

## An ERP correlate of metrical stress in spoken word recognition

KOEN B.E. BÖCKER,<sup>a,b</sup> MARCEL C.M. BASTIAANSEN,<sup>a</sup> JEAN VROOMEN,<sup>a</sup>  
CORNELIS H.M. BRUNIA,<sup>a</sup> AND BEATRICE DE GELDER<sup>a,b</sup>

<sup>a</sup>Psychonomics Section, Tilburg University, Tilburg, The Netherlands

<sup>b</sup>Experimental Psychology Department, Université Libre de Bruxelles, Brussels, Belgium

### Abstract

Rhythmic properties of spoken language such as metrical stress, that is, the alternation of strong and weak syllables, are important in speech recognition of stress-timed languages such as Dutch and English. Nineteen subjects listened passively to or discriminated actively between sequences of bisyllabic Dutch words, which started with either a weak or a strong syllable. Weak-initial words, which constitute 12% of the Dutch lexicon, evoked more negativity than strong-initial words in the interval between P2 and N400 components of the auditory event-related potential. This negativity was denoted as N325. The N325 was larger during stress discrimination than during passive listening. N325 was also larger when a weak-initial word followed a sequence of strong-initial words than when it followed words with the same stress pattern. The latter difference was larger for listeners who performed well on stress discrimination. It was concluded that the N325 is probably a manifestation of the extraction of metrical stress from the acoustic signal and its transformation into task requirements.

**Descriptors:** Speech recognition, Metrical stress, Rhythm, Prosody, Auditory event-related potentials

In understanding speech, the task of the listener is to transform the auditory sensory input into a meaningful representation of the spoken message. Even though the adult mental lexicon contains about 17,000 entries, or lemmas (according to a conservative estimate; D'Anna, Zechmeister, & Hall, 1991), this transformation is accomplished in a few hundreds of milliseconds (Grosjean, 1980). This task is further complicated because, unlike its written counterpart, spoken language contains no simple and reliable correlates of word boundaries, which would be equivalent to blank spaces in written language (e.g., Cutler & Norris, 1988). So the question is, how does the listener segment the continuous speech stream to isolate word candidates?

Recently, much research has been devoted to understanding how such segmentation might proceed (Cutler & Butterfield, 1992; Cutler & Norris, 1988; Vroomen & de Gelder, 1995, 1997; Vroomen, Van Zon, & de Gelder, 1996). In these studies the segmentation

process is presumed to be prelexical and part of the word recognition process itself, and not its consequence (see the TRACE model by McClelland & Elman, 1986, for an example of the latter). Rhythmic properties of the language have been found to be an important basis for segmentation. English and Dutch are so-called stress-timed languages. They distinguish between strong and weak syllables. Strong and weak syllables contain full and reduced vowels, respectively. In Dutch the reduced vowel is always a schwa (/ə/), such as the second vowel in "DANgər"<sup>1</sup> or in the Dutch word "BEzəm" (broom). In English there is a small proportion of weak syllables with other reduced vowels. The alternating sequence of weak and strong syllables lends English and Dutch their rhythmic character. Analysis of the English and Dutch CELEX (1990) lexicons showed that 81% and 88% of the entries in those lexicons, respectively, starts with a strong syllable (Vroomen & de Gelder, 1995). Taking word frequency into account, Cutler and Carter (1987) estimated that in everyday English speech 85–90% of the spoken words starts with a strong syllable. Cutler and Norris (1988) advanced the notion that listeners exploit this characteristic by relying on a metrical segmentation strategy. In essence, lexical access is attempted at every strong syllable according to this strategy. Because of the preponderance of strong initial syllables in English and Dutch, lexical access will be successful in most of the cases.

The results of psycholinguistic studies employing juncture misperception (Cutler & Butterfield, 1992; Vroomen et al., 1996) and

K.B.E. Böcker was supported by the Ministry of Education of the Belgian French-speaking Community, Concerted Research Action "Language processing in different modalities: Comparative approaches." At the time of preparation of the manuscript, K.B.E. Böcker was supported by the Dutch Organization for Scientific Research (NWO; grant PPS-95-027), and M.C.M. Bastiaansen by the Cooperation Centre Tilburg and Eindhoven Universities (SOBU; grant 1995-AD).

We thank Theo Popelier, for stimulus preparation, and Geert van Boxtel, Jyrki Tuomainen, Mireille Besson, and two anonymous reviewers, for their valuable comments on earlier versions of the manuscript.

Address reprint requests to: Koen Böcker, Psychopharmacology Department, Utrecht University, PO Box 80082, 3508 TB Utrecht, The Netherlands. E-mail: K.B.E.Böcker@pharm.uu.nl.

<sup>1</sup>Strong syllables are in capitals.

word spotting paradigms (Cutler & Norris, 1988; Vroomen & de Gelder, 1995) support the conclusion that English and Dutch listeners use metrical segmentation strategy. One type of juncture misperception that is predicted from this strategy is the insertion of a word boundary preceding a strong noninitial syllable. The other type entails the deletion of a word boundary preceding a weak-initial syllable, which is collated to the preceding ones. An example taken from Cutler and Butterfield (1992) is that "conDUCT asCENTS . . ." presented to the subject embedded in noise, was reported as "the DOCTOR SENDS. . ." This example shows two insertions of a word boundary (i.e., a disjunction) preceding strong syllables and the deletion of one (i.e., a junction) preceding a weak syllable. The typical results of word-spotting studies (taken from Cutler & Norris, 1988) is that "mint" is detected faster in "MIN-təsh" than in "MIN-TAYVE." According to the metrical segmentation strategy, this result stems from the fact that "mint" has to be reassembled across the segmentation point, which results from the second strong syllable in "MIN-TAYVE."

Rhythm not only plays an important role in word recognition, but presumably the rhythm of a language is acquired prior to, and is used for segmentation during, the acquisition of the lexicon (see Mehler, Bertocini, Dupoux, & Pallier, 1994 for a review). An example of data that support this view is that American 9-month-old infants listen significantly longer to words starting with a strong compared to a weak syllable (cf., the metrical segmentation strategy; Jusczyk, Cutler, & Redanz, 1993).

The present paper reports on an event-related potential (ERP) correlate of a rhythmic property of Dutch and other stress-timed languages, that is, metrical stress. By recording such an ERP correlate in future psycholinguistic research, it should be possible to study the temporal dynamics of the processing of rhythm during speech recognition. A particular advantage of using ERPs in such research is that recording ERPs does not necessarily require an extraneous task (Kutas & Van Petten, 1994).

The present study used an oddball-like paradigm and capitalized on the brain as a deviance detector (Donchin, Spencer, & Dien, 1997), to assess if ERPs reflect its capacity to discriminate stress patterns and, if so, at what latency. In one task condition, passive listening, native Dutch speakers listened to sequences of four bisyllabic Dutch nouns in which the last item either had the same or a different metrical stress pattern as the first three items. In the second task, subjects discriminated actively between sequences with same and different metrical stress patterns. This task required that metrical stress information be processed explicitly.

We expected to find three types of ERP components to be related to differences in stress pattern. First, the early exogenous components were expected to occur in all conditions because metrical stress is realized in the physical characteristics of the speech signal (Fear, Cutler, & Butterfield, 1995; Appendix B). The second and more interesting type of components, we expect, are endogenous modulations related to metrical stress. In our paradigm the presence of an endogenous stress effect could be signaled by three possible interactions of metrical stress with other independent variables. Because metrical stress is relevant in performing the stress discrimination task, possible ERP differences between words starting with a weak syllable and words starting with a strong syllable might be enhanced during this discrimination task versus the passive listening task. Given the oddball-like paradigm, such ERP differences might also depend on the congruence of the sequence, that is, whether the final word has the same or a different stress pattern as the former ones. And finally the ERP differences might correlate with the sensitivity of the subject ( $d'$ ) in the stress dis-

crimination task. The third type of components we expect to find are those that show endogenous modulations that do not interact with metrical stress, but that do vary with task condition and with the congruence of stress patterns in a sequence. An example of the latter are the mismatch negativity (MMN; Näätänen, Gaillard, & Mäntysalo, 1978) or N2b (Näätänen & Gaillard, 1983), in the passive and active task conditions, respectively. Other components that have been recorded with spoken (and/or written) word stimuli are the (early) left anterior negativity (LAN; Friederici & Mecklinger, 1996; King & Kutas, 1995; Müller, King, & Kutas, 1997; Neville, Nicol, Barss, Foster, & Garret, 1991), the phonological mismatch negativity (PMN; Connolly & Phillips, 1994), the lexical processing negativity (LPN; Kutas, 1997), the N400 (Kutas & Hillyard, 1980), and the late positive complex or component (LPC), which usually follows a sequence of words (Curran, Tucker, Kutas, & Posner, 1993; Kutas & Hillyard, 1980; Pritchard, Shapell, & Brandt, 1991). Whether these components are affected by rhythm or metrical stress is at present unknown.

We are particularly interested in the latency of a stress-sensitive component that shows endogenous modulations. Lexical activation for words presented in isolation in the auditory modality is estimated to start after the presentation of one or two phonemes, that is, after about 150 ms (Tyler & Wessels, 1983). The bottom-up selection of a lexical candidate for words presented in isolation would be finished after about 300 ms (Grosjean, 1980). The latency of lexical integration, the process by which the selected lexical entry is integrated with its context, can be estimated at about 400 ms, that is, the peak latency of the N400, which has been regarded as a manifestation of postlexical integration (e.g., Chwilla, 1996; Chwilla, Brown, & Hagoort, 1995; Connolly, Byrne & Dyan, 1995; Holcomb, 1993; Van Petten & Kutas, 1990). If the latency of endogenous modulations of a stress-sensitive component is between 150 and 300 ms, this component probably reflects processes at the interface of the automated representation of the acoustic speech signal (the exogenous components) and controlled processes such as lexical integration. This interface could include those processes that are involved in prelexical speech segmentation on the basis of rhythmic qualities such as metrical stress, as suggested by the metrical segmentation strategy (Cutler & Norris, 1988).

## Methods

### Subjects

Nineteen right-handed subjects (8 men and 11 women) aged 19–40 years ( $M = 24.7$  years) participated in the experiment. All were native Dutch speakers. Most of the subjects were undergraduate students. They were paid a small fee (Dfl. 7.50 per hour).

### Stimulus Material

A set of 20 pairs of words (see Appendix A) was selected from the automated lexicon of the Dutch Centre for Lexical Information (CELEX, 1990). They were monomorphemic, bisyllabic Dutch nouns in which each syllable started with a consonant or a consonant cluster. Each pair contained one item with a StrongWeak metrical stress pattern, and one item with a WeakStrong pattern. Because of their scarcity WeakStrong words were selected first. Within a pair the item with the StrongWeak pattern matched the WeakStrong stress pattern on initial consonant and (within limits) on word frequency. A digital recording (sampled at 22 kHz with 16-bit resolution) was made of the word lists as produced by a male native Dutch speaker. The stimulus length varied between

619 and 885 ms ( $M = 768 \pm 69$  ms). The prosodic features of the stimuli and their analyses are presented in Appendix B.

### Design and Procedure

Subjects were seated in a comfortable chair in a dimly illuminated, electrically shielded and sound-attenuating chamber. A response button was mounted on the right arm of the chair, under the right index finger.

A trial started with an auditory warning signal (500 Hz, 100 ms, 70 dB[A]). At the same time, a fixation LED in the middle of the subject's visual field was illuminated for 100 ms, to help the subjects minimize their eye movements. After 1,500 ms, a sequence of four words was presented. The stimulus onset asynchrony for subsequent words was also 1,500 ms. Because the average word length was 768 ms this asynchrony resulted in about 750-ms pauses between words. The word sequences were played at a comfortable listening level (approximately 75 dB[A]). All auditory stimuli were presented through a loudspeaker, mounted 1.50 m in front of the subject and 1.80 m above the floor. Finally, 2,500 ms after the onset of the last word the fixation LED was illuminated once again for 2 s, indicating to the subjects that a response could be made. The delayed response was chosen to prevent motor preparation processes to interfere with the task, and the stimulus related ERPs. The intertrial interval ranged from 3 to 5 s. Stimuli were presented by a PC with timer-counter, AD conversion, and sound cards, which also recorded the responses.

The first three words in a sequence all had the same stress pattern, that is, either StrongWeak or WeakStrong. The fourth word could have either the same (congruent) or the opposite (incongruent) stress pattern. In this way four types of sequences were created that differed in congruence and stress pattern (see Table 1). Each word was presented to each subject equally often in each position in the sequence.<sup>2</sup> Furthermore, congruent and incongruent sequences had an equal probability of occurrence. Different pseudorandom sequences of four words were presented in a different pseudorandom order to each subject, under the condition that a word never occurred twice in the same or two subsequent sequences.

A complete recording session lasted approximately 6 hr, including 2 hr for electrode montage and familiarization of the subject with the recording procedure. During the remaining time six experimental blocks of about 30 min were presented, which were followed by 5–10-min breaks and a longer lunch break, according to the needs of the individual subjects. Subjects performed two tasks. The first task was a passive listening task in which the subjects were instructed to listen carefully to the words presented to them. This first task comprised two blocks of 120 trials each. The second task was a stress discrimination task, which comprised four blocks of 120 trials each. This second task was a delayed Go-NoGo task in which the subjects were asked to discriminate between sequences of congruent and incongruent stress patterns. In one version subjects were asked to press the response button when the stress pattern of the final word was congruent with that of the first three, in the other version a response had to be given when the stress pattern of the final word was incongruent with that of the first three. The two versions of the task were presented in an alternating order to each subject. Half of the subjects began with one version and the other half with the other version, respectively.

<sup>2</sup>This design made it possible to study position effects (especially between final and nonfinal words) not contaminated by repetition effects (which generally cause more positive ERPs; Rugg, 1985) nor by averaging over different words at each within-sequence position.

**Table 1.** Examples of the Four Possible Stimulus Sequences

Condition	Word 1	Word 2	Word 3	Word 4	Congruence
1	SW bengəl	SW vogəl	SW mantəl	SW fran̩jə	Congruent
2	WS cəment	WS frəgat	WS gəlei	WS pləz̩ier	Congruent
3	WS kəbab	WS gəmeen	WS brətəl	SW venstər	Incongruent
4	SW sektə	SW gestə	SW sabəl	WS gəniəp	Incongruent

Note: SW = StrongWeak; WS = WeakStrong stress pattern.

The task conditions, however, were presented in a fixed order, starting with the passive listening condition, to prevent carryover effects from the discrimination to the passive condition.

### ERP Recordings

For the electroencephalogram (EEG) recordings, nonpolarizing Beckman 8 mm Ag-AgCl electrodes were affixed to the scalp at positions F3, F4, T3, C3, Cz, C4, T4, T5, P3, P4, and T6 of the international 10-20 system using Grass EC2 electrode paste. Electrode impedance was kept below 5 k $\Omega$ . Linked mastoids served as reference. The electrooculogram (EOG) was recorded using nonpolarizing Beckman 2.1 mm Ag-AgCl electrodes, in three bipolar derivations: one vertical EOG for each eye and a horizontal EOG across both eyes.

EEG and EOG signals were amplified by Nihon Kohden (4314F) amplifiers with a 30-s time constant and a high-frequency cut-off (−3 dB) of 70 Hz. Later, because of a predominance of alpha activity (which may have been a consequence of the long duration of the recording sessions), the data were digitally lowpass filtered at 5.97 Hz (FIR filter length: 63 samples; transition band 4.5–9.5 Hz; see, e.g., Connolly, Stewart, & Phillips, 1990, for similar lowpass filtering). EEG and EOG signals were digitized online using a 12-bit AD-converter (Keithly, DAS1600) with a sampling frequency of 256 Hz, for an 8,000-ms epoch starting at the warning signal and ending 2,000 ms after the onset of the last word. The signals were stored on a PC and written out on paper by the Nihon Kohden electroencephalograph for real-time visual inspection.

### Data Reduction and Analysis

**Behavioral data.** Behavioral data, in the form of percentages of errors, were available for the discrimination task only. Trials requiring a delayed response were considered correct if the response was executed in the response interval, that is, between 2,500 and 4,500 ms after the onset of the last word.

Behavioral data were averaged over experimental blocks, and a repeated-measures analysis of variance (ANOVA) was performed using the module P4V of the BMDP statistical package. Factors were congruence (same, different), stress pattern of the final word (StrongWeak, WeakStrong), and response status (Response Required, No Response Required). Note that only the latter trials (not requiring a response) were included in the ERP (see below). Finally, to examine the effect of discrimination accuracy on ERP amplitudes the signal detection measure of sensitivity ( $d'$ ) and bias ( $\beta$ ) were calculated from the percentage of errors on noise (con-

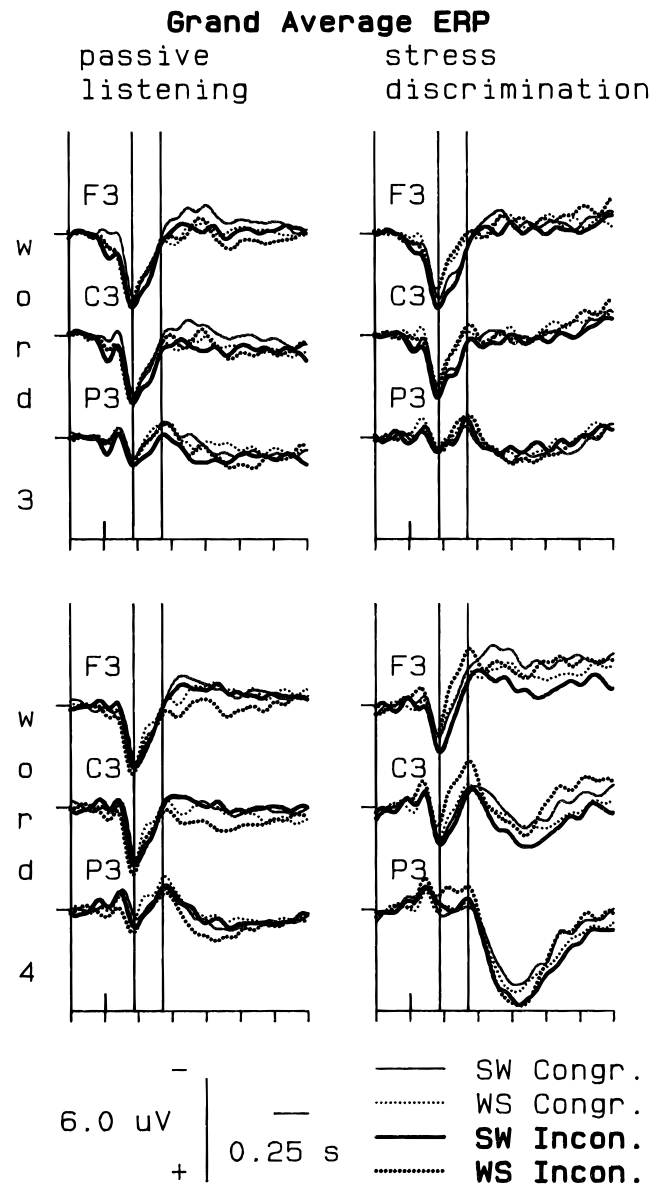
gruent) and signal (incongruent) trials respectively, that is, from the so-called false alarm and miss rates.

**Physiological data.** For the discrimination task only the trials not requiring an overt response were analyzed, to exclude any explanation in terms of motor preparation. First, trials were EOG-corrected using an autoregression model based on EOG calibration trials recorded before each experimental block (Van den Berg-Lenssen, Brunia, & Blom, 1989). Next, a semiautomatic artifact detection was performed, using the following criteria: (1) no spikes exceeding  $100 \mu\text{V}$  were present in the sampled epoch. After having met this criterion, data were 2-Hz lowpass filtered, for artifact detection only, and scanned for drift with respect to the baseline. Two criteria were used to define drift: (2) the minimum and maximum sample values should not differ from each other by more than  $80 \mu\text{V}$  and (3) the mean amplitude in four subsequently sampled intervals of equal length (in the present experiment 1,625 ms, starting from the onset of the first word) should not differ from baseline ( $-250$  to  $0$  ms before the first word) by more than  $35 \mu\text{V}$ .

Trials that met these criteria were divided in four parts of 1,750 ms each, the first 250 ms being a baseline interval before the onset of each word and the subsequent 1,500 ms covering the ERP on each separate word in the sequence. Thirty-two average ERPs were calculated per subject, that is, one for each word with a certain stress pattern (StrongWeak or WeakStrong) at a certain position (1, 2, 3, or 4) within a given sequence (congruent or incongruent) per task (passive listening or stress discrimination), respectively.

For a statistical analysis of the physiological data, we were interested in four measures of components that are often reported for speech stimuli and/or occur around the estimated time of lexical access (200–300 ms). Three measures, P2, N400 and LPC, were defined as the mean amplitude of three consecutive sample points (11.7 ms) measured baseline-to-peak. The latency of each peak was individually determined as the average latency of the local extreme at electrodes C3 and C4 in a 100-ms window (500 ms for the LPC) around the respective peak in the grand average (P2: 171–271 ms; N400: 381–481 ms; LPC: 566–1066 ms; Figure 1). Another peak that is often recorded with speech stimuli, that is, the N100, was not analyzed because it occurred before the estimated time of lexical access and because its amplitude (but not the amplitudes of the later peaks) was compromised by the digital lowpass filter.

The fourth measure was an area measure, the mean of the 210-ms interval between the grand average peak latencies of the P2 and the N400 (i.e., from 221 to 431 ms after word onset, see vertical lines in Figure 1). It was calculated from the 16 difference waves that resulted from the subtraction of words with a Strong-Weak stress pattern from words with a WeakStrong stress pattern, at each position within either congruent or incongruent sequences during each task, respectively. The subtraction in this direction allowed for a comparison of the infrequent versus the canonical stress patterns (StrongWeak stress patterns are more common in Dutch; Vroomen & De Gelder, 1995). The reason that a 210-ms interval was used to describe this WeakStrong-StrongWeak difference wave was that there was obviously a difference between ERPs on words with different stress patterns, but there was considerable variation (between subjects and conditions) in the latency at which this difference was maximal. Because of its polarity and the midpoint of the interval from which it is calculated, this difference measure will be denoted as N325. To aid comparison of N325 with a possible MMN or N2b, the difference ERP for the



**Figure 1.** Grand-average ( $N = 19$ ) event-related potential (ERP) waveforms in both tasks, on the third and the fourth word. Electrode positions as indicated. The small vertical marks indicate word onset. Note the surplus frontocentral negativity between the peaks of P2 and N400, indicated by the vertical lines, on the fourth word when it has an incongruent Weak-Strong stress pattern. WS = WeakStrong; SW = StrongWeak.

final word in incongruent minus congruent sequences was analyzed too.

On the P2, N400, and LPC, separate repeated-measures ANOVAs were performed using the module P4V of the BMDP statistical package. Factors included in the analyses were: task (passive, discrimination), congruence (same, different), stress pattern (Strong-Weak, WeakStrong), word position (1, 2, 3, 4), electrode position (F [F3 and F4], Tc [T3 and T4], C [C3 and C4], Tp [T5 and T6], and P [P3 and P4]), and hemisphere (left, right). The repeated-measures ANOVA on the WeakStrong-StrongWeak and Incongruent-Congruent area measures comprised the same factors, except for stress pattern and congruence, respectively. Significant interactions were clarified by breaking them down into simple effects.



**Table 2.** Error Percentages ( $\pm$  SD) of Both Performance Groups for Congruent and Incongruent Sequences Finishing With StrongWeak (SW) or WeakStrong (WS) Words Requiring a Response (R) or No Response (non-R)

	Good performers		Poor performers	
	SW	WS	SW	WS
Congruent				
R	0.56 ( $\pm$ 0.84)	2.04 ( $\pm$ 2.86)	9.26 ( $\pm$ 18.0)	9.07 ( $\pm$ 12.2)
non-R	0.74 ( $\pm$ 1.47)	1.11 ( $\pm$ 1.18)	5.19 ( $\pm$ 9.63)	4.07 ( $\pm$ 5.66)
Incongruent				
R	0.56 ( $\pm$ 1.18)	0.93 ( $\pm$ 1.69)	7.22 ( $\pm$ 7.27)	12.78 ( $\pm$ 5.62)
non-R	0.19 ( $\pm$ 0.56)	0.37 ( $\pm$ 0.74)	3.70 ( $\pm$ 4.06)	8.15 ( $\pm$ 4.28)

Note: Only non-R were included in the event-related potential analyses.

Where necessary, degrees of freedom were adjusted using the Greenhouse–Geisser epsilon to prevent an increase of Type I errors (Vasey & Thayer, 1987).

To examine the relationship between the behavioral data and the physiological measures, the subjects were split into two groups on the basis of the median sensitivity ( $d'$ ). This split resulted in a group of good performers ( $n = 9$ ) and a group of poor performers ( $n = 9$ ); the subject with the median sensitivity ( $d' = 3.8$ ) was dropped from this analysis.<sup>3</sup> Next, the same repeated-measures ANOVAs as before were applied, this time extended with the between-subjects factor group (good performers, poor performers).

## Results

### Behavioral Data

Table 2 presents average error percentages for both groups in all conditions. Subjects performed better when a trial ended with a StrongWeak stress pattern than when it ended with a WeakStrong stress pattern,  $F(1,18) = 7.46$ ,  $p < .05$ . They produced more errors on trials requiring a response than on those requiring no response (the latter were actually included in the average ERPs),  $F(1,18) = 5.24$ ;  $p < .05$ , although this tended to be true for the subgroup of poor performers only, Type  $\times$  Group:  $F(1,16) = 3.83$ ,  $p < .07$ . The poor performers also tended to make more errors on incongruent trials, Congruence  $\times$  Group:  $F(1,16) = 4.42$ ,  $p < .06$ , that is, in signal detection terms they made more misses than good performers. Finally, good performers tended to have a larger sensitivity,  $d' = 5.83$  vs. 3.26, respectively,  $t(16) = 4.10$ ,  $p < .001$ , and an equal bias,  $\beta = 1.00$  vs. 0.85, respectively,  $t(16) = 0.54$  (*ns*), for discriminating metrical stress patterns, compared with poor performers.

### Physiological Data

Figure 1 presents a general overview of the ERP waveforms for the third and the fourth word in both tasks. They were characterized by a N100 that was barely visible (partly due to filtering) and therefore not analyzed, followed by a P2 and N400 deflection, peaking on average at the latencies indicated by the vertical lines at 221 and 413 ms after word onset. The clearest effects of stress pattern and congruence occurred between those two peaks, as a surplus negativity for incongruent final WeakStrong words during stress discrimination (Figure 1, bottom right-hand panel). As shown by the

subtraction waveforms (Figure 2), the same negativity for WeakStrong compared with StrongWeak words was also present in the other conditions. Finally, the N400 and the LPC were also modulated by the experimental factors, for example, both are largest during stress discrimination for final, incongruent words. The distribution of the peaks is illustrated by the maps in Figure 3.

The ANOVA results will be presented below for each peak separately. Because the primary focus of this report is on endogenous stress effects, trends are discussed only if they bear relevance on this issue. For the same reason word position effects (except for the main effect) are reported only if they interact with congruence.

### P2

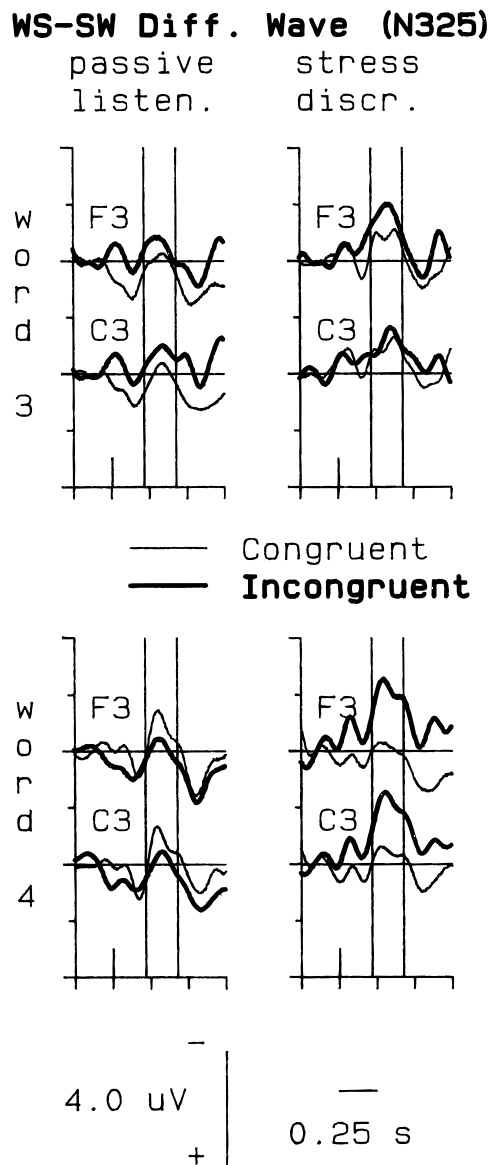
The P2, measured at the first vertical line in Figure 1, exhibited a frontocentral, bilaterally symmetrical maximum (Figure 3, first column) and decreased from the first to the last word in the sequence. These observations were confirmed by the ANOVA on P2 amplitudes, which was significant for the factors electrode,  $F(4,72) = 50.74$ ,  $p < .0001$ , Greenhouse–Geisser  $\epsilon = 0.40$ , and word position,  $F(3,54) = 13.63$ ,  $p < .0001$ ,  $\epsilon = 0.78$ , but not for hemisphere,  $F(1,18) < 1$ , *ns*.

In the passive listening condition WeakStrong words evoked a larger P2 than StrongWeak words (Figure 4a). This effect was absent during the stress discrimination task, during which the difference for the final word even tended to reverse (Figure 4b). Together these effects created marginally significant interactions between Task  $\times$  Stress Pattern,  $F(1,18) = 3.90$ ,  $p < .07$  and Task  $\times$  Stress Pattern  $\times$  Word Position,  $F(3,54) = 2.58$ ,  $p < .08$ ;  $\epsilon = 0.78$ . For the subgroup of good performers the reversal of the stress pattern effect during the discrimination task tended to occur for incongruent sequences only, and not for congruent sequences (Figure 4c), Task  $\times$  Congruence  $\times$  Stress Pattern  $\times$  Word Position  $\times$  Group:  $F(3,48) = 2.98$ ,  $p < .06$ ,  $\epsilon = 0.81$ . This difference also generated a couple of lower-order interactions and trends among congruence and stress pattern: Congruence  $\times$  Stress Pattern:  $F(1,18) = 3.17$ ,  $p < .10$  and Congruence  $\times$  Stress Pattern  $\times$  Group:  $F(1,16) = 10.63$ ,  $p < .01$ .

### WeakStrong-StrongWeak Difference Wave or N325

The WeakStrong-StrongWeak difference wave (Figure 2), showed an overall negativity (i.e., over tasks, conditions and word positions;  $F[1,18] = 28.00$ ,  $p < .0001$ ) with a frontocentral maximum (Figure 3, second column;  $F[4,72] = 10.67$ ,  $p < .001$ ,  $\epsilon = 0.42$ ),

<sup>3</sup> A median split on the average error rate (2.9%) led to identical groups.



**Figure 2.** Grand-average ( $N = 19$ ) WeakStrong-StrongWeak difference waves in both tasks, on the third and the fourth word. See Figure 1 for further legends.

in the interval between the peaks of P2 (grand-average latency 221 ms) and N400 (431 ms), which are indicated by the vertical lines in Figure 2. This ERP difference between words with a different stress pattern was small ( $1\text{--}2.5\ \mu\text{V}$ ) but robust in the present experiment. After the midpoint of the interval, which corresponded well with the average peak latency ( $320 \pm 35$  ms), we will denote this peak as N325.

In the discrimination task the N325 was lateralized to the left at frontocentral electrodes, where it attained larger amplitudes than during passive listening (Figure 3, second column), Task  $\times$  Electrode Position  $\times$  Hemisphere:  $F(4,72) = 3.60$ ,  $p < .05$ ,  $\epsilon = 0.70$ . The N325 was also affected by the congruence of the stress pattern of the last word with that of the first three words, in that it was larger for incongruent than for congruent sequences (Figure 2),  $F(1,18) = 5.88$ ,  $p < .05$ . However, congruence did not interact with word position, although visual inspection of the data sug-

gested that the N325 was largest on incongruent final words (see Figure 2). This apparent inconsistency could be resolved by comparing the results of both performance groups. As illustrated by Figure 5, the subgroup of good performers showed a differential N325 for congruent (no N325) and incongruent trials (relatively large N325) on the fourth word in the discrimination task, whereas the subgroup of poor performers produced an intermediate N325, with equal amplitudes in both task conditions, Task  $\times$  Congruence  $\times$  Word Position  $\times$  Group:  $F(3,48) = 3.41$ ,  $p < .05$ ,  $\epsilon = 0.92$ .

#### Incongruent-Congruent Difference Wave

Figure 6 shows the difference ERP for incongruent-congruent trials for StrongWeak and WeakStrong words separately. Overall the average amplitude within the interval from 221 to 431 ms (between both vertical lines) was not significantly different from baseline,  $F(1,18) < 1$ , *ns*. In fact, this difference wave was negative for WeakStrong words ( $-0.10\ \mu\text{V}$ ), but did not differ from 0, whereas for StrongWeak words the measure was significantly positive,  $+0.23\ \mu\text{V}$ ; stress pattern,  $F(1,18) = 5.84$ ,  $p < .05$ .

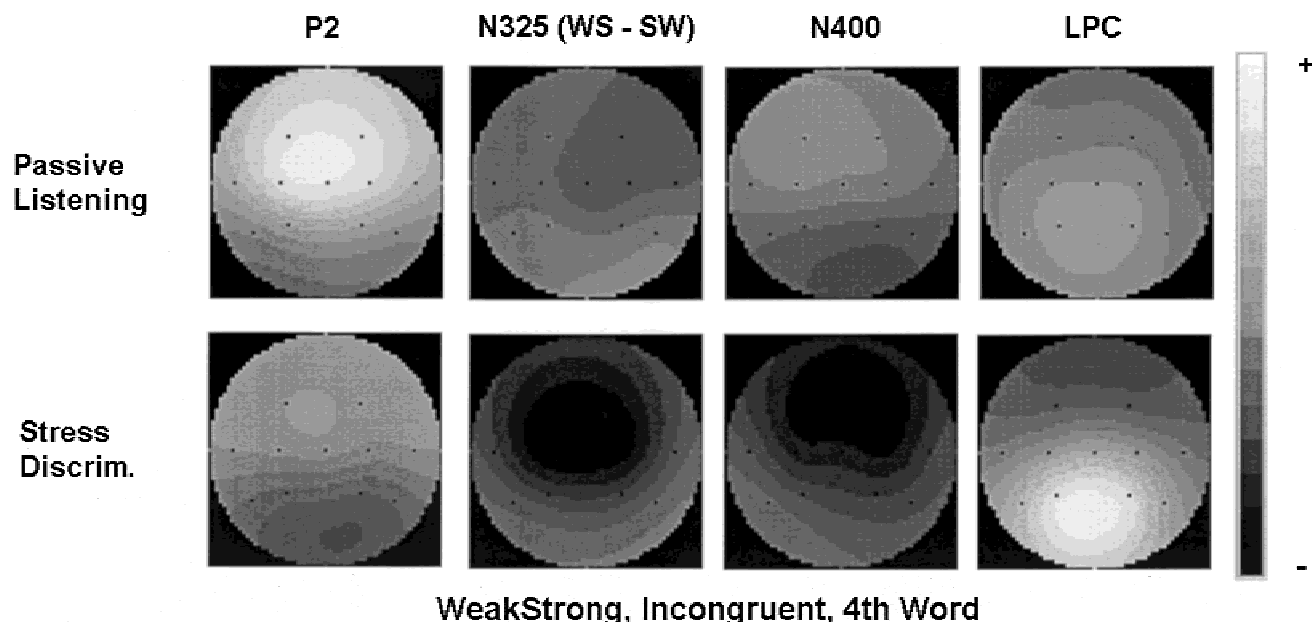
The impression that there was at least negativity for WeakStrong words on the final word (the stimulus that can be either congruent or incongruent) in the discrimination task was supported by a marginal significant interaction of Task  $\times$  Stress Pattern at the final word,  $F(1,18) = 3.94$ ,  $p < .07$ . It was again the group of good performers who showed the latter effect more reliably than the poor performers, Task  $\times$  Stress Pattern  $\times$  Word Position  $\times$  Group,  $F(3,48) = 3.39$ ,  $p < .05$ ,  $\epsilon = 0.66$ . The negativity for WeakStrong and the positivity for StrongWeak words at the final word in the discrimination task was significant only for the good performers; the poor performers showed a nonsignificant negativity for both types of stress patterns.

#### N400

With the exception of the final word in the discrimination task (see below), the N400 (Figure 1, second vertical line) showed a parietal maximum, which was slightly lateralized to the right (Figure 3, third column, top panel and Figure 7), electrode position:  $F(4,72) = 5.07$ ,  $p < .05$ ,  $\epsilon = 0.47$ ; hemisphere:  $F(1,18) = 10.56$ ,  $p < .01$ , Electrode  $\times$  Hemisphere:  $F(4,72) = 4.20$ ,  $p < .05$ ,  $\epsilon = 0.59$ , as is usually reported, and was largest on the final word,  $F(3,54) = 10.14$ ,  $p < .001$ ,  $\epsilon = 0.70$ .

The right hemisphere preponderance of the N400 was enhanced in the discrimination task (Figure 7), Task  $\times$  Hemisphere:  $F(1,18) = 7.75$ ,  $p < .05$ . Two effects were superimposed on this enhancement of right hemisphere amplitudes. First, over the left hemisphere during the discrimination task the N400 was larger on WeakStrong than on StrongWeak words (Figure 7, bottom row). Over the right hemisphere and during passive listening this difference was not significant, Task  $\times$  Stress Pattern  $\times$  Hemisphere:  $F(1,18) = 6.69$ ,  $p < .05$ . Second, in the discrimination task the N400 on the last word showed a frontocentral maximum, which deviated from the parietal maximum in the other conditions (Figure 3, third column, bottom panel versus top panel and Figure 7). This gave rise to a cluster of interactions comprising the factors task (T), word position (W), electrode position<sup>4</sup> (E), and hemisphere (H): W  $\times$  E:  $F(12,216) = 13.45$ ,  $p < .0001$ ,  $\epsilon = 0.31$ ;

<sup>4</sup>The interactions of electrode position with the other factors cannot be a consequence of the multiplicative effect warned against by McCarthy and Wood (1985), because a multiplicative effect cannot create a shift in the location of the maximum such as the one shown by the N325 (Figure 3, third column). Also refer to Haig, Gordon, & Hook (1997).



**Figure 3.** Topographic maps displaying the scalp distribution of the components recorded on final WeakStrong words following a sequence of StrongWeak words. Scaling is symmetrical and optimized for each component. Extreme values are 5  $\mu$ V (P2), 3  $\mu$ V (N325), 4  $\mu$ V (N400), and 7  $\mu$ V (LPC), respectively. During stress discrimination the resemblance between the frontocentral distribution of the N325 and the N400 suggests component overlap.

W  $\times$  E  $\times$  H:  $F(12,216) = 2.92$ ,  $p < .05$ ,  $\epsilon = 0.39$ ; T  $\times$  W:  $F(3,54) = 5.42$ ,  $p < 0.01$ ,  $\epsilon = 0.75$ ; T  $\times$  E:  $F(4,72) = 6.77$ ,  $p < .01$ ,  $\epsilon = 0.37$  and T  $\times$  W  $\times$  E:  $F(12,216) = 8.18$ ,  $p < .001$ ,  $\epsilon = 0.26$ .

Finally, there was a difference in stress pattern effects on the final word between both performance groups: for the good performers, it held true that final WeakStrong words tended to evoke a small N400 in incongruent series during passive listening and a large N400 during stress discrimination (Figure 8). No such differences were present for the poor performers, Task  $\times$  Congruence  $\times$  Stress Pattern  $\times$  Word Position  $\times$  Group:  $F(3,48) = 4.72$ ,  $p < .05$ ,  $\epsilon = 0.78$ . This difference in the group of good performers also entailed a marginally significance of this interaction in the total sample,  $F(3,54) = 2.66$ ,  $p < .09$ ,  $\epsilon = 0.68$ , whereas the marginally significant Congruence  $\times$  Stress Pattern interaction was in the same direction as in the discrimination task, that is, on incongruent trials WeakStrong words tended to evoke a larger N400 than on congruent trials,  $F(1,18) = 3.08$ ,  $p < .10$ .

Further interactions with word position and group were not broken down any further, because the group differences were present at the second word, which is not the focus of interest.

### LPC

The LPC peaked at 816 ms in the grand average (Figure 1) and showed a parietal maximum, slightly lateralized to the left (Figure 3, right-hand column), electrode position:  $F(4,72) = 13.87$ ,  $p < .0001$ ,  $\epsilon = 0.55$ ; hemisphere:  $F(1,18) = 13.5$ ,  $p < .01$ ; Electrode Position  $\times$  Hemisphere:  $F(4,72) = 8.80$ ,  $p < .001$ ,  $\epsilon = 0.78$ . The LPC was larger on incongruent than on congruent trials (Figure 1),  $F(1,18) = 6.94$ ,  $p < .05$ , and was clearly largest on the fourth word in the discrimination task to the point of being indiscernible in the passive listening task and on nonfinal word positions in the discrimination task (Figure 1). This produced several

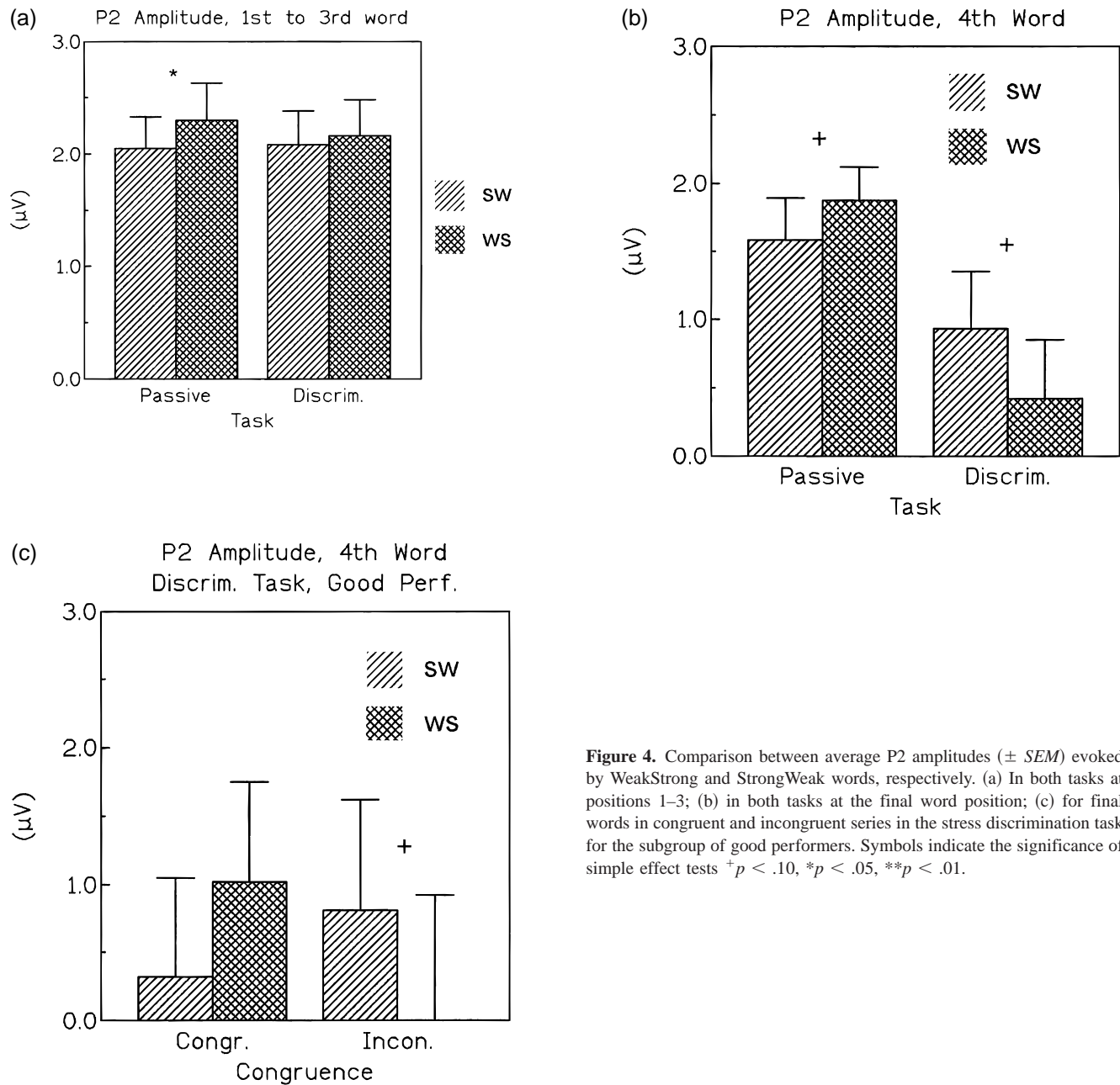
significant interactions among task (T) and word position (W) with electrode position (E) and hemisphere (H), T:  $F(1,18) = 8.96$ ,  $p < .01$ ; T  $\times$  W:  $F(3,54) = 5.12$ ,  $p < .05$ ,  $\epsilon = 0.46$ ; T  $\times$  E:  $F(4,72) = 16.60$ ,  $p < .001$ ,  $\epsilon = 0.39$ ; and T  $\times$  W  $\times$  E:  $F(12,216) = 12.91$ ,  $p < .0001$ ,  $\epsilon = 0.37$ .

StrongWeak words in congruent sequences tended to evoke relatively small LPC amplitudes over the left hemisphere in the discrimination task (Figure 1, bottom right-hand panel, electrode P3), Task  $\times$  Congruence  $\times$  Stress Pattern  $\times$  Hemisphere:  $F(1,18) = 4.37$ ,  $p < .06$ , and at electrode T3, Congruence  $\times$  Stress Pattern  $\times$  Electrode Position  $\times$  Hemisphere:  $F(4,72) = 2.43$ ,  $p < .09$ ,  $\epsilon = 0.64$ , respectively. These trends are difficult to interpret, because they could be broken down to a Congruence  $\times$  Stress Pattern interaction that does not interact with word position.

There were five between-group effects. First, for the good performers the left hemispheric dominance of the LPC was more pronounced in the discrimination task than in the passive task, Task  $\times$  Hemisphere  $\times$  Group:  $F(1,16) = 5.25$ ,  $p < .05$ , which was not the case for the poor performers. The remaining interactions with group occurred at temporal electrodes or at nonfinal words, which were neither the focus of the LPC nor of the N325.

### Discussion

Behavioral data have established that rhythmic properties of language are important in domains such as speech segmentation (e.g., Cutler & Norris, 1988; Vroomen et al., 1996) and language acquisition (Jusczyk et al., 1993; Mehler et al., 1994). We made an initial attempt to investigate whether the auditory ERP comprises a component that shows sensitivity to metrical stress, the rhythmic alternation of weak and strong syllables in stress-timed languages such as Dutch and English. To this end, Dutch listeners were presented with strings of Dutch words with opposite stress pat-



**Figure 4.** Comparison between average P2 amplitudes ( $\pm$  SEM) evoked by WeakStrong and StrongWeak words, respectively. (a) In both tasks at positions 1–3; (b) in both tasks at the final word position; (c) for final words in congruent and incongruent series in the stress discrimination task for the subgroup of good performers. Symbols indicate the significance of simple effect tests  $^+p < .10$ ,  $*p < .05$ ,  $**p < .01$ .

terns. The latency of such an ERP component could provide evidence about the time-course of rhythm-driven speech segmentation.

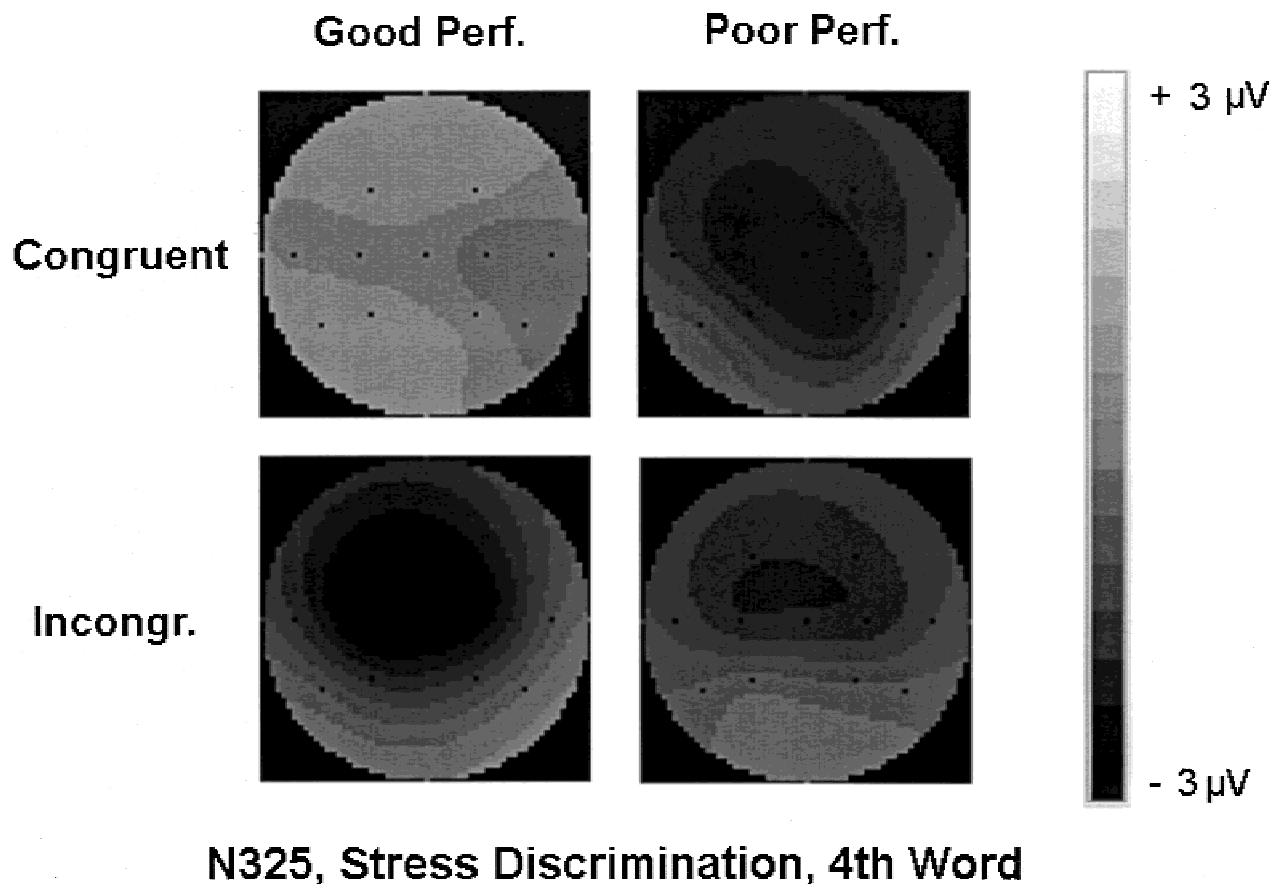
The ERP showed four peaks. Three of those are manifestations of components that are typical for auditory linguistic material: the P2, N400, and LPC (Figure 1). The fourth component, the N325, emerged in the WeakStrong-StrongWeak difference wave (Figure 2). Most of the endogenous stress effects were observed at this component. The P2 showed mainly exogenous effects and the N400 and LPC mostly endogenous task effects. Most of the endogenous effects were larger for that half of the subjects who showed the better performance in terms of discrimination accuracy ( $d'$ ).

#### Exogenous Metrical Stress Effects

Although the P2 amplitude on WeakStrong words was larger than on StrongWeak words in the passive listening condition only, and

not during the stress discrimination task (Figure 4a and 4b), this P2 enhancement is considered to be an exogenous effect. For an endogenous stress effect we would expect the opposite difference, that is, larger amplitude differences between stress patterns during stress discrimination (compare Picton & Hillyard, 1974). The disappearance of the P2 enhancement for WeakStrong words in the active task condition can be explained by temporal overlap with the N325 (the surplus negativity for WeakStrong words, between the peaks of the P2 and the N400), which was larger in the active task, at least at frontocentral electrodes, where both P2 and N325 reached their maximum amplitude. This interpretation is strengthened by the fact that in the one condition in which the good performers showed a very small N325, that is, on the final word in congruent trials (Figure 5), their P2 on WeakStrong was larger (although *ns*) than on StrongWeak words (Figure 4c), just as in the





**Figure 5.** Topographical maps of the WeakStrong–StrongWeak difference waves (N325) for both performance groups ( $n = 9$ ) for the final word in the discrimination task. Note the differences in surplus negativity for the good performers, whereas the negativity is constant over level of congruence for the poor performers.

passive condition (Figure 4a and 4b). This would imply that the ascending limb of the N325 started at, or before, the peak latency of the P2 (221 ms).

The WeakStrong–StrongWeak difference wave can also be considered to have an exogenous quality, because the N325 showed an overall significant negativity, irrespective of task condition, congruence, or word position (see, e.g., the top left-hand panel of Figure 2). The distribution of this effect was frontocentral (Figures 3 and 5). Given its exogenous quality, this negative shift is probably due to physical/acoustical differences between WeakStrong and StrongWeak stimuli (Fear et al., 1995; Appendix B) or the distribution of weak-initial and strong-initial entries in the Dutch lexicon (12% vs. 88%; Vroomen & de Gelder, 1995). The other panels of Figure 2 show that in the other conditions the negativity was generally larger. This exogenous quality distinguishes the N325 from all other components analyzed in the present paper, including the temporally overlapping P2 and N400, which did not show a main effect of metrical Stress Pattern.

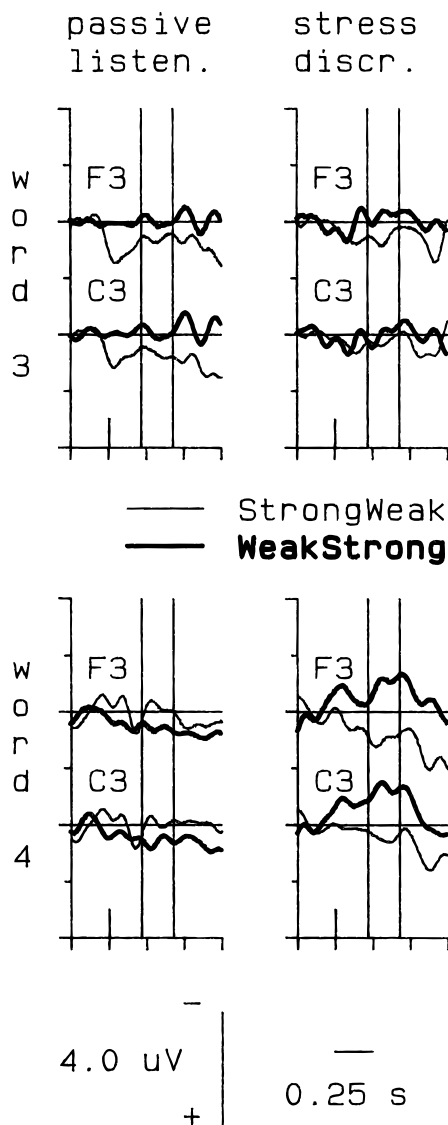
#### *Endogenous Metrical Stress Effects*

Next to an exogenous effect, the N325 showed all the signs of endogenous stress effects that were suggested in the Introduction. Remember that the N325 is a difference potential between WeakStrong and StrongWeak words, that is, the factor stress is part of all

effects by implication. First, there is the left-frontal lateralization during the discrimination task, which was absent in the passive task (Figure 3, second column). Second, the N325 was larger for sequences terminating with a change in stress pattern (Figure 2). Further analysis demonstrated that this effect was caused mainly by the subgroup of good performers who displayed a larger N325 on the final word in incongruent than in congruent sequences (Figure 5; bottom vs. top left-hand panel).

The N400 also showed endogenous stress effects: its amplitude was influenced by interactions between stress, task, hemisphere, electrode position, and word position. Together these effects shifted the commonly reported right parietal maximum of the N400 (Kutas & Van Petten, 1994) toward a frontal maximum on the final WeakStrong word in the discrimination task (compare the bottom with the top map in the third column of Figure 3 and with those in Figure 7). The frontal distribution of the N400 for final WeakStrong words in the discrimination task was almost identical to that of the N325 (compare the second and third map in the bottom row of Figure 3), and different from the distribution of the N400 in the other conditions (e.g., Figure 7), which suggests strongly that these effects on the N400 were caused by temporal overlap between N325 and N400. The enhancement of the N400 to final WeakStrong words in incongruent sequences for the good performers during stress discrimination (Figure 8), again mirrored differences in N325 amplitudes (Figure 5).

### Incongruent - Congruent Difference Wave



**Figure 6.** Grand-average ( $N = 19$ ) Incongruent–Congruent difference waves in both tasks, on the third and the fourth word. Note that the difference wave is dominated by positivity in the interval from 221 to 431 ms (vertical lines), with the exception of the difference waves for final WeakStrong words in the discrimination task (for the good performers, see text). See Figure 1 for further legends.

Given these endogenous stress effects we can give a positive answer to the main question of this paper. Our findings suggest strongly that a stress-sensitive ERP component exists. The WeakStrong–StrongWeak difference wave, measured as N325, shows both potentially exogenous and clear endogenous signs of stress-dependent modulations. First, exogenous effects, because the N325 reflects an extra overall negativity of WeakStrong over StrongWeak words, which is probably due to physical/acoustical correlates of metrical stress, although this finding might also reflect the distribution of metrical stress over initial syllables in Dutch. The choice between those possibilities should be based on cross-

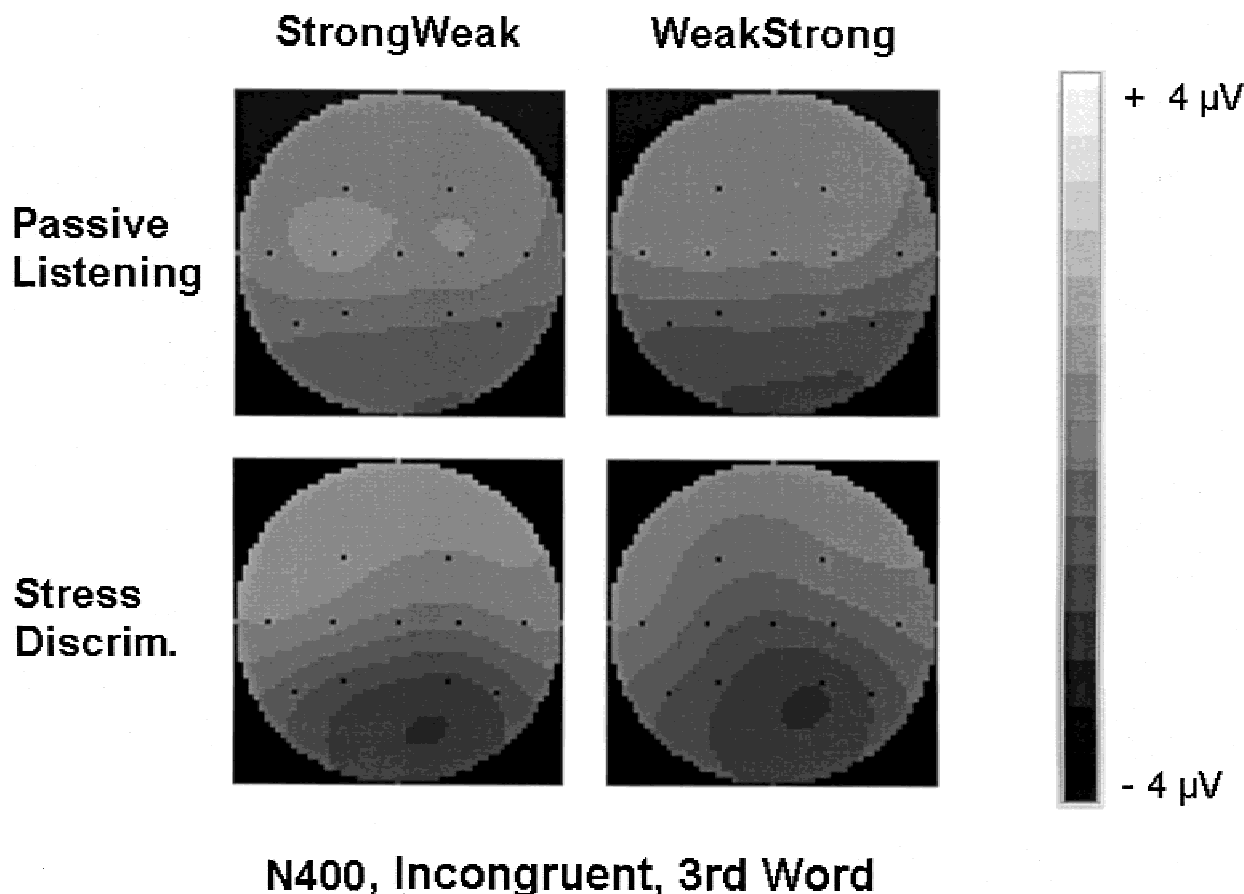
linguistic research. Second, endogenous effects, because the amplitude of the N325 is modulated by task condition, congruence, and performance level.

What might be the processing nature of the N325? The N325 overlaps in time with the N400 but its peak clearly precedes the peak of the N400 (Figure 2). Therefore, lexical integration as indexed by the N400 (e.g., Chwilla, 1996; Chwilla et al., 1995; Connolly, Byrne, et al., 1995; Holcomb, 1993; Van Petten & Kutas, 1990) is not a likely candidate. This leaves us with prelexical processes and lexical processes (e.g., initial access, competition, and selection) as candidates. Given that we study the ERP correlate of a phonological property (i.e., metrical stress) and that the present experiment comprises no lexical (e.g., word frequency, uniqueness point) manipulation, (initial) lexical access is a good candidate. This hypothesis is confirmed by the time window of the N325, starting around 225 ms, or even earlier, as indicated by the temporal overlap with the P2 (Figures 1 and 2; see Connolly & Phillips, 1994, for a similar argument about the processing nature of P2 in the passive listening condition (Figure 4a and 4b) and the exogenous part of the N325, that is, the small negativity that is present for all WeakStrong–StrongWeak contrasts (Figure 2), reflect the representation of the acoustic signal. The task-related endogenous modulations of the N325 (Figures 2, 3, and 5) reflect the extraction of relevant information from the acoustic signal. In the present experiment this was the metrical stress pattern per se, and its relation to the stress pattern of earlier words in the sequence, which was relevant for motor behavior. In listening to connected speech, metrical stress is relevant for speech segmentation (Cutler & Norris, 1988; even in the preverbal infant, Jusczyk et al., 1993; Mehler et al. 1994; Morgan, 1996).

If we accept that the N325 in the present experiment reflects metrical stress, a further matter of debate is just what aspect of metrical stress is reflected in the N325? Metrical stress is defined primarily in terms of vowel quality (see Appendix B also), that is, the frequency of the first and the second formants, which also define the identity of the vowel itself. However, reduced vowels are also characterized by certain prosodic characteristics. They are of shorter duration, less intense and have a lower pitch than full vowels (Fear et al., 1995; Appendix B). Recent research in our laboratory (Vroomen & de Gelder, 1997) has shown that, in Dutch listeners, the rhythm produced by word level stress, as realized by prosodic contrasts between syllables, might be more important than that produced by metrical stress. Future research should clarify whether the N325 is a reflection of vowel quality and/or prosody, which could not be differentiated in the present study.

#### Endogenous Task Effects

The only endogenous effect on the N400 that is totally independent of stress and temporal overlap with N325, was the enhancement of the right hemisphere preponderance during the active versus the passive task (Figure 7). The only prominent LPC was recorded on the final word in the discrimination task (Figure 1, bottom right-hand panel). In fact the large deflection in the grand average was largely due to the subgroup of good performers. In a similar vein Curran and co-workers (1993) reported a significant correlation between LPC amplitude and recognition memory. This correlation opens the possibility that the LPC is a late instance of P3, which is interpreted as a manifestation of categorization or context updating (Donchin, Ritter, & McCallum, 1978). Only those subjects who reliably discriminated WeakStrong from StrongWeak patterns showed the ERP manifestation of categorization. As should be



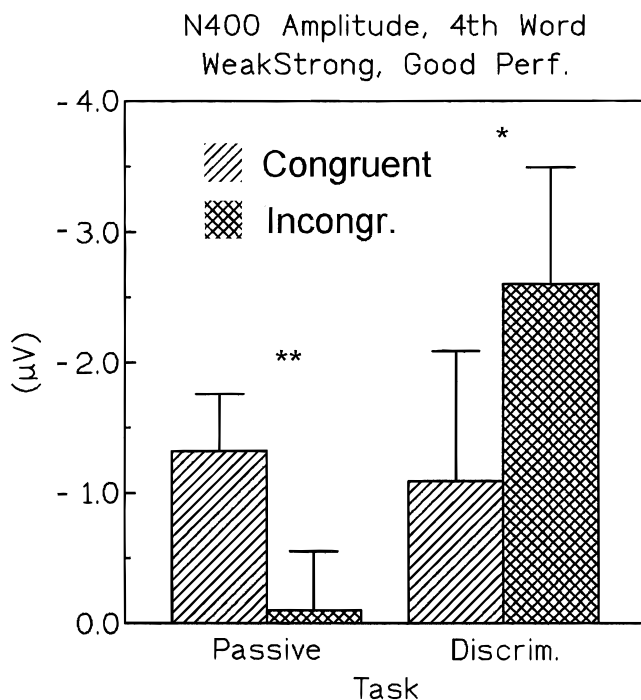
**Figure 7.** Scalp distribution of the grand-average ( $N = 19$ ) N400 amplitudes in both tasks, on the third word of incongruent series. The third word is displayed because it best indicates task effects on the distribution that are independent word position. See the deviant scalp distribution on the fourth word in the discrimination task (Figure 3, second column). Note: the position of the extrapolated occipital maximum in some of the panels is unreliable.

expected for such a component, the LPC was sensitive to congruence, which defined the response to be issued.

#### *The N325 in Relation to Other ERP Components*

We should consider whether the N325 is an instance of a known ERP component. The possibility that the N325 is an instance of N100 evoked by the second syllable of WeakStrong words (onset  $243 \pm 52$  ms) is considered unlikely. First, unlike the present N325, the N100 does not show left hemispheric dominant amplitudes during active discrimination conditions or likewise manipulations and is more closely related to stimulus detection per se (e.g., Näätänen, 1992, Chapter 4) than to discrimination sensitivity, which contrasts with the performance group effects on N325. This leaves open the possibility that the exogenous part of the N325 reflects the N100. However, in that case the N325, like the N100, should be reduced by the 6 Hz lowpass filtering, which it is not. Finally, in a recent experiment where, due to somewhat different recording conditions, lowpass filtering was unnecessary, we were able to study the (quite large) N100 evoked by the first syllable. Strong syllables in first position did not consistently evoke larger N100 amplitudes than weak syllables, and even if they did the difference was at maximum half that of the N325.

It remains possible that the N325 is not specific to language stimuli, but occurs in the processing of any discriminative auditory stimulus. Known ERP manifestations of such processing are the MMN (e.g., Näätänen et al., 1978), which has recently been implied in language-specific phoneme discrimination as well (Näätänen et al., 1997, Dehaene-Lambertz, 1997) and the N2b (Näätänen & Gaillard, 1983). Both are evoked by a change in a series of repetitive stimuli (the so-called oddball paradigm). The MMN is probably automatic in the sense of being independent of attentional resources, whereas it is overlapped by N2b when attention is drawn to the stimuli (Näätänen, 1992). Figure 6 shows the incongruent–congruent difference waves from which the MMN and N2b can be calculated, although usually the preceding standard is subtracted from the deviant stimulus (see however Sams, Alho, & Näätänen, 1983, for a similar and Dehaene-Lambertz, 1997, for an identical MMN subtraction). With the exception of the difference wave (of the good performers) on final WeakStrong words in the stress discrimination task, this difference wave was mostly zero or positive, both within and before the N325 analysis window (221–431 ms). The passive listening condition showed no sign of a negative difference wave (Figure 6, left-hand column). This finding implies that the N325



**Figure 8.** Average N400 amplitudes ( $\pm$  SEM) for the subgroup of good performers for final WeakStrong words in congruent and incongruent series during both tasks. See Figure 4 for further legends.

is not an instance of the MMN, because in that case it should not depend on attentional resources, which covaried with task condition in the present experiment. The N325 is also unlikely to be an instance of N2b, because the incongruent–congruent difference wave was negative only if WeakStrong words followed a sequence of StrongWeak words and not vice versa (Figure 6, bottom right-hand panel). Stated otherwise, the negativity around 325 ms varied more systematically with stress pattern than with congruence and therefore is not an instance of N2b.

Next, we consider a number of early (200–400 ms) frontal ERP manifestations of receptive language processing, starting with the PMN (Connolly & Phillips, 1994). The PMN is evoked between 270 and 300 ms after the onset of a sentence terminal word with an unexpected onset-phoneme (i.e., different from that of the highest cloze-probability word) and has a (left) frontocentral distribution. It is not followed by an N400 if the evoking stimulus is semantically congruous. The PMN was found to be specific to the auditory modality (Connolly, Phillips, & Forbes, 1995). The N325 in the present experiment could be an instance of PMN in as far as metrical stress is reflected in phonology. However, the N325 was also evoked by nonfinal words in the sequence, in which case the expectation about stress pattern (if any) was always confirmed. Furthermore, for the final word the N325 was more sensitive to stress pattern per se than to congruence (as argued in the previous paragraph, see Figure 6). Finally, the PMN is interpreted as a manifestation of an exclusionary process during lexical selection in the original Cohort model (Connolly & Phillips, 1994), whereas metrical stress, at least in English, does not constrain the cohort of possible word candidates (Cutler, 1986; Cutler, Dahan, & van Donselaar, 1997).

Of particular interest here is a report by Patel and Holcomb (1997) who recorded an early (150–200 ms) negativity preceding

the N400 (at Cz and Pz)<sup>5</sup> to anomalous words ending a non-rhythmical compared with a rhythmical sentence. Regular rhythmicity was probably produced by reducing the unstressed syllables. Thus, a less predictable stressed, strong-initial word evoked extra negativity, that could well be another instance of PMN. That negativity differs from the N325 described in the present paper, because it was evoked by strong-initial versus weak-initial words. Future studies should elucidate the relation between metrical stress, N325 and PMN.

The ERP correlates of another phonological phenomenon, that is, rhyme, have been studied more extensively (Praamstra, Meyer, & Levelt, 1994; Rugg, 1984). Nonrhyming words evoked a larger N400, with a posterior distribution, than rhyming words. For non-alliterating versus alliterating words (i.e., words that share onsets vs. codas), this effect occurred earlier (250–450 ms), and displayed a frontocentral distribution, not unlike that of the PMN and N325.

Another language-related ERP component in the relevant latency range is the so-called LPN (Kutas, 1997), which has a left anterior temporal distribution and a latency that increases with word-length and decreases with word frequency. If the N325 is an instance of LPN the implication would be that LPN amplitude is dependent on factors that influence lexical access, here metrical stress. The present latency (325 ms) conforms to expectation for a bisyllabic noun (Figure 2 in Kutas, 1997). The relation between LPN and N325 merits further research. These negativities are probably also related to the early part (250–350 ms) of the magnetic brain response from the vicinity of the left auditory cortex, to contextually inappropriate as well as appropriate but unexpected words, which have been recorded recently by Helenius, Salmelin, Service, and Connolly (1998). The authors related this activity to word recognition processes and the later part to postlexical integration (compare N400).

Another ERP manifestation of linguistic processing, more particularly of (a violation in) first-pass syntactic parsing (Friederici & Mecklinger, 1996; Neville et al., 1991), or the working memory load associated with parsing (King & Kutas, 1995; Müller et al., 1997) is the (early) left anterior negativity or (E)LAN. With respect to the relation between the N325 and the (E)LAN, which share latency, polarity and a left frontal distribution, it is interesting to note that durational cues (which also induce rhythm) are used to disambiguate syntax (Cutler et al., 1997). However, it is improbable that WeakStrong words pose a higher working memory load than StrongWeak words, especially during passive listening.

In sum, there is a variety of ERP components related to receptive language processing, with latencies of 150–350 ms, and an anterior (PMN, LPN, [E]LAN N325, alliteration-related N400) or posterior (rhyme-related N400 and “early N400”; Patel & Holcomb, 1997) scalp distribution. Most of these are related to word-level phonological or lexical processes by the authors, except for the (E)LAN, which is related to syntax or working memory. On the one hand, as argued above, the N325 differs from the other components in that it was observed under different antecedent conditions, and we do not want to identify it with one or the other component. On the other hand, the antecedent conditions of the N325 and most of the other language-related negativities concern factors known to influence lexical access, activation, or selection

<sup>5</sup>At lateral sites, the effect of rhythmicity is much smaller and less separable from N400. Patel and Holcomb (1997) interpreted the effect as advancing N400 latency.



and competition. This makes us hesitate to introduce another functional label, because the N325 and those other negativities might be regarded as one and the same component at a more abstract level. Another consequence of many antecedent conditions being related to similar early left-frontal negativities is that psycholinguistic experiments using ERPs should be designed carefully to avoid confounding factors (e.g., syntactic class, onset-phoneme, word frequency, metrical stress pattern) that might all influence the same observed negative peak (compare Cutler, 1981).

Finally, we reconsider the relation between N325 and N400. Following the arguments by Connolly et al. (1990; Connolly, Phillips, Stewart, & Brake, 1992) they can be separated on three bases. First, both components were observed in the present experiment, albeit that the N400 is clearer in the conventional ERP and the N325 in the subtracted wave forms, respectively. Second, their topographies are different (Figure 3): the N400 displayed a right parietal distribution in conditions where the N325 was small (Figure 7). The N325 displayed a left frontocentral maximum (Figure 5). This difference was enhanced during the discrimination task. Third, the N325 was evoked by any WeakStrong word and the N400 was only enhanced by final incongruent WeakStrong words. As argued before, the topography of the N400 is shifted to that of the N325 for these final WeakStrong words, which suggests component overlap. Finally, the N400 is generally interpreted as a manifestation of postlexical integration (e.g., Chwilla, 1996; Chwilla et al., 1995; Connolly, Byrne, et al., 1995; Holcomb, 1993; Van Petten & Kutas, 1990), whereas components in the latency range of the N325 (like LPN and PMN) are generally interpreted as manifestations of lexical processes.

## Conclusion

From the present experiment we conclude that, with respect to the processing nature of the observed components, the P2 is an exogenous component that reflects the physical/acoustic stimulus parameters (Appendix B), whereas the N400 and LPC are endogenous components that reflect task requirements. Irrespective of metrical stress, the right-hemisphere preponderance of the N400 was enhanced, and the LPC mainly reflected congruence (for the good performers at least), during stress discrimination. The N325 is at the interface between the exogenous P2 and the endogenous N400 and LPC and does show main as well as task-related stress effects. The overall surplus negativity for weak-initial words (which constitute only 12% of the Dutch lexicon) reflected stress pattern in an exogenous way, like the P2 enhancement). The endogenous part of the N325 manifests the interaction of stress pattern with task condition, congruence, and discrimination accuracy. During stress discrimination the N325 was larger for weak-initial words following a sequence of strong-initial words, at least for those subjects who discriminated best between metrical stress patterns. This finding suggests that the N325 is a manifestation of the process(es) that extract the metrical stress pattern from the acoustic signal and translate it into task requirements. This result is promising in the sense that not only did we record an ERP component reflecting metrical stress patterns in the expected pre- or peri-lexical latency range, but the N325 is also a manifestation of those underlying processes that are sensitive to task requirements. This result makes the N325 a good candidate for investigating the temporal dynamics of the role of rhythm in spoken word recognition.

## REFERENCES

- CELEX. (1990). *Dutch database (release N31)*. [On-line: <http://www.kun.nl/celex/>]. Available: Nijmegen, The Netherlands: Centre for Lexical Information.
- Chwilla, D. J. (1996). *Electrophysiology of word processing: The lexical processing nature of the N400 priming effect*. Unpublished doctoral dissertation, Nijmegen University, The Netherlands.
- Chwilla, D. J., Brown, C. M., & Hagoort, P. (1995). The N400 as a function of the level of processing. *Psychophysiology*, 32, 274–285.
- Connolly, J. F., Byrne, J. M., & Dywan, C. A. (1995). Assessing adult receptive vocabulary with event-related potentials: An investigation of cross-modal and cross-form priming. *Journal of Clinical and Experimental Neuropsychology*, 17, 548–565.
- 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, 256–266.
- Connolly, J. F., Phillips, N. A., & Forbes, K. A. K. (1995). The effects of phonological and semantic features on sentence-ending words on visual event-related brain potentials. *Electroencephalography and Clinical Neurophysiology*, 94, 276–287.
- 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.
- Curran, T., Tucker, D. M., Kutas, M., & Posner, M. I. (1993). Topography of the N400: Brain electrical activity reflecting semantic expectancy. *Electroencephalography and Clinical Neurophysiology*, 88, 188–209.
- Cutler, A. (1981). Making up materials is a confounded nuisance: Or will we be able to run any psycholinguistic experiments at all in 1990? *Cognition*, 10, 65–70.
- 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 the English vocabulary. *Computer Speech and Language*, 2, 133–142.
- 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., & Norris, D. (1988). The role of strong syllables in segmentation for lexical access. *Journal of Experimental Psychology: Human Perception and Performance*, 14, 113–121.
- D'Anna, C. A., Zechmeister, E. B., & Hall J. W. (1991). Toward a meaningful definition of vocabulary size. *Journal of Reading Behaviour*, 23, 109–122.
- Dehaene-Lambertz, G. (1997). Electrophysiological correlates of categorical phoneme perception in adults. *NeuroReport*, 8, 919–924.
- Donchin, E., Ritter, W., & McCallum, W. C. (1978). Cognitive psychophysiology: The endogenous components of the ERP. In E. Callaway, P. Tueting, & S. H. Koslow (Eds.), *Event-related brain potentials in man* (pp. 349–411). New York: Academic Press.
- Donchin, E., Spencer, K. M., & Dien, J. (1997). The varieties of deviant experience: ERP manifestations of deviance processors. In G. J. M. van Boxtel and K. B. E. Böcker (Eds.), *Brain and behavior: Past, present and future* (pp. 67–91). Tilburg, The Netherlands: Tilburg University Press.
- Fear, B. D., Cutler, A., & Butterfield, S. (1995). The strong/weak syllable distinction in English. *Journal of the Acoustical Society of America*, 97, 1893–1904.
- Friederici, A. D., & Mecklinger, A. (1996). Syntactic parsing as revealed by brain responses: First-pass and second-pass parsing. *Journal of Psycholinguistic Research*, 25, 157–176.
- Grosjean, F. (1980). Spoken word recognition processes and the gating paradigm. *Perception and Psychophysics*, 28, 267–283.
- Haig, A. R., Gordon, E., & Hook, S. (1997). To scale or not to scale: McCarthy and Wood revisited. *Electroencephalography and Clinical Neurophysiology*, 103, 323–325.

- Helenius, P., Salmelin, R., Service, E., & Connolly, J. F. (1998). Distinct time course of word and context comprehension in the left temporal cortex. *Brain*, 121, 1133–1142.
- Holcomb, P. J. (1993). Semantic priming and stimulus degradation: Implications for the role of the N400 in language comprehension. *Psychophysiology*, 30, 47–61.
- Juszyk, P. W., Cutler, A., & Redanz, N. J. (1993). Infants' preference for the predominant stress pattern of English words. *Child Development*, 64, 675–687.
- King, J. W., & Kutas, M. (1995). Who did what and when? Using word- and clause-level ERPs to monitor working memory usage in reading. *Journal of Cognitive Neuroscience*, 7, 376–395.
- Kutas, M. (1997). Views on how the electrical activity that the brain generates reflects the functions of different language structures. *Psychophysiology*, 34, 383–398.
- Kutas, M., & Hillyard, S. A. (1980). Reading senseless sentences: Brain potentials reflect semantic incongruity. *Science*, 207, 203–205.
- Kutas, M., & Van Petten, C. K. (1994). Psycholinguistics electrified (event-related brain potential investigations). In M. A. Gernsbacher (Ed.), *Handbook of psycholinguistics* (pp. 83–143). San Diego, CA: Academic Press.
- McCarthy, G., & Wood, C. C. (1985). Scalp distributions of event-related potentials: An ambiguity associated with analysis of variance models. *Electroencephalography and Clinical Neurophysiology*, 62, 203–208.
- McClelland, J. L., & Elman, J. L. (1986). The TRACE model of speech perception. *Cognitive Psychology*, 18, 1–86.
- Mehler, J., Bertocini, J., Dupoux, E., & Pallier, C. (1994). The role of suprasegmentals in speech perception and acquisition. *Dokkyo International Review*, 7, 343–376.
- Morgan, J. L. (1996). A rhythmic bias in preverbal speech segmentation. *Journal of Memory and Language*, 35, 666–688.
- Müller, H. M., King, J. W., & Kutas, M. (1997). Event-related potentials elicited by spoken relative clauses. *Cognitive Brain Research*, 5, 193–203.
- Näätänen, R. (1992). *Attention and brain function*. Hillsdale, NJ: Erlbaum.
- Näätänen, R., & Gaillard, A. W. K. (1983). The orienting reflex and the N2 deflection of the ERP. In A. W. K. Gaillard & W. Ritter (Eds.), *Advances in psychology: Vol. 10. Tutorials in event related potential research: Endogenous components* (pp. 119–141). Amsterdam: North-Holland.
- Näätänen, R., Gaillard, A. W. K., & Mäntysalo, S. (1978). Early selective-attention effect on evoked potential reinterpreted. *Acta Psychologica*, 42, 313–329.
- Näätänen, R., Lehtokoski, A., Lennes, M., Cheour, M., Huottilainen, M., Iivonen, A., Vainio, M., Alku, P., Ilmoniemi, R. J., Luuk, A., Allik, J., Sinkkonen, J., & Alho, K. (1997). Language-specific phoneme representations revealed by electric and magnetic brain responses. *Nature*, 385, 432–434.
- Neville, H., Nicol, J. L., Barss, A., Foster, K. I., & Garret, M. F. (1991). Syntactically based sentence processing classes: Evidence from event-related brain potentials. *Journal of Cognitive Neuroscience*, 3, 151–165.
- Patel, A. D., & Holcomb, P. J. (1997, March). *Semantic context, rhythmic regularity, and lexical processing in connected speech: An event-related potential (ERP) study*. Poster session presented at the annual meeting of the Cognitive Neuroscience Society, Boston MA.
- Picton, T. W., & Hillyard, S. A. (1974). Human auditory evoked potentials: II. Effects of attention. *Electroencephalography and Clinical Neurophysiology*, 36, 191–200.
- Praamstra, P., Meyer, A. S., & Levelt, W. J. M. (1994). Neurophysiological manifestations of phonological processing: Latency variation of a negative ERP component time-locked to phonological mismatch. *Journal of Cognitive Neuroscience*, 6, 204–219.
- Pritchard, W. S., Shapell, S. A., & Brandt, M. E. (1991). Psychophysiology of N200/N400: A review and classification scheme. *Advances in Psychophysiology*, 4, 43–106.
- Rugg, M. D. (1984). Event-related potentials in phonological matching tasks. *Brain and Language*, 23, 225–240.
- Rugg, M. D. (1985). The effects of word repetition and semantic priming on event-related potentials. *Psychophysiology*, 22, 642–647.
- Sams, M., Alho, K., & Näätänen, R. (1983). Sequential effects on the ERP in discriminating two stimuli. *Biological Psychology*, 17, 41–58.
- Tyler, L. K., & Wessels, J. (1983). Quantifying contextual contributions to word-recognition processes. *Perception and Psychophysics*, 34, 409–420.
- Van den Berg-Lenssen, M. M. C., Brunia, C. H. M., & Blom, J. A. (1989). Correction for ocular artifacts in EEGs using an autoregressive model to describe the EEG: A pilot study. *Electroencephalography and Clinical Neurophysiology*, 73, 72–83.
- Van Petten, C., & Kutas, M. (1990). Interactions between sentence context and word frequency in event-related brain potentials. *Memory and Cognition*, 18, 380–393.
- Vasey, W. M., & Thayer, J. F. (1987). The continuing problem of false positives in repeated measurements ANOVA in psychophysiology: A multivariate solution. *Psychophysiology*, 24, 474–486.
- Vroomen, J., & de Gelder, B. (1995). Metrical segmentation and lexical inhibition in spoken word recognition. *Journal of Experimental Psychology: Human Perception and Performance*, 21, 98–108.
- Vroomen, J., & de Gelder, B. (1997). Trochaic rhythm in speech segmentation. *Abstracts of the Psychonomic Society*, 2, 72.
- Vroomen, J., Van Zon, M., & de Gelder, B. (1996). Cues to speech segmentation: Evidence from juncture misperceptions and word spotting. *Memory and Cognition*, 24, 744–755.

(RECEIVED November 14, 1997; ACCEPTED January 5, 1999)

## APPENDIX A

The set of matched StrongWeak and WeakStrong Dutch bisyllabic monomorphemic nouns selected for the present study. The weak vowel was always realized as schwa (/ə/).

StrongWeak		WeakStrong	
bengel	(scamp)	beschuit	(Dutch rusk)
bezem	(broom)	beton	(concrete)
brasem	(bream)	bretel	(suspenders)
franje	(fringe)	fregat	(frigate)
geiser	(geyser)	gelei	(jelly)
geste	(gesture)	gemeen	(rabble)
gevel	(facade)	geniep*	(sly)
kavel	(lot)	kebab	(kebab)
mantel	(coat)	meloen	(melon)
mossel	(mussel)	miljard	(billion)
parel	(pearl)	pedaal	(pedal)
pater	(Reverend)	persoon	(person)
pleister	(plaster)	plezier	(fun)
ritme	(rhythm)	rebel	(rebel)
sabel	(saber)	seeing	(lilac)
sekte	(sect)	cement	(cement)
sheriff	(sheriff)	schlemiel	(wally)
skelter	(go-kart)	skelet	(skeleton)
venster	(window)	verdriet	(grief)
vogel	(bird)	verniss	(varnish)

\*as in “t geniep” (“on the sly”).

## APPENDIX B

The prosodic properties, that is, the duration of the syllables and the duration, pitch (F0), intensity and vowel quality (distance from the center of gravity in the normalized logarithmic F1-F2 plane of the vowels; see Fear et al., 1995 for a comparable analysis), of WeakStrong and StrongWeak words are presented in Table B.1. They have been analyzed by means of repeated measures ANOVAs with the within factor Stress (weak or strong syllable) and the between factor Stress Pattern (WeakStrong or StrongWeak). These factors showed an interaction for syllable duration,  $F(1,38) = 87.46$ ,  $p < .0001$ , because the weak syllable in WeakStrong words was shorter and the strong syllable longer than in StrongWeak words, in which both syllables were about equally

long. This effect was not due to differences in the duration of the vowel nucleus of the syllables. The latter showed only a main effect of the within factor stress, because weak syllables were shorter than strong syllables,  $F(1,38) = 31.91, p < .0001$ . For the pitch and intensity of the vowel, all main effects and the interactions were significant. These effects were due to the fact that weak syllables in StrongWeak words are of lower pitch and inten-

sity compared to all other syllables, interaction effects for pitch:  $F(1,38) = 227.15, p < .0001$  and intensity:  $F(1,38) = 140.17, p < .0001$ . Finally, weak vowels were indeed reduced when compared to full vowels, that is, the distance from the center of the F1-F2 plane was larger for strong than for weak vowels,  $F(1,38) = 41.80, p < .001$ , and eccentricity was smaller for WeakStrong than for StrongWeak words,  $F(1,38) = 5.82, p < .05$ .

**Table B.1.** Prosodic Features (Mean  $\pm$ SD) of the Syllables and Vowels in StrongWeak and WeakStrong Words

	StrongWeak		WeakStrong	
	Strong	Weak	Weak	Strong
Syllable				
duration (ms)	341 $\pm$ 62	386 $\pm$ 81	243 $\pm$ 52	531 $\pm$ 79
Vowel				
duration (ms)	159 $\pm$ 69	99 $\pm$ 26	85 $\pm$ 18	140 $\pm$ 44
Pitch				
(F0 in Hz)	133 $\pm$ 8	89 $\pm$ 5	109 $\pm$ 21	125 $\pm$ 8
Intensity				
(RMS in AU)	2.8 $\pm$ 0.3	1.5 $\pm$ 0.3	3.0 $\pm$ 0.5	2.6 $\pm$ 0.4
Vowel quality				
(eccentricity of z-scores)	1.7 $\pm$ 0.6	1.1 $\pm$ 0.7	0.6 $\pm$ 0.3	1.5 $\pm$ 0.6

Note: RMS = root mean square; AU = arbitrary units.