch manipulated words

4.1 Introduction

The present and the following chapter concern the processing of pitch manipulated speech material. Pitch manipulated words are in focus of the present chapter. In Chapter 5 experiments with pitch manipulated word fragments are discussed. The parameter pitch was manipulated for two reasons. First, the results of ERP-study I suggest that pitch plays an important role during spoken word recognition. Only the stress manipulation of initially stressed words, but not that of initially unstressed words elicit behavioral and electrophysiological responses. These differences might be caused by the different realization of the stress manipulation for both types of stress in the material used in ERP-study I. Incorrect versions of initially stressed words used in that study show a deviant pitch contour in comparison to their correct counterparts, whereas incorrect versions of initially unstressed words carry a correct pitch contour. Thus, pitch might be an important factor to establish a stress violation.

A second important ratio for the pitch manipulation is that it leaves temporal characteristics of spoken material constant. As the results of ERP-studies I and II reveal, identical time structures of spoken words appear to be a necessary precondition for a meaningful interpretation of ERP effects. If duration is not controlled for, latency jitter effects might either hinder effects from being obvious in the ERP (see results of ERP-study I and II) or evoke systematic latency differences across experimental groups (see results of ERP-study II). Due to the pitch modulation timing characteristics and amplitude envelope of the words constant were kept constant.

The processing of pitch as a single suprasegmental parameter has been addressed in behavioral studies with naturally spoken material. In these studies meaningless consonant-vowel strings, for instance /ba/ and /da/, were presented. English speaking subjects were instructed to judge consonant identity. That is, in the above example a decision whether the string contained a */b/* or a */d/* would be required. It was argued that if pitch is implicitly processed during spoken word recognition, responses for strings varying in pitch contour should be delayed as compared to strings carrying an identical pitch contour (Lee & Nusbaum, 1993; Miller, 1978; Repp & Lin, 1990). The proposed interaction of pitch and phoneme processing was indeed found for strings that vary in pitch contour (e.g. rising versus falling across the string), but not if pitch is lowered or heightened across the entire string (Lee & Nusbaum, 1993; Wood, 1974). Thus, although English does not use tone information to distinguish word meaning, English listeners are influenced in their processing of spoken words by pitch contours. These results have been interpreted to reflect that pitch contour is automatically processed to guide segmentation and recognition of fluent speech and to establish affective information in spoken material. However, no attempt has been undertaken to integrate pitch information into models of spoken word recognition.

The present study aimed at clarifying the role of pitch during spoken word recognition by recording behavioral and electrophysiological responses to pitch manipulated words. ERPs might allow to investigate when the brain distinguishes different pitch contours. The early categorization of pitch information as indicating a stressed or an unstressed syllable appears to be an important presumption for the early use of pitch during spoken word recognition. If the beginning of a word can be determined as stressed or unstressed on the basis of pitch information, this might modulate the recognition of that word. That is, if pitch information is incorrect for an actual word, recognition might be delayed. Consequently, RTs for words with incorrect pitch might be delayed. Furthermore, enhanced processing effort during lexical access, might be paralleled by a modulated ERP component as well. ERP-study I and ERP-study II revealed that incorrect stress information is correlated with enhanced N400a and N400b amplitudes. The N400a might be associated with the activation of competitors during lexical access which might cause enhanced selection effort (see section 2.3.4 for a discussion of this aspect). The N400b, in contrast, appears to be related to later strategic processing of incorrect prosodic information of a word. Both ERP deflections might be sensitive to the pitch manipulation.

4.2 ERP-study III

4.2.1 Method

Subjects Twenty-four right-handed students with normal hearing (12 female, age 18-30 years, M = 24.5 years) participated in the study. All were native speakers of German, who had not participated in ERP-study I.

Stimuli and Procedure 80 bisyllabic German words were selected from the material presented in the gating task and in ERP-study I (see Appendix A). All the words presented in the previous experiment were manipulated in their pitch contour. A speech editor (PRAAT version 3.2, ©Paul Boersma & David Weenink) was used to extract the F_0 contour of one strong-initial word ('AMboss' [anvil]) and one weak-initial word ('abTEI' [abbey]). On the basis of both pitch contours two versions of each word were resynthesized with an automatized procedure provided by PRAAT. Periodical parts of the signal were adjusted to realize one version of each word with a strong-initial, and another version of the same word with a weak-initial pitch contour (see Figure 4.1).

The words for the present experiment were selected from the corpus of all manipulated words according to the judgments of 20 subjects (different from those participating in the present study) who decided whether the words sounded nat-

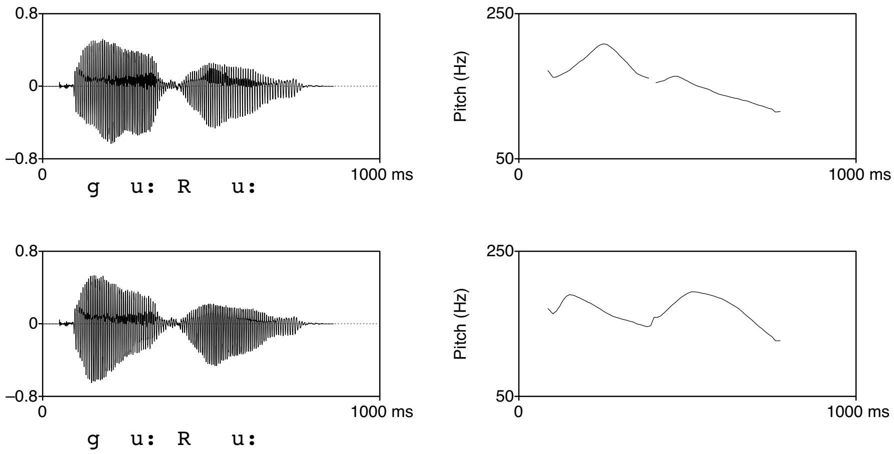

Figure 4.1: *Examples of the pitch manipulation for the initially stressed German word Guru' (Engl.: guru). The upper waveform illustrates the version with initially stressed pitch contour. This contour matches the expected stress pattern of Guru. The version of Guru with initially unstressed pitch, which does not match the expected stress pattern, is illustrated below.*

ural or not. Furthermore, words were selected so that half of them were initially stressed, half initially unstressed, and that half of each group belonged to the category animate, the other half to the category inanimate. In parallel to the tasks applied in ERP-study I, subjects were engaged in a semantic and in a prosodic judgment task. In the former task they decided whether the presented words were either part of the category living or non-living. In the latter task they evaluated the stress pattern (correct vs. incorrect) of the presented word. All subjects performed both tasks, half of them began with the semantic judgment, the other half with the prosodic judgment. Subjects heard all words in both versions in each task. Half of the subjects heard 20 initially stressed and 20 initially unstressed words with correct pitch and the other words with incorrect pitch first. The other half heard the alternative version of the former words first. The presentation order of correctly and incorrectly stressed words was randomized within subjects.

Stimulus presentation in both tasks was identical. Words were presented auditory via loudspeakers. Different than in ERP-study I, presentation rate was enhanced. This modification was chosen, as it became apparent from ERP-studies I and II that subjects did not require 2500 ms to make a decision. In the present experiment RTs were measured from the beginning of a word up to 1500 ms. A fixation cross in the center of the screen appeared 500 ms before the stimulus presentation. It remained on the screen until 1500 ms after the beginning of a word. Half of the subjects gave yes-responses with the thumb of the left hand and no-responses with the thumb of the right hand and vice versa. After a variable inter-stimulus interval (500-1000 ms) the next trial began.

Electrophysiological Recordings See section 2.3.2.

Data analysis Separate analyses for percentage of correct responses, for RTs and for ERPs were conducted. In both analyses effects of the factors *original stress* (initial stressed vs. initial unstressed words) and *pitch* (initial stressed pitch vs. initial unstressed pitch) and task (semantic judgment vs. prosodic judgment) were tested.

For the analysis of the ERP data two further factors were included to test for topographic differences of effects. As in ERP-study I and II the factor *hemisphere* contrasted left and right scalp regions. To account accurately for the regional difference of ERP effects in the present study, which differ from the effects of ERP-study I and II (see Figures 4.2 and 4.3), the factor *region* separated effects of the anterior, medial and posterior scalp region. This resulted in the following six regions of interest: anterior left (AF7, AF3, F9, F7, F5, F3, FT9, FT7), medial left (FC5, FC3, C5, C3, CP5, CP3, P5, P3), posterior left (A1, TP9, TP7, P9, P7, PO7, PO3, O1), anterior right (AF8, AF4, F10, F8, F6, F4, FT10, FT8), medial right (FC6, FC4, C6, C4, CP6, CP4, P6, P4) and posterior right (A2, TP10, TP8, P10, P8, PO8, PO4, O2, see Figure 2.2 for electrode sites).

4.2.2 Results

4.2.2.1 Behavioral data

Percent correct There was a significant effect of the factor *task* ($F[1,23] = 189.99$, $p < .001$), indicating more correct responses in the semantic judgment (M = 96.0 %, SD = 3.7) than in the prosodic judgment (M = 72.0 %, SD = 18.8). Furthermore, there were significant interactions of the factors *task* and *original stress* ($F[1,23] = 8.40$, $p < .01$), and of the factors *original stress* and *pitch* ($F[1,23] = 79.87$, $p < .001$) which were modulated by a significant three-way interaction of all factors, namely *task*, *original stress* and *pitch* ($F[1,23] = 72.30$, $p < .001$).

To analyze the three-way interaction, separate ANOVAS including the factors *original stress* and *pitch* were run for the semantic and the prosodic judgment. There was a main effect of the factor *original stress* for the semantic judgment. Only for the prosodic judgment a significant interaction of the factors *original stress* and *pitch* ($F[1,23] = 77.31$, $p < .001$) was found. Subjects categorized words with stress patterns that did not match their original stress more often as correctly stressed than they categorized words with correct pitch as incorrectly stressed. That is, initially stressed words with initially stressed pitch were judged as correctly stressed in 87.6 % of the cases, whereas the same words with initially unstressed pitch were judged as incorrectly stressed in only 59.8 % of the cases ($t[23] = 66.42$, $p < .001$). Vice versa, initially unstressed words with initially unstressed pitch were judged as correctly stressed in 81.4 % of the cases, whereas the same words with initially stressed pitch were judged as incorrectly stressed in 59.0 % of the cases ($t[23] = 25.98$, $p < .001$). Thus, although the categorization of correctly and incorrectly stressed words was clearly above chance, there was a tendency to judge all words as correctly stressed.

RTs Because of the high error rates in the prosodic task, all reaction times were corrected for outliers. Responses slower than one SD below the individual mean and faster than one SD above the individual mean were excluded from the analyses. This constraint excluded 0.6 % of the data. Missing data points, due to errors or outlying RTs, were not replaced. As for the analyses of correct responses, there was a significant effect of the factor *task* ($F[1,23] = 116.54$, $p < .001$). Subjects reacted more slowly (M = 1404 ms, SD = 201) in the prosodic judgment as in the semantic judgment (M = 1050 ms; SD=116). Furthermore, there was a significant main effect of the factor *original stress* ($F[1,23] = 64.34$, $p < .001$) indicating faster responses for initially stressed words (M = 1201 ms, SD = 223) than for initially unstressed words (M = 1254 ms, SD = 257).

Both main effects were modulated by a significant interaction of the factors *task* and *original stress* ($F[1,23] = 15.60$, $p < .01$). Additionally, a significant three-way interaction of the factors *task*, *original stress* and *pitch* reached significance ($F[1,23] = 10.24$, $p < .01$). To analyze the three-way interaction, separate ANOVAs of the factors *original stress* and *pitch* were conducted for the prosodic and the semantic decision task. Both two-way analyses revealed a significant main effect of the factor *original stress* (prosodic judgment: $F[1,23] = 39.91$, $p < .001$; semantic judgment: $F[1,23] = 33.43$, $p < .001$) and significant interactions of this factor with the factor *pitch* (prosodic judgment: $F[1,23] = 12.97$, $p = .001$; semantic judgment: $F[1,23] = 7.71$, $p < .001$). Subjects in both tasks responded more slowly if an initially unstressed word was presented with initially stressed pitch contour (prosodic judgment: M = 1402 ms, SD = 209; semantic judgment: M = 1053 ms, SD = 115) than if the same word was presented with initially unstressed pitch contour (prosodic judgment: M = 1486 ms, SD = 217, $t[23] = 12.96$, $p = .002$; semantic judgment: M = 1075 ms, SD = 121, $t[23] = 11.70$, $p = .002$). That is, incorrect pitch resulted in significantly delayed responses for initially unstressed words across both tasks. The same was true for initially unstressed words in the prosodic decision task. Those with initially stressed pitch were responded to faster (M = 1339 ms, SD = 178) than the same words with initially unstressed pitch (M = 1391 ms, SD = 183, $t[23] = 5.93$, $p = .02$). In contrast, there was no significant differences in RTs to initially stressed words with correct pitch (M = 1036 ms, SD = 110) and for the same words with incorrect pitch in the semantic decision task ($p > .10$).

Taken together, the behavioral responses were sensitive to the pitch manipulation. Subjects could indicate, whether a word carried the correct pitch contour or not. Incorrect pitch contours delayed subjects' responses. However, it appears from the behavioral data, that the pitch modulation was more effective for the initially unstressed words than for the initially stressed words.

4.2.2.2 ERPs

The ERP waveform elicited in the present experiment is characterized by an P1-N1-P2 complex followed by an N400 component. Additionally, there was a prominent late positivity elicited in the present experiment (see Figures 4.2 and 4.3). As it can be seen, the pitch manipulation modified an early ERP deflection, namely the P2. To analyze this effect a time window ranging from 200 to 280 ms was chosen according to visual inspection. In accordance to ERP-studies I and II, a time window ranging from 400 to 600 ms was analyzed for early N400 effects and a time window ranging from 600 to 1000 ms was analyzed for later N400 effects. Obviously, in this time window not only an N400b, but also a late positivity is modulated.

200 to 280 ms: P2 The analysis of mean amplitudes between 200 to 280 ms revealed a significant interaction of the factors *pitch* and *region* (F[1,46] =5.39, p = .01). Only the post-hoc comparisons for the posterior ROI revealed a significant difference (t[23] = 5.28; p = .03). Words with an initially unstressed pitch contour elicited a more negative amplitude in that time window (M = -3.6 μV, SD = 3.5) than the same words with initially stressed pitch contour (M = -2.9 μV, SD = 3.2, compare Figures 4.2 and 4.3). There was no main effect nor any interaction with the factor *original stress*.

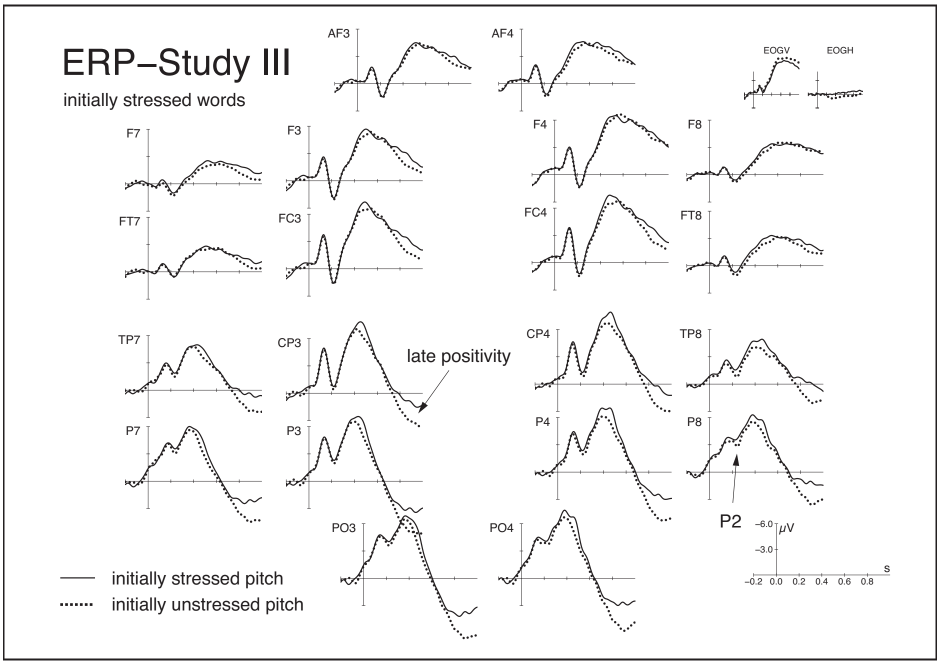

Figure 4.2: *ERPs for selected electrode sites. Solid lines indicate ERPs for initially stressed words with initially stressed pitch contour, dotted lines indicate ERPs for the same words with initially unstressed pitch contour.*

400 to 600 ms: N400a Again there was a significant interaction of the factors *pitch* and *region* (F[1,46] = 5.54, p = .01). This interaction was modulated by a significant interaction of the factors *pitch*, *original stress*, *region* and *hemisphere* (F [1,46] = 3.48, p = .05). Post-hoc comparisons revealed only significant effects for the pitch manipulation of initially stressed words. Those with initially unstressed pitch elicited a less negative amplitude of the ERP in this time window over the medial-left (M = -5.00 μV, SD = 4.6) and the posterior-left scalp region (M = -2.04 μV, SD = 4.8) than the same words with initially stressed pitch (M = -5.75 μV, SD = 5.0, t[23] = 4.88, p = .04 and M = -2.81 μV, SD = 5.2, t[23] = 5.58, p = .03).

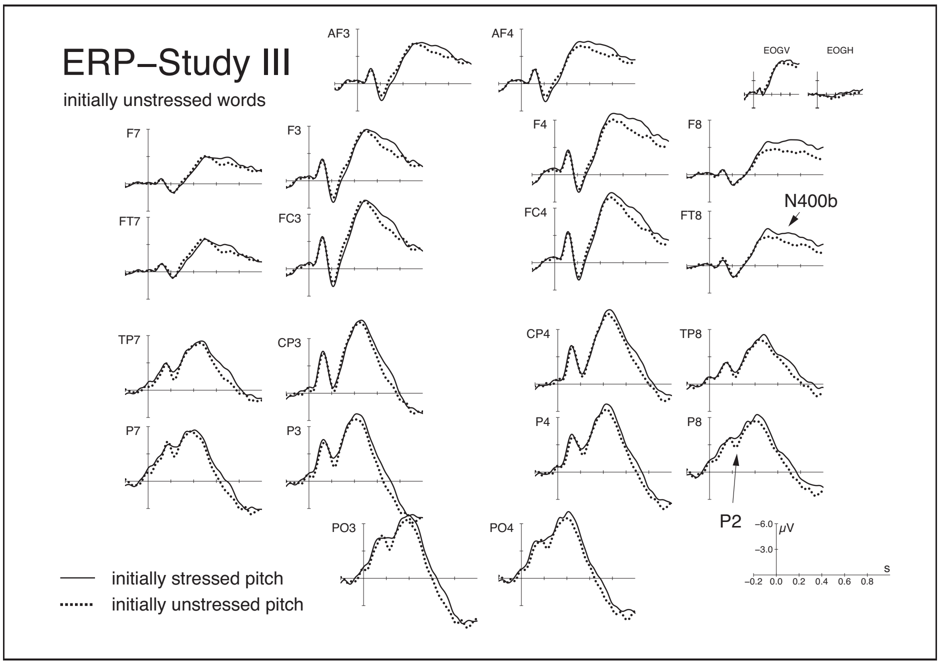

Figure 4.3: *ERPs for selected electrode sites. Solid lines indicate ERPs for initially unstressed words with initially stressed pitch contour, dotted lines indicate ERPs for the same words with initially unstressed pitch contour.*

600 to 1000 ms: N400b and late positivity There was a main effect of the factor *pitch* (F[1,23] = 6.98, p = .01) indicating that words with initially stressed pitch contour elicited more negative amplitudes in this time window (M = -1.03 μV, SD = 5.8) than the same words with initially unstressed pitch contour (M = -0.30 μV, SD = 5.9). Furthermore, there were significant interactions of the factors *pitch*, *original stress* and *hemisphere* (F[1,23] = 4.36, p = .05), and of the factors *pitch*, *original stress*, *hemisphere* and *region* (F[1,46] = 4.17, p = .02). Post-hoc comparisons revealed medial and posterior effects of the pitch manipulation for initially stressed words. Those with initially unstressed pitch elicited more positive amplitudes (medial left: M = 0.08 μV, SD = 6.0, medial right: M = -0.41 μV, SD = 5.1, posterior left: M = 2.98 μV, SD = 6.0, posterior right: M =

1.80 μV, SD = 4.4) than the same words with initially stressed pitch (medial left: M = -0.93 μV, SD = 5.7, t[23] = 6.86, p = .01, medial right: M = -0.41 μV, SD = 5.1, t[23] = 8.11, p = .01, posterior left: M = 2.98 μV, SD = 6.0, t[23] = 11.31, p $<$ .01, posterior right: M = 1.80 μV, SD = 4.4, t[23] = 6.90, p = .01). In contrast, initially unstressed words elicited frontal effects for the pitch manipulation. If they carried initially stressed pitch they elicited significantly more negative amplitudes over the right frontal ROI (M = -3.66 μV, SD = 3.3) than if they carried an initially unstressed pitch contour (M = -2.66 μV, SD = 4.0, t[23] = 4.21, p = .05).

Taken together, there was an early ERP effect for the stress manipulation which appears to be comparable for initially stressed and initially unstressed words. Those with initially unstressed pitch elicited an enhanced amplitude in the P2. Later on, the pitch manipulation for initially stressed words was reflected in a late positive ERP deflection, whereas the pitch manipulation for initially unstressed words was reflected in an N400b deflection.

4.2.3 Discussion

The aim of the present study was to explore behavioral and electrophysiological correlates for the processing of an isolated prosodic parameter, namely pitch, during spoken word recognition. The ERP results indicate that the brain differentiates initially stressed and initially unstressed pitch contours. This differentiation starts as early as 200 ms after the beginning of a word. It can be concluded from this result that already the pitch information of the first syllable of the presented words is extracted by the word processing system, as the mean duration of the first syllable of the words is 330 ms (367 ms for initially stressed words and 297 ms for initially unstressed words, see section 2.2.2). The early discrimination of pitch indicates, as well as the results of previous behavioral research (Lee & Nusbaum, 1993; Miller, 1978; Repp & Lin, 1990; Wood, 1974), that even in stress languages pitch contours are processed by the listener. Moreover, if pitch information is already available for the first syllable it can be used for lexical activation.

Word-initial syllables with unstressed pitch contour elicited a stronger P2 than the same syllables with stressed pitch contour. This result might be interpreted in the way that an initially unstressed pitch contour is detected as a mismatch in relation to the more common pattern of initial stress in German (see section 2.2.4). That is, the P2 modulation might be mainly modulated by a mismatch response. However, it clearly indicates that pitch information is available at this time.

According to the ERPs for initially stressed words it appears that the pitch effect lasts longer than the P2. More positive amplitudes for initially stressed words with initially unstressed than with initially stressed pitch were even observed in the N400a time window. This might result from the longer first syllables of initially stressed as compared to initially unstressed words. Thus, it might take longer to extract the pitch contour of the former as compared to the latter words. That is, the more positive amplitude in the N400a time window elicited by initially stressed words with initially unstressed pitch might well be caused by an overlap of the pitch discrimination effect in the P2 and the N400a for these words.

From the perspective of the present data it appears even more unlikely that the 'N325' is a marker of stress (see section 3.2). This ERP deflection is reported to be a surplus of negativity for weak-initial words as compared to strong-initial words (Böcker, Bastiaansen, Vroomen, Brunia & de Gelder, 1999). The authors interpreted the 'N325' as reflecting early prosodic processing. However, the 'N325' was not replicated in ERP-study I and ERP-study II for initially unstressed words versus initially stressed words with containing only full vowels. Furthermore, the 'N325' shows a completely reversed amplitude difference than the 'P2' effect in the present experiment. Now a surplus of positivity is elicited for words with initially unstressed pitch contours.

Regarding the processing of incorrect pitch, behavioral and electrophysiological data are conclusive. At least RTs for initially unstressed words with incorrect pitch were longer than for the same words with correct pitch across both tasks. Furthermore, incorrect pitch contours are differentiated from correct pitch contours in the ERPs. These results strongly suggest an effective role of pitch during spoken word processing in German. However, as the effects of the pitch manipulation are pronounced in the prosodic judgment and as the effects for initially stressed words were not as robust as that for initially unstressed words, one might argue that the pitch information might not play a critical role during natural spoken word processing. This assumption is supported by the poor discrimination performance in the prosodic task, but it contrasts to the finding of an early electrophysiological marker for pitch discrimination. An unified interpretation of these apparently conflicting results is that the extracted pitch information is processed automatically and that the stress decision requires a later controlled process.

Although, pitch is clearly available for lexical activation, an ERP component, which might be an indicator of selection effort during lexical access, namely the N400a (see section 2.3.4), was not sensitive to the pitch manipulation in the present experiment. The N400a was interpreted as a possible indicator of the effort needed to select the appropriate candidate from the number of activated entries. On the basis of the results of ERP-study I it is not clear, whether the enhanced N400a for the incorrectly stressed versions of initially stressed words is caused by the violation of pitch or by the violation of all prosodic parameters. The present finding suggest that only the violation of all prosodic parameters, but not the violation of pitch alone elicits an N400a effect. However, the behavioral data clearly indicate, that incorrect pitch modulates spoken word recognition. This result can be taken to conclude, that pitch is an important factor during early lexical processing. Behavioral research suggests that prosodic parameters are used for lexical activation and competition (see section 1.2). The question remains, whether modulated initial activation and/or due to pitch can be investigated with the ERP method if a different paradigm than that of ERP-study III. A cross-modal priming experiment which is presented in Chapter 5 directly concerns this issue. This paradigm is thought to reflect activation and competition processes during spoken word recognition. There are indeed effects of the pitch of the primes in that paradigm. These can not only be observed on the behavioral data, but also on the ERPs.

As in ERP-study I there are different results for a prosodic violation of initially stressed and initially unstressed words. However, there was a stronger effect of the stress violation for initially stressed than for initially unstressed words in the former study (see section 2.3), but the reversed pattern holds true for the present data. If these effects would be related to different lexical processes for initially stressed and initially unstressed words, one should expect that they vary similarly in both studies. As this is not the case, it appears that the different effects rather reflect acoustic properties of the materials presented in both experiments, than different lexical processing of initially stressed and initially unstressed words.

I argue that processes yielding to differences between initially stressed and initially unstressed words in the present study do not reflect lexical access. That is, pitch in relation to the remaining prosody information might be, especially in the first syllables, more informative for initially unstressed than for initially stressed words. Amplitude and duration might be strong predictors for a stressed syllable, but not for an unstressed syllable. Initial syllables of stressed words might be marked by duration and amplitude as stressed, whereas that of initially unstressed words might not be as strongly marked by these parameters. The behavioral results are in line with this assumption. Prosodic judgments appeared to be harder for initially stressed than for initially unstressed words. The bias towards a yes-response, as indicated in high percentage of yes-responses, which was pronounced for initially stressed words, might reflect the uncertainty underlying the prosodic judgment. They might rather result from later processes that integrate all available prosodic information of a word and that are engaged in prosodic decision making.

Even electrophysiological responses for the pitch violation of initially stressed and initially unstressed words diverged. There was only an N400b effect of the pitch manipulation for initially unstressed words, but not for initially stressed words. In contrast, the latter words elicited a difference in a late positivity (see Figures 4.2 and 4.3). The latency difference might result from the already mentioned enhanced effort to classify initially stressed words with unstressed pitch contour as incorrectly stressed. Thus, the present results indicate that that neural processes underlying the N400b effect do not provide all analyses that lead to the yes/no response in prosodic decision tasks. Judgments based on more complex or subtle information might be reflected in a later positive ERP deflection. However, the exact relation between processes which elicit the N400b and the late positive effect in the ERP can only be determined if pitch, duration and amplitude are systematically varied for initially stressed and initially unstressed words and effects on both deflection are compared. The results of the present study indicate that such effects might be highly interactive and complex.

In summary of the present results I would like to argue that pitch is early and automatically extracted by the listener regardless of the remaining prosody of the word. This discrimination process is reflected in the P2 component of the electrophysiological signal. As pitch information is available early, it may well be used to constrain lexical activation. However, there was no corresponding indicator in the ERP. It appears that the ERPs elicited in the present experiment method might not be sensitive to lexical activation. Clarity in this respect is provided by the results of the cross-modal priming experiment presented in Chapter 5. Moreover, the information provided by the early pitch discrimination appears to be only poorly used for prosodic decision making. It appears that the judgment required in the present experiment was more complicated than that of ERP-study I. This difference might be caused on the fact that subjects had to base their decisions on highly ambiguous information in the present experiment. The pitch contour is incorrectly applied to half of the words, whereas the remaining prosodic parameters are correct for all words. The behavioral data as well as the N400b and the late positivity in the ERP reflect this complexity. Both ERP deflections might be related to prosodic judgments and, presumably, to prosodic integration during spoken word recognition as they were observed across the prosodic and the semantic task. For initially unstressed words the prosodic decision was easier. The underlying neural processes might be reflected in the N400b effect. For initially unstressed words, later processes than that reflected in the N400b are possibly engaged in the prosodic decision.

Chapter 5

Pitch manipulated word fragments

5.1 Overview

In this chapter the role of *pitch contour* during spoken word recognition is further explored. Results of ERP-study I revealed that ERPs elicited by spoken words with incorrectly applied pitch, amplitude and duration information might harm lexical access as compared to words that bear incorrect duration and amplitude, but correct pitch. The question emerged, whether lexical access is sensitive only to a variation of all prosodic parameters or especially sensitive to pitch. Results of ERP-study III were not conclusive in that respect. They indicated that pitch is extracted early from the acoustic signal. However, there was no ERP deflection which could be related to a significant role of pitch during lexical access.

ERPs for entirely presented spoken words might not be sensitive to subtle differences in the activation of lexical entries in the mental lexicon. The lack of an ERP effect for lexical activation and the finding of an ERP index of enhanced selection effort for words with incorrectly applied pitch, amplitude and duration information do not conflict if one assumes that lexical activation and competition take different prosodic parameters into account. The present two experiments were aimed to explore the function of pitch for lexical activation and competition more directly. Especially the paradigm applied in ERP-study IV, namely cross-modal word fragment priming, has previously been discussed to tap into early stages of spoken word processing (Cooper, Cutler & Wales, 2002; Cutler & van Donselaar, 2001; Marslen-Wilson, 1990; Soto-Faraco, Sebastian-Galles & Cutler, 2001). On the basis of the available behavioral evidence, it cannot be distinguished whether prosody is used during activation and/or competition. The results of ERP-study IV should reveal whether pitch on one hand and duration on the other hand are differently exploited during lexical activation and competition.

The two studies which are explored in the present chapter differ in fundamental issues from the studies presented in the preceding chapters. Spoken words were presented in the former studies. Word fragments, more specifically word onset syllables, were presented in the two current experiments. Both experiments have in common that two versions of word onset syllables were presented. One version carries a pitch contour which is characteristic for a stressed syllable. Another version of each syllable carries an unstressed pitch contour. In both experiments half of the syllables originated from initially stressed words, half from initially unstressed words. That is, in both experiments duration and amplitude, which are referred to in the following as *remaining prosody*, were controlled for.

The first experiment presented in the present chapter is a word completion task. The logic behind this paradigm is in parallel to that of the gating task presented in Chapter 2 (see section 2.2). When subjects are required to complete a word onset syllable to a word, responses may well reflect whether or not the target's lexical entry is activated by the acoustic information of the fragment. As introduced in Chapter 2, results of word guessing experiments indicate that subjects use prosodic information to complete words. In the first experiment of the present chapter it is explored whether pitch, as a single prosodic parameter, modulates the stress pattern of word guesses in a word completion task. If so, subjects should produce more initially stressed words for fragments with a stressed pitch contour than for the same fragment with an unstressed pitch contour. If pitch information alone influences the stress pattern of word completion, it can be assumed that pitch is used for lexical activation.

A more direct investigation of activation and competition during spoken word recognition is investigated in the second experiment of the present chapter. In this experiment cross-modal word fragment priming was used to investigate the function of pitch. Auditory primes were taken from the fragments presented in the word completion task. Following the primes, visual targets were presented. Target words were taken from word guesses of the word completion task. They either match the pitch of the prime or not. Reaction times and ERPs were recorded for the targets. If pitch is used for lexical activation, responses to targets with a stress pattern that matches the pitch of the prime should be faster than targets with a stress pattern that does not match the pitch of the prime. If there is a behavioral facilitation for pitch match, the ERP signal for target words that match the pitch of the primes might be modulated in an access related component.

5.2 Word completion

5.2.1 Method

Subjects Twenty participants were paid for their participation. All were native German speakers with no reported hearing or neurological disabilities.

Materials The words presented in the gating task and in ERP-study I (see Appendix A) were reduced to their first syllables. These word onset syllables were presented to the subjects. Eighty fragments were taken from initially stressed German words (e.g., BAL from BALsam), and eighty fragments were taken from initially unstressed German words (e.g., ak from akTEUR). All words were bisyllabic and contained only full vowels (a description of the words is given in section 2.2.2).

The words, spoken by a female native speaker of German, were recorded onto Digital Audio tape and digitized at 44.1 KHz and a 16 bit sampling rate. The first syllable of each word was isolated with a sound editor (CoolEdit v. 1.52, ©Syntrillium Software Corp.). Then a speech editor (PRAAT version 3.2, ©Paul Boersma & David Weenink) was used to extract the F_0 contour of every syllable.

To establish the time course of F_0, the onset-, the highest- and the offset-measured value and their corresponding time points in relation to the syllable onset were determined and averaged for every syllable. Stressed syllables show an onset of the voiced signal at 50 ms with 161 Hz, a maximum at 200 ms with 209 Hz and an offset at 304 ms with 166 Hz. In contrast, unstressed syllables are characterized by an equal F_0 onset as the stressed syllables, but a weaker maximum with 191 Hz at already 96 ms and an earlier offset at 248 ms with 158 Hz.

Onset, maximum and offset F_0 point for stressed syllables were interpolated to create an averaged stressed as well as an averaged unstressed F_0 contour. Each interpolated F_0 contour was fitted to the voiced portion of each syllable. That is, the stressed F_0 contour was applied to one version of a syllable and the unstressed F_0 contour was applied to another version of the same syllable. The fitting to periodical parts of the acoustic signal was automatized with PRAAT. This automatic procedure left duration and amplitude of the syllables constant (see Figure 5.1 for an example of the pitch manipulation and Table 2.1 on page 37 for duration and amplitude measures). The re-synthesized syllables were presented in two pseudo-randomized orders. Both orders contained all 80 syllables in both versions. Stressed and the unstressed versions of each syllable were presented in counterbalanced order across subjects.

Procedure Syllables were presented binaurally via loudspeakers. Following each syllable, subjects were asked to write down a word beginning with that syllable. Subjects were not aware of the pitch manipulation. They were instructed to write down the first word that came to their mind beginning with that syllable, even if they had already produced this word in a former trial. After the written answer had been given, the subjects had to press a response key to start the presentation of the next syllable.

Data analyses The responses were scored by hand and classified as belonging either to the category of *initially stressed words* (including monosyllabic words) or as belonging to the category of *initially unstressed words*. The data of one

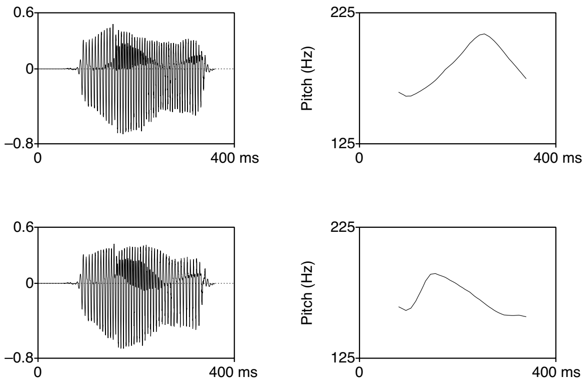

Figure 5.1: *Waveforms (left) and pitch contours (right) of the two re-synthesized versions of the syllable 'ME' from the initially stressed German word 'MEtro' (Engl.: metro). The version with stressed pitch contour is illustrated above, the version with unstressed pitch contour is illustrated below.*

group were completely redundant for statistical analysis. For instance, the production of initially unstressed words equals the production of initially stressed words subtracted from 100 %. Thus, a two-way ANOVA including the factors *pitch contour of the fragment*, *original stress of the fragment* was conducted for the initially stressed words only.

5.2.2 Results and Discussion

Subjects gave valid responses to each syllable. 78.3 % (SD = 6.6) of the words generated by the subjects words were classified as initially stressed. Only 21.7 % (SD = 6.7) of the words were classified as initially unstressed. The overwhelming percentage of initially stressed responses appears to reflect the predominance of this stress pattern in German (Dogil, 1999, see section 2.2.4 for a discussion of this aspect). Furthermore, the number of initial stressed words is pushed by the inclusion of monosyllabic words into this group.

The *pitch manipulation* showed the expected effect on the stress pattern of the generated words. After listening to syllables with stressed pitch subjects responded more often with initially stressed words (M = 80.4 %, SD = 9.0) than after listening to the same syllables with unstressed pitch (M = 76.2 %, SD = 11.0, t[1,19] = 22.85, $p < .001$, F[1,19] = 21.92, $p < .001$). As in the gating experiment presented in Chapter 2, there was a significant effect of the factor *original stress of the fragment* (F[1,19] = 104.45, $p < .001$). After listening to fragments taken from initially stressed words, subjects produced more initially stressed words (M = 84.7 %, SD = 6.9) than after listening to syllables taken from initially unstressed words (M = 71.9 %, SD = 9.0).

The main effects of the factors *pitch of the fragment* and *original stress of the fragment* were modulated by a significant interaction of both factors (F[1,19 = 10.4, p < = .001). However, significant effects in the direction of the main effect of the factor *pitch* were found for syllables originating from initially stressed words as well as for syllables originating from initially unstressed words. Originally stressed syllables with stressed pitch were responded to with 85.8 % (SD = 7.5) initially stressed words. The same syllables with unstressed pitch were responded to with only 83.5 % (SD = 6.3) initially stressed words (t[19] =5.32, p = .03). Originally unstressed syllables with stressed pitch evoked the production of 75.0 % (SD = 7.1) initially stressed words, whereas the same syllables with unstressed pitch were responded to with only 68.9 % (SD = 9.7, t = 17.22, $p < .001$) initially stressed words. Thus, regardless of the remaining prosody, word completion was more likely to be initially stressed and less likely to be initially unstressed if pitch was stressed than if pitch was unstressed.

These results of the word completion task confirm that pitch does play a role in the processing of stress, although the effect was less strong for syllables originating from initially stressed words. They clearly reveal that listeners can use a single prosodic feature to guide word completion and replicates previous findings that listeners use prosodic information in word completion tasks (Cooper, Cutler & Wales, 2002; Cutler & van Donselaar, 2001; Cutler & Otake, 1999; Mattys, 2000). Moreover, a single prosodic parameter, namely pitch, is powerful enough to slightly distort word completion.

The main effect for the factor *remaining prosody* is difficult to interpret. As well as in gating experiment presented in Chapter 2, cohort differences between the initially stressed and initially unstressed original words might have biased responses (see section *gating* for a discussion of this issue). Interestingly, however, the stronger effect of the pitch manipulation for initially unstressed words is in line with the stronger behavioral effect of the pitch manipulation for initially unstressed words in ERP-study III. The present result might suggest that the pitch manipulation is more prominent and does not conflict with other prosodic parameters for initial portions of initially unstressed words as compared to initially stressed words.

An unifying interpretation of the result of the gating task and the results of ERP-study III is that the present data reflect late processing stages of spoken word recognition as well as lexical activation. Accordingly, the subtle effect of the pitch manipulation (regarding only 4.2 % of the responses) might not necessarily attest that pitch plays a subordinate role for word completion in German. Pitch might not be as easy available to processes engaged in a word completion task or might be overwritten by effects of the remaining prosody which might have stronger effects on spoken word recognition. Only a paradigm which is supposed to be more *online* than word completion can answer accurately which processes are modulated by pitch and the remaining prosody. Such a paradigm is used in the following cross-modal word fragment priming experiment.

5.3 ERP-Study IV: Cross-modal fragment priming

5.3.1 Introduction

In the present experiment syllables from Experiment 1 were used as primes during cross-modal word fragment priming. Using this paradigm it was examined whether pitch and/or the remaining prosody modulate the activation status of visually presented targets with different stress patterns. As in the previous experiment each prime was presented in two versions: with a stressed pitch contour and with an unstressed pitch contour (see Figure 5.1). Again, the remaining prosodic features of the primes were controlled for by using the same number of syllables from initially stressed and initially unstressed base words.

Targets were presented visually. Two main types of targets were combined with the primes: segmentally matching targets and segmentally mismatching targets. The group of segmentally matching targets was taken from words produced in the word completion task. For instance, the word *Medium*, which was frequently produced after the presentation of the syllable *me* taken from the word *Metro*, was in ERP-study IV used as target for the prime *me*. A second group of targets was established by words whose initial segments were not identical to that of the primes. These words were matched in number of letters and frequency to the segmentally matching words. An example for a segmentally mismatching target is the word *Heizung* preceded by the prime *me*. For each prime an initially stressed and an initially unstressed segmentally matching word and an initially stressed and an initially unstressed segmentally mismatching word was selected. The resulting experimental conditions for the prime syllable *ME* with unstressed pitch contour are illustrated in Figures 5.2 and 5.3. The syllable *ME* used for illustration purpose in Figures 5.2 and 5.3 originates from the initially stressed word *MEtro*. Thus, the pitch of the prime signals an unstressed syllable, whereas the remaining prosodic features of the same prime signal a stressed syllable. As segmentally matching words *MEdium* and *mediZIN* were selected. Both were frequently produced guesses for the fragment ’ME’ in the word completion task. The words *HEIzung* and *kataLOG* do not match the segments of the prime. The combination of primes and targets resulted in three different match/mismatch conditions: next to the segmental match/mismatch condition, there was a pitch match/mismatch condition and a remaining prosody match/mismatch condition. The initially unstressed words *mediZIN* and *kataLOG* match the unstressed pitch of the prime. The initially stressed words *MEdium* and HEIzung do not match not the pitch of the prime (see Figure 5.2). However, the latter two words match the remaining prosody of the prime (see Figure 5.3). Note, that the pitch match/mismatch conditions are reversed for ME with stressed pitch contour, whereas the match mismatch conditions for the remaining prosody do not change.

Former experiments with word fragment priming studies have shown effects of suprasegmental information only for the combination of all prosodic features and only for targets with the same initial segments as the prime (Cooper, Cutler & Wales, 2002; Cutler & van Donselaar, 2001; Soto-Faraco, Sebastian-Galles &

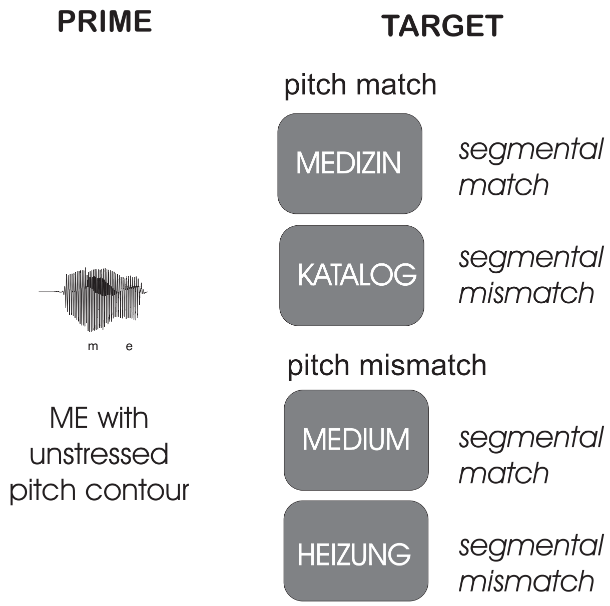

Figure 5.2: *Pitch match/mismatch conditions realized for the prime ME with unstressed pitch.*

Cutler, 2001, see section 1.2.2.2). The present study diverges from the previous ones on three main aspects. First, the stress pattern of all target words is controlled for in order to see whether prosodic features activate lexical representations independent from their segmental information. A second new aspect is, that effects of prosody in cross-modal word fragment priming were split into effects related to pitch and into effects related to the remaining prosody of the fragments. A third novelty is the use of a new methodology, namely the recording of ERPs during cross-modal word fragment priming.

The recording of ERPs provides a promising tool to investigate the time course during which segmental, pitch and remaining prosodic information is used by the listeners. With this method it can be explored, whether the segments, the pitch and the remaining prosody are used at the same time during lexical access. However, cross-modal word fragment priming only indirectly explores lexical access during spoken word recognition. ERP deflections elicited by the visual words that are sensitive to different aspects of the auditory prime might not necessarily be equal to those for access related deflections elicited by spoken words.

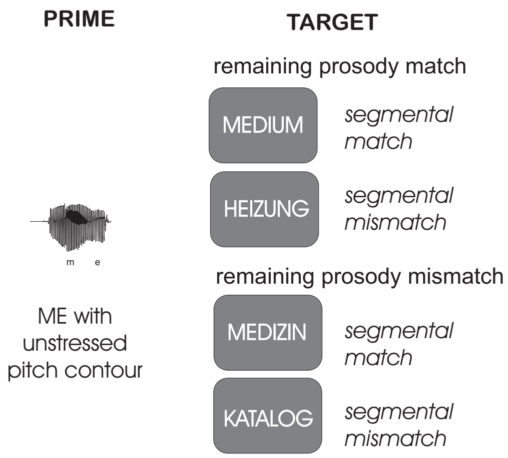

Figure 5.3: *Remaining prosody match/mismatch conditions realized for the prime ME with unstressed pitch.*

Cross-modal word fragment priming is assumed to tap activation in a modality-independent mental lexicon (Marslen-Wilson, 1990). There is current neurophysiological evidence of the time course of the use of such a hypothetical modality independent mental lexicon. A magnetoencephalographic (MEG) component named M350 is reported to vary in latency with the same stimulus properties that affect RTs in various word recognition tasks (Embick, Hackl, Schaeffer, Kelepir & Marantz, 2001; Pylkkänen, Stringfellow, Flagg & Marantz, 2000; Pylkkänen, Stringfellow & Marantz, 2002). Frequent words elicit shorter reactions and an earlier M350 latency than infrequent words (Embick, Hackl, Schaeffer, Kelepir & Marantz, 2001). Furthermore, the second presentation of a word elicits facilitated RTs and an earlier peak of the M350 than the first presentation of the same word (Pylkkänen, Stringfellow, Flagg & Marantz, 2000). However, in contrast to a behavioral inhibition for words with a high density of phonological neighbors, the latency of the M350 is not affected by neighborhood density (Pylkkänen, Stringfellow, Flagg & Marantz, 2000; Pylkkänen, Stringfellow & Marantz, 2002). The results are interpreted by means of a model of spoken word recognition, namely NAM (Neighborhood activation model by Vitevitch & Luce, 1998, 1999, see section 1.2.2.2). NAM assumes, as all current models of spoken word recognition do, that multiple lexical entries are activated by the acoustic input. Activation is faster for high frequent words and for pre-activated entries as it is the case for repeated presentation. The selection of the appropriate candidate among the activated entries depends on the number of activated competitors. According to NAM, the M350 latency might reflect initial activation as it is sensitive to frequency and repetition. In contrast, the M350 does not appear to reflect selection effort, as it is not sensitive to neighborhood density. Pylkkänen et al. (2000) concluded, that the M350 should be the first shared component between auditory and visual word recognition. Taken together, there is neurophysiological evidence, that activation in a modality independent mental lexicon takes place at 350 ms following the onset of the visual presentation of a word.

For the current experiment it was hypothesized that the primes segmental, pitch or remaining prosodic information affect the activation status of lexical representations for target words. Reaction times should be faster for targets that share segments, pitch or the remaining prosody with the prime as compared to targets that do not share the specific information. According to the M350 related to lexical activation, the modulated activation status in the present experiment should be reflected in the ERPs at about 350 ms after the onset of the visual target.

5.3.2 Method

Subjects Twenty-four right-handed subjects (12 male) aged 19 to 32 years (M = 24.5 years) participated in two sessions of the experiment in return for a small payment. All subjects were native speakers of German with no reported hearing or neurological problems. None of the subjects had participated in Experiment 1. The experimental sessions were separated by a minimum of two weeks.

Materials A complete list of the material used in the present experiment is listed in Appendix D. 80 syllables (half from initially stressed words, half from initially unstressed words) were taken from the word completion task (see Column I in Appendix D). For the group of segmentally matching words two- or three-syllabic nouns generated in Experiment 1 were selected. For each prime syllable an initially stressed and an initially unstressed segmentally matching word was selected (see column I and II in Appendix D). Not all of these words were exactly identical to the fragments in syllable structure, co-articulation at the end of the first syllable and vowel length. These differences are not critical for *pitch* effects as targets serve as their own controls for this condition. To warrant a meaningful interpretation of effects of the *remaining prosody* an attempt was made to distribute critical words equally across the match and mismatch conditions.

Furthermore, initially stressed and initially unstressed words were chosen so that across both groups word length, word frequency, and number of syllables were comparable. Note, that the restricted number of primes resulted from the restricted number of nouns that matched these selection criteria. The selected initially stressed words had a mean length of 6.2 letters (SD = 0.3), initially unstressed words had a mean length of 6,1 letters (SD = 0.3). This difference was not significant in a t-test for independent samples (t[1,358] = 1.15, p = .15). The word frequency ranged within the frequency categories 9 to 17 (as related to the most frequent German word, that is the article 'der' [English: the]). The median category index of the word frequency was 14 (range 11 to 17) for initially stressed and initially unstressed words (Quasthoff, 1998). Both groups consisted of 58 two-syllabic and 22 three-syllabic words.

Words that did not match the segments of the prime were selected from an online German database (Quasthoff, 1998). Each of these words was matched in terms of stress pattern, word length, word frequency and number of syllables to a segmentally matching word. This resulted in 80 initially stressed segmentally mismatching words (see column III in Appendix D) and 80 initially unstressed segmentally mismatching words (see column IV in Appendix D). No item was a compound word and no item with an initial syllable that forms a monosyllabic word was used. For each segmentally matching and each segmentally mismatching word a phonological legal pseudoword was created by interchanging letters of the last syllable (e.g. MEDIUM was changed to MEDIMU).

Two versions of each prime were presented: one with a rising pitch contour and one with a falling pitch contour (see method section of the previous experiment for details). Two experimental lists contained prime-target pairs in randomized order. In each list, half the primes had a rising pitch contour whereas the remaining primes had a falling pitch contour. A prime was repeated 8 times as it was paired with four words and four pseudowords. Four repetitions of a prime were included in one experimental list and the remaining four repetitions were included in the other experimental list. Subjects were tested for one version of the two lists in the first session of the experiment. The second version was tested in the second session of the experiment. Twelve subjects began with the first, twelve with the second list.

Procedure Participants were comfortably seated in an electrically and acoustically shielded chamber. Visual stimuli were presented on a computer screen in front of the subjects. Loudspeakers were placed to the left and to the right side of the screen and used to present word fragments.

An experimental trial began with the presentation of a fixation cross in the center of the screen. Participants were instructed to fixate the cross at any time it appeared on the screen. After 300 ms the prime syllable was presented via loudspeaker while the fixation cross remained on the screen. Immediately following the auditory prime, the fixation cross was replaced by a visual target. Targets were presented in uppercase white letters on a black background. The task was to respond as fast and as correct as possible whether the target was a word or not. Half the subjects made yes-responses with the left thumb and no-responses with the right thumb, for the remaining subjects response hands were reversed. Reaction times were measured from stimulus onset with a timeout of 1300 ms. The next trial started after a fixed inter-stimulus interval of 1000 ms.

Electrophysiological Recordings See section 2.3.2.

Data Analysis Behavioral responses to the words were analyzed in separate ANOVAs for RTs and error percentages. In each analysis three within-subject factors, namely, *segmental congruency* (segmental match vs. segmental mismatch), *pitch congruency* (pitch match vs. pitch mismatch) and *remaining congruency* (remaining prosodic match vs. remaining prosody mismatch), were tested (see table 5.1 for an illustration of the factorial design). Additionally, a factor *original stress of the fragment* was included in the analysis to control whether the effects are comparable for fragments originally taken from initially stressed words and for fragments originally taken from initially unstressed words.

prime: *me* with unstressed pitch, taken from the initially stressed word *MEtro*

target	*segments*	*pitch*	*remaining prosody*
MEDIUM (initially stressed)	match	mismatch	match
MEDIZIN (initially unstressed)	match	match	mismatch
HEIZUNG (initially stressed)	mismatch	mismatch	match
KATALOG (initially unstressed)	mismatch	match	mismatch

Table 5.1: *Example of the match and mismatch conditions in ERP-study IV realized for one prime syllable. Note, that for the same syllable with stressed pitch contour only the pitch match/mismatch conditions were reversed, whereas the remaining conditions were the same.*

ERP analyses included two additional factors. The factor *hemisphere* differentiates left and right electrode positions. The factor *region* differentiates anterior and posterior electrode positions. This resulted in the following regions of interest (ROIs): anterior left (F9, F7, F5, F3, FT9, FT7, FC5, FC3, T7, C5, C3), anterior right (F10, F8, F6, F4, FT10, FT8, FC6, FC4, T8, C6, C4), posterior left (TP9, TP7, CP5, CP3, P9, P7, P5, P3, PO7, PO3, O1) and posterior right (TP10, TP8, CP6, CP4, P10, P8, P6, P4, PO8, PO4, O2).

5.3.3 Results

5.3.3.1 Behavioral data

Tables 5.2 and 5.3 summarize the behavioral effects of the three experimental conditions of primary interest in ERP-study IV. The overall percentage of errors in the present experiment was 10.7 % (SD = 3.8).

	segments	*pitch*	*remaining prosody*
match	666 (91.9)	674 (89.3)	673 (89.9)
mismatch	687 (86.5)	678 (89.2)	679 (88.5)

Table 5.2: *Reaction times in ms (and percent correct) for the three experimental conditions in ERP-study IV*

Percent correct The three-way ANOVA of error percentages only revealed significant main effects of the factors *segmental congruency* ($F[1,23] = 36.21$, $p < .001$) and *remaining congruency* ($F[1,23] = 10.46$, $p < .01$) (see 5.2). More errors were made for segmentally mismatching words (M = 86.5 percent correct, SD = 6.6) than for segmentally matching ones (M = 91.9 percent correct, SD = 4.9). Accuracy was higher for words with a stress pattern as indexed by the remaining prosody of the prime (M = 89.9 %, SD = 6.3) than for words with a stress pattern different from the remaining prosody of the prime (M = 88.5 %, SD = 6.5). There was no significant effect of the factor *pitch congruency* nor any interaction with that factors for the correct responses.

There was no significant effect for the factor *original stress of the fragment* ($F[1,23] = 3.46$, $p = .08$). However, a significant interaction of the factors *segmental congruency*, *remaining congruency* and *original stress* ($F[1,23] = 7.29$, $p = .013$) revealed different effects for targets following fragments which were originally taken from initially stressed or initially unstressed words. Significantly reduced error rates for targets that match the primes remaining prosody were observed for targets that match the segments of the prime (initially stressed fragments: $t[23] = 13.46$, $p < .01$, initially unstressed fragments: $t[1,23] = 3.36$, $p < 0.79$). In contrast, for segmentally mismatching targets there was only a reduced error rate for targets that match the remaining prosody of fragments from initially unstressed words (t[23] = 4.35, p = .048), but not for fragments taken from initially stressed words (t [23] ¡ 1). That is, if the fragment was originally stressed, an initially stressed target with the same initial segments as the prime was responded to more correctly than an initially unstressed target with the same segments as the prime. This effect was not as strong for originally unstressed fragments followed by initially unstressed targets. However, even initially unstressed targets that do not match the segments of the prime were responded to more correct of they followed an originally unstressed fragment. No further interaction reached significance for the analysis of the correct responses.

	segmental match		*segmental mismatch*	
	P	RP	P	RP
match	663 (91.5)	658 (93.0)	687 (87.9)	690 (86.9)
mismatch	669 (92.5)	674 (90.8)	688 (84.1)	685 (84.2)

Table 5.3: *Reaction times in ms (and percent correct) for both prosodic conditions in ERP-study IV separated by words that match the segments of the prime and such that do not match the segments of the prime. P = pitch, RP = remaining prosody.*

RTs Because of the high error rate, RTs were corrected for outliers. RTs slower than 2 SD added to the individual mean of a subject and RTs faster than 2 SD subtracted from the individual mean of a subject were excluded. Missing data points, due to errors or outlying RTs, were not replaced. The exclusion criterion yielded to the rejection of 1.2 % of the data.

There were significant main effects of all four analyzed factors (*segmental congruency*: F[1,23] = 111.81, p < .001, *pitch congruency*: F[1,23]=4.74, p=.04, *remaining congruency*: F[1,23] = 14.80, p < .001, *original stress of the fragment*: F[1,23] = 4.96, p = .04). Segmentally matching words were responded to faster (M = 666 ms, SD = 61) than words that do not match the segments of the prime (M = 688 ms, SD = 58). Words with a stress pattern as indexed by the pitch of the prime were responded to faster (M = 675 ms, SD = 61) than words with a different stressed pattern (M = 679 ms, SD = 61). Targets with a stress pattern as indexed by the original prosody of the prime were responded to faster (M = 674 ms, SD = 61) than targets with a stress pattern different from the original prosody of the prime (M = 679, SD = 61). When fragments were taken from originally stressed words, targets were responded to faster (M = 675 ms, SD = 60) than when primes were taken from originally unstressed words (M = 679 ms, SD = 62).

A significant interaction of the factors *segmental congruency* and *remaining congruency* (F[1,23] = 26.54, p <.001) revealed that the facilitation effect of match between the primes' remaining prosody and the targets stress pattern was caused by segmentally matching targets (see Table 5.3). Segmentally matching targets showed a significant RT advantage (M = 658ms, SD = 60) when they matched the primes' remaining prosody than when they do not match the primes' remaining prosody (M = 674 ms, SD = 61, (t[23] = 59.55, p < .001). In contrast, RTs for segmentally mismatching targets were not facilitated by the original prosody of the prime. (match: M = 690 ms, SD = 59, mismatch: M = 685 ms, SD = 59, t[t23] = 3.11, p = .09). There was no significant interaction of the factors *segmental congruency* and *pitch congruency* (F[1,23] = 1.72, p = .20, see Table 5.3).

A significant interaction of the factors *remaining congruency* and *pitch congruency* (F[1,23] = 3.95, p = .058) indicated that targets which match the fragments pitch and their original prosody elicited shorter reaction times (M = 670 ms, SD = 61) than any other condition (pitch match/remaining prosody mismatch: M = 679 ms, SD = 60, t[23]= 24.03, p < .001; pitch mismatch/remaining prosody match: M = 677 ms, SD = 60, t[23] = 12.01, p = .002, pitch mismatch/remaining prosody mismatch: M = 680 ms, SD = 60, t[23] = 12.01, p = .002). No further interaction reached significance for the analysis of the reaction times.

In summary, the pitch manipulation affected RTs, but not the percentage of correct responses. As expected, a congruency between the primes' pitch contour and the stress pattern of the following targets facilitated RTs. This effect was not modulated by the segments of the prime. However, only targets that are congruent with all prosodic features of the prime, including pitch and the remaining prosody, were facilitated in RTs.

5.3.3.2 ERPs

Figure 5.4 plots the ERPs for words that match or mismatch the segments of the prime. As can be seen, the elicited waveforms equal that of other prime-target combinations (see Figures 5.5, 5.6, 5.7, 5.8, 5.9 and 5.10). The ERPs illustrations start with a baseline of 200 ms that precedes the onset of the visual presentation. The following 600 ms represent the electrophysiological response up to the beginning of the visual presentation.

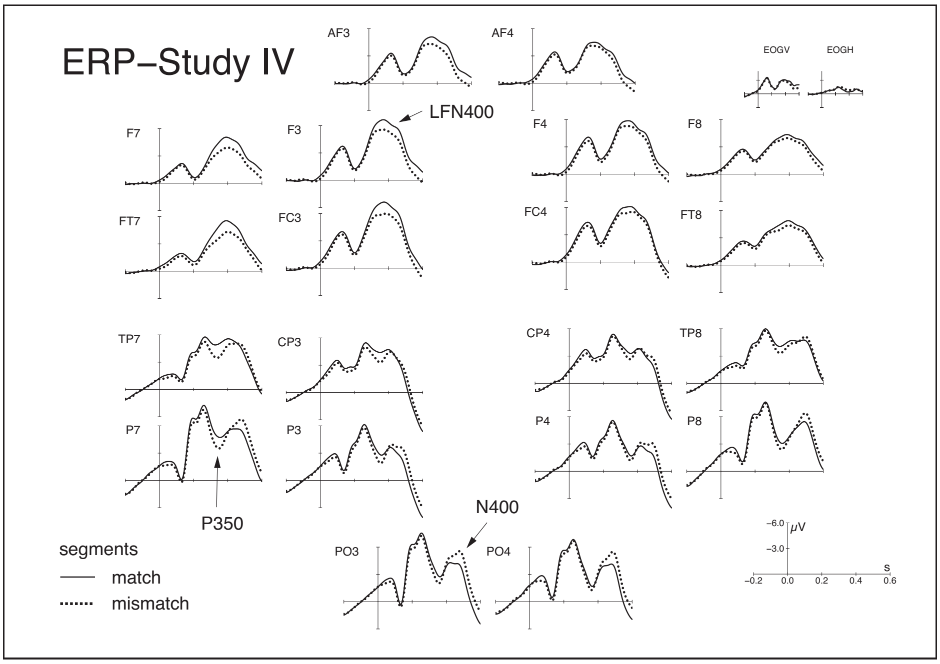

Figure 5.4: *ERPs for selected electrode sites. Solid lines: targets that match the segments of the prime, dotted lines: targets that do not match the segments of the prime.*

The first 200 ms of the ERP waveform are characterized by an N1-P2 complex with posterior scalp distribution. The differentiation for segmentally matching and segmentally mismatching starts at 300 ms (see Figure 5.4) with a positive-going ERP deflection over posterior scalp regions and a negative-going deflection over left-frontal scalp regions. Following the positive waveform, there is an N400 effect over posterior scalp region.

In the following the posterior waveform will be named as P350, referring to a positive going deflection peaking at approximately 350 ms. The P350 wave ranges from 300 to 400 ms after target onset. The frontal waveform is named LFN400, referring to its left frontal scalp distribution and its negative peak at 400 ms. The LFN400 ranges from 300 to 500 ms after target onset. The P350 is followed by a negative waveform over posterior scalp regions. This negative ERP deflections appears to be an N400 component, it ranged from 400 to 600 ms after target onset. As can be seen in Figures 5.5, 5.6, 5.7, 5.8, 5.9 and 5.10 the three ERP deflections described above are differently modulated by the primes' pitch information and by their remaining prosody.

In case of interactions of the factors region and hemisphere with the experimental manipulation, the description of the ERP analyses focuses on different ROIs for the P350, the LFN400 and the N400. To analyze the P350 effect, amplitudes in a time window of 300 to 400 ms after target onset were averaged. As the P350 overlaps in time with the LFN400 the description of P350 effects focuses on both posterior ROIs. The LFN400 was analyzed with a time window ranging from 300 to 500 ms after target onset. According to the apparent frontal distribution of the LFN400 deflection, the description of this ERP deflection focusses on both frontal ROIs. The third time window for the analysis of N400 effects ranged from 400 to 600 ms after target onset.

300 to 400 ms: P350 A five-way ANOVA of the first time window revealed a main effect of the factor *segmental congruency* ($F[1,23] = 26.73$, $p < .001$] which was modulated by interactions with the factor *hemisphere* ($F[1,23] = 41.62$, $p < .001$] and with the factors *hemisphere* and *region* ($F[1,23] = 10.17$, $p < .01$]. The separate analysis of both posterior ROIs revealed significant P350 effects for the factor *segmental congruency* (posterior left ($t[1,23] = 21.50$, $p < .001$], posterior right ($t[1,23] = 7.32$, $p = .013$]). Although mean amplitudes between 300 and 400 ms were negative, segmentally mismatching targets elicited a P350 with a more *positive* amplitude over both posterior ROIs (posterior left: M = -3.88 μV, SD = 4.09; posterior right: M = -3.18 μV, SD = 4.53) than segmentally matching targets (posterior left: M = -4.68 μV, SD = 4.12; posterior right: M = -3.63 μV, SD = 4.43, compare Figure 5.4).

Furthermore, a main effect of the factor *pitch congruency* (F[1,23] = 6.05, p = .02) modulated by an interaction the factor *region* (F[1,23] = 13.35, p $<$ .01) was found. There was a significant effect of the pitch manipulation only for the posterior scalp region (t[1,23] = 9.50, p $<$ .01). A target preceded by a prime with a pitch contour that matched the targets stress pattern elicited an enhnaced amplitude of the P350 (M = -3.54 μV, SD = 4.44) than the same target preceded by a prime with a mismatching pitch contour (M = -4.14 μV, SD = 4.45).

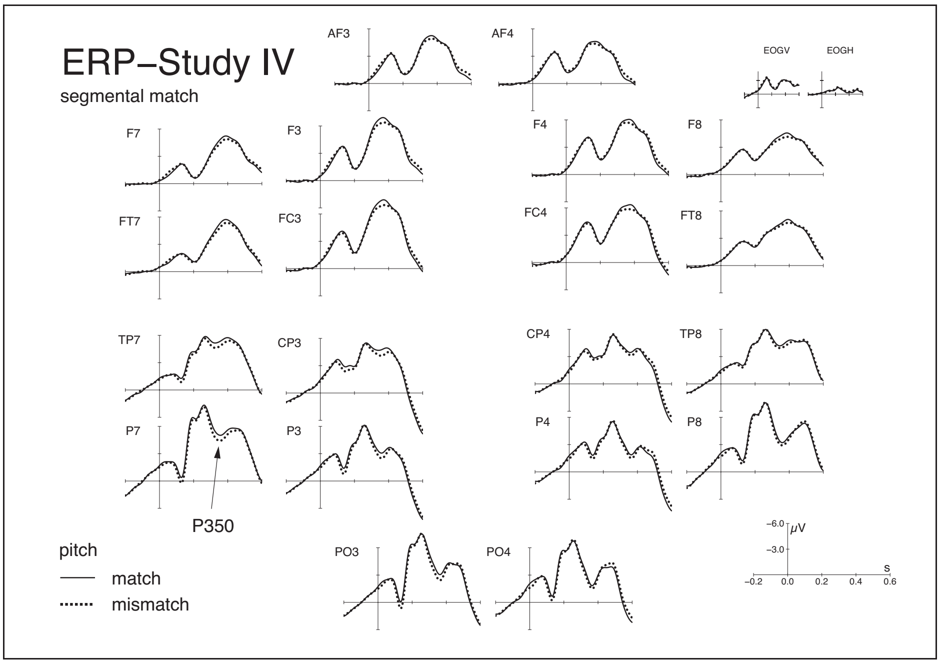

Figure 5.5: *ERPs for selected electrode sites for segmentally matching targets. Solid lines: targets with a stress pattern that matches the pitch of the prime, dotted lines: targets with a mismatching stress pattern.*

The three-way interaction of the factors *segmental congruency*, *pitch congruency* and *region* was not significant (F[1,23] = 3.04, p = .10). However, eyeballing indicates that the pitch effect was smaller for segmentally matching than for segmentally mismatching targets. Indeed, there was only a trend for the former (F[1,23] = 3.25, p = .08), but a significant effect for the latter targets (F[1,23] = 7.53, p = .01, compare Figures 5.5 and 5.6).

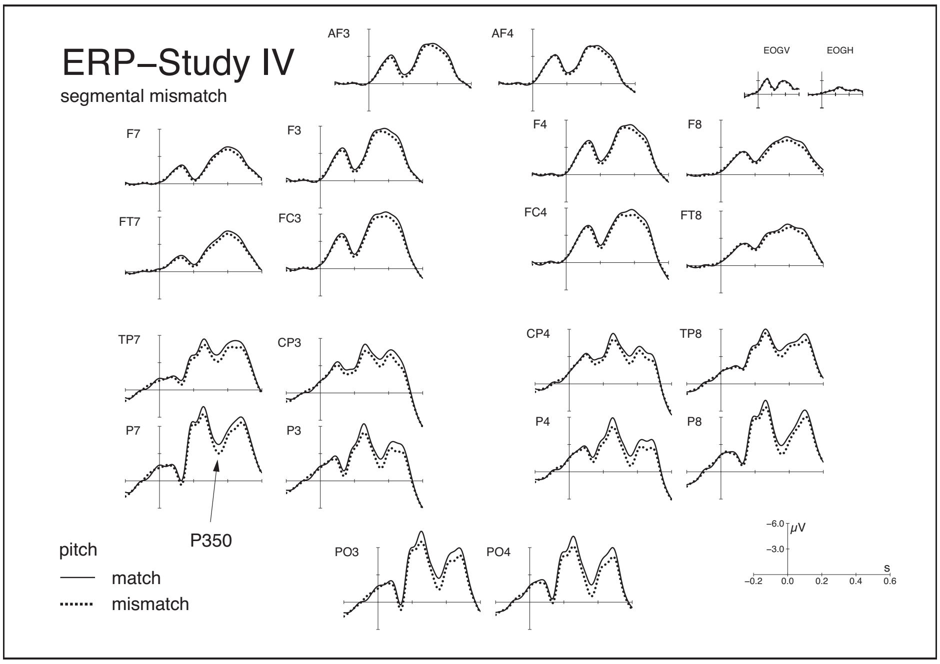

Figure 5.6: *ERPs for selected electrode sites for segmentally mismatching targets. Solid lines: targets with a stress pattern that matches the pitch of the prime, dotted lines: targets with a mismatching stress pattern.*

Note, that there were no other significant interactions with the factor *pitch congruency*, nor any interactions with the factors *remaining congruency* or *original stress of the fragment* (p > .10, compare Figures 5.7, 5.8, 5.9 and 5.10). Taken together, the results reveal that the amplitude of the positive-going P350 deflection in the ERP was significantly increased for targets that did not match the primes' segments or their pitch.

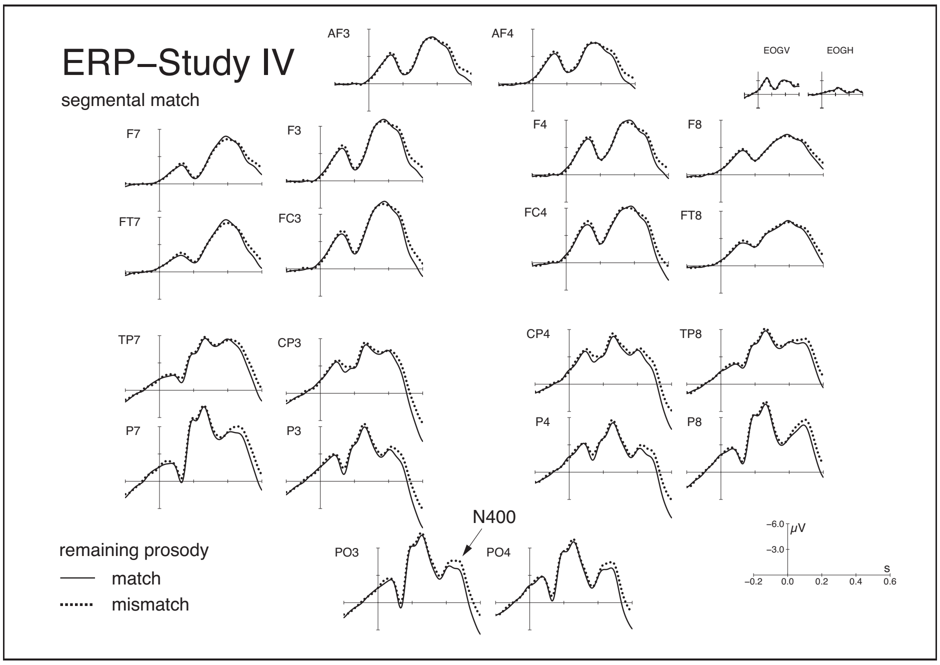

Figure 5.7: *ERPs for selected electrode sites for segmentally matching targets. Solid lines: targets with a stress pattern that matches the original prosody of the prime, dotted lines: targets with mismatching stress.*

300 to 500 ms: LFN400 A five-way ANOVA of the time window focusing on the LFN400 revealed a significant effect of the factor *segmental congruency* (F[1,23] = 5.30, p < .04), interactions of this factor with the factors *region* (F[1,23] = 19.22, p < .001] and *hemisphere* (F[1,23] = 16.18, p < .001) as well as a three-way interaction of the factors *segmental congruency*, *region* and *hemisphere* (F[1,23] = 27.59, p < .001]. Post-hoc comparisons for both frontal ROIs were conducted. The results indicated significant effects for both the anterior left ROI (F[1,23] = 47.29, p < .001) and the anterior right ROI (F[1,23] = 5.02, p < .04). Targets that matched the primes' segments elicited a LFN400 with higher amplitude (anterior left: M = -5.06 μV, SD = 1.98, anterior right: M = -4.43 μV, SD =

2.20) than mismatching targets (anterior left: M = -4.12 μV, SD = 2.22, anterior right: M = -4.07 μV, SD = 2.60, compare Figure 5.4). There was no effect of the prosodic manipulation nor any interaction with one of the factors Pitch Congruency or Remaining Congruency for the LFN400.

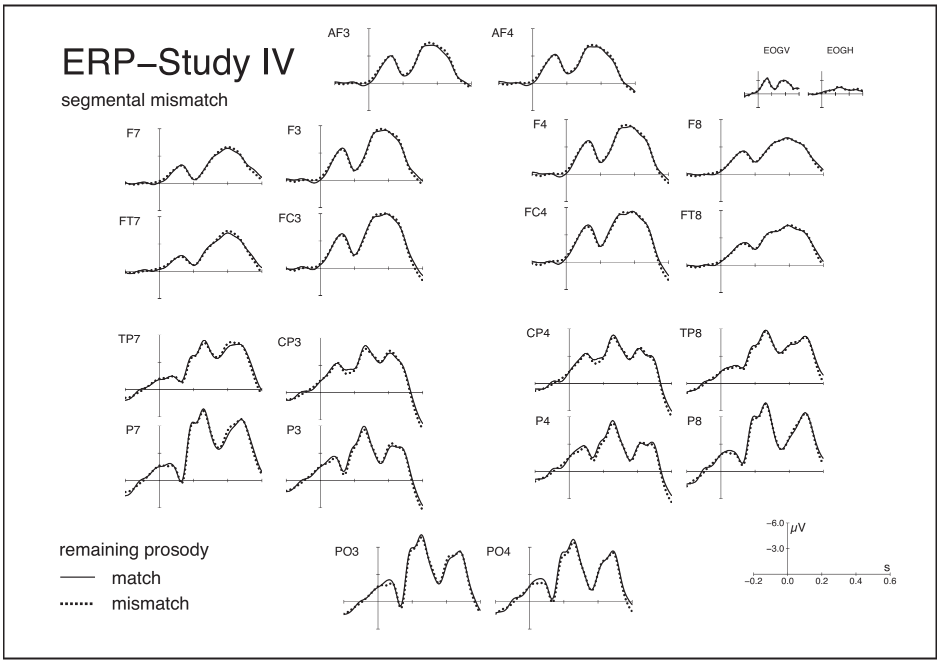

Figure 5.8: *ERPs for selected electrode sites for segmentally mismatching targets. Solid lines: targets with a stress pattern that matches the original prosody of the prime, dotted lines: targets with mismatching stress.*

400 to 600 ms: N400 A five-way ANOVA of the time window of the N400 revealed no significant main effect of the experimental factors. There was a significant two-way interaction of the factors *segmental congruency* and *region* (F[1,23] = 54.08, p $<$ .001) and a significant three-way interaction of the factors *segmental congruency*, *hemisphere* and *region* (F[1,23] = 25.71, p $<$ .001). Post-hoc comparisons for the posterior ROIs revealed significantly higher amplitudes (posterior left ROI: t[1,23] = 10.37, $p < .01$, posterior right: t[1,23] = 9.39; $p < .01$) for segmentally mismatching targets (posterior left: -3.72 μV, SD = 4.85, posterior right: -3.60 μV, SD = 5.00) as compared to segmentally matching ones (posterior left: -2.89 μV, SD = 4.80, posterior right: -2.89 μV, SD = 4.83, compare Figure 5.4).

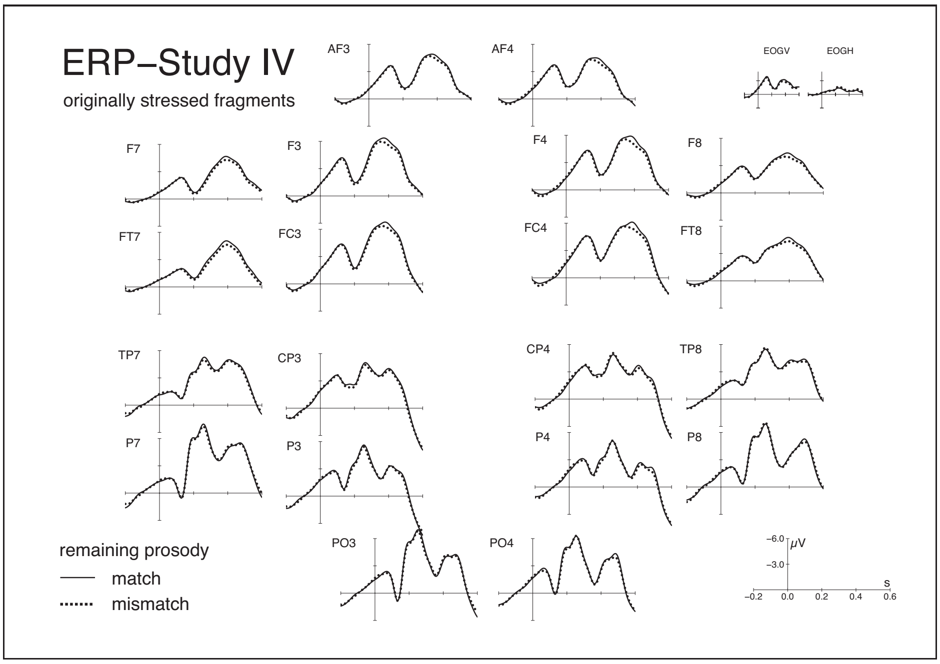

Figure 5.9: *ERPs for selected electrode sites for targets following originally stressed primes. Solid lines: initially stressed targets, dotted lines: initially unstressed targets.*

There was a two-way interaction of the factors *segmental congruency* and *remaining congruency* (F[1,23] = 4.66, $p < .05$). Post-hoc comparisons indicated that the ERPs for segmentally matching targets that did not match the remaining prosody of the prime were more negative (M = -3.32 μV, SD = 3.92) than for ones that matched the remaining prosody (M = -2.58 μV, SD = 3.80, t[1,23] = 6.33, $p < .02$, compare Figure 5.7). No effect of the factor *remaining congruency* was found for segmentally mismatching words (see Figure 5.8). Notice, that no main effect nor any interaction with the factor Pitch Congruency was found for the N400 (all $p > .10$).

Furthermore, there was a significant interaction of the factors *remaining congruency* and *original stress of the fragment* (t[23] = 12.91, $p < .01$). This interaction indicates that effects of the remaining prosody on the N400 are pronounced for target words that follow originally unstressed primes. Initially unstressed targets following originally unstressed primes show a reduced negativity (M = -2.25 μV, SD = 3.50) as compared to initially stressed targets (M = -3.22 μV, SD = 3.61, see Figure 5.10). No reliable effect of the factor *remaining prosody* was found for segmentally matching and mismatching targets following originally stressed primes (t[23] < 1, see Figure 5.9).

There was no significant effect of the factor *pitch congruency* nor any interaction with this factor in the time window of the N400.

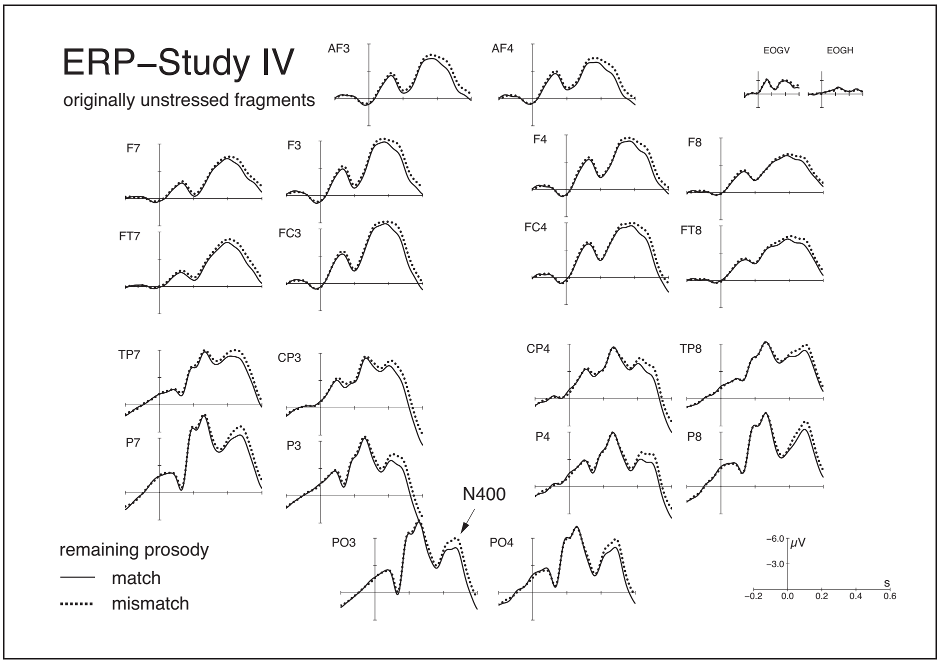

Figure 5.10: *ERPs for selected electrode sites for targets following originally unstressed primes. Solid lines: initially unstressed targets, dotted lines: initially stressed targets.*

In summary, ERPs reveal three waveforms which are sensitive to experimental manipulations in cross-modal word fragment priming. A positive-going posteriorly distributed deflection, named P350, peaks earlier than a left-lateralized frontally distributed negative-going potential, named LFN400. Both are followed by a posteriorly distributed N400 component.

Most importantly, the P350 is sensitive to a mismatch between the primes' pitch contour and the targets stress pattern (see Figures 5.5 and 5.6). Furthermore, the P350 differs in amplitude for targets that match or mismatch the primes' segments (see Figure 5.4). The LFN400, is modulated only by the primes' segmental information (see Figure 5.4). The N400 differs in amplitude for segmentally matching as compared to segmentally mismatching targets (see Figure 5.4). Furthermore, the N400 appears to be sensitive to a complex interaction of segmental and prosodic information. Segmentally matching targets with a stress pattern as indexed by the primes' remaining prosody elicit a reduced N400 as compared to segmentally matching targets with a stress pattern that does not match the original prosody of the primes (see Figure 5.7). Additionally, N400 effects are pronounced for targets that follow primes taken from originally unstressed words (see Figures 5.9 and 5.10).

5.3.4 Discussion

The present results confirm the hypothesis that the primes' pitch contour modulates responses to targets in cross-modal word fragment priming. RTs were slightly faster for targets with a stress pattern that matches the pitch contour of the primes than for targets with a stress pattern that mismatches the primes' pitch contour. The picture of the processing of prosodic information which emerges from the present data is highly complex. ERPs aid to disentangle the behavioral results.

Most importantly, the ERPs reveal new insights into cross-modal word fragment priming. The peak of the P350 temporally coincides with that of the M350 reported by Pylkkänen and colleagues (Pylkkänen, Stringfellow, Gonnerman & Marantz, 2001; Pylkkänen, Stringfellow & Marantz, 2002). In parallel to the interpretation of the M350, the P350 might reflect activation in a modality independent mental lexicon. The LFN400 peaks later than the P350. However, this component, as well as the P350 starts at 300 ms. Thus, the present data are not conclusive whether the P350 or the LFN400, or both ERP deflections reflect lexical activation. In terms of the function of pitch, this is a critical question. Only the P350 is sensitive to the primes' pitch information. If this deflection reflects lexical activation, pitch can be interpreted as a parameter that is used during the initial contact of the acoustic input with stored lexical representations.

The N400, on the other hand appears to be too late to reflect lexical activation. The time window of the N400 in the present experiment is in parallel to the N400a effect in ERP-study I. Following the interpretation of the N400a, both components might reflect the effort to select the appropriate word among activated candidates. Crucially, the primes' remaining prosody affects only that component. The functional interpretation of the ERP deflections elicited in cross-modal word fragment priming, at least that of the P350 and the N400 which were differentially sensitive to prosodic aspects of the prime is a crucial aspect of the present study. To further investigate these effects we conducted two follow-up experiments in which effects of segmental overlap were tested systematically. Both experiments are shortly described in the following paragraph which deals with effects of the segmental information. Thereafter, the present effects of the prosodic information of the primes are discussed in terms of segmental effects in cross-modal word fragment priming. Separate paragraphs focus on the pitch and on the remaining prosodic information of the primes.

Segments The behavioral results of Experiment 2 regarding the processing of segmental information of the primes replicate previous findings. Faster reactions to targets that match the primes' segments as compared to mismatching targets can be observed (Cooper, Cutler & Wales, 2002; Cutler & van Donselaar, 2001; Marslen-Wilson, 1990; Soto-Faraco, Sebastian-Galles & Cutler, 2001). The behavioral facilitation is accompanied by effects in three ERP deflections. Targets that match the primes' segments elicit a lower amplitude in the P350 and in the

N400, but a higher amplitude in the LFN400 as compared to segmentally mismatching targets. To further investigate functional correlates of all three ERP deflections we conducted two related experiments with word fragment priming (Friedrich, Kotz & Gunter, 2001). Both experiments were designed to explore which ERP deflection is related to the initial activation and which might be correlated to competition or selection of the appropriate word from the group of activated candidates.

Again we tested ERP effects for visual target words (e.g., *Amboss* or *Zebra*) preceded by prime fragments that match the initial segments of the prime or not (e.g., *am*). The amount of segmental information was varied by means of fragments length (e.g., *am*, *amb* or *ambo*). Lexical activation not necessarily depends on the amount of acoustic information. *Amboss* should be activated by *am*, *amb* as well as *ambo*. However, the competition process and the resulting activation pattern for the selection process should be sensitive to fragment length. The longer a fragment, the more effectively the activated entry that clearly matches the information of the fragment competes with simultaneously activated candidates. Thus, the selection of the appropriate word should require less effort for longer fragments than for shorter fragments. As in the present study, auditory primes were presented in one experiment. In another experiment, however, visual primes were presented, because an ERP deflection related to lexical activation in a modality-independent mental lexicon should be found for primes of the auditory modality, as well as for primes of the visual modality.

The results of both experiments were conclusive with regard to the functional interpretation of the P350 and the N400. P350 effects were elicited by primes of both modalities. In both experiments the P350 was the ERP deflection with the earliest peak at about 350 ms that was sensitive to the experimental variation. Crucially, the P350 effect did not vary with prime length. N400 effects were also elicited across both tasks. The N400 peaked at around 500 ms. In contrast to the constant P350 effect across fragments of different length, N400 effects varied with the amount of segmental information of the fragments. The longer the fragment, the stronger the reduction in the N400 amplitude for segmentally matching targets.

The findings of both follow-up studies indicate, that word-fragment priming reliably elicits P350 and N400 effects. We concluded that the P350 is most likely to reflect lexical activation, whereas N400 effects appear to reflect the competition process and the resulting effort to select the appropriate candidate among the activated entries. That is, the longer the fragments, the more precise is the activation in the mental lexicon. This activation pattern is modulated by the initial activation and the competition among the activated candidates, respectively the de-activation in the competition process (see general introduction, section 1.2.2.3). The present results are in line with the interpretation that N400 effects do not reflect automatic spreading activation (Brown & Hagoort, 1993; Brown, Hagoort & Chwilla, 2000; Chwilla, Brown & Hagoort, 1995; Holcomb, 1993). Although, the N400 is clearly influenced by the initial activation, it is also modulated by the following competition and selection process. Specifically in cross-modal word fragment priming, initial lexical activation appears to be reflected in the P350. However, a number of priming studies with word primes discussed the N400 as reflecting post-lexical integrative processing (Brown & Hagoort, 1993; Brown, Hagoort & Chwilla, 2000; Chwilla, Brown & Hagoort, 1995; Holcomb, 1993). This is unlikely to account for the present data, as targets were immediately preceded by primes. In such designs post-lexical strategic effects are not likely to occur (Neely, 1991). Furthermore, subjects were not able to improve their performance due to strategy, as initial overlap with the primes occurred for half of the target words as well as for half of the pseudowords.

LFN400 effects were elicited in the follow-up experiments only for short auditory and visual fragments. Crucially, these fragments frequently represent the first syllables of the target words. In combination with the results of the present experiment in which only syllabic primes were used, the LFN400 might be interpreted as reflecting syllable processing. Processing related to a mental syllabary is assumed, for example, by a recent model of word production (Levelt & Wheeldon, 1994). A link between the LFN400 and syllabic processing needs, of course, further investigation. However, in terms of the interpretation of the present results it appears unlikely, that the LFN400 reliably reflects lexical activation or competition.

Taken together, the following discussion of the ERP effects of the primes' pitch and their remaining prosodic features is based on the assumptions that the P350 deflection reflects lexical activation, whereas the N400 reflects the competition process respectively the activation pattern that results from competition used for the selection of the appropriate candidate. Segmental information of the primes is clearly used by all three processes. As it is supposed for this type of information it modulates the activation of the targets' lexical entries as reflected in the P350. Furthermore, segmental information of the primes enhances the power of the targets words in the competition process and, thus, facilitates the selection of the targets. Additionally, segmental information is used by processes underlying the LFN400 deflection, but the exact nature of these processes can yet not be defined.

Pitch. An important outcome of the present experiment is that effects of prosodic information of the primes are not only restricted to segmentally matching words as it has been reported so far (Cooper, Cutler & Wales, 2002; Cutler & van Donselaar, 2001; Soto-Faraco, Sebastian-Galles & Cutler, 2001). That is, for instance, the syllable *ME* with stressed pitch contour elicited slightly faster reactions and a smaller amplitude of the P350 not only for *MEdium* as compared to *mediZIN*, but also for *HEIzung* as compared to *kataLOG*. Vice versa, *ME* with unstressed pitch contour elicited faster reactions and a smaller P350 for *mediZIN* and for *kataLOG* as compared to *MEdium* and *HEIzung*. Following the argumentation developed in the previous section, the P350 effect indicates that pitch is used to activate stored lexical representations during spoken word recognition. Thus, the present results shed further light onto the ongoing debate as to whether lexical access units are feature-based or not (Dahan, Magnuson, Tanenhaus & Hogan, 2001; Norris, McQueen & Cutler, 2000; McQueen, Norris & Cutler, 1999, see section 1.2) by means of a single prosodic feature. They suggest that pitch is directly mapped onto lexical representations regardless of segmental information. That is, the acoustic information of the targets is split into phonemic and pitch information, which are both used for lexical activation.

Furthermore, the initial activation due to pitch, as indicated in the P350, appears not to be modulated by the remaining prosodic features. However, the behavioral data suggest, that pitch is not only used in a general fashion for all targets. Reaction times were fastest for targets that match the pitch and the remaining prosody. That this effect is not apparent in the N400 might be caused by the opposite polarity of the P350 and the N400. Thus, the P350 effect for pitch might have canceled out a pitch related N400 effect. Two arguments underline this proposition. First, the pitch related P350 effect appears to be smaller for segmentally matching than for segmentally mismatching targets (see Figure 5.5 and 5.6). This may well be caused by a small N400 effect for pitch congruent versus incongruent segmentally matching words. Second, behavioral facilitation due to pitch was slightly stronger for segmentally matching than for segmentally mismatching targets (see Table 5.3). This might indicate cumulative facilitation due to modulated activation and competition for the former, but not for the latter targets. Thus, the present results suggest that pitch is exploited by the listeners to constrain lexical activation. Thereafter pitch is used in a complex interactive way together with segmental information as well as duration and amplitude in the competition and selection process.

The present results confirm the assumption based on the results of ERP-study I, that pitch might be the critical factor for lexical activation. In line with the results of ERP-study III, the present results suggest that pitch is not of unique importance during spoken word recognition. Only a complex interplay of all prosodic parameters appears to significantly harm lexical access. However, the present results indicate that duration and amplitude are used later than pitch.

Remaining prosody. In contrast to pitch, which modulated behavioral and brain responses for all targets, the remaining prosody mainly affected responses for targets that match the primes' segments. Those targets brought on faster responses and elicited smaller N400 amplitudes when their stress pattern matched the primes' remaining prosody. That is, *ME* taken from *MEtro* modulated responses to *MEdium* as compared to *mediZIN*, but not for *HEIzung* as compared to *kataLOG*. Additionally, effects of the remaining prosody depended on the original stress of the fragment. Pronounced effects of duration and amplitude were found when the primes were taken from initially unstressed words. According to the interpretation deducted from the ERP effects for segmental match, the reduced N400 for targets with a stress pattern that matches the remaining prosody of the prime implicates, that duration and amplitude are exploited during the competition and selection process in spoken word recognition.

The different results for pitch and for the remaining prosody suggest that prosodic parameters are exploited in a complex fashion during spoken word recognition. Furthermore, they support the assumption that behavioral effects of cross-modal word fragment priming are not only a manifestation of the modulation of a single sub-process in spoken word recognition. Taken together the picture emerges, that pitch is used for the activation of lexical entries, whereas the remaining prosody is used to deactivate entries in the competition process.

The contribution of duration to the competition process appears plausible if one considers the function of duration as a marker for speech segmentation. Word onset syllables and word final syllables are frequently lengthened in continuous speech (Beckman & Edwards, 1990; Gow & Gordon, 1995). Segmentation of the acoustic input is assumed to be guaranteed by the competition process. During competition activated candidates that do not yield to an optimal segmentation of the input are penalized (Norris, McQueen, Cutler & Butterfield, 1997). Thus, duration might be calculated in relation to possible word boundaries in the competition process. The N400 effects found in ERP-study IV reveal, that targets following fragments from initially stressed words are not as strongly penalized in the competition process as targets following fragments from initially unstressed words (see Figures 5.9 and 5.10). These results suggest that the processing system supposes that it is very unlikely that a short syllable is the beginning of an initially stressed word. Thus, these entries are penalized in the competition process. In contrast, long syllables are handled as equally likely to be word onset regardless whether they begin an initially stressed or an initially unstressed word.

Amplitude is so far not reported to be a marker for segmentation. Furthermore, amplitude appears not to be an marker for stress in German as duration and amplitude are reported to be the most prominent parameters for stress distinctions (Dogil, 1999). The present study does not allow to disentangle the role of duration and amplitude, as both varied together within the factor *remaining prosody*. It can not be stated from the present data whether duration and amplitude contribute to the pattern of activation available to the selection process or whether only duration is of importance. Thus, further research is needed to determine the role of duration and amplitude for the behavioral facilitation and the N400 effect in cross-modal word fragment priming separately.

The results of ERP-study IV clearly favor one interpretation of the results of ERP-study I. In ERP-study I the N400a of the ERP for spoken words was interpreted as an electrophysiological correlate of selection. The question emerged whether the selection process is harmed by the violation of the pitch information for initially stressed words with incorrect prosody or by the violation of all prosodic parameters for these words. The results of ERP-study III showed that the violation of pitch alone is not an adequate presumption for the elicitation of an N400a effect. The results of the present study underline the assumption of a complex interplay of all prosodic parameters during lexical access. Thus, the N400a effect for initially stressed words in ERP-study I, which was absent for initially unstressed words with incorrect stress appears to be caused by the violation of pitch and the remaining prosody. The present results reveal, that pitch is the prosodic parameter which is initially exploited for the activation of lexical entries. Additionally, however, the results underline, that duration and possibly amplitude also serve an important function during spoken word recognition, namely for the competition process and the selection of the appropriate word candidate.

Chapter 6

Summary and Conclusions

The present set of experiments focuses on electrophysiological correlates of the processing of prosody during spoken word recognition. Results of previous behavioral research suggested that word prosody is used during lexical access by English (Cooper, Cutler & Wales, 2002), Dutch (Cutler & van Donselaar, 2001) and Spanish (Soto-Faraco, Sebastian-Galles & Cutler, 2001) listeners. To enable lexical access, the acoustic speech signal is thought to activate multiple entries stored in the listener's mental. The activated entries are thought to compete with each other for recognition. Prosodic information appears to be a variable that modulates the activation status of matching entries (see section 1.2). This empirical evidence from behavioral experiments was used to design electrophysiological experiments, that explore the time course of lexical activation and competition in spoken word recognition.

The results of a **gating task** and of **ERP-study I** reveal that incorrect stress delays subjects' behavioral responses and evokes an enhanced ERP-responses between 400 and 1000 ms, namely the **N400a**. This component was found to be enhanced for incorrectly stressed words. However, the results show that the processing of prosody is more complex than expected given previous findings from behavioral research. It appears, that the acoustic parameters - pitch, duration and amplitude - differently contribute to spoken word recognition. This is suggested by different ERP effects for different violations of the stress manipulation by the speaker. Only if pitch, duration and amplitude are violated, but not if only duration and amplitude are incorrectly applied, an N400a effect is elicited.

The results of **ERP-study II** revealed, as well as that of ERP-study I, that the extraction of the stress pattern of a word is not reflected in the ERP. This might be caused by temporal variability of the single syllables of initially stressed and initially unstressed words.

In **ERP-study III** pitch was modulated systematically by keeping duration and amplitude contours constant. The results of that study are conclusive in that pitch is extracted early from the speech signal. An early ERP component, namely the **P2**, varies in amplitude for words with stressed pitch as compared to the same words with unstressed pitch. However, although behavioral responses are delayed for words with incorrect pitch, there is no ERP deflection indicating the time course of the use of pitch information after it's early extraction. That is, an N400a effect can not be elicited if only pitch is incorrectly applied to a word.

To directly investigate the function of pitch during spoken word access processes, a cross-modal word fragment priming experiment was conducted in **ERP-study IV**. The results of the study confirm the assumption that pitch, duration and amplitude differently affect spoken word recognition. It seems that pitch is an important parameter for the initial activation of lexical entries, which is presumably reflected in the **P350**. In contrast, duration and amplitude appear to play a critical role in the competition process, which is presumably reflected in the **N400** component elicited by visual targets in cross-modal word fragment priming.

In the following I will discuss the two major aspects of the present thesis, that is, how prosodic information is processed during spoken word recognition, and how the ERP waveform elicited by spoken words might be modulated by processes engaged in spoken word recognition. The discussion is related to ERP correlates of spoken word processing as well as to ERP correlates of word fragment priming found in the present set of experiments.

6.1 ERP correlates of spoken word recognition

6.1.1 Discrimination of Prosody (P2)

One main goal of the present set of experiments was to determine the time point at which prosodic information of a word is available to the speech recognition system. A major finding in this respect was observed in ERP-study III. The results reveal that the *P2* component of the ERP waveform evoked by spoken words is sensitive to pitch processing. So far, the P2 has not been discussed to be sensitive to aspects of spoken word processing. However, following interpretations of studies which presented visual material, the P2 is considered to be sensitive to physical properties of a stimulus (see section 1.3.1). The current finding suggests that these properties might even be highly abstract parameters such as a stressed versus an unstressed pitch contour. The P2 effect indicates that neural correlates that provide pitch differentiation operate already at an early point in time starting at approximately 200 ms after word onset. According to this result it can be assumed that pitch information is available for spoken word access processes.

The P2 effect for pitch clearly contradicts the *'N325'* effect reported by Böcker et al. (1999). The 'N325' describes a surplus of negativity for a subgroup of initially unstressed words, namely weak-initial words, as compared to initial stressed words. In contrast, the P2 effect in ERP-study III is characterized as a surplus of positivity for initially unstressed words. That is, from the point of pitch processing it appears unlikely that an ERP marker for prosodic discrimination is reflected in an enhanced negative amplitude for initially unstressed words. Moreover, the results of ERP-study II indicate that the 'N325' might rather be caused by systematic timing differences between strong- and weak-initial words than by different processing of both types of words. In any case, ERP correlates for the discrimination of duration and amplitude information of differently stressed spoken words have to be the focus of further studies before final conclusions can be drawn.

6.1.2 Selection effort (N400a)

The first ERP study was guided by the assumption that incorrect prosodic information hampers lexical access. It was suggested that words with incorrect prosody activate their lexical representations to a lesser amount than than correctly stressed words. The incorrect activation pattern should result in enhanced effort to select the appropriate word among the activated candidates. Surprisingly, only a prosodic violation of initially stressed words elicited an enhanced negativity in the ERP between 400 and 600 ms post stimulus onset (N400a).

With regard to the results of ERP-study IV, the N400a can indeed be interpreted as reflecting an enhanced lexical selection effort for the incorrect versions of initially stressed words. These words are, in contrast to the incorrect versions of initially unstressed words, characterized by a modulated pitch contour next to the variation of duration and amplitude. To account for the effects observed in ERP-study I it can on the basis of ERP-study IV be inferred that an initially unstressed pitch contour, as realized for the incorrect versions of initially stressed words, let to a stronger activation of initially unstressed competitors than an initially stressed pitch contour. Thus, there was already an incorrect activation pattern at the stage of initial activation in the mental lexicon if subjects heard the incorrect versions of initially stressed words in ERP-study I.

The incorrect activation pattern for incorrect versions of initially unstressed words was extended during the competition of activated candidates. As the results of ERP-study IV show, the competition process mainly focusses on hints in the acoustic signal that indicate initially unstressed words. Duration and amplitude of incorrect versions of initially stressed words point to initially unstressed words. Thus, initially stressed entries were penalized in the competition process when subjects listened to incorrect versions of initially stressed words in ERP-study I. The hampered initial activation and the misled deactivation in the competition process resulted in a less definite activation pattern of the finally recognized words. The N400a effect appears to be a correlate of the enhanced selection effort due to this inconclusive activation pattern.

In contrast, incorrectly stressed versions of initially unstressed words did not activate a wrong set of candidates in ERP-study I. The acoustical analyses revealed that the pitch information of these words correctly does not vary from their correct counterparts. Thus, they activated initially unstressed entries stronger than initially stressed entries. Furthermore, as revealed by the results of ERP-study IV, duration and amplitude information of initially stressed words are not used as strong hints to restrict activation in the competition process. As duration and amplitude of incorrect versions of initially unstressed words point to initially stressed words, the activation of initially unstressed lexical entries was not strongly modulated during competition. Taken together, the initial activation as well as the competition process for incorrect versions of initially unstressed words in ERP-study I were not powerful to restrict activation to initially stressed words. Therefore, these words did not elicited an enhanced N400a as compared to their correct counterparts.

In light of the results of ERP-studies III and IV it can be concluded that the initial lexical activation is not reflected in the ERPs for entirely presented words. As the results of ERP-study III indicate, there is no difference in ERPs for words with incorrect pitch as compared to the same words with correct pitch. The N400a appears to be the earliest ERP correlate related to spoken word access processes. The present data suggest that the N400a reflects a process different from the P350 for pitch in cross-modal word fragment priming. In contrast to the P350, the N400a is not modulated by variation of pitch alone. As the P350 appears to reflect the initial activation, the interpretation is strengthened that the N400a as reflects the selection process, which operates with an activation pattern not only modulated by pitch. What is not completely clear from the present results is whether the N400a results from a combination of misled activation and competition or mainly from the deactivation in the competition process. However, the behavioral results of ERP-study IV can be taken as a hint for the former assumption. Lexical selection appears to be most strongly affected if pitch and the remaining prosodic information is distorted.

The interpretation of the N400a effect as reflecting selection effort is confirmed by ERP-study II as well. Weak-initial words elicited the smallest N400a amplitude as compared to words starting with full vowels. This can be interpreted as a consequence to the fact that weak-initial words represent only a small group of German words. In contrast, words in the other conditions, which start with full vowels, activate a large number of competitors beginning with the same full vowel. As weak initial words activate fewer competitors, the selection effort of the appropriate word is reduced. An important conclusion of this outcome is that ERP studies regarding the processing of spoken words have to take into account that the ERP signal is sensitive to cohort size in a time window between 400 and 600 ms.

6.1.3 Post-lexical processing of incorrect prosody (N400b)

Across the present experiments, an ERP deflection named N400b, was repeatedly associated with prosodic violation. The time window of the N400b (600 and 1000 ms) suggests, that this component is related to processes following the selection of a word, as reflected in the N400a. The N400b might be a correlate for the detection of a mismatch between a word's prosody and the stress pattern information stored for that word.

In ERP-study I a higher amplitude of the N400b was found for a prosodic violation of initially stressed words in the semantic and in the prosodic judgment and for a prosodic violation of initially unstressed words in the prosodic judgment. Thus, it appears, that the prosodic judgment might be powerful to elicit an N400b effect for prosodically violated words, even if the violation does not include pitch contour. The different results for initially stressed and unstressed words suggest that a prosodic judgment requires processes underlaying the N400b, whereas during a semantic judgment such processes are only initiated if access process are harmed and a reanalysis is necessary.

In ERP-study III only a pitch violation of initially unstressed words, but not a pitch violation of initially stressed words elicited an enhanced N400b amplitude. On the one hand, this might be caused by a stronger salience of pitch in relation to the remaining prosodic parameters as argued in the discussion of ERP-study III. In lights of the results of ERP-study II, however, the enhanced N400b for incorrect versions of initially unstressed words might result from stronger conflicts in lexical access for these words, that led to a prosodic reanalysis. As the results of ERP-study IV suggest, especially prosodic hints for initially unstressed words are calculated in the competition process. Thus, a mismatch between the activated candidates and the information used by the competition process may lead to an earlier initiation of prosodic reanalysis than incorrect pitch information for initially stressed words.

A further N400b effect was observed for initially unstressed words with full first syllables as compared to initially stressed words and weak initial words in ERP-study II. As in ERP-studies I and III, a prosodic judgment was required in that experiment. The results suggest that post-lexical prosodic processing can also be initiated for correctly stressed words. It appears that processes which compare the prosodic information of the heard word and the prosodic information stored for that word are consulted if the stress pattern reveals ambiguous information. Initially unstressed words with full vowels, differ less clear from initially stressed words than weak-initial words with reduced vowels. Thus, the prosodic judgment might have been less easy for these words as compared to weak-initial words and initially stressed words. However, further studies will have to take into account that the N400b might also be related to experimental variables such as stimulus presentation rate or the proportion of words including full and reduced vowels in unstressed position as indicated by a comparison of the results of ERP-studies I and II.

6.1.4 Post-lexical processing of incorrect prosody (late positivity)

A further ERP deflection elicited by incorrect prosodic information was observed in ERP-study III. Initially stressed words with incorrect pitch elicited an enhanced *late positivity* as compared to the same words with correct pitch. This deflection shows that prosodic reanalysis for initially stressed words with incorrect pitch occurs later than for initially unstressed words. Possible explanations for the delayed reanalysis were already discussed in the former section. On the one hand this might be caused by the weaker salience of the pitch manipulation for initially stressed words. On the other hand incorrect pitch of initially unstressed words might have been detected already during spoken word access processes and thus a prosodic reanalysis might have been initiated for the incorrect versions of initially unstressed words. The late positivity appears to reflect a later reanalysis than that reflected in the N400b. However, it appears that the processes underlying both components are involved in decision making in prosodic judgments rather than in natural spoken word recognition.

The present results clearly indicate, that incorrect prosody on the word level is processed in a highly complex fashion. They underline the need for carefully controlled material. This presumption is hard to realize with naturally spoken words. As a consequence, artificially manipulated material might be used to explore the processing of word prosody. Even if the violation of prosodic features is comparable for initially stressed and initially unstressed words, there are possible differences for both types of violations. These differences signify the complexity of processing single prosodic parameters and their interplay, which causes various effects on the behavioral results as well as for the ERPs.

6.2 ERP correlates of word fragment priming

The cross-modal fragment priming paradigm is an indirect method to investigate spoken word recognition. In this paradigm responses to visual targets preceded by fragments of spoken words are explored. Of course, the ERPs evoked by visual words differ from those evoked by auditory words. Thus, there is only limited transferability from the ERPs observed for spoken words to those observed for visual targets in cross-modal fragment priming and vice versa. Furthermore, to my knowledge ERP-study IV and the two shortly introduced follow-up experiments were the first ERP experiments on cross-modal fragment priming. Consequently, the interpretation of results remains preliminary until further research has systematically explored effects observed with this paradigm. So far it appears plausible to assume that ERPs elicited by targets in cross-modal word fragment priming can be dissected into a deflection of initial activation due to the prime (P350) and the selection effort that results from the activation pattern, which is modulated by the prime via the initial activation stage and/or the competition process (N400).

6.2.1 Initial lexical activation (P350)

The ERP signal elicited by visual words in cross-modal fragment priming is sensitive to segmental and prosodic aspects of the prime at 300 ms. The associated ERP deflection, named P350 can be related to initial activation in a modality-independent mental lexicon. It was argued, that the slightly reduced RTs for words that fit the primes' pitch together with the reduced P350 in ERP-study IV might indicate that lexical activation is modulated by pitch. This clearly reveals, that prosodic information is used for the initial activation of lexical entries. However, as there is no P350 effect for the congruency between the target words' stress pattern and the primes' duration and amplitude it can be argued that these parameters might be used for later processes than the initial lexical activation.

There was no coinciding P350 deflection in the ERPs recorded for the processing of entirely presented spoken words, revealing that word fragment priming is a more sensitive paradigm to explore this process. On the other hand the behavioral data observed for targets in cross-modal fragment priming do not clearly distinguish between the initial activation and later processes, indicating that the recording of ERPs during cross-modal word fragment priming is more sensitive to modulations in the initial activation than the behavioral data alone. Thus, the combination of behavioral and ERP results should be employed to investigate the initial activation of entries in the listeners mental lexicon. The amplitude of the P350 appears to directly reflect the targets pre-activation due to the prime. With this method, effects on the initial activation and the later competition and selection process can be disentangled.

For instance, it can be investigated whether phonetic features modulate the P350 amplitude comparable to pitch. So far there is heterogeneous empirical evidence for the role of phonetic features during lexical activation (see general introduction section 1.2). ERP effects related to a congruency in the phonetic features between primes and targets can be compared to ERP effects related to a congruency in the phoneme information. Given the present results of cross-modal word fragment priming phonetic features and phonemes might differently modulate the amplitude of the P350 and the N400. The P350 might be sensitive to both types of information, whereas the N400 might be more sensitive to phoneme information of the primes. According to the preliminary interpretation of the present results, this would suggest that phonetic features are used during the initial lexical activation. Phoneme information on the other hand might affect competition process, as reflected in the N400, and the behavioral responses stronger. These considerations denote the promising potential of the recording of ERPs during cross-modal word fragment.

6.2.2 Selection effort (N400)

In the ERPs elicited by targets in cross-modal word fragment priming the P350 is followed by an N400 component. Thus, the N400 presumably reflects later processing of the visual target. Therewith information of the primes that modulates the N400 can be assumed to be used later than information of the primes that modulates the P350. A possible explanation of the N400 in cross-modal word fragment priming, is that it is - similar to the N400a for spoken words - related to the effort needed to select the appropriate word among the activated candidates. The clearer the activation pattern favors a winner, the less selection effort is needed. Most importantly in terms of the research interest of the present thesis, prosodic features despite of pitch influence only the amplitude of the N400. It can thus be argued that duration and amplitude play a critical role during the competition process.

The N400 observed for visual targets in cross-modal word fragment priming coincides with the N400a effect observed for spoken words. Both can be interpreted as reflecting selection effort, suggesting that the selection function is shared by spoken as well as visual word recognition. An advantage of the N400 in cross-modal fragment priming, as compared to the N400a for spoken words, is that the former can be tested in a variety of designs with different primes. Of course, further research is needed, to clearly demonstrate, that the N400 in cross-modal word fragment priming indeed reflects spoken word access processes. Such studies should have designs which explicitly test the competition process. For instance, targets that lead to an impossible segmentation of the prime should affect the N400 amplitude, as they are de-activated in the competition process. In contrast, segmentally identical targets which make a phonotactically legal segmentation of the prime possible should not be penalized in the competition process and thus not affect the N400 amplitude. According to the interpretation of the present results, however, both types of targets, those leading to a possible as well as those leading to an impossible segmentation of the prime, should elicit reduced P350 amplitudes as compared to segmentally unrelated targets.

6.3 Implications for models of spoken word recognition

As repeatedly pointed out throughout the discussion, the results of the present thesis contradict the interpretation of a unified exploitation of prosodic information during spoken word recognition. They reveal that at least pitch, and the remaining prosodic features, namely duration and amplitude, might be processed differently. The behavioral as well as the electrophysiological data suggest an early lexical relevance of pitch information. The strongest support regarding this issue, comes from ERP-study IV. In this study a behavioral facilitation of a match between the prime's pitch and the target's stress pattern, and a reduced amplitude of the P350 to pitch match were found. However, also the results of ERP studies I and III can be meaningful integrated into this interpretation (see previous section).

By combining the present results, the following picture of pitch processing during spoken word recognition emerges: The initial pitch contour is extracted automatically from the acoustic input and used for the initial activation of word candidates in the mental lexicon. It appears that the pitch contour is acts independently from segmental information, that is, is activates all representations with a stress patter that matches this feature. Other suprasegmental parameters such

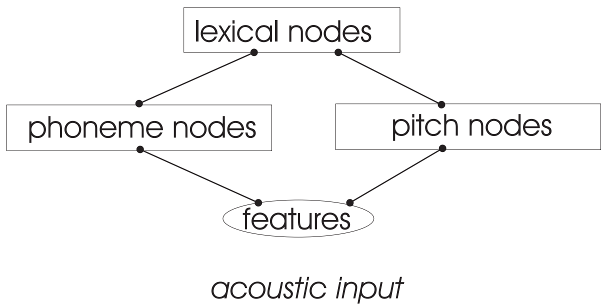

Figure 6.1: *Model of spoken word recognition with a prelexical level that codes phoneme and pitch information separately (modified from Ye and Connine, 1999)*

as amplitude and duration come into play at later processing stages, presumably during the competition process. Thus, it appears that pitch contour is a crucial parameter not just in tonal languages, but also stress languages. Of course, this assumption has to be replicated by further research.

To account for the present data, one might assume separate pitch nodes on a prelexical level, as it has been already suggested for spoken word recognition in Chinese (Ye & Connine, 1999). Figure 6.1 shows a modified version of the model provided by Ye and Connine (see general introduction section 1.2). In this model prelexical nodes code properties of the acoustic signal. According to the present results, separate pitch nodes coexist with nodes coding phonetic information also for German as a stress language. Lexical representations can be activated by phonetic nodes as well as by pitch nodes. Alternatively, one might assume that pitch directly activates lexical entries. According to this view no mediating prelexical nodes are required to activate entries in the mental lexicon (e.g. Gaskell & Marslen-Wilson, 1997, see general introduction section 1.2). It has to be focus of further studies whether pitch is used in the same global fashion also in other stress languages such as English or Dutch.

From the results of ERP-study IV it can be concluded that the remaining prosodic parameters, namely duration and amplitude, are not used during lexical activation. A match between the prime's remaining prosody and the target's stress pattern was reflected later in the ERP than pitch match. Thus, it can be concluded that these parameters are used later than pitch. Furthermore, they appear to be more closely related to the processing of segmental information than pitch, as effects were more robust for segmentally matching than for segmentally mismatching targets. To account for the present results the assumption of a lexical selection function that is based on slightly different information than provided by the initial lexical activation appears necessary. Following current models of spoken words recognition, the activation pattern which is used by the selection function is modulated by the competition process. Accordingly, duration and amplitude might modulate activation during competition. This is in line with the assumption that the competition process enables speech segmentation, as previous research revealed that duration is an important marker for word boundaries (see discussion of ERP-study IV).

Furthermore, initially stressed and initially unstressed words appear to be differently treated during competition. The results suggest that it are mainly markers for initial unstressed words which are attended to by the competition process. This might be related to the assumption that competition may take stress as a probabilistic variable into account (e.g. Cutler & Butterfield, 1992, see general introduction section 1.2 for a detailed discussion). Most of German words are initially stressed. That is, initially stressed syllables are most likely to be word onsets. Therefore activation might not be modulated when duration and amplitude signify an initially stressed syllable. In contrast, if duration and amplitude index an initially unstressed word, the competition process may have a closer look to the activated candidates and may penalize initially stressed words. This interpretation is in line with the assumption of different processing for initially stressed and initially unstressed words as it was repeatedly suggested by Mattys and Samuel (see Mattys, 1997; Mattys & Samuel, 1997; Mattys, 2000; Mattys & Samuel, 2000). However, in contrast to the interpretation of these authors, who allocate different processing for both types of words during lexical activation, the data of ERP-study IV point to different processing at the level of the competition or selection process. Nonetheless, the different results for fragments originating from initially stressed and initially unstressed words further strengthen the assumption that prosodic information is processed in a complex fashion during spoken word recognition.

An important outcome of the present set of experiments is that ERPs recorded during cross-modal word fragment priming tell us more about spoken word processing than behavioral data. They are more precise about what is going on during target processing than the behavioral data are. Interpretations relying only on behavioral data draw a different picture of the processing of word prosody than interpretations that take ERP effects into account. According to the ERPs, crucial processing, such as that of pitch information, goes on before the behavioral reaction is initialized, whereas the behavioral data clearly favor the importance of duration and amplitude over that of pitch.

Taken together, the ERP results of the present set of experiments suggest the following picture of the time course of the processing of spoken words. The extraction of the pitch contour of a word takes place 200-280 ms following word onset. The initial lexical activation of candidates in a modality-independent mental lexicon starts 300 ms after the beginning of the presentation, whereas the competition process and the selection of the appropriate candidate is reflected in the electrophysiological responses between 400 and 600 ms. Following this stages of lexical access, post-lexical prosodic reanalysis is carried out in a time window of 400 to 600 ms post stimulus onset. Of course, future studies have to be done to relate the electrophysiological response of the brain more closely to the underlying mechanisms of word recognition.

Bibliography

Attias, J. & Pratt, H. (1992). Auditory event related potentials during lexical categorization in the oddball paradigm. *Brain and Language*, *43*, 230–239.

Baayen, R. H., Piepenbrock, R. & van Rijn, H. (1995). *The CELEX lexical database (Release 2, CD-ROM)*. Nijmegen: Center for lexical information. Max-Planck-Institute for Psycholinguistics.

Barrett, S. E. & Rugg, M. D. (1989). Asymmetries in event-related potentials during rhyme-matching: Confirmation of the null effects of handedness. *Neuropsychologia*, *27*, 539–548.

Beckman, M. E. & Edwards, J. (1990). Lengthenings and shortenings and the nature of prosodic constituency. In J. K. . M. E. Beckman (Ed.), *Papers in laboratory phonology I: Between the grammar and physics of speech* (pp. 152–178). Cambridge: Cambridge University Press.

Bentin, S. & Deouell, L. Y. (2000). Structural encoding and identification in face processing: ERP evidence for separate mechanisms. *Cognitive Neuropsychology*, *17*, 35–54.

Bentin, S., Kutas, M. & Hillyard, S. A. (1995). Semantic processing and memory for attended and unattended words in dichotic listening: Behavioral and electrophysiological evidence. *Journal of Experimental Psychology: Human Perception and Performance*, *21*, 54–67.

Bentin, S., McCarthy, G. & Wood, C. C. (1985). Event-related potentials. lexical decision and semantic priming. *Electroencephalography and clinical Neurophysiology*, *60*, 343–355.

Besson, M., Faita, F., Czternasty, C. & Kutas, M. (1997). What's in a pause: Event-related potential analysis of temporal disruptions in written and spoken sentences. *Biological Psychology*, *46*, 3–23.

Böcker, B. E., Bastiaansen, M. C. M., Vroomen, J., Brunia, C. H. M. & de Gelder, B. (1999). An ERP correlate of metrical stress in spoken word recognition. *Psychophysiology*, *36*, 1–15.

Bond, R. & Small, L. H. (1983). Voicing, vowel and stress mispronunciation in continuous speech. *Perception and Psychophysics*, *34*, 470–474.

Brendel, D. & Kiefer, M. (2002). Die Rolle der Aufmerksamkeit beim unbewussten automatischen Priming. In M. Baumann, A. Klinath & J. F. Krems (Eds.), *Experimentelle Psychologie. Abstracts der 44. Tagung experimentell arbeitender Psychologen. Chemnitz, 25. - 27. März 2002* (p.9̃0). Lengerich: Pabst Science Publishers.

Brent, M. R. & Cartwright, T. A. (1996). Distributional regularity and phonotactic constraints are useful for segmentation. *Cognition*, *61*, 93–125.

Brown, C. & Hagoort, P. (1993). The processing nature of the N400: Evidence from masked priming. *Journal of Cognitive Neuroscience*, *5*, 34–44.

Brown, C. M., Hagoort, P. & Chwilla, D. (2000). An event-related brain potential analysis of visual word priming effects. *Brain and Language*, *72*, 158–190.

Chwilla, D. J., Brown, C. M. & Hagoort, P. (1995). The N400 as a function of the level of processing. *Psychophysiology*, *32*, 274–285.

Clark, J. & Yallop, C. (1995). *An introduction to Phonetics and Phonology (Second Edition)*. Oxford: Blackwell Publishers Ltd.

Cobianchi, A. & Giaquinto, S. (1997). Event-related potentials to Italian spoken words. *Electronencephalography and Clinical Neurophysiology*, *104*, 213–221.

Coninne, M. C., Clifton, C. & Cutler, A. (1987). Effects of lexical stress on phonetic categorization. *Phonetica*, *44*, 133–146.

Connine, C. M., Blasko, D., Brandt, R. & Layer, J. K. (1992). Idiomatic processing: Syntactic frozeness and subjective familiarity. *Psychological Research*, *54*, 225–232.

Connine, C. M. & Titone, D. (1996). Phoneme monitoring. *Language and Cognitive Processes*, *11*, 635–645.

Connolly, J. F. & Phillips, N. A. (1994). Event-related potential components reflect phonological and semantic processing of the terminal word of spoken sentences. *Journal of Cognitive Neuroscience*, *6*, 265–266.

Connolly, J. F., Phillips, N. A., Stewart, S. H. & Brake, W. G. (1992). Event-related potential sensitivity to acoustic and semantic properties of terminal words in sentences. *Brain and Language*, *43*, 1–18.

Connolly, J. F., Stewart, S. H. & Phillips, N. A. (1990). The effects of processing requirements on neurophysiological responses to spoken sentences. *Brain and Language*, *39*, 302–318.

Cooper, N., Cutler, A. & Wales, R. (2002). Constraints of lexical stress on lexical access in English: Evidence from native and nonnative listeners. *Language and Speech*, *45*, 207–228.

Cutler, A. (1986). Forbear is a homophone: Lexical prosody does not constrain lexical access. *Language and Speech*, *29*, 201–220.

Cutler, A. & Butterfield, S. (1992). Rhythmic cues to speech segmentation: Evidence from juncture misperception. *Journal of Memory and Language*, *31*, 218–236.

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

Cutler, A. & Chen, H. (1997). Lexical tone in Cantonese spoken-word processing. *Perception and Psychophysics*, *59*, 165–179.

Cutler, A. & Clifton, C. J. (1984). The use of prosodic information in word recognition. In H. Bouma & D. G. Bouwhuis (Eds.), *Attention and Performance X: Control of Language Processes* (pp. 183–196). Hillsdale, NJ: Erlbaum.

Cutler, A. & Clifton, C. J. (1999). Comprehending spoken language: A blueprint of the listener. In C. M. Brown & P. Hagoort (Eds.), *The Neurocognition of Language* (pp. 123–166). New York: Oxford University Press.

Cutler, A., Dahan, D. & van Donselaar, W. (1997). Prosody in the comprehension of spoken language: A literature review. *Language and Speech*, *40*, 141–201.

Cutler, A. & van Donselaar, W. D. (2001). *Voornaam* is not (really) a homophone: Lexical prosody and lexical access in Dutch. *Language and Speech*, *44*, 171–195.

Cutler, A., McQueen, J., Norris, D. & Butterfield, S. (2000). Cross-modal associative priming which disappears in sentence context. *C. U. N. Y. Conference on Human sentence Processing, March 30 - April 1 2000, La Jolla California*, *13*, 91.

Cutler, A., Mehler, J., Norris, D. & Segui, J. (1987). Phoneme identification and the lexicon. *Cognitive Psychology*, *19*, 141–177.

Cutler, A. & Norris, D. (1988). The role of strong syllables in segmentation for lexical access. *Journal of Experimental Psychology: Human Perception & Performance*, *14*, 113–121.

Cutler, A. & Otake, T. (1999). Pitch accent in spoken-word recognition in Japanese. *Journal of the Acoustic Society of America*, *105*, 1877–1888.

Dahan, D., Magnuson, J. S., Tanenhaus, M. K. & Hogan, E. M. (2001). Subcategorial mismatches and the time course of lexical access: Evidence for lexical. *Language and cognitive processes*, *16*, 507–534.

D'Arcy, R. C. N., Connolly, J. F. & Crocker, S. F. (2000). Latency shifts in the n2b component track phonological deviations in spoken words. *Clinical Neurophysiolgoy*, *111*, 40–44.

Davidson, R. J., Jackson, D. C. & Larson, C. L. (2000). Human electroencephalography. In J. T. Cacioppo, L. G. Tassinary & G. G. Bernston (Eds.), *Handbook of Psychophysiology* (pp. 27–52). Cambridge: Cambridge University Press.

Davis, M. H. D., Marslen-Wilson, W. D. & Gaskell, M. G. (2002). Leading up the lexical garden-path: Segmentation and ambiguity in spoken word recognition. *Journal of Experimental Psychology; Human Perception and Performance*, *28*, 218–244.

Deacon, D., Hewitt, S., Yang, C. & Nagata, M. (2000). Event-related potential indices of semantic priming using masked and unmasked words: Evidence that the N400 does not reflect a post-lexical process. *Cognitive Brain Research*, *9*, 137–146.

Dogil, G. (1999). The phonetic manifestation of word stress. In H. van der Hulst (Ed.), *Word prosodic systems in the languages of Europe* (pp. 273–334). New York: De Gruyter.

Donchin, E. & Coles, M. (1988). Is the P300 component a manifestation of context updating? *Behavioral and Brain Sciences*, *11*, 357–374.

Dumay, N., Benraiss, A., Barriol, B., Colin, C., Radeau, M. & Besson, M. (2001). Behavioral and electrophysiological study of phonological priming between bisyllabic spoken words. *Journal of Cognitive Neuroscience*, *13*, 121–143.

Eimer, M. (2000). Event-related potentials distinguish processing stages involved in face perception and recognition. *Clinical Neurophysiology*, *11*, 694–705.

Embick, D., Hackl, M., Schaeffer, J., Kelepir, M. & Marantz, A. (2001). A magnetoencephalographic component whose latency reflects lexical frequency. *Cognitive Brain Research*, *10*, 345–348.

Fabiani, M., Gratton, G. & Coles, G. H. (2000). Event-related brain potentials: Methods, theory and applications. In J. T. Cacioppo, L. G. Tassinary & G. G. Bernston (Eds.), *Handbook of Psychophysiology* (pp. 53–84). Cambridge: Cambridge University Press.

Fear, B. D., Cutler, A. & Butterfield, S. (1995). The strong/weak syllable distinction in English. *Journal of the Acoustic Society of America*, *97*, 1893–1904.

Federmeier, K. D. & Kutas, M. (2001). Meaning and modality: Influences of context, semantic memory organization, and perceptual predictability on picture processing. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *27*, 202–224.

Federmeier, K. D. & Kutas, M. (2002). Picture the difference: Electrophysiological investigations of picture processing in the two cerebral hemispheres. *Neuropsychologia*, *40*, 730–747.

Fery, C. (1998). German word stress in optimality theory. *Journal of Comparative Germanic Linguistics*, *2*, 101–142.

Frauenfelder, U. H. & Floccia, C. (1999). The recognition of spoken words. In A. D. Friederici (Ed.), *Language Comprehension: A biological perspective* (pp. 1–40). Heidelberg: Springer-Verlag.

Friederici, A. D. (2002). Towards a neural basis of auditory sentence processing. *Trends in Cognitive Sciences*, *6*, 78–84.

Friederici, A. D., Pfeifer, E. & Hahne, A. (1993). Event-related brain potentials during natural speech processing: Effects of semantic, morphological and syntactic violations. *Cognitive Brain Research*, *1*, 183–192.

Friedrich, C. K., Kotz, S. A. & Gunter, T. C. (2001). Event-related evidence of word fragment priming: New correlate for language processing? *Psychophysiology*, *38 (Supplement 1)*, 42.

Ganong, W. F. (1980). Phonetic categorization in auditory word perception. *Journal of Experimental Psychology: Human Perception and Performance*, *6*, 110–125.

Gaskell, G. & Marslen-Wilson, W. D. (1997). Integrating form and meaning: A distributed model of speech perception. *Language and Cognitive Processes*, *12*, 613–656.

Gaskell, G. M. & Marslen-Wilson, W. D. (2001). Lexical ambiguity resolution and spoken word recognition: Bridging the gap. *Journal of Memory and Language*, *44*, 325–349.

Goldinger, S. D. (1996). Words and voices: Episodic traces in spoken word identification and recognition memory. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *22*, 1166–1183.

Goldinger, S. D. (1998). Echoes of echoes? An episodic theory of lexical access. *Psychological Review*, *105*, 251–279.

Gow, D. W. & Gordon, P. (1995). Lexical and prelexical influence on word segmentation: Evidence from priming. *Journal of Experimental Psychology: Human Perception and Performance*, *21*, 344–359.

Grosjean, F. (1980). Spoken word recognition processes and the gating paradigm. *Perception and Psychophysics*, *28*, 267–283.

Grosjean, F. (1996). Gating. In F. Grosjean & U. H. Frauenfelder (Eds.), *A guide to spoken word recognition paradigms* (pp. 597–604). Hove, U.K.: Psychology Press Ltd.

Hagoort, P. & Brown, C. M. (2000). ERP effects of listening to speech: Semantic ERP effects. *Neuropsychologia*, *38*, 1518–1530.

Halliday, M. (1967). *Intonation and Grammar in British English*. Mouton: The Hague.

Hayashi, R., Imaizumi, S., Koichi, M., Niimi, S., Ueno, S. & Kiritani, S. (2001). Elicitation of N400m in sentence comprehension due to lexical prosody. *Neuroreport*, *12*, 1753–1756.

Herning, R. I., Jones, R. T. & Hunt, J. S. (1987). Speech event related potentials reflect linguistic content processing level. *Brain and Language*, *30*, 116–129.

Hillyard, S. A., Luck, S. J., & Mangun, G. R. (1994). The cuing of attention to visual field locations: Analysis with ERP recordings. In H.-J. Heinze, T. F. Münte & G. R. Mangun (Eds.), *Cognitive Electrophysiology* (pp. 1–25). Boston: Birkhaeuser.

Holcomb, P. J. (1993). Semantic priming and stimulus degradation: Implications for the role of the N400 in language processing. *Psychophysiology*, *30*, 47–61.

Holcomb, P. J. & Anderson, J. E. (1993). Cross-modal semantic priming: A time-course analysis using event-related brain potentials. *Language and Cognitive Processes*, *8*, 379–411.

Holcomb, P. J. & Neville, H. J. (1990). Auditory and visual semantic priming in lexical decision: A comparison using evoked potentials. *Language and Cognitive Processes*, *5*, 281–312.

Holcomb, P. J. & Neville, H. J. (1991). Natural speech processing: An analysis using event-related brain potentials. *Psychobiology*, *19*, 286–300.

Jongenburger, W. (1996). *The role of lexical stress during spoken-word processing*. Dordrecht (the Netherlands): unpublished doctoral dissertation.

Kiefer, M. (2002). The N400 is modulated by unconsciously perceived masked words: Further evidence for an automatic spreading activation account of N400 priming effects. *Cognitive Brain Research*, *13*, 27–39.

Koster, M. & Cutler, A. (1997). Segmental and suprasegmental contributions to spoken-word recognition in Dutch. *Proceedings of the 5th European Conference on Speech Communication and Technology 1997, Rhodes, Greece* (pp. 2167–2170).

Kutas, M. (1997). Views on how the electrical activity that the brain generates reflects the functions of different language structures. *Psychophysiology*, *38*, 383–398.

Kutas, M. & Federmeier, K. D. (2000). Electrophysiology reveals semantic memory use in language comprehension. *Trends in Cognitive Science*, *4*, 463–470.

Kutas, M. & Hillyard, S. A. (1980). Reading senseless sentences: Brain potentials reflect semantic incongruity. *Science*, *207*, 203–205.

Lahiri, A. & Marslen-Wilson, W. (1991). The mental representation of lexical form: A phonological approach to the recognition lexicon. *Cognition*, *38*, 254–294.

Lee, L. & Nusbaum, H. C. (1993). Processing interactions between segmental and suprasegmental information in native speakers of English and Mandarin Chinese. *Perception and Psychophysics*, *53*, 157–165.

Levelt, J. M. & Wheeldon, L. (1994). Do speakers have access to a mental syllabary?. *Cognition*, *50*, 239–269.

Lindfield, K. C., Wingfield, A. & Goodglass, H. (1999). The role of prosody in the mental lexicon. *Brain and Language*, *68*, 312–317.

Luce, P. A. & Cluff, M. S. (1998). Delayed commitment in spoken word recognition: Evidence from cross-modal priming. *Perception and Psychophysics*, *60*, 484–490.

Luce, P. A. & Piosini, D. (1998). Recognizing spoken words: The neighborhood activation model. *Ear and Hearing*, *19*, 1–36.

Marslen-Wilson, W. (1990). Activation, competition and frequency in lexical access. In G. T. M. Altmann (Ed.), *Cognitive models of speech processing. Psycholinguistic and computational perspectives* (pp. 148–172). Massachusetts: MIT Press.

Marslen-Wilson, W., Moss, H. E. & van Halen, S. (1996). Perceptual distance and competition in lexical access. *Journal of Experimental Psychology: Human Perception and Performance*, *22*, 1376–1392.

Marslen-Wilson, W. & Warren, P. (1994). Levels of perceptual representation and processes in lexical access: Words phonemes and features. *Psychological Review*, *101*, 653–675.

Marslen-Wilson, W. D. & Welsh, A. (1978). Processing interactions and lexical access during word recognition in continuous speech. *Cognitive Psychology*, *10*, 29–63.

Mattys, S. & Samuel, A. G. (2000). Implications of stress-pattern differences in spoken-word-recognition. *Journal of Memory and Language*, *42*, 571–596.

Mattys, S. L. (1997). The use of time during lexical processing and segmentation: A review. *Psychonomic Bulletin & Review*, *4*, 310–329.

Mattys, S. L. (2000). The perception of primary and secondary stress in english. *Perception and Psychophysics*, *62*, 253–265.

Mattys, S. L. & Samuel, A. (1997). How lexical stress affects speech segmentation and interactivity: Evidence from the migration paradigm. *Journal of Memory and Language*, *36*, 87–116.

McCallum, W. C., Farmer, S. F. & Pocock, P. V. (1984). The effects of physical, semantic incongruities on auditory event-related potentials. *Electroencephalography and Clinical Neurophysiology*, *59*, 477–488.

McClelland, J. L. & Elman, J. L. (1986). The TRACE model of speech perception. *Cognitive Psychology*, *18*, 1–86.

McQueen, J. & Cutler, A. (1997). Cognitive processes inn speech perception. In w. J. Hardcastle & J. Laver (Eds.), *The Handbook of Phonetic Sciences* (pp. 566–585). Oxford: Blackwell.

McQueen, J. M. (1996). Phonetic categorization. *Language and Cognitive Processes*, *11*, 655–664.

McQueen, J. M. & Cutler, A. (2001). Spoken word access processes: An introduction. *Language and Cognitive Processes*, *16*, 469–490.

McQueen, J. M., Norris, D. & Cutler, A. (1994). Competition in spoken word recognition: Spotting words in other words. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *20*, 621–638.

McQueen, J. M., Norris, D. & Cutler, A. (1999). Lexical influence in phonetic decision making: Evidence from subcategorical mismatches. *Journal of Experimental Psychology: Human Perception and Performance*, *25*, 1363–1389.

Miller, J. L. (1978). Interactions in processing segmental and suprasegmental features of speech. *Perception and Psychophysics*, *24*, 175–180.

Mimura, M., Verfaellie, M. & Milberg, W. (1997). Repetition priming in an auditory lexical decision task: Effects of lexical status. *Memory and Cognition*, *25*, 819–825.

Neely, J. H. (1991). Semantic priming effects in visual word recognition: A selective review of current findings and theories. In D. Besner & G. Humphreys (Eds.), *Basic processes in reading: Visual word recognition* (pp. 264–336). Hillsdale, NJ: Erlbaum.

Norris, D. (1994). Shortlist: A connectionist model of continuous speech recognition. *Cognition*, *52*, 189–234.

Norris, D., McQueen, J. M. & Cutler, A. (2000). Merging information in speech recognition: Feedback is never necessary. *Behavioral and Brain Sciences*, *23*, 299–370.

Norris, D., McQueen, J. M. & Cutler, A. (2002). Bias effects in facilitatory phonological priming. *Memory and Cognition*, *30*, 399–411.

Norris, D., McQueen, J. M., Cutler, A. & Butterfield, S. (1997). The possible word constraint in the segmentation of continuous speech. *Cognitive Psychology*, *34*, 191–243.

Osterhout, L., McLaughlin, J. & Bersick, M. (1997). Event-related brain potentials and human language. *Trends in Cognitive Sciences*, *1*, 203–209.

Otake, T. & Cutler, A. (1999). Perception of suprasegmental structure in non-native dialect. *Journal of Phonetics*, *27*, 229–253.

Pallier, C., Colome, A. & Sebastian-Galles, N. (2001). The influence of native-language phonology on lexical access: Exemplar based versus abstract lexical entries. *Psychological Science*, *12*, 445–449.

Patel, A. D., Gibson, E., Ratner, J., Besson, M. & Holcomb, P. J. (1998). Processing syntactic relations in language and music: An event-related potential study. *Journal of Cognitive Neuroscience*, *10*, 717–733.

Picton, T. W. & Hillyard, S. A. (1988). Endogenous event-related potentials. In T. W. Picton (Ed.), *Human Event-related Potentials. EEG Handbook*, Volume 3 (pp. 361–426). Amsterdam: Elsevier.

Pisoni, D. B. (1997). Some thoughts on 'normalization' in speech perception. In K. Johnson & J. W. Mullennix (Eds.), *Talker Variability in Speech Processing* (pp. 9–33). San Diego: Academic Press.

Pivik, R. T., Broughton, R. J., Coppola, R., Davidson, R. J., Fox, N. & Nuwer, M. R. (1993). Guidelines for the recording and quantitative analysis of electroencephalographic activity in research contexts. *Psychophysiology*, *30*, 547–558.

Polich, J., Ellerson, P. C. & Cohen, J. (1996). P300, stimulus intensity, modality and probability. *International Journal of Psychophysiology*, *23*, 55–62.

Praamstra, P., Meyer, A. S. & Levelt, W. J. M. (1994). Neurophysiological manifestations of phonological processing: Latency variation of a negative ERP component timelocked to phonological mismatch. *Journal of Cognitive Neuroscience*, *6*, 204–219.

Praamstra, P. & Stegemann, D. F. (1993). Phonological effects on the auditory N400 event-related brain potential. *Cognitive Brain Research*, *1*, 73–86.

Price, C., Indefrey, P. & van Turennout, M. (1999). The neural architecture underlying the processing of written and spoken word forms. In C. M. Brown & P. Hagoort (Eds.), *The Neurocognition of Language* (pp. 211–240). New York: Oxford University Press.

Pylkkänen, L., Stringfellow, A., Flagg, E. & Marantz, A. (2000). A neural response sensitive to repetition and phonotactic probability: MEG investigations of lexical access. *Presented at Biomag 2000, 12th International Conference on Biomagnetism, Helsinki, Finnland*.

Pylkkänen, L., Stringfellow, A., Gonnerman, L. & Marantz, A. (2001). Are inhibitory effects in lexical decision lexical or post-lexical? An MEG study. *Journal of Cognitive Neuroscience*, *Supplement*, 125.

Pylkkänen, L., Stringfellow, A. & Marantz, A. (2002). Neuromagnetic evidence for the timing of lexical activation: An MEG component sensitive to phonotactic probability but not to neighborhood density. *Brain and Language*, *81*, 666–678.

Quene, H. & Koster, M. L. (1998). Metrical segmentation in dutch: Vowel quality or stress? *Language and Speech*, *41*, 185–202.

Radeau, M., Besson, M. Fonteneau, E. & Castro, S. L. (1998). Semantic, repetition and rime priming between spoken words: Behavioral and electrophysiological evidence. *Biological Psychology*, *48*, 183–204.

Repp, B. H. & Lin, H.-B. (1990). Integration of segmental and tonal information in speech perception: A cross-linguistic study. *Journal of Phonetics*, *18*, 481–495.

Rugg, M. D. (1984a). Event related brain potentials and the phonological processing of words and non-words. *Neuropsychologia*, *22*, 435–443.

Rugg, M. D. (1984b). Event-related potentials in phonological matching tasks. *Brain and Language*, *23*, 225–240.

Rugg, M. D. & Barrett, S. E. (1987). Event-related potentials and the interaction between orthographic and phonological information in a rhyme-judgment task. *Brain and Language*, *32*, 336–361.

Schirmer, A., Kotz, S. A. & Friederici, A. D. (2001). Sex differentiates the role of emotional prosody during word processing. *Cognitive Brain Research*, *14*, 228–233.

Schreuder, R. & Baayen, R. H. (1994). Prefix stripping re-revisited. *Journal of Memory and Language*, *33*, 357–375.

Sekiguchi, T. & Nakajima, Y. (1999). The use of lexical prosody for lexical access of the Japanese language. *Journal of Psycholinguistic Research*, *28*, 439–454.

Slowiaczek, L. M. (1990). Effects of lexical stress in auditory word recognition. *Language and Speech*, *33*, 47–68.

Slowiaczek, L. M. (1991). Stress and context in auditory word recognition. *Journal of Psycholinguistic Research*, *20*, 465–481.

Soares, I., Collet, L. & Duclaux, R. (1991). Electrophysiological correlates of auditory lexical decision: An attempt to test the 'Cohort model'. *International Journal of Neuroscience*, *57*, 111–122.

Soto-Faraco, S., Sebastian-Galles, N. & Cutler, A. (2001). Segmental and suprasegmental mismatch in lexical access. *Journal of Memory and Language*, *45*, 412–432.

Spencer, K. M., Vila Abad, E. & Donchin, E. (2000). On the search for the neurophysiological manifestation of recollective experience. *Psychophysiology*, *37*, 494–506.

Steinhauer, K., Alter, K. & Friederici, A. D. (1999). Brain potentials indicate immediate use of prosodic cues in natural speech processing. *Nature Neuroscience*, *2*, 191–195.

Swinney, D. A. & Cutler, A. (1979). The access and processing of idiomatic expressions. *Journal of Verbal Learning and Behavior*, *18*, 523–534.

Titone, D. & Connine, C. M. (1994). Comprehension of idiomatic expressions: Effects of predictability and literality. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *20*, 1126–1138.

Tyler, L. K. & Wessels, J. (1983). Quantifying contextual contributions to word-recognition processes. *Perception and Psychophysics*, *34*, 409–420.

van den Brink, D., Brown, C. M. & Hagoort, P. (2001). Electrophysiological evidence for early contextual influences during spoken-word recognition: N200 versus N400 effects. *Journal of Cognitive Neuroscience*, *13*, 967–985.

van Donselaar, W., Koster, M. & Cutler, A. (submitted). Exploring the role of lexical stress in lexical recognition.

van Heuven, V. & Hagman, P. (1988). Lexical statistics and spoken word recognition in Dutch. In P. Coopmans & A. Hulk (Eds.), *Linguistics in the Netherlands* (pp. 59–68). Dordrecht: Foris.

van Heuven, V. J. (1985). Perception of stress pattern and word recognition: Recognition of dutch words with incorrect stress position. *Journal of the Acoustical Society of America*, *78*, 21.

Van Petten, C., Coulson, S., Rubin, S., Plante, E. & Parks, M. (1999). Time course of word identification and semantic integration in spoken language. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *25*, 394–471.

Van Petten, C. & Rheinfelder, H. (1995). Conceptal relationships between spoken words and enviromental sounds: Event-related brain potential measures. *Neuropsychologia*, *33*, 485–508.

Vitevitch, M. S. & Luce, P. (1998). When words compete: Levels of processing in perception of spoken words. *Psychological Science*, *9*, 325–329.

Vitevitch, M. S., Luce, P., Charles-Luce, J. & Kemmerer, D. (1997). Phonotactics and syllable stress: Implications for the processing of spoken nonsense words. *Language and Speech*, *40*, 47–62.

Vitevitch, M. S. & Luce, P. A. (1999). Probabilistic phonotactics and neighborhood activation in spoken word recognition. *Journal of Memory and Language*, *40*, 374–408.

Vroomen, J., Zon, M. v. & de Gelder, B. (1996). Cues to speech segmentation: Evidence from juncture misperception and word spotting. *Memory and Cognition*, *24*, 744–755.

Wang, Y. & Connolly, J. F. (2001a). An ERP study of the segmental and tonal processes in spoken Chinese Mandarin. *International Journal of Psychophysiology*, *41*, 229.

Wang, Y. & Connolly, J. F. (2001b). Subphonetical and tonal processes in reading Chinese Mandarin: Evidence from event-related brain potentials. *Psychophysiology*, *38 (Supplement 1)*, 98.

West, W. C. & Holcomb, P. J. (2002). Event-related potentials during discourse-level semantic integration of complex pictures. *Cognitive Brain Research*, *13*, 363–375.

Wingfield, A., Goodglass, H. & Lindfield, K. C. (1997). Word recognition from acoustic onsets and acoustic offsets: Effects of cohort size and syllabic stress. *Applied Psycholinguistics*, *18*, 85–100.

Wood, C. C. (1974). Parallel processing of auditory and phonetic information in speech discrimination. *Perception and Psychophysics*, *15*, 501–508.

Woodward, S. H., Owens, J. & Thompson, L. W. (1990). Word-to-Word variations in ERP component latencies: Spoken words. *Brain and Language*, *38*, 488–503.

Ye, Y. & Connine, C. M. (1999). Processing spoken Chinese: The role of tone information. *Language and Cognitive Processes*, *14*, 609–630.

Yip, M. C. W. (2001). Phonological priming in Cantonese spoken-word processing. *Psychologica*, *44*, 223–229.

Zwitserlood, P. (1989). The locus of the effects of sentential-semantic context in spoken-word processing. *Cognition*, *32*, 25–64.

Zwitserlood, P. (1996). Form priming. *Language and Cognitive Processes*, *11*, 589–596.

Appendix A

Materials for ERP-study I

Initially stressed words

Animated.

Ahorn (maple), Baby (baby), Bischof (bishop), Bussard (buzzard), Curry (curry), Diva (star), Fakir (fakir), Fötus (fetus), Guru (guru), Hering (herring), Kaktus (cactus), Kiwi (kiwi), Krokus (crocus), Lady (lady), Lama (lama), Leutnant (lieutenant), Mammut (mammoth), Mango (mango), Pascha (pasha), Pony (pony), Poree (leek), Rabbi (rabbi), Safran (saffron), Sultan (sultan), Tabak (tobacco), Wirsing (Savoy cabbage), Zebra (zebra)

Inanimated.

Balsam (balsam), Bottich (tub), Cola (coke), Diskus (discus), Essig (vinegar), Gulasch (goulash), Gummi (rubber), Honig (honey), Jogurt (yogurt), Käfig (cage), Kaviar (caviar), Kajak (kayak), Kanu (canoe), Kaschmir (cashmere), Kilo (kilo), Kino (cinema), Komma (comma), Kompass (compass), Konto (account), Krimi (thriller), Lasso (lasso), Lotto (lotto), Marmor (marble), Mensa (canteen), Metro (metro), Mofa (small moped), Monstrum (monster), Müsli (muesli), Opus (opus), Pasta (paste), Pfennig (cent), Pizza (pizza), Plastik (plastic), Prisma (prism), Pulli (sweater), Salto (somersault), Samba (samba), Schampoo (shampoo), Sofa (sofa), 65. Tango (tango), Taxi (taxi), Teddy (teddy), Tennis (tennis), Teppich (carpet), Ticket (ticket), Toga (toga), Tonic (tonic), Tuba (tuba), Tundra (tundra), Turban (turban), Villa (villa), Whisky (whisky), 79. Yoga (yoga)

Initially unstressed words

Animated.

Akteur (protagonist), Armee (army), Bankier (banker), Chauffeur (chauffeur), Dekan (dean), Delfin (dolphin), Fasan (pheasant), Friseur (hairdresser), Gepard (cheetah), Kakao (cocoa), Kalif (caliph), Kamel (camel), Kurier (messenger), Mandant (client), Major (major), Monarch (monarch), Monteur (fitter), Notar (notary public), Organ (organ), Patient (patient), Pilot (pilot), Pirol (oriole), Vampir (vampire), Vikar (curate)

Inanimated.

Abtei (abbey), Allee (avenue), Altar (altar), Balkon (balcony), Bankett (banquet), Basalt (basalt), Benzin (petrol), Beton (concrete), Brikett (briquette), Büro (office), Butan (butane), Choral (chant), Depot (depot), Fabrik (factory), Figur (figure), Flanell (flannel), Frisur (hairstyle), Juwel (jewel), Kamin (chimney), Kanal (canal), Karton (card), Klavier (piano), Konfekt (confectionery), Kopie (copy), Kostüm (costume), Likör (liqueur), Lokal (pub), Lotion (lotion), Magnet (magnet), Menthol (menthol), Methan (methane), Modell (model), Morast (mire), Muskat (nutmeg), Notiz (note), Orkan (hurricane), Palast (palace), Papier (paper), Parkett (parquet), Phantom (phantom), Plakat (poster), Plateau (plateau), Propan (propane), Quadrat (quad), Rasur (shave), Regal (shelves), Rubin (ruby), Signal (signal), Smaragd (emerald), Sopran (soprano), Taifun (typhoon), Titan (titanium), Trikot (shirt), Ventil (valve), Vulkan (volcano), Waggon (freight car)

Appendix B

Materials for ERP-study II

Initially stressed words with /a/ in the first syllable

Balkan (the Balkans), Balken (beam), Banner (banner), Brandung (breakers), Dampfer (damper), Drachen (dragon), Farmer (farmer), Fasching (carnival), Fassung (setting), Galgen (gallows), Gammler (layabout), Gasse (lane), Gattin (wife), Gattung (genus), Haftung (liability), Halle (hall), Hammer (hammer), Handlung (plot), Kaktus (cactus), Kante (border), Kapsel (capsule), Karte (card), Katze (cat), Klammer (peg), Klappe (flap), Krabbe (shrimp), Kratzer (scratch), Landung (landing), Lanze (lance), Laster (truck), Magma (magma), Mannschaft (team), Partner (partner), Plasma (plasma), Plastik (plastic), Ranke (tendril), Ratte (rat), Spannung (tension), Tango (tango), Tante (aunt), Wandel (change)

Initially stressed words with /ɛ/ in the first syllable

Bremse (brake), Brenner (burner), Bresche (breach), Delle (dent), Denker (thinker), Ferse (heel), Festung (fortress), Fetzen (shred), Freske (fresco), Gangster (gangster), Gerste (barley), Ghetto (ghetto), Gletscher (glacier), Häftling (prisoner), Hälfte (half), Hefter (directory), Hektar (hectare), Hektik (trouble), Hemmung (inhibition), Herrschaft (power), Kämpfer (fighter), Keller (cellar), Kelter (winepress), Kenntnis (knowledge), Kerker (dungeon), Kerze (candle), Kette (chain), Menge (amount), Merkmal (feature), Messe (mess), Messing (brass), Messung (measuring), Nenner (denominator), Nennung (naming), Pelle (skin), Pensum (workload), Perle (pearl), Renner (runner), Tänzer (dancer), Tempel (temple), Wechsel (change)

Initially unstressed words with /a/ in the first syllable

Balkon(balcony), Bankett (banquette), Barbar (barbarian), Basar (bazar),Fabrik (factory), Garant (garant), Granit (granite), Kalkül (calculation), Kamel (camel), Kamin (chimney), Kanal (canal), Kartei (card file), Kartell (cartel), Karton (card), Klavier (clavier), Magnet (magnet), Mandat (mandate), Manier (manner), Marxist (Marxist), Masseur (masseur), Natur (nature), Paket (pile), Palast (palace), Papier (paper), Parfum (parfum), Patent (patent), Partei (party), Partie (part), Pfarrei (parish), Plakat (poster), Planet (planet), Rabatt (discount), Radar (radar), Salat (salad), Saturn (Saturn), Spalier (trellis), Statut (statute), Tabu (tabu), Talent (talent), Tarif (rate), Transport (transport), Trapez (trapezium)

Initially unstressed words with /ə/ in the first syllable

Bedacht (with care), Befund (results), Beruf (profession), Begier (desire), Beginn (beginning), Belang (importance), Beschluss (decision), Bestand (continuance), Bewuchs (vegetation), Bezirk (district), Gebet (prayer), Gebot (rule), Gebrüll (bellowing), Gebühr (charge), Gebüsch (bushes), Gefährt (carriage), Gefäss (vessel), Geflecht (network), Gehalt (salary), Gehirn (brain), Gehör (hearing), Gemüt (mind), Genuss (consumption), Gepäck (luggage), Geräusch (sound), Gesang - (song), Geschick (skill), Geschlecht (gender), Geschmack (taste), Geschöpf (creature), Geschoss (projectile), Geschütz (gun), Geshwulst (growth), Gespann (yoke), Gestell (stand), Gestrüpp (undergrowth), Getier (creatures), Getränk (drink), Gewand (garment), Gewehr (rifle), Revue (revue), Velours (velours)

Words presented with incorrect stress

Originally stressed words. Basis (base), Chaos (chaos), Datum (data), Fazit (conclusion), Grafik (graphic), Habicht (goshawk), Ladung (stowage), Lava (lava), Mahnung (reminder), Nahrung (nutrition), Panik (panic), Skala (scale), Zahlung (settlement)

Originally unstressed words. Ferment (ferment), Membran (membrane), Prägnanz (conciseness), Reptil (reptile), Tendenz (tendency), Termin (deadline), Terrain (terrain), Ventil (valve), Vernunft (rationality), Verbot (proscription), Verbund (network), Verrat (treason), Verstoss (violation), Zäsur (caesura)

Appendix C

Materials for ERP-study III

Note that the words were selected from the materials presented in Appendix A.

Initially stressed words

Animated. Baby, Bischof, Bussard, Diva, Fakir, Fötus, Guru, Hering, Kaktus, Kiwi, Krokus, Lady, Lama, Leutnant, Mammut, Pascha, Mango, Rabbi, Sultan, Zebra

Inanimated. Amboss, Balsam, Cola, Gulasch, Gummi, Honig, Käfig, Kajak, Kilo, Konto, Krimmi, Lotto, Marmor, Müsli, Pfennig, Plastik, Samba, Sofa, Teppich, Whiskey

Initially unstressed words

Animated. Akteur, Bankier, Chauffeur, Dekan, Delphin, Fasan, Friseur, Gepard, Kalif, Kamel, Kurier, Major, Mandant, Monarch, Monteur, Notar, Patient, Pilot, Vampir, Vikar

Inanimated. Abtei, Bankett, Basalt, Brikett, Juwel, Kamin, Karton, Konfekt, Kopie, Kostüm, Likör, Magnet, Morast, Papier, Parkett, Quadrat, Rasur, Regal, Signal, Vulkan

Appendix D

Materials for ERP-study IV

column I: word where the prime was taken from (see Appendix A). The end of the prime is indicated by a dot. column II: segmentally matching initially stressed word; column III: segmentally matching initially unstressed word, column IV: segmentally mismatching initially stressed word, column V: segmentally mismatching initially unstressed word

Primes from initially stressed words

I	II	III	IV	V
A.HORN	AMEISE (ant)	AGENT (agent)	KOMIKER (comedian)	EKLAT (sensation)
BAL.SAM	BALTIKUM (Baltic states)	BALKON (balcony)	PUPPE (doll)	INFARKT (infarct)
BI.SCHOF	BISCHOF (bishop)	BESTECK (flatware)	ALKOHOL (alcohol)	PALAST (palace)
CO.LA	CHOLERA (cholera)	KOHLRABI (kohlrabi)	KASCHMIR (cashmere)	PARASIT (parasite)
DIS.KUS	DISKUS (discus)	DISKONT (discount)	LASSO (lasso)	TRABANT (satellite)
ES.KIMO	ESKIMO (eskimo)	ESSENZ (essence)	KIESEL (pebble)	ZÄSUR (caesura)
FA.KIR	FADEN (string)	FABRIK (factory)	HEKTAR (hectare)	KREDIT (credit)
GU.RU	GURKE (cucumber)	GOURMET (gourmet)	FASCHING (carnival)	DELPHIN (dolphin)
HE.RING	HERING (herring)	HEROIN (heroin)	KORDEL (cord)	SOLDAT (soldier)
JO.GURT	JUGEND (youth)	JURIST (jurist)	ANZAHL (number)	CHEMIE (chemistry)
KÄ.FIG	KEHRICHT	KEROSIN	MATRIX	MATROSE

I	II	III	IV	V
	(sweepings)	(kerosene)	(matrix)	(sailor)
KA.JAK	KAMERA	KAMIN	RISIKO	NOTIZ
	(camera)	(chimney)	(risk)	(note)
KAK.TUS	KAKTUS	KAKAO	PLURAL	GEBET
	(cactus)	(cocoa)	(plural)	(prayer)
KA.NU	KARIES	KARAT	TOMBOLA	VIKAR
	(caries)	(carat)	(tombola)	(curate)
KI.NO	KIMONO	KINETIK	TONIKA	KARUSELL
	(kimono)	(kinetic)	(tonic)	(carrousel)
KOM.PASS	KOMPASS	KOMMERZ	BUSSARD	FRISEUR
	(compass)	(commerce)	(buzzard)	(hairdresser)
KON.TO	KONTEXT	KONZERT	KRIMI	DATEI
	(context)	(concert)	(thriller)	(file)
LA.DY	LASTER	LABOR	KOBOLD	PRODUKT
	(truck)	(lab)	(goblin)	(product)
LO.TTO	LOTTO	LOTTERIE	ALGEBRA	FLAMENCO
	(lotto)	(lottery)	(algebra)	(flamenco)
MAN.GO	MANGOLD	MANGAN	DOGMA	KONSERVE
	(mangle)	(manganese)	(dogma)	(can)
MAR.MOR	MARMOR	MARZIPAN	LIBERO	AFFEKT
	(marble)	(marzipan)	(sweeper)	(affect)
ME.TRO	MEDIUM	MEDIZIN	HEIZUNG	KATALOG
	(medium)	(medicine)	(heating)	(catalogue)
MO.FA	MORPHIUM	MOMENT	LEUTNANT	AKTION
	(morphine)	(moment)	(lieutenant)	(action)
O.PUS	OPIUM	OPTIMIST	KRABBE	KADAVER
	(opium)	(optimist)	(shrimp)	(cadaver)
PA.SCHA	PANTHER	PATENT	LUZIFER	KALIF
	(panther)	(patent)	(Lucifer)	(calif)
PAS.TA	PASTOR	PASSANT	FISKUS	MONARCH
	(vicar)	(passer-by)	(treasury)	(monarch)
PI.ZZA	PILLE	PIRAT	LABSAL	AKTEUR
	(pill)	(pirate)	(refreshment)	(protagonist)
PO.REE	PORTO	PORTAL	MANKO	SATURN
	(postage)	(portal)	(deficit)	(Saturn)
PRIS.MA	PRISMA	PRESTIGE	KANADA	KOMPANIE
	(prism)	(prestige)	(Canada)	(company)
RA.BBI	RADIO	RABINER	TURBAN	ELEKTRA
	(radio)	(rabbi)	(turban)	(Electra)
SAF.RAN	SAFRAN	SAPHIR	KOMMA	GEPARD
	(saffron)	(sapphire)	(comma)	(cheetah)
SAL.TO	SALBEI	SALAT	KORPUS	FIGUR
	(sage)	(salad)	(corpus)	(figure)
SO.FA	SOHLE	SOPRAN	FLAMME	LAWINE
	(insole)	(soprano)	(flame)	(avalanche)
SUL.TAN	SULTAN	SULFAT	KOBRA	BARBAR
	(sultan)	(sulfate)	(cobra)	(barbarian)
TA.BAK	TABAK	TAPETE	MONITOR	INSEKT
	(tobacco)	(wallpaper)	(monitor)	(insect)

I	II	III	IV	V
TAN.GO	TANGO (tango)	TANGENTE (tangent)	ROUTE (route)	MERIDIAN (meridian)
TE.DDY	TEPPICH (carpet)	TERMIN (date)	LASER (laser)	SIGNAL (signal)
TE.NNIS	TENNIS (tennis)	TENDEN (tendency)	ZUSAGE (confirmation)	KOMET (comet)
TI.CKET	TICKET (ticket)	TITAN (titan)	PATHOS (pathos)	VERSAND (dispatch)
TU.BA	TUNDRA (tundra)	TOUPET (toupee)	WHISKY (whisky)	PROPAN (propane)

Primes from initially unstressed words

I	II	III	IV	V
AB.TEI	APFEL (apple)	ABTEI (abbey)	DOSIS (dose)	EMOTION (emotion)
BAL.KON	BALDRIAN (valerian)	BALLADE (ballad)	KÖRPER (body)	MAGNET (magnet)
BAN.KETT	BANDE (gang)	BANKETT (banquet)	ZUGABE (extra)	EMIGRANT (emigrant)
BAN.KIER	BANNER (flag)	BANKIER (banker)	ZEBRA (zebra)	OFFERTE (offer)
BA.SALT	BASIS (basis)	BASAR (bazar)	FÖTUS (fetus)	NOTAR (notary public)
BU.TAN	BUTTER (butter)	BUTAN (butane)	WILLE (will)	RASUR (shave)
DE.KAN	DEMUT (humility)	DEKOR (decor)	BALKAN (the Balkans)	KAMEL (camel)
DEL.PHIN	DELTA (delta)	DELIRIUM (delirium)	BRESCHE (breach)	VANILLE (vanilla)
FA.SAN	FAZIT (result)	FASAN (pheasant)	BROKKOLI (broccoli)	PEDAL (pedal)
FI.GUR	FIGARO (hairdresser)	FINALE (finale)	HELIUM (helium)	SMARAGD (smaragd)
FLA.NELL	FLADEN (flat cake)	FLANELL (flanell)	RETINA (retina)	KONFEKT (confectionery)
FRI.SUR	FRISCHE (freshness)	FRISUR (hairstyle)	PLENUM (plenum)	BENZOL (benzol)
JU.WEL	JUNGE (boy)	JUWEL (jewel)	KIOSK (kiosk)	PLAKAT (poster)
KA.MEL	KARPFEN (carp)	CHAOT (chaot)	HEKTAR (hectare)	FUSION (fusion)
KA.MIN	CABRIO (convertible)	KABINE (cabin)	MOBILE (mobile)	RUBIN (ruby)
KA.NAL	KAVIAR (caviar)	KANONE (gun)	BODDEN (mud-flats)	TUMULT (commotion)
KAR.TON	KARTE (card)	KARTEI (card index)	SHAMPOO (shampoo)	METHAN (methane)
KON.FEKT	KONTO (account)	KONZIL (council)	SALTO (somersault)	LAGUNE (lagoon)
KO.PIE	KOBALT (cobalt)	KOKAIN (cocaine)	PASCHA (pasha)	ALLEE (avenue)
KOS.TUEM	KOSMOS (cosmoa)	KOSTÜM (costume)	ASIEN (Asia)	GARANT (guarantor)
KU.RIER	KORDEL (cord)	KOLLEGE (colleague)	NEKTAR (nectar)	TURNIER (tournament)
LI.KOER	LIBIDO (libido)	LIKÖR (liqueur)	DOMINO (domino)	REPTIL (reptile)
LO.KAL	LOTUS (lotus)	LOKAL (pub)	MENSA (canteen)	ALTAR (altar)

I	II	III	IV	V
MAG.NET	MACKE (quirk)	MAGIE (magic)	THEKE (bar)	DEKAN (dean)
ME.THAN	METRO (metro)	METHODE (method)	TIGER (tiger)	ISLAM (Islam)
MO.DELL	MODEM (modem)	MOTIVATION (motiv)	KAJAK (kayak)	BERUF (occupation)
MON.TEUR	MONSTRUM (monster)	MONTEUR (fitter)	CHARISMA (charisma)	COURAGE (courage)
OR.GAN	ORGIE (orgy)	ORGAN (organ)	MÜSLI (muesli)	DEBUT (debut)
OR.KAN	ORDNER (directory)	ORKAN (hurricane)	GHETTO (ghetto)	ENZYM (enzyme)
PA.LAST	PAVILLON (pavilion)	PATRIOT (patriot)	KLINIKUM (clinic)	ROSINE (raisin)
PA.PIER	PAPRIKA (paprika)	PAPIER (paper)	INSASSE (passenger)	SERIE (series)
PLA.TEAU	PLASTE (plastic)	PLANTAGE (plantation)	KÜRBIS (pumpkin)	SELLERIE (celeriac)
PRO.PAN	PROBE (test)	PROLET (pleb)	PANAMA (Panama)	KAROSSE (limousine)
QUA.DRAT	QUADER (cuboid)	QUADRAT (quad)	PENSUM (brigade)	BRIGADE (pensum)
RA.SUR	RADLER (biker)	RADAR (radar)	SEELE (soul)	KURIER (messenger)
RE.GAL	REGEL (rule)	REGAL (shelves)	GATTIN (wife)	CHORAL (chant)
SIG.NAL	SINGULAR (singular)	SIGNATUR (signature)	FAKIR (fakir)	DEFEKT (defect)
TRI.KOT	TRAINGEL (triangle)	TRIKOT (shirt)	BRANDUNG (surf)	DEPOT (depot)
VAM.PIR	WAMPE (paunch)	VAMPIR (vampire)	BRECHUNG (breaking)	GENIE (genius)
WA.GGON	VAKUUM (vacuum)	WAGGON (freight car)	LUXUS (luxury)	PRIMAT (priority)

Curriculum Vitae

Claudia Katrin Friedrich

25.05.1974	geboren in Berlin
1992	Abitur am Käthe-Kollwitz-Gymnasium, Berlin
1993 - 1999	Studium der Psychologie an der Technischen Universität (TU) Berlin
1995	Vordiplom in Psychologie
1996	Forschungspraktikum an der TU Berlin, Abteilung Physiologische Psychologie bei Prof. Dr. Gisela Erdmann
1996 - 1998	Wissenschaftliche Hilfskraft mit Lehraufgaben an der TU Berlin, Abteilung Physiologische Psychologie bei Prof. Dr. Gisela Erdmann
1998	Forschungspraktikum an der TU Dresden, Institut für Klinische, Diagnostische und Differentielle Psychologie bei Prof. Dr. Burckhard Brocke
1999	Diplom in Psychologie
seit April 1999	Doktorandin am Max-Planck-Institut für neuropsychologische Forschung, Leipzig, Abteilung Neuropsychologie bei Prof. Dr. Angela Friederici

MPI Series in Cognitive Neuroscience

1 **Anja Hahne**
Charakteristika syntaktischer und semantischer Prozesse bei der auditiven Sprachverarbeitung: Evidenz aus ereigniskorrelierten Potentialstudien

2 **Ricarda Schubotz**
Erinnern kurzer Zeitdauern: Behaviorale und neurophysiologische Korrelate einer Arbeitsgedächtnisfunktion

3 **Volker Bosch** Das Halten von Information im Arbeitsgedächtnis: Dissoziationen langsamer corticaler Potentiale

4 **Jorge Jovicich**
An investigation of the use of Gradient- and Spin-Echo (GRASE) imaging for functional MRI of the human brain

5 **Rosemary C. Dymond**
Spatial Specificity and Temporal Accuracy in Functional Magnetic Resonance Investigations

6 **Stefan Zysset**
Eine experimentalpsychologische Studie zu Gedächtnisabrufprozessen unter Verwendung der funktionellen Magnetresonanztomographie

7 **Ulrich Hartmann**
Ein mechanisches Finite-Elemente-Modell des menschlichen Kopfes

8 **Bertram Opitz**
Funktionelle Neuroanatomie der Verarbeitung einfacher und komplexer akustischer Reize: Integration haemodynamischer und elektrophysiologischer Maße

9 **Gisela Müller-Plath**
Formale Modellierung visueller Suchstrategien mit Anwendungen bei der Lokalisation von Hirnfunktionen und in der Diagnostik von Aufmerksamkeitsstörungen

10 **Thomas Jacobsen**
Characteristics of processing morphological structural and inherent case in language comprehension

11 **Stefan Kölsch**
Brain and Music. A contribution to the investigation of central auditory processing with a new electrophysiological approach

12 **Stefan Frisch**
Verb-Argument-Struktur, Kasus und thematische Interpretation beim Sprachverstehen

13 **Markus Ullsperger**
The role of retrieval inhibition in directed forgetting - an event-related brain potential analysis

14 **Martin Koch**
Measurement of the Self-Diffusion Tensor of Water in the Human Brain

15 **Axel Hutt**
Methoden zur Untersuchung der Dynamik raumzeitlicher Signale

16 **Frithjof Kruggel**
Detektion und Quantifizierung von Hirnaktivität mit der funktionellen Magnetresonanztomographie

17 **Anja Dove**
Lokalisierung an internen Kontrollprozessen beteiligter Hirngebiete mithilfe des Aufgabenwechselparadigmas und der ereigniskorrelierten funktionellen Magnetresonanztomographie

18 **Karsten Steinhauer**
Hirnphysiologische Korrelate prosodischer Satzverarbeitung bei gesprochener und geschriebener Sprache

19 **Silke Urban**
Verbinformationen im Satzverstehen

20 **Katja Werheid**
Implizites Sequenzlernen bei Morbus Parkinson

21 **Doreen Nessler**
Is it Memory or Illusion? Electrophysiological Characteristics of True and False Recognition

22 **Christoph Herrmann**
Die Bedeutung von 40-Hz-Oszillationen für kognitive Prozesse

23 **Christian Fiebach**
Working Memory and Syntax during Sentence Processing. A neurocognitive investigation with event-related brain potentials and functional magnetic resonance imaging

24 **Grit Hein**
Lokalisation von Doppelaufgabendefiziten bei gesunden älteren Personen und neurologischen Patienten

25 **Monica de Filippis**
Die visuelle Verarbeitung unbeachteter Wörter. Ein elektrophysiologischer Ansatz.

26 **Ulrich Müller**
Die katecholaminerge Modulation präfrontaler kognitiver Funktionen beim Menschen

27 **Kristina Uhl**
Kontrollfunktion des Arbeitsgedächtnisses über interferierende Information

28 **Ina Bornkessel**
The Argument Dependency Model: A Neurocognitive Approach to Incremental Interpretation

29 **Sonja Lattner**
Neurophysiologische Untersuchungen zur auditorischen Verarbeitung von Stimminformationen

30 **Christin Grünewald**
Die Rolle motorischer Schemata bei der Objektrepräsentation: Untersuchungen mit funktioneller Magnetresonanztomographie

31 **Annett Schirmer**
Emotional Speech Perception: Electrophysiological Insights into the Processing of Emotional Prosody and Word Valence in Men and Women

32 **André J. Szameitat**
Die Funktionalität des lateral-präfrontalen Cortex für die Verarbeitung von Doppelaufgaben

33 **Susanne Wagner**
Verbales Arbeitsgedächtnis und die Verarbeitung ambiger Wörter in Wort- und Satzkontexten

34 **Sophie Manthey**
Hirn und Handlung: Untersuchung der Handlungsrepräsentation im ventralen prämotorischen Cortex mit Hilfe der funktionellen Magnet-Resonanz-Tomographie

35 **Stefan Heim**
Towards a Common Neural Network Model of Language Production and Comprehension: fMRI Evidence for the Processing of Phonological and Syntactic Information in Single Words