# Psychophysiology in everyday life:

## The Concealed Information Test

松田　いづみ／ロビン・オータイ
Izumi MATSUDA *／ Robin ORTHEY *

### Introduction

The Concealed Information Test (CIT) is a physiological tool to detect hidden crime knowledge in suspects. It is predominantly used by the Japanese police with over 5,000 examinations per year (Hira, & Furumitsu, 2002; Osugi, 2011) upon request by the lead investigator. In the CIT, a suspect is confronted with factual questions about the crime, for example: “What was the murder weapon?”. The suspects do not have to answer the question themselves, but are presented with different answer alternatives one after another, e.g. “Gun”, “Knife”, “Hammer”, etc. One of these answers is always correct while the others are plausible but incorrect. During this procedure several physiological measures are recorded. Typically, autonomic nervous system measures, i.e. sweat, heart-rate, and respiration, are used, but brain activity, ocular motion, as well as behavioural and cognitive measures are also suitable substitutes. The underlying principle of this test relies on a simple but well-established phenomenon in psychophysiology, the oddball paradigm (e.g. Squires, Squires, & Hillard, 1975; Ferrari, 2010). In essence, the oddball paradigm consists out of a set of stimuli, for example tones, and one of the stimuli deviates significantly from the others. The stimulus that stands out among the set is accompanied by an enhances physiological response. The CIT makes use of this paradigm by posing questions and answers that only the perpetrator of the crime, but not innocent bystanders, know. Hence, we expect the perpetrator to recognize the correct answer among the answer alternatives, leading to an elevated response to that specific answer relative to the rest. In contrast, to an innocent examine who has no knowledge of the crime all answers appear equally plausible, leading to a similar response to all answer alternatives.

One must not confuse the CIT with the “polygraph” test that is often depicted in movies and is commonly used in the United States of America. Although the CIT in Japan is also referred to as polygraph by Japanese forensic examiners, the tests used in the USA and Japan

* These two authors contributed equally to this work and share first authorship.

are completely different. The polygraph test used in the USA is also referred to as the Control Question Test (CQT). In the CQT, suspects are confronted with two types of questions, for example “Did you murder your spouse?” or “Have you ever done anything for which you now are ashamed?” The former would be specific about the crime and is a ‘relevant question’, the latter is a general statement that applies to everyone and is a ‘control question’. Prior to the CQT examinees are subjected to a pre-interview that has two goals. First, it is to convince to the examinee that test cannot fail and will detect every lie. Second, the examiner implies that answering ‘yes’ to any of the questions leads to an automatic failure of the test. The reasoning behind this procedure is that examinees will deny all questions, but the control questions are designed so that by answering ‘no’ everyone would lie. CQT proponents believe that examinees are afraid of being caught lying, which leads to an elevated state of arousal. Innocent examinees should only be afraid of the control questions, thus leading to a larger response for control questions as opposed to relevant questions, while guilty examinees are expected to be more afraid of the relevant questions than the control questions. That at least is the general idea behind the test, however, the validity of the CQT is questioned by the general scientific community and we refer readers to a detailed discussion on the limitations of the CQT here: National Research Council (2003); Patrick, & Iacono (1991); Iacono, & Ben-Shakhar (2019).

In contrast to the CQT, the validity of the CIT is well established as demonstrated over numerous studies and two meta-analyses (Ben-Shakhar, & Elaad, 2003; Meijer, et al. 2014). Furthermore, evidence suggests that a CIT conducted in experimental research conditions does not differ from CIT examinations in the field (Zaitsu, 2016), which lends the extensive CIT literature further credibility. Most CIT research is conducted with autonomic measures, but other measures, such as brain waves (e.g. Farwell, & Donchin, 1991), ocular metrics (Lancry-Dayan, et al. 2018), or response times (Verschuere, et al. 2015) are also viable. Yet, these alternative measures do not outperform autonomic measures (Meijer et al., 2014), with Electrodermal Activity (EDA), reflecting psychological sweating, being the most potent. Autonomic measures are relatively easy to measure, which makes the CIT easy to study in the laboratory and apply in the field.

In practice, the CIT is a crucial element in the criminal investigation in Japan (Osugi, 2011). Typically, it is used after initial evidence of the crime has been collected. A CIT examination is requested by investigators and conducted by a forensic examiner. In most cases, the purpose of the CIT examination is to confirm whether a suspect has knowledge of the crime that an innocent person should not have. If the CIT examiner detects that the suspect recognizes the correct answers, the investigators conduct an interview to clarify why that person has this knowledge. In some cases, the CIT has another purpose. If the investigators are sure that the suspect is the real perpetrator, a CIT can be conducted to extract information that the police not yet knows. Instead of developing answer alternatives based on the available evidence the forensic examiner makes reasonable guesses. The logic behind this approach is that if the correct answer is among the reasonable guesses, the correct answer can be identified based on the physiological response (see Matsuda, et al. 2012; Matsuda, et al. 2019). Hence, the CIT is used in practice as either a confirmation tool for crime memory or as an intelligence gathering tool.

We have seen that the CIT is a well validated technique with a vast and active research community and considerable practical impact. Yet, a mostly neglected limitation of conducting CIT research is that a standard CIT requires sophisticated and expensive equipment that is generally only found in dedicated psychophysiology laboratories. Thus limiting the accessibility of this research field. However, recent advances in wearable sensor technology have produced affordable alternatives that have already been applied in consumer research settings. For example, Kim and Fesenmaier (2015) recorded the EDA of two travellers during a four-day trip to Philadelphia with an unobtrusive wristband. The travellers attended serval tourist locations and activities and their EDA was measured each second. After the trip the travellers were debriefed about their experiences and subjective measures were taken for each experience. The results show that the emotional states varied considerably during the activities and this variation was not captured by the traditional subjective measures. Similarly, Shoval, Schvimer and Tamir (2018) conducted a similar experiment with 68 tourists in Jerusalem. They recorded geolocation and EDA data using a device akin to a smartwatch at various tourist spots in and around Jerusalem in addition to standard self-reported measures. The resulting data is complex, but when aggregated one can identify locations and activities that are emotionally evocative, thus of interest for the purpose of tourist activity and memorability. The same devices could be used to conduct CIT studies as well, which would increase the accessibility of CIT research and thereby accelerating advancements in this field. In the following we describe an example of a low cost CIT experiment set-up and present two cases that we recorded with this set-up.

## CIT experimental set-up & methodology

There are three key elements to a psychophysiological set-up: (1) stimulus presentation software, (2) the recording device for the physiological data, and (3) a way to synchronize trigger codes from the presentation software with the physiological data. We present an example of a low-cost experimental set up, but this example is not meant to be exhaustive and there are alternatives for each of the three elements.

First, we must find a way to synchronize the raw data we collect with trigger codes send from the stimulus presentation software. The easiest way is to use a parallel port connection between the stimulus and acquisition computer. However, parallel ports are non-existent in modern computer, so we need another way. Many commercial recording systems come with a software that is able of producing trigger codes on its own, but most of these cannot be linked to stimulus presentation software, trigger codes cannot be customized, and/or triggers must be added manually during the experiment. A circumvent this problem is to send trigger codes to an auxiliary channel of the recording device. This requires a trigger box that converts the digital signal into an analogue signal that can be captured with the auxiliary channel. Trigger boxes, due to their niche use, tend to be costly, and hence are not ideal. Another way would be to include a light sensor to the set up and attach it to a corner of the presentation screen. When a stimulus is presented triggers are send by changing the background colour in the corner of the light switch. This solution however requires an additional sensor. Fortunately, there is a software-based alternative, Lab Stream Layer (LSL; Stenner et al., 2022). Essentially, it is a code library that enable communication between computers using local networks. It can send different data types as streams to the network and it includes a recording program that can capture streams from various sources in the network. Our setup includes a stream for the EDA data and a stream for the trigger codes. The synchronized data is recorded in a single file that can be further processed in, for example, MATLAB.

We have chosen PsychoPy (Peirce et al., 2019) as stimulus presentation software, because it is free to use and comes in two modes. Experiments can be built with a drag-and-drop style builder mode that does not require any coding experience, or experiments can be programmed from scratch. PsychoPy is popular among scientists and it has been extensively validated (Bridges, Pitiot, MacAskill, & Peirce, 2020). It is also capable to interface and send trigger codes via LSL.

As of today, there are several low-cost EDA recording devices available (e.g., under 200,000JPY) . Here we use the Schimmer GSR+ sensor (Schimmer 2022), which can record EDA and Photoplethysmogram (PPG). The device features a python API that allows us to access the raw data via bluetooth and send it via LSL to the network. Data was recorded at a 100Hz sampling rate.

## Verification

We demonstrate the effectiveness of the above set-up using two CITs: A standard CIT with an Inter Stimulus Interval (ISI) of 20-23 seconds and four sets and a short CIT with an ISI of 6-8 seconds and ten sets. Both CITs followed the same make-up and procedure. Each CIT question featured six answer alternatives. One correct answer, four incorrect answers, and one answer to absorb the novelty response, which is an elevated physiological response simply because the stimulus sequence has started. The novelty response item was always presented first in the set of six answer and ignored for analysis purposes. The remaining five answers were presented in random sequence. The participants were instructed to verbally deny each answer, but otherwise remain calm and relaxed.

We transformed the EDA data as follows. First, we extracted the trough-to-peak difference for each answer. We searched for the minimum value in the 1-3 seconds interval following stimulus presentation and the for the maximum value in the interval starting at the minimum up to 5 seconds after stimulus presentation. Then we standardized these values with each set of answer alternatives to eliminate individual difference in EDA responsivity. We also computed an average for each answer alternative over all presentations. We expect that the correct answer will produce elevated EDA relative to all answer alternatives.

**Case 1**

Our first subject was a 26 year old female. The standard CIT question was about her first name and the short CIT was about the first name of her significant other. In figure 1 we can see the standardized responses to each answer over all repetitions for the standard and short CIT. The outcome is clear, the subject's own name elicited a larger response relative to all other names. Similarly, when asked about the name of her significant other we found the same pattern in a CIT with a shorter ISI, but ten repetitions. Again, her significant other's name elicited the largest response overall.

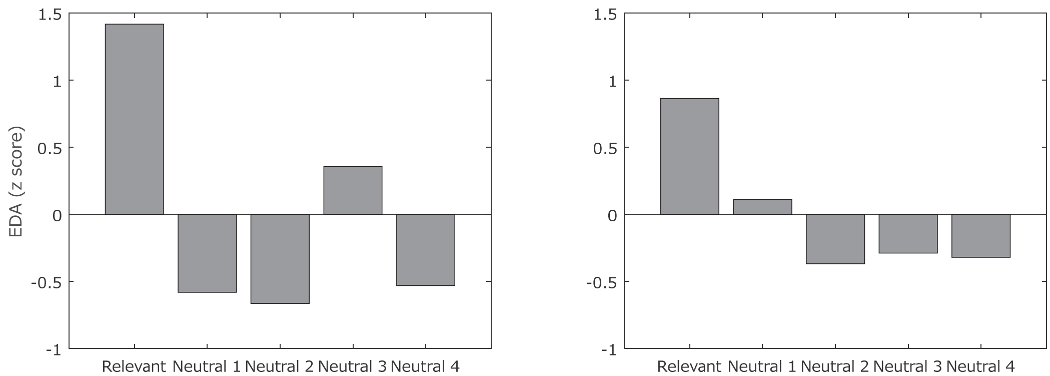


Figure 1. Average EDA in the standard CIT on her own name (left) and the short CIT on a name of her significant other (right) in the first subject

**Case 2**

Our second subject was a 31 year old female. The standard CIT question was about her first name and the short CIT was about the first name of her hometown. The results are displayed in figure 2. The same pattern emerged for this subject. One's own name and the name of one's hometown elicited the largest response among the set of answer alternatives.

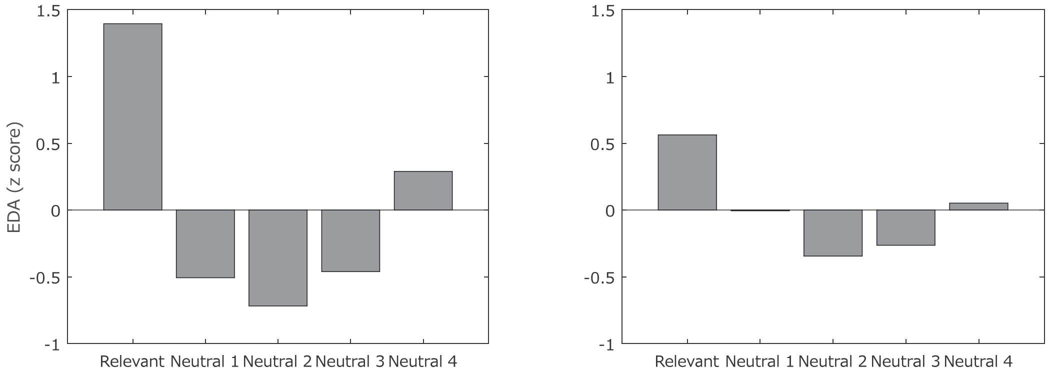


Figure 2. Average EDA in the standard CIT on her own name (upper) and the short CIT on her hometown (lower) in the second subject

## Conclusion

The CIT is a well validated tool, a vibrant research community and has considerable impact in the criminal investigation in Japan. Despite that, the accessibility to conduct CIT experiments is limited due to the high expenses of psychophysiological laboratories. Here, we demonstrate a low cost set-up to conduct a CIT experiment and validate it in two example cases. It is our hope to ease the accessibility to conduct CIT research.

## References

Bridges, D., Pitiot, A., MacAskill, M.R., &Peirce, J. (2020). The timing mega-study: comparing a range of experiment generators, both lab-based and online. *Brain, Cognition, and Mental Health, 8:e9414*, DOI: 10.7717/peerj.9414

Iacono, W.G., & Ben-Shakhar, G. (2019). Current status of forensic lie detection with the comparison question technique: An update of the 2003 national academy of sciences report on polygraph testing. *Law and Human Behaviour, 43(1), 86 – 98*. DOI: 10.1037/lhb0000307

Farwell, L.A., & Donchin, E. (1991). The truth will out: interrogative polygraph ("lie detection") with event-related potentials. *Psychophysiology, 28(5), 531 – 547*.

Ferrari, V.J. (2010). Detecting Novelty and Significance. *Journal of Cognitive Neuroscience. 22(2): 404–411*. DOI:10.1162/jocn.2009.21244.

Kim, J., & Fesenmaier, D.R. (2015). Measuring emotions in real time: Implications for tourism experience design. *Journal of Travel Research, 54, 419 – 429*.

Lancry-Dayan, O.C., Nahari, T., Ben-Shakhar, G., & Pertzov, Y. (2018). Do you know him? Gaze dynamics toward familiar faces on a concealed information test. *Journal of Applied Research in Memory and Cognition, 7, 291 – 302. DOI: 10.1016/j.jarmac.2018.01.011*

Matsuda, I., Nittono, H., & Allen, J.J. (2012). The current and future status of the concealed information test for field use. *Frontiers in Psychology, 3, 532. DOI: 10.3389/fpsyg.2012.00532*

Matsuda, I., Ogawa, T., & Tsuneoka, M. (2019). Broadening the use of the concealed information test in the field. *Frontiers in Psychiatry, 10, 24. DOI: 10.3389/fpsyt.2019.00024*

National Research Council 2003. The Polygraph and Lie Detection. *Washington,DC: The National Academies Press*. DOI: 10.17226/10420.

Osugi A. Daily application of the concealed information test: Japan. In: Verschuere B, Ben-Shakhar G, Meijer E, editors. *Memory Detection: Theory and Application of the Concealed Information Test*. Cambridge: Cambridge University Press (2011). p. 253–75. DOI: 10.1017/CBO9780511975196.015

Patrick, C.J., & Iacono, W.G. (1991). Validity of the control question polygraph test: The problem of sampling bias. *Journal of Applied Psychology, 76(2), 229-238*.

Peirce, J. W., Gray, J. R., Simpson, S., MacAskill, M. R., Höchenberger, R., Sogo, H., Kastman, E., Lindeløv, J. (2019). PsychoPy2: experiments in behavior made easy. *Behavior Research Methods.* 10.3758/s13428-018-01193-y

Shoval, N., Schvimer, Y., Tamir, M. (2018). Real-time measurement of tourists` objective and subjective emotions in time and space. *Foundations of Tourism: A special Series, 57, 3-16*.

Squires, N.K, Squires, K.C., & Hillyard, S.A. (1975). Two varieties of long-latency positive waves evoked by unpredictable auditory stimuli in man. Electroencephalography and Clinical Neurophysiology, 38(4), 387–401. DOI: 10.1016/0013-4694(75)90263-1.

Tristan Stenner, Chadwick Boulay, Matthew Grivich, David Medine, Christian Kothe, tobiasherzke, Giso Grimm, xloem, Arthur Biancarelli, Boris Mansencal, chausner, Jérémy Frey, kyucrane, Samuel Powell, Pierre Clisson, & phfix. (2022). sccn/liblsl: v1.16.0 (v1.16.0). Zenodo. DOI: 10.5281/zenodo.6387090

Verschuere, B., Kleinberg, B., & Theocharidou, K. (2015). RT-based memory detection: Item saliency effects in the single-probe and the multiple-probe protocol. *Journal of Applied Research in Memory and Cognition, 4, 59 – 65*. DOI: 10.1016/j.jarmac.2015.01.001

Zaitsu, W. (2016). External validity of concealed information test experiment: Comparison of respiration, skin conductance, and heart rate between experimental and field card tests. *Psychophysiology, 53, 1100 – 1107*. DOI: 10.1111/psyp.12650