Tilburg University

Event-related potentials as indirect measures of recognition memory

van Hooff, J.C.; Brunia, C.H.M.; Allen, J.J.

Published in:
International Journal of Psychology

Publication date:
1996

Link to publication in Tilburg University Research Portal

Citation for published version (APA):
van Hooff, J. C., Brunia, C. H. M., & Allen, J. J. (1996). Event-related potentials as indirect measures of recognition memory. *International Journal of Psychology*, *21*, 15-31.

General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

Take down policy
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Download date: 12. May. 2021

ELSEVIER

International Journal of Psychophysiology 21 (1996) 15–31

INTERNATIONAL
JOURNAL OF
PSYCHOPHYSIOLOGY

Event-related potentials as indirect measures of recognition memory

Johanna C. van Hooff [a,b,*], Cornelis H.M. Brunia [b], John J.B. Allen [c]

[a] *Co-operation Center Tilburg and Eindhoven Universities, Tilburg, The Netherlands*
[b] *Tilburg University, Tilburg, The Netherlands*
[c] *University of Arizona, Tucson, AZ, USA*

Received 10 April 1995; revised 26 September 1995; accepted 4 October 1995

Abstract

Event-related potentials (ERPs) were recorded during an auditory word-recognition task to determine whether they can be used as indirect measures of recognition memory, defined as the ability to differentiate learned from unlearned material when no overt recognition response from the subject is required. A modified version of the two-choice reaction time task developed by Allen, Iacono and Danielson (Allen et al., 1992) was used. In three recognition tasks, administered on two consecutive days, subjects were instructed to indicate recognition of recently learned words. These words were presented along with unlearned words and along with previously learned words which both required a non-recognition response. Recently learned target words as well as previously learned nontarget words elicited a centro-parietal positivity around 500–1000 ms post-stimulus. The size and onset of this late positivity (P300) were affected by the requirement of an overt recognition response. The results suggest that ERPs are sensitive to differences between learned and unlearned words, to some extent independently of the behavioral response. ERPs may therefore be used as indirect measures of recognition memory. In addition, because the present results held for stimuli presented in the auditory modality and because recognition indices were still observed after a one-day interval between learning and testing, this procedure might prove useful in various applications when the integrity of memory is in question.

Keywords: Auditory ERPs; P300; Recognition; Memory; Indirect measures

1. Introduction

Event-related Potentials (ERPs) could prove to be a useful way to assess memory for events which subjects do not explicitly report remembering. Persons may, for example, report that they do not remember previous events because they are unable (e.g., persons with amnestic syndromes) or because they are unwilling (e.g., malingering). Additionally, suboptimal conditions during initial presentation can produce impaired conscious recollection despite evidence of having encoded the stimuli (Eich, 1984; Kemp-Wheeler and Hill, 1988). ERPs may supplement traditional behavioral measures of (implicit) memory for several reasons. ERPs have a time-resolution on the order of milliseconds and can therefore provide temporal correlates of the stages of informa-

---

[*] Corresponding author. At: Physiological Psychology Section (P604), Tilburg University, PO Box 90153, 5000 LE Tilburg, The Netherlands. Tel.: +31-13-4662115; Fax: +31-13-4662370; E-mail: J.C.vHooff@KUB.NL

0167-8760/96/$15.00 © 1996 Elsevier Science B.V. All rights reserved
SSDI 0167-8760(95)00043-7

tion processing between stimulus and response. Their scalp topography may provide indications of which brain areas may underlie memory performance in a variety of tasks. Furthermore, compared to behavioral measures, ERPs are less vulnerable to effects of malingering (Rosenfeld et al., 1995) and can be more easily adapted to individual assessment procedures (e.g., Wasserman and Bockenholt, 1989; Allen et al., 1992). Finally, ERP recording does not necessarily require that subjects behaviorally respond, but only that they cognitively discriminate between experimental conditions, which may be valuable for studies on different patient populations in particular.

The present study examined whether ERPs are sensitive to memories for previously learned material, regardless of whether or not recognition is reported. The utility of ERPs as indirect measures of preserved or concealed memories has been demonstrated in several studies. Farwell and Donchin (1986, , 1991) and Rosenfeld et al. (1991) have shown in deception-detection paradigms that it was possible, on the basis of ERP responses, to correctly identify material for which subjects had knowledge, even when no behavioral indication of recognition was provided. These deception-detection studies were administered in the format of an oddball paradigm, i.e., the critical 'guilty' items appeared less frequently than did the non-relevant items. The studies were based upon the assumption that P300s, or P300-like components, can be reliably elicited by stimuli which stand out as distinct and which possess special significance for the subject. In a similar fashion to these studies, Allen et al. (1992) developed an ERP memory- assessment procedure that validly differentiated learned from unlearned material, independently of overt recognition responses. In contrast to the studies of Farwell and Donchin (1991) and of Rosenfeld et al. (1991), Allen et al. (1992) used stimuli that were not associated with 'crimes' or antisocial acts, but rather involved simple list learning. Subjects were instructed to respond affirmatively to items from a recently learned list of six words and to respond negatively to items of an earlier learned list of six words and to 30 novel words. They were encouraged to deceive the experimenters by hiding the fact that they had knowledge about the earlier learned list. ERP results from three experiments, differing in motivational manipulations, showed that recently learned words as well as previously learned words elicited a large positivity which was generally absent for the novel words. These results suggest that ERP-based procedures may have assessment applications beyond lie-detection, extending to memory assessment in a variety of domains. For example, in a case study of prosopagnosia (Renault et al., 1989), it has been demonstrated that a similar recognition oddball task may hold potential for clinical assessment protocols. Renault et al. (1989) reported that the P300 component was sensitive to familiar faces despite the prosopagnostic patient's inability to consciously recognize them. Likewise, such ERP-based procedures may be valuable for evaluation of memory functions with different types of amnesia. It should be similarly possible to study preserved memories for stimuli which are presented below detection threshold (subliminal perception) or during periods of minimal encoding (sleep or anesthesia).

ERPs are frequently studied in relation to memory processes, revealing a close relationship between late positive components and both encoding and retrieval processes. For example, studies interested in the initial encoding of information have shown that ERPs elicited by words that were later recalled or recognized were more positive around 400 ms after word-onset than those of words that were later forgotten (Fabiani et al., 1986; Paller and Kutas, 1987). More importantly for the present study, recognition studies have demonstrated that ERPs elicited by correctly recognized old words were characterized by a large positivity, starting around 200–300 ms following stimulus-onset and persisting over several hundreds of milliseconds, which was largely absent for new words (Friedman, 1990; Neville et al., 1986; Smith and Guster, 1993). It is suggested that this positivity may index the involvement of processes functionally related to the ability to discriminate old items from new items, although it remains unclear to what extent these processes are independent from overt behavioral responses. Sanquist et al. (1980) and Neville et al. (1986) analyzed ERPs to correctly and incorrectly identified words finding late positivities only for recognized old words, concluding that the ERP effects were dependent on both the repetition of words as well as on explicit recognition. However, Leiphart et al. (1993) reported that ERP effects of prior learning were affected by response correctness in a yes-no recognition task but not in a perceptual identification task. This may suggest that in particular conditions larger positivities could be found for old versus new words without a correct behavioral classification or explicit recognition response. This latter suggestion is supported by experiments showing ERP repetition effects when subjects were engaged in tasks other than yes-no recognition, such as lexical or semantic decision. In these tasks, involving repetition of words after different numbers of intervening items, repeated words provoked more positive going ERPs compared to first presentations (Rugg, 1987; Rugg et al., 1988; Bentin and Peled, 1990). Because subjects did not have to make decisions about whether they recognized previously presented items, it is suggested that memory search and overt recognition processes did not contribute to the elicited positivity. Smith and Guster (1993) have emphasized the multicomponent nature of the ERP memory effects associated with different types of retrieval processes, one being more related to conscious recollection and one being more related to repetition. Additionally, as has been described earlier, recollection or recognition processes appeared to be difficult to suppress at the ERP level (Allen et al., 1992; Farwell and Donchin, 1991; Rosenfeld et al., 1991) and may even occur without phenomenological awareness (Renault et al., 1989).

The present study is modelled after that of Allen et al. (1992). It is designed to extend the utility of existing ERP memory assessment procedures to allow future investigations involving auditory presentation and longer delays between learning and testing. Our study had three specific objectives. First, in the studies of Farwell and Donchin (1991) and Allen et al. (1992), subjects were instructed to deceive the experimenters by hiding the fact that they had knowledge about a mock crime or an earlier learned list of words respectively. Although in these studies it was possible to identify familiar material on the basis of ERP responses, such a deception instruction limits the utility of the procedure and would prove difficult in experiments that aim to reveal implicit memories of which subjects are unaware. Moreover, it could be hypothesized that the instruction to intentionally withhold knowledge has been largely responsible for the reported results, because it might provide extra significance to the critical items (those that had to be kept hidden). The present study examined whether ERPs are still able to differentiate learned from unlearned material without a deception instruction.

Second, almost all ERP deception and memory studies have reported results using visually presented material. Because many applications may require an auditory format (e.g., presentation during sleep or anesthesia, presentation to visually impaired populations) it is essential to examine whether comparable results can be obtained with auditorily presented material. Ferlazzo et al. (1993) have reported that an old-new ERP effect could be observed during an auditory recognition task using two-syllable words. Whether comparable effects on auditory ERPs could also be obtained when no explicit recognition response is given is investigated in the present study.

Third, the time course over which an ERP-based procedure can be effective is unknown. The studies to date have typically used a short interval from learning to testing. Other applications may require a longer interval, e.g., from learning trials during sleep to testing after waking several hours later, from learning trials during anesthesia to testing several hours or even days later after recovery. Because memory traces could be expected to decline over time, the present study investigated whether ERP effects could still be found after a one-day interval between learning and testing.

In our study, subjects learned different lists of six words and then participated in a yes-no recognition task. During the recognition tasks, unlearned words were presented together with recently learned words and previously learned words. Only the recently learned words (learned targets) required a YES-response to indicate recognition. Both the unlearned words (unlearned nontargets) and the earlier learned words (learned nontargets) required a NO-response. The tasks were constructed such that learned words appeared infrequently against a background of unlearned material. The rationale was that if subjects had learned two or three lists of words, of which only the most recent one required an overt recognition response, subjects should nonetheless recognize all previously learned lists of words. ERPs, and in particular the P300, were hypothesized to be sensitive to such recognition. The amplitude of the P300 is inversely proportional to stimulus-probability and is directly related to stimulus-relevance (Johnson, 1986). Consequently, if learned words were perceived as a distinct and rare class of stimuli, it is likely that they would elicit a P300-like component. If a P300 could also be found for the learned words which needed no overt indication of recognition, this would support the hypothesis that ERPs can serve as indirect measures of recognition memory. Furthermore, longer RTs were expected for the learned targets, because their infrequent occurrence would produce a strong bias to respond negatively. Longer RTs were also expected for the learned nontargets, because of a likely response-conflict (i.e., needing to indicate NO to a learned item). After a one-day interval, ERP and RT effects for the words which were learned the day before were expected to decrease in magnitude as a result of declined memory traces and reduced stimulus significance.

2. Method

2.1. Subjects

Twenty volunteers (11 female 9 male students) participated in this study. They were paid *ƒ*7.50 (approximately $4.50 US) per hour. Two females were left-handed whereas all other subjects were right-handed. They were aged from 18–35 with a mean of 23.7 years (SD ± 4.7). All subjects were native Dutch speakers (one bilingual) and had no hearing impairments. The main purpose of the study and the complete procedure were not revealed to subjects until after they had completed their participation.

2.2. Stimuli and apparatus

Subjects were seated in a comfortable chair in a dimly illuminated, sound-attenuated, and electrically shielded chamber. In consecutive study and recognition phases, spoken words were presented to the subjects through stereo headphones at a comfortable listening level (approximately 60 dB). The words were recorded by a native male speaker of Dutch onto a Digital Audio Tape. Before storage on an IBM-type 486 Personal Computer (sample frequency 20 kHz, 12 bit resolution), each word was examined, and when necessary amplified, using a speech editor. The PC was provided with a LabMaster AD/DA board so that it could be used for stimulus presentation, experimental control and data acquisition. All

Table 1
Overview of the word categories, their mean presentation time (ms) and mean frequency of occurrence for subjects in groups A and B

Category	Presentation Time	Frequency of Occurrence	Test 1		Test 2		Test 3	
			A	B	A	B	A	B
1. Animals	535	15–2	X		X	O	X	
2. Landscape elements	528	26–4	X		X	O	X	
3. Parts of the body	418	45–5	X		X	O	X	
4. Clothing	497	17–5		X	O	X		X
5. Kitchen utensils	476	26–6		X	O	X		X
6. Furniture	460	48–6		X	O	X		X
7. Food	502	15–1			O	O		
8. Forms of transportation	469	42–9			O	O		
9. Weather elements	512	26–3					O	O
10. Materials	540	18–7	O	O			O	O
11. Means for storage	468	23–6	O	O			O	O
12. Green	508	38–4	O	O			O	O
13. Parts of the house	473	50–8	O	O			O	O

X indicates the learned categories, O indicates the unlearned categories. For any single subject, one of the three X-marked categories has been learned in Test 1, two of the three X-marked categories has been learned in Test 2 (one before Test 1 and one before Test 2), and in Test 3 all three X-marked categories have been learned (one before Test 1, one before Test 2 and one before Test 3). The mean occurrence frequency for each category is calculated from a frequency-list composed by Uit den Boogaart (1975). The first number indicates the mean frequency in written Dutch based upon counting of 600 000 words from five different types of printed language. The second number indicates the mean frequency of spoken Dutch with an extent of approximately 120 000 words.

words were one-syllable nouns beginning and ending with a consonant and comparable in presentation time, intensity and frequency of occurrence in the Dutch language (see Table 1). The words were divided into 13 different semantic categories of six words each. All categories contained three- to six-letter words of which at least four had different vowels or pairs of vowels. In the study as well as in the test phases, the stimulus onset asynchronies (SOA) were approximately 2 s, varying between 1850 ms and 2300 ms. Each test consisted of 12 blocks with different pseudo-random word orders. In all blocks, each word was presented only once. Because the words were identical for each test-block, all words were repeated 12 times. Words from the learned lists were never presented successively or as the last word. Because of possible influences of an orienting response at the beginning of the test blocks, each block began with three out of a set of six non-relevant words, which were discarded from further analysis. Two small tubes with push-buttons on top served as response buttons which recorded subject responses and latency of responses.

The Electroencephalogram (EEG) was recorded by non-polarizing Ag-AgCl electrodes which were fixed with collodion to the scalp at Fz, Cz, Pz and two lateral positions C5 and C6, located midway between T3-C3 and T4-C4 respectively. Linked mastoids served as the reference. Interelectrode impedances were less than 3 kΩ. To monitor eye-movements, EOG was recorded by three pairs of electrodes, two pairs for vertical movements (above and below each eye) and one pair for horizontal movements (at the outer canthi). EEG and EOG signals were amplified and written out on paper by a 14-channel Nihon Kohden electroencephalograph (timeconstant 6.6 s, low-pass filter −3 dB cut-off at 35 Hz). A calibration pulse (peak-peak amplitude 100 μV) was recorded before and after each test. The amplified signals were digitized on-line at 125 Hz (resolution 12 bit). EEG data were corrected for eye-movement artifacts off-line before further analyses (Van den Berg-Lenssen et al., 1989).

2.3. Procedure

Because this experiment was also intended to select stimulus material for further study, subjects were assigned alternately to two separate groups (A and B), which differed only in which particular words were learned. Table 1 gives an overview of the learned and unlearned categories used in each test. Group A ($N = 10$) had to learn in three separate study phases the categories "animals", "landscape-elements" and "parts of the body" and group B ($N = 10$) the categories "clothing", "kitchen utensils" and "furniture". Within each group, the order of the three categories to be learned were counterbalanced between subjects. Although complete counterbalancing was not performed across all lists, this arrangement made it unlikely that any obtained effects were due solely to stimulus-specific factors.

Subjects were first instructed to memorize six auditorily presented words (first study list) from one of the three categories to be learned. The words were presented repeatedly until the subjects could produce the list of words in the order presented and in reversed order after a 1-min break. In the next recognition test (Test 1), the words of the learned category were presented pseudo-randomly among words from four other unlearned categories. Because the learned words appeared infrequently ($p = 1/5$) compared to the unlearned words ($p = 4/5$) this recognition test was essentially an oddball task. The subjects' task was to respond as accurately and quickly as possible with the thumb of their dominant hand when they recognized the word presented as one of the words they had learned (YES-response) and with the thumb of their other hand when they did not (NO-response). To obtain a sufficient number of trials for a proper signal-to-noise ratio, the test was administered in twelve blocks with 1–2 min rest periods between blocks.

Subsequently, after a 15-min break, the subjects unexpectedly learned another list of six words (second study list) belonging to one of the two remaining categories to be learned. The learning procedure was the same as described for the first study phase. After learning this list, a second recognition oddball test was administered in which the words from the first and second study list were presented pseudo-randomly among words from five other unlearned categories, with the probability of occurrence being 1/7 for each of the two learned categories and 5/7 for the unlearned words. The subjects were instructed to give a YES-response for only the recently learned words and to give a NO-response for all other words. It was not explicitly mentioned to subjects that the words from the first study list were included in this test. Because the words from the earlier-learned study list did not require a YES-response, these are referred to as 'learned nontargets'. The recently learned words which had to be recognized are referred to as 'learned targets'. The second test also consisted of twelve blocks with 1–2 min rest periods between blocks. The next day, at the same time as the day before, subjects again learned six auditorily presented words (third study list) with the same criterion for learning. These were the words from the last category to be learned (from among the three, see Table 1). A third recognition test was then administered in which all three learned categories were pseudo-randomly intermixed with five unlearned categories, one previously unused category and four which had served as foils in the first recognition test (see Table 1). As a result, the learned words of the three study lists had a probability of occurrence of 1/8 each and the unlearned words had a probability of occurrence of 5/8. In this task, administered also in twelve blocks, the subjects had to give a YES-response for the recently learned words only (learned targets). NO-responses were required for all other words. It was not explicitly mentioned that the words which were learned the day before (learned nontargets, the items from study list 1 and 2) would be presented in this test.

Compared to the first test, the probabilities of the learned words requiring a YES-response to indicate recognition (learned targets) were smaller in the second and third recognition test (1/5 versus 1/7 and 1/8 for the first, second and third test respectively), whereas the probabilities of all learned words together (learned targets + learned nontargets) were somewhat larger in the second and third test (1/5 versus 2/7 and 3/8). Because these probabilities cannot be equal in all tests, we have chosen these ratios because they form a compromise solution. In addition, it should be noted that we were primarily interested in ERP differences within tests between the distinct classes of stimuli, rather than in ERP differences between tests. Furthermore, we did not have a specific purpose in re-using the unlearned categories of the first test in the third test, but this was a necessity because it was not possible to construct other matched categories in terms of presentation time and frequency of occurrence.

After the third recognition test, the subjects were asked to write down the words which were learned in all three study phases.

2.4. Data analysis

Although in the following analyses repeated measures analysis of variance designs were used, the multivariate test of significance (Jennings, 1987; Vasey and Thayer, 1987) was selected to circumvent the problem of violating the sphericity assumption. In each case, the approximate F-value associated with the multivariate test is reported.

Mean RTs for the words presented in the recognition tests were calculated for each study list and for the unlearned words. Mean RTs for all unlearned words were averaged together and treated as a single list. Responses faster than 250 ms after word-onset were considered to be guesses and were therefore excluded from statistical analysis, as were Misses and False alarms. To test possible training effects, a Multivariate Analysis of Variance (MANOVA) with repeated measures (SPSS-PC +) was carried out for the overall mean RTs. Separate MANOVAs were carried out for each test with study list as within subjects factor, supplemented with pairwise comparisons when main effects were found (the significance level was Bonferroni corrected). For the recall data, a MANOVA was carried out with study list as within subjects factor.

Eye-movement-corrected EEG signals were automatically screened for artifacts. Accepted signals were averaged for each channel time-locked to the onset of the stimuli. The total length of the averaging epoch was 2 s extending from 500 ms preceding word-onset to 1500 ms following word-onset. The 200 ms period preceding the onset of the stimulus was used for base-line correction. ERP waveforms were obtained for each study list separately. Those for the unlearned words were averaged together and treated as a single list. Only trials with correct responses were included in these averages. Peak amplitude of the late positivity was identified as the most positive value in the 450–950 ms post-stimulus interval, measured in respect to the average voltage of the 200 ms pre-stimulus baseline. This positivity is considered to correspond with, and is further referred to as, P300 because of its centro-parietal dominance, its sensitivity to stimulus probability and because it was elicited by relevant items in a categorical decision task. In order to assess the time-course of post-stimulus slow-potential shifts, mean amplitudes were determined for five consecutive periods of 150 ms each, starting 200 ms after word onset. This was done solely for the midline electrode positions. P300 peak amplitude, P300 peak latency and mean epoch amplitudes were used as dependent variables in statistical analysis.

For each test and ERP parameter, MANOVAs with repeated measures were carried out with study list and electrode position (midline electrode positions) and study list and hemisphere (lateral electrode positions) as within subject factors. When main effects were found, these MANOVAs were supplemented with Bonferroni corrected pairwise comparisons. To evaluate possible confounding effects of target probability and recurrence, MANOVAs were carried out on P300 amplitude and latency for learned targets and unlearned nontargets respectively with test as within subjects factor.

3. Results

3.1. Performance

Table 2 summarizes the performance data for each recognition test as a function of study list. Errors were made almost exclusively for the learned targets (Misses), indicating a strong response-bias in the direction of negative responses. The average percentage of errors across all categories was about 2.22% for all three recognition tests. Because of a decrease in mean RTs from the first to the third test it was evident that an effect of training was present ($F(2,18) = 6.43$, $p < 0.01$). As can be seen in Table 2, training influenced only the RTs of the NO-responses. When the RTs of the NO-responses were averaged and compared with those of the YES-responses, a significant Test × Response interaction was present ($F(2,18) = 4.01$, $p < 0.05$). The RTs for the NO-responses decreased from the first to the third test ($F(2,18) = 7.88$, $p < 0.01$) whereas the RTs for the YES-responses remained approximately the same.

In the first recognition test, RTs were signifi-

Table 2
Mean reaction times, errors and P300 peak amplitudes at Pz for each recognition test as a function of study list

		Unlearned	List 1	List 2	List 3
Test 1	**Response**	**NO**	**YES**		
	RT (ms)	582 (62)	627 (68)		
	Errors (%)	1.05 (2.1)	6.50 (3.71)		
	P300 amplitude (μV)	4.95 (2.24)	13.10 (5.31)		
	P300 latency (ms)	706 (115)	680 (95)		
Test 2	**Response**	**NO**	**NO**	**YES**	
	RT (ms)	547 (74)	595 (92)	628 (92)	
	Errors (%)	0.55 (1.3)	0.95 (1.5)	10.15 (6.3)	
	P300 amplitude (μV)	4.92 (2.15)	7.91 (2.57)	15.32 (5.80)	
	P300 latency (ms)	711 (120)	745 (104)	668 (108)	
Test 3	**Response**	**NO**	**NO**	**NO**	**YES**
	RT (ms)	530 (71)	536 (69)	547 (70)	622 (78)
	Errors (%)	0.90 (2.2)	0.20 (0.5)	0.80 (1.1)	11.15 (5.0)
	P300 amplitude (μV)	5.10 (1.80)	7.66 (2.40)	7.47 (2.79)	14.17 (4.58)
	P300 latency (ms)	702 (125)	690 (66)	732 (104)	674 (95)
	Recall		5.35 (0.75)	4.90 (1.17)	5.75 (0.44)

Standard Deviations are given in parentheses. Errors represent % misses and false alarms of all responses. Note that for Test 2, mean RTs as well as P300 peak amplitudes were different for the unlearned words and the earlier learned words (List 1) ($p < 0.001$), although both required the same behavioral response. For Test 3, the mean RTs were no longer different for the unlearned words and the words learned earlier (Lists 1 and 2). In contrast, the P300 peak amplitudes for words from List 1 ($p < 0.001$) and List 2 ($p < 0.01$) were still larger than those for unlearned words.

cantly faster for unlearned words than for learned words ($F(1,19) = 28.90$, $p < 0.001$). In the second recognition test, a significant main effect of study list was also present ($F(2,18) = 38.87$, $p < 0.001$). Pairwise comparisons revealed that RTs for unlearned nontargets were significantly faster than to learned targets (List 2; $F(1,19) = 80.64$, $p < 0.001$) and to learned nontargets (List 1; $F(1,19) = 34.81$, $p < 0.001$). Mean RT for learned nontargets was in turn faster than for learned targets ($F(1,19) = 16.32$, $p < 0.001$). In the third test, a significant overall effect of study list ($F(3,17) = 36.43$, $p < 0.001$) was the result of longer RTs for the recently learned words (List 3) compared to those for unlearned words ($F(1,19) = 98.21$, $p < 0.001$) and those for previously learned words from study list 1 ($F(1,19) = 55.06$, $p < 0.001$) and study list 2 ($F(1,19) = 33.06$, $p < 0.001$). Other pairwise comparisons showed no significant differences, revealing that after a one-day interval between learning and testing mean RTs no longer differentiated between unlearned words and those learned words which did not require a YES-response indicating recognition.

Only three subjects could recall all 18 learned words after completion of the study. The number of words from each study list that could be recalled afterwards differed significantly between lists ($F(2,18) = 5.76$, $p < 0.05$) (see Table 2). Pairwise comparisons demonstrated that this resulted from poorer recall of the words from the second study list (4.90 ± 1.17) than those from the third study list (5.75 ± 0.44) ($F(1,19) = 11.16$, $p < 0.01$). Words from the first study list (5.35 ± 0.75) were somewhat better recalled, although not significantly, than those from the second study list ($F(1,19) = 2.82$, $p = 0.11$) but worse than those from the third study list ($F(1,19) = 4.08$, $p = 0.06$).

3.2. Event-related potentials

For each recognition test, grand average ERP waveforms are presented in Fig. 1 as a function of study list. All words provoked clear N1-P2-N2 waveforms. A sustained fronto-central negativity was present in the ERP waveforms to unlearned words and earlier learned words which both served as

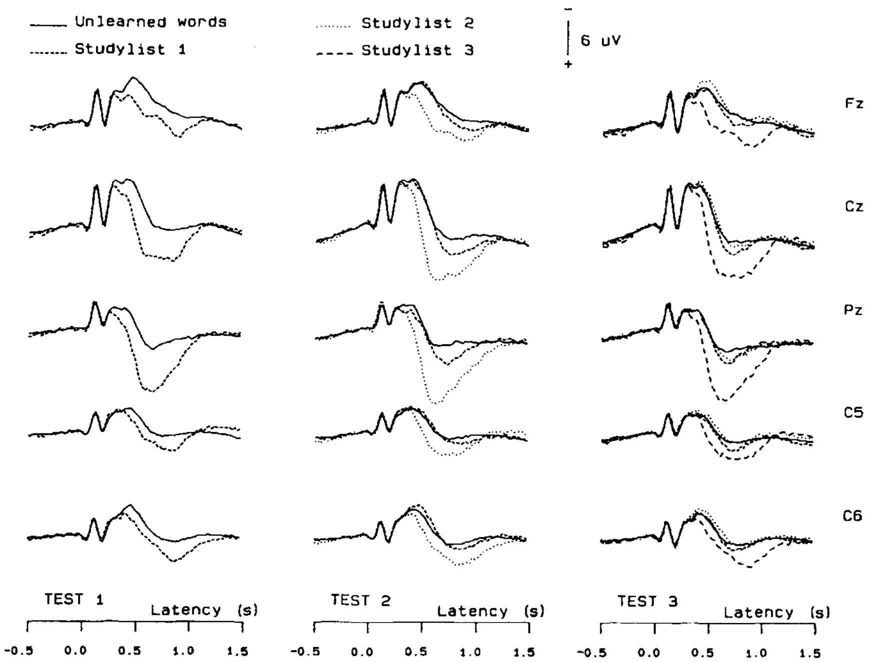

Fig. 1. Averaged ERP waveforms from the midline and lateral electrode positions for each test as a function of study list. Word-onset was at 0.0 s.

nontargets. This negative component was smaller and of shorter duration for the overtly recognized words. The recently learned as well as previously learned words elicited a centro-parietal positivity around 500–1000 ms post-stimulus, which was larger than those elicited by unlearned words. The size and onset of this positivity (P300) were clearly affected by the requirement of an overt (YES) recognition response, showing a larger amplitude and an earlier onset for the overtly recognized words.

3.2.1. P300 peak amplitude and latency

MANOVA results for P300 peak amplitude and latency (max. 450–950 ms) recorded from the midline electrode positions are shown in Table 3. Mean values at Pz are indicated in Table 2. For each test (Test 1, 2 and 3) there were significant main effects of study list on P300 amplitude (midline and lateral: $p < 0.001$). Pairwise comparisons executed for the midline electrode positions revealed that all learned words, regardless of whether they required an overt (YES) recognition response, elicited larger P300s than did the unlearned words (see Table 3). These results were significant for each test even after Bonferroni correction (Test 2: $p < 0.017$; Test 3: $p < 0.008$). Pairwise comparisons executed for the lateral electrode positions showed essentially the same results. However, the P300 amplitude difference between study list 2 and unlearned words did not reach significance in the third test ($t = 1.89$, $p = 0.074$) at these lateral sites. P300 amplitude was largest at all sites for the learned target words which were overtly recognized (all pairwise comparisons with the recently learned list: $p < 0.001$). In the third test, P300 peak amplitudes for list 1 and list 2 (both learned nontargets) were not significantly different from each other (midline: $t = 1.44$, ns; lateral: $t = 2.20$, ns). For the midline analyses, the significant effects of electrode position (p's < 0.001) indicate that the P300 had a centro-parietal dominance. The Study list × Electrode position interactions (p's < 0.05) indicate that the effects of study list were the largest on the central and parietal electrode positions (see Table 3). There were no significant hemisphere differences in the appearance of P300 nor in terms of interaction with study list.

P300 peak latency was significantly affected by the factor study list in the third test (midline: $F(3,17) = 4.43$, $p < 0.05$; lateral: $F(3,17) = 4.15$, $p < 0.05$). Pairwise comparisons executed for the midline electrode positions revealed that this was mainly the result from a P300 latency difference between list 1

Table 3

Results from MANOVAs and pairwise comparisons carried out for each test separately on the P300 peak amplitude and P300 peak latency at midline electrodes

	TEST 1		TEST 2		TEST 3	
	df	*F*	df	*F*	df	*F*
P300 PEAK AMPLITUDE						
Study list (S)	1,19	76.09 ***	2,18	24.24 ***	3,17	17.98 ***
Electrode (E)	2,18	32.01 ***	2,18	49.55 ***	2,18	71.06 ***
S × E	2,18	9.52 **	4,16	7.67 **	6,14	3.81 *
Paired comparisons (t values):						
List 1 vs Unlearned				5.79 ***		5.91 ***
List 2 vs Unlearned				7.15 ***		3.08 **
List 3 vs Unlearned						7.44 ***
P300 PEAK LATENCY						
Study list (S)	1,19	2.65	2,18	2.93	3,17	4.15 *
Electrode (E)	1,19	9.76 **	1,19	13.22 ***	1,19	6.72 **
S × E	1,19	1.52	2,18	1.25	3,17	1.65
Pairwise comparisons (t values):						
List 1 vs Unlearned				< 1		2.88 *
List 2 vs Unlearned				1.87		< 1
List 3 vs Unlearned						1.72

Outcomes from the pairwise comparisons for the midline electrodes are indicated by t-values. $^{*} = p < 0.05$, $^{**} = p < 0.01$, $^{***} = p < 0.001$.

and unlearned words ($t = 2.32$, $p < 0.05$) and between List 1 and List 2 ($t = 2.88$, $p < 0.05$), showing a shorter P300 latency for list 1. However, after Bonferroni correction these comparisons were no longer significant. Pairwise comparisons executed for the lateral electrode positions revealed that P300 latency was shorter for list 1 than for list 3 ($t = 3.38$, $p < 0.008$). In each test, a significant effect of electrode position was present for P300 latency ($p < 0.01$). Mean latency values indicated that P300 reached its maximal amplitude earliest at Pz and latest at Fz. A significant hemisphere effect in the second ($F(1,19) = 12.36$, $p < 0.01$) and third test ($F(1,19) = 6.62$, $p < 0.01$) revealed that during these tests P300 latency was longer at C6 (right hemisphere) than at C5 (left hemisphere).

3.2.2. Mean epoch amplitudes

Mean amplitudes of five consecutive 150 ms epochs as a function of test and study list are given in Table 4. Significant amplitude differences between the distinct classes of stimuli are also indicated in this table. The negativity, mainly present for the words that needed a NO-response, encompassed the first three 150 ms epochs and shifted from a central towards a frontal maximum. The subsequent positivity, prominently present for the learned words, appeared in the period of the last two or three epochs and had a parietal maximum.

MANOVAs carried out on the mean amplitudes obtained during the first test revealed significant effects of electrode position ($p < 0.001$) and study list ($p < 0.05$) for all 150 ms epochs. Mean ERP

Table 4
Mean amplitudes of five 150 ms epochs for the first, second and third recognition test as a function of Study list

Latency range		Test 1		Test 2			Test 3			
		List 0 (NO)	List 1 (YES)	List 0 (NO)	List 1 (NO)	List 2 (YES)	List 0 (NO)	List 1 (NO)	List 2 (NO)	List 3 (YES)
200–350 ms	Fz	−4.01	−3.66	−4.21	−4.33	−3.92	−3.65	−3.32	−3.92	−2.98
	Cz	−5.95	−5.34	−6.50	−6.22	−5.93	−5.83	−5.52	−5.82	−5.42
	Pz	−3.06	−2.25	−3.73	−3.16	−3.23	−3.13	−2.97	−2.87	−2.80
		(a)	(b)	(a)	(a)	(a)	(a)	(a)	(a)	(a)
350–500 ms	Fz	−6.94	−4.14	−6.58	−6.87	−4.21	−5.94	−5.88	−7.08	−2.64
	Cz	−7.51	−3.34	−7.41	−7.02	−3.26	−6.25	−6.37	−7.05	−2.50
	Pz	−2.80	−1.52	−3.46	−2.40	1.25	−2.47	−2.30	−2.41	1.35
		(a)	(b)	(a)	(a)	(b)	(a)	(a)	(a)	(b)
500–650 ms	Fz	−5.68	−1.42	−4.88	−5.11	−0.24	−3.98	−3.44	−5.35	0.90
	Cz	−2.76	4.43	−1.51	−1.26	7.40	−0.43	0.53	−1.30	6.94
	Pz	1.94	9.70	1.62	2.65	10.52	2.40	3.54	2.67	10.69
		(a)	(b)	(a)	(a)	(b)	(a)	(a1) [a]	(a2) [a]	(b)
650–800 ms	Fz	−2.90	−0.24	−2.05	−0.82	1.68	−1.66	−0.30	−1.77	1.85
	Cz	0.43	5.55	1.82	4.34	8.64	2.38	4.25	3.03	8.38
	Pz	2.76	9.76	2.65	5.56	11.29	3.23	4.88	4.63	10.97
		(a)	(b)	(a)	(b)	(c)	(a)	(b)	(a)	(c)
800–950 ms	Fz	−1.27	2.12	−0.50	0.99	2.86	−0.61	0.06	−0.90	3.63
	Cz	0.66	5.26	1.90	4.30	7.58	2.22	3.10	2.59	7.65
	Pz	1.90	6.27	2.18	4.00	8.06	2.75	3.09	3.58	7.84
		(a)	(b)	(a)	(b)	(c)	(a)	(a)	(a)	(b)

List 0 corresponds with the unlearned list of words. For each test: the mean amplitudes of the Study lists which were significantly different from each other within each interval across all sites are indicated with a different letter. Significance levels were Bonferroni corrected: Test 1: $\alpha = 0.05$, Test 2: $\alpha = 0.05/3$, Test 3: $\alpha = 0.05/6$. The application of this Bonferroni correction procedure did not contribute to the difference between tests in the pattern of list-related effects, i.e., if the same Bonferroni correction is applied to Test 2 and Test 3, these results are unchanged.

[a] a1 differed from a2 but neither differed from a.

amplitudes for the learned words were less negative (200–350 ms and 350–500 ms) or more positive (500–650 ms, 650–800 ms and 800–950 ms) than those for the unlearned words. Significant Study list × Electrode position interactions were present for the mean amplitudes of all but the first epoch ($p < 0.05$), referring to a larger effect of study list (learned versus unlearned words) on the central and parietal than on the frontal electrode positions.

For the second recognition test, significant effects of electrode position ($p < 0.001$) were found for all mean amplitudes. Significant main effects of study list (i.e., across the three midline positions) were found for the mean amplitudes of the last four 150 ms epochs ($p < 0.001$). Pairwise comparisons revealed that mean ERP amplitudes for the learned targets (List 2) were less negative (350–500 ms epoch) or more positive (500–650 ms, 650–800 ms and 800–950 ms) than those for both the learned nontargets (List 1) and unlearned nontargets ($p < 0.001$). In turn, mean ERP amplitudes for the learned nontargets appeared to be more positive than those for unlearned nontargets after about 600 ms post stimulus (see Fig. 1). Compared to the results of the overtly recognized words (List 2) this effect was less pronounced and was significant only for the 650–800 ms ($F(1,19) = 19.01$, $p < 0.001$) and 800–950 ms ($F(1,19) = 15.68$, $p < 0.01$) epochs. The effects of study list were largest for the central and parietal sites as could be deduced from Table 4 and a significant Study list × Electrode position interaction for all but the first 150 ms epoch ($p < 0.05$).

In the third recognition test, significant effects of electrode position ($p < 0.001$) were also found for all mean amplitudes. The factor study list significantly affected the mean amplitudes of the last four 150 ms epochs ($p < 0.001$). As in the second test, pairwise comparisons revealed that from the 350–500 ms epoch the mean ERP amplitudes for words from the recently learned list (List 3) were more positive than those for unlearned words ($p < 0.001$) and those for words from the earlier learned lists ($p < 0.001$). In the 650–800 ms epoch, mean ERP amplitudes across the three midline positions for words from the first study list were more positive than those for unlearned words ($F(1,19) = 18.23$, $p < 0.001$). For the same interval, a significant overall difference between mean ERP amplitudes for words from the second study list and those for unlearned words was not found, although they differed significantly on the Pz electrode position ($F(1,19) = 11.42$, $p < 0.01$; Study list × Electrode-position interaction: $F(6,14) = 6.47$, $p < 0.01$). Mean ERP amplitudes for words from the first and second study list differed from each other in the 500–650 ms and 650–800 ms epoch ($p < 0.01$). In these periods the mean ERP amplitudes for words from the first study list were more positive than those for words from the second study list. Significant Study list × Electrode position interactions were present for the mean amplitudes of all but the first 150 ms epoch ($p < 0.05$).

3.2.3. Evaluation of possible confounding effects

Because distinct unlearned categories were assigned together to one list, it could be argued that latency-jitter may have contributed to the reported ERP differences. Further, the three tests differed with respect to target probability which may have contributed to the observed effects. Finally, because the items were not fully rotated and because categories from the first test were re-used in the third test, item-specific confounds and implicit learning effects have to be taken into account. Figs. 2–4 are provided to illustrate the plausibility of these possibilities. Supplementary analyses were performed to evaluate effects of stimulus probability and recurrence of stimuli.

Figs. 2 and 3 present parietally recorded (Pz) ERPs to the learned nontargets and each unlearned category separately. Consequently, the number of trials of which an ERP was composed are approximately equal. Visual inspection of these figures reveals that positivities for each unlearned category were absent or smaller than those for the learned categories, with only one possible exception for category 13 in the third test. Nevertheless, considering the fact that the learned categories were not the same for all subjects but instead were composed of three or six different categories in the second and third test respectively (see Table 1), latency-jitter can be presumed to play no or only a minor role in the origin of the results. As can also be observed in these figures and corraborated by MANOVAs, there were no significant differences between the distinct unlearned categories, making interpretations in terms of item-specific responses less plausible.

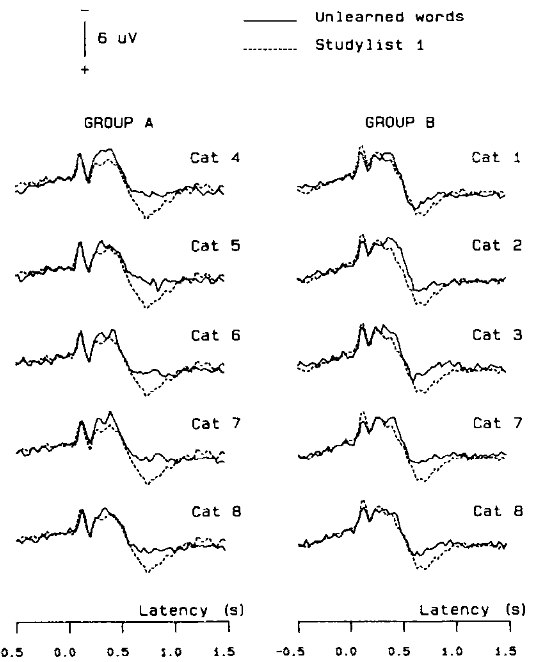

Fig. 2. Test 2. Averaged ERP waveforms recorded from Pz for the previously learned categories which did not require a YES-response to indicate recognition (Study list 1: dashed lines) and for the separate unlearned categories (solid lines). For Group A the learned categories were 1,2 or 3 and the unlearned categories were 4,5,6,7 and 8. For Group B the learned categories were 4, 5, or 6 and the unlearned categories were 1,2,3,7 and 8. Word-onset was at 0.0 s.

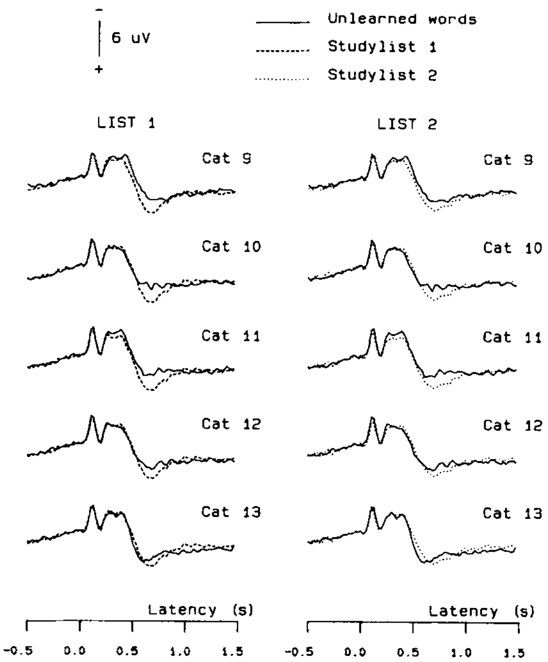

Fig. 3. Test 3. Averaged ERP waveforms recorded from Pz for the previously learned categories which did not require a YES-response to indicate recognition (Study list 1: dashed lines; Study list 2: dotted lines) and for the separate unlearned categories (solid lines). The learned categories were 1,2 or 3 for Group A, and 4,5 or 6 for Group B (see also Table 1). The unlearned categories were 9,10,11,12 and 13 for both Groups A and B. Word-onset was at 0.0 s.

To determine effects of target probability, MANOVAs were carried out on P300 peak amplitude and latency for the learned targets, with test and electrode position (midline electrodes) and test and hemisphere (lateral electrode positions) as within subject factors. Because we were specifically interested in distinctions between all three tests, these MANOVAs were supplemented with planned pairwise comparisons. P300 peak amplitude recorded from the midline electrode positions differed significantly between tests ($F(2,18) = 6.23$, $p < 0.01$). As can be seen in Fig. 4, this was due to a smaller P300 amplitude in the first test compared to those in the second test ($t = 3.34$, $p < 0.01$) and the third test ($t = 2.42$, $p < 0.05$). Because target probabilities in test 1–3 were 1/5, 1/7, and 1/8 respectively, this result corresponded well with differences in target probability. There were no P300 amplitude differ-

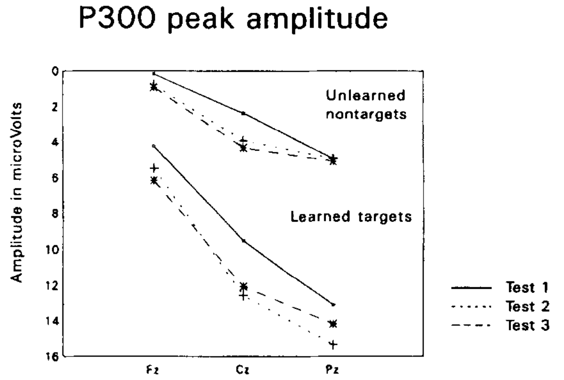

Fig. 4. Mean P300 peak amplitude (in μV) for unlearned words and overtly recognized words as a function of test. Note that for both the unlearned and overtly recognized words, the P300 peak was smaller in the first test than in the second and third test, especially at the Cz electrode position.

ences between the second and third test. A Test × Electrode position interaction ($F(4,16) = 3.52$, $p < 0.05$) revealed that the P300 amplitude difference between tests was larger at Cz ($p < 0.01$) than at the other midline electrode positions ($p < 0.05$). P300 peak amplitude recorded from the lateral electrode positions did not differ significantly between tests. P300 latency was not affected by test.

The fact that words that served as foils in the first test, were used as unlearned words in the third test may have influenced the results. To determine the extent of this confounding effect of recurrence, MANOVAs, similar as those described above, were conducted on P300 peak amplitude and latency for the unlearned nontargets. A significant effect of test on P300 amplitude was present for the midline electrode positions ($F(2,18) = 3.62$, $p < 0.05$). Pairwise comparisons revealed that this was due to a smaller P300 amplitude in the first test than in the second test ($t = 2.21$, $p < 0.05$) and in the third test ($t = 2.12$, $p < 0.05$). A Test × Electrode position interaction ($F(4,16) = 16.18$, $p < 0.001$) revealed a shift in potential distribution between Tests. Analysis for each electrode position showed a significant effect of test for the Cz electrode position ($F(2,18) = 10.65$, $p < 0.01$) which was absent for Fz and Pz electrode positions. A significant effect of test was not found for P300 peak amplitude recorded from the lateral electrode positions. P300 latency differed between tests at the lateral electrode positions ($F(2,18) = 4.61$, $p < 0.05$). Pairwise comparisons showed that this was due to longer P300 latencies in test 1 than in test 2 ($t = 2.46$, $p < 0.05$) and in test 3 ($t = 2.69$, $p < 0.05$), with no effect of hemisphere.

In short, these results show that P300 differences between tests were mainly the result of dissimilarities between test 1 on the one hand and test 2 and 3 on the other hand (see also Fig. 4). Because these P300 amplitude test effects were comparable for learned targets and for unlearned nontargets, it appeared that target probability alone had no substantial influence on the previously reported results. In addition, because there were no P300 amplitude differences between the second test (no recurrence) and the third test (recurrence), whereas both differed from test 1 (no recurrence), it seems evident that the re-use of word categories had no important influence on the results. Apparently, these differences between the three tests are due to another factor, possibly the presence of earlier learned material, which introduced a new task aspect in the second and third tests (see discussion).

4. Discussion

To summarize the results, a late central-parietal positivity was present for recently learned as well as earlier-learned words, even though the latter did not require an overt indication of recognition. This positivity was largely absent for unlearned words. Furthermore, in the second recognition test mean RTs to both types of learned words were slower than those to unlearned words. These results support the notion that ERP as well as RT measures may be a useful way to assess memory in absence of overt indications of recognition, provided that testing can occur close to the time of the study-period. With larger intervals between learning and testing, the present data suggest that only ERP measures can be used as accurate indices of covert recognition.

The main purpose of the current study was to establish whether an ERP-based procedure could detect traces of learned material, independently of overt recognition responses. It can be considered as a first step towards assessing implicit memory effects. Whether the procedure works when subjects are unaware of previously presented material is a subject for further study (Van Hooff et al., 1995). If the procedure is not sensitive to such material in amnesic patients the procedure may be valuable for diagnosing patients with explicit memory but without the ability to report memories because of neurological or functional impairment, or for demonstrating the ability of explicit retention during periods of minimal encoding, e.g., during sleep or dual task performance. If, on the other hand, the procedure is sensitive to previously learned material in the absence of phenomenological awareness, the procedure may prove useful for assessing implicit memory effects, e.g., learning during anesthesia. Because the purpose of this study was more circumscribed – to extend the utility of existing ERP memory assessment procedures – the results will be discussed with respect to the three specific aims discussed previously.

4.1. Three specific objectives of the current study

4.1.1. Necessity of the instruction to deceive

It was examined whether an instruction to deceive is a necessary prerequisite to obtain ERP and RT differences between learned and unlearned items in an assessment paradigm such as those used in deception-detection studies. The present study made use of a paradigm that strongly resembled that of Allen et al. (1992), (but the instruction to intentionally conceal previously learned material was omitted. Subjects were not told that earlier learned words would re-appear in the following tests but were simply instructed to respond affirmatively upon the most recently learned words. Allen et al. (1992) reported that ERP positivities to learned words which had to be concealed were larger than those to unlearned words. A similar effect could also be observed in our study, although the amplitude difference between positivities to learned nontargets and those to unlearned nontargets did not reach the same magnitude as that reported by Allen et al. (1992). Further, there was a substantial difference in onset and amplitude between the P300s for learned targets (overtly recognized) and learned nontargets (not overtly recognized). These effects are most probably due to the fact that we did not direct the subjects to intentionally withhold knowledge about the previously learned items, which decreased the significance and relevance of these items. In contrast, a deception instruction might provide extra significance to learned words causing them to be relatively more distinctive as compared to unlearned words. Nevertheless, our results suggest that a deception instruction, although presumably augmenting the effects, is not a prerequisite for obtaining ERP differences between learned (but nontarget) and unlearned material. The results suggest that the ERP memory assessment procedure may prove useful in clinical investigations, as for example, the assessment of persistent memory functions during anesthesia or in patients with amnesia.

4.1.2. Auditory modality

With visual presentation the entire stimulus is simultaneously presented on the screen whereas with auditory presentations the word is presented as a sequence of sounds. Further, in contrast to visually presented words, spoken words are more variable in duration and intensity-patterns, causing a larger variance in the single trial ERPs. However, we obtained clear ERP waveforms in the present study and observed similar effects as those reported in the study of Allen et al. (1992), using visually presented words. That the ERP procedure works well with auditory stimuli suggests that the procedure might be applicable to the detection of preserved memories for material presented during general anesthesia or sleep. A further implication is that standard visual memory paradigms could be extended to the auditory modality, as has been demonstrated also by Ferlazzo et al. (1993).

4.1.3. Temporal factors affecting the procedure

The ERP memory assessment procedure could still differentiate, albeit less robustly, learned from unlearned words after a one-day interval between learning and testing. In the third recognition test, unlearned words were intermixed with words that were learned the day before and six words that were learned just before the test. Only the recently learned words required an overt indication of recognition. The one-day interval between learning and testing diminished the learning effects for P300 and eliminated them for RTs. In the third test, P300 evidence of covert recognition was still found for the words that were learned the day before but, compared to the corresponding P300s to learned nontargets in the second test, the P300 effect was smaller in amplitude, encompassed a more restricted time period, and was present at a more limited set of scalp sites. The decrease in the size of the P300 amplitude effects might be explained by an effect of implicit learning, because the unlearned nontargets from the first test were re-used in the third test. As previously discussed in the result section, however, it appeared that recurrence of unlearned word categories did not contribute substantially to the reduced ERP effects. Moreover, the difference in ERP effects observed in the second test and those observed in the third test can only be explained by a decrease in P300 amplitude for learned nontargets because the P300 amplitude for unlearned nontargets were largely the same in both tests. The decrease in amplitude therefore most probably resulted from a decline in the strength of the memory trace and/or from decreased stimulus significance. These same factors likely caused the disappearance of the RT effect. The fact that the ERP effects persisted while the RT effect disappeared after a one-day interval suggests that ERPs can provide more sensitive (i.e., longer lasting) indices of covert recognition than do RT data.

Farwell and Donchin (1991) also tested their subjects one day after participation in a mock espionage scenario. Slowed RTs to memory-relevant items were still found, but the authors suggested that RTs may not be suitable as a measure of guilt or innocence because they can easily be voluntary manipulated. Allen et al. (1992) administered their second recognition task after a 30-min break. They reported that implicit behavioral measures (mean RTs for correct responses and the number of incorrect responses) were equally effective in identifying the critical list as was the ERP-based procedure. Therefore, they emphasized that the superiority of ERP recordings over implicit behavioral measures has yet to be demonstrated. In our study, this superiority has been demonstrated in memory-testing across time, suggesting ERPs have utility for the assessment of memory over longer study-test intervals. It is unclear over what durations the procedure may prove effective beyond the one day examined in the present study.

4.2. General discussion

The ERP effects may be explained by the fact that both types of learned words had special significance for the subjects (by virtue of previous learning) and appeared with a lower probability than did the unlearned words. Apparently, these two aspects were sufficient conditions for eliciting ERP signatures that differed from unlearned words without the additional requirement of an overt recognition response. As has been noted by Rosenfeld et al. (1991), it should be explicitly stated here that factors such as (involuntary) attention, relevance, subjective meaning, and other attributes may explain the P300-evoking property of learned items in addition to low stimulus probability. Moreover, all word-categories had the same subjective probability, thus only when learned items were perceived as a distinct rare category – that is recognized as having been learned previously — would they possess the ability to evoke a large P300.

Onset of amplitude differences between the ERPs for learned targets and learned nontargets suggests that different processes might be responsible for the observed effects. ERPs in response to words requiring an overt indication of recognition started to differ from those in response to words requiring no overt indication of recognition in the 200–350 ms (first test) or 350–500 ms (second and third test) interval, which implies that the discrimination of words requiring a YES or NO response has taken place prior to these intervals. In visual recognition tasks, comparable divergence latencies are found (Friedman, 1990; Neville et al., 1986; Smith and Guster, 1993). This apparently fast discriminating process does not automatically imply that a correct word identification is possible without additional processing. The sustained negativity, recorded for the words that needed a NO-response, might be related to this additional processing. It might be comparable with the Processing Negativity (Näätänen, 1982) or Search Negativity (Okita et al., 1985) recorded in the same time window in selective attention paradigms. The shorter duration of the negative component for the words that needed a YES-response might reflect the fact that these words could be identified more quickly and required no additional processing. An alternative explanation for the reduced negativity for recognized words could be that it resulted from an overlap with the P300 elicited by these words as a consequence of its target status.

In contrast with the P300 for overtly recognized words, the positive shift recorded for the learned nontarget words did not start until after the response was given. This suggests that it presumably reflects post-decision processes. The late appearance of the P300 might be associated with the character of the current paradigm. Our subjects were confronted with words to which they initially had to respond with YES and subsequently with NO. Consequently, if in addition to the word itself, the associated response also was remembered (c.f., Bentin and Peled, 1990), this would probably have caused a response conflict, which might have been subject to post-decision evaluation. Furthermore, the occurrence of a response conflict might also be responsible for the P300 amplitude and scalp distribution differences between test 1 (no response conflict) and test 2 and 3 (presence of response conflict). Because the P300 for overtly recognized words lasted until long after the response was given it might be assumed that post-decision processes also affected this ERP component.

The ERP divergence between unlearned nontargets and learned nontargets was smaller in the third test than in the second test, which might be the consequence of a decline in the strength of the memory trace. As is described earlier in studies of Bentin and Moscovitch (1990) and Bentin et al. (1992), the size of the current P300 was influenced by both the recency and the number of presentations of the learned words. The sensitivity to the recency of learned nontargets was reflected by the decrease in P300 size from the second to the third test, as can be revealed from the mean interval amplitudes. In test 2 (administered 15 min after learning the first study list), the P300 to words from study list 1 was significantly larger than the P300 to unlearned words in the 650–800 ms and 800–950 ms epochs. In contrast, in test 3 (administered 24 hours after learning the first study list), the difference in P300 size was significant only for the 650–800 ms epoch. The sensitivity to the number of presentations was reflected by the fact that in the third test, words from the first study list, which were presented in two preceding tests, elicited larger P300s than words from the second study list, which were presented in only one preceding test. This holds, however, for the mean amplitude in the 650–800 ms epoch but not for the P300 peak amplitude. Assuming that a larger number of presentations induced a stronger memory trace that, in turn, decays over time, it might be concluded that the recorded positivity was sensitive to the strength of the memory trace. In addition, this interpretation of the ERP data is in agreement with the recall performance since the words from the most recent study list (List 3) could be recalled the best and the words from the least repeated study list (List 2) the worst (see Table 2).

5. Conclusion

In conclusion, during an auditory word recognition task, learned words reliably elicited larger P300s than unlearned words even when no overt indications of recognition were required. A deception instruction appeared not to be a necessary prerequisite to obtain these discriminating effects. After a one-day interval between learning and testing, ERPs could still differentiate, although less robustly, learned from unlearned words. These results support the notion that ERPs can be used as indirect measures of recognition memory. For the clinical practice, this might imply that with the help of ERP measures the presence of information in long-term memory could be reliably assessed, perhaps even in situations in which patients are unable to report information of which they might have some knowledge.

Acknowledgements

This study was funded, in part, by the Co-operation Center Tilburg and Eindhoven Universities, and by a grant from the NATO International Scientific Exchange Programmes (CRG 940305). The authors would like to thank Greet van den Berg-Lenssen and Pierre Cluitmans for technical and programming support.

References

Allen, J.J., Iacono, W.G. and Danielson, K.D. (1992) The identification of concealed memories using the Event Related Potential and Implicit Behavioral measures: A methodology for prediction in the face of Individual differences. Psychophysiology, 29: 504–522.

Bentin, S. and Moscovitch, M. (1990) Psychophysiological indices of implicit memory performance. Bull. Psychon. Soc., 28: 346–352.

Bentin, S. and Peled, B. (1990) The contribution of task-related factors to ERP repetition effects at short and long lags. Mem. Cognition, 18: 359–366.

Bentin, S., Moscovitch, M. and Heth, I. (1992) Memory with and without awareness: Performance and Electrophysiological evidence of savings. J. Exp. Psychol. Learn. Mem. Cogn., 18: 1270–1283.

Eich, E. (1984) Memory for unattended events: remembering with and without awareness. Mem. Cogn., 12: 105–111.

Fabiani, M., Karis, D. and Donchin, E. (1986) P300 and recall in an Incidental Memory paradigm. Psychophysiology, 23: 298–308.

Farwell, L.A. and Donchin, E. (1986) The 'Brain detector: ' P300 in the detection of deception. (Abstract) Psychophysiology, 25: 445.

Farwell, L.A. and Donchin, E. (1991) The truth will out: Interrogative polygraphy (Lie detection) with event-related brain potentials. Psychophysiology, 28: 531–547.

Ferlazzo, F., Conte, S. and Gentilomo, A. (1993) Event-Related Potentials and recognition memory: the effect of word imagery value. Int. J. Psychophysiol., 15: 115–122.

Friedman, D. (1990) ERPs during continuous recognition memory for words. Biol. Psychol., 30: 61–87.

Jennings, J.R. (1987) Editorial policy on analyses of variance with repeated measures. Psychophysiology, 24: 474–475.

Johnson, R.J. (1986) A triarchic model of P300 amplitude. Psychophysiology, 23: 367–384.

Kemp-Wheeler, S.M. and Hill, A.B. (1988) Semantic Priming without Awareness: Some Methodological Considerations and Replications. Q. J. Exp. Psychol., 40A: 671–692.

Leiphart, J., Rosenfeld, J.P. and Gabrieli, J.D. (1993) Event-related potential correlates of implicit priming and explicit memory tasks. Int. J. Psychophysiol., 15: 197–206.

Näätänen, R. (1982) Processing Negativity: An evoked potential reflection of selective attention. Psychol. Bull., 92: 605–640.

Neville, H.J., Kutas, M., Chesney, G. and Schmidt, A.L. (1986) Event-related brain potentials during initial encoding and recognition memory of congruous and incongruous words. J. Mem. Lang., 25: 75–92.

Okita, T., Wijers, A.A., Mulder, G. and Mulder, L.J.M. (1985) Memory search and visual spatial attention: An event related brain potential analysis. Act. Psychol., 60: 263–292.

Paller, K.A., Kutas, M. and Mayes A.R. (1987) Neural correlates of encoding in an incidental learning paradigm. Electroencephalogr. Clin. Neurophysiol., 67: 360–371.

Renault, B., Signoret, J., Debruille, B., Breton, F. and Bolgert, J. (1989) Brain potentials reveal covert facial recognition in prosopagnosia. Neuropsychologia, 27: 905–912.

Rosenfeld, J.P., Angell, A., Johnson, M. and Qian, J. (1991) An ERP-based, control-question lie detector analog: Algorithms for discriminating effects within individuals' average waveforms. Psychophysiology, 28: 319–335.

Rosenfeld, J.P., Ellwanger, J. and Sweet, J. (1995) Detecting simulated amnesia with event-related brain potentials. Int. J. Psychophysiol., 19: 1–11.

Rugg, M.D. (1987) Dissociation of semantic priming, word and nonword repetition effects by event-related potentials. Mem. Cognition, 15: 473–481.

Rugg, M.D., Furda, J. and Lorist, M. (1988) The effects of task on the modulation of Event-Related Potentials by word repetition. Psychophysiology, 25: 55–63.

Sanquist, T.F., Rohrbaugh, J.W., Syndulko, K. and Lindsey, D.B. (1980) Electrocortical signs of levels of processing: perceptual analysis and recognition memory. Psychophysiology, 17: 568–576.

Smith, M.E. and Guster, K. (1993) Decomposition of recognition memory event-related potentials yields target, repetition, and retrieval effects. Electroencephalogr. Clin. Neurophysiol., 86: 335–343.

Uit den Boogaart, P.C. (1975) Woordfrequenties in geschreven en gesproken Nederlands, Oosthoek, Scheltema & Holkema, Utrecht (in Dutch).

Van den Berg-Lenssen, M.M.C., Brunia, C.H.M. and Blom, J.A. (1989) Correction of ocular artifacts in EEGs using an autoregressive model to describe the EEG; a pilot study. Electroencephalogr. Clin. Neurophysiol., 73: 72–83.

Van Hooff, J.C., De Beer, N.A.M., Brunia, C.H.M., Cluitmans, P.J.M., Korsten, H.H.M., Tavilla, G. and Grouls, R. (1995) Information processing during cardiac surgery: An Event Related Potential study. Electroencephalogr. Clin. Neurophysiol., 96: 433–452.

Vasey, M.W. and Thayer, J.F. (1987) The continuing problem of false positives in repeated measures ANOVA in psychophysiology: a multivariate solution. Psychophysiology, 24: 479–486.

Wasserman, S. and Bockenholt, U. (1989) Bootstrapping: applications to psychophysiology. Psychophysiology, 26: 208–221.