Clinical Neurophysiology 114 (2003) 662–672
www.elsevier.com/locate/clinph

Electrophysiological assessment of language function following stroke

Ryan C.N. D'Arcy[a,b,*], Yannick Marchand[a], Gail A. Eskes[c,d,e], Edmund R. Harrison[c], Stephen J. Phillips[e], Alma Major[a], John F. Connolly[a,d,f]

[a]Cognitive/Clinical Neuroscience Unit, Department of Psychology, Dalhousie University, Halifax, MB R3B 1Y6, Canada
[b]National Research Council's Institute for Biodiagnostics (Atlantic), Halifax, NS, Canada
[c]Nova Scotia Rehabilitation Centre, Queen Elizabeth II Health Sciences Centre, Halifax, NS, Canada
[d]Department of Psychiatry, Dalhousie University, Halifax, NS, Canada
[e]Department of Medicine (Division of Neurology), Halifax Infirmary, Queen Elizabeth II Health Sciences Centre, Halifax, NS, Canada
[f]Pediatrics (Neurology), IWK Health Centre, Dalhousie University, Halifax, NS, Canada

Accepted 30 December 2002

Abstract

Objective: Event-related brain potentials (ERPs) were used to assess language function after stroke and demonstrate that it is possible to adapt neuropsychological tests to evaluate neurocognitive function using ERPs. Prior ERP assessment work has focused on language in both healthy individuals and case studies of aphasic neurotrauma patients. The objective of the current study was to evaluate left-hemisphere stroke patients who had varying degrees of receptive language impairment. It was hypothesized that ERPs would assess receptive language function accurately and correlate highly with the neuropsychological data.

Methods: Data were collected from 10 left-hemisphere stroke patients; all were undergoing rehabilitation at the time of testing. Each patient received a battery of neuropsychological tests including the Peabody Picture Vocabulary Test-Revised (PPVT-R; Minnesota: American Guidance Service, 1981). ERPs were recorded during a computerized PPVT-R, in which pictures are presented followed by digitized spoken words that are either congruent or incongruent with the pictures.

Results and conclusion: Incongruent spoken words *within* an individual's vocabulary level elicited well-known ERP components. One of the components (the N400) could be utilized as a marker of intact semantic processing. The ERP results were subsequently quantified and N400 derivative scores correlated highly with the neuropsychological findings. The results provided a clear demonstration of the efficacy of ERP-based assessment in a neurological patient group.

Significance: Language function in stroke patients can be evaluated, independent of behavior, using electrophysiological measures that correlate highly with traditional neuropsychological test scores.

Keywords: N400; Phonological mismatch negativity; Stroke; Event-related brain potentials; Language; Neuropsychological assessment

1. Introduction

The ability to evaluate accurately whether or not a patient can understand speech is essential in determining clinical intervention strategies during acute care and rehabilitation. Unfortunately, a myriad of neurological disorders, including stroke, can result in behavioral and/or communicative limitations that render the assessment of comprehension abilities difficult or impossible. For example, patients with left hemisphere stroke typically present with deficits in both motor and language functions, thus limiting the efficacy of neuropsychological assessment methods (Enderby and Philipp, 1986; Enderby et al., 1987; Marquardsen, 1969; Morse and Montgomery, 1992; Pedersen et al., 1995; Wade et al., 1986).

Recently, we have developed and validated the Innovative Methods of Assessment Program, which utilizes event-related brain potentials (ERPs) to assess language functions (Connolly and D'Arcy, 2000; Connolly et al., 2000). Similar to evoked potentials, ERPs are derived from electroencephalographic recordings that are time-locked to a stimulus event and represent the brain's 'on-line' response to a series

* Corresponding author. Institute for Biodiagnostics, National Research Council, 435 Ellice Avenue, Winnipeg, NS R3B 1Y6, Canada. Tel.: +1-204-984-6973; fax: +1-204-984-7036.
E-mail address: ryan.d'arcy@nrc-cnrc.gc.ca (R.C.N. D'Arcy).

1388-2457/03/$30.00 © 2003 International Federation of Clinical Neurophysiology. Published by Elsevier Science Ireland Ltd. All rights reserved.
doi:10.1016/S1388-2457(03)00007-5

of stimuli (Knight, 1997). Historically, evoked potentials have been used successfully as a clinical measure to evaluate the functional integrity of sensory systems (e.g. auditory, visual, and somatosensory; Chiappa, 1997). In contrast, ERPs have been regarded as too variable for use in the clinical evaluation of cognitive functions, but have been investigated intensively in research contexts. Due to recent advances, it is now possible to utilize ERPs for neuropsychological assessment of neurological patient populations.

Our group has developed techniques that utilize ERPs, instead of behavioral responses, to assess cognitive functions. One of the primary goals of this research has been to provide physiological tools that can be used with communication-impaired populations (e.g. aphasic patients). The initial investigations have focused on the assessment of language functions, as they represent one of the most common sequelae of neurotrauma. This clinical use of ERPs as an assessment tool has been developed by: (1) adapting standardized neuropsychological tests for computer presentation with simultaneous ERP recordings; (2) recording ERPs in healthy individuals in order to show that differences in the waveforms correspond with the range of normal performance on the test; and (3) using ERPs to evaluate patients with language deficits secondary to brain damage using a case study approach. The strength of this work is derived from the fact that it combines both the current knowledge of ERPs with knowledge of the brain-behavior relationships obtained from standardized neuropsychological testing (Connolly and D'Arcy, 2000; Connolly et al., 2000).

In a series of normative studies with adult participants, we have demonstrated the ability to adapt neuropsychological tests for computer presentation and simultaneous ERP recordings. Topics that have been addressed by this research include tests of vocabulary knowledge (visual and auditory; Connolly et al., 1995a, 1999a) and sentence comprehension (visual and auditory; D'Arcy and Connolly, 1999; D'Arcy et al., 2000).

The Peabody Picture Vocabulary Test – Revised (PPVT-R; Dunn and Dunn, 1981) has been normed for both children and adults using ERPs (Byrne et al., 1995a,b; Connolly et al., 1995a). The computerized PPVT-R is divided into 3 levels of vocabulary difficulty and individual test items consist of a picture paired with a digitized spoken word that does or does not accurately describe the accompanying picture. For a given test item, a picture (e.g. car) is presented followed by a spoken word that is either congruent or incongruent (e.g. 'car' or 'plate'). ERPs are recorded to the onset of the spoken words. In healthy individuals, the incongruent spoken words elicit two distinct components: an early phonological mismatch negativity (PMN) and a N400. While still under active investigation, the PMN is a small negative-going waveform (250–350 ms range), which is thought to reflect phonological processing of unexpected speech input (Connolly et al., 1995a, 2001; Connolly and Phillips, 1994; Dehaene-Lambertz et al., 2000; Hagoort and Brown, 2000; Praamstra and Stegeman, 1993; van Den Brink et al., 2001). The N400 is a well-known negative-going waveform (peaking around 400 ms) and has been linked to semantic analysis in both reading and speech processing (Connolly and Phillips, 1994; Connolly et al., 1995b; Holcomb and Neville, 1990; Kutas and Van Petten, 1994). The findings of the initial PPVT-R studies in healthy subjects have demonstrated that it was possible to measure auditory linguistic processing and comprehension on the basis of the presence or absence of the PMN and N400 components (and independent of behavior). That is, the incongruent spoken words reliably elicited PMN and N400 components, indicating intact comprehension of the picture-spoken word mismatches. Of particular importance, the ERP differentiation occurred only for items *within* a participant's vocabulary knowledge level. No ERP differences were observed to congruent or incongruent words when the subject failed to comprehend differences between the test items (because the words were beyond the participant's psychometrically determined vocabulary level).

These ERP assessment measures have been used in single case studies in order to begin evaluating their clinical efficacy. The preliminary results have been positive. To assess written sentence comprehension in 3 left-hemisphere stroke patients, we used a computerized version of the reading comprehension subtest of the Psycholinguistic Assessments of Language Processing in Aphasia (PALPA; Kay et al., 1992). The 3 patients were characterized by varying degrees of language impairment with all performing more than 1 SD below the normative values on the standardized test (Connolly et al., 2000). Examination of the data revealed that the expected ERP differentiation matched behavioral performance on the PALPA. The only case in which ERP differentiation failed to occur corresponded with the most severely impaired patient. The patient with the highest score on the reading subtest of the PALPA showed the most pronounced ERP differentiation. The results provided initial signs of a correspondence between ERP and neuropsychological findings in patients. However, detailed examination of this relationship requires a larger group of patients and a more comprehensive neuropsychological evaluation.

Additional evidence for the efficacy of ERP assessment has been derived from two separate case studies. In one case study (Connolly et al., 1999b), a 21-year-old male suffered a traumatic brain injury after being stabbed through the head with a 12-inch knife. He survived the attack and was admitted to a rehabilitation hospital for assessment. He was globally aphasic and his physical impairments prevented reliable movements of any type. Traditional assessment was impossible, which severely limited any judgments about his level of functioning. He was deemed 'vegetative' and without the requisite cognitive abilities to be a candidate for rehabilitation. Contrary to his clinical presentation, ERP assessment demonstrated intact speech comprehension abilities. As a result, he was subsequently admitted to a rehabilitation program that resulted in a near complete recovery (with the exception of expressive abilities) and return to normal daily activities.

The second case (Byrne et al., 1995a) involved assessing the receptive single word auditory abilities of a youth with spastic quadriplegic cerebral palsy using the PPVT-R (Byrne et al., 1995a). Data were recorded from the patient and 3 age-matched controls. The patient showed the same ERP patterns as the matched controls. That is, a PMN and a N400 were present to incongruent picture descriptions at the basic and age-appropriate vocabulary levels and were clearly differentiated from the waveforms to the congruent picture descriptions. This study demonstrated a graded, age-appropriate level of functioning in a young communication-impaired cerebral palsy patient.

The research-to-date has led to a number of conclusions (for a review, see Connolly and D'Arcy, 2000; Connolly et al., 2000). Namely, behavioral performance comparisons have proven that the computer-adapted ERP measures maintain the essential features of the standardized neuropsychological tests. Also, ERP differentiation is independent of the ability of the participant to provide behavioral responses. These results can be obtained reliably at the individual subject level in all cases, and the differentiation disappears when the individual's abilities are exceeded by the test demands. Finally, the ERP measures obtained in these tests provide valuable information about the nature of the cognitive process necessary for successful performance (beyond that which can be gained from neuropsychological assessment alone). In the current study, we investigated these techniques in left hemisphere stroke patients, a group for which accessibility through traditional testing methods is frequently compromised.

1.1. Objectives and hypotheses

The objective of this study was to extend ERP neurocognitive testing beyond case studies to a larger group study format. While previous ERP findings have fit well with performance on the standardized test, this study sought to quantify the ERP differentiation enabling a more direct comparison with traditional neuropsychological test scores. Thus, the goal of this experiment was to empirically evaluate the receptive language function in left-hemisphere stroke patients using electrophysiological measures. We administered the computerized PPVT-R to a group of patients and recorded ERPs simultaneously. The ERP data were integrated with behavioral data from traditional neuropsychological measures of language and other cognitive functions.

The N400 was of specific interest because it is a relatively large response that is readily observed in ERP waveforms. The N400 has been the subject of an extensive literature and its relationship to specific cognitive functions is established. A number of studies have also begun to identify the neuroanatomical generators of the N400. Sources have been localized within left perisylvian regions (Helenius et al., 1998, 2002) and anterior medial temporal regions (McCarthy et al., 1995). However, there is evidence that the sources may be more distributed and vary between individuals (Haan et al., 2000). Because the N400 has been well characterized, it was selected as the primary measure of language function in left hemisphere stroke patients.

The patient sample consisted of individuals who were previously fluent in English and had been admitted to hospital following a left-hemisphere stroke. The patients were selected to capture a range of language dysfunction. It was hypothesized that N400 responses on the computerized PPVT-R would accurately evaluate a patient's ability to understand spoken words (across varying levels of language impairment as measured by standardized neuropsychological testing).

2. Methods

2.1. Patients

Ten English-speaking patients (8 males and 2 females), enrolled in rehabilitation programs in a rehabilitation hospital, volunteered for a study on language assessment techniques following stroke. All patients were admitted to hospital following the diagnosis of a left-hemisphere stroke and had undergone diagnostic imaging (CT and/or MRI scans) in order to identify infarct locations. All patients were dextral (Edinburgh Handedness Inventory Laterality Quotient range: 66.7–100; Oldfield, 1971), with a mean age of 61.8 ($SD = 13.2$) years and mean education of 11.1 ($SD = 2.5$, range = 5–14) years. Only one patient (S09) had been hospitalized previously for stroke. A detailed description of their infarct locations and levels of functioning is provided in Section 3.

For the experiment, each patient participated in two separate 1.5 h testing sessions (ERP and Neuropsychological, counterbalanced), both of which always occurred over a 3 day testing period. The mean number of weeks between the stroke and 1st day of testing was 7.6 ($SD = 3.9$, range 3.9–14.6). This study had ethical committee approval and informed consent was obtained prior to testing. Upon completion of the experiment, the patients and next of kin were debriefed fully and all questions were answered.

2.2. Electrophysiological methods

Electroencephalogram (EEG) was recorded (bandpass = 0.05–30 Hz, digitally filtered to 0.1–20 Hz; sampled at 1000 Hz) from 100 ms pre-stimulus to 1000 ms post-stimulus using 3 midline sites that were referenced to linked ears (Fz, Cz, and Pz). Trials with electrooculogram artifacts ± 75 μV between -100 and 750 ms were rejected from the analysis. Across all participants, 78.5%

(SD = 15.9) of trials were accepted for analysis. EEG data were averaged by experimental condition (Congruent and Incongruent) for all 3 levels of difficulty in the computerized PPVT-R. Average amplitude values for the 350–650 ms period (100 ms intervals) were derived from the individual waveforms for subsequent statistical analysis. The analysis focused on the N400, with the peak being scored as the most negative point in the 350–650 ms interval. The PMN was excluded from the analysis because its amplitude is much smaller than the N400 and its phonological nature was not central to evaluating semantic comprehension in stroke patients. Individual waveforms were averaged together in order to derive grand average waveforms for the experimental conditions.

2.3. Neuropsychological assessment

A battery of tests was selected to evaluate both expressive and receptive aspects of language (Lezak, 1995; Spreen and Strauss, 1998). In addition, other cognitive abilities that might be expected to influence performance independent of language functions were assessed, including verbal and visual attention, and working memory, as well as depressive symptomatology. The total time required to administer these tests was approximately 90 min. The brief description of the tests is presented below.

2.3.1. Language measures

The 'Short Version' of the Token Test (De Renzi and Faglioni, 1968) was used to assess receptive auditory sentence comprehension. The PPVT-R (Dunn and Dunn, 1981; Form L) was given to assess auditory comprehension of picture names. The Written Sentence Comprehension section of the PALPA (Kay et al., 1992) evaluated reading comprehension of simple sentences. Controlled Oral Word Association (Benton and Hamsher, 1983) was used to assess the spontaneous production of words (i.e. word fluency). The Boston Naming Test (Kaplan et al., 1983) was given to evaluate the ability to name pictured objects. Repetition of words and phrases was evaluated using the sub-tests from the Boston Diagnostic Aphasia Exam (Goodglass and Kaplan, 1987); this tested connections between receptive and expressive speech.

2.3.2. Attention and memory

Digit Span (Forward and Backward) and Visual Span (Forward and Backward) from the Wechsler Memory Scale – Revised (Wechsler, 1987) were given to assess attention and working memory. A line cancellation task was used to evaluate spatial inattention.

2.3.3. Emotion

Specifically, mood and depression were assessed using a 15-item Geriatric Depression Scale (Fountoulakis et al., 1999; Yesavage et al., 1983).

2.4. Computerized PPVT-R

The computerized PPVT-R (Form M) was used to assess auditory comprehension of vocabulary using ERPs (Byrne et al., 1995a,b; Connolly et al., 1995a). Like the original PPVT-R (Dunn and Dunn, 1981), the computerized PPVT-R has graded levels of difficulty: Level 1 (25 items) contained words from a Preschooler level (2.5–5 years); Level 2 (40 items) contained words from a Child level (10–17 years); and Level 3 (25 items) contained words from an Adult level (advanced).

Each trial began with a picture presentation (duration = 1700 ms), the onset of which was followed 700 ms later by a digitized spoken word (duration = 750–1000 ms) that was either semantically congruent or incongruent (0.5 probability) in relation to the picture. Thus, a picture of a car would be followed by the word 'car' or the word 'plate.' The monochromatic pictures ($N = 90$) were adapted directly from the PPVT-R and paired with congruent and incongruent spoken words ($N = 180$), which were presented on headphones (90 dB, 12 KHz digitization rate). The inter-trial interval was 5.2 s.

Patients were tested in a quiet room and stimuli were presented on a computer monitor positioned between 75 and 100 cm away (depending on the patient's visual acuity). Pictures subtended a visual angle of approximately 7° (at an 85 cm distance). Behavioral responses were not required in order to minimize the task complexity and attentional demands.

2.5. Statistical analyses

ERP data were analyzed using a repeated measures analysis of variance (ANOVA) and where appropriate, post hoc analyses (Tukey's Honestly Significant Difference, HSD) were used (alpha levels ≤ 0.05 were required for statistical significance). The ANOVA was conducted as an omnibus test with Greenhouse-Geisser corrections being applied to the degrees of freedom (Greenhouse and Geisser, 1959). All corrected probabilities are reported. The analyses included Word Type (Congruent and Incongruent), Difficulty (Levels 1, 2, and 3), Time (350–450, 450–550, and 550–650 ms), and Site (Fz, Cz, and Pz) as factors.

The study involved patients who had varying degrees of language deficits. In light of this fact, the results could not be assumed to represent a normally distributed sample. While an ANOVA followed by post-hoc comparisons represented the standard statistical approach in a normally distributed sample, further analyses were required. Correlational analyses, which accommodated both parametric and non-parametric assumptions, were also done and all results were interpreted within the context of both statistical approaches.

The objective of correlational analyses was to quantify predicted ERP effects and correlate them with behavioral performance on the standard PPVT-R. Similar to previous work (Byrne et al., 1995a; Connolly et al., 1999a; D'Arcy and Connolly, 1999; D'Arcy et al., 2000), differences between the Congruent and Incongruent conditions in individual waveforms were analyzed using serial *t*-scores. However, in the current study serial *t*-scores were used to provide an innovative quantification of the ERP differentiation between experimental conditions. This procedure allowed for the calculation of N400 derivative scores.

Marchand et al. (2002) provides a full description of the N400 derivative scoring procedure along with additional empirical data. Briefly, the *t*-scores are particularly useful in cases where a priori predictions exist concerning ERP waveform differentiation. The *t*-scores were used to account for variance in the individual EEG trials that comprised the average (*t*-score: Congruent-Incongruent/[pooled variance/sq root of (n-1)], where $n =$ the total number of trials accepted). Scoring the maximum negative-going peak in the Incongruent condition (350–650 ms) identified the N400. Using the peak latency, a single *t*-score was obtained for each of the 3 levels of difficulty (for all 3 electrode sites). For individual averages in which there were fewer than 20 trials included in the signal averaging process (i.e. low signal-to-noise, or SNR), the *t*-score values were weighted by a scaling factor (*t*-score × [accepted trials/20-trial cut-off]). The purpose of the scaling factor was to adjust the N400 derivative scores compromised by a low SNR (i.e. to control for spurious values). In all, scaling factors were applied to 30 of the 90 *t*-scores (10 waveforms × 3 levels × 3 sites). The sums of the *t*-scores across difficulty levels and/or sites were used to obtain an index of N400 differentiation. For example, N400 differentiation across more than one difficulty level (and more than one site) resulted in higher N400 derivative scores, while the absence of reliable differentiation at any difficulty level (or site) resulted in low N400 derivative scores. This cumulative analysis procedure effectively permitted the N400 derivative scores to be evaluated subsequently in correlational analyses.

The N400 derivative scores were compared to the neuropsychological data using correlational analyses (parametric and non-parametric). All significant correlation coefficients reported were based on a parametric analysis (Pearson *r*) and a non-parametric analysis (Spearman's rho). The N400 derivative scores were correlated with performance on measures of receptive comprehension, the PPVT-R (raw and standard scores) and Token Test (raw scores). Only correlations that were significant for both PPVT-R raw and standard scores are reported. Correlations were also done using measures of attention and working memory (i.e. Digit and Visual Span; WMS-R) in order to verify that the relationship was specific to language, and did not reflect more general impairments (e.g. which would have also affected performance on either Digit Span or Visual Span).

3. Results

3.1. Neurological and neuropsychological data

The diagnostic imaging results and functional rating scores are summarized in Table 1. The patients all had histories involving major risk factors for stroke (e.g. hypertension, hypercholesterolemia, and/or smoking). Also, all 10 patients presented with right-sided motor impairments and 5 patients had signs of sensory impairment (S03, S07, S08, S09, and S10). Three patients presented with some signs of visual impairment (S03, S07, and S09), but their deficits did not interfere with the task of picture identification.

The results of the neuropsychological assessment are summarized in Table 2. Inspection of the neuropsychological data revealed that there was considerable variation in language functions. The findings corroborated existing clinical aphasia classifications for the patients. Four patients (S01, S02, S05, and S08) had some expressive difficulties, but their receptive language functions were relatively intact (although S08 had a moderately low score on the Token Test). The expressive impairments in these patients were revealed by their performance on the Controlled Oral Word Association task, which assessed their ability to spontaneously produce words. In contrast, 5 of the patients (S03, S04, S07, S09, and S10) had both expressive and pronounced receptive deficits. Critically important is the fact that these patients had the lowest scores on the PPVT-R and the Token Test, the two measures that had been selected to assess receptive comprehension. There was one individual (S06) who had no expressive deficits, but did have some difficulties with reading comprehension (PALPA) and lower scores on the PPVT-R. However, this individual's performance on the Token Test (a measure of auditory sentence comprehension, and not vocabulary level) was not impaired. While this pattern of results may have stemmed in part from a low education level (Grade 5 level, lowest in the sample), the discrepancy made it difficult to classify the level of receptive language comprehension for this individual. Overall, the neuropsychological findings were fairly consistent with the lesion data and allowed for an enhanced characterization of the patients' aphasic syndromes (Tables 1 and 2). In light of these findings, the relationship between the ERP and neuropsychological data was investigated further (see Section 3.2.2).

3.2. Electrophysiological data analyses

In order to evaluate the ERP data, two analyses were run: (1) a standard analysis to demonstrate the group effects that are typically examined in ERP studies; and (2) correlational analyses on the ERP and neuropsychological findings in order to validate the N400 differentiation as a measure of receptive comprehension deficits in these patients.

Table 1
Neurological data

Patient	Hemisphere	Region	NIHSS[a]	BI[b]
S01	L	Lacunar, internal capsule, caudate nucleus	5	71/100
S02	L	Frontal, subcortical (hemorrhage)	4	74/100
S03	L	Posterior frontotemporal, basal ganglia, (left carotid occlusion, middle cerebral artery distribution)	13	20/100
S04	L	Frontal-posterior, parietal (middle cerebral artery)	7	57/100
S05	L	Frontal	6	25/100
S06	L	Lacunar, pons	4	45/100
S07	L	Basal ganglia, deep hemisphere, anterior lateral ventricle, sub-cortical	12	35/100
S08	L	No focal abnormality (left carotid occlusion, middle cerebral artery distribution)	9	45/100
S09	L	Frontal-posterior, insular cortex (middle cerebral artery), old right parietal-occipital infarct	8	62/100
S10	L	Parietal lobe (left carotid occlusion)	9	46/100

[a] National Institute of Health Stroke Scale at admission ($M = 7.7$; SD = 3.1).
[b] Barthel Index at admission ($M = 48/100$; SD = 18.2).

3.2.1. Standard analysis

The grand average waveforms from Congruent and Incongruent words at all 3 levels of difficulty are presented in Fig. 1. Examination of the waveforms revealed that Incongruent words elicited both a PMN (*) and a N400 (↓) at Level 1 (Cz and Pz). The mean N400 peak latencies were 507 ms (SD = 46.4, range = 415–576 ms) and 475 ms (SD = 50.4, range = 416–531 ms) for Cz and Pz, respectively. The PMN peaked at approximately 417 ms (Cz) and 414 ms (Pz) in the grand average waveforms. The ERP differentiation did not occur in the grand average waveforms at Levels 2 and 3. A repeated measures ANOVA was conducted with Word Type (Congruent and Incongruent), Difficulty (Levels 1–3), Time (350–450, 450–550, and 550–650 ms), and Site (Fz, Cz, and Pz) as factors.

There were no significant main effects. There was a significant Word Type × Site interaction ($F(2, 18) = 6.023$, $P < 0.05$, $\epsilon = 0.571$) and post-hoc analyses revealed that the N400 was characterized by a centro-parietal distribution (largest at Pz). There was also a significant Time × Site interaction ($F(4, 36) = 6.014$, $P < 0.01$, $\epsilon = 0.614$). However, the results are best interpreted within the significant Word Type × Time × Site interaction ($F(4, 36) = 4.037$, $P < 0.05$, $\epsilon = 0.473$). The largest N400 amplitudes were in

Table 2
Neuropsychological data[a]

	S01	S02	S03	S04	S05	S06	S07	S08	S09	S10
Token (Raw/36)	33	32	12	9.5	36	31.5	13.5	24	14.5	11.5
PPVT-R (Standard)	86	80	50	59	99	66	40	85	44	70
PALPA (Raw/30)	29	25	17	–	30	22	15	17	11	13
COWA (% ile)	11	38	–	–	25	96	–	1	1	–
BNT (Raw 60)	55	49	–	–	53	46	–	50	–	–
Repetition (Raw 26)	25	23	–	–	26	22	17	25	21	6
DS (F)	6	7	–	–	5	6	4	6	5	4
DS (B)	5	5	–	–	3	4	2	3	3	2
VS (F)	5	5	5	5	6	3	4	6	4	4
VS (B)	4	4	4	4	5	4	4	4	3	3
Line Canc. (Errors L:R)	00	00	00	00	00	00	10	00	00	77
GDS (Raw 15)	3	1	10	5	1	4	0	11	7	7

[a] PPVT-R, Peabody Picture Vocabulary Test-Revised; PALPA, Psycholinguistic Assessments of Language Processing in Aphasia – Reading Comprehension sub-test; COWA, Controlled Oral Word Association; BNT, Boston Naming Test; DS, Digit Span; VS, Visual Span; and GDS, Geriatric Depression Scale. Blanks denote that the test was not given or was discontinued due to impairment.

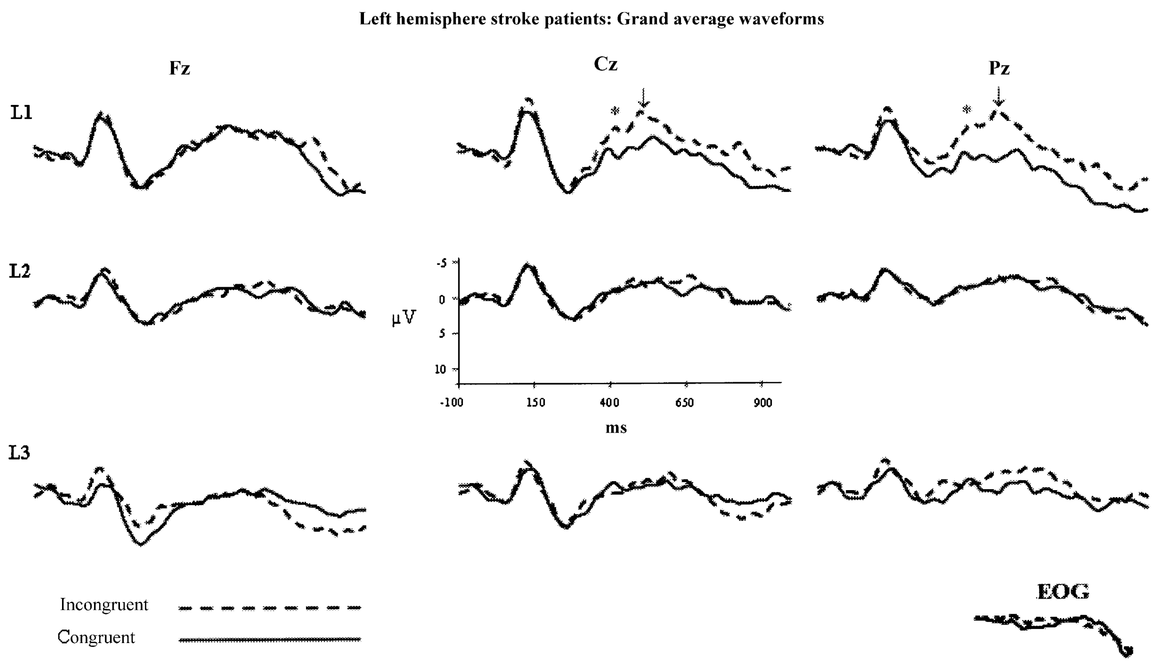

Fig. 1. Grand average waveforms from left-hemisphere stroke patients from the computerized PPVT-R. The computerized PPVT-R contains 3 levels of difficulty (L1, L2, and L3). For an individual trial, a picture is presented and then a digitized spoken word follows that is either Congruent or Incongruent (e.g. a picture of a car; 'car' or 'plate'). Incongruent spoken words (dashed line) were differentiated from the Congruent spoken words (solid line) on the basis of PMN (*) and N400 (↓) components for items in Level 1. Time (ms) is on the x-axis and Amplitude (μV) is on the y-axis. Negative is plotted up.

the Incongruent condition (across all time intervals). The N400 to Incongruent words began during Time 1 (350–450 ms) and was largest during Time 2 (450–550 ms). The N400 distribution was characterized by a significant centro-parietal topography (Cz and Pz), with the largest amplitudes at Pz. This finding was consistent with the typical timing and distribution of the N400. N400 amplitudes in the Congruent condition remained relatively stable across the Time and Site factors.

3.2.2. ERP and neuropsychological convergence

There was a high degree of correspondence between the ERP and neuropsychological data. The individual waveforms for all 10 patients (Levels 1 and 2) are presented in Fig. 2. There was no reliable N400 differentiation at Level 3 (cf. Fig. 1). Visual examination of the waveforms showed that, while there was variance, the N400 differentiation generally matched performance on the PPVT-R. In order to quantify the N400 differentiation, derivative scores were calculated for each individual, these measured the difference between the Congruent and Incongruent conditions. In light of the centro-parietal distribution, the N400 derivative scores comprised the sum of Cz and Pz for each difficulty level (L1, L2, and L3). Also, combined N400 derivative scores were then obtained for Levels 1 and 2 (L12) and L1–L3 (L123). Correlations were calculated between the N400 derivative scores and the scores for the PPVT-R as well as the Token Test. Scores from the Digit Span and Visual Span tasks were also examined (see Section 2 for details).

There was significant correlation between L1 and PPVT-R scores (Pearson $r = 0.693$, $P < 0.05$), revealing that higher scores on the PPVT-R were related to higher N400 derivative scores at L1. Fig. 3 shows a scatter plot of the positive correlation between L1 and the PPVT-R scores; a linear relationship is evident (Marchand et al., 2002). The correlations for L2 and L3 with the PPVT-R scores were not significant. However, it is noteworthy that the coefficients were distributed in a pattern matching the graded levels of difficulty (L1 = 0.693; L2 = 0.479; and L3 = 0.177). There was also a significant correlation between L12 and the PPVT-R scores (Pearson $r = 0.735$, $P < 0.05$). In contrast to the previous correlation, the scatter plot may reflect either a linear or a curvilinear relationship (e.g. quadratic curve estimation; $F = 4.67$, $P = 0.051$). Examination of the scatter plot reveals a distinct cluster of patients who had both high N400 derivative and PPVT-R scores (Fig. 4). Remarkably, these patients corresponded to those individuals who had been identified previously on the basis of the neuropsychological data as having predominately expressive deficits (i.e. S01, S02, S05, and S08). In contrast, the patients who presented with more pronounced receptive deficits (with the exception of S09) clustered in the lower range. The correlation between L123 and the PPVT-R scores was not significant (Pearson $r = 0.627$, $P = 0.052$). This result is not surprising because N400 derivative scores at L3 reflected primarily waveform variance.

In addition to the PPVT-R, there was also a significant correlation between N400 derivative scores at L1 and the Token Test scores (Pearson $r = 0.658$, $P < 0.05$). This finding reflected the fact that high N400 derivative scores

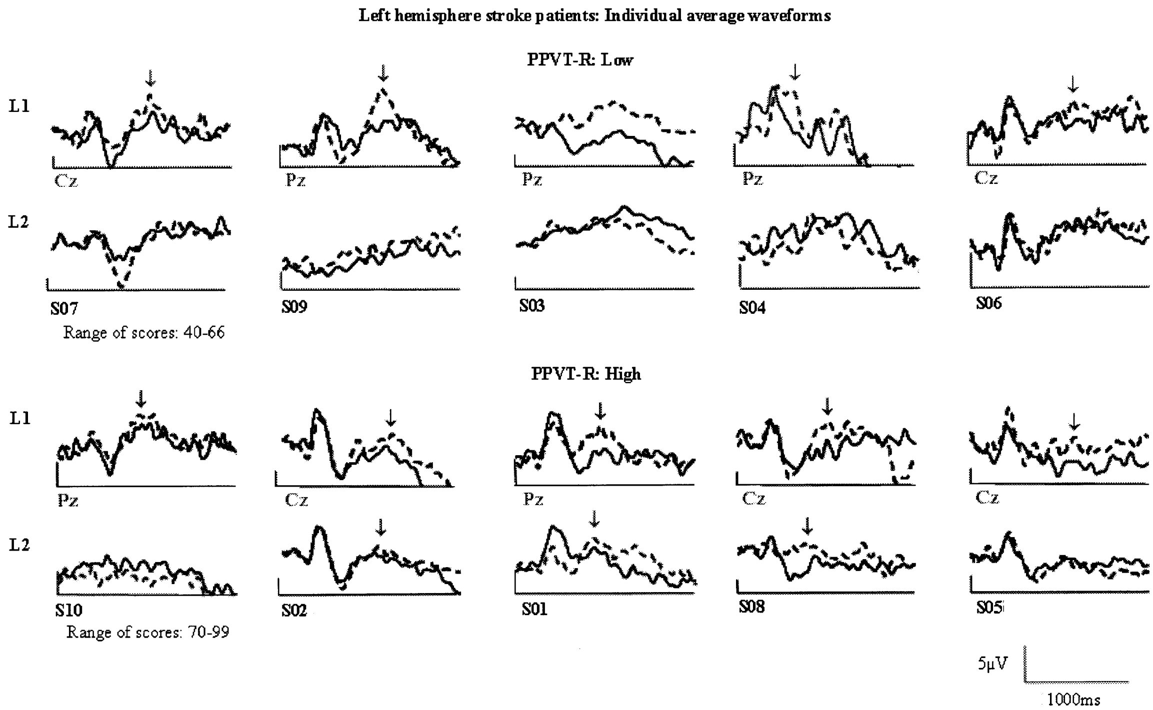

Fig. 2. Individual average waveforms (all 10 patients) for Levels 1 and 2. Data from Level 3 are not shown because there was no reliable N400 differentiation at this level (see Section 3). The sample is divided into Low and High groups on the basis of the PPVT-R performance (standard scores). Note the general correspondence between N400 differentiation (↓) and PPVT-R performance. However, clear identification of the N400 differentiation was not straightforward in all instances (e.g. S09). Accordingly, this relationship was subsequently quantified and confirmed using N400 derivative scores (Figs. 3 and 4). All other details as for Fig. 1.

were related to better performance on the Token Test. Similar to the prior results, the coefficient values were reduced in a graded manner as the level of difficulty increased (L1 = 0.658; L2 = 0.430; and L3 = 0.260), confirming the relationship. In contrast to the PPVT-R standard scores, the Token Test correlations for L12 were not reliably significant across both parametric and non-parametric analyses. There was no significant correlation

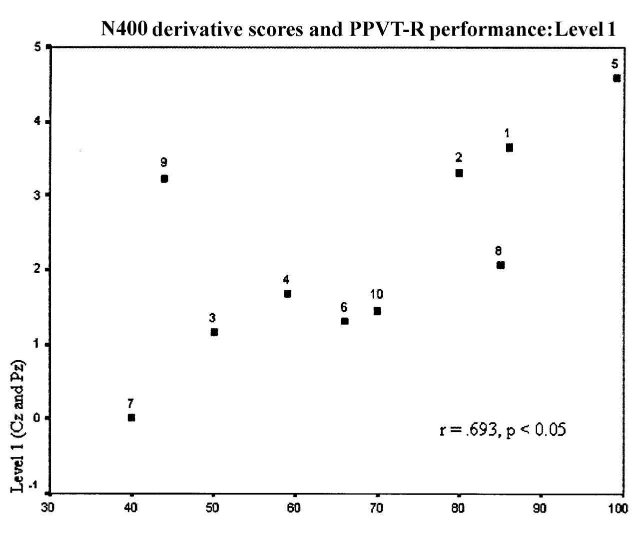

Fig. 3. N400 derivative scores from Level 1 (y-axis) were positively correlated with standard scores from the PPVT-R (x-axis). The scatter plot shows a linear function in which N400 derivative scores correlate with higher standard scores. See text for details.

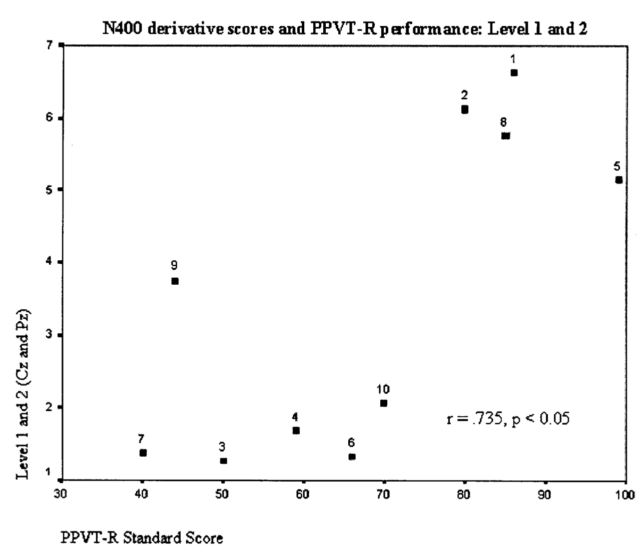

Fig. 4. N400 derivative scores summed across Levels 1 and 2 (y-axis) were also positively correlated with standard scores from the PPVT-R (x-axis). The scatter plot reveals that summing N400 derivative scores from both Level 1 and 2 resulted in the ability to identify patients with expressive deficits and relatively intact receptive function as a distinct cluster (i.e. a bi-modal distribution). Whereas, the patients who had more pronounced receptive impairments (with the exception of one outlier, S09) clustered in the lower range. Note that the y-axis scale has increased as a result of the cumulative N400 derivative scoring procedure. See text for details.

between the N400 derivative scores and performance on the Digit Span (e.g. Pearson $r = 0.490$, $P = 0.218$, $N = 8$, L1 and Total score) and Visual Span (e.g. Pearson $r = 0.483$, $P = 0.158$, $N = 10$, L1 and Total score). Overall, the pattern of results indicated that the N400 differentiation was highly specific to receptive language comprehension.

4. Discussion

The findings supported the hypothesis that ERP responses on the computerized PPVT-R can evaluate accurately a left-hemisphere stroke patient's ability to understand spoken words. Specifically, the incongruent spoken words elicited a N400 component, with a characteristic centro-parietal distribution (Fig. 1). The N400 has been linked to semantic analysis associated with the processing of the incongruent word (Kutas and Van Petten, 1994). Visual inspection of the waveforms revealed that the presence of the N400 generally matched vocabulary comprehension on an individual patient basis (Fig. 2). However, the peak latencies were slightly delayed relative to the prototypical N400 latency and there was notable individual variance. The latency delay and the variance were likely associated with the fact that the sample was comprised of older stroke patients. While not the central focus of this study, a PMN was also elicited by the unexpected phonological speech input and further validated the ERP results (Connolly et al., 1995a, 2001; Connolly and Phillips, 1994).

In order to quantify the ERP differentiation, derivative scores were calculated for the N400. N400 derivative scores from Level 1 correlated positively with overall behavioral performance on the PPVT-R in a patient group with varying degrees of language impairment following stroke. The N400 derivative scores from Level 1 increased linearly as a function of performance on the PPVT-R (Fig. 3). Further, there was a graded reduction in the correlations across the increasing levels of vocabulary difficulty. These results were supported by complimentary evidence for a strong relationship between the N400 derivative scores and the Token Test, which provided an additional measure of receptive speech comprehension. In contrast, the N400 derivative scores did not correlate with attention and working memory performance. This pattern of results showed that N400 differentiation was specific to receptive language comprehension in these patients.

The clinical evidence (neurological and neuropsychological data; Tables 1 and 2, respectively) revealed that two general sub-groups of patients existed: one that had expressive problems with relatively intact performance on receptive measures; and a second that had variable levels of *both* expressive and receptive deficits. The latter sub-group had the lowest scores on the PPVT-R and the Token Test. When the N400 derivative scores were summed across Levels 1 and 2, they also correlated significantly with performance on the PPVT-R. Indeed, this summed measure (across the first two levels of vocabulary difficulty) appeared to be particularly sensitive to the nature of aphasic deficits. Examination of the scatter plot (Fig. 4) showed that data points for the sub-group of patients with relatively intact receptive function were separated as a cluster of scores in the higher range of both N400 derivative measures and PPVT-R scores. In contrast, the sub-group who performed poorly on the receptive measures clustered, for the most part, in the lower ranges of the N400 derivative and PPVT-R indices. This result provided a valuable demonstration of the potential for ERP assessment of receptive language impairments. Ongoing work is directed at better characterizing the sensitivity and specificity of this ERP differentiation (Marchand et al., 2002).

Presumably, this electrophysiological measure of receptive comprehension relates in some manner to the functional status of N400 intracranial generators. The left perisylvian area has been implicated with respect to primary N400 generators in speech processing, particularly the superior temporal cortex in the vicinity of the auditory cortex (Helenius et al., 2002). However, there is additional evidence from the visual modality that N400 sources can be distributed throughout the cortex and that the foci for these sources may vary between individuals (Haan et al., 2000; Helenius et al., 1998; McCarthy et al., 1995). In future work, it will be critical to continue integrating lesion data with N400 source data in order to better understand the changes in ERPs as a function of aphasia (Friederici et al., 1999).

Two caveats concerning the correlational results exist. First, the findings provided clear evidence that the sample was not normally distributed, but was in fact bi-modal. While the non-parametric nature of the sample was taken into account a priori, future studies should examine patients with only expressive or only receptive deficits. Second, while research with patient populations typically involves lower sample sizes, the small sample size used in this study ($N = 10$) warrants replication in order to verify the correlations between the N400 derivative scores and behavioral performance. However, the consistent pattern of results between the PPVT-R and the Token Test suggests that the present findings are representative. The above points notwithstanding, the results indicated that the N400 differentiation (previously demonstrated to reflect intact processing of picture-spoken word congruence) could be quantified and linked to behavioral performance in a stroke patient population.

The evidence for the clinical efficacy of ERP assessment (subsumed under the Innovative Methods of Assessment Program, see Section 1) has been amply demonstrated in both normative and single case studies (Connolly and D'Arcy, 2000; Connolly et al., 2000). The present experiment provides confirmation of those reports in the first empirical study of a patient group. It is noteworthy that this technique contributes invaluable information about the level of functioning, particularly when the clinical question revolves around issues related to speech comprehension. For example, one of the most common referral questions we receive is 'Does this patient understand speech?' A positive result is critical with respect to this question – it provides an answer that can significantly influence subsequent intervention.

With respect to the implementation of this technique, it is necessary to mention some key diagnostic guidelines. If ERP differentiation is present (either visually or based on statistical analyses), then interpretation of the results must be made within the wider clinical/neuropsychological evaluation. If this differentiation does not occur, then standard approaches must be followed, including the administration of additional tests aimed at clarifying the negative result. From a practical perspective, the implementation, administration, and analysis of this technique are completely feasible within a hospital setting. And, while the results of this investigation address issues of language comprehension, it should be noted that work on applying ERP assessment methods to a wider range of cognitive functions (e.g. attention and memory) has begun.

Acknowledgements

This work was funded by the Scottish Rite Charitable Foundation of Canada and the Natural Sciences and Engineering Research Council of Canada (JFC and RCND). G. Eskes was also supported by a Department of Psychiatry Faculty Development Award and the Department of Health Designated Mental Health Research Fund. The authors wish to thank D. Benedict and C. Striemer for valuable assistance with this study.

References

Benton AL, Hamsher K. Multilingual aphasia examination. Iowa City: AJA, 1983.

Byrne JM, Dywan CA, Connolly JF. An innovative method to assess the receptive vocabulary of children with cerebral palsy using event-related brain potentials. J Clin Exp Neuropsychol 1995a;7:9–19.

Byrne JM, Dywan CA, Connolly JF. Assessment of children's receptive vocabulary using brain event-related potentials: development of a clinically valid test. Child Neuropsychol 1995b;1:211–23.

Chiappa KH. Evoked potentials in clinical medicine, 3rd ed. New York: Lippincott-Raven, 1997.

Connolly JF, D'Arcy RCN. Innovations in neuropsychological assessment using event-related brain potentials. Int J Psychophysiol 2000;37: 31–47.

Connolly JF, Phillips NA. Event-related potential components reflect phonological and semantic processing of the terminal word of spoken sentences. J Cogn Neurosci 1994;6:256–66.

Connolly JF, Byrne JM, Dywan CA. Event-related brain potentials reflect the receptive vocabulary of individuals as measured by the Peabody Picture Vocabulary Test-Revised: a study of cross-modal and cross-form priming. J Clin Exp Neuropsychol 1995a;17:548–65.

Connolly JF, Phillips NA, Forbes KAK. The effects of phonological and semantic features of sentence-ending words on visual event-related brain potentials. Electro Clin Neuro 1995b;94:276–87.

Connolly JF, Major AM, Allen SL, D'Arcy RCN. Performance on WISC-III and WAIS-R NI vocabulary subtests assessed with event-related brain potentials: an innovative method of assessment. J Clin Exp Neuropsychol 1999a;21:444–64.

Connolly JF, Mate-Kole CC, Joyce BM. Global aphasia: an innovative assessment approach. Arch Phys Med Rehabil 1999b;80:1309–15.

Connolly JF, D'Arcy RCN, Newman RL, Kemps R. The application of cognitive event-related brain potentials in language-impaired individuals: review and case studies. Int J Psychophysiol 2000b;38:55–70.

Connolly JF, Service E, D'Arcy RCN, Kujula A, Alho K. Phonological aspects of word recognition as revealed by high-resolution spatio-temporal brain mapping. Neuroreport 2001;12:237–43.

D'Arcy RCN, Connolly JF. An event-related brain potential study of receptive speech comprehension using a modified Token Test. Neuropsychologia 1999;37:1477–89.

D'Arcy RCN, Connolly JF, Eskes GA. Evaluation of reading comprehension with neuropsychological and event-related brain potential (ERP) methods. J Int Neuropsychol Soc 2000;6:552–63.

Dehaene-Lambertz G, Dupoux E, Gout A. Electrophysiological correlates of phonological processing: a cross-linguistic study. J Cogn Neurosci 2000;12:635–47.

De Renzi E, Faglioni P. Normative data and screening power of a shortened version of the token test. Cortex 1968;14:41–9.

Dunn LM, Dunn LM. Peabody picture vocabulary test-revised. Minnesota: American Guidance Service, 1981.

Enderby P, Philipp R. Speech and language handicap: towards knowing the size of the problem. Br J Disord Commun 1986;21:151–65.

Enderby P, Wood VA, Wade DT, Hewer RL. Aphasia after stroke: a detailed study of recovery in the first 3 months. Int Rehabil Med 1987;8: 162–5.

Fountoulakis KN, Tsolaki M, Iacovides A, Yesavage J, O'Hara R, Kazis A, Ierodiakonou C. The validation of the short form of the Geriatric Depression Scale (GDS) in Greece. Aging (Milano) 1999;11:367–72.

Friederici AD, von Cramon DY, Kotz SA. Language related brain potentials in patients with cortical and subcortical left hemisphere lesions. Brain 1999;122:1033–47.

Goodglass H, Kaplan EF. The assessment of aphasia and related disorders, 2nd ed. Philadelphia: Lea & Febiger, 1987.

Greenhouse SW, Geisser S. On methods in the analysis of profile data. Psychometrika 1959;24:95–112.

Haan H, Streb J, Bien S, Rösler F. Individual cortical current density reconstructions of the semantic N400 effect: using a generalized minimum norm model with different constraints (L1 and L2 norm). Hum Brain Mapp 2000;11:178–92.

Hagoort P, Brown C. ERP effects of listening to speech: semantic ERP effect. Neuropsychologia 2000;38:1518–30.

Helenius P, Salmelin R, Service E, Connolly JF Distinct time courses of word and context comprehension in the left temporal cortex. Brain 1998;121:1133–42.

Helenius P, Samelin R, Service E, Connolly JF. Leinonen S, Lyytinen H. Cortical activation during spoken-word segmentation in non-reading impaired and dyslexic adults. J Neurosci 2002;22:2936–44.

Holcomb PJ, Neville HJ. Auditory and visual semantic priming in lexical decision: a comparison using evoked potentials. Lang Cogn Process 1990;5:281–312.

Kaplan EF, Goodglass H, Weintraub S. The Boston Naming Test, 2nd ed. Philadelphia: Lea & Febiger, 1983.

Kay J, Lesser R, Coltheart M. Psycholinguistic assessments of language processing in aphasia. East Sussex, UK: Lawrence Erlbaum Associates, 1992.

Knight RT. Electrophysiological methods in behavioral neurology and neuropsychology. In: Feinberg TE, Farah MJ, editors. Behavioral neurology and neuropsychology. New York: McGraw-Hill, 1997. p. 101–20.

Kutas M, Van Petten C. Psycholinguistics electrified: event-related brain potential investigations. In: Gernsbacher MA, editor. Handbook of psycholinguistics. San Diego: Academic Press, 1994. p. 83–143.

Lezak M. Neuropsychological assessment. New York: Oxford University Press, 1995.

Marchand Y, D'Arcy RCN, Connolly JF. Linking neurophysiological and neuropsychological measures for aphasia assessment. Clin Neurophysiol 2002;113:1715–22.

Marquardsen J. The natural history of acute cerebrovascular disease. A retrospective study of 769 patients. Acta Neurol Scand Suppl 1969;38: 67–8.

McCarthy G, Nobre AC, Bentin S, Spencer DD. Language-related field potentials in the anterior-medial temporal lobe: I. Intracranial distribution and neural generators. J Neurosci 1995;15:1080–9.

Morse PA, Montgomery CE. Neuropsychological evaluation of traumatic brain injury. In: White RF, editor. Clinical syndromes in adult neuropsychology: the practitioner's handbook. Amsterdam: Elsevier, 1992. p. 86–176.

Oldfield RC. The assessment and analysis of handedness: The Edinburgh Inventory. Neuropsychologia 1971;9:97–113.

Pedersen PM, Jørgensen HS, Nakayama G, Raaschou HO, Olsen TS. Aphasia in acute stroke: incidence, determinants, and recovery. Ann Neurol 1995;38:659–66.

Praamstra P, Stegeman DF. Phonological effects on the auditory N400 event-related brain potential. Brain Res Cogn Brain Res 1993;1:73–86.

Spreen O, Strauss E. A compendium of neuropsychological tests: administration norms and commentary. New York: Oxford University Press, 1998.

van Den Brink D, Brown C, Hagoort P. Electrophysiological evidence for early contextual influences during spoken-word recognition: N200 versus N400 effects. J Cogn Neurosci 2001;13:967–85.

Wade DT, Hewer RL, David RM, Enderby PM. Aphasia after stroke: natural history and associated defects. J Neurol Neurosurg Psychiatry 1986;49:11–16.

Wechsler D. Memory Scale-Revised. New York: Psychological Corporation, 1987.

Yesavage JA, Brink TL, Rose TL, Lum O, Huang V, Adey M, Leirer VO. Development and validation of a geriatric depression rating scale: a preliminary report. J Psychiat Res 1983;17:37–49.