

# The Assessment of Reliability of Cognitive Evoked Potential in Normal Person

Ji Seong Hong, MD, Ji Hoon Lee, MD, Yeo Hoon Yoon, MD, Jung Hwa Choi, MD,  
Jae Eun Shin, MD, Sun Mi Kim, MD, Yoon Ghil Park, MD

Department of Rehabilitation Medicine and Rehabilitation, Institute of Neuromuscular Disease,  
Yonsei University College of Medicine, Seoul, Korea

**Objective** To evaluate intra-tester reliability of P300 more precisely, this study was designed. Event-related potential (ERP) is the result of endogenous brain response following cognitive stimulus. The P300 component of the human ERP is a positive wave with a latency of 300 ms or greater. Our purpose of this study was to estimate reliability of P300 latency and amplitude with 30 normal persons without head injury, as well as to set up them as the reference values in the event that they would be found to be highly reliable.

**Methods** ERP was performed at three separate times on 30 normal adults in their 20s and 30s. We measured P300 latency and amplitude among ERP.

**Results** P300 latency show excellent reliability with intraclass correlation coefficient (ICC) of 0.81. As to P300 amplitude, reliability was good to fair with ICC of 0.53. Average value of P300 latency was  $311.3 \pm 37.0$  ms, shorter than reference value of previous study in Korea.

**Conclusion** P300 latency revealed higher reliability than P300 amplitude, although reliability of P300 was confirmed in both component. After further study including precise mechanism, influence factor on measurement and method standardization, it is expected to be an objective indicator to assess the cognitive state and predict prognosis

**Keywords** Event-related potentials, P300 component, Reliability

## INTRODUCTION

Neural activity is an electric action, and thus, cerebral responses to stimuli can be measured through brainwav-

es. In particular, the waveform drawn from cerebral electrical activities processed by repetitive stimuli, including specific information, is named event-related potential (ERP) [1], which is measured by a non-invasive analysis of electrophysiological phenomena that are caused by stimuli and arise in the cerebral cortex, with the calculation of the average. This potential is known to be generated in a process of cerebral perception or recognition [2,3].

Clinically, the ERP can be used in patients with dementia, depression, Parkinson disease, hydrocephalus, schizophrenia, metabolic encephalopathy, and central auditory processing disorder [4,5].

---

Received June 25, 2012; Accepted October 18, 2012  
Corresponding author: Yoon Ghil Park  
Department of Rehabilitation Medicine, Gangnam Severance Hospital,  
211 Eonju-ro, Gangnam-gu, Seoul 135-720, Korea  
Tel: +82-2-2019-3493, Fax: +82-2-3463-7585, E-mail: [drtlc@yuhs.ac](mailto:drtlc@yuhs.ac)

© This is an open-access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0>) which permits unrestricted noncommercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

Copyright © 2013 by Korean Academy of Rehabilitation Medicine

The waveform of ERP is made up of various components, e.g., N100, N200, P300, N400, and P600; especially, the most well-known is the P300, positive potential with the average latency of 300 ms, which is known to be useful in assessing the cognitive disorder [6-8]. The amplitude of P300 is related to the quantity of nerve resources used to process data, i.e., regeneration of memory, and its latency is a reflection of the time it takes the brain to register the meaning of the stimulus and is considerably more apparent for rare target stimuli than for frequent non-target stimuli [9].

In 1993, Segalowitz and Barnes [10] suggested that total variance of an ERP measure can be expressed as a function of 1) trait variance that represents stable subject characteristics, such as age, gender, cognitive capacity, working memory, intelligence, learning disability or dementia; 2) stimulus variance that is systematically manipulated in the experimental paradigm, such as target frequency or intensity; 3) variance due to subject's psychological or physiological state independent of stimulus properties, such as arousal, nicotine, and caffeine use; and 4) measurement error. Among them, state variance and measurement error are critical in determining the basic reliability of the ERP measure itself and present an upper bound on its validity.

In previous studies that use normal population as subjects, P300 latency was highly reliable, but reliability of P300 amplitude showed comparative controversial results. Moreover, test-retest reliability was assessed only twice, not assessed by the multiple repetitive tests [10-12].

Our aim of this study is to investigate the intra-rater reliability of the amplitude and latency of P300, confirmed by an agreement of the test score from one occasion to the others if multiple repetitive tests were done in normal young adults. The other aim is to establish them as the reference values in the event that they are found to be highly reliable.

## MATERIALS AND METHODS

### Subjects

This study was conducted on men and women in their 20s and 30s (age range, 25 to 34 years), who were working at the general hospital as medical doctors without trouble hearing and central nerve system, without medical histories of psychotic disorder or active substance abuse, serving at hospitals, and expressed their intentions to

**Table 1.** General characteristics

Characteristic	Male (n=17)	Female (n=13)
Age (20s:30s)	9:8	13:0

participate in this study. They were composed of 17 men and 13 women, whose average age was  $28.2 \pm 2.99$  years (Table 1). Written informed consent was obtained from all participants.

### Methods

ERP was measured at three separate times by the same researcher, where after the reliability of P300 latency and amplitude was analyzed. Subjects in headphones were given auditory stimuli by use of oddball paradigms, bilaterally. The oddball paradigm is to familiarize subjects with experimental conditions without needing attention, and then to make them pay attention to distinguish target stimuli from non-target stimuli to process ERP. The test was conducted by the use of Medelec Synergy (VIASYS HealthCare, Surrey, UK) electromyography machine. The non-target stimulus (standard stimulus) was given once per second (1 peak particle velocity) in a low note at 1,000 Hz, and the target stimulus was given in a high note at 2,000 Hz. The ratio of non-target stimuli to target stimuli (the oddball ratio) was 4 to 1. The target stimulus was given 60 times to all at a stimulus intensity of 70 dB and a stimulus rate of 1 time/sec. Sweep speed and sensitivities were adjusted at 500 ms/division and 2  $\mu$ V/division, respectively and data were recorded by averaging technique. The signal was initially filtered at 0.1–50 Hz.

Each subject was allowed to hear both stimuli and was given a full explanation in advance of the test. While undergoing the test, each of them was lying on a medical bed with relaxed mind. In addition, random stimuli were given to subjects lying straight on their back with their eyes closed in order to prevent prediction of when the target stimulus would be given, and they were asked to count the total number of target stimuli to themselves to encourage them to pay attention.

To reduce interference from electrode response, we used electrodes that were made of silver chloride plated with silver (Ag/AgCl), which were placed on the head. Standard electrodes were placed on both mastoid processes (A1 and A2) and the active electrodes and the ground electrode were placed on 'Cz' and the glabella close to the active electrode.

**Table 2.** Intraclass correlation coefficient (ICC) for latency

	ICC	95% CI		F-test			
		Lower limit	Upper limit	Value	df 1	df 2	p-value
Single measures	0.584	0.381	0.755	5.215	29	58	0.000
Average measures	0.808	0.648	0.902	5.215	29	58	0.000

CI, confidence interval; df, degree of freedom.

**Table 3.** Intraclass correlation coefficient (ICC) for amplitude

	ICC	95% CI		F-test			
		Lower limit	Upper limit	Value	df 1	df 2	p-value
Single measures	0.268	0.046	0.510	2.099	29	58	0.008
Average measures	0.524	0.127	0.757	2.099	29	58	0.008

CI, confidence interval; df, degree of freedom.

Reliability was evaluated by using intraclass correlation coefficient (ICC). ICC can be used when between two or more different target groups measure the level of internal consistency, so it is suitable for identifying. To calculate the outcomes, we applied modern ICC definition. ICC means ratio of the relative variance of intra-rater with relation to variance of intra-rater and residual on the assumption that it had no interaction between the raters. Suppose that the observation X for the *i*th object in the *j*th measurement is  $x_{ij}=m+r_i+c_j+rc_{ij}+e_{ij}$ , where  $i=1, \dots, n, j=1, \dots, k$ , *m* (the population mean for all observations) is constant,  $r_i$  (the object effects) are random, independent, and normally distributed with the mean 0 and variance  $\sigma_r^2$ ,  $c_j$  (the measurement effects) are random, independent, and normally distributed with the mean 0 and variance  $\sigma_c^2$ ,  $rc_{ij}$  (the interaction effects between objects and measurements) are random, independent, and normally distributed with the mean 0 and variance  $\sigma_{rc}^2$ , and  $e_{ij}$  (residual effects) are random, independent, and normally distributed with the mean 0 and variance  $\sigma_e^2$ . Calculating a single score or average score ICCs for consistency in two-way models is given in  $\frac{\sigma_r^2}{\sigma_r^2+(\sigma_{rc}^2+\sigma_e^2)}$  and  $\frac{\sigma_r^2}{\sigma_r^2+\frac{(\sigma_{rc}^2+\sigma_e^2)}{k}}$ .

Calculating a single score or average score ICCs for absolute agreement in two-way models is given in  $\frac{\sigma_r^2}{\sigma_r^2+(\sigma_c^2+\sigma_{rc}^2+\sigma_e^2)}$  and  $\frac{\sigma_r^2}{\sigma_r^2+\frac{(\sigma_c^2+\sigma_{rc}^2+\sigma_e^2)}{k}}$  (Tables 2, 3). Larger ICC resulted in higher reliability. The data were analyzed by using the program SPSS ver. 18.0 (SPSS, Chicago, IL, USA) and calculated by confirming the ICC within a reliability analysis in the menu.

**Table 4.** Means, standard deviation and intra-rater reliability of P300 latency and amplitude

	Latency (ms)	Amplitude (mV)
Latency 1st measurement	300.9±36.8	5.25±3.22
Latency 2nd	315.9±39.8	4.59±3.24
Latency 3rd	317.1±33.1	5.02±3.66
Average latency	311.3±37.0	4.95±3.35
ICC between three separate measurement	0.81	0.53

ICC, intraclass correlation coefficient.

**RESULTS**

The mean latency of P300 indicated 300.9±36.8, 315.9±39.8, and 317.1±33.1 ms at the first, second, and third studies, respectively; thus, the average latency was worked out at 311.3±37.0 ms (Table 4). In a previous study on the latency of P300 in normal Koreans, for reference, the normal values of P300 were 324.10±15.55 and 327.77±15.63 ms in subjects in their 20s and 30s, respectively [13]. In this study, the average latency in subjects in their mid-20s to mid-30s was shorter than both reference values of 20s and 30s.

The mean amplitude of P300 indicated 5.25±3.22, 4.59±3.24, and 5.02±3.66 mV at the first, second, and third studies, respectively, and thus, the average amplitude was worked out at 4.95±3.35 mV (Table 4).

The reliability of P300 latency was rated excellent with ICC of 0.81, whereas, the reliability of P300 amplitude was to be fair to good with ICC of 0.53 (Table 4). The

**Table 5.** Analysis of variance table for latency

	Sum of squares	Degree of freedom	Mean square	False	Significance
Between groups	84,840.156	29	2925.523	-	-
Within groups					
Regression	4,884.206	2	2,442.103	4.353	0.017
Residual	32,535.961	58	560.965	-	-
Total	37,420.167	60	623.669	-	-
Total	122,260.322	89	1,373.711	-	-

**Table 6.** Analysis of variance table for amplitude

	Sum of squares	Degree of freedom	Mean square	False	Significance
Between groups	508.765	29	17.544	-	-
Within groups					
Regression	6.755	2	3.378	0.404	0.669
Residual	484.803	58	8.359	-	-
Total	491.558	60	8.193	-	-
Total	1,000.323	89	11.240	-	-

process of statistical hypothesis test for P300 latency ICC value was described in Tables 3, 5. And P300 amplitude ICC value was described in Tables 3, 6.

## DISCUSSION

The study on the cognitive process and function of humankind is mightily important in the comprehension of high-level mental activities and the sequelae of brain diseases, as well as in the setup of therapeutic plans. The measurement of ERP, generated by specific stimuli in the brain-based mechanisms involved in the cognitive processes, has made such studies more objective and reliable.

In recent times, studies have been made of endogenous potentials, especially with P300 that are evoked not by physical stimuli directly, but by resulting mental activities, such as memory, perception and attention [14]. In 2003, Hodo [15] reported that the amplitude of P300 was lower in patients with schizophrenia than in the normal control group, but the latency was longer in the patients.

In this study, it was found that the reliability of P300 latency in the auditory oddball paradigm had a high agreement of the test score with an ICC of 0.81. The reliability optimizes the differences of P300 latency caused by trait variance like diseases, such as schizophrenia and dementia, and is supportive of its use in clinical and experimental studies as a psychophysiological screening tool.

P300 amplitude was found to have a fair to good reliability with an ICC of 0.53. Thus, P300 latency is presumed to be more reliable than P300 amplitude in clinical applications.

Regarding the reliability of P300 in the auditory oddball paradigm, in 1993 Segalowitz and Barnes [10] assessed the test-retest reliability in 19 growing adolescents twice at intervals of 2 years, and the ICCs of latency and amplitude were worked out at 0.76 and 0.61, respectively. On the authority of the results, they checked the temporal stability along with the test score agreement with a view to ascertain whether the stability of P300 maintained in changes during development or degree of variability over time. In this study, on the other hand, the test was conducted on adults in their 20s and 30s three times without time interval, which was focused on the low measurement error and agreement of the test score.

In 2000, Lee et al. [16] conducted the test on 30 patients with encephalopathy four times and found that P300 latency had a high reliability with the alpha value of 0.977. However, they did not measure P300 amplitude and dealt with patients with encephalopathy. In the methodology; therefore, their study is different from this study conducted on the normal.

In 2006, Hall et al. [11] assessed the test-retest reliability in 19 monozygotic twins twice at the intervals of 7 to 56 days, and as a result, the ICCs of latency and amplitude were worked out at 0.88 and 0.85, respectively. However,

their study was different from this study in sample size, frequency and time interval. In this study, moreover, subject was given 60 target stimuli trial in each session, whereas 80 target stimuli trial was given in each session in their study; in case the target stimuli increase in number, invalid variance is reduced [17]. In the study of Hall et al. [11] P300 amplitude and latency were all excellent in reliability, whereas in this study amplitude was inferior to the latency in reliability. Amplitude varies depending upon the arousal state of the subject. The P300 latency component is only 20 ms longer with severe drowsiness; whereas the amplitude changes dramatically for the state components as the subject's arousal level drops [18]. Thus, it is possible to explain the reason why amplitude is inferior to the latency in reliability in the process of repetitive tests.

In 1986, Polich [12], who assessed the test-retest reliability twice in 100 college students (mean age, 20.4 years), reported that P300 amplitude had a high reliability with a Pearson correlation of 0.71. However, due to data presented as Pearson correlation, it represented inter-subject stability rather than score agreement. However, this study was focused on score agreement as the data were presented in ICC.

As aforesaid, this study was conducted with more repetitive tests compared to the previous studies in which test-retest reliability was assessed in the normal, and moreover, was focused on the agreement of the test score. In result, latency was excellent in the reliability. Amplitude was inferior to the latency in reliability, but had a fair to good reliability. As such, this study is significant. In other words, measurement error in the experimental condition, normal variation in P300 latency and amplitude that is due to subject state, such as fatigue and arousal level, are not sufficient to exceed the normal range of intrarater reliability in the normal adults group.

In 1997, Kim et al. [13] made a comparative analysis according to age in normal adult subjects in their 10s to 60s. According to their report, the latency of ERP was shortest in subjects aged 19, but tended to be longer in proportion to age. The amplitude did not show significant changes in the subjects who are in their 10s, 20s, and 30s, but tended to be decreased rapidly in subjects aged 50 and older. To minimize interference caused by age variance on P300 latency [19], we recruited healthy young subjects in their mid-20s to mid-30s. In this study, average

latency was  $311.3 \pm 37.0$  ms. In comparison with the reference value presented by Kim et al. [13] ( $324.10 \pm 15.55$  and  $327.77 \pm 15.63$  ms in adults in their 20s and 30s), the average latency of subjects in their mid-20s to mid-30s in this study was shorter and the standard deviation was wider. The reason may be explained by several differences, e.g., sample size, frequency, filter setting, and stimulus variance.

There were some limitations of this study. It is not certain that the same reliability can be derived from other studies of similar experimental condition and stimulus design. The reason is that not all the factors, controlling the variability of ERP components, can be identified. Accordingly, there is the need to conduct a further study on factors affecting the reliability, including laboratorial temperature and the difficulty level of task.

## CONFLICT OF INTEREST

No potential conflict of interest relevant to this article was reported.

## REFERENCES

1. Pfefferbaum A, Ford JM, Wenegrat BG, Roth WT, Koppell BS. Clinical application of the P3 component of event-related potentials. I. Normal aging. *Electroencephalogr Clin Neurophysiol* 1984;59:85-103.
2. Goodin DS, Squires KC, Starr A. Long latency event-related components of the auditory evoked potential in dementia. *Brain* 1978;101:635-48.
3. Diner BC, Holcomb PJ, Dykman RA. P300 in major depressive disorder. *Psychiatry Res* 1985;15:175-84.
4. Pfefferbaum A, Wenegrat BG, Ford JM, Roth WT, Koppell BS. Clinical application of the P3 component of event-related potentials. II. Dementia, depression and schizophrenia. *Electroencephalogr Clin Neurophysiol* 1984;59:104-24.
5. Porjesz B, Begleiter H, Samuely I. Cognitive deficits in chronic alcoholics and elderly subjects assessed by evoked brain potentials. *Acta Psychiatr Scand Suppl* 1980;286:15-29.
6. Olbrich HM, Nau HE, Lodemann E, Zerbin D, Schmit-Neuerburg KP. Evoked potential assessment of mental function during recovery from severe head injury. *Surg Neurol* 1986;26:112-8.

7. Olbrich HM, Nau HE, Zerbin D, Lanczos L, Lodemann E, Engelmeier MP, et al. Clinical application of event related potentials in patients with brain tumours and traumatic head injuries. *Acta Neurochir (Wien)* 1986;80:116-22.
8. Papanicolaou AC, Levin HS, Eisenberg HM, Moore BD, Goethe KE, High WM Jr. Evoked potential correlates of posttraumatic amnesia after closed head injury. *Neurosurgery* 1984;14:676-8.
9. Johnson R Jr. A triarchic model of P300 amplitude. *Psychophysiology* 1986;23:367-84.
10. Segalowitz SJ, Barnes KL. The reliability of ERP components in the auditory oddball paradigm. *Psychophysiology* 1993;30:451-9.
11. Hall MH, Schulze K, Rijdsdijk F, Picchioni M, Ettinger U, Bramon E, et al. Heritability and reliability of P300, P50 and duration mismatch negativity. *Behav Genet* 2006;36:845-57.
12. Polich J. Normal variation of P300 from auditory stimuli. *Electroencephalogr Clin Neurophysiol* 1986;65:236-40.
13. Kim HD, Kim GK, Lim YJ, Kim TS, Rhee BA, Leem W. Normal value of cognitive evoked potentials in Koreans. *J Korean Neurosurg Soc* 1997;26:1190-6.
14. Knight RT, Hillyard SA, Woods DL, Neville HJ. The effects of frontal and temporal-parietal lesions on the auditory evoked potential in man. *Electroencephalogr Clin Neurophysiol* 1980;50:112-24.
15. Hodo DW. Kaplan and Sadock's synopsis of psychiatry: behavioral sciences clinical psychiatry. *JAMA* 1996;275:883-4.
16. Lee YH, Park JM, Lee JM, Kim SH, Kim KW, Kang SJ, et al. Correlation between P300 latency and cognitive capacity screening examination in patients with brain lesion. *J Korean Acad Rehabil Med* 2000;24:836-41.
17. Secrest L. Reliability and validity. In: Bellack AS, Hersen M, editors. *Research methods in clinical psychology*. 2nd ed. New York: Pergamon; 1984. p. 24-54.
18. Segalowitz SJ, Ogilvie RD, Simons IA. An ERP state measure of arousal based on behavioral criteria. In: Horne J, editor. *Sleep '90: proceedings of the tenth European Congress on Sleep Research*. Bochum: Pontenagel Press; 1990. p. 23-5.
19. Goodin DS, Squires KC, Henderson BH, Starr A. Age-related variations in evoked potentials to auditory stimuli in normal human subjects. *Electroencephalogr Clin Neurophysiol* 1978;44:447-58.