

The false positive rate of P300-based concealed information test*

Jin-Sup Eom¹, Hajung Jeon², Jin-Hun Sohn^{2*}

¹Department of Psychology, Chungbuk National University,

²Department of Psychology, Chungnam National University

P300 concealed information test (P300 CIT) assumes that all stimuli except the target stimuli show the same P300 amplitudes in innocent participants. However, according to the P300 theory, the assumption of the equality of stimuli can be easily violated, thus false positive rate of P300 CIT can be increased. A series of studies were conducted to confirm this possibility. In study 1, we identified how much the assumption of the equality of stimuli was violated in single-probe protocol. This assumption was invalidated in about 30% of the participants in the study using a mock crime, but it was almost met in the study using autobiographical information. In study 2, we conducted an experimental study to evaluate the assumption of equality of stimuli in multiple-probe protocol. It turned out that the assumption was not met in about 25% of participants. In study 3, a Monte Carlo study was conducted to estimate the false positive rates by the level of non-equality of stimuli (0, 1, 2, 3, 4, and 5), the number of probes (1, 3, and 6), the number of trials (30, 48, and 66), and the significance levels (.10 and .05). When the level of non-equality of stimuli was set to 0, the false positive rates were properly controlled regardless of the number of probes, the number of trials, and the significance levels. However, the false positive rates was increased in all conditions when the level of non-equality of stimuli was 1 or higher. The false-positive rates got higher when the level of non-equality of stimuli got higher, the number of probes got smaller, and the number of trials got higher. When the level of non-equality of stimuli was 2, the false positive rates ranged from 12% to 25% at the significance level .10. While the level of non-equality of stimuli was 5, it fell within the ranged of 18% to 38%. In the discussion, we proposed methods to control false positive rates in the presence of non-equality of stimuli.

Keywords: P300, concealed information test, guilty knowledge test, lie detection, false positive rate

1 차원고접수 18.06.25; 수정원고접수 18.07.12; 최종게재결정 18.07.19

범죄와 관련된 구체적 사실들이 일반인에게 공개되지 않은 경우에 범죄 용의자에게 거짓말 탐지의 한 방법인 숨긴정보 검사(concealed information test: CIT)¹⁾를 실시할 수 있다. CIT에서는 범죄와 직접 관련된 사실인 관련자극과 범죄와 직접 관련이 없는 다수의 무관련자극을 섞은 후 범죄 용의자에게 하나씩 제시하면서 생리적 반응을 관찰한다. CIT는 정향반응(orienting response)과 습관화(habituation)라는 견고한 이론적 근거를 가지는 것으로 평가되고 있다

(Ben-Shakhar & Elaad, 2002). 용의자에게 첫 번째 무관련 자극을 제시하면 정향반응에 의해 생리적 반응이 크게 나타나지만 무관련자극들을 계속 제시하면 습관화에 의해서 생리적 반응이 점차 감소한다. 이러한 과정 중에 관련자극을 제시하면 유죄인 용의자와 무죄인 용의자의 반응이 서로 달라진다. 유죄인 용의자는 관련자극을 알아볼 것이므로 정향반응이 다시 크게 나타나지만, 무죄인 용의자는 관련자극과 무관련자극이 동일한 가치를 가지므로 습관화 현상이 지속될

* 이 논문은 2017년도 충남대학교 학술연구지원사업의 연구비 지원을 받아 수행된 연구임(2017-2010-01).

† 교신저자: 손진훈, 충남대학교 심리학과, (34134) 대전광역시 유성구 대학로 99, E-mail: jhsohn@cnu.ac.kr

1) 유죄지식검사(guilty knowledge test: GKT)라고도 한다.

것이다. 이러한 CIT는 특히 무죄인 용의자를 유죄로 판단하는 오류 긍정률(false positive rate)이 낮은 것으로 알려져 있지만, 상황에 따라서 유죄인 용의자를 유죄로 옳게 판단하는 민감도(sensitivity)가 낮을 수 있다는 단점이 있다 (Verschuere, Ben-Shakhar, & Meijer, 2011).

최근에 생리적 반응을 측정하는 CIT의 대안으로 사건관련 전위(event-related potential: ERP)의 P300 성분을 사용하는 CIT(즉 P300 CIT)에 대한 연구가 활발히 수행되고 있다 (Deng, Rosenfeld, Ward, & Labkovsky, 2016; Osugi & Ohira, 2017). 일반적으로 P300 CIT에서는 관련자극과 무관련자극에 목표자극(피검사자가 검사자극에 주의를 집중하도록 하기 위한 용도를 가진)을 섞어 피검사자에게 하나씩 제시하면서 목표자극과 나머지 자극을 변별하여 반응하도록 지시한다.

P300 CIT를 고안한 Farwell과 Donchin(1991)은 이 검사에 대한 논리적 근거를 P300의 일반적 특성으로 설명하고 있다. 바로 두 범주의 자극이 반복해서 제시될 때 출현빈도가 더 낮은 범주의 자극에서 P300 진폭이 더 크게 나타난다는 점이다. 유죄인 피검사자와 무죄인 피검사자 모두 목표자극을 나머지 자극으로부터 구별해서 반응해야 하므로 목표자극을 한 범주로, 나머지 자극을 다른 한 범주로 구분하게 된다. 일반적으로 목표자극과 관련자극, 무관련자극의 비율을 1:1:4로 구성하므로, 목표자극의 출현빈도(약 17%)가 나머지 자극의 출현빈도(약 83%)보다 낮고 따라서 목표자극에서 P300 진폭이 크게 나타날 것이다. 유죄인 피검사자에게 관련자극은 개인적으로 중요한 의미를 가지는 반면 무관련자극은 그렇지 못하므로 두 종류의 자극은 서로 다른 범주로 구별될 것이며 출현빈도가 더 낮은 관련자극(약 17%)의 P300 진폭이 출현빈도가 높은 무관련자극(약 67%)의 P300 진폭

보다 더 크게 나타날 것이다. 반면에 무죄인 피검사자에게 관련자극과 무관련자극은 모두 같은 범주의 자극일 뿐이므로 두 자극의 P300 진폭이 다르지 않을 것이다. 따라서 어떤 피검사자의 P300 CIT 결과에서 관련자극에 대한 P300 진폭이 무관련자극에 대한 P300 진폭보다 더 크면 관련자극을 알고 있는 것으로 즉 유죄로 판단하며, 관련자극과 무관련자극에 대한 P300 진폭이 다르지 않다면 관련자극을 알지 못하는 것으로 즉 무죄로 판단한다.

P300 CIT의 논리적 근거에서 가장 중요한 가정은 피검사자가 무죄인 경우에 관련자극과 무관련자극들이 동일한 크기의 P300 진폭을 유발한다는 자극의 동질성 가정이다. 1개의 관련자극과 4개의 무관련자극을 사용하는 조건에서 자극의 동질성 가정이 충족되는 경우와 파기되는 경우에 대한 가상의 예가 Figure 1에 제시되어 있다. 자극의 동질성 가정은 Figure 1의 (A)처럼 5개 자극의 P300 진폭이 모두 비슷할 것이라는 가정이다. 정확하게 표현하면 5개 자극의 P300 진폭은 동일하지만 표본오차 때문에 약간은 달라질 수 있다. 이러한 경우에는 5개 자극 중 어떤 것이 관련자극이라 하더라도 관련자극과 나머지 무관련자극 간에 P300 진폭은 통계적으로 유의하지 않을 것이다. 그러나 자극의 동질성 가정이 깨어지면 무죄인 용의자를 유죄로 잘못 판단하는 오류 긍정률이 증가할 것이다. 즉 Figure 1의 (B) 또는 (C)와 같이 무죄인 용의자에게 5개의 자극들이 서로 다른 크기의 P300 진폭을 유발한다면 관련자극의 P300 진폭이 무관련자극의 P300 진폭보다 통계적으로 유의하게 클 확률이 증가한다. 5개의 자극 중 한 자극을 구성하는 관련자극이 나머지 자극들보다 가장 클 확률은 20%이고, 첫 번째 또는 두 번째로 클 확률은 40%나 된다. 관련자극이 첫 번째나 두 번째로 크면 무관련자극들 전체의 평균보다 크게 되며, 비동질성 수준이

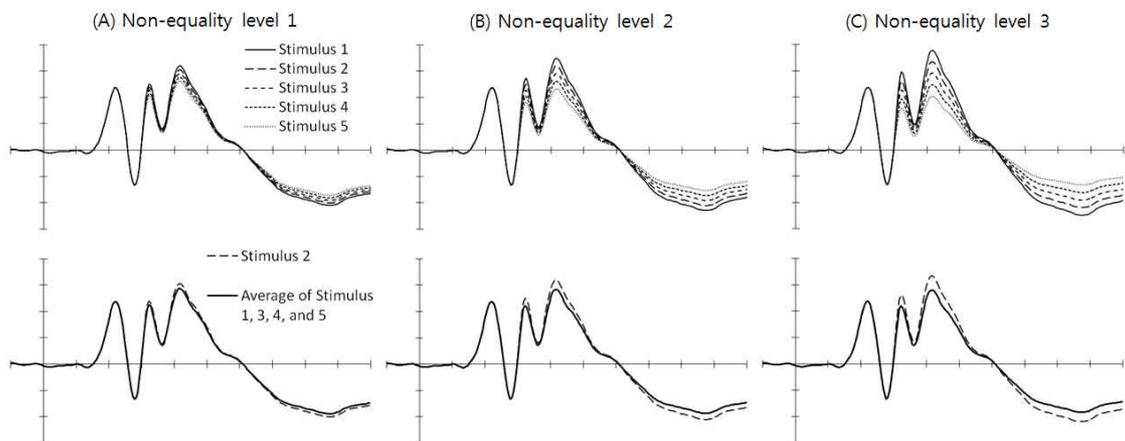


Figure 1. The hypothetical example of non-equality of stimuli. (A) low level of non-equality of stimuli. (B) moderate level of non-equality of stimuli. (C) high level of non-equality of stimuli.

증가함에 따라서 이 차이가 통계적으로 유의할 가능성도 함께 증가하게 될 것이다. 이론적으로 비동질성 수준이 아주 크다면 오류 긍정률은 40%에 근접할 것이다. 법률적 판단에서 무죄인 용의자를 유죄로 잘못 판단하는 것은 심각한 오류이므로 P300 CIT의 기본 가정을 검토하는 일은 중요하다고 할 수 있다. 그럼에도 불구하고 아직까지 이러한 가정, 즉 자극들의 동질성 가정을 체계적으로 평가한 연구를 찾아볼 수 없다.

우선 이론적 수준에서 자극들 간의 P300 진폭이 서로 동질적인지 평가하기 위하여 Johnson(1993)의 삼요인 모형을 살펴보았다. P300의 진폭에 영향을 미치는 요인들을 정리한 삼요인 모형에 의하면, P300 진폭은 주관적 확률(subjective probability)과 자극의 의미(stimulus meaning), 피실험자에게 전달된 정보량(information transmission)의 함수로 결정된다. 이 모형은 다시 수식 (1)처럼 세분화된다.

$$P300 \text{ 진폭} = f(T \times ((1/GP + 1/SE) + (TC + SC + SV))) \quad (1)$$

T: 피실험자에게 전달된 정보량, GP: 전반적 확률,
SE: 순서적 기대, TC: 과제의 복잡성,
SC: 자극의 복잡성, SV: 자극의 가치

피실험자에게 전달된 정보량(T)은 자극으로부터 피실험자가 받아들인 정보의 양을 의미한다. 주관적 확률은 앞으로 나타날 자극에 대한 불확실성을 반영하는 것으로 전반적 확률(global probability: GP)과 순서적 기대(sequential expectancies: SE)에 의해 결정된다. 전반적 확률은 각 자극의 출현빈도를 의미하며 순서적 기대는 앞으로 어떤 자극이 나올지에 대한 주관적 기대를 말한다. 자극의 의미는 자극의 정보처리에 관한 것으로 과제의 복잡성(task complexity: TC)과 자극의 복잡성(stimulus complexity: SC), 자극의 가치(stimulus value: SV)에 의해서 결정된다. 과제의 복잡성은 피검사가 수행해야 할 과제가 얼마나 복잡한지를 의미하는 것으로, 예를 들면 자극에 반응을 하지 않는 것보다 버튼 반응을 하는 것이 P300 진폭을 더 크게 만든다. 자극의 복잡성은 자극이 얼마나 복잡한지에 대한 것으로, 예를 들면 단순한 청각 자극보다는 시각자극에서 P300 진폭이 더 크게 나타난다. 자극의 가치는 자극의 중요성을 의미하는 것으로 개인적인 중요성이 클수록 자극에 대한 인지적 처리가 증가하고 더 많은 주의자원이 할당되므로 P300 진폭이 증가한다.

P300 CIT는 Johnson(1993)의 삼요인모형에서 주관적 확률의 요소 중 하나인 전반적 확률(GP)에 이론적 근거를 두

고 있다고 볼 수 있다. 무죄인 용의자에게 P300 진폭에 영향을 미치는 나머지 요소들 중 순서적 기대(SE)와 과제의 복잡성(TC), 자극의 복잡성(SC), 전달된 정보량(T)은 다수의 자극들(즉, 관련자극과 무관련자극들) 간에 동일할 것이다. 그러나 자극의 가치(SV)는 무죄인 용의자의 개인적 경험에 따라 다수의 자극들 간에 다를 수 있으며, 이로 인하여 자극들 간에 P300 진폭이 다르게 나타날 수 있다. 본 연구에서는 이 문제를 ‘자극의 비동질성 문제’로 부르기로 한다. 예를 들어, 범죠행장에서 범행도구로 보이는 부엌칼이 발견되었다면 관련자극 1개와 목표자극 1개, 무관련자극 4개를 사용하는 단일관련자극 프로토콜을 사용하여 P300 CIT를 실시할 수 있다. 즉 부엌칼을 관련자극으로 하고 과도를 목표자극으로, 회칼을 무관련자극 1로, 잭나이프를 무관련자극 2로, 중국식칼을 무관련자극 3으로, 단검을 무관련자극 4로 구성하여 용의자에게 P300 CIT를 실시할 수 있다. 이때 이 용의자가 진실로는 무죄이지만 과거에 식당을 운영한 경험이 있다면, 부엌칼과 중국식칼, 회칼에 특별한 의미가 있을 수 있고 이로 인해 자극들 간에 P300 진폭이 크게 달라질 수 있다.

본 연구에서는 자극의 비동질성 문제가 P300 CIT의 오류 긍정률에 어떠한 영향을 미치는지 파악하였다. P300 CIT를 구성하는 세부사항들은 고정된 것이 아니라 P300 CIT를 사용하는 사람과 상황에 의해서 다양하게 설정되므로, 본 연구에서도 다양한 P300 CIT의 특성이 오류 긍정률에 미치는 영향을 평가하였다. 구체적으로 세 가지 세부사항에 따라 오류 긍정률을 평가하였다. 첫 번째는 P300 CIT에 사용하는 관련자극의 개수이다. P300 CIT는 검사에 사용되는 관련자극의 개수에 따라서 단일관련자극 프로토콜과 중다관련자극 프로토콜로 나눌 수 있다(Rosenfeld, Shue, & Singer, 2007). 단일관련자극 프로토콜에서는 관련자극 1개와 목표자극 1개, 무관련자극 4개를 섞어서 각 자극의 제시횟수가 적어도 30회 이상이 되도록 반복하여 제시한다. 중다관련자극 프로토콜에서는 관련자극이 4개면 목표자극을 4개, 무관련자극을 16개 만들어 사용한다. 이 경우에는 관련자극 전체의 제시횟수가 적어도 30회 이상이 되도록 반복하여 제시한다. 두 번째는 시행 수이다. P300 CIT에서는 일반적으로 관련자극의 시행 수가 30회 이상이 되도록 실시하는데(Rosenfeld, Biroshak, & Furedy, 2006), 연구자들에 따라서 70회 이상 실시하기도 한다(Farwell & Donchin, 1991). 세 번째는 유의수준이다. 유의수준으로 .10을 사용하는 연구자들도 있으며(Rosenfeld et al., 2007), .05를 사용하는 연구자들도 있다(Lefebvre, Marchand, Smith, & Coonolly, 2009).

본 연구의 목적은 P300 CIT에서 자극의 비동질성 문제가

어느 정도 나타나는지, 그리고 이러한 자극의 비동질성 문제가 존재할 때 무죄인 조사대상자를 유죄로 잘못 판단하는 오류 긍정률이 어느 정도 나타나는지 확인하는 것이다. P300 CIT에서 자극의 비동질성 문제가 존재한다면 각 조사대상자 내에서 자극들 간의 P300 진폭이 통계적으로 유의하게 다를 것이다. 또한 자극의 비동질성 문제가 존재한다면 오류 긍정률은 증가할 것이다. 연구 1과 연구 2에서는 각각 실험자료를 이용하여 관련자극을 1개 사용하는 단일관련자극 프로토콜과 관련자극을 5개 사용하는 중다관련자극 프로토콜에서 자극의 비동질성 문제가 어느 정도 나타나는지 확인하였다. 연구 3에서는 자극의 비동질성 수준에 따라서 오류 긍정률이 나타나는 정도를 확인하였다. 이때, 관련자극의 개수와 전체 시행 수, 유의수준을 함께 고려하여 오류 긍정률이 어떻게 달라지는지 확인하였다.

연구 1

연구 1의 목적은 관련자극을 1개만 사용하는 단일관련자극 프로토콜에서 자극의 비동질성 문제가 발생하는지 또한 자극의 비동질성 문제가 발생한다면 그 수준은 어느 정도인지 평가하는 것이다. 단일관련자극 프로토콜에서 자극의 비동질성 문제가 존재한다면 각 실험참가자 내에서 자극들 간의 P300 진폭이 유의하게 다를 것이다. 또한 자극의 비동질성 수준은 자극들 간의 변량추정치를 이용하여 평가할 수 있다.

연구 1에서는 기존에 발표된 실험연구의 자료를 이용하였다. 연구 1의 목적이 단일관련자극 프로토콜에서 자극의 비동질성 문제가 존재하는지 확인하는 것이므로, 무죄조건에서 1개의 관련자극과 4개의 무관련자극들이 동질적인 P300 진폭을 유발하는지 검토해야 할 것이다. 그러나 무죄인 조사대상자에게 관련자극과 무관련자극은 서로 구별되지 않으므로 모두 무관련자극이라고 생각하여도 무방할 것이며, 따라서 원래의 무관련자극 4개만 이용하여 자극의 비동질성 문제를 파악하여도 큰 문제는 발생하지 않을 것으로 추론할 수 있다. 또한 4개의 무관련자극들은 무죄조건과 유죄조건에서 다르게 기능할 이유가 없으므로, 유죄조건에서 4개의 무관련자극들을 이용하여 자극의 비동질성 문제를 파악하는 것도 가능할 것이다. P300 CIT의 특성상 유죄조건만 포함하는 실험연구가 많으며(Dietrich, Hu, & Rosenfeld, 2014; Verschuere, Rosenfeld, Winograd, Labkovsky, & Wiersema, 2009), 본 연구자들이 수집할 수 있는 실험연구의 자료도 유죄조건만 포함하고 있었다. 이에 따라 연구 1에서는 선행연구에서 유죄조건에 포함된 4개의 무관련자극들에

대한 뇌파를 이용하여 자극의 비동질성 문제가 존재하는지, 그리고 비동질성의 정도는 어떠한지 평가하였다.

P300 CIT에 대한 실험연구들은 유죄조건을 구성할 때 두 가지 방법을 많이 이용한다. 한 가지는 가상의 범위를 수행하도록 한 후에 P300 CIT를 실시하는 것이며(Osugi & Ohira, 2017; Rosenfeld et al., 2007), 다른 한 가지는 거짓으로 기억상실을 주장하는 상황에 자기참조적인 정보에 대한 P300 CIT를 실시하는 것이다(Deng et al., 2016; Rosenfeld et al., 2006). 따라서 본 연구에서도 두 가지 실험자료를 분석하였다. 모의범위를 이용한 실험자료는 Eom과 Park(2014)의 연구에서 가져왔으며, 기억상실을 이용한 실험자료는 Lee와 Hwang, Park, Sohn, Eom(2013)이 수행한 연구에서 가져왔다.

방 법

실험참가자

모의범죄 연구(Eom & Park, 2014)에서는 대학생 24명이 실험에 자원하여 참가하였다. 성별은 남녀 각각 12명씩이었으며 평균 연령은 23.8세(범위 18~40세)였다. 기억상실 연구(Lee et al., 2013)에서는 대학생 26명이 실험에 자원하여 참가하였다. 성별은 남자가 14명, 여자가 12명이었다. 평균 연령은 23.9세(범위 20~28세)였다. 모든 실험참가자들은 실험이 끝난 후에 금전적인 보상을 받았다.

실험자극

모의범죄 연구의 P300 CIT에서 검사자극은 총 6가지로 반지, 목걸이, 팔찌, 시계, 귀걸이 및 브로치였다. 목표자극은 팔찌였으며, 관련자극은 모의범죄에서 실험참가자가 훔친 물건(반지 또는 목걸이)이었다. 실험참가자의 받은 검사자극을 사진으로 제시받았으며, 나머지 받은 문자로 제시받았다. 기억상실 연구의 P300 CIT에서도 검사자극은 6가지로 실험참가자의 이름을 관련자극으로 사용하였으며, 실험참가자에게 특별한 의미를 가지지 않는 이름 4개를 무관련자극으로, 또 다른 이름 하나를 목표자극으로 사용하였다.

실험절차

모든 실험참가자들은 실험에 대한 전반적인 사항이 적혀있는 실험참가동의서를 읽고 서명한 후에 실험에 참가하였다. 모의범죄 연구에서 실험참가자들은 P300 CIT를 실시하기 전에 모의범위를 수행하였다. 모의범죄는 아무도 없는 실험실에 들어가서 책상서랍에 있는 반지 또는 목걸이를 훔쳐 나오는

것이였다. 책상서랍에는 반지와 목걸이 중 하나가 들어 있었으며 실험참가자의 반은 책상서랍에서 반지를 훔쳤으며, 나머지 반은 목걸이를 훔쳤다. 모의범죄를 수행한 후 실험참가자가 돌아오면 P300 CIT를 실시하였다. 실험참가자에게 목표자극이 무엇인지 알려준 후에, 목표자극이 제시되면 ‘예’ 버튼(마우스 왼쪽버튼)을 누르도록 하였고, 목표자극 이외의 자극이 제시되면 ‘아니오’ 버튼(마우스 오른쪽 버튼)을 누르도록 하였다. ‘예’ 버튼과 ‘아니오’ 버튼은 실험참가자별로 균형화하였다. 실험참가자들 중 반은 검사자극을 사진으로 제시받았으며, 나머지 반은 문자로 제시받았다. 목표자극과 관련자극, 무관련자극을 무선적인 순서로 하나씩 제시하였으며, 자극은 실험참가자 앞에 놓인 모니터 중앙에 제시하였다. 사진자극은 흑백이였으며, 수평시각도와 수직시각도는 모두 2.01° 였다. 문자자극의 수평시각도는 2.58° 였으며 수직시각도는 0.86° 였다. 자극은 3초에 하나씩 제시하였으며, 사진자극은 500ms 동안 제시하였고, 문자자극은 300ms 동안 제시하였다. 한 블록에 각 자극 당 30번씩 총 180회 자극을 제시하였으며, 네 블록 동안 총 720회 자극을 제시하였다. 자극의 제시와 반응의 기록은 SuperLab 4.0을 이용하였다.

기억상실 연구에서 자신의 이름이 기억나지 않는다고 거짓으로 주장하는 상황에서 기억상실이 거짓인지를 평가하는 P300 CIT를 실시하였다. 실험참가자가 실험실에 도착하면 미리 준비한 무관련자극 이름 4개를 실험참가자에게 보여 주어 특별한 의미를 가진 이름이 있는지 물어본 후 특별한 의미를 가진 이름이 있었던 경우에는 다른 이름으로 바꾸었다. 실험참가자에게 목표자극이 무엇인지 알려준 후에 목표자극이 제시되면 가능한 빨리 마우스 왼쪽 버튼을 누르도록 지시하였다. P300 CIT에서 6개의 자극을 섞어서 무선적으로 하나씩 제시하였으며, 자극은 실험참가자 앞에 놓여 있는 모니터에 흰색 바탕에 검은색 글씨로 제시하였다. 수평시각도는 2.36° 였으며, 수직시각도는 0.72° 였다. 자극은 300ms동안 제시하였고, 3초에 하나씩 제시하였다. 자극당 40번씩 총 240회 자극을 제시하였다. 자극의 제시와 반응의 기록은 SuperLab 4.5를 이용하였다.

뇌파기록

모의범죄 연구와 기억상실 연구에서 뇌파는 Grass Model 12 Neurodata acquisition system(Grass Instruments, Quincy, MA, USA)을 이용하여 측정하였다. Ag/AgCl 전극을 국제 10-20 체계에 따라 Fz, Cz, 및 Pz에 부착하였으며, 양쪽 귓볼에 기준전극을, 이마에 접지전극을 부착하였다. 모의범죄 연구에서는 왼쪽 눈의 위와 아래에 전극을 부착하여

안전도를 측정하였으며, 기억상실 연구에서는 오른쪽 눈의 위와 아래에 그리고 양쪽 눈의 옆에 전극을 부착하여 안전도를 측정하였다. 뇌파는 0.3~30Hz 대역여과한 후 20000배 증폭하였으며, 안전도는 5000배 증폭하였다. 뇌파를 200Hz의 표집률로 A/D 변환하여 컴퓨터에 저장하였다.

분석

뇌파를 분석한 방법은 모의범죄 실험자료와 기억상실 실험자료에서 동일하였다. 모의범죄 실험자료는 자극을 그림으로 제시한 조건과 문자로 제시한 조건이 있었지만, 본 연구에서는 P300 CIT에서 전반적으로 오류 긍정률이 나타나는 정도를 파악하는 것이 목적이므로 실험조건별로 구분하여 분석하지는 않았다. P300 CIT에서는 대부분 P300 진폭이 가장 크게 발생하는 Pz 영역의 뇌파만 이용하므로, 본 연구에서도 Pz 영역에서 측정된 뇌파만 분석하였다. 먼저 Semlitsch와 Anderer, Schuster, Presslich(1986)가 제안한 방법으로 눈감 박임과 안구운동에 의한 뇌파의 변형을 교정하였다. 다음으로 각 자극의 제시 전 200ms부터 자극제시 후 1400ms까지 총 1600ms 구간으로 나누어 분석하였다. 분석구간에서 전위가 $\pm 75\mu V$ 를 넘는 시행은 분석에서 제외하였다. 모의범죄 실험자료에서 분석에서 제외된 시행의 비율은 0.2%였으며, 기억상실 연구에서는 0.5%였다. 본 연구는 무관련자극들이 서로 동질적인지 파악하는 것이 목적이므로, 4개의 무관련자극별로 평균하여 4개의 ERP를 산출한 후 P300 진폭을 계산하였다. ERP를 산출할 때에는 자극제시 전 200ms구간의 평균 전위를 이용하여 기저선을 교정하였다. P300 진폭은 P300 CIT에 가장 적합하다고 알려진 정점-정점 방법을 이용하여 계산하였다(Soskins, Rosenfeld, & Niendam, 2001). 정점-정점 방법은 P300의 전위와 P300 이후에 나타나는 부적 정점 전위값 간의 차이를 이용하는 방법이다. P300의 전위는 자극제시 후 300ms부터 700ms 사이에 100ms 구간의 평균 전위가 가장 큰 값을 이용하였다. 부적 정점은 P300 이후부터 1400ms까지 100ms 구간의 평균전위가 가장 작은 값을 이용하였다.

무관련자극들 간의 P300 진폭이 통계적으로 유의하게 다른지 검증하기 위하여 각 시행에서 P300 진폭을 계산하였다. 각 시행의 P300 진폭은 ERP에서 P300의 전위가 나타난 시점과 P300 이후 부적 정점이 나타난 시점을 각 시행의 뇌파에 적용하여 계산하였다. 각 실험참가자별로 자극들 간에 P300 진폭이 통계적으로 유의하게 다른지 검증하기 위하여 일원변량분석을 실시하였다. 다수의 일원변량분석을 실시하기 때문에 전체 연구의 1종 오류가 증가할 것이지만, 자극들

간의 비동질성 문제에 대한 판단은 1종 오류보다는 2종 오류가 더 위험한 상황이므로 Bonferroni 교정을 사용하지 않고 일반적인 유의수준 .05를 사용하였다.

자극의 비동질성 문제가 존재하는지 검증하는 것에 더불어, 자극의 비동질성 문제가 어느 정도 수준으로 나타나는지 파악하기 위하여 자극의 비동질성 수준을 수량화 하였다. 자극의 비동질성 수준을 수량화하는 이유는 연구 3에서 사용할 기초 자료를 얻기 위해서이다. 자극의 비동질성 수준은 자극들이 유발하는 P300 진폭이 서로 다른 정도를 의미하므로, P300 진폭이 서로 다른 정도를 수량화하면 된다. 통계학에서 측정치들이 서로 다른 정도는 변량(variance)으로 수량화하므로, 본 연구에서도 자극의 비동질성 수준을 변량으로 측정하였다. 즉, 자극 간 변량을 측정하였다. 표본자료는 무선오차를 포함하고 있기 때문에 P300 진폭으로 변량을 직접 계산하면 변량이 과대 추정된다. 오차변량이 제거된 변량을 추정하기 위하여 변량분석에서 산출되는 집단 간 평균자승과 집단 내 평균자승을 이용하였다. 일원변량분석에서 집단 간 평균자승의 기댓값은 $E(MSB) = \sigma_c^2 + n\sigma_A^2$ 로 오차변량(σ_c^2)과 집단 간 변량(σ_A^2)에 사례 수(n)를 곱한 값으로 정의되며, 집단 내 평균자승의 기댓값은 $E(MSE) = \sigma_c^2$ 로 오차변량(σ_c^2)만을 포함하고 있다(Park, 2003). 따라서 본 연구에서 산출하고자 하는 자극 간 변량 즉, 집단 간 변량(σ_A^2)을 수식 (2)와 같이 추정할 수 있다.

$$\sigma_A^2 = \frac{MSB - MSE}{n} \quad (2)$$

결 과

행동반응

실험참가자들이 검사자극에 적절한 주의를 기울였는지 평가하기 위하여 정확반응률과 반응시간을 산출하였다. 모의범죄

연구에서 목표자극과 무관전자극의 정확반응률 평균은 각각 88.4%(SD=12.4%)와 99.6%(SD=0.4%)였으며, 목표자극과 무관전자극의 반응시간 평균은 각각 547ms(SD=127ms)와 450ms(SD=126ms)였다. 기억상실 연구에서 목표자극의 정확반응률 평균은 99.5%(SD=1.2%)였으며, 목표자극에 대한 반응시간 평균은 592ms(SD=108ms)였다.

사건관련전위

모의범죄 실험자료와 기억상실 실험자료에 대한 Pz 영역의 ERP가 Figure 2에 제시되어 있다. 모의범죄 실험자료에서 목표자극에 대한 P300은 자극 제시 후 약 500ms에서 최대 값을 보였으며, 무관전자극에 대한 P300은 400~500ms에서 가장 큰 값을 보였다. 기억상실 실험자료에서도 목표자극에 대한 P300은 자극 제시 후 약 500ms에서, 무관전자극에 대한 P300은 400~500ms에서 가장 큰 값을 보였다.

모의범죄 실험자료에서 각 실험참가자별 무관전자극들에 대한 P300 진폭과 자극 간 변량 추정치, 일원변량분석결과가 Table 1에 제시되어 있다. 실험참가자별로 4개의 무관전자극에 대한 P300 진폭이 통계적으로 유의하게 다른지 일원변량분석을 실시한 결과, 전체 24명 중 7명에서 통계적으로 유의한 평균차이가 나타났다($F(3,474)=5.83, p=.001$; $F(3,476)=2.85, p=.037$; $F(3,476)=3.79, p=.010$; $F(3,472)=3.44, p=.017$; $F(3,476)=5.11, p=.002$; $F(3,476)=3.10, p=.026$; $F(3,475)=6.30, p<.001$). 즉, 7명은 자극의 비동질성 문제가 통계적으로 유의한 수준에서 나타났다. 비동질성 문제가 있었던 7명 중에서 P300 진폭이 가장 큰 자극을 확인한 결과, 무관전자극 1이 2명, 무관전자극 2가 1명, 무관전자극 3이 3명, 무관전자극 4가 1명이었다. 각 실험참가자의 자극 간 변량의 추정치는 -1.11~3.46의 범위를 보였다. 자극의 비동질성 문제가 통계적으로 유의하였던 경우에 자극 간 변량의 추정치는 1.21~3.46이었다. 전체 실험참가자의 평균 P300 진폭은 8.36 μ V였으며, 각 시행에 대한 P300 진폭의 표준편차

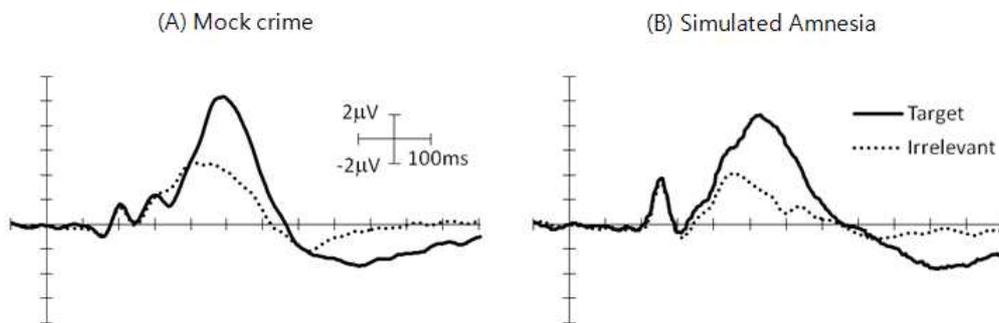


Figure 2. Grand averaged ERPs at Pz site in Study 1. (A) ERPs from the experiment using mock crime information. (B) ERPs from the experiment using self-referring information.

Table 1. P300 amplitude difference among stimuli in single probe protocol using mock crime

ID	Irrelevant 1	Irrelevant 2	Irrelevant 3	Irrelevant 4	MSB	MSW	$\hat{\sigma}_A^2$	<i>F</i>	<i>p</i>
1	10.05	10.46	9.57	7.65	185.84	86.00	0.83	2.16	.092
2	10.76	9.69	9.94	6.65	387.89	66.48	2.68	5.83	.001
3	13.85	14.36	14.84	13.83	27.48	160.62	-1.11	0.17	.916
4	5.35	4.55	6.49	3.25	222.98	78.20	1.21	2.85	.037
5	8.02	8.89	6.95	9.85	183.18	76.45	0.89	2.40	.068
6	7.66	9.46	7.41	7.44	116.20	67.64	0.40	1.72	.162
7	5.62	6.47	3.87	6.29	168.66	71.61	0.81	2.36	.071
8	10.13	10.05	10.85	8.32	139.02	82.04	0.47	1.69	.167
9	10.38	11.59	10.23	10.80	44.26	92.59	-0.40	0.48	.698
10	5.04	7.67	4.29	6.86	294.51	77.63	1.81	3.79	.010
11	11.76	11.84	10.98	13.53	138.14	94.52	0.36	1.46	.224
12	4.40	6.14	4.71	7.64	265.88	77.28	1.57	3.44	.017
13	8.11	4.76	8.24	7.52	318.70	123.18	1.63	2.59	.052
14	11.92	8.75	7.36	7.70	516.99	101.27	3.46	5.11	.002
15	7.85	8.29	9.17	6.78	118.95	60.29	0.49	1.97	.117
16	6.54	5.76	7.14	5.06	98.62	85.13	0.11	1.16	.325
17	20.23	18.10	20.71	20.52	175.88	131.31	0.37	1.34	.261
18	5.96	7.40	9.72	6.37	340.64	109.82	1.92	3.10	.026
19	3.18	3.65	2.68	1.27	127.11	64.15	0.52	1.98	.116
20	9.52	8.91	8.89	8.16	37.10	96.73	-0.50	0.38	.765
21	9.22	8.35	9.03	8.45	22.27	53.43	-0.26	0.42	.741
22	5.06	4.42	4.22	4.95	19.57	76.23	-0.47	0.26	.857
23	6.08	7.67	9.78	6.11	366.54	58.14	2.57	6.30	<.001
24	7.83	6.07	5.04	7.64	212.12	92.20	1.00	2.30	.076
Total		8.36				86.79			

Note. $\hat{\sigma}_A^2$ is the estimate of variance among stimuli

(즉, MSW의 제곱근)는 평균 9.32 μ V였다.

기억상실 실험자료에서 각 실험참가자별 무관련자극에 대한 P300 진폭과 자극 간 변량 추정치, 일원변량분석결과가 Table 2에 제시되어 있다. 실험참가자별로 4개의 무관련자극에 대한 P300 진폭이 통계적으로 유의하게 다른지 검증한 결과, 전체 26명 중 1명에서 통계적으로 유의한 평균차이가 나타났다($F(3,156)=3.65$, $p=.014$). 각 실험참가자의 자극 간 변량의 추정치는 -4.25~5.10의 범위를 보였다. 비동질성 문제가 통계적으로 유의하였던 경우에 자극 간 변량의 추정치는 5.10이었다. 전체 실험참가자의 평균 P300 진폭은 6.59 μ V였으며, 각 시행에 대한 P300 진폭의 표준편차(즉, MSW의 제곱근)는 평균 9.97 μ V였다.

논 의

연구 1에서는 단일관련자극 프로토콜을 사용한 P300 CIT에서 자극의 비동질성 문제가 나타나는지 확인하였다. 모의범죄 실험자료에서 목표자극에 대한 반응오류율은 약 10%내외로 선행연구들과 큰 차이가 없었다(Cutmore, Djakovic, Kebbell, & Shum, 2009; Rosenfeld Soskins, Bosh, & Ryan, 2004). 기억상실 실험자료에서 목표자극에 대한 반응오류율은 0.4%로 매우 낮았는데, 목표자극에만 반응을 하도록 하였기 때문인 것으로 생각된다. 목표자극에 대한 반응시간은 500~600ms 내외로 선행연구들과 유사하였다(Hu, Pornpattananangkul, & Rosenfeld, 2013). 이러한 결과는

Table 2. P300 amplitude difference among stimuli in single probe protocol using simulated amnesia

ID	Irrelevant 1	Irrelevant 2	Irrelevant 3	Irrelevant 4	MSB	MSW	$\hat{\sigma}_A^2$	F	p
1	10.34	9.95	15.03	11.10	217.17	141.00	1.90	1.54	.206
2	3.87	4.21	4.20	5.40	18.07	67.95	-1.25	0.27	.850
3	3.50	6.02	8.17	3.56	200.40	112.69	2.19	1.78	.154
4	8.66	7.15	5.50	6.48	70.57	93.75	-0.58	0.75	.522
5	4.73	5.64	4.70	5.45	9.49	136.83	-3.18	0.07	.976
6	5.71	4.16	4.47	3.94	25.06	92.25	-1.68	0.27	.846
7	3.75	6.95	2.97	5.60	129.84	61.33	1.71	2.12	.100
8	7.60	6.58	11.10	10.55	194.73	133.63	1.53	1.46	.228
9	6.23	9.83	4.90	9.25	224.67	88.05	3.42	2.55	.058
10	4.46	3.44	3.28	4.17	12.88	82.31	-1.74	0.16	.925
11	5.68	3.61	3.99	4.11	33.50	98.80	-1.63	0.34	.797
12	7.41	3.38	2.43	3.85	190.03	124.98	1.63	1.52	.211
13	7.60	7.24	6.36	8.90	44.44	106.44	-1.55	0.42	.741
14	11.65	11.33	8.60	10.46	75.10	67.27	0.20	1.12	.344
15	11.44	8.07	6.37	5.59	269.92	134.66	3.38	2.00	.116
16	6.42	10.26	9.24	8.00	109.79	113.49	-0.09	0.97	.410
17	5.20	7.71	4.68	5.20	74.19	244.24	-4.25	0.30	.823
18	12.55	8.21	7.82	7.00	247.28	108.76	3.46	2.27	.082
19	7.68	3.77	7.13	5.95	120.24	84.53	0.89	1.42	.238
20	8.96	11.72	12.56	12.13	105.57	85.24	0.51	1.24	.298
21	3.52	4.54	7.25	5.53	100.56	65.06	0.89	1.55	.205
22	8.03	2.22	3.72	6.60	280.80	76.90	5.10	3.65	.014
23	5.27	6.23	5.44	5.98	8.14	78.63	-1.76	0.10	.958
24	3.61	3.88	3.39	2.73	9.69	42.80	-0.83	0.23	.878
25	5.18	4.28	4.39	2.51	51.02	73.08	-0.55	0.70	.555
26	10.59	9.25	9.68	9.06	18.49	69.76	-1.28	0.27	.851
Total		6.59				99.40			

Note. $\hat{\sigma}_A^2$ is the estimate of variance among stimuli

실험참가자들이 실험자극에 적절한 주의를 기울였다는 것을 의미한다.

모의범죄 실험자료에서 24명 중 7명에서 4개의 무관련자극들 간에 비동질성 문제가 있는 것으로 나타났다. 즉, 본 연구 결과에 따르면 실험참가자의 약 30%에서 통계적으로 유의한 자극의 비동질성 문제가 나타난 것이다. 자극의 비동질성 문제가 자극 자체의 특성 때문에 발생하는지 파악하기 위하여 자극의 비동질성 문제가 통계적으로 유의하였던 7명에서 P300 진폭이 가장 컸던 자극을 살펴보았다. 그 결과 무관련자극 1부터 4까지 비교적 고르게 분포하였다. 즉, 자

극의 비동질성 문제가 자극 자체의 특성 때문이 아니라 실험참가자의 특성(개인적 경험에 의한 자극의 의미) 때문에 발생하는 것으로 생각된다.

모의범죄 실험자료와 달리 기억상실 실험자료에서는 26명 중 1명만 무관련자극들 간의 P300 진폭이 통계적으로 유의하게 다른 것으로 나타났다. 즉, 기억상실 실험자료에서는 자극의 동질성 문제가 뚜렷하게 나타나지는 않았다. 이러한 결과가 나타난 이유를 실험절차에서 찾을 수 있다. 기억상실 실험에서는 P300 CIT를 실시하기 전에 실험참가자에게 무관련자극을 보여주고 개인적으로 의미가 있는 자극이 있는 경

우에 이 자극을 개인적으로 의미가 없는 자극으로 바꾸는 것이 일반적이며(Verschuere et al., 2009), Lee 등(2013)도 동일한 절차를 사용하였다. 즉, 모든 자극이 동일한 정도로 의미가 없도록 조작하는 절차가 있었기 때문에 자극의 비동질성 문제가 현저히 감소하였을 것이다. 그러나 실제 거짓말 탐지에서는 P300 CIT를 실시하기 전에 검사 자극들을 보여 주고 개인적으로 의미가 있는 자극을 의미가 없는 다른 자극으로 바꾸는 절차를 사용하기가 어렵다. 왜냐하면 피검사자가 관련자극을 개인적으로 의미가 있는 것이라고 거짓으로 응답할 수 있기 때문이다. 따라서 기억상실 실험자료는 생태학적 타당도가 낮으므로 모의 범죄 실험자료만 이용하여 자극의 비동질성 문제를 평가하는 것이 타당할 것이다.

실제 거짓말 탐지 환경과 더 유사한 모의범죄 실험자료에서는 약 30%에서 자극의 비동질성 문제가 나타났으므로, 단일관련자극 프로토콜에서 자극의 비동질성 문제가 존재한다고 결론을 내릴 수 있을 것이다. 자극의 비동질성 수준은 모집단에서 자극 간 변량으로 정의할 수 있을 것이다. 모의범죄 자료와 기억상실 자료를 종합하면 자극 간 변량은 최대 3~5 μ V 정도일 것으로 추정된다.

연구 1의 목적이 피검사자가 무죄인 경우에 1개의 관련자극과 4개의 무관련자극이 모두 동질적인지 평가하는 것이었으므로, 무죄 조건의 실험자료에서 1개의 관련자극과 4개의 무관련자극을 평가하여야 한다. 그러나 본 연구에서는 유죄 조건의 실험자료에서 4개의 무관련자극을 평가하였기 때문에 두 가지 문제가 있을 수도 있다. 한 가지는 무관련자극만 평가한 점이다. 그러나 이 문제는 결과에 큰 영향을 미치지 않았을 것이다. 무죄인 피검사자의 경우에 관련자극은 제 5의 무관련자극일 뿐이므로 무관련자극들 간의 P300 차이만 분석하였다 하더라도 결과가 크게 다르지는 않을 것이다. 다른 한 가지는 유죄인 피검사자의 실험자료를 분석했다는 점이다. 유죄인 피검사자의 주의를 목표자극과 관련자극에 집중되어 있을 것이지만(Farwell & Donchin, 1991), 무죄인 피검사자의 경우에는 목표자극에만 주의를 집중될 것이다. 따라서 무관련자극에 주어지는 주위의 양은 무죄인 경우보다 유죄인 경우에 더 작을 것이다. P300 진폭은 자극에 주어지는 주위의 양에 비례하므로(Johnson, 1993), 본 연구의 결과는 자극의 비동질성 문제를 과소평가했을 가능성이 있다.

연구 2

연구 2의 목적은 관련자극을 여러 개 사용하는 중다관련자극 프로토콜에서 자극의 비동질성 문제가 발생하는지 또한

자극의 비동질성 문제가 발생한다면 그 수준은 어느 정도인지 평가하는 것이다. 중다관련자극 프로토콜에서 자극의 비동질성 문제가 존재한다면 각 실험참가자별로 자극들 간의 P300 진폭이 유의하게 다를 것이다. 또한 자극의 비동질성 수준은 자극들 간의 변량추정치를 이용하여 평가할 수 있다.

관련자극을 1개만 사용하는 단일관련자극 프로토콜에서는 모든 자극이 동일한 범주의 자극이지만, 중다 관련자극 프로토콜에서는 서로 다른 범주의 자극들이 사용되므로 자극의 비동질성 문제가 다르게 발생할 수도 있다. 따라서 무죄 조건의 실험참가자들에게 중다관련자극 프로토콜을 사용한 P300 CIT를 실시한 후에 자극의 비동질성 문제를 평가하였다. 기억상실 상황에 대한 실험연구는 생태학적 타당도가 낮으므로 연구 2에서는 가상의 절도 사건에 대한 실험연구를 수행하였다.

방 법

실험참가자

대학생 17명이 실험에 자원하여 참가하였다. 실험참가자들은 모두 뇌손상을 입은 경험이나 정신과적 질환으로 인한 진료를 받은 적이 없다고 응답하였다. 성별은 남자 9명과 여자 8명이었으며, 평균연령은 22.7세(범위 18~26세)였다. 실험이 끝난 후 실험참가자들은 금전적인 보상을 받았다.

실험자극

가상의 범죄 사건으로부터 P300 CIT에 사용할 자극을 구성하였다. 자극은 5개 범주로 구성하였으며 각 범주에 6개의 자극을 포함하였다. 5개 범주와 자극은 도난 사건이 일어난 층(1층, 2층, 3층, 5층, 6층, 7층), 도난 사건이 일어난 장소(실습실, 연구실, 행정실, 사무실, 회의실, 강의실), 도난당한 물건(반지, 팔찌, 목걸이, 시계, 귀걸이, 브로치), 도난당한 금액(40만원, 45만원, 50만원, 55만원, 60만원, 65만원), 도난당한 물건과 돈이 들어 있던 곳(가방, 지갑, 책상, 서랍, 핸드백, 잠바)이었다. 이 자극들 중 목표자극은 2층, 연구실, 팔찌, 65만원, 지갑이었다. 본 연구는 무죄조건만 포함하고 있었기 때문에 관련자극을 특정하지는 않았다.

실험절차

실험참가자들은 실험에 대한 전반적인 사항이 적혀있는 실험 참가동의서를 읽고 서명한 후에 실험에 참가하였다. 실험참가자들은 모두 무죄 조건에서 실험에 참가하였다. 즉, 실험 참가자들은 절도를 하지 않은 상황에서 실험에 참가하였다.

실험참가자에게 “당신은 무죄인 상황에서 도난 사건에 대한 용의자로 거짓말 탐지 검사를 받을 것입니다.”라고 알려주었다.

목표자극이 무엇인지는 P300 CIT를 실시하기 전에 구두로 불러주었으며, 목표자극이 무엇인지 묻고 답하는 과정을 통하여 실험참가자가 목표자극을 잘 암기하였는지 확인하였다. P300 CIT가 시작되면 목표자극이 제시되었을 때 ‘예’ 버튼(마우스 왼쪽 버튼)을, 목표자극 이외의 자극이 제시되었을 때 ‘아니오’ 버튼(마우스 오른쪽 버튼)을 가능한 빠르고 정확하게 누르도록 지시하였다. 목표자극과 무관련자극 30개를 섞어서 하나씩 무선적으로 제시하였으며, 자극은 실험참가자 앞에 놓여있는 컴퓨터 모니터 중앙에 제시하였다. 글자는 신명조체로 제시하였으며, 자극의 수평시각도는 최대 2.58°, 수직시각도는 0.86° 였다. 자극은 300ms 동안 제시하였으며, 자극 간 제시 간격은 평균 2700ms(범위 2200~3200ms)였다. 즉, 평균적으로 3초에 한 번씩 자극을 제시하였다. 30개의 자극 세트를 무선적으로 제시하는 것을 9회 반복하여 총 270회 자극을 제시하였다. 자극의 제시와 반응의 기록은 SuperLab 5.0을 이용하였다.

뇌파기록

뇌파측정위치가 F3, Fz, F4, Cz, P3, Pz, 및 P4였다는 점과 A/D 변환을 위한 표본율이 250Hz였다는 점을 제외하고, 뇌파를 측정하고 노이즈를 제거하는 방법은 연구 1과 동일하였다. 분석에서 제외된 시행의 비율은 0.7%였다.

분석

각 자극의 제시 전 200ms부터 자극제시 후 1400ms까지 총 1600ms 구간으로 나누어 분석하였다. 무관련자극에 대한 뇌파를 평균하여 1개의 ERP를 산출한 후 P300 진폭을 계산하였다. 자극제시전 200ms구간의 평균전위를 이용하여 ERP 기저선을 교정하였다. P300 진폭을 계산하는 방법은 연구 1과 동일하였다. 자극 별로 ERP를 산출하지 않은 이유는 각 자극의 시행 수가 9회에 불과하기 때문에 안정적인 ERP를 산출할 수가 없었기 때문이다. 자극들 간의 P300 진폭이 유의하게 다른지 검증하기 위하여 각 시행에서 P300 진폭을 계산하였다. 각 시행의 P300 진폭은 전체 ERP에서 P300의 전위가 나타난 시점과 P300 이후 부적인 정점이 나타난 시점을 이용하여 계산하였다. 중다관련자극 프로토콜에서는 자극들이 범주로 나누어지므로, 25개 자극들 간의 P300 진폭이 다른지 검증하게 되면 자극의 효과에 범주의 효과가 포함되게 된다. 따라서 자극 범주를 동지된 요인으로 하고 각 자

극을 동지된 요인으로 한 동지요인설계에 의한 변량분석을 실시하였으며, 동지된 요인과 동지된 요인을 모두 무선요인으로 간주하였다.

자극의 비동질성 수준을 수량화하기 위하여 자극들 간의 변량을 계산하였다. 동지요인 설계에서 자극은 범주에 조건화되어 있으므로, 범주에 조건화된 자극 간 변량인 범주 내 자극 간 변량을 추정하였다. 동지요인 설계에서 범주 내 자극 간 평균자승의 기댓값은 $E(MSB_{B:A}) = \sigma_e^2 + n\sigma_{B:A}^2$ 으로 오차 변량(σ_e^2)과 범주 내 자극 간 변량($\sigma_{B:A}^2$)에 한 자극의 시행 수(n)를 곱한 값으로 구성되어 있다. 집단 내 평균자승의 기댓값은 $E(MSW) = \sigma_e^2$ 으로 오차변량(σ_e^2)만으로 구성되어 있다 (Park, 2003). 따라서 범주 내 자극 간 변량($\sigma_{B:A}^2$)을 수식 (3)과 같이 추정할 수 있다.

$$\sigma_{B:A}^2 = \frac{MSB_{B:A} - MSW}{n} \tag{3}$$

결 과

행동반응

실험참가자들이 검사자극에 적절한 주의를 기울였는지 평가하기 위하여 정확반응률과 반응시간을 산출하였다. 목표자극과 무관련자극의 정확반응률 평균은 각각 87.8%(SD=10.2%)와 99.8%(SD=1.8%)였으며, 개인별 목표자극의 정확반응률은 모두 70%이상이었다. 목표자극과 무관련자극의 반응시간 평균은 각각 770ms(SD=141ms)와 690ms(SD=164ms)였다.

사건관련전위

Fz, Cz, 및 Pz에서 측정된 ERP가 Figure 3에 제시되어 있다. P300 진폭이 가장 크게 나타나는 Pz에서 가장 큰 정적 전위는 자극 제시 후 약 600ms에서 나타났으며, 관련자극에 대한 P300 진폭이 무관련자극에 대한 P300 진폭보다 더 크게 나타났다. 관련자극과 무관련자극에 대한 P300 진폭은 Cz와 Fz로 갈수록 작아졌지만, 관련자극에 대한 P300 진폭은 무관련자극에 대한 P300 진폭보다 더 큰 경향은 유지되었다. P300 CIT에서는 P300 진폭이 가장 크게 나타나는 Pz의 ERP를 이용한다. 따라서 본 연구에서도 Pz영역의 뇌파를 이용하여 자극 간 동질성 여부를 검증하였다.

각 실험참가자별 무관련자극 전체의 평균 P300 진폭과 범주 내 자극 간 변량의 추정치, 변량분석결과가 Table 3에 제시되어 있다. 개인별로 범주 내 자극에 대한 P300 진폭이 통계적으로 유의하게 다른지 검증하기 위하여 동지요인설계

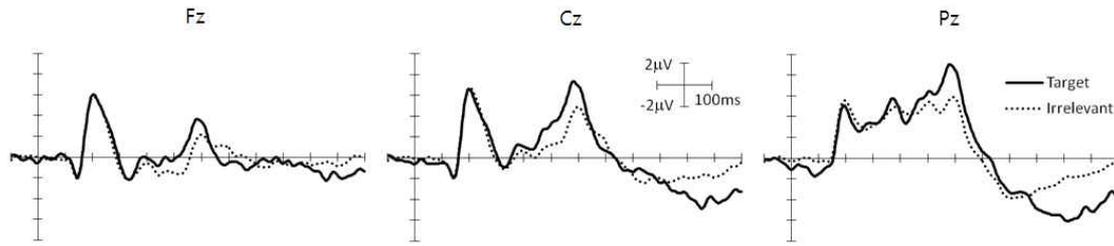


Figure 3. Grand averaged ERPs at Fz, Cz, and Pz site in Study 2. The solid lines represent the target stimuli and dotted lines represent the irrelevant stimuli.

Table 3. P300 amplitude difference among stimuli in multiple probe protocol

ID	P300	MSBA	MSBBIA	MSW	$\hat{\sigma}_{BA}^2$	Category		Stimulus within category	
						<i>F</i>	<i>p</i>	<i>F</i>	<i>p</i>
1	11.90	732.29	163.13	155.12	0.89	4.49	.009	1.05	.404
2	10.56	21.18	239.51	199.11	4.49	0.09	.985	1.20	.255
3	11.60	128.34	315.75	127.72	20.89	0.41	.802	2.47	.001
4	8.46	476.34	98.65	120.22	-2.40	4.83	.007	0.82	.687
5	7.04	117.10	179.81	168.22	1.29	0.65	.633	1.07	.385
6	14.31	273.11	272.19	157.60	12.73	1.00	.429	1.73	.032
7	8.70	668.56	225.57	144.83	8.97	2.96	.045	1.56	.066
8	5.56	29.06	164.85	147.72	1.90	0.18	.948	1.12	.335
9	9.86	319.66	143.93	205.60	-6.85	2.22	.103	0.70	.824
10	18.05	274.55	180.32	94.64	9.52	1.52	.234	1.91	.014
11	9.05	177.02	82.00	122.62	-4.51	2.16	.111	0.67	.854
12	10.54	331.32	248.86	155.65	10.36	1.33	.293	1.60	.056
13	10.03	127.92	84.21	91.09	-0.76	1.52	.235	0.92	.557
14	20.44	633.66	367.75	249.89	13.10	1.72	.184	1.47	.094
15	12.92	765.71	367.74	283.13	9.40	2.08	.121	1.30	.183
16	5.17	174.89	105.13	86.39	2.08	1.66	.198	1.22	.243
17	18.22	680.36	331.52	148.98	20.28	2.05	.125	2.23	.003
Total	11.32			156.38					

Note. P300 is the mean P300 of 25 stimuli. $\hat{\sigma}_{BA}^2$ is the estimate of variance among stimuli within category.

에 의한 변량분석을 실시하였다. 분석결과 전체 17명 중 4명에서 통계적으로 유의한 평균차이가 나타났다($F(20,199)=2.47$, $p=.001$; $F(20,200)=1.73$, $p=.032$; $F(20,200)=1.91$, $p=.014$; $F(20,200)=2.23$, $p=.003$). 각 실험참가자의 범주 내 자극 간 변량의 추정치는 $-6.85\sim 20.89$ 의 범위를 보였다. 비동질성 문제가 통계적으로 유의하였던 경우에 범주 내 자극 간 변량의 추정치는 $9.52\sim 20.89$ 였다. 개인별로 범주 간 P300 진폭이 통계적으로 유의하게 다른 경우는 전체 17명 중 3명이 있었다($F(4,20)=4.49$, $p=.009$; $F(4,20)=4.83$, $p=.007$; $F(4,20)=2.96$, $p=.045$). 전체 실험참가자의 평균 P300

진폭은 $11.32\mu V$ 였으며, 각 시행에 대한 P300 진폭의 표준편차(즉, MSW의 제곱근)는 평균 $12.51\mu V$ 였다.

논 의

중다관련자극 프로토콜을 사용한 P300 CIT에서 자극의 비동질성 문제가 나타나는지 확인하기 위하여 실험연구를 수행하였다. 목표자극에 대한 반응오류율은 약 10%내외로 선행연구들과 큰 차이가 없었으며(Cutmore et al., 2009; Rosenfeld et al., 2004), 본 연구의 반응시간은 관련자극을

3개 사용한 연구(Mertens & Allen, 2008)와 비슷하였으며, 관련자극을 6개 사용한 연구(Farwell & Donchin, 1991) 보다는 짧았다. P300 CIT가 예측하듯이 목표자극에 대한 P300 진폭이 무관련자극의 P300 진폭보다 더 컸다. 이러한 결과는 전체적으로 실험연구가 적절하게 진행되었음을 의미한다.

전체 17명 중 4명에서 범주 내 자극의 비동질성 문제가 있는 것으로 나타났다. 즉, 약 25%에서 자극의 비동질성 문제가 있었다. 자극의 비동질성 문제가 통계적으로 유의하였던 4명에서 P300 진폭이 가장 컸던 자극을 살펴본 결과, 동일한 자극에서 P300 진폭이 가장 큰 경우는 없었다. 즉, 자극의 비동질성 문제가 자극 자체의 특성 때문에 발생한 것은 아니었다. 5개의 범주 간 P300 진폭차이는 전체 17명중 3명에서 통계적으로 유의한 결과가 나타났지만, 범주 간 P300 진폭차이는 P300 CIT의 판단에 큰 영향을 끼치지 않을 것이다. 왜냐하면 범주의 효과가 관련자극과 무관련자극에 동일하게 나타날 것이기 때문이다.

자극의 비동질성 문제는 단일관련자극 프로토콜과 중다관련자극 프로토콜에서 전체 실험참가자의 약 25%~30%에서 나타났다. 그러나 단일관련자극 프로토콜에서는 전체 시행수가 많았기 때문에 중다관련자극 프로토콜에서보다 검증력이 훨씬 높았을 것이다. 이러한 점을 감안하면 단일관련자극 프로토콜보다 중다관련자극 프로토콜에서 자극의 비동질성 문제가 더 크게 나타날 가능성이 있다. 자극 간 변량의 추정치도 이러한 생각을 지지해 준다. 단일관련자극 프로토콜에서 자극 간 변량의 추정치는 최대 3.46인 반면에 중다관련자극 프로토콜에서 자극 간 변량의 추정치는 최대 20.89였다. 만약 중다관련자극 프로토콜에서 자극의 비동질성 문제가 더 크게 나타나는 것이 사실이라면, 그것은 자극의 종류가 많아짐에 따라서 자극이 가지는 의미의 폭이 더 넓어지기 때문일 것이다.

연구 3

연구 3의 목적은 자극의 비동질성 수준이 오류 긍정률에 미치는 영향을 확인하는 것이다. 연구 1과 2에서 확인하였듯이 자극의 비동질성 문제는 개인마다 다르게 나타나며, 자극의 비동질성 수준(즉, 자극들간의 변량 추정치)도 개인마다 다르다. 따라서 연구 3에서는 다양한 비동질성 수준에서 오류 긍정률이 나타나는 정도를 확인하였다. 자극의 비동질성 수준이 높아질수록 오류 긍정률이 높아질 것으로 예상할 수 있다. 또한 P300 CIT에서 사용하는 관련자극의 개수, 시

행 수, 유의수준이 다양하므로, 비동질성 수준과 관련자극의 개수, 시행 수, 유의수준에 따라서 오류 긍정률이 어떻게 달라지는 파악하였다. 다양한 조건에 따른 대규모 실험연구를 수행하는 것은 사실상 불가능하므로 본 연구에서는 가상의 자료를 이용하는 몬테 카를로(Monte Carlo) 연구를 수행하였다.

방 법

조건

비동질성 수준과 관련자극의 수, 시행 수, 유의수준이 오류 긍정률에 미치는 영향을 평가하기 위하여, 몬테 카를로 연구에 사용할 비동질성 수준과 관련자극의 수, 시행 수, 유의수준을 정하였다. 비동질성 수준은 자극 간 변량의 제곱근, 즉 표준편차로 정의하였다. 연구 1의 모의범죄 자료에서 자극 간 변량의 추정치는 최대 3.46이며, 연구 2에서 범주 내 자극 간 변량의 추정치는 최대 20.89의 값을 보였다. 자극 간 변량을 표준편차로 환산하면 연구 1은 최대 2 μ V, 연구 2는 최대 5 μ V의 값을 가진다. 따라서 비동질성 수준을 0, 1, 2, 3, 4, 5 μ V로 정하였다. 단일관련자극 프로토콜에서는 관련자극을 1개만 사용한다. 중다관련자극 프로토콜에서 관련자극의 개수는 특정 사건에서 얻을 수 있는 관련자극의 수에 따라 결정된다. 선행연구들에서는 최대 6개의 관련자극을 사용하고 있다(Farwell & Donchin, 1991). 이에 따라 본 연구에서는 관련자극의 개수를 1개, 3개, 6개를 사용하였다. 무관련자극의 개수는 대부분의 경우에 관련자극 개수의 4배를 사용하므로, 본 연구에서도 동일하게 하였다. 즉, 관련자극의 개수가 1개인 경우에 무관련자극의 개수는 4개(전체 자극 개수는 5개), 관련자극의 개수가 3개인 경우에 무관련자극의 개수는 12개(전체 자극개수는 15개), 관련자극의 개수가 6개인 경우에는 무관련자극의 개수는 24개(전체 자극개수는 30개)였다. P300 CIT에서는 일반적으로 관련자극에 대한 시행수가 30회 이상이 되도록 하며(Rosenfeld et al., 2006), 선행연구 중에는 72회가 가장 많은 시행 수였다(Farwell & Donchin, 1991). 따라서 본 연구에서도 30회에서 72회까지의 시행 수 중에서 3가지를 선택하였다. 이 때 조건들 간의 비교가능성을 유지하기 위하여 관련자극의 개수 1, 3, 6개의 공통배수가 되는 30회, 48회, 66회를 선정하였다. 유의수준은 선행연구에서 주로 사용하는 .10과 .05를 사용하였다(Lefebvre et al., 2009; Rosenfeld et al., 2007).

무선자료산출

실험조건(비동질성 수준, 관련자극의 수, 시행 수)에 따라 관련자극과 무관련자극의 P300 진폭에 대한 무선자료를 산출하기 위하여, P300 진폭의 모수(parameter)와 개별 시행에서 P300 진폭의 표준편차에 대한 모수를 결정하였다. P300 진폭은 개인차이가 있기는 하지만, 연구 1과 연구 2의 평균 P300 진폭은 $8.36\mu V$ 와 $11.32\mu V$ 로 약 $10\mu V$ 로 나타나 P300 진폭에 대한 모수는 $10\mu V$ 으로 정하였다. 개별 시행에 대한 P300 진폭의 표준편차(Table 1과 Table 3에서 MSW의 제곱근)는 평균 약 $10\mu V$ 로 나타나 표준편차에 대한 모수는 $10\mu V$ 로 정하였다.

개별 시행에 대한 P300 진폭이 정규분포한다고 가정할 수 있는지 평가하기 위하여 시행 수가 가장 많았던 연구 1의 모의범죄 자료를 검토하였다. 각 실험참가자에 대해서 단일 시행의 P300 진폭에 대한 왜도와 첨도, 히스토그램과 P-P 도표를 작성하였다. 24명의 실험참가자에 대해서 왜도는 $-0.183\sim 0.487$ 의 범위를 보였고, 첨도는 $-0.221\sim 1.440$ 의 범위를 보였으며 히스토그램과 P-P 도표를 살펴본 결과 정규 분포에서 크게 이탈한 사례는 없었다. 따라서 개별 시행에 대한 P300 진폭은 정규분포한다고 가정하였다.

비동질성 수준이 $1\sim 5\mu V$ 인 경우에 자극들 간 P300 진폭의 분포가 정규분포한다고 가정할 수 있는지 평가하였다. 이 가정의 타당성을 검토하기 위해서는 많은 수의 자극에 대한 P300 진폭이 필요하기 때문에 연구 2의 자료를 이용하였다. 그러나 연구 2의 자료에서 한명의 실험참가자에게 얻을 수 있는 P300 진폭은 25개(자극의 개수)에 불과하기 때문에, 모든 실험참가자의 전체 자극들에 대한 P300 진폭을 이용하여 왜도와 첨도를 산출하였다. 왜도는 0.49였으며 첨도는 0.22로 약간 정적 편포하는 것으로 나타났다. 히스토그램과 P-P 도표도 정적편포하는 경향이 있었지만, 그 정도가 심하지 않아서 자극들 간 P300 진폭의 분포가 정규분포한다고 가정하였다.

무죄인 경우에는 관련자극과 무관련자극 간에 P300 진폭의 차이가 없다고 가정하므로 동일한 조건에서 관련자극과 무관련자극에 대한 무선자료를 추출하였다. 비동질성 수준이 0인 경우에 평균 10과 표준편차 10인 정규분포로부터 관련자극과 무관련자극의 개수와 시행 수만큼 무선자료를 추출하였다. 비동질성 수준이 1인 경우에는 평균 10, 표준편차 1인 정규분포로부터 관련자극과 무관련자극의 개수만큼 무선자료를 추출하여 각 자극의 P300 진폭에 대한 모수를 결정하였다. 그런 다음, 각 자극에 대하여 해당 P300 진폭과 표준편차 10인 정규분포로부터 정해진 시행 수만큼 무선자료를 추

출하였다. 비동질성 수준이 2~5인 경우도 동일한 절차를 이용하여 무선자료를 추출하였다.

특정 조건에 해당하는 무선자료를 추출한 후에는 다음 항에 설명할 부트스트랩 절차를 이용하여 관련자극에 대한 P300 진폭이 무관련자극에 대한 P300 진폭보다 유의하게 더 큰지 평가하였으며, 이 과정을 10만 번 반복하여 유의수준 .10과 .05에서 오류 긍정률을 산출하였다.

분석방법

P300 CIT에서 유/무죄를 판단하는 통계적 기법들은 크게 두 가지로 분류될 수 있다. 한 가지는 관련자극과 무관련자극의 P300 파형 차이를 이용하여 유죄 집단과 무죄 집단을 판별하는 방정식을 도출한 후, 이 방정식을 사용하여 유/무죄를 판단하는 방법이다(예, Abootalebi, Moradi, & Khalizadeh, 2006; Mettens & Allen, 2008). 그러나 P300 파형은 검사자극에 따라서 달라지므로(Johnson, 1993), 판별 방정식은 검사자극에 따라서 달라져야만 한다. 그러나 P300 CIT에 사용되는 검사자극은 사례별로 모두 다르기 때문에 사례별로 판별 방정식을 매번 산출하여 거짓말 탐지에 이용하는 것은 거의 불가능하다.

다른 한 가지는 관련자극과 무관련자극 간의 P300 파형이 얼마나 유사한지를 분석하여 유/무죄를 판단하는 방법이다. 이 방법은 다시 두 가지로 세분된다. 첫 번째는 Farwell과 Donchin(1991)이 고안한 방법으로 부트스트랩 절차를 사용하여 관련자극에 대한 P300 파형이 목표자극에 대한 P300 파형과 무관련자극에 대한 P300 파형 중 어떤 것과 더 비슷한지를 통계적으로 평가하여 유/무죄를 판단하는 것이다. 그러나 이 방법은 유죄인 피검사자인 경우 목표자극에 대한 P300 파형이 관련자극에 대한 P300 파형과 비슷하다고 가정하는데, 실제로 목표자극과 관련자극에 대한 P300 파형이 다르게 생긴 경우에 유/무죄 판단에 문제가 발생할 수 있다(Rosenfeld, 2005). 또한 목표자극의 특성과 과제의 특성에 따라서 목표자극의 P300 파형이 영향을 받기 때문에 널리 사용되지는 않는다. 두 번째는 부트스트랩 P300 진폭 차이 검증으로 관련자극에 대한 P300 진폭이 무관련자극에 대한 P300 진폭보다 통계적으로 유의하게 더 큰지를 평가하는 방법이다(Rosenfeld, Ward, Meijer, & Yuhnenko, 2017). 이 방법은 목표자극의 특성과 과제의 특성에 관계없이 모든 경우에 사용할 수 있기 때문에 P300 CIT 연구에서 가장 널리 사용되고 있다. 따라서 본 연구에서는 부트스트랩 P300 진폭 차이검증을 이용하여 P300 CIT의 오류 긍정률을 평가하였다.

부트스트랩 P300 진폭 차이검증은 다음과 같은 과정을 거쳐 수행되었다. 먼저 관련자극에 대한 N1개의 표본으로부터 N1개의 표본을 복원추출하고, 무관련자극에 대한 N2개의 표본으로부터 N2개의 표본을 복원추출한다. 복원추출을 하기 때문에 원래의 표본과 복원추출된 표본은 약간의 차이를 가지게 된다. 복원추출된 표본으로부터 관련자극과 무관련자극에 대한 P300 진폭(즉, 평균)을 계산한다. 만약 관련자극에 대한 P300 진폭이 무관련자극에 대한 것보다 더 크면 부트스트랩 지수에 1을 더한다. 이 과정을 1000번 반복한다. 유의수준 .10에서는 부트스트랩 지수가 900 이상이면 유죄 판단을 하며(Rosenfeld et al., 2006), 유의수준 .05에서는 부트스트랩 지수가 950 이상이면 유죄 판단을 한다(Lefebvre et al., 2009). 관련자극이 1개인 경우(단일관련자극 프로토콜)에 Lefebvre 등(2009)은 오류 긍정률을 줄이기 위하여 관련자극에 대한 P300 진폭을 무관련자극 중 P300 진폭이 가장 큰 것과 비교하는 것을 권하였다. 그러나 이 절차가 오류 긍정률을 어느 정도 감소시키는지 알려진 바가 없다. 따라서 본 연구에서는 단일관련자극 프로토콜에서 관련자극에 대한 P300 진폭을 무관련자극 중 P300 진폭이 가장 큰 것과 비교하는 부트스트랩 P300 진폭 차이검증을 한 번 더 수행하였다.

결 과

비동질성 수준과 관련자극의 개수, 시행 수, 유의수준에 따라 오류 긍정률을 산출한 결과가 Table 4에 제시되어 있다. 비동질성 수준이 0 μ V인 경우, 즉 자극의 비동질성 문제가 존재하지 않는 경우에는 관련자극의 개수와 시행 수에 관계없이 유의수준과 매우 유사한 오류 긍정률이 나타났다. 유의수준으로 .10을 사용하는 경우에 오류 긍정률은 .101~.107의 범위를 보였으며, 유의수준으로 .05를 사용하는 경우에 오류 긍정률은 .050~.057의 범위를 보였다.

비동질성 수준이 1 μ V 이상인 경우에는 모든 경우에 설정된 유의수준보다 높은 오류 긍정률이 나타났다. 특히 관련자극의 개수가 적을수록 오류 긍정률이 높았고 시행 수가 많을수록 오류 긍정률이 높아져, 관련자극의 개수가 6개이고 관련자극의 시행 수가 30회일 때 오류 긍정률이 상대적으로 낮았고 관련자극의 개수가 1개이고 시행 수가 66회일 때 오류 긍정률이 가장 높았다. 유의수준으로 .10을 사용했을 때, 비동질성 수준 1 μ V에서 오류 긍정률은 .110~.163의 범위를 보였으며, 비동질성 수준이 2에서는 .122~.252의 범위를, 비동질성 수준 3 μ V에서는 .139~.311의 범위를, 비동질성 수준 4 μ V에서는 .160~.344의 범위를, 비동질성 수준이 5 μ V에서는

Table 4. The results of Monte Carlo study

Comparison	NP	NT	Level of non-equality											
			0		1		2		3		4		5	
			Sig.	Sig.	Sig.	Sig.	Sig.	Sig.	Sig.	Sig.	Sig.	Sig.	Sig.	
			.10	.05	.10	.05	.10	.05	.10	.05	.10	.05	.10	.05
P vs Iall	1	30	.107	.057	.135	.082	.199	.139	.257	.200	.300	.250	.332	.288
		48	.104	.053	.151	.192	.229	.171	.290	.239	.331	.287	.361	.323
		66	.103	.050	.163	.105	.252	.194	.311	.260	.344	.303	.383	.348
P vs Iall	3	30	.107	.056	.116	.063	.143	.086	.173	.114	.208	.148	.235	.176
		48	.103	.053	.120	.065	.159	.099	.203	.144	.239	.182	.268	.215
		66	.101	.052	.126	.071	.173	.116	.225	.166	.265	.210	.294	.244
P vs Iall	6	30	.105	.055	.110	.059	.122	.069	.139	.083	.160	.101	.176	.119
		48	.102	.053	.112	.059	.130	.077	.158	.101	.186	.126	.209	.150
		66	.103	.053	.116	.059	.138	.084	.179	.118	.209	.152	.236	.179
P vs Imax	1	30	.011	.003	.016	.007	.033	.018	.054	.035	.073	.052	.089	.068
		48	.010	.003	.019	.008	.043	.025	.070	.048	.090	.068	.106	.086
		66	.009	.002	.026	.011	.052	.032	.081	.060	.097	.077	.118	.099

Note. P = probe, Iall = the mean P300 amplitude of all irrelevants, Imax = the largest maximum irrelevant stimulus P300. Sig = significance level, NP = the number of probes, NT = the number of trials for probe. The number of trials for irrelevants is the same as the number of trials for probe multiplied by 4.

.176~.383의 범위를 보였다. 유의수준으로 .05를 사용했을 때에는 전체적으로 .059~.348의 범위를 보였다.

관련자극이 1개인 경우, 추가적으로 관련자극에 대한 P300 진폭과 무관련자극 중 P300 진폭이 가장 큰 값과 비교하였다. 비동질성 수준이 0인 경우, 유의수준 .10에서 오류 긍정률은 시행 수에 관계없이 약 .010이었으며, 유의수준 .05에서 약 .003의 오류 긍정률을 보였다. 비동질성 수준이 높아질수록 오류 긍정률도 높아져서, 비동질성 수준이 $3\mu V$ 이고 시행 수가 66회, 유의수준 .05인 경우에는 설정한 유의수준보다 높은 .060의 오류 긍정률을 보였다.

논 의

연구 3에서는 가상의 자료를 이용하여 자극의 비동질성 문제와 P300 CIT의 특성들이 오류 긍정률에 미치는 영향을 검증하였다. 비동질성 수준이 $0\mu V$ 인 경우 오류 긍정률은 유의수준 보다 약간 큰 경향이 있었지만 1% 이내였다. 비동질성 수준이 $1\mu V$ 이상인 경우에 오류 긍정률은 적게는 유의수준보다 1% 더 높았으며 많게는 30% 가까이 더 높았다. 즉 상황에 따라서 자극의 비동질성 문제가 오류 긍정률에 큰 영향을 미치는 것으로 나타났다.

연구 1의 단일관련자극 프로토콜에서는 비동질성 수준이 최대 약 $2\mu V$ 로 나타났다. 이 값을 단일관련자극 프로토콜에서 비동질성 수준이 가장 높은 경우로 정의하면, 단일관련자극 프로토콜에서는 오류 긍정률이 최대 25%까지 높아질 수 있는 것으로 나타났다. 연구 2의 중다관련자극 프로토콜에서는 비동질성 수준이 최대 약 $5\mu V$ 로 나타났다. 이 값을 중다관련자극 프로토콜에서 비동질성 수준이 가장 높은 경우로 정의하면, 중다관련자극 프로토콜에서는 오류 긍정률이 최대 38%까지 높아질 수 있는 것으로 나타났다.

비동질성 수준이 $1\mu V$ 이상인 경우 관련자극의 개수가 증가함에 따라서 오류 긍정률이 감소하였다. 유명한 통계적 정리인 중심극한정리(central limit theorem)에서 이러한 결과가 나타난 이유를 찾을 수 있다. 중심극한 정리에 따르면 표본의 크기가 클수록 표본 평균은 모집단 평균에 가까워진다. 즉, 평균을 산출하는데 사용되는 표본의 크기가 클수록 평균들 간의 변산성이 작아진다. 따라서 중다관련자극 프로토콜처럼 3개(또는 6개)의 관련자극과 12개(또는 24개)의 무관련자극에서 P300 진폭을 계산하는 경우가 단일관련자극 프로토콜처럼 1개의 관련자극과 4개의 무관련자극에서 P300 진폭을 계산하는 경우보다 변산성이 더 작아지게 되고, 결과적으로 오류 긍정률이 감소했을 것이다.

비동질성 수준이 $1\mu V$ 이상인 경우 시행 수가 많아질수록 오류 긍정률이 증가하였다. 이러한 결과도 중심극한정리에 의해서 잘 설명된다. 시행 수가 많을수록 표준오차가 작아지는데, 표준오차가 작아지면 관련자극과 무관련자극에 대한 P300 진폭 차이의 유의확률이 작아지고, P300 진폭 차이가 통계적으로 유의할 가능성이 증가하게 된다. 결과적으로 관련자극의 개수와 시행 수가 오류 긍정률에 미치는 영향은 모두 동일한 원인에 의해서 나타나는 것으로 생각된다.

관련자극의 P300 진폭을 무관련자극들 중 P300 진폭이 가장 큰 값과 비교하는 경우에 오류 긍정률이 극적으로 감소하였다. 비동질성 수준이 $0\mu V$ 이고 유의수준이 .10인 경우에 오류 긍정률은 약 1%였다. 관련자극의 P300 진폭을 무관련자극 전체의 P300 진폭과 비교하는 경우에 비교하여 약 1/10로 줄어든 것이다. 유의수준 .05를 사용하는 경우에는 더욱 감소하여 오류 긍정률이 0.3%에 불과하였다. 자극의 비동질성 수준이 $1\mu V$ 이상인 경우에는 오류 긍정률이 다소 증가하였지만, 비동질성 수준 $3\mu V$ 까지는 연구자가 설정한 1종 오류보다는 낮았다고 볼 수 있다. 단일관련자극 프로토콜에서 비동질성 수준이 $3\mu V$ 까지 발생하는 경우가 드물다고 보면, 관련자극의 P300 진폭을 무관련자극들 중 P300 진폭이 가장 큰 값과 비교하는 경우에 오류 긍정률은 유의수준 이하로 통제된다고 하여도 무방할 것이다.

전체 논의

본 연구에서는 P300 CIT에서 자극의 비동질성 문제가 나타나는지 그리고 자극의 비동질성 문제가 오류 긍정률에 어떠한 영향을 미치는지 확인하였다. 연구결과, 단일관련자극 프로토콜과 중다관련자극 프로토콜에서 모두 자극의 비동질성 문제가 나타났으며, 자극의 비동질성 문제가 심각할수록 오류 긍정률이 증가하였다.

Lefebvre 등(2009)은 24명이 포함된 무죄조건에서 부트스트랩 진폭차이검증 결과 오류 긍정률이 29%에 달했다고 보고하고 있지만, P300 CIT를 메타분석한 논문(Meijer, Selle, Elber, & Ben-Shakhar, 2014)에 따르면 모의범죄를 이용한 5편의 연구(총 사례수 40명, 유의수준 .10)에서 부트스트랩 진폭차이검증의 오류 긍정률은 .05로 비교적 낮은 값을 보였다. 자극의 비동질성 문제가 오류 긍정률을 증가시키에도 불구하고 실제 연구에서 그 효과가 나타나지 않았던 이유는 메타분석에 사용한 연구의 수와 전체 사례수가 적었기 때문일 수 있다. Meijer 등(2014)이 메타분석에 포함시킨 5편의 연구와 Lefebvre 등(2009)의 연구를 합하여 오류 긍정률을 산

출하면 14%가 된다. 유죄조건과 무죄조건을 포함하고 있지만 개인별 부트스트랩 진폭 차이검증 결과를 보고하지 않은 연구(Abootalebi et al., 2006; Hu, et al., 2013)들은 약 75%의 정확 판별률을 보고하고 있다. Meijer 등(2014)이 보고한 모의범죄를 사용한 유죄조건인 평균 정확 판별률 .67를 적용하여 오류 긍정률을 추정하면 약 20%가 나온다. 또한 Jung과 Kim(2013)의 Table 5에서 무죄집단에 대한 부트스트랩 지수에 유의수준 .10을 적용할 경우 58%의 높은 오류 긍정률이 나타나기도 하였다. 이상의 결과들을 종합하면 P300 CIT를 사용한 연구의 오류 긍정률은 유의수준을 상회할 가능성이 높다.

선행 연구에서 오류 긍정률이 뚜렷하게 높아지지 않았던 다른 이유도 있을 것이다. 즉 비동질성 문제가 모든 피검사자에서 나타나는 것이 아니며 비동질성의 문제가 조건(관련 자극의 개수, 시행 수 등)에 따라 작을 수 있기 때문일 것이다. 본 연구결과에 의하면 모의범죄를 이용한 경우에 전체 사례의 약 30%에서 자극의 비동질성 문제가 나타났으며 유의수준 .10을 사용하는 경우에 자극의 비동질성 문제는 오류 긍정률을 1~30% 정도 증가시켰다. 따라서 자극의 비동질성 문제는 전체 연구의 오류 긍정률을 약 1~10% 정도 증가시키므로, 사례수가 소수인 실험연구에서는 자극의 비동질성 문제를 명확하게 관찰하기 어려울 수 있다. 연구 상황에서 오류 긍정률이 눈에 띄게 증가하지 않는다고 해서 자극의 비동질성 문제가 중요하지 않다는 것은 아니다. 거짓말 탐지하는 한 개인을 판단하는 문제이므로 P300 CIT가 특정 조건에서 오류 긍정률이 30% 이상 높아질 수 있다는 것은 중대한 결과를 가져올 수도 있기 때문이다.

단일관련자극 프로토콜에서 자극의 비동질성 문제를 해결하는 방법으로 Lafebvre 등(2009)이 사용하였던 절차를 이용할 수 있다. 관련자극에 대한 P300 진폭을 P300 진폭이 가장 큰 무관련자극과 비교하는 것이다. Lafebvre 등(2009)은 무관련자극 전체의 P300 진폭을 사용한 경우에 오류 긍정률이 29%였지만, P300 진폭이 가장 큰 무관련자극을 사용하는 경우에 오류 긍정률이 0%였다고 보고하고 있다. 본 연구결과에 의하면 비동질성 수준이 2이하에서 관련자극의 P300 진폭을 P300 진폭이 가장 큰 무관련자극과 비교하는 경우에 오류 긍정률은 유의수준 .10에서 .016~.052의 범위에 있는 것으로 나타났다. 즉, 단일관련자극 프로토콜에서 높은 수준의 비동질성에서도 유의수준보다 낮은 오류 긍정률을 보였다. 유죄인 사람을 무죄로 잘못 분류하는 오류 부정률보다는 무죄인 사람을 유죄로 잘못 분류하는 오류 긍정률이 더 위험한 상황임을 고려하면 연구자가 설정한 유의수준보다 실제

오류 긍정률이 더 낮은 것은 바람직한 결과라고 할 수 있다.

중다관련자극 프로토콜에서는 각 자극에 대한 시행 수가 많지 않기 때문에 각 자극에 대한 P300 진폭을 산출하는 것이 어렵다. 따라서 관련자극의 P300 진폭을 P300 진폭이 가장 큰 무관련자극의 P300과 비교하는 것이 불가능하다. 이 경우에 두 가지 방법을 이용하여 오류 긍정률을 통제할 수 있을 것으로 생각된다. 한 가지는 중다관련자극 프로토콜을 사용하지 않고 단일관련자극 프로토콜을 관련자극의 개수만큼 실행하는 것이다. 단일관련자극 프로토콜을 여러 번 사용하는 것이 중다관련자극 프로토콜을 사용하는 것보다 P300 CIT의 민감도를 높이는 경향이 있다는 연구도 있으므로 (Rosenfeld et al., 2007) 일석이조라고 할 수 있다. 그러나 전체 검사시간이 그만큼 길어진다는 단점이 있으며 여러 번의 단일관련자극 프로토콜에서 나온 결과를 통합하는 방법의 문제도 있다. 다른 한 가지 방법은 수식 (3)을 이용하여 자극의 비동질성 수준을 측정하는 다음 Table 4를 이용하여 유의수준을 교정하는 것이다²⁾. 예를 들어 관련자극의 개수가 6개, 무관련자극의 개수가 24개이고 자극 당 시행 수가 8회인 P300 CIT에서 수식 (3)으로 계산한 비동질성 수준이 3.00라고 하자. 그리고 연구자는 유의수준을 .10을 사용한다고 하자. 유의수준 .10인 경우 이론적인 오류 긍정률은 .158고 유의수준 .05인 경우 이론적인 오류 긍정률은 .101이므로, 유의수준을 .05로 사용하면 실제로는 오류 긍정률이 10%로 통제될 것이다.

피검사자가 유죄인 경우에도 자극의 비동질성 문제는 P300 CIT의 결과에 영향을 미칠 것이지만, 그 영향을 무죄인 경우처럼 체계적으로 논의하기는 어렵다. 왜냐하면 피검사자가 실제로 유죄일 때 관련자극의 P300 진폭이 무관련자극의 P300 진폭보다 얼마나 더 클지 알 수 없기 때문이다. 그러나 한 가지 분명한 것은 자극의 비동질성 문제에 대해 위에서 언급한 적극적인 대처를 취할 경우 유죄인 사람을 유죄로 옳게 판단하는 민감도가 낮아진다는 점이다. 실제로 단일관련자극 프로토콜에서 관련자극의 P300 진폭을 무관련자극 전체의 P300 진폭과 비교하는 경우에 비하여 관련자극의 P300 진폭을 P300 진폭이 가장 큰 무관련자극과 비교하는 경우에 민감도가 낮아진다는 연구결과들이 있다. Lefebvre 등(2009)의 연구에서는 민감도가 .90에서 .70으로, Meixner와 Rosenfeld(2014)의 연구에서는 .75에서 .33으로, Lee와 Eom, Eum, Sohn(2015)의 연구에서는 .86에서 .57로 낮아졌

2) 이때에는 무관련자극들만 이용하여 자극의 비동질성 수준을 측정하여야 한다. 왜냐하면 피검사자가 실제로 유죄라면 관련자극의 P300은 무관련자극의 P300 진폭과 다를 것이기 때문이다.

다. 이 문제는 오류 긍정률의 문제와 독립적으로 해결책을 모색해야 할 것이다.

연구 3에서 비동질성 수준 0인 경우에 오류 긍정률은 유의수준과 같은 값이 산출되어야 하지만, 유의수준보다 약간 높은 값을 보였다. 이것과 유사한 현상을 Rosenfeld와 Donchin(2015)도 보고하였다. 이 문제는 표본 수가 적을 때 발생하는 부트스트랩 검증의 고유한 문제점으로 보인다 (Good, 2004). 비록 오류의 정도가 작은 수준이기는 하지만 오류가 없는 정확한 판단 방법을 개발할 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.
- Ben-Shakhar, G., & Eiland, E. (2002). The Guilty Knowledge Test (GKT) as an application of psychophysiology: Future prospects and obstacles. In Murray Kleiner (Ed.), *Handbook of polygraph testing*. San Diego: Academic Press.
- 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.
- Deng, X., Rosenfeld, J. P., Ward, A., & Labkovsky, E. (2016). Superiority of visual (verbal) vs. auditory test presentation modality in a P300-based CIT: The Complex Trial Protocol for concealed autobiographical memory detection. *International Journal of Psychophysiology*, 105, 26-34.
- Dietrich, A. B., Hu, X., & Rosenfeld, J. P. (2014). The effects of sweep numbers per average and protocol type on the accuracy of the P300-based concealed information test. *Applied psychophysiology and Biofeedback*, 39, 67-73.
- Eom, J. S. & Park, K. (2014). Effects of the number of trials on test results in P300-based concealed information test. *Korean Journal of Social and Personality Psychology*, 28, 61-79.
- Farwell, L. A. & Donchin, E. (1991). The truth will out: Interrogative polygraphy ("lie detection") with event-related brain potentials. *Psychophysiology*, 28, 531-547.
- Good, P. I. (2004). *Permutation, parametric, and bootstrap tests of hypotheses*. New York: Springer.
- Hu, X., Pornpattananangkul, N., & Rosenfeld, J. P. (2013). N200 and P300 as orthogonal and integrable indicators of distinct awareness and recognition processes in memory detection. *Psychophysiology*, 50, 454-464.
- Johnson, R. (1993). On the neural generators of the P300 component of the event related potential. *Psychophysiology*, 30, 90-97.
- Jung, E. K. & Kim Y. Y. (2013). A comparison of detection accuracy of P300-based guilty knowledge test: Based on bootstrap approach. *The Korean Journal of Cognitive and Biological Psychology*, 25, 75-92.
- Lee, B. H., Hwang S. T., Park, K., Sohn, J. H., & Eom, J. S. (2013). P300-based concealed information test and inter-stimulus intervals (ISIs): A comparison among 500ms, 800ms, and 3000ms ISIs. *Korean Journal of Social and Personality Psychology*, 27, 87-107.
- Lee, M. H., Eom, J. S., Eum, Y. J., & Sohn, J. H. (2015). Effects of countermeasures on P300-based concealed information test with short inter-stimulus interval. *Korean Journal of Social and Personality Psychology*, 29, 91-108.
- 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.
- 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.
- Meixner, J. B., & Rosenfeld, J. P. (2014). Detecting knowledge of incidentally acquired, real-world memories using a P300-based concealed-information test. *Psychological Science*, 25, 1994-2005.
- 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.
- Osugi, A., & Ohira, H. (2017). Emotional arousal at memory encoding enhanced P300 in the Concealed Information Test. *Frontiers in Psychology*, 8, 2334.
- Park, K. (2003). *Analysis of variance and regression*. Seoul: Hakjisa.
- Rosenfeld, J. P. (2005). Brain fingerprinting: A critical analysis. *The Scientific Review of Mental Health Practice*, 4, 20-37.
- Rosenfeld, J. P., Biroshak, 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., 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., 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). The case for peak-to-peak measurement of P300 recorded at 0.3 hz high pass filter settings in detection of deception. *International Journal of Psychophysiology*, *40*, 173-180.
- Verschuere, B., Ben-Shakhar, G., & Meijer, E. (Eds.). (2011). *Memory detection: Theory and application of the Concealed Information Test*. Cambridge University Press.
- 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.

P300 숨긴정보검사의 오류 긍정률

엄진섭¹, 전하정², 손진훈^{2*}

¹충북대학교 심리학과, ²충남대학교 심리학과

P300 숨긴정보검사(P300 CIT)는 무죄조건에서 목표자극을 제외한 나머지 자극들이 모두 동일한 P300을 유발할 것이라고 가정한다. 그러나 P300 이론에 따르면 자극의 동질성 가정이 충족되지 않을 수 있으며, 따라서 P300 CIT의 오류 긍정률이 증가할 수 있다. 이러한 가능성을 확인하기 위하여 일련의 연구를 수행하였다. 연구 1에서는 단일관련자극 프로토콜에서 자극의 동질성 가정이 파기되는 정도를 확인하였다. 모의범죄를 이용한 연구에서는 전체 사례의 약 30%에서 자극의 동질성 가정이 파기되었으며, 기억상실을 이용한 연구에서는 자극의 동질성 가정이 대부분 유지되었다. 연구 2에서는 중다관련자극 프로토콜에서 자극의 동질성 가정이 파기되는 정도를 확인하였다. 연구결과 전체 사례의 약 25%에서 자극의 동질성 가정이 파기되었다. 연구 3에서는 자극의 비동질성 수준(0, 1, 2, 3, 4, 5)과 관련자극의 개수(1, 3, 6), 시행 수(30, 48, 66), 유의수준(.10, .05)에 따른 오류 긍정률을 추정하기 위하여 몬테 카를로 연구를 수행하였다. 자극의 비동질성 수준이 0인 경우에는 관련자극의 개수와 시행 수, 유의수준에 관계없이 오류 긍정률이 적절하게 통제되었다. 그러나 자극의 비동질성 수준이 1 이상인 경우에는 모든 조건에서 오류 긍정률이 증가하였으며, 자극의 비동질성 수준이 증가할수록, 관련자극의 개수가 적을수록, 시행 수가 증가할수록 오류 긍정률이 증가하였다. 유의수준 .10에서 비동질성 수준이 2인 경우에는 오류 긍정률이 12%~25%의 범위를 보였으며, 비동질성 수준이 5인 경우에는 18%~38%의 범위를 보였다. 논의에서 자극이 비동질적인 경우에 오류 긍정률을 통제하는 방법을 제안하였다.

주제어: P300, 숨긴정보검사, 유죄지식검사, 거짓말 탐지, 오류 긍정률