

P300-CIT 부트스트랩 분석 비교*

송 인 옥 김 혜 민 이 경 은 장 은 희 김 현 택†
고려대학교 심리학과

거짓말 탐지 분야에서 널리 사용되고 있는 P300-CIT(Concealed Information Test)는 주로 목표, 탐침, 무관련자극을 사용하여 P300 사건관련전위 성분을 구하고, 비교하여 거짓/진실 판정에 사용한다. 연구자들은 각 자극의 P300이 유의하게 다른지 판단하기 위한 주된 방법으로 부트스트랩을 활용한 분석들을 사용하고 있다. 하지만 이러한 분석방법들의 분류정확도를 비교한 연구는 매우 부족하다. 선행 연구를 토대로 비교적 널리 사용되고 있는 기법 여섯 가지(상관부트스트랩, 기저선-정점 진폭부트스트랩, 정점-정점 진폭부트스트랩, 수정된 형태의 상관부트스트랩, 수정된 형태의 정점-정점 또는 기저선-정점 진폭부트스트랩)를 선정하고 이를 종합적으로 비교하였다. 거짓집단과 진실집단으로 할당된 총 30명의 참가자들의 P300을 사용하여 여러 부트스트랩 방법들이 참가자들을 얼마나 각 집단으로 맞게 판정하는지 비교하였다. 거짓/진실 판정 기준으로 가장 많이 사용되는 90% 유의수준을 기준으로 하였을 때, 정점-정점 진폭부트스트랩 방법의 탐지정확률이 가장 높게 나타났다. 반면, 상관부트스트랩 방법은 비교적 낮은 정확률을 보였다. 또한 수정된 진폭부트스트랩 방법들은 판단불능을 포함한 판정기준들(90%/30%유의수준, 60%/40%유의수준)을 적용하였을 때, 낮은 수준의 판단불능률 및 높은 수준의 탐지정확률을 보여주었다.

주요어 : 숨김정보검사, 거짓말 탐지, 부트스트랩, P300, 탐지정확률

* 2014학년도 고려대학교 문과대학 특별연구비에 의하여 수행되었음(Supported by a Korea University Grant).

† 교신저자 : 김현택, 고려대학교 심리학과, 서울특별시 성북구 안암동 5가 고려대학교 구법학관 205호

Tel : 02-3290-2065, E-mail : neurolab@korea.ac.kr

거짓말 탐지를 위한 다양한 심리학적 접근이 이루어지고 있는 가운데, 심리생리학 분야에서 현재 가장 널리 사용되고 있는 방법 중 하나는 뇌파(Electroencephalography; EEG)를 사용하는 검사이다. 기존의 폴리그래프가 자율신경계의 변화로 인한 생리적 반응을 측정하는 것과 달리, 뇌파는 중추신경계(뇌)에서 나타나는 변화를 측정한다. 이 때 뇌파 측정과 함께 가장 많이 사용되는 검사 기법은 숨김정보검사(Concealed Information Test; CIT)로, 국내외에서 이 검사를 사용한 많은 연구가 실시되었다. 특히 측정된 뇌파를 평균하여 사건관련전위(event-related potential; ERP) 분석을 실시, 두정중심영역(Pz)에서 가장 크게 나타나는 300ms 이후의 성분을 이용하는 연구가 활발히 진행되었다. 해당 ERP 성분은 P300으로 불리는 성분으로, 다양한 인지적 활동을 반영하는 것으로 알려져 있다(Luck, 2014; Polich, 2007). 대다수의 CIT 연구는 측정된 P300에 기반하여(P300-CIT) 해당 피검자의 거짓 또는 진실 여부를 판별한다.

만약 거짓말이 범죄와 관련이 있는 경우, CIT는 거짓말을 하는 사람이 특정 범죄와 관련된 정보를 가지고 있음을 가정한다. CIT는 용의자가 이 특정 정보를 알아차리는지 아닌지를 바탕으로 거짓말을 판별한다. 일반적으로 CIT는 세 종류의 자극으로 구성되는데, ‘목표(Target)’, ‘탐침(Probe)’, ‘무관련(Irrelevant)’ 자극이 그것이다. 탐침자극은 거짓인 사람 및 피해자, 경찰관 등 소수의 사람만이 알 수 있는 범죄관련 자극이다. 예를 들어, 누군가를 찌른 특정한 칼은 해당 사건 관련자만이 알 수 있는 탐침자극이다. 무관련자극은 탐침자극과 같은 범주에 있을 수 있는 범죄와 관련이 없는 자극을 말한다. 예를 들어, 같은 칼의

범주에 해당하지만 사건과 관련이 없는 다른 칼들이 자극으로 제시될 경우 무관련자극의 역할을 하게 된다. 목표자극은 무관련자극 중에서 특정반응을 하도록 지시받은 자극을 말한다. 예를 들어, 피검사자는 탐침자극과 무관련자극이 나올 때는 왼쪽버튼을 누르지만, 목표자극이 나올 때는 오른쪽버튼을 누르도록 지시받을 수 있다. 목표자극은 피검사자가 과제에 집중을 유지하도록 요구하며, 거짓인 사람이 탐침자극을 볼 때 크게 나타나는 P300 진폭과 비교하기 위한 참조틀(reference)로도 사용된다.

CIT는 보통 적은 개수의 목표자극과 탐침자극, 그리고 많은 개수의 무관련자극을 제시한다. 희소하며 유의미한 자극인 목표자극과 탐침자극은 강한 P300 진폭을 발생시키며, 반면 빈번하게 제시되며 중요하지 않은 무관련자극은 낮은 P300 진폭이 발생된다. P300-CIT의 목적은 제시한 범죄관련자극, 즉 탐침자극으로 인한 P300 진폭을 통해 피검사자가 탐침자극에 관한 정보를 알고 있는지 가려내는 데에 있다.

각 시행을 평균하여 ERP를 구하는데, 이를 통해 개별 시행들의 노이즈와 변인이 감소된다. 그런데 ERP는 개별 시행들의 평균이므로 각 자극 유형(목표, 탐침, 무관련)에 대해서 하나의 파형 및 성분(예, P300)값만 가지게 된다. 이로 인하여, 탐침자극의 P300 진폭값이 무관련자극의 진폭값보다 유의하게 상승하였는지 통계적으로 판단하기 어렵다. 만약 ERP를 구하지 않고 각각의 개별 시행들로 진폭을 구해 통계분석을 실시할 경우, 각 시행들의 노이즈로 인하여 검정력이 낮아지게 된다(Rosenfeld, Ward, Meijer, & Yukhenko, 2017). 이러한 한계를 극복하고 효과적인 거짓/진실 판별을 위해

많은 연구들이 진행되었다. 대표적으로 부트스트랩(Bootstrap) 방법, 베이지안(Bayesian) 방법, 웨이블릿(Wavelet) 방법, Support Vector Machine (SVM) 방법 등이 활용되었다.

다양한 방법 중에서 가장 널리 사용되는 것은 부트스트랩 방법으로, 복원추출(with replacement)을 반복적으로 실시하여 새로운 재표집 샘플을 만들어내는 방법이다. 즉, 원 시행들에서 반복적으로 n 개의 시행을 복원추출하여 새로운 ERP 파형을 만들 수 있다. P300-CIT에서 사용되는 부트스트랩은 특정 최대값들을 복원추출하는 것이 아니라, 각 시행들의 뇌파 데이터들을 복원추출하여 새로운 ERP파형을 만들어 사용한다(Rosenfeld & Donchin, 2015). P300-CIT에서는 진폭부트스트랩(bootstrapped amplitude difference; BAD), 상관부트스트랩(bootstrapped correlation difference; BCD), RT-BOOT(bootstrapped analysis of reaction time) 등이 부트스트랩 방법으로 사용된다.

BAD는 각각 탐침자극과 무관련자극에서 복원추출로 재표집한 ERP파형을 구하고, 탐침자극의 P300 진폭과 무관련자극의 P300 진폭의 차이(P-I)를 구하는 과정을 반복한다. 이 과정을 통하여 P-I의 분포를 만들고 90%의 신뢰수준에서 탐침자극의 P300 진폭값이 무관련자극의 P300 진폭값보다 유의미하게 큰지 판단한다. 즉, 진폭 차이 분포의 90% 신뢰구간의 하한선(-1.29 표준편차)이 0보다 크면 거짓으로 판단하는 기준을 사용하거나(Rosenfeld, Soskins, Bosh, & Ryan, 2004; Rosenfeld et al., 2008), 반복시행 과정에서 만들어진 P-I 차이값들의 90%가 0보다 크지를 판단한다(Meixner & Rosenfeld, 2010). 후자의 방법은 반복시행 과정에서 탐침자극의 진폭이 무관련자극 진폭보다 클 경우 부트스트랩 지수(Bootstrap Index)에 1

을 추가하는 방식으로 계산된다(정은경, 김영운, 2013). 본 연구에서는 부트스트랩 지수로 결과를 표기하였다. 여기서 90%는 임의로 설정된 기준으로, 기존의 연구들은 90% 기준을 많이 사용하였다(Farwell & Donchin, 1991; Mertens & Allen, 2008; Rosenfeld et al., 2004). BAD방법은 P300 진폭을 계산하기 위한 방법에 따라 기저선-정점(base-to-peak; b-p)과 정점-정점(peak-to-peak; p-p)으로 구분된다. 기저선-정점은 자극 제시 전의 기저선과 P300 파형의 최대값의 차이로 P300 진폭을 정의하며, 정점-정점은 P300 파형의 최대값 이후 나타나는 최저값과의 차이로 P300 진폭을 명명한다(Soskins, Rosenfeld, & Niendam, 2001). 기존의 연구들은 BAD방법, 특히 BAD(p-p)방법이 높은 탐지정확률을 보임을 다수 보고하였다(Mertens & Allen, 2008; Rosenfeld et al., 2004; Soskins et al., 2001). 또한 P300 성분이 나타나는 지연시간(latency)이 자극마다 다를 경우에도 효과적으로 기능한다는 장점이 있다(Rosenfeld et al., 2008).

BCD는 목표자극과 탐침자극 그리고 무관련자극을 부트스트랩으로 재표집하고, 목표자극과 탐침자극의 상관($C_{T,P}$)과 탐침자극과 무관련자극의 상관($C_{P,I}$)을 비교한다. 만약 거짓 피검사자라면 탐침자극의 P300 진폭이 상승하여 목표자극-탐침자극의 상관이 탐침자극-무관련자극의 상관보다 클 것이라는 점을 전제한다(Farwell & Donchin, 1991). RT-BOOT는 뇌파가 아닌 반응시간(Reaction Time; RT)을 사용하여 부트스트랩을 실시한다. 거짓말 및 정보를 숨기는 과정은 더 많은 인지과정을 요구하기 때문에 반응시간이 늦어지게 되는데, 이러한 점을 이용하여 탐침자극의 RT가 무관련자극의 RT보다 유의하게 길 경우, 거짓으로 판정되게 된다. BCD법은 진실인 사람을 거짓으로 잘못

판정하는 허위긍정의 오류를 줄일 수 있는 가능성이 높다고 보고되었다(정은경, 김영운, 2013).

Abootaeli, Moradi와 Khalilzadeh(2006)는 이러한 부트스트랩 방법들을 변형하여 사용하였다. 보통 BAD법이 탐침자극과 무관련자극의 진폭만을 사용하는 것과는 달리 목표자극의 진폭을 함께 고려하여 계산하였다. 우선 목표자극과 탐침자극의 진폭 차이를 구한 후($D_1 = T-P$), 목표자극과 탐침자극의 차이를 구하고($D_2 = -I$), D_1 이 D_2 보다 더 클 경우 부트스트랩 지수에 1을 추가하였다. 또한 BCD방법이 자극 제시 전 구간부터 P300 파형이 나타나는 구간 범위에서 상관계수를 구하였던 것과 달리 300ms 부터 900ms 사이의 상관계수를 구하였다. 수정된 방법들은 본 연구에서 기존의 방법들과 비교되었으며, 수정된 BAD 및 짧은 구간을 사용한 BCD(BCD-short)로 지칭하였다. Abootalebi 등(2006)은 해당 논문에서 웨이블릿(Wavelet)을 사용한 방법도 제시하였는데, B-spline wavelet을 사용하여 시간-주파수 분석을 실시해 거짓/진실을 판별하고자 하였다.

연구자들은 부트스트랩 방법 외에 다른 방법들을 적용하여 각 파형이 유의하게 다른지 알아보하고자 하였다. Mertens와 Allen(2008)은 베이지안(Bayesian)법을 제시하였다. ERP파형의 특정한 특징들을 추출하여, 주어진 ERP가 어떠한 패턴을 보이는지에 따라 확률적으로 구분하였다. 즉, ERP파형이 전형적인 형태와 유사한지 여부를 계산하여 거짓/진실을 구분하고자 하였다. Lui와 Rosenfeld(2009)는 부트스트랩이 아닌 순열검정(permutation)을 이용하였는데, 부트스트랩을 사용한 결과와 거의 동일한 결과를 얻었다. 최근에는 SVM을 사용하여 최적의 구분을 시도하였다(Gao, Yan, Sun, &

Zheng, 2011; Zhao, Zheng, & Zhao, 2012). SVM 방법은 90%를 상회하는 높은 분류정확도가 보고되고 있으나, 아직까지 널리 사용되고 있지는 않다.

부트스트랩을 사용한 방법들은 P300-CIT 분야에서 폭넓게 사용되고 있다. 국내에서는 주로 BAD(p-p)법을 사용한 연구가 보고되었다(엄진섭, 박광배, 2014; 엄진섭, 박광배, 손진훈, 2012; 엄진섭, 한유화, 박광배, 손진훈, 2008; 전하정, 손진훈, 박광배, 엄진섭, 2017). 정은경과 김영운(2013)은 BCD와 BAD(b-p)을 비교하였는데, BCD의 탐지정확률이 더 높게 나타났다. 이는 Abootalebi 등(2006)이 보고한 것과 유사한데, BCD, BAD(p-p), 웨이블릿 순으로 탐지정확도가 높았다. 하지만 Abootalebi 등(2006)의 연구는 BCD와 BAD를 변형하여 사용하였으므로 다른 연구와 직접적인 비교를 하기 어렵다. 정은경과 김영운(2013)의 연구에서도 가장 널리 사용되는 BAD(p-p)는 직접적으로 비교되지 않았으며, 탐지정확률이 50% 이하로 나타났다고만 보고되었다.

본 연구는 현재 사용되는 부트스트랩을 사용한 거짓/진실 판별 방법들을 폭넓게 비교하기 위하여 BCD, BAD(b-p), BAD(p-p)를 모두 사용하였다. 뿐만 아니라, Abootalebi 등(2006)이 사용한 수정된 방법을 사용하여 함께 비교하고자 하였다. 또한 거짓으로 판정되는 통계적인 유의수준이 각 논문마다 일관적이지 않게 사용된다는 점을 고려하여, 여러 거짓/진실 판정 기준을 적용하여 비교해 보고자 하였다. 판정 기준은 가장 널리 사용되는 90기준(10회 부트스트래핑 시 부트스트랩 지수가 90을 넘어야 거짓으로 판정)을 사용할 때와 (Abootalebi et al., 2006; Mertens & Allen, 2008; Rosenfeld et al., 2004), Farwell과 Donchin이 BCD

를 사용할 때 적용하던 기준(90/30기준, 100회 부트스트래핑 시 부트스트랩 지수 90이상은 거짓, 30미만은 진실, 90미만 30이상은 판단불능), 정은경과 김영운(2013)이 사용한 60/40기준(100회 부트스트래핑 시 부트스트랩 지수 60이상은 거짓, 40미만은 진실, 60미만 40이상은 판단불능)의 세 가지를 사용하였다.

방 법

참가자

31명의 오른손잡이 참가자(평균나이 = 25.6, 표준편차 = 2.14, 15명 여성)가 모집되었다. 모든 참가자들은 온라인 공고를 통해 모집되었으며 시력, 정신적 병력, 약물복용에 문제가 없었다. 모집된 참가자들은 거짓집단 또는 진실집단에 무선적으로 할당되었다. 참가자들의 과제 몰입감을 높이고 모의 범죄 시나리오가 가지는 낮은 생태학적 타당도를 보완하기 위해 지시 사항을 추가하였다. 즉, 참가자는 자신이 속한 집단에 관계없이 이어지는 검사에서 진실판정을 받기 위해 노력하도록 지시받았다. 만약 참가자가 거짓말 탐지 검사에서 지시를 잘 따르고 진실로 판정될 경우 40,000원을 지급받지만, 지시를 잘 따르지 않거나 거짓으로 판정될 경우는 20,000원으로 보상금이 차감되고 검사 이후에 검사실에 30분 더 머물러야 한다는 설명을 들었다. 실제로는 실험 종료 후 모든 참가자에게 40,000원이 지급되었다. 참가자들은 실험에 관한 설명을 듣고 실험참가 동의서를 작성한 후에 실험에 참여하였다. 실험에 참여한 참가자 중 1명은 EEG 측정에서 과도한 노이즈로 인하여 분석에서

제외되었다. 최종적으로 총 30명의 데이터가 분석에 사용되었다. 본 실험은 고려대학교 IRB의 승인 하에 실시되었다(KU-IRB-15-155-A-1-(E-P-1)).

실험 집단

참가자들은 무선적으로 진실집단 또는 거짓집단에 할당되었다. 각 집단은 각각 15명으로 구성되었으며, 할당된 집단에 따라 ‘진실 임무(mission)’ 또는 ‘거짓 임무’를 수행하였다(모의 범죄). 진실 임무는 실험 진행자에게 지시를 받아, 교수 연구실로 이동하여, 선물을 포장하는 과정으로 구성되었다. 구체적으로, 교수 연구실에 있는 ‘작은 선물상자’를 다른 큰 상자 안에 카드와 함께 넣고 리본으로 포장하는 과정이 임무로 제시되었다. 거짓 임무는 이 포장 과정에서 ‘작은 선물상자’를 열고 안의 내용물(반지 또는 시계)을 훔치는 과정이 추가되었다.

실험 절차

실험 참가자들은 실험의 목적과 절차에 대한 설명을 듣고, 실험참가 동의서에 자필로 서명 후 실험에 참여하였다. 연구자는 이 실험은 참가자의 거짓말 여부를 알아내기 위한 실험이며, 이후 실시될 EEG 검사에서 거짓/진실 여부가 판단될 것이라 언급하였다. 만약 EEG 거짓말 탐지 검사에서 거짓으로 판정될 경우, 앞서 언급된 불이익을 받게 될 것이라는 점도 언급되었다.

이에 동의한 참가자들은 자기보고 설문지에 응답하였다. 자기보고 설문지는 EEG 결과에 영향을 줄 수 있는 변인들을 측정하기 위해

실시되었으며, 집단 간 유의한 차이는 나타나지 않았다(표 1). 자기보고 설문지에서 하위요인이 있는 설문의 경우(예, BIS, PANAS, PPI-R), 하위요인에 대해서도 집단 간 비교가 실시되었으며, 집단 간 유의한 차이는 발견되지 않았다. 주로 사용하는 손을 통일하기 위하여 오른손잡이 검사(handness scale)를 실시하였으며, 보상에 대한 뇌의 반응성과 연관이 있다고 알려진 행동억제 체계 및 행동활성화 체계 척도(BIS/BAS; Carver & White, 1994)를 실시하였다. 실험 참가 당시의 정서에 의한 영향을 배제하기 위하여 정적-부적 정서 척도(PANAS; Watson, Clark, & Tellegen, 1988), Beck 우울검사 (BDI; Beck, 1985), Beck 불안검사(BAI; Beck, 1967)를 실시하였고, 거짓말 행동과 관련이 있는 심리적 요인으로 알려진 자기 검색 척도(Self-monitoring; Snyder & Gangestad, 1986), 마키아벨리즘 성향척도(Mach IV; Christie, Geis, & Berger, 1970), 위험

감수성 척도(Risk-Taking; Knowles, Cutter, Walsh, & Casey, 1973)를 실시하였다. 또한 거짓말을 할 때 생리적 반응이 적다고 알려진 정신병질 성향을 확인하기 위해 정신병질 성격평가(PPI-R; Lilienfeld & Widows, 2005)를 실시하였다.

설문지 응답을 마친 후, 연구자는 실험 참가자에게 두 개의 봉투를 제시하였다. 하나는 진실 임무 내용이, 다른 하나는 거짓 임무가 담겨 있다고 설명하였으나 실제로는 할당된 집단에 해당하는 같은 설명문이 들어있었다. 연구자는 실험 참가자에게 임무의 내용을 설명하고 설명문 하단의 빈칸에 앞으로 수행할 내용을 간단히 적으며 숙지하도록 하였다. 간단한 질문을 통해 숙지가 완료되었다고 판단되면, 참가자는 임무를 수행하도록 교수 연구실 앞으로 안내되었다. 교수 연구실에서 참가자는 자신이 받았던 설명문에 따라 ‘진실 임무’ 또는 ‘거짓 임무’를 수행하였다.

임무를 수행한 후, 참가자는 뇌파 검사실로

표 1. 자기보고 설문지 결과

	거짓집단 평균	진실집단 평균	<i>t</i>	<i>df</i>	<i>p</i>
Handness Scale	28.7 (5.1)	33.4 (13.3)	-1.285	28	.209
BIS	20 (2.3)	19 (1.6)	1.016	28	.318
BAS	11 (2.0)	12 (1.7)	-1.357	28	.186
PANAS-positive	24 (5.5)	24 (7.2)	-.142	28	.888
PANAS-negative	15 (4.0)	14 (2.9)	.733	28	.470
Self-Monitoring	8.2 (2.5)	7.8 (2.1)	.478	28	.636
Mach-IV	58 (4.1)	60 (5.6)	-1.447	28	.159
Risk-Taking	59 (7.0)	63 (5.3)	-1.709	28	.098
BAI	7 (5.2)	6.7 (4.2)	.153	28	.879
BDI	8.2 (4.2)	5.7 (4.6)	1.534	28	.136
PPI-R	58 (11.0)	55 (8.8)	-1.422	28	.160

• 각 집단의 평균값. 우측의 괄호 안은 표준편차(SD).

이동하였으며, 검사실에 있는 검사관(다른 연구자)과 간단한 면담을 진행하였다. 이후 참가자는 뇌파 캡을 착용 후 P300-CIT 검사를 수행하였다. 검사가 완료된 후, 검사관은 참가자에게 연구내용에 대해 설명하였으며 자료이용 동의서를 받았다.

검사자극

본 연구에서는 한국어 문장을 이용한 P300-CIT가 실시되었다. 문장은 각각 단어로 분리되어 화면에 제시되었으며, 한 문장이 끝나면 참가자가 버튼을 누르도록 지시하였다. 문장은 “당신은”-“XXX을”-“흡쳤습니까?” 순서로 구성되었다(그림 1). “당신은”과 “XXX을” 부분은 각각 1000ms 동안 제시되었으며, “흡쳤습니까?”는 500ms 동안 제시되었다. “당신은”과 “XXX을” 사이의 자극 간 간격은 1050, 1150, 1250, 1350ms 중에서 한 간격이 무선적으로 선택되었으며, “XXX을” 및 “흡쳤습니까?” 사이의 간격은 500ms였다. “흡쳤습니까?” 이후 참가자는 최대 1000ms 이내에 반응을 입력할 수 있었으며, 이보다 오래 걸린 반응을 보인 시행은 분석에서 제외되었다. 반응은 키보드

의 ‘예’) 또는 ‘(아니오)’ 버튼을 누르도록 하였으며, 중간의 휴식시간을 기준으로 키를 바꾸어 균형화하였다. 참가자는 해당 자극을 흡치지 않았다는 의미로 ‘아니오’ 버튼을 눌러야 하였으나, 목표자극에 대해서만은 검사자의 지시로 ‘예’ 버튼을 눌렀다.

‘XXX을’ 부분에 실제 분석에 사용될 자극이 제시되었으며, 검사자극은 시계, 반지, 지갑, 목걸이, 향수, 만년필, 벨트가 사용되었다. 해당 자극들은 선행 사전설문을 통하여 글자수가 비슷하며, 연상되는 값어치가 비슷한 수준의 자극이 선정되었다. 이 중 탐침자극은 참가자가 임무에서 흡친 물건(시계 또는 반지)이었으며, 진실집단 참가자의 경우 시계 또는 반지 중 하나를 무선적으로 선택하여 탐침자극으로 분석하였다. 나머지 자극들 중 하나는 목표자극으로 제시되었다. 목표자극은 각 참가자별로 다른 자극을 사용하여 균형화하였다. 목표자극과 탐침자극을 제외한 나머지 5개의 자극들은 무관전자극으로 사용되었다.

목표자극과 탐침자극 그리고 5개의 무관전자극이 무선적인 순서로 제시되었으며, 11번 반복되어 하나의 블록을 구성하였다. 이 때, 목표자극 및 탐침자극은 같은 자극이 연속해

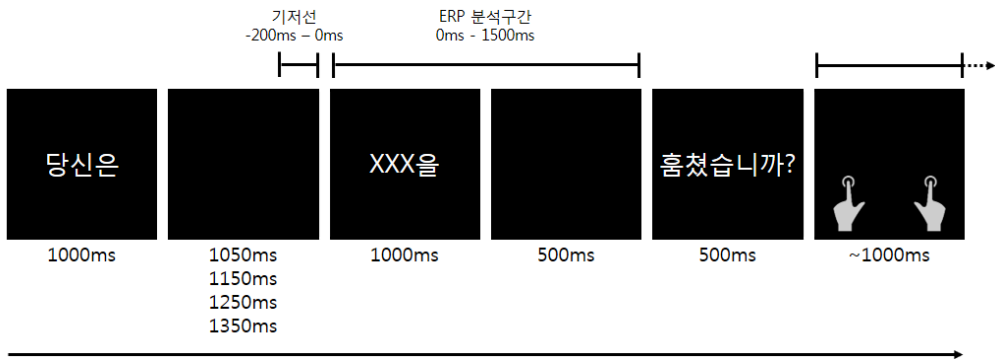


그림 1. 한국어 문장을 사용한 P300-CIT 자극제시절차. ‘XXX을’ 부분에 각 목표, 탐침, 무관전자극이 제시되었다.

서 제시되지 않도록 하였다. 검사는 총 6개의 블록으로 구성되었으며, 각 블록 후에는 30초의 휴식시간을 가지도록 하였다. 또한 세 번째 블록과 네 번째 블록 사이(검사의 절반이 진행된 부분)에서는 5분 동안 휴식하도록 지시하였다. 따라서 검사는 총 462시행으로 구성되었다(1목표+1탐침+5무관련) × 11회 반복 × 6블록).

제시된 자극은 검은색 배경에 밝은 회색 글씨로 제시하였으며, 자극의 시각도(visual angle)는 2.1°였다. 자극들은 모니터 주사율로 인한 지연을 최소화하기 위하여 17인치 CRT 모니터로 제시되었다.

뇌파와 생리신호 측정 및 분석

뇌파는 NeuroScan사의 SynAmps RT 64-channel EEG system을 사용하여 측정되었다(NeuroScan Compumedics, Charlotte, NC, USA). 기준전극은 양 귓볼에 부착되었으며, 평균된 값이 뇌파 측정의 기준으로 사용되었다(linked earlobe). 접지전극은 FPz전극과 Fz전극 사이에 위치하였으며, 왼쪽 눈의 위와 아래 및 양 눈의 좌우 바깥 쪽에 4개의 전극을 부착하여 안전도(Electrooculogram; EOG)를 측정하였다. 각 EOG 전극들은 이후 독립성분분석(Independent Component Analysis; ICA)의 정확도를 높이기 위하여 단극(monopolar)으로 측정되었다(Bigdely-Shamlo et al., 2015). 각 전극은 5k Ω 이하의 임피던스를 유지하였으며 1000Hz의 표집률로 기록되었다.

뇌파 측정과 동시에 심전도(Electrocardiogram; ECG) 측정이 이루어졌다. ECG 측정은 Biopac System(model MP150, Santa Barbara, CA, USA)를 사용하였으며 각 전극은 양 쇄골 밑 부분과

옆구리 부분에 부착되었다. 표집률은 뇌파와 같은 1000Hz로 표집되었으며. ECG 신호는 뇌파에 심박이 미치는 영향을 제거하기 위하여 측정되었다(Park, Correia, Ducorps, & Tallon-Baudry, 2014).

측정된 뇌파는 250Hz로 다운샘플링 되었으며, 0.1~30Hz 대역여과 되었다(Luck, 2014; Mertens & Allen, 2008). ECG 또한 250Hz로 다운샘플링 되었으며, 1~40Hz로 대역여과 되었다(Park et al., 2014). 필터링은 EEGLAB의 FIR 필터를 사용하였다(Delorme & Makeig, 2004). 여과된 뇌파는 ICA 분석을 실시하여 EOG, ECG 신호와 상관성이 높은 ICA 성분을 제거함으로써 보정하였다(Chaumon, Bishop, & Busch, 2015). 보정된 뇌파는 'XXX을'(목표, 탐침, 무관련) 자극 제시 전 200ms에서 자극 제시 후 1500ms까지 분석구간으로 나누어졌다. 나누어진 각 시행의 진폭 크기가 $\pm 75\mu V$ 를 넘을 경우 해당 시행은 분석에서 제외되었다.

남겨진 시행들에서 두정중심영역(Pz)의 신호를 사용하여 P300 성분 분석 및 개인별 부트스트랩 분석이 실시되었다. P300 분석은 모의 범죄의 임무 수행으로 인한 집단 간의 처치 차이가 적절하게 이루어졌는지 확인하기 위하여 실시되었다. P300의 진폭을 자극 유형(목표, 탐침, 무관련) 별로 구하여, 반복측정 변량 분석을 실시하였다. 집단 간 변인은 실험집단(거짓집단, 진실집단)이었으며, 집단 내 변인은 자극 유형이었다. 분자의 자유도가 2인 경우는 Greenhouse-Geisser ϵ 으로 교정된 통계량을 사용하였다. P300의 진폭은 널리 사용되는 정점-정점 방법을 이용하여 구하였다(Soskins et al., 2001). 정점-정점 방법은 P300 파형의 가장 높은 정적값과 그 이후에 나타나는 가장 낮은 부적 전위값의 차이를 구하여 진폭으로 사용

하는 방법이다. 본 연구에서는 자극 제시 후 300ms부터 900ms 사이 100ms 구간 윈도우의 평균값이 가장 높은 값을 찾았으며, 가장 높은 값이 나타난 지점부터 1500ms까지 범위에서 100ms 구간 윈도우 평균값이 가장 낮은 값을 찾아 사용하였다.

개인별 분석

개인별 수준에서 실험 참가자의 거짓/진실 판단을 위하여 부트스트랩 방법을 사용하였다. 부트스트랩 분석은 BAD법과 BCD법을 실시하였으며, 분석을 위해 작성된 MATLAB 코드를 사용하였다. 후술된 부트스트랩 분석은 보다 신뢰로운 결과를 위하여 10번 반복하여 평균 값을 최종 부트스트랩 지수로 사용하였다.

BAD법

BAD법은 크게 기저선-정점을 사용한 방법(BAD(b-p))과 정점-정점을 사용한 방법(BAD(p-p))으로 구분된다. 이 방법들은 관련자극과 무관련자극의 P300 진폭을 측정하여 비교하는 방법이다. 또한 Abootalebi 등(2006)은 이에 대한 변형으로, 탐침자극의 진폭에서 무관련자극의 진폭을 뺀 값($D_1 = P-I$)과 목표자극의 진폭에서 탐침자극의 진폭을 뺀 값($D_2 = T-P$)의 차이를 계산하여(D_1-D_2) 개인별 판단을 실시하였다. 즉, 목표자극의 진폭을 함께 고려한 변형방법을 사용하였다. 본 연구에서는 이 방법을 수정된(modified) 방법으로 지칭하였다.

우선 분석할 피험자의 모든 시행들을 자극 유형별로(목표, 탐침, 무관련) 구분한다. 구분된 시행들은 각 자극 유형별로 N개의 복원추출(with replacement)을 통하여 재표집되었다. 이때 복원추출의 시행횟수(N)는 자극 유형 중

적은 시행횟수를 가지는 자극을 기준으로 하였다(Rosenfeld & Donchin, 2015; Rosenfeld et al., 2017). 재표집된 시행들을 평균하여 새로운 ERP파형을 만들고 나면, 자극 제시 후 300ms부터 900ms까지의 범위에서 100ms 구간 윈도우가 최대치인 값을 찾는다. 즉 251ms부터 950ms 구간 사이에서, 100ms 크기의 윈도우를 이동시켜가며 평균값을 구하고, 그 평균값이 가장 높은 윈도우 구간을 찾는다. 해당 최대값 지점에서의 평균 진폭값과 자극 제시 전 100ms 구간의 평균 진폭을 뺀 값을 구하여 P300 진폭으로 명명한다(b-p). 또한 최대값을 보인 100ms 윈도우의 가운데 지점을 지연시간이라 하면, 지연시간에서 1500ms까지 범위에서 최저값을 보이는 100ms 윈도우 구간을 찾는다(NEG; Soskins et al., 2001). 이전에 계산된 최대값과 최저값의 차이가 P300 진폭으로 명명된다(p-p).

기존의 BAD(b-p)방법과 BAD(p-p)방법은 재표집된 탐침자극과 무관련자극의 P300 진폭을 비교하였다. 탐침자극의 P300 진폭이 무관련자극의 P300 진폭보다 크다면 부트스트랩 지수에 1을 추가한다. 이 과정을 100번 반복하여, 부트스트랩 지수가 90이상이면 거짓으로, 90미만이면 진실로 판단하였다. 수정된 방법 또한 같은 방법으로 계산되었다. 각각 BAD(b-p)방법과 BAD(p-p)방법에 대해 목표자극의 진폭을 함께 계산하여 D_1 값과 D_2 값의 차이를 구하여 부트스트랩 지수의 증가를 결정하였다.

BCD법

BCD법은 전통적으로 널리 사용되었던 Farwell과 Donchin(1991)이 사용한 방법과, P300 ERP 파형이 나타나는 부분에서 상관을 구한 방법(Abootalebi et al., 2006) 두 가지를 구하였다.

우선 각 참가자의 전체 자극유형(목표, 탐침, 무관련) 시행들을 평균하여 ‘전체 평균’을 계산한다. 이 ‘전체 평균’을 부트스트랩 과정에서 재표집된 각 자극유형의 새로운 ERP파형 표본에서 빼서 나온 파형을 비교에 사용한다(double-centering). 부트스트랩은 BAD방법과 동일한 방법으로 복원추출 되었다. ‘전체 평균’을 빼 각각의 목표자극, 탐침자극, 무관련자극을 사용하여 상관계수를 구하였다. 즉, 목표자극과 탐침자극 간의 상관(T-P), 탐침자극과 무관련자극 간의 상관(P-I)을 구하여 T-P상관이 P-I상관보다 높다면 부트스트랩 지수에 1을 추가하였다. 이 과정을 100회 반복하여 부트스트랩 지수가 90이상이면 거짓으로, 30미만이면 진실로 판정하였다. 30이상 90미만의 지수가 나올 경우에는 판단불능으로 간주하였다.

상관계수를 구하는 범위는 두 종류를 사용하였다. 하나는 자극 제시 전 부분부터 P300 파형의 진폭이 기저선 수준으로 줄어드는 범위로, 자극 제시 전 100ms 부터 자극 제시 후 900ms까지를 사용하였다(정은경, 김영운, 2013; Farwell & Donchin, 1991). 본 연구에서는 이 범위를 사용하는 방법을 BCD-long으로 지칭하였다. 다른 방법은 P300 파형이 나타나는 부분부터 가라앉는 범위로, 자극 제시 후 300ms 부터 900ms까지를 사용하였다(Abootalebi et al., 2006). 해당 방법은 BCD-short로 지칭하였다.

ROC곡선 분석

부트스트랩을 사용한 개인별 진단 방법들의 전체적인 분류효과성을 알아보기 위하여 ROC 곡선(Receiver operating characteristic curve) 분석을 실시하였다(Abootalebi et al., 2006; Ben-Shakhar & Elaad, 2003). 진실집단을 특이도

(specificity) 기준 집단으로 설정하였으며, ROC 곡선 종속측정치에는 부트스트랩 분석에서 산출된 부트스트랩 지수가 사용되었다. ROC곡선 방식은 모수적(parametric) ROC를 사용하였다. 각 진단 방법들의 분류효과성은 AUC(Area Under Curve)값으로 제시되었다. AUC 면적값이 1일 경우 완벽한 분류효율성을 의미하며, 0.5일 경우 우연 수준의 분류효율성을 의미한다. 이후 Hanley와 McNeil(1983)의 방법을 사용하여 AUC 면적의 95% 신뢰구간을 구하고 AUC 면적을 서로 비교하였다.

결 과

P300 진폭

각 집단의 Pz에서의 평균 ERP가 그림 2에 제시되어 있다. 예상대로 두 집단 모두 목표 자극은 P300 파형이 높게 올라갔으며, 무관련 자극의 진폭의 크기는 목표자극 보다 작았다. 거짓집단의 탐침자극은 목표자극 수준으로 진폭이 상승하였으며, 진실집단의 탐침자극은 무관련자극과 비슷한 진폭 크기를 보였다(표 2). 반복측정 변량분석 결과(표 3) 자극유형에 의한 주효과가 유의미하였으며($F(2,56)=47.541$, $p<.01$, $\eta^2=.629$), 실험 집단에 의한 주효과도 유의미하였다($F(1,28)=6.816$, $p<.05$, $\eta^2=.196$). 자극유형과 실험집단에 의한 상호작용 또한 유의미하게 나타났다($F(2,56)=8.253$, $p<.05$, $\eta^2=.228$). 즉, 자극의 유형에 따라 각 집단에서 P300의 진폭값이 다르게 나타났다. 사후분석을 위하여 자극유형별로 실험집단 간 독립표본 t 검정을 실시한 결과, 탐침자극에 대한 P300 진폭이 집단 간 유의미하게 달랐음을 확

표 2. 거짓집단과 진실집단의 P300 진폭(μV)의 평균(표준편차)

	목표자극	탐침자극	무관련자극
거짓집단	13.2(6.2)	11.7(5.1)	6.6(2.7)
진실집단	11.9(4.3)	5.2(1.6)	4.8(2.1)

표 3. P300 진폭(μV)의 반복측정 변량분석 결과 및 사후분석 결과

변인	F	df	p	η^2
자극유형	47.541	(2,56)	.000***	.629
실험집단	6.816	(1,28)	.014*	.196
자극유형 × 실험집단	8.253	(2,56)	.002**	.228

자극유형	t	df	p
목표자극	0.626	(28)	.536
탐침자극	4.706	(28)	.000***
무관련자극	1.896	(28)	.068

* $p < .05$, ** $p < .01$, *** $p < .001$

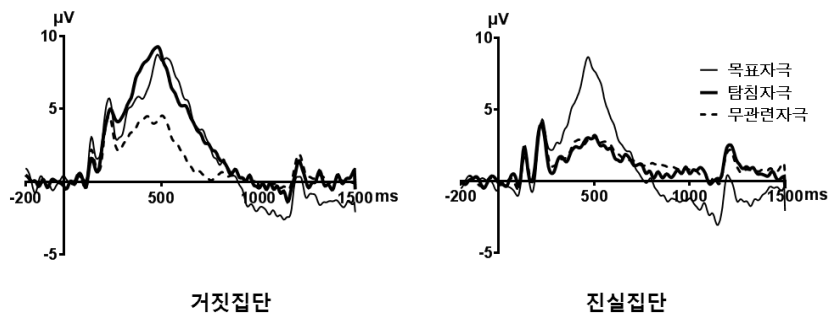


그림 2. 거짓집단과 진실집단의 Grand average ERP

인하였다($t(28)=4.706, p < .01$).

개인별 분석 결과

BAD법

표 4는 BAD법을 사용하여 개인별 부트스트랩 분석을 실시한 결과이다. 90기준을 적용할

때, 기존에 사용되던 BAD(b-p)방법과 BAD(p-p) 방법은 높은 탐지정확률을 보여주었다. BAD(b-p)방법은 거짓집단 15명 중 11명을 거짓으로 판단하였다. BAD(p-p)방법은 거짓집단 15명 중 12명을 거짓으로 판단하였다. 두 방법 모두 진실집단 15명을 모두 진실로 옳게 판정하였다. 즉, BAD(b-p)방법은 86.6%의 탐지정확률

표 4. BAD 개인별 분석 결과

번호	거짓집단		진실집단	
	BAD(b-p)	BAD(p-p)	BAD(b-p)	BAD(p-p)
1	100	100	55.7	42.5
2	49.6	43.5	39.3	29.7
3	99.4	99.6	67.3	89.3
4	71	72.6	34.1	58.4
5	99.9	99.9	67.7	41.8
6	99	98.3	58.2	81.3
7	99.6	100	23.3	30.6
8	95.2	90.9	35	66.5
9	66.7	90.1	53.8	78.4
10	100	100	65.3	38.9
11	96	90.1	24.6	56.2
12	99.1	98.5	10.4	29.8
13	77.2	65.9	69.6	83.1
14	100	100	11.8	9.5
15	99.9	99.7	71.4	73.8
분류결과 (거짓판정인원 / (판단불능) / 진실판정인원)				
90기준	11 / 4	12 / 3	0 / 15	0 / 15
90/30기준	11 / 4 / 0	12 / 3 / 0	0 / 11 / 4	0 / 12 / 3
60/40기준	14 / 1 / 0	14 / 1 / 0	5 / 3 / 7	6 / 4 / 5
집단별 탐지정확률 (판단불능 제외, %)				
90기준	73.3	80	100	100
90/30기준	73.3 (100)	80 (100)	26.7 (100)	20 (100)
60/40기준	93.3 (100)	93.3 (100)	93.3 (100)	33.3 (45.5)
종합 탐지정확률 (판단불능 제외, %)				
90기준		86.6		90
90/30기준	BAD(b-p)	50 (100)	BAD(p-p)	50 (100)
60/40기준		70 (80.7)		63.3 (76)

• 개인별 숫자는 10회 반복하여 평균된 부트스트랩 지수를 의미함.

을 보였으며, BAD(p-p)방법은 90%의 탐지정확률을 보였다.

거짓/진실 판별을 구분하는 기준을 90에서 변경하여, BCD법처럼 30미만을 진실로 판정하고 30이상 90미만을 판단불능으로 판별할 경우(90/30기준), BAD(b-p)방법은 거짓집단에서 4명을 판단불능으로 11명을 거짓으로 판정하였으며 진실집단에서는 11명을 판단불능으로 4명을 진실로 판정하였다. 즉, 거짓집단에서 진실로 잘못 판정된 참가자들은 판단불능의 범위에 해당되어 허위 부정(false negative)을 줄였으나, 진실집단의 다수 참가자가 30이상 90미만의 부트스트랩 지수를 보여 판단불능이 급격히 늘어나게 되었다. 이 때 판단불능을 포함한 탐지정확률은 50%, 판단불능을 제외한 탐지정확률은 100%가 되었다. 판단불능을 포함하여 BAD(p-p)방법을 사용할 경우에는 거짓집단에서 3명을 판단불능으로 12명을 거짓으로 판정하였으며, 진실집단에서는 12명을 판단불능으로 3명을 진실로 판정하였다. 이 때 판단불능을 포함한 탐지정확률은 50%, 판단불능을 제외한 탐지정확률은 100%가 되었다. 즉, BAD(b-p)방법과 BAD(p-p)방법 모두 판별 기준점을 변경하였을 때, 진실집단에서 판단불능의 경우가 많이 나타났으며, 판단불능이 아닐 경우에는 거짓/진실을 정확히 맞추었다.

정은경과 김영운(2013)의 기준을 사용하여 부트스트랩 지수 60이상(상위 40%)을 거짓으로 간주하고 40미만(하위40%)을 진실로 간주하며 40이상 60미만의 범위를 판단불능으로 하였을 경우(60/40기준), BAD(b-p)방법은 거짓집단 15명 중 1명을 판단불능으로 14명을 거짓으로 판정하였으며 진실집단 15명 중 3명을 판단불능, 5명을 거짓, 7명을 진실로 판정하였

다. 판단불능을 포함한 탐지정확률은 70%, 판단불능을 포함하지 않은 탐지정확률은 80.7%였다. BAD(p-p)방법은 거짓집단 15명 중 1명을 판단불능으로, 14명을 거짓으로 판정하였으며 진실집단 15명 중 4명을 판단불능, 6명을 거짓, 5명을 진실로 판정하였다. 판단불능을 포함한 탐지정확률은 63.3%, 판단불능을 포함하지 않은 탐지정확률은 76%였다. 즉, 60/40기준을 사용할 경우 거짓집단에서의 판정은 큰 변화가 없었으나, 진실집단을 거짓으로 판정하는 허위 긍정(false positive)이 증가하였다.

수정된 BAD법

표 5는 수정된(modified) BAD법을 사용하여 부트스트랩 분석을 실시한 결과이다. 90기준을 적용할 때, 목표자극을 함께 고려한 수정된 BAD법은 BAD(b-p) BAD(p-p) 두 방법 모두 기존의 방법에 비하여 낮아진 탐지정확도를 보였다. 수정된 BAD(b-p)방법은 거짓집단 15명 중 7명을 거짓으로 판정하였으나 8명을 진실로 잘못 판정하였다. 진실집단에서는 15명을 모두 진실로 판정하였다. 수정된 BAD(b-p)방법은 거짓집단에서 6명을 거짓으로 판정하였으나 9명을 진실로 잘못 판정하였고, 진실집단에서는 15명을 모두 진실로 옳게 판정하였다. 다시 말해, 수정된 방법으로 거짓/진실 판단을 하였을 경우 거짓집단에서 허위 부정이 크게 증가하였다. 수정된 BAD(b-p)의 탐지정확률은 73.3%, 수정된 BAD(p-p)의 탐지정확률은 70%였다.

90/30기준을 사용하여 판단불능을 고려할 경우, 수정된 BAD(b-p)는 거짓집단 15명 중 7명을 판단불능, 7명을 거짓, 1명을 진실로 판정하였으며, 진실집단에서는 15명 중 11명을 판단불능, 4명을 진실로 판정하였다. 판단불능

표 5. 수정된(modified) BAD 개인별 분석 결과

번호	거짓집단		진실집단	
	수정된 BAD(b-p)	수정된 BAD(p-p)	수정된 BAD(b-p)	수정된 BAD(p-p)
1	97.5	97.1	12.5	0.1
2	77.6	29.4	4.7	0.5
3	99.8	95.4	45.9	56.7
4	50.2	70.9	40.4	28.3
5	99.3	99.5	9.3	0
6	92.6	88.9	40.5	51.6
7	92.6	98.1	2.4	1.4
8	75.9	68.4	27.7	43
9	28.3	35.1	5.4	2.2
10	94.1	99.1	0	0
11	79.5	44.9	0.7	0.2
12	78.7	8.3	0	0
13	52.8	52.2	7	0.8
14	99.4	100	1.2	2.3
15	75.6	66.4	33.5	3.1
분류결과 (거짓판정인원 / (판단불능) / 진실판정인원)				
90기준	7 / 8	6 / 9	0 / 15	0 / 15
90/30기준	7 / 7 / 1	6 / 7 / 2	0 / 4 / 11	0 / 3 / 12
60/40기준	12 / 2 / 1	10 / 2 / 3	0 / 3 / 12	0 / 3 / 12
집단별 탐지정확률 (판단불능 제외, %)				
90기준	46.7	40	100	100
90/30기준	46.7 (87.5)	40 (75)	73.3 (100)	80 (100)
60/40기준	80 (92.3)	66.7 (76.9)	80 (100)	80 (100)
종합 탐지정확률 (판단불능 제외, %)				
90기준		73.3		70
90/30기준	수정된 BAD(b-p)	60 (94.7)	수정된 BAD(p-p)	60 (90)
60/40기준		80 (96)		73.3 (88)

• 개인별 숫자는 10회 반복하여 평균된 부트스트랩 지수를 의미함.

을 포함한 탐지정확률은 60%, 판단불능을 제외한 탐지정확률은 94.7%가 되었다. 수정된 BAD(p-p)는 거짓집단 15명 중 7명을 판단불능, 6명을 거짓, 2명을 진실로 판정하였으며, 진실 집단에서는 15명 중 3명을 판단불능 12명을 진실로 판정하였다. 판단불능을 포함한 탐지정확률은 60%, 판단불능을 제외한 탐지정확률은 90%였다.

60/40기준을 적용했을 경우, 수정된 BAD(b-p)는 거짓집단 15명 중 2명을 판단불능, 12명을 거짓, 1명을 진실로 판정하였으며, 진실 집단에서는 3명을 판단불능, 12명을 진실로 판정하였다. 탐지정확률은 판단불능 포함 시 80%, 판단불능 미포함 시 96%였다. 수정된 BAD(p-p)는 거짓집단 15명 중 2명을 판단불능, 10명을 거짓, 3명을 진실로 판정하였으며, 진실 집단에서는 3명을 판단불능, 12명을 진실로 판정하였다. 판단불능 포함 시 탐지정확률은 73.3%, 판단불능 미포함 시 탐지정확률은 60%였다.

BCD법

표 6은 BCD법을 사용하여 부트스트랩 분석을 실시한 결과이다. 90기준을 적용하였을 때, BCD법은 BAD법 및 수정된 BAD법과 비교하여 더 낮은 탐지정확률을 보였다. BCD-long은 거짓집단 15명 중 오직 2명만을 거짓, 13명을 진실로 판정하여 높은 허위 부정을 보였다. 진실집단에서는 15명을 모두 진실로 판정하였다. 탐지정확률은 56.6%였다. BCD-short도 거짓집단 15명 중 2명을 거짓로, 13명은 진실로 판별하였으며 진실집단은 모두 진실로 판별하여 동일한 탐지정확률을 보였다.

기존의 BCD연구에서 널리 사용되던 90/30기준을 사용하여 판단불능을 고려할 경우, BCD

-long은 거짓집단 15명 중 5명을 판단불능, 2명을 거짓, 8명을 진실로 판정하여 마찬가지로 높은 허위 부정을 보였다. 진실집단에서는 15명을 모두 진실로 판정하였다. 탐지정확률은 56.6%였으며 판단불능을 포함하지 않았을 경우는 68%였다. BCD-short은 BCD-long보다 더 낮은 정확도를 보였다. BCD-short은 거짓집단 15명 중 4명을 판단불능, 2명을 거짓, 9명을 진실로 판정하였으며, 진실집단 15명 중 2명을 판단불능, 13명을 진실로 판정하였다. 두 집단을 모두 고려한 탐지정확률은, 판단불능을 포함 시 50%였으며 판단불능을 제외 시 62.5%였다. 즉, 비교한 모든 방법 중에서 가장 낮은 탐지정확률을 보였다.

60/40기준을 적용했을 경우, BCD-long은 15명의 거짓집단 중 3명을 판단불능, 2명을 거짓, 10명을 진실로 판정하였으며, 15명의 진실 집단은 모두 진실로 판정하였다. 판단불능을 포함한 탐지정확률은 56.6%, 판단불능을 제외한 탐지정확률은 62.9%였다. BCD-short의 경우, 거짓집단 15명 중 1명을 판단불능, 4명을 거짓, 10명을 진실로 판정하였으며, 15명의 진실 집단 중 1명은 판단불능, 14명은 진실로 판정하였다. 판단불능 포함 탐지정확률은 60%, 판단불능 제외 탐지정확률은 64.2%였다. 종합적으로, 어떠한 기준을 사용하더라도 BCD가 BAD보다 낮은 탐지정확률을 보였다.

결과비교

각 부트스트랩 방법과 판정기준(90기준, 90/30기준, 60/40기준)에 따른 판단불능률, 탐지정확률, 허위긍정률, 허위부정률을 표 7에 제시하였다. 판단불능률은 모든 집단에서 판단불능으로 판별된 인원의 비율을 나타내며, 탐지정확률 또한 모든 집단에서 거짓/진실을

표 6. BCD 개인별 분석 결과

번호	거짓집단		진실집단	
	BCD-long	BCD-short	BCD-long	BCD-short
1	95.5	93.5	0	0
2	19.6	57.8	0.3	1.4
3	38.7	60.5	28.7	57.1
4	37.5	26.4	8	7.3
5	29.7	3.2	2.4	0.9
6	20.9	12.9	19.7	21.7
7	92.4	96.8	1.1	0.3
8	45.8	37.7	15.8	37.7
9	1	1.1	0.2	0.3
10	46.7	65.7	0	0
11	3.4	2	2.4	3.1
12	7.5	1	0.4	0.8
13	18.9	5.6	1.6	1.7
14	43.4	25.5	12	7.9
15	7.4	1.9	2.2	1.8
분류결과 (거짓판정인원 / (판단불능) / 진실판정인원)				
90기준	2 / 13	2 / 13	0 / 15	0 / 15
90/30기준	2 / 5 / 8	2 / 4 / 9	0 / 0 / 15	0 / 2 / 13
60/40기준	2 / 3 / 10	4 / 1 / 10	0 / 0 / 15	0 / 1 / 14
집단별 탐지정확률 (판단불능 제외, %)				
90기준	13.3	13.3	100	100
90/30기준	13.3 (20)	13.3 (18.2)	100 (100)	86.7 (100)
60/40기준	13.3 (16.7)	26.7 (28.6)	100 (100)	93.3 (100)
종합 탐지정확률 (판단불능 제외, %)				
90기준		56.7		56.7
90/30기준	BCD-long	56.7 (68)	BCD-short	50 (62.5)
60/40기준		56.7 (63)		60 (64.3)

• 개인별 숫자는 10회 반복하여 평균된 부트스트랩 지수를 의미함.

맞게 판별한 인원의 비율을 나타낸다. 허위긍정률과 허위부정률은 각각 진실집단에서 거짓으로 잘못 판별되거나, 거짓집단에서 진실로 잘못 판별된 인원의 비율을 의미한다.

기존 연구(Abootalebi et al., 2006; Mertens & Allen, 2008; Rosenfeld et al., 2004)에서 널리 사용되던 90기준을 사용할 경우 가장 높은 정확도를 보여준 방법은 BAD(p-p)방법으로, 거짓집단 판별에서 허위부정이 발생(3명)하였으나 가장 높은 탐지정확률을 보였다. BAD(b-p)방법 또한 소수의 허위부정(4명) 판별인원을 제외하면 정확한 판별능력을 보여주었다. 수정된 BAD방법들과 BCD방법들은 BAD(p-p) 및 BAD(b-p)방법보다는 낮은 탐지정확률을 보였다. 수정된 BAD방법들은 허위긍정으로 잘못 판별된 경우는 없었으나 허위부정률이 50%이상으로 나타나 거짓집단을 잘 판별하지 못하는 경향을 보였다. 또한 BCD방법들에서도 허위긍정의 경우는 없었으나 허위부정률이 각각 86.7%로 나타났다.

90/30기준을 적용하여 판단불능을 고려할 경우, BAD(p-p) 및 BAD(b-p)방법은 판단불능을 제외할 경우 100%의 탐지정확률을 보였다. 다만 이 방법들은 각각 50%의 가장 높은 판단불능률을 보였다. 수정된 BAD방법들은 30%대의 판단불능률을 보였으나, 판단불능을 제외할 경우는 수정된 BAD(b-p)가 94.7%, 수정된 BAD(p-p)가 90%의 비교적 높은 탐지정확률을 보였다. BCD방법들은 수정된 BCD방법들보다 낮은 판단불능률을 보였지만, 여전히 높은 허위부정률을 보였다.

60/40기준을 적용하였을 때는 BAD방법들의 판단불능률이 13.3%, 16.7%로 낮아졌지만 진실집단을 오판하는 허위긍정의 경우가 나타났다. 판단불능의 경우를 제외하였을 때, BAD(b-

p)는 41.7%, BAD(p-p)는 54.5%의 허위긍정률을 보였다. 반면 수정된 BAD방법들은 허위긍정은 나타나지 않았으며 90/30기준을 적용하였을 때와 비슷하거나(수정된 BAD(p-p)) 더 낮은(수정된 BAD(b-p)) 허위부정률을 보였다. 또한 90/30기준을 적용하였을 경우보다 더 낮아진 판단불능률(각각 16.7%)을 보여, 비교적 낮은 판단불능률 수준에서 높은 탐지정확률을 보였다. BCD방법들은 매우 낮은 수준의 판단불능률을 보였지만(BCD-long 10%, BCD-short 6.7%), 여전히 높은 허위부정률을 보였다.

종합적으로, 90기준의 비교적 높은 유의수준의 판별기준을 적용할 경우 기존의 BAD방법들의 판별정확률이 높게 나타났으며, 허위긍정의 경우가 나타나지 않았다. BAD(p-p)는 90기준 적용 시 판단불능이 없으면서도 90%의 탐지정확률을 보였다. 하지만 판별기준을 변경할 경우 판단불능률이 크게 상승하거나 허위긍정이 나타나는 문제가 있었다. 수정된 BAD방법들은 90기준에서는 높은 허위부정률을 보였으나, 판단불능을 고려한 기준들을 적용하였을 때는 허위부정률이 감소하여 전반적인 탐지정확률이 높은 수준으로 올라갔다. 수정된 BAD(b-p)방법은 60/40기준을 적용할 때, 16.7%의 판단불능률 하에서 96%의 탐지정확률을 보였다. 그러나 BCD방법들은 어떠한 기준을 적용하였을 경우라도, 다른 방법들보다 우위를 점하지 못하였다. BCD방법들은 53.3%에서부터 86.7%까지의 허위부정률을 보여, 진실집단은 잘 판별하지만, 거짓집단은 잘 판별하지 못하는 경향성이 강하게 나타났다.

ROC곡선

부트스트랩 방법들의 전체적인 분류효과성을 알아보기 위하여 ROC곡선 분석을 실시한

표 7. 부트스트랩 방법들과 판정기준에 따른 판단불능률, 탐지정확률, 허위긍정률, 허위부정률

		판단불능포함				판단불능제외		
		판단 불능률	탐지 정확률	허위 긍정률	허위 부정률	탐지 정확률	허위 긍정률	허위 부정률
BAD(b-p)	90기준	-	86.6	0	26.7	-	-	-
	90/30기준	50	50	0	0	100	0	0
	60/40기준	13.3	70	33.3	0	80.7	41.7	0
BAD(p-p)	90기준	-	90	0	20	-	-	-
	90/30기준	50	50	0	0	100	0	0
	60/40기준	16.7	63.3	40	0	76	54.5	0
수정된 BAD(b-p)	90기준	-	73.3	0	53.3	-	-	-
	90/30기준	36.7	60	0	6.7	94.7	0	12.5
	60/40기준	16.7	80	0	6.7	96	0	7.7
수정된 BAD(p-p)	90기준	-	70	0	60	-	-	-
	90/30기준	33.3	60	0	13.3	90	0	25
	60/40기준	16.7	73.3	0	20	88	0	23.1
BCD-long	90기준	-	56.7	0	86.7	-	-	-
	90/30기준	16.7	56.7	0	53.3	68	0	80
	60/40기준	10	56.7	0	66.7	63	0	83.3
BCD-short	90기준	-	56.7	0	86.7	-	-	-
	90/30기준	20	50	0	60	62.5	0	81.8
	60/40기준	6.7	60	0	66.7	64.3		71.4

• 각 숫자는 비율(%)을 의미함.

결과, 각각의 부트스트랩 방법들은 전체적으로 높은 분류효과성을 보여주었다. 표 8에 ROC곡선의 AUC 면적과 95%신뢰구간 및 비교 결과를 나타내었다. 그림 3에는 각 부트스트랩 방법들의 ROC곡선이 제시되어 있다. 개인별 진단에서 특정 기준(90기준, 90/30기준, 60/40기준)에서는 기존의 BAD방법들이 좋은 결과를 보였으나, 전체적인 분류효과성은 수정된 BAD방법들이 더 높게 나타났다. 그러나

그 차이가 유의미한 수준은 아니었다($p=n.s.$). BCD-long과 BCD-short은 ROC곡선 분석에서도 가장 낮은 분류효과성을 보였다. 특히 BCD-short는 AUC의 신뢰구간이 0.5(우연수준)에 근접하였다. 각 부트스트랩 방법들의 AUC 면적을 통계적으로 비교한 결과, 수정된 BAD(b-p)방법이 BCD-long($z=1.970, p<.05$), BCD-short($z=2.588, p<.01$)방법보다 유의미하게 AUC 면적이 크게 나타났으며, 수정된 BAD(p-p) 방법이

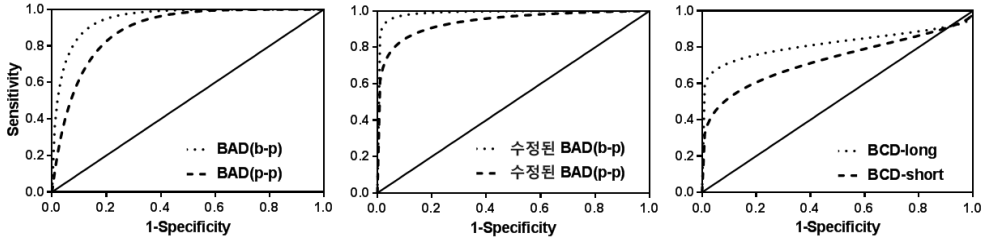


그림 3. 부트스트랩 방법들의 ROC곡선. 각 방법들이 진실집단과 거짓집단을 얼마나 효과적으로 분류할 수 있는지를 나타낸다.

표 8. 부트스트랩 방법들의 AUC 면적과 95% 신뢰구간 및 비교결과

부트스트랩 방법	BAD(b-p)	BAD(p-p)	수정된 BAD(b-p)	수정된 BAD(p-p)	BCD-long	BCD-short
AUC	0.950	0.889	0.991	0.941	0.819	0.730
95% 신뢰구간	0.868~ 0.991		0.949~	0.852~	0.739~	0.532~
		0.976	0.999	0.999	0.957	0.870
비교결과	BAD(b-p)	BAD(p-p)	수정된 BAD(b-p)	수정된 BAD(p-p)	BCD-long	BCD-short
BAD(b-p)	-	0.705 (.481)	1.074 (.283)	0.000 (1.000)	1.078 (.281)	1.782 (.075)
BAD(p-p)		-	1.541 (.123)	0.612 (.541)	0.763 (.445)	1.603 (.109)
수정된 BAD(b-p)			-	1.283 (.199)	1.970 (.049*)	2.588 (.009**)
수정된 BAD(p-p)				-	1.531 (.126)	2.350 (.019*)
BCD-long					-	2.085 (.037*)
BCD-short						-

• 비교결과의 값은 z-statistics이며 괄호 안은 유의수준(Hanley & McNeil, 1982)

* p<.05, ** p<.01

BCD-short방법보다 유의하게 크게 나타났다 (z= 2.350, p<.05). 또한 BCD-long방법이 BCD-short보다 유의하게 크게 나타났다(z=2.085, p<.05).

논 의

본 연구는 한국어를 사용한 P300-CIT를 실시하여, CIT검사에서 대표적으로 사용되는 개

인 거짓/진실 판별 방법인 부트스트랩 방법들을 수행하였다. 본 연구에서 수행된 부트스트랩 방법들은 널리 사용되는 방법(BAD(b-p), BAD(p-p), BCD-long) 외에도 수정된 방법들(수정된 BAD(b-p), 수정된 BAD(p-p), BCD-short)을 포함하였다. 또한 선행 연구들에서 다르게 사용되었던 거짓/진실 판단 기준점들을 적용한 결과를 제시하였다. 실제 측정된 뇌파 데이터를 사용하여 각 부트스트랩 방법의 분류정확도를 도출하여 비교한 결과, 기존 연구에서 널리 사용되던 90기준에서 BAD(b-p)와 BAD(p-p)가 높은 탐지정확률을 보였으며, 90/30기준에서는 기존 BAD방법들 및 수정된 BAD 방법들이 높은 탐지정확률을 보였다. 그러나 이 경우 판단불능률이 높게 나타났다. 60/40기준에서는 기존의 BAD방법들에서 허위긍정이 나타났다으며, 수정된 BAD방법들이 낮은 수준의 판단불능률 하에서 높은 탐지정확률을 보였다. BCD방법들은 모든 판별기준에서 기존 BAD방법 및 수정된 BAD방법보다 낮은 정확도를 보였다.

기존 BAD방법들의 탐지정확률이 BCD방법들의 탐지정확률보다 높게 나타난 것은 Rosenfeld 등(2004)의 결과와 일치하는 것으로, 거짓/진실 판단 기준으로 90기준을 사용하였을 때 BAD(p-p)의 탐지정확률이 가장 높다고 보고한 연구들과도 일치한다(Mertens & Allen, 2008; Rosenfeld et al., 2004; Soskins et al., 2001). 반면에 60/40기준처럼 ‘완화된’ 판별기준점을 사용할 경우, 기존 BAD 방법들은 높은 허위긍정률을 나타냈다. 60/40기준을 사용한 정은경과 김영운(2013)의 연구에서도 BAD(b-p)방법은 12명의 진실집단을 2명의 판정불능과 10명의 거짓으로 판정하여 높은 허위긍정률을 보였는데, 본 연구에서도 허위긍정률이 높아지

는 비슷한 경향이 나타났다. 본 연구에서는 BAD(b-p)방법은 진실집단 15명 중 5명, BAD(p-p)방법은 15명 중 6명을 거짓으로 판정하여 허위긍정률의 수치는 상대적으로 낮게 나타났다. 그러나 형사사법적 맥락에서 허위긍정의 오류가 거짓오판으로 이어질 수 있다는 위험성을 고려해 볼 때, 기존 BAD방법을 사용할 때는 ‘엄격한’ 판별 기준을 사용하는 것이 적합하다고 생각된다.

수정된 BAD방법을 제안한 Abootalebi 등(2006)의 연구에서는 탐지정확률이 가장 높게 나타나는 최적점을 판별기준점으로 선택하여 결과를 보고하였기 때문에 기존에 사용되던 판별기준점에 따른 탐지정확률과 직접적인 비교를 하기는 어려우나, 약 40기준으로 판별이 이루어졌다는 점을 볼 때, 60/40기준과 비교가 가능하다고 생각된다. Abootalebi 등(2006)은 수정된 BAD(p-p)방법을 사용하여 74%의 탐지정확률을 보고하였는데, 본 연구에서는 판단불능을 포함할 때 73.3%의 탐지정확률을 보여서 비슷한 수준으로 나타났다. 반면에 Abootalebi 등이 제안한 BCD-short의 경우 최대 80%의 탐지정확률을 나타냈지만, 본 연구에서는 60/40기준에서 60%(판단불능포함), 64.3%(판단불능 제외)의 탐지정확률을 보여 차이가 나타났다. BCD-long을 사용한 연구들은 다소 상이한 결과를 보고하고 있는데, Rosenfeld 등(2004)은 90기준에서 77% 탐지정확률을 보고하였으나 Mertens와 Allen(2008)은 35.5%(95기준), 정은경과 김영운(2013)은 65.2%(60/40기준)를 보고하였다. 본 연구에서 BCD-long은 56.7%에서 68%까지의 탐지정확률을 보여서 정은경과 김영운의 연구에 가장 가깝게 나타났다.

다양한 판단기준점을 적용한 판별정확도는 ROC곡선 분석으로도 재확인되었다. ROC곡선

분석으로 전체적인 분류효과성을 검증한 결과, 유의한 수준의 차이는 아니나 수정된 BAD(b-p)법의 AUC 면적이 가장 높게 나타났으며, BAD(b-p)와 수정된 BAD(p-p)방법이 그 뒤를 이었다. 반면, BCD-long과 BCD-short 방법은 BAD방법들에 비해 낮은 AUC 면적을 보였으며, 특히 BCD-short 방법은 AUC 면적의 95% 신뢰구간이 우연수준(0.5)에 근접하였다. Abootalebi 등(2006)이 제안한 수정된 BAD방법들은 널리 사용되고 있지 않으나, 적절한 판별기준점을 찾는 후속연구가 병행될 필요가 있다고 보인다.

본 연구는 P300-CIT에서 사용되는 다양한 부트스트랩 방법들과 함께 다양한 판별기준점을 적용하여 비교한 연구라는 점에서 의의를 지닌다. 지금까지 부트스트랩 분류법을 비교한 연구는 다수 있었으나, 서로 다른 판별기준점을 사용하여 부트스트랩 분류법들의 결과를 직접적으로 비교하기 어려웠다. 분류법들을 비교한 상당수의 연구는 90기준 또는 95기준을 판별기준으로 사용하였는데(Farwell & Donchin, 1991; Meixner & Rosenfeld, 2010; Mertens & Allen, 2008; Rosenfeld et al., 2004; Soskins et al., 2001), 일부의 연구는 탐지정확률이 가장 높은 최적점에 기반하여 보고하거나(Abootalebi et al., 2006), 60/40기준과 같은 완화된 기준을 사용하였기 때문이다(정은경, 김영운, 2013). 또한 보고된 연구들은 판별기준을 하나만 적용하였기 때문에, 기준을 변화시킬 경우 부트스트랩 방법의 결과 변화를 예측하기 어려웠다. 예를 들어, 기존의 BAD(b-p) 또는 BAD(p-p)방법에 90/30기준을 적용할 경우 판단불능률이 크게 상승하거나 60/40기준처럼 완화된 기준을 사용할 경우 허위긍정률이 높아진다는 점은 기존의 연구들에서 명확히 나

타나지 않았다.

한편 본 연구는 한국어를 사용한 P300-CIT 검사에 다양한 부트스트랩 방법들을 적용하여 비교한 연구라는 점에서도 의의가 있다. 정은경과 김영운(2013)의 연구에서는 현재 가장 널리 사용되는 BAD(p-p)법이 자세히 보고되지 않았다. 뿐만 아니라, Abootalebi 등(2006)이 제안한 수정된 방법을 사용한 연구는 아직까지 보고된 바 없다. 국내에서 실시되는 P300-CIT 검사는 한국어로 실시된다는 점에서, 해외의 연구들에 대한 타당도와 신뢰도 검증이 필요한 바, 본 연구는 한국어 P300-CIT에 적합한 분류법 탐색에 기여할 수 있다고 생각된다. 하지만 본 연구가 통제하지 못한 한계점들이 있으며, 다음 한계점들에 대해 후속연구가 필요하다고 생각된다.

첫째로, 본 연구는 P300-CIT에서 개인별 거짓/진실 판단을 위해 가장 널리 사용되는 부트스트랩 방법들 및 그 수정된 방법들을 포함하여 비교하였다. 그러나 최근 제안되고 있는 PCA 분석법 또는 SVM 방법을 포함한 보다 다양한 비교 분석이 필요하다고 생각된다. 특히 SVM 방법은 90%이상의 높은 분류정확도를 보고하고 있다는 점에서 부트스트랩 방법과 함께 비교, 검증해야 할 필요성이 있다(Gao et al., 2011; Zhao et al., 2012). 둘째로, 본 연구는 FIR필터와 0.1Hz~30Hz의 대역통과 필터를 사용하여 뇌파 신호를 처리하였다. 최근에 Rosenfeld를 비롯한 다른 연구자들이 BAD(p-p) 방법을 사용할 때, 0.3Hz~30Hz 또는 0.1Hz~30Hz 필터링을 추천하고 있는 점으로 미루어 볼 때(Verschuere, Ben-Shakhar, & Meijer, 2011), 본 연구에서 사용된 대역통과 필터 범위는 적절한 수준이라 생각된다. 하지만 실제 부트스트랩 분석이 P300 진폭 및 ERP 파형의 상관계

수를 구한다는 점을 생각해 볼 때, 뇌파의 전처리 과정이 차후 부트스트랩 결과에 미치는 영향을 분석한 연구가 필요하다. 기존의 연구들은 측정된 뇌파 신호를 보다 낮은 주파수로 저역통과 시켜 분석에 사용하였으며(Rosenfeld et al., 2004; Soskins et al., 2001), 필터의 종류에 따라 ERP 파형에 영향이 있었을 수 있다(Luck, 2014).

본 연구는 위와 같은 한계점을 지니고 있으며, 또한 각 분류법의 탐지정확률의 우위를 가리기에는 표본이 충분치 않았다. 그러나 한국어로 시행된 P300-CIT를 바탕으로 부트스트랩 분류법들을 비교한 국내 연구가 아직 부족하다는 점에서 본 연구에 의의를 둘 수 있다. 요약하면, 본 연구에서는 널리 사용되는 부트스트랩 방법 및 그 변형들을 함께 비교하였으며, 결과적으로 기존의 엄격한 90기준을 사용할 때, BAD(b-p)방법과 BAD(p-p)방법이 높은 탐지정확률을 보임을 확인하였다. 기존의 연구와 달리, BCD방법들은 종합적으로 가장 낮은 탐지정확률을 보였다. 또한 90/30기준 또는 60/40기준을 적용할 때 수정된 BAD방법들이 효과적인 판별 방법임을 보였으며, ROC곡선 분석에서 재확인되었다. 후속연구에서는 최적의 판별기준점을 탐색하는 연구와 더불어 부트스트랩 방법 외의 판별 방법을 포함한 비교연구가 필요할 것으로 보인다.

참고문헌

강기영, 김영운 (2010). 문장을 이용한 P300-기반 거짓지식검사. *한국심리학회지: 사회 및 성격*, 24(4), 19-41.
엄진섭, 박광배 (2014). P300 숨김정보검사에서의

시행수가 검사결과에 미치는 영향. *한국심리학회지: 사회 및 성격*, 28(3), 61-79.
엄진섭, 박광배, 손진훈 (2012). ERP 와 동공 반응을 이용한 숨김정보검사. *감성과학*, 15(2), 259-268.
엄진섭, 한유화, 박광배, 손진훈 (2008). P300 숨김정보검사의 민감도: 목표자극을 이용하는 측정 패러다임과 두 자극에 대한 범주판단 및 재인검사를 이용하는 측정 패러다임의 비교. *한국심리학회지: 일반*, 27(3), 653-673.
전하정, 손진훈, 박광배, 엄진섭 (2017). 반응양식이 P300 숨김정보검사의 정확도에 미치는 영향. *감성과학*, 20(3), 109-118.
정은경, 김영운 (2013). P300-기반 거짓지식검사의 탐지정확률 비교. *한국심리학회지: 인지 및 생물*, 25(1), 75-92.
Abootalebi, V., Moradi, M. H., & Khalilzadeh, M. A. (2006). A comparison of methods for ERP assessment in a P300-based GKT. *International Journal of Psychophysiology*, 62(2), 309-320.
Beck, A. T. (1967). *Depression: Clinical, experimental, and theoretical aspects (Vol.32)*: University of Pennsylvania Press
Beck, A. T. (1985). *Anxiety disorders and phobias: A cognitive perspective*. New York: Basic Books.
Ben-Shakhar, G., & Elaad, E. (2003). The validity of psychophysiological detection of information with the Guilty Knowledge Test: A meta-analytic review. *Journal of Applied Psychology*, 88(1), 131.
Bigdely-Shamlo, N., Mullen, T., Kothe, C., Su, K. M., & Robbins, K. A. (2015). The PREP pipeline: standardized preprocessing for large-

- scale EEG analysis. *Frontiers in neuroinformatics*, 9, 16.
- Carver, C. S., & White, T. L. (1994). Behavioral inhibition, behavioral activation, and affective responses to impending reward and punishment: the BIS/BAS scales. *Journal of personality and social psychology*, 67(2), 319.
- Chaumon, M., Bishop, D. V., & Busch, N. A. (2015). A practical guide to the selection of independent components of the electroencephalogram for artifact correction. *Journal of neuroscience methods*, 250, 47-63.
- Christie, R., Geis, F. L., & Berger, D. (1970). *Studies in machiavellianism*: Academic Press NewYork
- Delorme, A., & Makeig, S. (2004). EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of neuroscience methods*, 134(1), 9-21.
- Farwell, L. A., & Donchin, E. (1991). The truth will out: Interrogative polygraphy ("Lie detection") with Event Related brain potentials. *Psychophysiology*, 28(5), 531-547.
- Gao, J., Yan, X., Sun, J., & Zheng, C. (2011). Denoised P300 and machine learning-based concealed information test method. *Computer methods and programs in biomedicine*, 104(3), 410-417.
- Knowles, E. S., Cutter, H. S., Walsh, D. H., & Casey, N. A. (1973). Risk-taking as a personality trait. *Social Behavior and Personality: an international journal*, 1(2), 123-136.
- Lilienfeld, S. O., & Widows, M. R. (2005). Psychological assessment inventory revised (PPI-R). Lutz, FL: Psychological Assessment Resources.
- Luck, S. J. (2014). *An introduction to the event-related potential technique*. MIT press.
- Lui, M., & Rosenfeld, J. P. (2009). The application of subliminal priming in lie detection: Scenario for identification of members of a terrorist ring. *Psychophysiology*, 46(4), 889-903.
- Meixner, J. B., & Rosenfeld, J. P. (2010). Countermeasure mechanisms in a P300 based concealed information test. *Psychophysiology*, 47(1), 57-65.
- Mertens, R., & Allen, J. J. (2008). The role of psychophysiology in forensic assessments: Deception detection, ERPs, and virtual reality mock crime scenarios. *Psychophysiology*, 45(2), 286-298.
- Park, H. D., Correia, S., Ducorps, A., & Tallon-Baudry, C. (2014). Spontaneous fluctuations in neural responses to heartbeats predict visual detection. *Nature neuroscience*, 17(4), 612.
- Polich, J. (2007). Updating P300: an integrative theory of P3a and P3b. *Clinical neurophysiology*, 118(10), 2128-2148.
- Rosenfeld, J. P., & Donchin, E. (2015). Resampling (bootstrapping) the mean: A definite do. *Psychophysiology*, 52(7), 969-972.
- Rosenfeld, J. P., Labkovsky, E., Winograd, M., Lui, M. A., Vandenboom, C., & Chedid, E. (2008). The Complex Trial Protocol (CTP): A new, countermeasure resistant, accurate, P300 based method for detection of concealed information. *Psychophysiology*, 45(6), 906-919.
- Rosenfeld, J. P., Soskins, M., Bosh, G., & Ryan,

- A. (2004). Simple, effective countermeasures to P300 based tests of detection of concealed information. *Psychophysiology*, 41(2), 205-219.
- Rosenfeld, J. P., Ward, A., Meijer, E. H., & Yukhnenko, D. (2017). Bootstrapping the P300 in diagnostic psychophysiology: How many iterations are needed?. *Psychophysiology*, 54(3), 366-373.
- Snyder, M., & Gangestad, S. (1986). On the nature of self-monitoring: Matters of assessment, matters of validity. *Journal of personality and social psychology*, 51(1), 125.
- Soskins, M., Rosenfeld, J. P., & Niendam, T. (2001). Peak-to-peak measurement of P300 recorded at 0.3 Hz high pass filter settings in intraindividual diagnosis: complex vs. simple paradigms. *International Journal of Psychophysiology*, 40(2), 173-180.
- Verschuere, B., Ben-Shakhar, G., & Meijer, E. (Eds.). (2011). *Memory detection: Theory and application of the Concealed Information Test*. Cambridge University Press.
- Watson, D., Clark, L. A., & Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: the PANAS scales. *Journal of personality and social psychology*, 54(6), 1063.
- Zhao, M., Zheng, C., & Zhao, C. (2012). A new approach for concealed information identification based on ERP assessment. *Journal of medical systems*, 36(4), 2401-2409.

1 차원고접수 : 2018. 06. 15.

심사통과접수 : 2018. 07. 12.

최종원고접수 : 2018. 07. 16.

The comparison of Bootstrapping analyses in P300-CIT

Inuk Song Hyemin Kim Kyoung Eun Lee
Eunhee Chang Hyun Taek Kim

Department of Psychology, Korea University

The Concealed Information Test based on P300 (P300-CIT), which has been widely used in detecting deception, uses mainly target, probe, and irrelevant stimuli to obtain P300 event-related potentials (ERPs) and compares them to judge guilty or innocent. Researchers have used the applications of bootstrapping as the useful methods for determining significant difference between P300 ERP components of the stimuli. However, relatively few research comparing the classification accuracy of the bootstrapping analyses have been reported. On the basis of previous research, six widely used bootstrapping methods (bootstrapped correlation difference (BCD), base-to-peak bootstrapped amplitude difference (BAD) peak-to-peak BAD, modified BCD, modified base-to-peak BAD, modified peak-to-peak BAD) were selected and compared. Thirty participants were randomly assigned to the guilty and innocent groups. The P300 components were acquired to compare the classification accuracies by different bootstrapping analyses. Using 90% confidence level, which has been most widely used, as a guilty criterion of bootstrapping index, the peak-to-peak BAD showed the highest accuracy. On the other hand, the BCD methods showed relatively lower accuracy. Furthermore, the modified BAD methods applied to the other criteria including indeterminate category(90%/30%, 60%/40% confidence level) showed high level of accuracy with reduced rate in indeterminate category.

Key words : *Concealed Information Test, Lie detection, Bootstrap, P300, Classification Accuracy*