

The Accuracy of Permutation Test in P300-based Concealed Information Test*

Hajung Jeon¹, Jin-Hun Sohn¹, Jin-Sup Eom^{2†}

¹Department of Psychology, Chungnam National University

²Department of Psychology, Chungbuk National University

The purpose of this study is to show the P300 amplitude of probe stimulus is overestimated more than that of irrelevant stimulus because of the difference of the number of trials between two stimuli in the P300-based concealed information test, and to demonstrate that the type 1 error rate of bootstrap method is higher than the significance level, but that of permutation test is the same as the significance level. The type 1 error rate and the statistical power of bootstrap method and permutation test were estimated by using Monte Carlo study in the case of using 30 Hz, 10 Hz, or 5 Hz low-pass filter, and using P300 peak amplitude or P300 mean amplitude of 100 ms interval. As a result, despite using low-pass filter or P300 mean amplitude of 100 ms interval, the extent of overestimation for the P300 amplitude of probe stimulus was greater at least about $1 \mu V$ than that of irrelevant stimulus. For this reason the type 1 error rate of bootstrap method was higher than the significance level. However, the type 1 error rate of permutation test was the same as the significance level. The power of permutation test was highest when 5 Hz low-pass filter and P300 means amplitude of 100 ms interval were used. The results of this study can be used in all cases where the P300 amplitudes of two stimuli with different number of trials are compared.

Keywords: P300, concealed information test, permutation test, type 1 error, power

1 차원고접수 19.09.18; 수정본접수: 19.12.09; 최종게재결정 19.12.10

범죄사건 현장에 범죄와 직접 관련된 정보(예, 범행에 사용된 칼)가 남아 있는 경우에 이 정보를 이용하여 용의자가 범인인지 아닌지 판단하기 위한 P300 숨긴정보검사(P300-based concealed information test: P300 CIT)를 실시할 수 있다. 일반적인 P300 CIT에서는 조사대상자의 뇌파(electroencephalography: EEG)를 기록하면서 범행에 사용된 칼(관련자극; probe)과 범행에 사용되지 않은 여러 개의 칼(무관련자극; irrelevant), 범행에 사용되지 않은 특정한 칼(목표자극; target)을 섞어서 무선으로 하나씩 제시한다. 보통 관련자극과 무관련자극, 목표자극을 1:4:1의 비율로 구성

하며, 신뢰로운 P300 진폭을 측정하기 위하여 관련자극이 30회 이상 제시되도록 전체 자극들을 반복 제시한다. 즉, 관련자극을 30회, 무관련자극을 120회, 목표자극을 30회 제시한다. 조사대상자에게는 목표자극이 제시되면 '예' 반응을 하고, 목표자극 이외의 자극이 제시되면 '아니요' 반응을 하도록 지시한다. 자극제시가 모두 끝나면 관련자극과 무관련자극 각각에 대한 사건관련전위(event-related potential: ERP)를 산출하여 P300 진폭을 측정한다.

조사대상자가 범인이라면 두 가지 이유로 관련자극에 대한 P300 진폭이 무관련자극에 대한 P300 진폭보다 더 크게

* 이 논문은 2017년 대한민국 교육부와 한국연구재단의 지원을 받아 수행된 연구임(NRF-2017S1A5A8021834).

† 교신저자: 엄진섭, 충북대학교 심리학과, (28644) 충북 청주시 서원구 충대로 1

E-mail: jseom2003@hanmail.net

발생하게 된다. 한 가지는 개인에게 가치가 있는 자극에서 P300 진폭이 더 크게 발생하기 때문이다(Johnson, 1993). 범인에게 관련자극은 무관련자극보다 더 가치가 있는 자극이지만 무고한 조사대상자에게 관련자극과 무관련자극은 동일한 가치를 가지게 된다. 다른 한 가지는 다수의 자극들 중 제시빈도가 더 작은 자극범주에서 P300 진폭이 더 크게 발생하기 때문이다(Johnson, 1993). 무고한 조사대상자에게 관련자극과 무관련자극은 서로 구별되지 않는 동일한 범주인 반면, 범인에게 관련자극과 무관련자극은 서로 구별되는 범주가 되며 관련자극 범주의 제시빈도가 무관련자극 범주의 제시빈도보다 더 적게 된다(Farwell & Donchin, 1991). P300 CIT에서 관련자극과 무관련자극의 개수를 1:4로 구성하는 이유가 여기에 있다.

그러나 P300 CIT에서는 두 가지 해결해야 할 문제가 있다. 한 가지는 시행수가 서로 다른 두 자극의 P300 진폭을 비교하는 방법에 관한 것이며, 다른 한 가지는 관련자극의 P300 진폭이 무관련자극의 P300 진폭보다 얼마나 더 커야 조사대상자를 범인이라고 판단할 수 있는가에 관한 것이다.

한 시행에서 측정된 뇌파는 자극과 관련된 뇌파와 자극과 관련이 없는 뇌파(즉, 배경뇌파 또는 노이즈)가 합해진 것이다. 자극과 관련된 뇌파는 배경뇌파보다 진폭이 작기 때문에 한 번의 시행에서 자극과 관련된 뇌파를 자세히 확인하기는 어려우며 주로 평균화 기법을 통해서 확인한다(Luck, 2014). 평균화 기법은 동일한 자극을 여러 번 반복 제시한 후 자극 제시 시점을 기준으로 단일 시행 뇌파들(single sweeps), 즉 각 자극을 제시하였을 때의 뇌파를 평균하여 ERP를 산출하는 것이다. 배경뇌파는 자극과 관련된 뇌파와 독립적이며 평균이 0이므로 여러 번 측정된 뇌파들을 평균하면 ERP에 포함된 배경뇌파의 양은 감소한다. 이론적으로 무한대의 시행수를 거치면 ERP에서 배경뇌파를 완전히 제거할 수 있다. 이 말은 시행수가 유한한 경우 ERP에는 항상 배경뇌파가 어느 정도 포함되며, 시행수가 많아질수록 ERP에 포함된 배경뇌파의 양도 더 많이 감소한다는 의미이다.

ERP에 포함된 고주파 성분의 배경뇌파는 P300 정점을 더 크게 만든다. 그림 1에서 볼 수 있는 바와 같이, 순수한 P300 파형에 고주파의 배경뇌파가 추가되면 P300의 정점이 더 크게 측정된다. 뿐만 아니라 그림 1의 (A)처럼 시행수가 작으면 그림 1의 (B)와 같이 시행수가 더 큰 경우에 비하여 고주파 배경뇌파의 양이 더 많고, 결과적으로 P300의 정점이 더 크게 측정된다(Luck, 2014). P300 CIT에서 관련자극의 시행수가 무관련자극의 시행수보다 4배 더 작으므로 관련자극의 ERP에 더 많은 고주파 배경뇌파가 포함된다. 따라

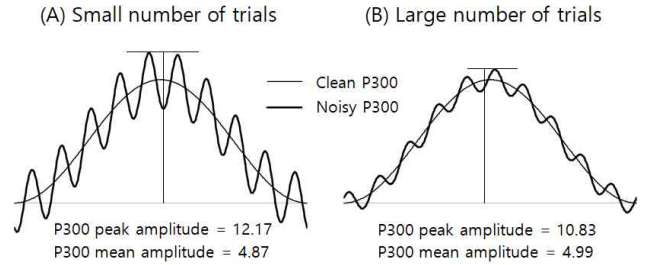


Figure 1. Examples of a clean P300 and noisy P300s

서 무죄인 용의자에게 P300 CIT를 실시한 후 관련자극과 무관련자극의 P300 정점을 측정하면, 관련자극의 P300 정점이 무관련자극의 P300 정점보다 더 크게 나타날 가능성이 높다. 즉 무죄인 용의자를 범인으로 잘못 판단할 가능성이 높아지는 문제가 발생한다. 앞으로 이 문제를 '시행수 차이로 인한 고주파 배경뇌파의 문제'로 부르기로 한다.

Luck(2014)은 이 문제에 대하여 두 가지 해결방법을 제안하였다. 한 가지는 P300 정점을 사용하는 대신 P300이 나타나는 구간의 평균을 사용하는 것이다. 즉, P300 파형에서 가장 높은 지점의 값을 사용하는 대신에 P300이 나타나는 구간의 평균 진폭을 사용하는 것이다. 그림 1의 (A)와 (B)에서 볼 수 있는 것처럼 P300 진폭이 시작하는 지점에서부터 끝나는 지점까지의 평균 진폭은 시행수에 따라 거의 달라지지 않는다. 다른 한 가지는 시행수가 더 많은 조건의 시행들 중 일부만 선택해서 두 조건의 시행수를 동일하게 만들어주는 것이다. 그러나 이 방법은 전체 표본 자료 중 일부만 사용하므로 통계적 검증력이 낮아진다(Luck, 2014).

P300 CIT에 관한 선행연구에서는 이 문제와 관련된 두 가지 방법을 사용하고 있다. 한 가지 방법은 Luck(2014)의 제안과 같이 P300의 정점을 사용하는 대신에 일정 구간의 평균 진폭을 사용하는 것이다(Lefebvre, Marchand, Smith, & Connolly, 2009; Rosenfeld, Soskins, Bosh, & Ryan, 2004). 그러나 이 방법은 구간의 폭을 얼마로 정해야 할지에 대한 문제가 있다. 구간의 폭을 너무 좁게 설정하면 고주파 배경뇌파가 영향을 미칠 것이며 너무 넓게 설정하면 P300이 아닌 성분이 측정값에 포함될 것이다. 선행연구에서는 대부분 P300 정점 전후 100 ms 구간의 평균전위를 사용하고 있기 때문에(Lefebvre et al., 2009; Rosenfeld et al., 2004), 10 Hz 보다 낮은 고주파 배경뇌파는 여전히 P300 진폭에 영향을 미칠 가능성이 남아 있다. 다른 한 가지 방법은 뇌파에 저역통과필터(low-pass filter)를 적용하여 고주파 성분을 제거한 후 ERP를 산출하는 것으로(Farwell & Donchin, 1991; Rosenfeld, Shue, & Singer, 2007), 평균 진폭을 사용하는 것과 본질적으로 동일한 것이다(Luck, 2014). 그러나

낮은 주파수(약 2~4 Hz) 성분까지 제거될 정도로 낮은 저역 통과필터를 사용하면 뇌파가 왜곡되는 부작용이 나타나며 (Luck, 2014), 이보다 높은 저역통과필터를 사용하면 고주파 성분이 완전히 제거되지 않는 문제점이 있다. 결과적으로 P300 CIT의 선행연구들에서 P300 평균 진폭을 사용하거나 저역통과필터를 사용하고 있지만, 여전히 시행수 차이로 인한 고주파 배경뇌파의 문제가 남아있을 가능성이 높다.

관련자극의 P300 진폭이 무관련자극의 P300 진폭보다 의미 있게 더 큰지를 판단하는데 여러 가지 방법들이 사용되었다. 이 방법들은 크게 세 가지 범주로 나누어진다. 첫 번째는 유죄집단과 무죄집단을 가장 잘 변별하는 P300의 특성을 찾아서, 유무죄의 기준으로 사용하는 것이다(Abootalebi, Moradi, & Khalilzadeh, 2006; Arasteh, Moradi, & Janghorbani, 2016; Mertens & Allen, 2008). 그러나 P300 파형은 자극의 종류에 따라 달라지므로(Curmore, Djakovic, Kebbell, & Shum, 2009) 자극이 달라지면 유무죄의 기준점도 달라져야 한다는 문제점을 가지고 있다. P300 CIT에 사용되는 자극은 조사 중인 사안에 따라 거의 항상 달라지므로, P300 CIT를 실시할 때마다 판단을 위한 기준점을 새로 산출해야만 한다. 따라서 이 방법의 실용성이 문제가 될 수 있다. 두 번째는 관련자극의 ERP가 목표자극의 ERP와 무관련자극의 ERP 중 어느 것과 더 유사한지를 평가하는 방법이다(Farwell & Donchin, 1991; Wasserman & Bockenholt, 1989). 이 방법은 범인인 경우 관련자극과 목표자극의 ERP가 유사하다고 가정하는데, 이 가정의 타당성이 문제가 될 수 있다(Rosenfeld, 2005). 마지막 세 번째는 관련자극의 P300 진폭이 무관련자극의 P300 진폭보다 통계적으로 유의하게 더 큰지를 평가하는 방법이다. 이 방법은 앞의 두 방법과 달리 논리적 문제가 없으며 P300 CIT에서 가장 많이 사용되고 있다(Ambach, Bursch, Stark, & Vaitl, 2010; Bowman, Filetti, Alsufyani, Janssen, & Su, 2014; Ellwanger, Rosenfeld, Sweet, & Bhatt, 1996; Kubo & Nittono, 2009; Lefebvre et al., 2009; Lukács et al., 2016). 통계적 방법으로 독립표본 t 검증(independent samples t test)과 부트스트랩 방법(bootstrap method), 순열 검증(permutation test)을 사용할 수 있다. 일반적으로 세 가지 통계적 방법의 검증력은 크게 다르지 않다(Good, 2005; Robinson, 1987). 초기 P300 CIT 연구에서는 독립표본 t 검증이 사용되기도 하였고(Ellwanger et al., 1996; Rosenfeld, Angell, Johnson, & Qian, 1991; Rosenfeld, Reinhart, Bhatt, Ellwanger, Gora, Sekera, & Sweet, 1998; Rosenfeld, Sweet, Chuang, Ellwanger, & Song, 1996), 최

근에는 P300 CIT에서 다중비교에 의한 1 종 오류의 증가를 통제하기 위한 용도로 순열 검증이 사용된 적이 있었지만 (Bowman et al., 2014), 부트스트랩 방법이 압도적으로 많이 사용되고 있다.

P300 CIT에서 관련자극의 P300 진폭이 무관련자극의 P300 진폭보다 통계적으로 유의하게 더 큰지를 판단하기 위한 부트스트랩 검증¹⁾이 그림 2에 제시되어 있으며, 다음과 같은 절차로 진행된다. ① 관련자극에 대한 30개의 단일 시행 뇌파에서 30개를 복원추출하고, 무관련자극에 대한 120개의 단일 시행 뇌파에서 120개를 복원추출한다. ② 복원추출된 뇌파를 이용하여 관련자극과 무관련자극의 ERP를 각각 산출한다. ③ ERP에서 P300 진폭을 측정 후 관련자극의 P300 진폭에서 무관련자극의 P300 진폭을 뺀 차이값을 구한다. ④ ①~③의 과정을 1,000번 반복하여 P300 진폭 차이값에 대한 표집분포를 만든다. ⑤ 부트스트랩 표집분포에서 0보다 작은 면적의 비율을 계산하여($p = .03$) 이 면적이 사전에 정한 유의수준($\alpha = .05$)보다 작으면 관련자극의 P300 진폭이 무관련자극의 P300 진폭보다 통계적으로 유의하게 더 큰 것으로 판단한다.

P300 CIT에서 사용할 수 있는 순열 검증의 방법이 그림 3에 제시되어 있으며, 다음과 같은 절차로 진행된다. ① 관련자극과 무관련자극 간의 P300 진폭 차이값을 구한다($4.9 \mu V$). ② 관련자극에 대한 단일 시행 뇌파들과 무관련자극에 대한 단일 시행 뇌파들을 모두 섞는다. ③ 합쳐진 전체 단일 시행 뇌파에서 30개의 뇌파를 무선표본추출하여 관련자극의 뇌파로 간주한다. 나머지 120개의 뇌파는 무관련자극의 뇌파로 간주한다. ④ 재표본추출된 뇌파를 이용하여 관련자극과 무관련자극의 P300 진폭 차이값을 계산한다. ⑤ ③~④의 과정을 1,000번 반복하여 표집분포를 산출한다. ⑥ 표집분포에서 원래 표본의 P300 진폭 차이값 $4.9 \mu V$ 보다 더 큰 영역의 비율을 계산한다($p = .055$). ⑦ 이 비율이 사전에 정한 유의수준($\alpha = .05$)보다 작으면 관련자극의 P300 진폭이 무관련자극의 P300 진폭보다 통계적으로 유의하게 더 큰 것으로 판단하며, 그렇지 않으면 P300 진폭 차이가 통계적으로 유의하지 않다고 판단한다.

앞서 설명한 바와 같이 부트스트랩 검증은 부트스트랩 표집분포에서 P300 진폭 차이값이 '0'보다 작은 경우의 수를 세어서 가설검증에 활용한다. 이때 '0'을 기준으로 하는 이

1) 부트스트랩 방법을 이용한 가설검증이 더 정확한 용어이지만 본 연구에서는 편의상 부트스트랩 검증으로 간략하게 부르려고 한다. 부트스트랩 검증을 수행하는 구체적인 절차는 여러 가지가 있지만 본 연구에서는 가장 많은 연구자들이 사용하는 방법을 적용하였다.

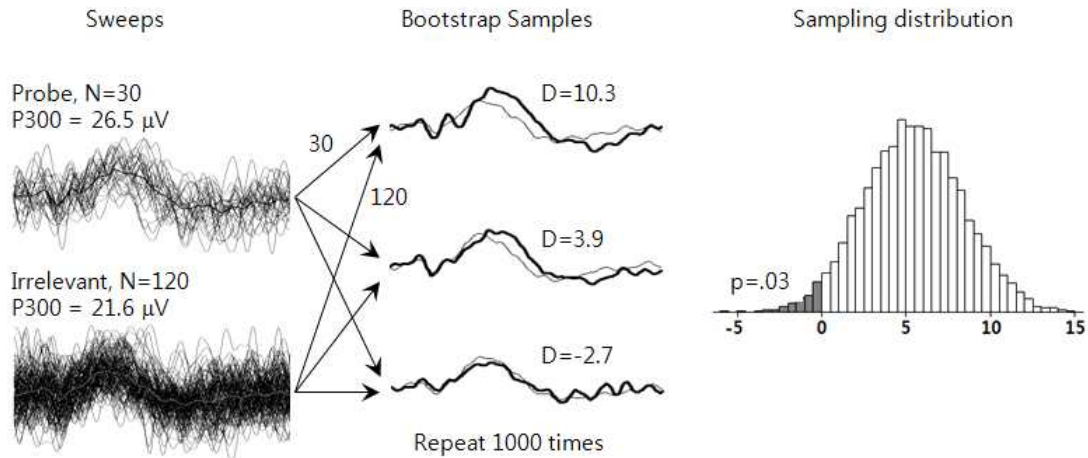


Figure 2. The bootstrap method for the difference between probe and irrelevant P300 amplitudes

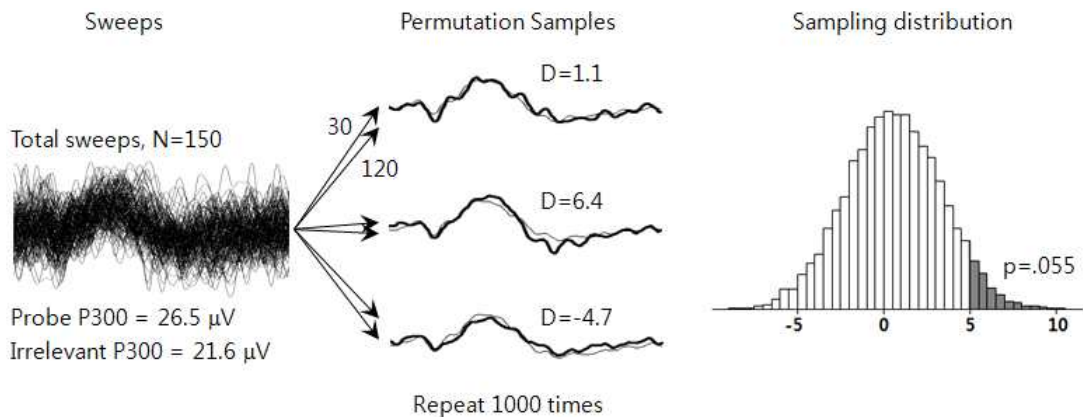


Figure 3. The permutation test for the difference between probe and irrelevant P300 amplitudes

유는 영가설이 참일 때 P300 진폭 차이가 '0'이라고 가정하기 때문이다. 그러나 실제로는 시행수 차이로 인한 고주파 배경뇌파의 문제 때문에 영가설이 참일 때 P300 진폭 차이는 0보다 큰 값을 가지게 된다. 따라서 P300 CIT에서 부트스트랩 검증으로 P300 진폭 차이에 대한 유의확률을 계산하면 실제보다 더 작은 확률값이 산출될 것이다. 즉, 시행수 차이로 인한 고주파 배경뇌파의 문제가 부트스트랩 검증의 1종 오류율을 증가시킬 것으로 예측된다.

부트스트랩 검증은 영가설이 참일 때 P300 진폭 차이가 0이라는 현실적이지 못한 가정을 하는 반면, 순열 검증에는 이러한 가정이 개입하지 않는다. 순열 검증에서는 영가설이 참일 때 P300 진폭 차이의 표집분포를 경험적으로 만들고, 이 표집분포에서 관찰된 P300 진폭 차이값이 차지하는 위치를 파악한다. P300 진폭 차이의 표집분포를 구성하는 값들에 시행수 차이로 인한 고주파 배경뇌파의 문제가 포함되어 있으며, 관찰된 P300 진폭 차이값에도 시행수 차이로 인한 고주파 배경뇌파의 문제가 포함되어 있다. 표집분포와 관찰된 P300 진폭 차이값 모두에 동일한 원천의 + 오차가 포함

되어 있으므로, 표집분포에서 P300 진폭 차이값이 차지하는 위치는 이 오차에 영향을 받지 않을 것이다. 따라서 순열 검증에서 산출하는 유의확률은 시행수 차이로 인한 고주파 배경뇌파의 문제에 영향을 받지 않을 것으로 예상된다.

본 연구의 목적은 P300 CIT에서 시행수 차이로 인한 고주파 배경뇌파의 문제가 지역통과필터와 P300 평균 진폭을 사용하여도 여전히 남아있는 지를 살펴보고, 이러한 문제가 부트스트랩 검증의 1종 오류에 영향을 미치는지 검증하고, 고주파 배경뇌파의 문제가 존재하여도 순열 검증의 1종 오류에는 영향을 미치지 않는지 확인하는 것이다. 더불어 두 통계적 절차의 검증력도 함께 확인하였다. 이를 위하여 몬테카를로(Monte Carlo) 연구를 수행하였다. 일반적인 몬테카를로 연구에서는 가상의 모집단(즉 평균, 표준편차, 분포의 모양)을 정의하고 이 모집단으로부터 무선표본을 추출하여 통계적 절차의 1종 오류와 검증력을 평가한다. 그러나 고주파 배경뇌파가 포함된 단일 시행 뇌파의 모집단을 정의하는 것이 매우 어렵기 때문에, 본 연구에서는 실제로 측정된 단일 시행 뇌파들로 모집단을 구성하였다. 연구 1에서는 실험

연구를 통하여 단일 시행 뇌파들로 유한한 크기의 모집단을 구성하였으며, 연구 2에서는 연구 1에서 구성한 모집단을 이용하여 부트스트랩 검증과 순열 검증의 1 종 오류율 및 검증력을 산출하였다.

연구 1. 단일 시행 뇌파의 모집단 구성

연구 1에서는 P300 CIT를 사용한 실험연구를 통하여 몬테카를로 연구에 사용할 단일 시행 뇌파의 모집단을 구성하였다. 한명의 실험참가자에게서 충분한 수의 단일 시행 뇌파를 얻기 어렵기 때문에 다수의 실험참가자에게서 측정한 뇌파를 모아서 단일 시행 뇌파의 모집단을 구성하였다. 각 검증방법의 1 종 오류율을 추정하기 위하여 무죄조건에서 단일 시행 뇌파의 모집단을 구성하였으며, 각 검증방법의 검증력을 추정하기 위하여 유죄조건에서 단일 시행 뇌파의 모집단을 구성하였다. 선행연구들에서는 30 Hz 저역통과필터를 사용하거나(Kubo & Nittono, 2009), 10 Hz 내외의 저역통과필터를 사용하거나(Bergström, Anderson, Buda, Simons, & Richardson-Klavehn, 2013; Mertens & Allen, 2008), 5 Hz 내외의 저역통과필터를 사용하였다(Matsuda, Nittono, & Ogawa, 2011; Rosenfeld, Hu, & Pederson, 2012; Verschuere, Rosenfeld, Winograd, Labkovsky, & Wiersema, 2009). 따라서 본 연구에서도 30 Hz와 10 Hz, 5 Hz 저역통과필터를 사용한 세 가지 종류의 무죄조건 모집단과 유죄조건 모집단을 구성하였다. 추가적으로 유한 모집단의 크기와 실험참가자들 간 ERP(구체적으로 P300 잠재시간)의 이질성이 연구결과에 영향을 미칠 수 있으므로, 모집단 크기와 P300 잠재시간의 이질성이 큰 모집단과 모집단 크기와 P300 잠재시간의 이질성이 작은 모집단을 구성하였다. 결과적으로 실험자료를 이용하여 실험 조건 2 수준(유죄, 무죄)과 저역통과필터 3 수준(30 Hz, 10 Hz, 5 Hz), 모집단의 크기 2 수준을 모두 곱하여 총 12개의 모집단을 구성하였다.

방 법

실험참가자

대학교 홈페이지 게시판에 실험참가자 모집을 공고하여 정신과 병력과 뇌손상 경험이 없는 실험참가자들을 모집하였다. 남녀 대학생 34명이 실험에 참가하였다. 성별은 남자 16명 여자 18명이었으며, 평균연령은 21.8세(범위 19-26세)였다. 34명중 17명은 무죄조건에, 나머지 17명은 유죄조건에 참가

하였다. 실험참가자는 실험참가에 대한 금전적 보상을 받았다. 본 연구는 충북대학교 생명윤리위원회 승인을 받았다(IRB No. CBNU-201706-SBETC-487-01).

실험자극

P300 CIT에서는 실험참가자에게 모의범죄를 수행하도록 한 후 모의범죄와 관련된 정보를 실험자극으로 이용하거나(Abootalebi et al., 2006; Mertens & Allen, 2008), 허위 기억상실을 가장한 상황에서 자기참조적인 정보를 실험자극으로 이용한다(Rosenfeld et al., 2012, Verschuere et al., 2009). 본 연구에서는 자기참조적 정보를 이용하여 P300 CIT에 사용할 자극을 구성하였다. 유죄조건에서는 실험참가자의 이름, 생일, 출신고등학교, 전공학과, 거주지역을 관련자극으로 사용하였으며, 무죄조건에서는 실험참가자와 관련이 없는 이름, 생일, 고등학교, 학과, 지역을 관련자극으로 사용하였다. 유죄조건과 무죄조건 모두에서 무관련자극은 실험참가자와 관련이 없는 이름, 생일, 고등학교, 학과, 지역을 각각 4개씩 만들어 총 20개를 구성하였다. 목표자극은 실험참가자와 관련이 없는 이름, 생일, 고등학교, 학과, 지역을 각각 1개씩 만들어 사용하였다.

실험절차

실험참가자가 실험실에 도착하면 실험참가동의서를 읽고 서명을 한 후에 유죄조건 또는 무죄조건에서 실험에 참가하도록 하였다. 실험참가자가 어떤 조건에 참가할지는 사전에 정해진 순서에 따라서 결정하였다. 실험참가자의 두피에 전극을 부착하고 자기참조적 정보에 대한 P300 CIT를 실시하였다. 실험참가자에게 목표자극이 무엇인지 구두로 알려주었으며, 잘 암기하였는지 확인하였다. 이후 무죄조건 실험참가자에게는 “지금부터 기억검사를 실시하겠습니다.”라는 안내와 함께 P300 CIT를 실시하였으며, 유죄조건 실험참가자에게는 아래의 지시문을 읽어주고 P300 CIT를 실시하였다.

“지금부터 기억검사를 실시하겠습니다. 본 실험은 실험참가자가 사고 후에 더 큰 피해 보상금을 얻기 위하여 허위로 기억 손상을 주장하는 상황을 가정합니다. 그러므로 실험참가자께서는 본인과 관련된 어떠한 정보도 기억나지 않는다고 주장해야 합니다. 검사가 시작되고 실험참가자와 관련된 정보가 모니터에 제시되었을 때, 자신과 관련된 정보가 아닌 것처럼 생각하고 검사에 반응해야 합니다.”

P300 CIT에서 관련자극 5개와 무관련자극 20개, 목표자극 5개를 섞어서 무선으로 하나씩 제시하였으며, 실험참가자에게 목표자극이 제시되면 ‘예’ 버튼(마우스 왼쪽 버튼)을 가능한 빠르게 누르고 다른 자극(즉 관련자극과 무관련자극)이 제시되면 ‘아니오’ 버튼(마우스 오른쪽 버튼)을 가능한 빠르게 누르라고 지시하였다. 자극은 실험참가자로부터 1m 앞에 놓여 있는 컴퓨터 모니터 중앙에 1.5cm 높이의 신명조체 한글로 제시하였다. 자극은 300 ms 동안 제시하였으며, 자극 간 제시 간격은 평균 2,700 ms(범위 2,200 ~ 3,200 ms)였다. 30개의 자극을 9회 반복 제시하여 총 270회 자극을 제시하였다. 자극의 제시와 반응의 기록은 SuperLab 5.0을 이용하였다.

뇌파기록 및 분석

뇌파는 Grass Model 12 Neurodata Acquisition System (Grass Instruments, Quincy, MA, USA)을 이용하여 측정하였다. Ag/AgCl 전극을 국제 10-20 체계에 따라 F3, Fz, F4, Cz, P3, Pz, 및 P4에 부착하여 뇌파를 기록하였으며, 양쪽 귓볼에 연결된 기준전극을 부착하였고, 이마에 접지전극을 부착하였다. 왼쪽 눈의 위와 아래에 전극을 부착하여 안전도 (electrooculogram: EOG)를 측정하였다. 뇌파는 0.3 ~ 100 Hz 대역여과한 후 20,000배 증폭하였으며, 안전도는 5,000 배 증폭하였다. 증폭된 뇌파를 250 Hz 표집률로 디지털 변환하여 컴퓨터에 저장하였다.

P300 CIT에서는 P300 진폭이 가장 크게 측정되는 Pz에서 측정된 뇌파를 주로 사용하므로, 본 연구에서도 Pz에서 측정된 뇌파만 분석에 사용하였다. 먼저 전체 뇌파에 대해서 EOG를 이용하여 눈깜박임과 안구운동에 의한 뇌파의 변형을 교정하였다(Semlitsch, Anderer, Schuster, & Presslich, 1986). 다음으로 전체 뇌파에 30 Hz 저역통과필터와 10 Hz 저역통과필터, 5 Hz 저역통과필터를 각각 적용하여 고주파 성분이 서로 다르게 포함된 세 가지 종류의 뇌파를 구성하였다. 마지막으로 전체 뇌파를 자극 제시 전 100 ms부터 자극 제시 후 1,400 ms까지 총 1,500 ms 구간으로 나누어, 관련

자극과 무관련자극, 목표자극에 대한 단일 시행 뇌파를 구분하였다.

결 과

유죄조건에 참가한 2명과 무죄조건에 참가한 2명은 전체 시행의 50% 이상에서 눈 깜박임을 보였거나 목표자극에 대한 정답률이 50% 미만이어서 분석에서 제외하였다. 무죄조건에서는 관련자극과 무관련자극이 모두 실험참가자에게 관련이 없는 자극이므로, 관련자극과 무관련자극을 합하여 무관련자극으로 분류하였다.

반응오류율과 반응시간

실험참가자들이 검사자극에 적절한 주의를 기울였는지 평가하기 위하여 표 1과 같이 반응오류율과 반응시간을 산출하였다. 무죄조건과 유죄조건에서 목표자극의 평균 반응오류율은 10% 미만이었으며, 관련자극과 무관련자극에 대한 평균 반응오류율은 1% 내외로 목표자극에 대한 반응오류율보다 낮았다. 유죄조건에서 목표자극과 관련자극, 무관련자극에 대한 평균 반응시간은 통계적으로 유의하게 다른 것으로 나타났다, $F(2,28) = 22.362, p < .001$. Bonferroni 교정을 사용한 다중비교 결과 목표자극에 대한 평균 반응시간 794 ms가 관련자극과 무관련자극에 대한 반응시간 709 ms와 723 ms보다 통계적으로 유의하게 더 느렸다. 무죄조건에서 목표자극에 대한 반응시간 804 ms는 무관련자극에 대한 반응시간 722 ms보다 통계적으로 유의하게 더 길었다, $F(1,14) = 48.118, p < .001$.

단일 시행 뇌파의 모집단 구성

여러 명의 실험참가자들에게서 측정된 뇌파를 합하여 단일 시행 뇌파의 모집단을 구성하였다. 이때 유한 모집단의 크기와 실험참가자들 간 P300 잠재시간의 이질성을 통제하기 위하여 유죄조건과 무죄조건 각각에서 실험참가자 15명의 자료를 모두 포함한 모집단(‘모집단 크기 15’ 조건)과 P300 잠

Table 1. Error rates (%) and reaction times (ms)

		Target	Probe	Irrelevant
Error rates	Guilty	9.9 (7.6)	0.4 (1.2)	1.0 (1.4)
	Innocent	8.3 (7.1)	-	1.1 (1.1)
Reaction times	Guilty	794 (106)	709 (113)	723 (130)
	Innocent	804 (95)	-	722 (108)

Note. standard deviations are in parentheses.

재시간이 서로 비슷한 실험참가자 10명의 자료만 포함한 모 집단(모집단 크기 10' 조건)을 구성하였다. 유죄조건에서 모 집단 크기 15 조건에 포함된 15명의 관련자극에 대한 P300 잠재시간의 표준편차는 84 ms였으며, 무관련자극에 대한 P300 잠재시간은 표준편차는 96 ms였던 반면, 모집단 크기 10 조건에 포함된 10명의 실험참가자에 대한 관련자극의 P300 잠재시간 표준편차는 57 ms였으며, 무관련자극에 대한 P300 잠재시간 표준편차는 56 ms였다. 무죄조건에서 모집단 크기 15 조건에 포함된 15명의 관련자극과 무관련자극에 대한 P300 잠재시간 표준편차는 80 ms였던 반면, 모집단 크기 10 조건에 포함된 10명의 참가자에 대한 P300 잠재시간 표준편차는 49 ms였다.

모집단 크기 15인 유죄조건에서는 관련자극에 대한 675개(45개 × 15명)의 단일 시행 뇌파로 관련자극에 대한 단일 시행 뇌파의 모집단을 구성하였으며, 무관련자극에 대한 2,700개(180개 × 15명)의 단일 시행 뇌파로 무관련자극에 대한 단일 시행 뇌파의 모집단을 구성하였다. 모집단 크기 15인 무죄조건에서는 관련자극과 무관련자극이 모두 실험참가자와 관련이 없는 자극이므로 관련자극과 무관련자극을 합하여 3,375개(225개 × 15명)의 단일 시행 뇌파로 모집단을 구성한 후 관련자극과 무관련자극의 단일 시행 뇌파의 모집단으로 사용하였다. 모집단 크기 10인 유죄조건에서는 관련자극에 대한 450개(45개 × 10명)의 단일 시행 뇌파로 관련자극에 대한 단일 시행 뇌파의 모집단을 구성하였으며, 무관련자극에 대한 1,800개(180개 × 10명)의 단일 시행 뇌파로 무관련자극에 대한 단일 시행 뇌파의 모집단을 구성하였다. 모집단 크기 10인 무죄조건에서는 관련자극과 무관련자극을 합하여 2,250개(225개 × 10명)의 단일 시행 뇌파로 모집단

을 구성한 후 관련자극과 무관련자극의 단일 시행 뇌파의 모집단으로 사용하였다.

30 Hz, 10 Hz, 및 5 Hz 저역통과필터를 각각 적용한 세 개의 뇌파에 대해서 모집단 크기 15인 조건과 10인 조건의 모집단을 각각 만들어 몬테 카를로 연구에 사용하였다. 그림 4에 각 모집단에 대한 ERP가 제시되어 있다. 모든 조건에서 약 400 ms ~ 700 ms 사이에 큰 정적 정점이 나타났으며, 이 정점을 P300으로 파악하였다. 유죄조건과 무죄조건 모두 목표자극에 대한 P300 진폭이 가장 크게 나타났다. 유죄조건에서는 관련자극의 P300 진폭이 무관련자극의 P300 진폭보다 더 크게 나타났다. 유죄조건과 무죄조건 모두에서 30 Hz 저역통과필터를 적용한 경우에 비하여 10 Hz 저역통과필터와 5 Hz 저역통과필터를 적용한 경우 ERP에서 고주파 성분이 점차 제거된 것을 볼 수 있다. 무죄조건에서는 모집단 크기 15인 조건과 모집단 크기 10인 조건 간에 ERP에서 큰 차이가 없었다. 유죄조건에서는 모집단 크기 10인 경우에 비하여 모집단 크기 15인 경우에 관련자극의 P300 진폭이 약간 더 넓게 분포하는 것으로 보인다. 모집단 크기 15인 경우가 모집단 크기 10인 경우보다 P300 잠재시간의 이질성이 더 크기 때문에 나타난 현상으로 보인다.

연구 2. 몬테 카를로 연구

시행수 차이로 인한 고주파 배경뇌파의 문제가 부트스트랩 검증과 순열 검증에 미치는 영향을 확인하기 위하여 연구 1에서 구성한 단일 시행 뇌파의 모집단을 이용한 몬테 카를로 연구를 수행하였다.

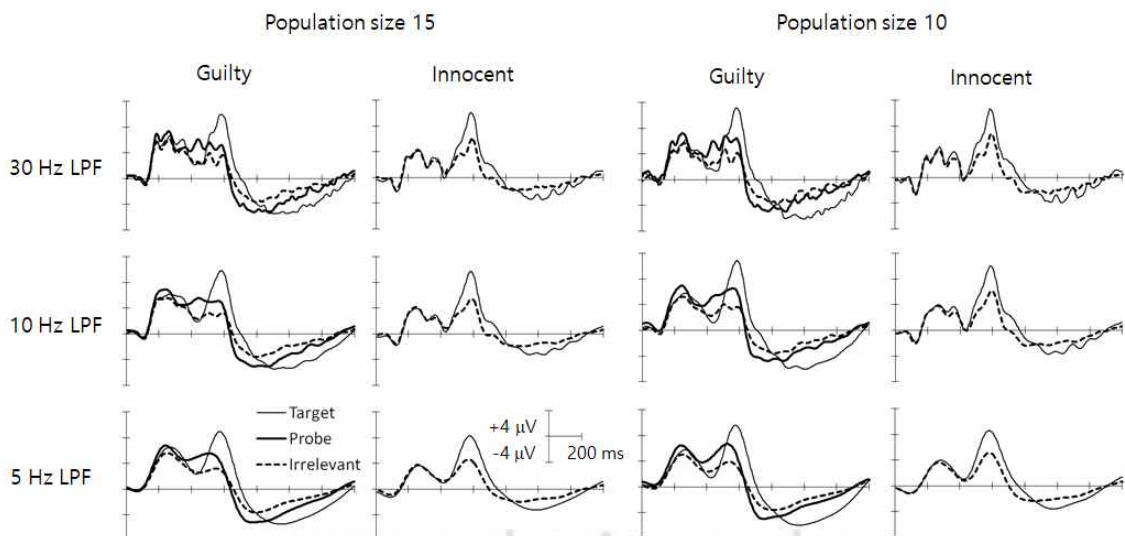


Figure 4. The ERPs for each condition

방 법

조건

시행수 차이로 인한 고주파 배경뇌파의 문제가 P300 CIT의 뇌파에 나타나는지, 저역통과필터의 사용과 100 ms 구간의 P300 평균 진폭을 사용하는 것이 이러한 영향을 제거하는지, 시행수 차이로 인한 고주파 배경뇌파의 문제가 부트스트랩 검증과 순열 검증의 결과에 영향을 미치는지 확인하기 위한 조건을 구성하였다. 먼저 각 검증방법의 1 종 오류율과 검증력을 확인하기 위하여 무죄조건과 유죄조건을 구성하였다. 두 번째로 선행연구에서 주로 사용하는 세 가지 저역통과필터 즉, 30 Hz 저역통과필터와 10 Hz 저역통과필터, 5 Hz 저역통과필터를 적용한 조건을 구성하였다. 세 번째로 P300의 정점 진폭을 측정하는 조건과 P300의 평균 진폭을 측정하는 조건을 구성하였다. P300의 평균 진폭을 사용한 선행 연구들은 대부분 P300 정점 전후 100 ms 구간의 평균을 사용하므로(Bergström et al., 2013; Marchand, Inglis-Assaff, & Lefebvre, 2013; Rosenfeld et al., 2007), 본 연구에서도 100 ms 구간의 평균 진폭을 사용하였다. 여기에 모집단 크기와 P300 잠재시간의 이질성이 큰 모집단 조건과 모집단 크기와 P300 잠재시간의 이질성이 작은 모집단 조건을 추가하여 최종적으로 24개(2 × 3 × 2 × 2)의 실험조건을 구성하였다. 이 조건들 중 P300 진폭을 측정하는 방법에 대한 조건에서는 모집단이 다르게 구성되지는 않으며 표본에서 P300을 측정하는 방법만 구분되는 조건이다.

표본추출

연구 1에서 구성한 단일 시행 뇌파의 모집단에서 표본을 추출하여 부트스트랩 검증과 순열 검증을 실시하였다. 선행연구에 따라서 관련자극의 표본 수는 30, 무관련자극의 표본 수는 120으로 정하였다(Rosenfeld et al., 2004). 반복 시행 수는 선행 몬테 카를로 연구들(Janssen & Pauls, 2005; Konietzschke & Pauly, 2014)과 컴퓨터 연산 시간을 고려하여 10,000번으로 정하였다. 즉, 실험조건 2 수준(유죄, 무죄)과 저역통과필터 3 수준(30 Hz, 10 Hz, 5 Hz), 모집단 크기 2 수준(15, 10)을 모두 곱하여 총 12개의 모집단 각각에서 관련자극의 단일 시행 뇌파 30개와 무관련자극 단일 시행 뇌파 120개를 표본추출하여 분석하는 과정을 10,000번 반복하여 부트스트랩 검증과 순열 검증의 1 종 오류율과 검증력을 산출하였다.

분석방법

표본추출된 단일 시행 뇌파들을 관련자극과 무관련자극 각각에 대하여 평균하여 관련자극의 ERP와 무관련자극의 ERP를 산출하였으며, ERP로부터 P300 진폭을 계산하였다. P300 진폭은 P300 CIT에서 가장 유용하며(Soskins, Rosenfeld, & Niendam, 2001), 널리 사용되는 정점-정점(peak-to-peak) 방법을 사용하였다. 정점-정점 방법은 P300 정점과 P300 이후에 나타나는 부적 정점 간의 차이로 정의한다(Rosenfeld et al., 2004). 실험설계에 따라서 두 가지 방법으로 정점-정점 P300 진폭을 계산하였다. 한 가지는 정점의 값을 사용하는 것이며 다른 한 가지는 100 ms 구간의 평균을 사용하는 것이다. 정점을 사용하는 P300 진폭은 자극 제시 후 400 ms ~ 700 ms 사이에 가장 큰 정적 정점의 값에서 P300 정점이 나타난 이후부터 1,400 ms 사이에 가장 큰 부적 정점의 값을 뺀 값으로 정의하였다. 100 ms 구간의 평균을 사용하는 P300 진폭은 자극 제시 후 400 ms ~ 700 ms 사이에 100 ms 구간의 평균값이 가장 큰 값에서 P300 정점이 나타난 이후부터 1,400 ms 사이에 100 ms 구간의 평균값이 가장 작은 값을 뺀 값으로 정의하였다.

관련자극의 P300 진폭이 무관련자극의 P300 진폭보다 통계적으로 유의하게 더 큰지를 판단하기 위하여 서론에서 자세히 기술한 부트스트랩 검증과 순열 검증을 실시하였다. 유의수준은 선행연구에서 주로 사용하는 .10(Rosenfeld et al., 2007; Verschuere et al., 2009)과 .05(Matsuda et al., 2011; Meijer, Smulders, Merckelbach, & Wolf, 2007; Mertens & Allen, 2008)를 사용하였다. 무죄조건에서는 각 실험조건별로 10,000번 반복한 검증결과에서 영가설이 기각된 비율을 계산하여 1 종 오류율을 산출하였으며, 유죄조건에서는 각 실험조건별로 10,000번 반복한 검증결과에서 영가설이 기각된 비율을 계산하여 검증력을 산출하였다.

결 과

P300 진폭

각 모집단에서 무선표본을 추출하여 ERP의 P300 진폭을 계산하는 과정을 10,000번 반복하여 실험조건별로 P300 진폭의 평균을 산출한 결과가 표 2에 제시되어 있다. 먼저 무죄 조건을 살펴보면, 모집단의 크기와 저역통과필터의 종류, P300 진폭의 측정 방법에 관계없이 관련자극의 P300 진폭에서 무관련자극의 P300 진폭을 뺀 결과가 모두 +값으로 나타났다. 모집단 크기 15와 모집단 크기 10인 경우 모두, 30 Hz 저역통과필터와 P300 정점 진폭을 사용한 경우에 P300

Table 2. P300 amplitudes (μV) for each conditions

PS	Group	P300 measure	30 Hz LPF			10 Hz LPF			5 Hz LPF		
			Probe	Irrelevant	P-I	Probe	Irrelevant	P-I	Probe	Irrelevant	P-I
15	Innocent	Peak	10.24	7.66	2.58	8.24	6.39	1.85	8.32	7.14	1.18
		Mean	7.09	5.76	1.33	6.84	5.59	1.25	7.65	6.64	1.00
	Guilty	Peak	15.93	9.35	6.58	13.84	8.11	5.73	12.50	7.48	5.02
		Mean	12.28	7.19	5.09	12.13	7.10	5.03	11.80	6.99	4.81
10	Innocent	Peak	13.61	10.92	2.69	11.23	9.56	1.67	9.20	8.10	1.10
		Mean	9.21	8.04	1.17	9.07	7.99	1.07	8.45	7.51	0.94
	Guilty	Peak	16.48	10.09	6.38	14.36	8.63	5.72	13.11	7.94	5.17
		Mean	12.83	7.77	5.06	12.66	7.66	5.00	12.36	7.55	4.81

Note. PS = population size, LPF = low-pass filter, Peak = peak amplitude, Mean = mean amplitude, P-I = probe P300 amplitude - irrelevant P300 amplitude.

진폭 차이 값이 2.58 μV 과 2.69 μV 로 가장 컸으며, 5 Hz 저역통과필터와 P300 평균 진폭을 사용하는 경우에 P300 진폭 차이 값이 1.00 μV 와 0.94 μV 로 가장 작았다. 결과적으로 5 Hz 저역통과필터를 사용하거나 P300 평균 진폭을 사용하여도 시행수 차이로 인한 고주파 배경뇌파의 문제가 여전히 존재하는 것으로 나타났다. 유죄조건에 유죄조건과 거의 비슷한 양상을 보였다.

1 종 오류율

몬테 카를로 연구를 통하여 추정된 부트스트랩 검증과 순열 검증의 1 종 오류율이 표 3에 제시되어 있다. 먼저 모집단 크기가 15일 때 부트스트랩 검증의 1 종 오류율은 30 Hz 저역통과필터를 사용하고 P300 정점 진폭을 사용하는 경우

에 유의수준 .05에서 .44, 유의수준 .10에서 .61로 매우 높았다. 더 낮은 Hz의 저역통과필터를 사용하거나 P300 평균 진폭을 사용하는 경우에 1 종 오류율은 점차 낮아져서, 5 Hz 저역통과필터와 P300 평균 진폭을 사용하는 경우에 1 종 오류율은 유의수준 .05에서 .16, 유의수준 .10에서 .27을 보였다. 모집단 크기 10일 때에도 비슷한 형태의 결과가 나타났다. 결과적으로, 시행수 차이로 인한 고주파 배경뇌파의 문제가 부트스트랩 검증에 영향을 미쳐서 1 종 오류율을 대폭 상승시켰다. 순열 검증의 1 종 오류율은 모집단 크기와 사용한 저역통과필터, P300 진폭의 측정 방법에 관계없이 유의수준과 동일한 값을 보였다. 즉, 시행수 차이로 인한 고주파 배경뇌파의 문제가 순열 검증의 1 종 오류율에는 전혀 영향을 미치지 않았다.

Table 3. Type 1 error rates of bootstrap and permutation test

PS	α	P300 measure	30 Hz LPF		10 Hz LPF		5 Hz LPF	
			B	P	B	P	B	P
15	.05	Peak	.44	.05	.30	.05	.18	.05
		Mean	.21	.05	.19	.05	.16	.05
	.10	Peak	.61	.10	.45	.10	.30	.10
		Mean	.34	.10	.32	.10	.27	.10
10	.05	Peak	.40	.05	.25	.05	.17	.05
		Mean	.18	.05	.17	.05	.15	.05
	.10	Peak	.57	.10	.39	.10	.28	.10
		Mean	.30	.10	.27	.10	.25	.10

Note. PS = population size, LPF = low-pass filter, B = bootstrap test, P = permutation test, Peak = peak amplitude, Mean = mean amplitude.

검증력

몬테 카를로 연구를 통하여 추정된 부트스트랩 검증과 순열 검증의 검증력이 표 4에 제시되어 있다. 모집단 크기 15일 때, 부트스트랩 검증의 검증력은 30 Hz 저역통과필터와 P300 정점 진폭을 사용할 때 가장 높았으며($\alpha = .05$ 에서 .92, $\alpha = .10$ 에서 .97), 5 Hz 저역통과필터와 P300 평균 진폭을 사용할 때 가장 낮았다($\alpha = .05$ 에서 .70, $\alpha = .10$ 에서 .81). 즉, 부트스트랩 검증의 검증력은 유의수준에 관계없이 낮은 주파수의 저역통과필터를 사용하고 P300 평균 진폭을 사용할 때 낮아졌다. 그러나 순열 검증의 검증력은 부트스트랩 검증과는 완전히 반대로 나타났다. 즉, 30 Hz 저역통과필터와 P300 정점 진폭을 사용할 때 검증력이 가장 낮았으며($\alpha = .05$ 에서 .38, $\alpha = .10$ 에서 .53), 5 Hz 저역통과필터와 P300 평균 진폭을 사용할 때 검증력이 가장 높았다($\alpha = .05$ 에서 .47, $\alpha = .10$ 에서 .61).

전체적으로 부트스트랩 검증의 검증력이 순열 검증의 검증력 보다 더 높게 나타났다. 부트스트랩 검증의 검증력이 더 높게 나타난 이유는 시행수 차이로 인한 고주파 배경뇌파의 문제 때문에 1 종 오류율이 과도하게 높아졌기 때문일 것이다. 이러한 가능성을 확인하기 위하여 순열 검증의 유의수준을 부트스트랩 검증의 1 종 오류율까지 높인 결과, 순열 검증의 검증력이 부트스트랩 검증의 검증력과 매우 유사한 수준까지 높아졌다(예, 5 Hz 저역통과필터와 P300 평균 진폭을 사용한 조건에서 유의수준으로 부트스트랩 검증의 1 종 오류율인 .16을 사용하면, 순열 검증의 검증력은 부트스트랩 검증의 검증력과 매우 비슷한 .71을 보였다).

모집단 크기 10인 경우의 부트스트랩 검증과 순열 검증의 검증력은 모집단 크기 15인 경우와 거의 비슷한 형태를 보

였다.

논 의

시행수가 서로 다른 두 조건의 P300 진폭을 비교하면 시행수가 적은 조건의 P300 진폭이 과대추정되는 문제가 발생한다. P300 CIT의 선행연구에서는 저역통과필터와 P300 평균 진폭을 사용하여 이 문제를 해결하려한 것으로 보인다. 본 연구에서는 실험자료를 이용한 몬테 카를로 연구를 통하여 P300 CIT에서 저역통과필터와 P300 평균 진폭을 사용하더라도 이러한 문제가 여전히 남아있으며, 순열 검증으로 이 문제가 극복될 수 있는지 확인하였다.

몬테 카를로 연구에 사용할 모집단을 구성하기 위하여 실험참가자들을 대상으로 P300 CIT를 실시하였다. 목표자극에 대한 반응오류를 평균은 무죄조건과 유죄조건에서 모두 10% 미만이었으며 목표자극에 대한 반응시간이 무관련자극에 대한 반응시간 보다 길었다. 다수의 목표자극을 사용한 선행연구(Mertens & Allen, 2008, Rosenfeld et al., 2004)에서도 목표자극에 대한 오류율은 10% 내외였고, 목표자극에 대한 반응시간은 800 ms 초반으로 본 연구의 결과가 비슷하였다. 이러한 결과는 본 연구에서 P300 CIT가 적절하게 수행되었음을 보여준다.

몬테 카를로 연구 결과, 무죄조건에서 관련자극의 P300 진폭이 무관련자극의 P300 진폭보다 더 크게 나타났다. 무죄조건에서 관련자극의 모집단과 무관련자극의 모집단은 같으므로 두 자극의 P300 진폭은 동일하게 측정되어야 한다. 그러나 관련자극의 시행수가 무관련자극의 시행수보다 작기 때문에 관련자극의 P300 파형에 고주파 노이즈가 더 크게

Table 4. Statistical powers of bootstrap and permutation test

PS	α	P300 measure	30 Hz LPF		10 Hz LPF		5 Hz LPF	
			B	P	B	P	B	P
15	.05	Peak	.92	.38	.85	.43	.74	.44
		Mean	.76	.46	.75	.46	.70	.47
	.10	Peak	.97	.53	.92	.57	.84	.59
		Mean	.86	.61	.85	.61	.81	.61
10	.05	Peak	.90	.34	.84	.40	.72	.44
		Mean	.75	.43	.74	.43	.68	.44
	.10	Peak	.96	.48	.92	.55	.84	.58
		Mean	.85	.58	.85	.58	.80	.58

Note. PS = population size, LPF = low-pass filter, B = bootstrap test, P = permutation test, Peak = peak amplitude, Mean = mean amplitude.

작용하여 P300 진폭이 과대추정된 것이다. 더욱이 이러한 문제점은 5 Hz 저역통과필터와 100 ms 폭의 P300 평균 진폭을 사용하여도 여전히 남아 있었다. 즉, 5 Hz 저역통과필터와 100 ms 폭의 P300 평균 진폭을 사용하여도 P300 진폭에 영향을 미치는 고주파 배경뇌파가 관련자극의 뇌파에 더 많이 남아 있다는 의미이다.

무죄조건에서 관련자극과 무관련자극의 P300 진폭 차이에 대한 부트스트랩 검증 결과, 모든 실험조건에서 유의수준보다 더 높은 1 종 오류율을 보였다. 5 Hz 저역통과필터와 100 ms 폭의 P300 평균 진폭을 측정하는 경우에도 유의수준 .05에서는 .15 ~ .16의 1 종 오류율을 보였으며, 유의수준 .10에서는 .25 ~ .27의 1 종 오류율을 보였다. 즉, 저역통과필터와 P300 평균 진폭을 사용하여도 ‘시행수 차이로 인한 고주파 노이즈의 문제’가 부트스트랩 검증의 1 종 오류율을 유의수준보다 10% 이상 증가시켰다.

P300 CIT에서 부트스트랩 검증이 널리 사용되었음에도 불구하고, 시행수 차이로 인한 고주파 노이즈의 문제 때문에 부트스트랩 검증의 1 종 오류율이 유의수준보다 높다는 것을 언급한 연구는 찾아볼 수 없었다. 아마도 부트스트랩 검증을 사용한 P300 CIT 연구들이 무죄조건을 포함하지 않은 경우가 많기 때문에 이러한 사실을 경험적으로 알기는 어려웠을 것이다. 그러나 무죄조건을 포함한 네 편의 선행 연구 결과(Bergstrom et al., 2013; Matsuda et al., 2011; Mertens & Allen., 2008; Rosenfeld et al., 2004)를 통합하여 산출한 부트스트랩 검증의 1 종 오류율은 유의수준 .05에서 약 .11(71명중 8명)이었으며, 유의수준 .10에서는 약 .33(24명중 8명)으로 유의수준보다 높은 1 종 오류율을 보였다. 비록 연구의 수가 적기는 하지만 본 연구의 결과와 유사한 결과가 나타났다.

무죄조건에서 관련자극과 무관련자극의 P300 진폭 차이에 대한 순열 검증 결과, 예상한 바와 같이 저역통과필터와 P300 측정방법에 관계없이 모든 조건에서 유의수준과 동일한 1 종 오류율이 나타났다. 즉, ‘시행수 차이로 인한 고주파 노이즈의 문제’가 P300 진폭에 영향을 미치더라도 순열 검증은 이를 반영하여 1 종 오류율을 적절하게 통제하는 것으로 밝혀졌다.

선행 연구들은 순열 검증과 부트스트랩 검증의 주된 차이점으로 표본을 복원추출하는지 또는 비복원추출하는지를 꼽고 있지만(Efron & Tibshirani, 1994; Rosenfeld, Ward, Meijer, & Tukhnenko, 2017), 본 연구에서 나타난 두 검증 방법 간의 1 종 오류율 차이는 두 검증에서 사용하는 표집분포의 차이가 만들어낸다. 두 검증 모두 표집분포를 표본자

료로부터 경험적으로 만들어 내지만, 부트스트랩 검증에서 만드는 표집분포는 검증 통계치들의 분포인 반면 순열 검증에서 만드는 표집분포는 영가설이 참일 때의 분포이다 (Zoumpoulaki, Alsufyani, & Bowman, 2015). 부트스트랩 검증에서 사용하는 표집분포에는 영가설이 참일 때 기대되는 값이 제공되지 않으므로, 단순히 ‘0’이라고 가정한다. 반면에 순열 검증의 표집분포에는 영가설이 참일 때 기대되는 값 (즉, P300 진폭 차이값)이 제공되는데, 표집분포의 평균이 바로 그 값이다. 이 표집분포의 평균은 시행수 차이로 인한 배경뇌파의 문제가 포함되어 있는 값이므로, 이 문제를 반영한 정확한 확률산출이 가능한 것이다.

유죄조건에서 저역통과필터와 P300 진폭의 측정방법에 관계없이 부트스트랩 검증의 검증력은 순열 검증의 검증력보다 더 높게 나타났다. 그러나 부트스트랩 검증의 결과는 ‘시행수 차이로 인한 고주파 노이즈의 문제’에 영향을 받아 검증력이 과대추정된 것이다. 실제로 저역통과필터와 P300 평균 전위를 사용하여 ‘시행수 차이로 인한 고주파 노이즈의 문제’를 줄여 나갈수록 부트스트랩 검증의 검증력은 점차 감소하였다. 따라서 높은 주파수의 저역통과필터를 사용하고 P300 정점 진폭을 측정하는 경우에 부트스트랩 검증의 검증력은 상당히 높은 수준으로 과대추정될 것이다. P300 CIT에서 부트스트랩 검증을 사용한 12편의 논문(Bergstrom et al., 2013; Cutmore et al., 2009; Kubo & Nittono, 2009; Lukacs, Grządziel, Kempkes, & Ansorge, 2019; Marchand et al., 2013; Matsuda et al., 2011; Meijer et al., 2007; Mertens & Allen, 2008; Rosenfeld, Biroshak, & Furedy, 2006; Rosenfeld et al., 2007; Rosenfeld et al., 2004; Verschuere et al., 2009)에서는 저역통과필터나 P300 진폭의 측정방법에 따라 검증력이 체계적으로 달라지지는 않았다. 숨긴정보의 종류(숨긴정보모의범죄, 자기참조적 정보, 카드 테스트, 학습한 단어, 얼굴사진)나 관련자극의 개수(1개, 6개, 12개), 자극의 형태(글자, 그림) 등이 P300 CIT의 결과에 영향을 미치는 것으로 알려져 있는데(Cutmore et al., 2009; Meijer, Selle, Elber, & Ben-Shakhar, 2014; Rosenfeld et al., 2007), 12편의 연구들은 숨긴정보의 종류와 관련자극의 개수, 자극의 형태 등이 달라서 저역통과필터의 효과를 체계적으로 관찰하기 어려웠을 것이다.

유죄조건에서 순열 검증의 결과는 부트스트랩 검증의 결과와 정반대로 나타났다. 즉, 낮은 주파수의 저역통과필터를 사용할수록 검증력이 증가하였으며, P300 정점을 사용하는 경우보다 P300 평균 진폭을 사용하는 경우에 검증력이 더 높았다. P300 CIT에서 순열 검증의 1 종 오류율은 적절하게

통제되었으므로 순열 검증의 검증력 또한 과대추정되지는 않았을 것이다. 따라서 5 Hz 저역통과필터와 P300 평균 진폭을 사용하는 경우에 산출된 순열 검증의 검증력은 과대추정되지 않은 가장 높은 검증력으로 추정된다. 무죄조건에서 저역통과필터와 P300 진폭의 측정방법에 관계없이 모든 조건에서 1 종 오류율이 동일했던 반면, 유죄조건에서 저역통과필터와 P300 진폭의 측정방법에 따라 검증력이 달라진 이유는 낮은 주파수의 저역통과필터를 사용하고 P300 평균 진폭을 측정함에 따라서 신호 대 잡음 비율(signal-to-noise ratio)이 증가하였기 때문일 것이다. 뇌파에서 잡음의 비율이 감소하면 통계적 검증력은 증가한다(Luck, 2014).

5 Hz 저역통과필터와 P300 평균 진폭을 사용하는 경우에 부트스트랩 검증의 검증력이 과대추정되는 정도가 가장 작았으며, 순열 검증의 검증력은 가장 높았다. 둘 사이의 검증력 차이는 .20 이상으로 순열 검증의 검증력이 낮았다. 뇌파가 아닌 일반적인 자료에서 부트스트랩 검증과 순열 검증의 검증력이 거의 비슷하다는 점을 고려하면(Good, 2005; Robinson, 1987), P300 CIT에서 과대추정되지 않은 최고의 검증력은 본 연구에서 산출한 순열 검증의 검증력과 부트스트랩 검증의 검증력 사이에 있을 것이다.

본 연구의 결과는 P300 CIT에서 부트스트랩 검증과 순열 검증의 1 종 오류율 및 검증력이 비슷하게 나타난 Rosenfeld와 Donchin(2015)의 연구결과와 불일치하는 것이다. 이러한 불일치가 발생한 이유는 Rosenfeld와 Donchin(2015)의 몬테 카를로 연구에서 관련자극과 무관련자극의 시행수를 모두 30으로 동일하게 설정하였기 때문이다. 즉, Rosenfeld와 Donchin(2015)의 연구에서는 시행수 차이로 인한 고주파 노이즈의 문제가 없는 조건에서 몬테 카를로 연구를 수행하였으므로, 부트스트랩 검증과 순열 검증의 1 종 오류율 및 검증력이 비슷하게 나타날 수밖에 없었다.

몬테 카를로 연구에서는 일반적으로 무한한 크기의 모집단을 가정하며, 본 연구에서도 무한 모집단을 가정하는 것이 타당할 것이다. 그러나 본 연구에서는 현실적인 문제로 유한한 크기의 모집단을 구성하여 부트스트랩 검증과 순열 검증의 1 종 오류율 및 검증력을 산출하였다. 통계적 검증의 근간이 되는 표집분포는 모집단으로부터 표본을 복원추출하는 경우에 무한 모집단과 유한 모집단에서 동일하게 정의되는데(Dowdy, Wearden, & Chilko, 2011), 본 연구에서도 유한 모집단으로부터 표본을 복원추출하였으므로 본 연구의 결과는 무한 모집단에서도 동일하게 유지될 것이다. 더욱이 15명의 자료를 포함한 모집단과 10명의 자료를 포함한 모집단에서 매우 비슷한 형태의 결과가 나타났으므로, 이러한 추론이

경험적으로도 지지되는 것으로 생각된다.

ERP의 형태는 개인차이가 상당히 크며(Luck, 2014), 특히 인지처리 시간을 반영하는 P300 잠재시간도 개인차이가 나타날 수밖에 없다(Polich, 2007). 본 연구에서는 개인차이가 존재하는 여러 명의 뇌파들을 합하여 모집단을 구성하였다. 즉, 이질적인 자료들을 합하여 모집단을 구성하였다. 그러나 이러한 절차가 연구결과를 왜곡시키지는 않았을 것이다. P300 잠재시간은 개인 내에서도 매 시행마다 달라지므로(Luck, 2014) 개인 간 잠재시간의 평균이 서로 다른 것도 같은 맥락으로 이해한다면, 개인 간 P300 잠재시간의 이질성이 연구결과에 영향을 미치지 않았을 것으로 추론된다. 또한 본 연구에서 모집단 크기와 더불어 P300 잠재시간의 이질성 정도를 조작한 결과에서도 두 모집단 간에 의미 있는 차이가 없었다.

연구설계에서 모집단 크기와 P300 잠재시간의 이질성이 공변하였기 때문에 개별 효과를 확인하기 어렵다고 평가할 수도 있다. 그러나 두 모집단(모집단 크기와 이질성이 큰 모집단과 모집단 크기와 이질성이 작은 모집단)에서 결과가 매우 유사하였기 때문에 모집단 크기와 P300 잠재시간의 이질성이 결과에 영향을 미치지 않았을 가능성이 높다.

결과적으로, P300 CIT에서 부트스트랩 검증은 '시행수 차이로 인한 고주파 노이즈의 문제'에 영향을 받아 1 종 오류율이 높은 반면에 순열 검증은 이 문제에 영향을 받지 않았다. 따라서 P300 CIT에서는 부트스트랩 검증 대신에 순열 검증을 사용하는 것이 더 좋을 것이다. 이러한 결과는 P300 진폭뿐만 아니라 사건관련전위의 다른 성분을 비교할 때에도 활용될 수 있을 것이다.

본 연구는 실험을 통하여 측정된 단일 시행 뇌파를 이용하여 몬테 카를로 연구에 사용할 모집단을 구성하였다. 이러한 연구방법은 본 연구자들이 고안한 것으로 추후 반복 연구를 통하여 이러한 연구방법의 타당성이 확립될 필요가 있다. 그러나 한편으로 ERP에 대한 몬테 카를로 연구 방법을 제시하였다는 가치를 가진다.

앞서 논의하였듯이 P300 CIT에서 과대추정되지 않은 최고의 검증력은 본 연구에서 산출한 순열 검증의 검증력과 부트스트랩 검증의 검증력 사이에 있을 것이다. 따라서 추후 연구에서는 과대추정되지 않은 최고의 검증력을 가지는 통계적 방법 또는 조건을 탐색할 필요성이 있다.

관련자극과 무관련자극의 시행수를 늘리면 신호 대 잡음의 비율이 \sqrt{N} 에 비례하여 증가하므로 시행수 차이로 인한 고주파 배경뇌파의 문제가 감소할 것이다. 그러나 표준오차(standard error)는 \sqrt{N} 에 반비례하므로 시행수가 늘어나면

검증력이 증가하게 된다. 결과적으로 관련자극과 무관련자극의 시행수를 늘려도 시행수 차이로 인한 고주파 배경뇌파의 문제가 여전히 남아 있을 것으로 예상된다. 추후 연구에서 이러한 예측에 대한 경험적 확인이 필요하다.

Rosenfeld 등(2008)은 원래의 부트스트랩 방법과 약간 다른 변형된 부트스트랩 방법을 사용하기 시작하였으며, 이후로 Rosenfeld와 동료들은 대부분의 P300 CIT연구에서 변형된 부트스트랩 방법을 사용하고 있다. 그러나 통계적 방법으로써 부트스트랩 방법과 순열 검증의 1종 오류율과 검증력은 여러 연구들에서 확인된 반면 변형된 부트스트랩 방법의 통계적 타당성이 검증된 적이 없다. 이러한 이유로 변형된 부트스트랩 방법을 본 논문의 내용에 포함시키지 않았다. 추후 연구에서는 변형된 부트스트랩 방법의 타당성과 P300 CIT에서 이 방법의 정확성이 확인될 필요성이 있다.

References

- Aboutalebi, 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*, 309-320.
- Ambach, W., Bursch, S., Stark, R., & Vaitl, D. (2010). A concealed information test with multimodal measurement. *International Journal of Psychophysiology*, *75*, 258-267.
- Arasteh, A., Moradi, M. H., & Janghorbani, A. (2016). A novel method based on empirical mode decomposition for P300-based detection of deception. *IEEE Transactions on Information Forensics and Security*, *11*, 2584-2593.
- Bergström, Z. M., Anderson, M. C., Buda, M., Simons, J. S., & Richardson-Klavehn, A. (2013). Intentional retrieval suppression can conceal guilty knowledge in ERP memory detection tests. *Biological psychology*, *94*, 1-11.
- Bowman, H., Filetti, M., Alsufyani, A., Janssen, D., & Su, L. (2014). Countering countermeasures: Detecting identity lies by detecting conscious breakthrough. *PloS one*, *9*(3), e90595.
- Cutmore, T. R. H., Djakovic, T., Keibell, M. R., & Shum, D. H. K. (2009). An object cue is more effective than a word in ERP-based detection of deception. *International Journal of Psychophysiology*, *71*, 185-192.
- Dowdy, S., Wearden, S., & Chilko, D. (2011). *Statistics for research (Vol. 512)*. Hoboken, NJ: John Wiley & Sons.
- Efron, B., & Tibshirani, R. J. (1994). *An introduction to the bootstrap*. New York: CRC press.
- Ellwanger, J., Rosenfeld, J. P., Sweet, J. J., & Bhatt, M. (1996). Detecting simulated amnesia for autobiographical and recently learned information using the P300 event-related potential. *International Journal of Psychophysiology*, *23*, 9-23.
- Farwell, L. A. & Donchin, E. (1991). The truth will out: Interrogative polygraphy ("lie detection") with event-related Potentials. *Psychophysiology*, *28*, 531-547.
- Good, P. I. (2005). *Permutation, parametric, and bootstrap tests of hypotheses*. New York: Springer.
- Janssen, A., & Pauls, T. (2005). A Monte Carlo comparison of studentized bootstrap and permutation tests for heteroscedastic two-sample problems. *Computational Statistics*, *20*, 369-383.
- Johnson, R. (1993). On the neural generators of the P300 component of the event-related potential. *Psychophysiology*, *30*, 90-97.
- Konietschke, F., & Pauly, M. (2014). Bootstrapping and permuting paired t-test type statistics. *Statistics and Computing*, *24*, 283-296.
- Kubo, K., & Nittono, H. (2009). The role of intention to conceal in the P300-based concealed information test. *Applied Psychophysiology and Biofeedback*, *34*, 227-235.
- Lefebvre, C. D., Marchand, Y., Smith, S. M., & Connolly, J. F. (2009). Use of event-related brain potentials (ERPs) to assess eyewitness accuracy and deception. *International Journal of Psychophysiology*, *73*, 218-225.
- Luck, S. J. (2014). *An introduction to the event-related potential technique*. Cambridge, MA: MIT Press.
- Lukács, G., Grządziel, A., Kempkes, M., & Ansorge, U. (2019). Item roles explored in a modified P300-Based CTP concealed information test. *Applied Psychophysiology and Biofeedback*, *44*, 195-209.
- Lukács, G., Weiss, B., Dalos, V. D., Kilencz, T., Tudja, S., & Csifcsák, G. (2016). The first independent study on the complex trial protocol version of the P300-based concealed information test: Corroboration of previous findings and highlights on vulnerabilities. *International Journal of Psychophysiology*, *110*, 56-65.
- Marchand, Y., Inglis-Assaff, P. C., & Lefebvre, C. D. (2013). Impact of stimulus similarity between the probe and the irrelevant items during a card-playing deception detection task: The "irrelevants" are not irrelevant. *Journal of Clinical*

- and Experimental Neuropsychology*, 35, 686-701.
- Matsuda, I., Nittono, H., & Ogawa, T. (2011). Event related potentials increase the discrimination performance of the autonomic based concealed information test. *Psychophysiology*, 48, 1701-1710.
- Meijer, E. H., Selle, N. K., Elber, L., & Ben-Shakhar, G. (2014). Memory detection with the concealed information test: A meta analysis of skin conductance, respiration, heart rate, and P300 data. *Psychophysiology*, 51, 879-904.
- Meijer, E. H., Smulders, F. T., Merckelbach, H. L., & Wolf, A. G. (2007). The P300 is sensitive to concealed face recognition. *International Journal of Psychophysiology*, 66, 231-237.
- Mertens, R. & Allen, J. J. (2008). The role of psychophysiology in forensic assessments: Deception detection, ERPs, and virtual reality mock crime scenarios. *Psychophysiology*, 45, 286-298.
- Polich, J. (2007). Updating P300: An integrative theory of P3a and P3b. *Clinical Neurophysiology*, 118, 2128-2148.
- Robinson, J. (1987). Nonparametric confidence intervals in regression: The bootstrap and randomization methods. In M. L. Puri, J. P. Vilaplana, & W. Wertz (Eds.), *New perspectives in theoretical and applied statistics* (pp. 243-255). New York: Wiley and Sons.
- Rosenfeld, J. P. (2005). Brain fingerprinting: A critical analysis. *The Scientific Review of Mental Health Practice*, 4, 20-37.
- Rosenfeld, J. P., Angell, A., Johnson, M., & Qian, J. H. (1991). An ERP based, control question lie detector analog: Algorithms for discriminating effects within individuals' average waveforms. *Psychophysiology*, 28, 319-335.
- Rosenfeld, J. P., Biroschak, J. R., & Furedy, J. J. (2006). P300-based detection of concealed autobiographical versus incidentally acquired information in target and non-target paradigms. *International Journal of Psychophysiology*, 60, 251-259.
- Rosenfeld, J. P., & Donchin, E. (2015). Resampling (bootstrapping) the mean: A definite do. *Psychophysiology*, 52, 969-972.
- Rosenfeld, J. P., Hu, X., & Pederson, K. (2012). Deception awareness improves P300-based deception detection in concealed information tests. *International Journal of Psychophysiology*, 86, 114-121.
- 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, 906-919.
- Rosenfeld, J. P., Reinhart, A. M., Bhatt, M., Ellwanger, J., Gora, K., Sekera, M., & Sweet, J. (1998). P300 correlates of simulated malingered amnesia in a matching-to-sample task: topographic analyses of deception versus truth-telling responses. *International Journal of Psychophysiology*, 28, 233-247.
- Rosenfeld, J. P., Shue, E., & Singer, E. (2007). Single versus multiple probe blocks of P300-based concealed information tests for autobiographical versus incidentally learned information. *Biological Psychology*, 74, 396-404.
- Rosenfeld, J. P., Soskins, M., Bosh, G., & Ryan, A. (2004). Simple effective countermeasures to P300-based tests of detection of concealed information. *Psychophysiology*, 41, 205-219.
- Rosenfeld, J. P., Sweet, J. J., Chuang, J., Ellwanger, J., & Song, L. (1996). Detection of simulated malingering using forced choice recognition enhanced with event-related potential recording. *The Clinical Neuropsychologist*, 10, 163-179.
- Rosenfeld, J. P., Ward, A., Meijer, E. H., & Yukhnenko, D. (2017). Bootstrapping the P300 in diagnostic psychophysiology: How many iterations are needed?. *Psychophysiology*, 54, 366-373.
- Semlitsch, H. V., Anderer, P., Schuster, P., & Presslich, O. (1986). A solution for reliable and valid reduction of ocular artifacts, applied to the P300 ERP. *Psychophysiology*, 23, 695-703.
- 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, 173-180.
- Verschuere, B., Rosenfeld, J. P., Winograd, M. R., Labkovsky, E., & Wiersema, R. (2009). The role of deception in P300 memory detection. *Legal and Criminological Psychology*, 14, 253-262.
- Wasserman, S., & Bockenholt, U. (1989). Bootstrapping: Applications to psychophysiology. *Psychophysiology*, 26, 208-221.
- Zoumpoulaki, A., Alsufyani, A., & Bowman, H. (2015). Resampling the peak, some dos and don'ts. *Psychophysiology*, 52, 444-448.

P300 숨긴정보검사에서 순열 검증의 정확도

전하정¹, 손진훈¹, 엄진섭²

¹충남대학교 심리학과, ²충북대학교 심리학과

본 연구의 목적은 P300 숨긴정보검사에서 저역통과필터나 100 ms 구간의 P300 평균전위를 사용하더라도 관련자극과 무관련 자극의 시행수 차이로 인하여 관련자극의 P300 진폭이 과대추정된다는 것을 밝히고, 부트스트랩 방법의 1 종 오류율은 유의수준보다 크게 나타나지만 순열 검증의 1 종 오류율은 적절하게 통제된다는 것을 증명하는 것이다. 몬테 카를로 연구를 이용하여 30 Hz, 10 Hz, 또는 5 Hz 저역통과필터를 사용하는 경우와 P300 정점 또는 100 ms 구간의 P300 평균전위를 사용하는 경우에 부트스트랩 방법과 순열 검증의 1 종 오류율과 통계적 검증력을 산출하였다. 연구 결과, 저역통과필터를 사용하거나 100 ms 구간의 P300 평균전위를 사용하더라도 관련자극의 P300 진폭이 과대추정되는 정도가 무관련자극의 P300 진폭이 과대추정되는 정도보다 약 1 μV 이상 더 컸다. 이로 인하여 부트스트랩 방법의 1 종 오류율은 유의수준보다 더 높게 나타났다. 그러나 순열 검증의 1 종 오류율은 유의수준과 동일하였다. 순열 검증의 검증력은 5 Hz의 저역통과필터를 사용하고 100 ms 구간의 P300 평균전위를 사용하는 경우에 가장 높았다. 본 연구결과는 P300 숨긴정보검사 뿐만 아니라 시행수가 서로 다른 자극의 P300 진폭을 비교하는 모든 경우에도 활용될 수 있을 것이다.

주제어: P300, 숨긴정보검사, 순열 검증, 1 종 오류, 검증력