ERP correlates of focus accentuation in Dutch

Diana V. Dimitrova [1,2], *Laurie A. Stowe*[1,2], *Gisela Redeker*[1], *John C.J. Hoeks*[1,2]

[1] Center for Language and Cognition, University of Groningen, The Netherlands
[2] NeuroImaging Center Groningen, University Medical Center Groningen, The Netherlands
d.dimitrova@rug.nl, l.a.stowe@rug.nl, g.redeker@rug.nl, j.c.j.hoeks@rug.nl

Abstract

The present ERP study investigated the on-line interaction of prosody and information structure in Dutch. More specifically, we looked into how pitch accents, which are either congruous or incongruous with respect to the discourse context, are processed. Our results show that listeners process prosodic information immediately and check it for congruity within the discourse. Inappropriate prosody elicited right lateralized centro-parietal negativity (N400) and a late positivity (P600) for superfluous accents on background elements. In contrast, missing accents on focused elements triggered a late positive effect (P600). We suggest that although unexpected accents are identified faster than unexpected missing accents, both types of prosodic mismatch are re-interpreted and integrated in the discourse.

1. Introduction

Prosody comprises an impressive and complex interplay of melody and rhythm in language. Often referred to as speech melody, prosody carries universal extra-linguistic information (i.e., emotions) but may also have language-specific linguistic functions. For Germanic languages such as Dutch, distinctions in the information structure of utterances (focus-background) as well as in the information status of discourse referents (given-new) are expressed by means of prosody [1]. Dutch speakers assign the highest prosodic prominence to the most informative part of the utterance, i.e., the focus constituent. Less informative or repeated information that belongs to the background usually does not receive an accentual marking [2].

The perception of prosodic prominence in Dutch follows the regularities of pitch accent assignment in spoken language: Listeners reject focus elements as unacceptable if they do not bear a pitch accent [3]. In contrast, prosodic violations such as superfluous accents on background elements were found to be perceptually more acceptable [3]. These findings reveal that listeners judge the appropriateness of pitch accent placement based on the linguistic functions of prosody in Dutch – namely, the marking of focus information.

Despite the large number of behavioral studies on prosody, little is known about the neural mechanisms of prosody processing, especially in the absence of an explicit acceptability judgment. Because prosodic functions interact simultaneously with multiple levels of linguistic structure (i.e., phonology, semantics, syntax), the investigation of the role of accentuation poses a complex challenge.

Most neuro-imaging studies so far have concentrated on the neural correlates of other grammar levels such as semantic and syntactic processes [4]. For instance, semantic processing is reflected in a negative fluctuation of the Event-Related Potentials waveforms (ERPs), which peak around 400 ms after stimulus onset. The N400 component reflects word meaning integration and violations of semantic expectations in the context, among various other non-linguistic processes. A late positive component, the P600, has been attributed to the processing of syntactic structure and to general processes of reanalysis and re-interpretation. As for prosody, various and often conflicting ERP components have been suggested as correlates of prosody processing [5, 6, 7, 8, 9]. The current study aims to provide insights in the time course and electrophysiological nature of prosody integration in Dutch.

2. Neural correlates of prosody

The preferences for pitch accent placement reported in early behavioral studies [3] have been found to reflect the processing of prosody in neuro-imaging experiments as well. In an ERP study with question-answer pairs in German, [5] have shown that focus elements with *correctly* assigned pitch accents trigger a late positive component around 500 ms after target onset, the Closure Positive Shift (CPS). Initially reported as a correlate of prosodic phrasing [6], the CPS in this study was interpreted as a correlate of expected and appropriately accented focus information. By contrast, focus elements without a corresponding pitch accent evoked the N400 component that represents semantic integration difficulties. Surprisingly, superfluous accents on background elements did not elicit differences in the ERP waveforms [5]. Yet different correlates of prosody processing have been provided by [7]: All types of prosodic violations in the context were found to elicit the CPS, with focus violations being behaviorally easier to identify than background violations.

A different experiment [8] tested whether expectations for upcoming pitch accents are created on-line in German. Instead of using questions, expectations for pitch accents were triggered by a focus particle (i.e., *sogar [even]*) [8]. Surprisingly, both types of prosodic mismatch (i.e., unaccented focus and accented background) elicited the N400 component in medial sentence position. The authors argue that the processing of pitch accents might depend on the sentence position of the corresponding elements rather than on the type of prosodic violation (missing vs. superfluous pitch accents).

Further evidence for the importance of sentence position of prosodic mismatches comes from [9]. The ERP study compared the processing of focus accentuation in medial and final position in French dialogues. The authors showed that pitch accents are processed based on their sentence position (medial vs. final). Missing accents on focus constituents as well as superfluous accents on background information in medial position elicited positive fluctuations in the ERP, interpreted as the P300 component. This early surprise-related positivity is argued to represent a reaction to the surprising accentual pattern in medial position in [9]. In contrast, inappropriate accents in final position were to evoke the N400 component reported in earlier studies [5]. According to the authors, surprise effects arise only in medial position due to mismatched expectations. In sentence-final position, the

Table 1: *Experimental conditions. Pitch accents are capitalized. The analysis is performed on targets in medial position (in* **bold***).*

	Question = Context		Condition	Answer
Dutch	**Q:** *Heeft de club een premie (of een boete) aan de speler of aan de trainer gegeven?* **A:** *Ze hebben een premie aan de speler gegeven.*			
1. OBJ focus	Did the club give a bonus or a fine to the player?	1A	*congruous accent*	They gave a **BONUS** to the player.
		1B	*incongruous lack of accent*	They gave a **bonus** to the PLAYER.
2. PP focus	Did the club give a bonus to the player or to the trainer?	2A	*congruous lack of accent*	They gave a **bonus** to the PLAYER.
		2B	*incongruous accent*	They gave a **BONUS** to the player.

negativity may be an instance of the Contingent Negative Variation (CNV) for the anticipation of upcoming expected violations.

In sum, listeners use contextual information on-line for the prediction of the presence and position of pitch accents in the upcoming utterance. However, existent studies present conflicting results regarding the role of contextual information for the appropriateness of pitch accents (i.e., missing and superfluous accents are processed differently in [5] but identically in [8, 9]). In addition, prosodic mismatches appear to trigger various distinct ERP components and their processing is claimed to be determined by sentence position [9]. In the present experiment, we tested how Dutch listeners process pitch accents which vary in their contextual congruity.

3. The Experiment

The current experiment follows the paradigm introduced in [9].

3.1. Participants

Thirty-four native speakers of Dutch (age 18-29; 13 male) were paid to participate in the ERP study. All participants were right-handed and none of them had any neurological, psychiatric, hearing or language impairment. Informed written consent was obtained from all participants. Five participants were discarded from further analysis because data loss exceeded 40 % in one or more electrodes in at least one of the experimental conditions. The analysis was performed on twenty-nine participants' data.

3.2. Stimuli

Experimental stimuli were short dialogues by a male and a female speaker and recorded in an acoustically shielded studio at the University of Groningen. In order to avoid any unnatural emphasis during stimulus production, speakers were not instructed about the purpose of the experiment. As a result, stimuli were produced with a rather high speech rate. Special care was taken to ensure that experimental stimuli did not contain any disruptions (e.g., hesitations or phrase boundaries). All recorded sentences were normalized in loudness.

Experimental stimuli were 120 dialogue items that were recorded as congruous dialogues in two variants each (n=240): Focus on the object and focus on the prepositional phrase (see Table 1). All dialogues were duplicated (n=480) and all questions and answers were cut out from the original recordings. Questions and answers were then re-combined in order to create a *prosodically congruous* (accented (1A) and unaccented (2A)) and a *prosodically incongruous* (accented (1B) and unaccented (2B)) version of each dialogue. All 480 dialogues were divided in four lists of 120 items using the Latin square design. In each list, there were 30 items per condition (4 x 30), presented in a pseudo-randomized order. None of the participants listened to the same question-answer pair more than once and there were no more than two successive dialogues of the same condition set per list.

For each dialogue, a question introduced a choice to be made between two contrastive elements. In the answer, one of the elements was chosen. In half of the questions of each list, the contrastive focus was on the object (OBJ focus, n=60, *Did the club give a bonus or a fine to the player?*). In the other half of the list, the contrastive focus was on the prepositional phrase (PP focus, n=60, *Did the club give a bonus to the player or to the trainer?*). In the answers, the contrastive focus elements received either a congruous matching pitch accent (n=30 for OBJ; n=30 for PP) or an incongruous non-matching pitch accent (n=30 for OBJ; n=30 for PP). All experimental conditions are presented in Table 1.

All sentences were matched for sentence length (measured in number of words), sentence plausibility, sentence structure, and average word frequency for target and non-target words. Word frequency was taken from the CELEX corpus [11]. Special attention was paid to the phonological form of the stimuli: All target words (both OBJ and PP) had a lexical stress on the first syllable and were thus matched for association with the pitch accent. In addition, an offline pretest (96 participants) measured the expectedness of each target and non-target element. All these factors were taken into account when attributing the items to the four conditions. In this paper, we report only results for focused objects in medial position.

3.3. Acoustic analysis

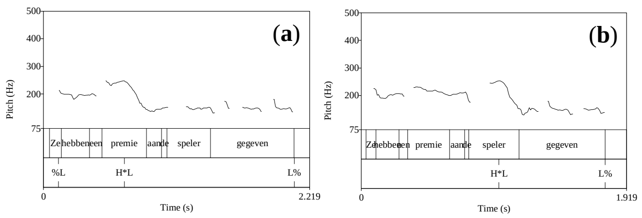
For all sentences, we measured the acoustic realization of direct objects with respect to the presence or absence of accent (see Table 2).

Table 2*: Acoustic data for contrastive object targets.*

means and SD for	accented OBJ	unaccented OBJ
duration OBJ	359 ms (SD=63)	273 ms (SD=54)
sentence length	1788 ms (SD=170)	1719 ms (SD=188)
f0 min	165 Hz (SD=27)	184 Hz (SD=16)
f0 max	253 Hz (SD=20)	212 Hz (SD=16)

Accented objects had a longer mean duration (by 86 ms) and a higher fundamental frequency (by 41 Hz) than unaccented ones. Pitch accents had the typical H*L contour for focus accents in Dutch, whereas background information was not prosodically prominent (see Figure 1).

Figure 1: Example of pitch contours: (a) focus accent on object; (b) focus accent on prepositional phrase.



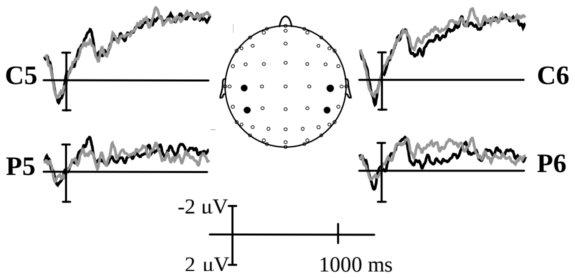


*Figure 2: **Accented direct objects**. Black line: Congruous accented focus (1A), grey line: Incongruous unaccented background (i.e., superfluous accents) (2B).*

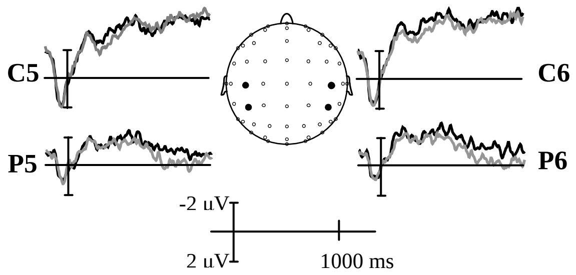


*Figure 3: **Unaccented direct objects**. Black line: Congruous unaccented background (2A), grey line: Incongruous unaccented focus (i.e., missing accents) (1B).*

3.4. Procedure

After applying the electrodes, participants were seated in an electrically shielded room in front of a computer. Following a trial session, participants completed the actual experiment which was divided in two blocks of 60 dialogues with an approximate duration of 15 minutes per block. Stimuli were presented auditorily via loudspeakers. To minimize blinking, participants were instructed to fixate a black cross on a grey computer screen while listening to the dialogues.

Each trial began with a fixation cross, followed by 100 ms of silence, a question, 500 ms silence, an answer, and 1200 ms of silence. After the answer, in some trials a word appeared on the screen and participants had to judge by key press whether its' meaning was related to the preceding dialogue. The comprehension task was used to guarantee attention to the stimulus materials. After the response (or otherwise after the end of the answer), a blinking period of 2000 ms was induced.

3.5. EEG recording and analysis

The EEG was recorded from an elastic cap with 64 Ag/AgCl electrodes according to the International 10-20 system (Electro Cap International). Electrodes were referenced online to the average of all electrodes and re-referenced offline to the algebraic average of left and right mastoid electrodes. Blinks as well as vertical eye movements were monitored via electrodes placed below and above the left eye. Horizontal eye movements were recorded from electrodes at the left and right cantus of each eye. Impedances were kept below 5 Ω. The EEG was digitalized online with a sampling frequency of 250 Hz and filtered offline with a band-pass filter of 0.01 – 30 Hz.

Trials containing movement artifacts, ocular artifacts or electrode drifts were rejected. We conducted an individual channel analysis and discarded all participants with more than a 40% data loss for one or more of selected electrodes in at least one single condition. As a result, five participants were not included. ERPs were time-locked to the onset of the target word in the answer which corresponds to the onset of the target's accented syllable. We performed a baseline correction relative to a 200 ms pre-stimulus baseline and conducted average waves in a time window ranging from 200 ms prior to 1300 ms post the onset of the critical word. Processing costs were always measured by comparisons of conditions with physically identical stimuli (accented 1A-2B or unaccented 2A-1B).

3.6. Results

A visual inspection of the grand averages of all 29 participants revealed two time windows where the EEG signal showed differences between the conditions tested. We identified a negativity between 300-500 ms and a positivity between 700-1000 ms.

We computed repeated measures ANOVAs for both identified time windows: 300-500 ms and 700-1000 ms post stimulus onset. The statistical analysis was performed separately for lateral and midline electrodes. For lateral electrodes, ANOVAs were calculated with the four within-subject factors *Accent* (accented vs. unaccented element), *Congruity* (contextually congruous prosody vs. contextually incongruous prosody), *Topography* (frontal vs. central vs. parietal regions), and *Lateralization* (left vs. right sites). For ANOVAs on midline electrodes, all factors were included except for *Lateralization*. All reported *p* values are adjusted with the Huynh-Feldt correction for nonsphericity. Single electrodes were grouped in Regions of Interests (ROIs) for *midline* electrodes: frontal (Fpz, AFz, Fz), central (FCz, Cz, CPz), parietal (Pz, POz, Oz), and for *lateral* electrodes: left frontal (Fp1, AF3, AF7, F3, F5, F7), right frontal (Fp2, AF4, AF8, F4, F6, F8), left central (FC3, FC5, C3, C5, CP3, CP5), right central (FC4, FC6, C4, C6, CP4, CP6), left parietal (P3, P5, P7, PO3, PO7, O1), right parietal (P4, P6, P8, PO4, PO8, O2). All statistical tests are performed on mean voltage values.

3.6.1. Results in the 300 – 500 ms time window

For lateral electrodes, we observed a main effect of the factor *Accent* ($F(1,28)=6.207$, $p<.05$) which was qualified by a significant interaction between *Accent*, *Congruity* and *Lateralization* ($F(1,28)=4.416$, $p<.05$). Follow-up analyses showed a significant interaction of *Congruity* and *Lateralization* for *accented* elements only ($F(1,28)=11.807$, $p<.001$) but no such interaction for *unaccented* elements ($p=.57$). The effect of *Congruity* was significant only over the right sites for *accented* elements ($F(1,28)=4.8$, $p<.05$), but not over left sites ($p=.66$). In other words, incongruously accented elements (i.e., superfluous accents on background elements) elicited more negative fluctuations of the EEG on right sites as compared to congruously accented elements (i.e., accents on focus elements). The negative effect is presented in Figure 2. For midline electrodes, there was a main effect of *Accent* ($F(1,28)=9.726$, $p<.01$) that did not reveal any further significant interactions.

3.6.2. Results in the 700 – 1000 ms time window

For lateral electrodes, the factor *Congruity* was involved in a marginal three-way interaction with *Accent* x *Lateralization* ($F(1,28)=2.912$, $p=.09$). Follow-up analyses examined a significant interaction of *Congruity* x *Lateralization* for accented elements ($F(1,28)=5.258$, $p<.05$) but not for unaccented elements ($p=.91$). Post-hoc tests did not reveal any differences between left and right sites: Incongruously accented elements (superfluous accents) triggered positive waveforms while there was no effect for missing accents. This positivity may belong to the P600 family and is displayed in Figure 3. In addition, there was an *Accent* x *Topography* interaction (F(2,56)=9.292, p<.01) which revealed a marginal effect of *Accent* for parietal regions (p=.06) but not for frontal or central regions (all p's>.1). For midline electrodes, there was a marginal effect of *Congruity* (F(1,28)=3.667, p=.06) indicating that incongruous elements (both accented and unaccented ones) triggered more positive waveforms than congruous elements. This late positivity may be an instance of a P600 for all incongruous elements. In addition, there was an *Accent* x *Topography* interaction (F(2,56)=3.378, p<.05), which failed to show a significant effect of *Accent* for one of the tested ROIs.

4. Discussion

The present study investigated whether Dutch listeners process prosodic information on-line according to its congruity in the discourse context and irrespective of the performance of a prosodic task. Furthermore, the study tested whether different types of prosodic violations such as missing and superfluous accents are manifested in distinct neural correlates.

Our results provide evidence for the impact of discourse on the processing of pitch accents in Dutch. Although listeners in the current experiment were not involved in a prosody judgment task, they processed incongruous prosody differently from congruous prosody. Furthermore, listeners appeared to process prosodic violations depending on the type of prosodic mismatch. That is, superfluous accents elicited different neural correlates as compared to missing accents.

Unlike [9] and similar to [8], a right-lateralized negativity, presumably an N400, was found for the processing of background elements which received a contrastive pitch accent. In addition, incongruously accented elements elicited a late positive component that might be an instance of a P600. We suggest that the N400 for superfluous accents reflects the processing of contextually unexpected accentuation. In a later time window, the initially unexpected superfluous accent triggers processes of re-analysis and re-interpretation of the focus structure, which in turn is indicated by the P600. Our results question earlier assumptions [3, 5] that listeners accept superfluous accents because they would not attend to their prosody. Our results show that listeners are sensitive to unexpected prosodic realizations of background information.

Moreover, listeners have predictions about the congruous accentuation of focus constituents as well. In our study, missing accents on focus elements elicited late positivities, probably belonging to the P600 family for processes of re-analysis. Our results are in line with [7] who report a CPS effect for the processing of missing (and superfluous) accents in German. However, we did not find an N400 effect for missing accents on focus elements as [5, 8] did. The N400 has been argued to reflect semantic integration difficulties triggered by mismatched expectations of accented focus. Several inconsistencies in the previous experiments such as time-locking at sentence onset [5], uncontrolled plausibility and frequency for target words [5,8] as well as uncontrolled lexical stress position [5, 8] represent a challenge for their interpretations of the negative effect. We suggest that missing accents in our study did not trigger early expectation negativities because the accent is not completely missing; it is rather the contrastive prominence of the accent that is not present. As indicated by the P600, listeners become aware of the lack of prosodic prominence in a late time window.

In other words, the lack of strong prominence on focus elements may initially not be interpreted as a mismatch of prosodic expectations. This assumption is further supported by the fact that, in contrast to the majority of previous studies, listens in the present experiment did not perform a prosodic task. As a result, listeners' attention was not directly driven to the appropriateness of prosody but rather to the semantic content of the dialogues as a whole. The nature of the task has been shown to affect the neural correlates of prosody [10]: A P800 indicating the processing of prosodic mismatches was only elicited if listeners performed a prosodic task but not a semantic task. We believe that listeners in our study may have paid less attention to the prosodic realization of expected structures such as focus elements and may have initially 'filled in' accents where they were expected. By contrast, accents that were unexpected by the context may have been identified immediately. Hence, even though unexpected information may be identified faster than expected information, listeners process and re-interpret both types of mismatches in order to create a coherent discourse representation.

5. Conclusions

Listeners of Dutch directly use contextual information during spoken language comprehension even if they do not perform a prosodic task. They process prosodic information on the basis of contextual expectations. That is, contextually unexpected prosodic prominence is detected and integrated immediately as revealed by the bipolar N400-P600 pattern for superfluous accents. Contextual expectations for focus elicit late processes of re-analysis of missing accents (P600). Our study provides evidence that listeners process contextually unexpected prosodic realizations of focus and background elements during semantic-task oriented spoken language comprehension.

6. References

[1] C. Gussenhoven, *The Phonology of Tone and Intonation*. Cambridge: Cambridge University Press, 2004.

[2] M. Swerts, E. Krahmer, and C. Avesani, "Prosodic marking of information status in Dutch and Italian: a comparative analysis," *Journal of Phonetics*, vol. 30, pp. 629-654, 2002.

[3] S. G. Nooteboom, and J.G. Kruyt, "Accents, focus distribution, and the perceived distribution of given and new information: An experiment," *Journal of the Acoustic Society of America*, vol. 82, pp. 1512-1524, 1987.

[4] M. Kutas, C.K. van Petten, and R. Kluender, "Psycholingusitics Electrified II," in *Handbook of Psycholinguistics, (2nd ed.), M.J.* Traxler and M.A. Gernsbacher, Eds. Amsterdam: Academic Press, pp. 659–724, 2006.

[5] C. Hruska, K. Alter, K. Steinhauer, and A. Steube, "Misleading dialogues: Human's brain reaction to prosodic information," in *Orality and Gestures. Interactions et comportements multimodaux dans la communication*, Ch. Cavé, I. Guaïtella, and S. Santi, Eds. Paris, pp. 425-430, 2001.

[6] K. Steinhauer, K. Alter, and A.D. Friederici, "Brain potentials indicate immediate use of prosodic cues in natural speech processing," *Nature Neuroscience*, vol. 2, pp. 191-196, 1999.

[7] U. Toepel, A. Pannekamp, and K. Alter, "Catching the news: Processing strategies in listening to dialogues as measured by ERPs", *Behavioral and Brain Functions*, vol. 3(53), 2007.

[8] S. Heim and K. Alter, "Focus on focus: The brain's electrophysiological response to focus particles and accents in German," in *Interface and Interface Conditions*, A. Späth, Ed. Berlin: Walter de Gruyter, pp. 277-298, 2007.

[9] C. Magne, C. Astesano, A. Lacheret-Dujour, M. Morel, K. Alter, and M. Besson, "On-line procesing of 'pop-out' words in spoken French dialogues," *Journal of Cognitive Neuroscience*, vol. 17, pp. 740-756, 2005.

[10] C. Astesano, M. Besson, and K. Alter, "Brain potentials during semantic and prosodic processing in French," *Cognitive Brain Research*, vol. 18, pp. 172-184, 2004.

[11] H.R. Baayen, R. Piepenbrock, and H. van Rijn, *The CELEX lexical database*. CD-ROM. Philadelphia, PA, 1993.