International Journal of Psychophysiology xxx (2015) xxx–xxx

Contents lists available at ScienceDirect

International Journal of Psychophysiology

journal homepage: www.elsevier.com/locate/ijpsycho

# Evidence suggesting superiority of visual (verbal) *vs.* auditory test presentation modality in the P300-based, Complex Trial Protocol for concealed autobiographical memory detection

J. Peter Rosenfeld *, Anne Ward, Vincent Frigo, Jesse Drapekin, Elena Labkovsky

*Department of Psychology, Institute for Neuroscience, Northwestern University, Evanston, IL, USA*

ARTICLE INFO

*Article history:*
Received 23 September 2014
Received in revised form 19 February 2015
Accepted 20 February 2015
Available online xxxx

*Keywords:*
Psychophysiological detection of deception
P300
Event-related potentials
Guilty knowledge tests
Concealed information tests
Lie detection
Credibility assessment
Visual modality superiority
Memory detection

ABSTRACT

One group of participants received a series of city name stimuli presented on trials of the Complex Trial Protocol (CTP) version of a P300-based, concealed information test (CIT). Stimuli were presented on alternating trials in either auditory or visual presentation modality. In 1/7 of the trials the participant's home town (probe) repeatedly appeared in a series of 6 other (irrelevant) repeated city names. In both modalities, probe stimuli produced larger P300s than irrelevant stimuli. Visual stimuli produced shorter behavioral reaction times and P300 latencies, as well as larger P300 probe amplitudes, probe–irrelevant amplitude differences, and individual diagnostic accuracies than the same stimuli presented in the auditory modality. Possible reasons for these effects are discussed, and subject to discussed limitations, the applied conclusion reached is that in all CITs, visual presentation of stimuli, if feasible, should be preferentially used.



## 1. Introduction

In concealed information tests (CITs), an investigator typically seeks to learn if an individual recognizes one or more crime-related items of information (called *probes* or key items), as it is assumed that only guilty (knowledgeable) persons (but not innocent, un-knowledgeable persons) can recognize such information, and thus involuntarily respond to it with an enhanced physiological response. To optimize concealed information detection tests (CITs), which have evolved over the years using increasingly sophisticated methodologies (Rosenfeld et al., 2012a), it is nevertheless important to pin down fundamental testing parameters which may vary across protocols using differing dependent measures; autonomic, imaging, and electroencephalographic. One of the most basic of these fundamental parameters involves the modality chosen for presentation of CIT questions.

In the field, questions are typically put to suspects acoustically, although more recently, and especially in the laboratory use of event-related potentials (ERPs) and imaging, questions are usually presented verbally on a display screen (*e.g.*, Rosenfeld, 2011). To our knowledge, there has been no previous comparison (in a CIT) of the visual–verbal and auditory test modalities, a comparison even more relevant to field use, in as much as field testing now is (as just noted) mostly in the auditory modality. This modality comparison is what we study here.

A specific protocol—the P300-based, Complex Trial Protocol (CTP; Rosenfeld et al., 2008)—was used to detect concealed information in the present study. Its elements are detailed in the Methods section and in Fig. 1. It was chosen for use here as it has been the most accurate and countermeasure-resistant protocol published (Rosenfeld, 2011) based on P300 amplitude size.

We expect the visual modality to produce larger P300s. This is because all parts of a visual stimulus phrase can be exposed completely on a display screen essentially simultaneously, whereas a finite amount of time is required for presentation of a spoken phrase. Thus it should take longer to process the spoken phrase. This means that visual neural activity in response to one stimulus onset, reaches visual cortex more rapidly and synchronously than protracted auditory neural activity reaches auditory cortex. In consequence, there will likely be more P300 latency jitter in the auditory mode, which should reduce average P300 amplitude and onset slope, but increase P300 latency, which is a function of stimulus processing time — among other influences; Duncan-Johnson (1981), Verleger (1997), Leuthold and Sommer (1998). Behavioral response time should also be increased, with longer stimulus processing and response selection. That is, if it takes longer for auditory stimuli to be processed, the total reaction time must increase.

* Corresponding author.

http://dx.doi.org/10.1016/j.ijpsycho.2015.02.026
0167-8760/

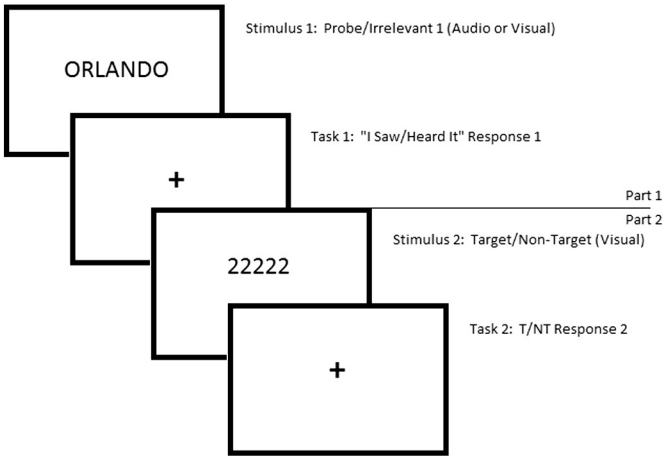


**Fig. 1.** Event sequence in the present Complex Trial Protocol. The two parts of the trial are separated by a horizontal line.

The only paper we found somewhat relevant to the present study utilized an "old–new" paradigm in which words were presented in either an auditory versus visually presented item series (block) that occasionally repeated some of the words ("old"), while ERPs were recorded; this was a paper by Kayser et al. (2003). Obviously, such old words are not, like our home town semantic stimuli, autobiographical items, and are closer to episodic memories, however these data were the most relevant we could find to our situation, and are thus worth brief mention here. Kayser et al. (2003) found that there was no modality difference regarding participants' behavioral abilities to distinguish old and new words. However, the mean response latency was significantly shorter in the visual versus auditory block series. This was attributed to processing time differences as we described above, and supports our expectation of faster visual than auditory RTs in the present study. They also found a parallel difference in what they identified as P300 peak latency (via temporal principal component analysis) which also agrees with our prediction for the present study, likewise based on the view (Duncan-Johnson, 1981) that P300 latency also represents stimulus evaluation time, as well as the duration of perceptual processing (Verleger, 1997). Kayser et al. (2003) also reported larger P300 amplitudes (their "factor 520") for both old and new words in the visual than in the auditory modality. Yet there are important differences between old–new and the present Complex Trial Protocols: 1) old and new are far from being exactly comparable with probe and irrelevant (respectively), and 2) our probe probability is 12.5%, versus the usually equal probabilities of old and new words in an "old–new" paradigm. 3) As already noted, the type of memory we study (autobiographical/semantic) is very different than that studied in old–new protocols (incidentally acquired/episodic). 4) Kayser et al. (2003) had separate auditory and visual blocks whereas we alternated modalities on trials within blocks. 5) Finally, our inter-trial interval (SOA) was at 4 s twice the length of that used by Kayser et al. Thus, this paper is the most relevant we identified, and its results are mostly consistent with our expectations, but the differences just enumerated discourage specific hypothesis building based on its results.

## 2. Methods

### 2.1. Participants

This study had 10 participants, seven females. It was approved by the Northwestern IRB. Participants ranged from 20–30 years of age, and were screened for serious mental disease illness and current use of psychoactive drugs. Five of the participants were Northwestern students in an upper-level physiological psychology lab course. Four of the remaining participants, also northwestern undergraduates, were friends of and recruited by members of this same class. The one remaining participant was a PhD candidate in Northwestern University's Psychology Department. For all participants, involvement was completely voluntary. No participants were disqualified due to failure to follow directions or for reasons relating to lack of response/attention, or excessive artifacts. The study was completely within participant, with all of the participants run in both modality conditions.

### 2.2. Protocol/procedures

All participants were seated in a darkened room, with bridge of the nose approximately 1 m from a display monitor. Visually presented stimulus words (hometowns) were 1.9 cm tall by 5 cm wide, white on black. For auditory presentation, speakers were placed approximately 2 m behind the participants, within easy listening distance, and set to a comfortable volume (about 72 dB) adequate for hearing. The auditory stimuli were presented by a voice synthesizer using Audacity™ software, and had stimulus durations of 100–300 ms, according to the Audacity™ software. This software was used to control for the variability in emotionality and intonation that occurs in a human voice.

A specific protocol—the P300-based, Complex Trial Protocol (CTP; Rosenfeld et al., 2008)—was used to detect concealed information in the present study. This protocol has two parts per trial. In the first part, the participant is presented with either a rare probe or frequent irrelevant. This is responded to with the same, left hand button press (perception acknowledgement) regardless of whether probe or irrelevant was presented. Attention is enforced with unpredictable tests on stimulus identity every 7–12 trials. Participants were warned prior to the run that the unpredictable tests would occur, and that failure on more than one such test would end participation. (This has happened only once in our running of more than 200 participants, and never here.) In the second part of the trial, number strings were presented, either a target (11111) or a non-target (22222, 33333,…55555) was presented and participants were required to respond on a right hand, left mouse button for targets, and on the other mouse button for non-targets. This target discrimination task helped maintain attention as in the original "3-stimulus protocol" (Rosenfeld, 2011). The target/non-target conditional probabilities following either a probe or an irrelevant were .2 and .8 respectively but whether a target or non-target was presented was randomly determined. The CTP was developed to defeat countermeasures of the type typically used to defeat the older, P300-based "3-stimulus protocol" for detecting concealed information (explained in Rosenfeld et al., 2008), and has been largely successful in defeating these countermeasures (Rosenfeld, 2011; Rosenfeld et al., 2013). That is why we chose to use this protocol here.

In this CTP study, participants were accurately told that they would be presented with alternating audio and visual stimuli over successive trials. For visual stimulation, they were instructed to silently read the probe or irrelevant words (cities) as they appeared on the screen and press a button on a computer mouse held in their left hand whenever one appeared, whether it was a city/town they recognized as their home town or not—this perception acknowledgement is referred to as the "I saw it" response. Participants were also told to press this same mouse button whenever they *heard* a word (spoken by an easily understood, computerized voice) from the speakers behind them, again, regardless of whether the word/city was familiar to them or unknown—the "I heard it" response. The probe stimulus was a given participant's hometown. The irrelevant stimuli were other cities: Atlanta, Buffalo, Orlando, Pittsburgh, Stockton, and Wichita. Each of these stimuli, probe and (each of six) irrelevant, was presented (auditorily or visually) on approximately 1/7 of trials in the first ("Part 1") portion of each trial. Again, the modality changed on every trial from auditory to visual, in a regularly alternating series, and the participants had been told before the run about the regularly alternating modality. As noted above, to enforce attention, the participants were repeatedly tested at random intervals (5–15 trials) about the stimulus

Please cite this article as: Rosenfeld, J.P., et al., Evidence suggesting superiority of visual (verbal) vs. auditory test presentation modality in the P300-based, Complex Trial Protocol for concealed..., Int. J. Psychophysiol. (2015), http://dx.doi.org/10.1016/j.ijpsycho.2015.02.026

on the first part of the trial, as the screen went dark following probe/irrelevant presentation. Inter-trial interval (time between first stimuli) here was 4 s, from trial start (first stimulus presentation) to trial start.

For "Part 2" of a trial, participants had a second mouse under their other (right) hand. They were instructed to press the right ("Yes") button on this mouse whenever they saw the target ("11111") on the screen before them, and press the left ("No") button on the mouse when they saw any other numeric stimulus ("22222," "33333", "44444", or "55555". See Fig. 1). Each of the target/non-target stimuli was randomly presented on approximately 1/5 of trials in the delayed ("Part 2") portion of each series. The probability of a target following a probe or any single irrelevant was .2 (a symmetric protocol; Rosenfeld et al., 2009).

Participants were asked to limit their blinking to the intervals between stimuli, when a fixation cross would always appear on the screen before them (see Fig. 1). They were not instructed to employ any countermeasures or in any manner attempt to "defeat" the test. No probe or given irrelevant stimulus was ever immediately repeated during testing. In other words, a visual probe or irrelevant was never followed by the audio version of that same city, and likewise for auditory stimuli.

### 2.3. Data acquisition

EEG recording was taken using Ag/AgCl electrodes attached to the scalp at sites Fz, Cz, and Pz. The electrodes were referenced to linked mastoids. EOG was differentially recorded with Ag/AgCl electrodes above and below the right eye. Eyeblinks were removed using an algorithm based on Semlitsch et al. (1986). Remaining eye movement artifacts were detected, marked, and all trial data containing 100 μV (or more) signals in any EEG or EOG channel were dropped. The diagonal placement of the eye electrodes ensured that both vertical and horizontal eye movements would be picked up, as verified in previous studies; (see Rosenfeld et al., 2004, 2008). The forehead was connected to the chassis of the isolated side of the amplifier system ("ground"). Signals were passed through a Mitsar 19 channel amplifier with a .16 Hz high pass filter setting, and low pass filters set at 30 Hz. Amplifier output was passed to a 16-bit Mitsar A/D converter sampling at 500 Hz. For all analyses and displays, single sweeps and averages were digitally filtered off-line to remove higher frequencies, with the digital filter set to pass frequencies from 0–6 Hz using a "Kaiser" filtering algorithm.

P300 amplitude was measured at Pz, where P300 amplitudes are known to be maximal, (Fabiani et al., 1987), using the peak-to-peak ("p–p") method. (The p–p method has repeatedly been confirmed as the most sensitive in P300-based deception investigations: See Meijer et al., 2007; Soskins et al., 2001). This method searches from 300–650 ms for the maximally positive 100 ms segment; the midpoint of this maximum positive segment is defined as the P300 latency. The average amplitude of the segment from the pre-stimulus baseline is given as the base-peak (b-p) value. For p–p, the algorithm also searches for the maximally negative 100 ms segment between P300 latency and 1300 ms and then subtracts the average absolute amplitude of that segment from that of the maximally positive segment described above. Although other (but similar) search windows have been used in other studies, we, and other researchers, believe it is a poor idea to choose a window for novel studies with novel protocols based on those used in previous studies with different protocols and P300 latencies (Labkovsky and Rosenfeld, 2014). Our present choice of a search window was made based on a grand average of all present participants in all conditions, a procedure recommended by Keil et al. (2014).

ANOVA methods are used here for analysis of group/condition effects, and two effect size estimates are provided, partial eta squared ($\eta_p^2$) and "classical" eta squared ($\eta^2$). Richardson (2011) observed that although the former is preferred currently for many reasons, one cannot compare $\eta_p^2$ values for independent variables in a mixed, higher order ANOVA, for which purpose he suggested using $\eta^2$.

#### 2.3.1. Within individual analysis: bootstrapped amplitude difference method

To determine whether or not the P300 evoked by one stimulus is greater than that evoked by another *within an individual*, the bootstrap method (Efron, 1979) was used on the Pz site where P300 is typically largest. This will be illustrated with an example of a probe response being compared with an irrelevant response. The type of question answered by the bootstrap method is: Is the probability more than 90 in 100 that the true difference between the average probe P300 and the average irrelevant P300 is greater than zero? For each participant, however, each one has only one available average probe P300 and one average irrelevant P300. Answering the statistical question requires separate distributions of average probe and irrelevant P300 waves, and these actual distributions are not available unless one repeats the experiment multiple times which is not feasible. One thus bootstraps these distributions, in the bootstrap variation used here, as follows: A computer program goes through the combined (probe-followed-by target, as well as probe-followed-by non-target) set (all single sweeps) and draws at random, with replacement, a set of n1 probe waveforms. It averages these and calculates P300 amplitude from this single average using the maximum segment selection method as described above for the p–p index. Then a set of n2 waveforms is drawn randomly with replacement from the combined irrelevant set, from which an average P300 amplitude is calculated. The number n1 is the actual number of accepted probe sweeps for that participant, and n2 is the actual number of accepted irrelevant sweeps for that participant multiplied by a fraction (about .125 on average across participants in the present report) which randomly reduces the number of irrelevant trials to within one trial of the number of probe trials. The ranges of n1 and n2 were both 25–40 here. The calculated irrelevant mean P300 is then subtracted from the comparable probe value, and one thus obtains a difference value to place in a distribution which will contain 100 values after 100 iterations of the process just described. Multiple iterations will yield differing (variable) means and mean differences due to the sampling-with-replacement process. (We also use the mean of this distribution here as one dependent variable, as described below.)

In order to state for a given participant with 90% confidence (the criterion used in most preceding studies, *e.g.*, Farwell and Donchin, 1991; Soskins et al., 2001; Rosenfeld et al., 1991, 2004) that probe and irrelevant evoked ERPs are indeed different for a given participant, we require that the value of zero difference or less (a negative difference) not be >− 1.29 SDs below the mean of the distribution of differences. In other words, the lower boundary of the 90% confidence interval for the difference would be greater than 0. It is further noted that a one-tailed 1.29 criterion yields a $p < .1$ confidence level within the block because the hypothesis that the probe evoked P300 is greater than the irrelevant evoked P300 is rejected either if the two are equal or if the irrelevant P300 is found larger. (T-tests on single sweeps are too insensitive to use to compare mean probe and irrelevant P300s within individuals; see Rosenfeld et al., 1991).

We emphasize that optimizing diagnostic accuracy is not our main concern in this report. Here we focus mainly on comparison of auditory versus visual test modalities: The bootstrap measures are used here primarily as dependent variables now described.

### 2.4. Dependent variables

In evaluating the group effects of the critical independent variables of interest, three different (though interrelated) dependent variables were utilized here. First, and obviously, is the Pz p–p P300 amplitude difference in microvolts (P300DF) between probe and irrelevant P300 averages, expected to be large in knowledgeable, but not in unknowledgeable participants. This is a mean computed directly from the present sample of participant data. Additionally, in the intraindividual bootstrapping diagnostic procedure we use (detailed above), means of the iterated bootstrapped average p–p P300s for probe and irrelevant

Please cite this article as: Rosenfeld, J.P., et al., Evidence suggesting superiority of visual (verbal) vs. auditory test presentation modality in the P300-based, Complex Trial Protocol for concealed..., Int. J. Psychophysiol. (2015), http://dx.doi.org/10.1016/j.ijpsycho.2015.02.026

**Table 1**
ANOVA data.

| Measure | Main effect | ANOVA data | Power |
|---|---|---|---|
| P300, p–p Pz | | | |
| | Modality | F(1,9) = NS | |
| | Stim. type | F(1,9) = 66.2, p < .001, $\eta_p^2$ = .88, $\eta^2$ = .66 | .999 |
| | Interaction | F(1,9) = 11.4, p < .009, $\eta_p^2$ = .53, $\eta^2$ = .09 | .299 |
| P300DF | Modality | F(1,9) = 11.8, p < .008, $\eta_p^2$ = .57, $\eta^2$ = .57 | .998 |
| BSMEAN | Modality | F(1,9) = 10.65, p < .02, $\eta_p^2$ = .54, $\eta^2$ = .54 | .997 |
| BSITERS | Modality | F(1,9) = 5.39, p < .05, $\eta_p^2$ = .375, $\eta^2$ = .375 | .991 |
| RT | Modality | F(1,9) = 79.8, p < .001, $\eta_p^2$ = .90, $\eta^2$ = .79 | .999 |

Note: *Power is post hoc as computed from *G*Power 3.1.9.2*.

items are produced in each participant for each iteration, and the mean of these sample means also estimates the *population* mean P300s for probes and irrelevants (Efron, 1979). Thus our second dependent measure utilized here is the difference between such estimated population means for probe and irrelevant (BSMEAN; it correlates >.95 with P300DF). Finally, the most direct measure of diagnostic accuracy in these studies is the number of bootstrapped iterations out of the 100 performed in which the bootstrapped probe P300 (p–p) at Pz for an iteration is greater than that obtained for the bootstrapped irrelevant P300 (p–p) for the same iteration. (This is sometimes called the P > I value, and is here abbreviated to BSITERS.) It is evident that for a *knowledgeable* versus *unknowledgeable* decision to be made in these studies, one usually specifies a *criterion* number of P > I values that must be reached for a *knowledgeable* decision, and thus the higher the P > I value, the greater the likelihood of a *knowledgeable* decision, as described above. In many recent P300 studies (Rosenfeld, 2011), the criterion has been defined as .9; that is, at least 90 out of 100 iterations must yield P > I for a *knowledgeable* decision, although other criteria may be used in some situations; see Rosenfeld et al. (2013). Again, we are not here concerned with diagnostic accuracy *per se*, but with test modality comparisons.

## 3. Results

### 3.1. Behavioral

The mean reaction times ("I saw it" response times) for auditory probe and irrelevant, and visual probe and irrelevant were 697 (sd = 196), 668 (sd = 142), 440 (sd = 112), and 437 (sd = 101) ms, respectively. A completely within-participant 2 (test modality: auditory *vs.* visual) × 2 (stimulus type: probe *vs.* irrelevant) ANOVA on these data revealed an effect (Table 1, row 5) of test modality but no other effects. The "I saw it" response is simply a perception acknowledgement, and does not involve much cognitive effort, so it is unsurprising that probe *vs.* irrelevant would have no effect (see Verschuere and De Houwer, 2011). On the other hand, regarding the test modality effect, the visual stimuli are completely on screen from onset, whereas the acoustic stimuli require 100–300 ms to fully expose (as shown by Audacity™ software during stimulus creation) so that one would expect faster processing and stimulus evaluation for visual stimuli, accounting for the effect of modality obtained.

Stimulus durations were 100–300 ms, according to Audacity™ software.

### 3.2. ERP results, qualitative

The ERP waveforms for both test modalities and stimulus types are shown in Fig. 2 in which the auditory and visual irrelevants appear similar, and both smaller than probe P300s, but the probe–irrelevant differences for visual P300s seem much greater than for auditory P300s, suggesting that visual probe P300s are larger than auditory counterparts. Probe P300 latencies also appear earlier for visual P300s.

### 3.3. ERP results, quantitative

Fig. 3 shows the plotted, computer calculated p–p P300 amplitudes at Pz for the ERPs in Fig. 2. It is evident that the probe P300s are indeed larger for visual than auditory test presentation, whereas the irrelevant P300s are larger for the auditory modality. There is thus a cross-over interaction, suggesting no main effect of modality, and the usual main effect of stimulus type, with probe > Irrelevant P300s. These expectations were confirmed in another, completely within-participant 2 (test modality: auditory *vs.* visual) × 2 (stimulus type: probe *vs.* irrelevant) ANOVA on the four points plotted in Fig. 3. The results are in the top row of Table 1, where effects of stimulus type and interaction are reported, but no main effect of test modality.

To confirm and decompose the interaction, we ran repeated measures t-tests comparing probe P300 and the average of all irrelevant P300s (Iall) in both modalities. In the visual modality, the p–p probe P300 was 17.4 μV. The p–p irrelevant P300 was 7.9 μV. T(9) = 7.1,

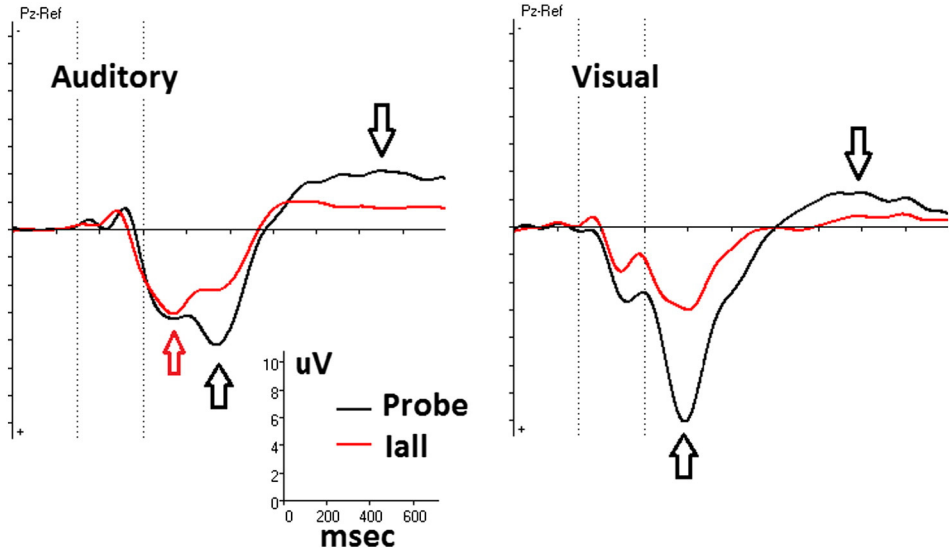


**Fig. 2.** ERPs evoked on alternating trials by auditory (left) and visual(right) stimuli. Probe ERPs shown in black, the average of all irrelevants (Iall) in red. Up arrows show b-p P300 peaks (which differ for probe and irrelevant in auditory modality, as shown by 2 up arrows). Down arrows show computer selected peaks of associated late negative waves (Soskins et al., 2001). P–p P300 is the difference between these peaks. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Please cite this article as: Rosenfeld, J.P., et al., Evidence suggesting superiority of visual (verbal) vs. auditory test presentation modality in the P300-based, Complex Trial Protocol for concealed..., Int. J. Psychophysiol. (2015), http://dx.doi.org/10.1016/j.ijpsycho.2015.02.026

$p < .0001$. The mean difference was 9.5 μV. The auditory average probe p–p P300 was 14.9 μV, the irrelevant average p–p average was 10.4 μV. $T(9) = 4.99$, $p < .001$. The mean difference was 4.48 μV. That the mean differences were different is also shown by the significant interaction in the ANOVA.

Three one-way, repeated measures, ANOVAs were performed on the three major dependent variables, all indexing p–p probe–irrelevant P300 difference; P300DF, BSMEAN, BSITERS. The results are also in Table 1, rows 2, 3, and 4 respectively, and show expected significant effects of modality on the three related indices. (These are all simple 1-way, 2-level ANOVAs so that simple and partial eta squared values are identical.) It is noted that the results (excepting for $\eta^2$) for P300DF are very similar to those for the 2 × 2 ANOVA interaction in row 1. This redundancy is expected since the interaction measures the difference between the probe–irrelevant differences, which defines P300DF. Those differences in microvolts were 8.698 μV for the visual modality versus 3.915 μV for the auditory modality. (Table 1 also shows post hoc achieved power (*G*Power 3.1.9.2*).

Fig. 4 indicates the P300 probe latency differences between the auditory and visual modalities. A 1-way repeated measures ANOVA on these values (auditory = 749 ms versus visual = 690 ms) yielded $F(1,9) = 4.7$, $p = .058$, partial and simple eta squared = .34. This latency effect is also consistent with the stimulus processing time view articulated above regarding reaction time.

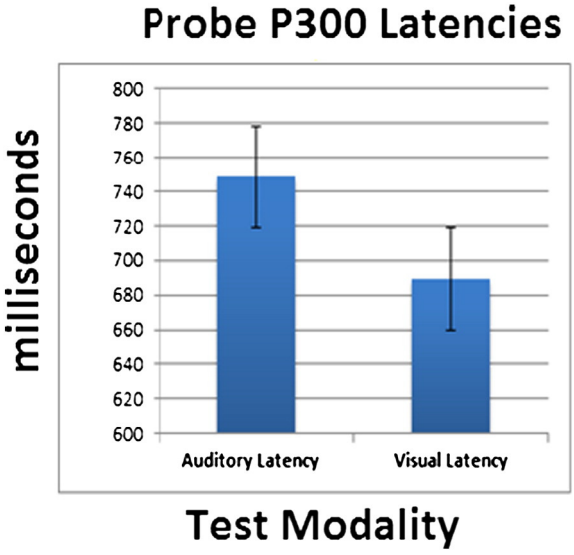


**Fig. 4.** Combined probe and lall P300 latencies in auditory and visual modalities.

*3.3.1. Individual diagnostic effects*

Table 2 shows within each participant the numbers of probe > irrelevant p–p P300 amplitude iterations for the two presentation modalities. These points are plotted in Fig. 5. The auditory mean is 82.8, and the visual mean is 97.7. As shown in Table 1, row 4, a 1-way repeated measures ANOVA yielded $F(1,9) = 5.39$, $p < .05$, partial and simple eta squared = .375. Thus the visual superiority effect is reflected in this most direct measure (BSITERS) of diagnostic accuracy. More intuitively, in only two of the 10 cases (5 and 7) was the criterion-reaching, iteration number greater for auditory presentation, and in both cases, the number for visual presentation was quite close to that of auditory presentation, and both were above the 90% criterion for a guilty diagnosis. In contrast, there were four participants (cases 1, 2 , 6, 10) whose auditory numbers fell below a 90% criterion, and in three of those (1, 2, and 6), the visual number was >.9.

## 4. Discussion

The concluding remarks to follow are qualified by two limitations of the present design: 1) Because the modalities of first stimuli regularly alternated here on each trial, but the target/non-target stimuli were always presented in the visual modality, there was a modality switch preceding auditory but not visual city name presentation. Such modality switching has a cost (*e.g.*, Spence et al., 2000). Thus the modality switching for the auditory stimuli here could have impaired the processing represented by associated probe P300s, which confounds the simple latency jitter interpretation, suggested above, although the findings of visual superiority themselves stand with the present protocol. 2) There were only 10 participants run in the present study (though each in two modality conditions). *Theoretically* oriented P300 studies (which this was not) typically have cell numbers of 12–25. One might then question generality here. However, in many P300 CITs, cell sizes of 10–12 have been deemed adequate; (*e.g.*, Rosenfeld et al., 2012b, also had a cell size of 10 as published in this journal). Moreover, the effect of stimulus type (probe versus irrelevant) is quite reliable and familiar in the visual modality, after dozens of replications (Rosenfeld, 2011; Rosenfeld et al., 2013), and really requires no further support. Indeed, 10 of 10 participants here showed visual probe P300 > irrelevant P300, (and only one of 10 participants in the auditory mode failed to show this relationship). Regarding the modality effect, in only two of 10 participants was the auditory probe–irrelevant P300 difference larger than the visual. Thus, the large effect sizes (as defined by Richardson, 2011) and associated post hoc power values shown in Table 1 were obtained. We agree that this study would have been stronger with a larger cell size, yet the present results seem quite clear.

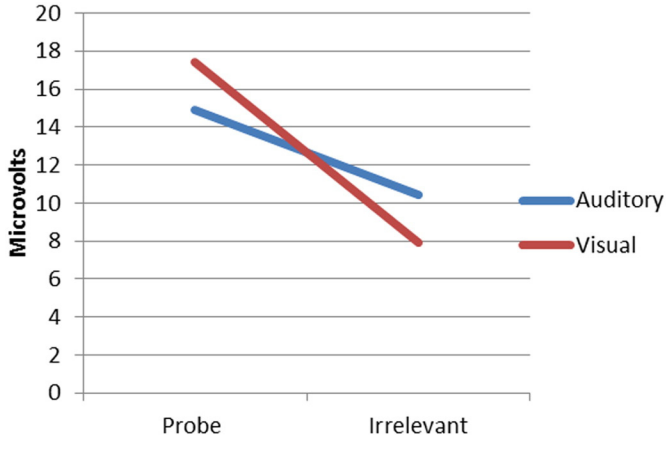


**Fig. 3.** Computer calculated P300 amplitudes in response to probes and lalls presented in auditory (blue) and visual (red) modalities. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

**Table 2**
Numbers (out of 100) of probe > irrelevant P300 iterations, for both test modalities.

| Participant | Significant iterations (audio) | Significant iterations (visual) |
|---|---|---|
| P001 | 49 | 98 |
| P002 | 72 | 100 |
| P003 | 93 | 100 |
| P004 | 93 | 100 |
| P005 | 100 | 96 |
| P006 | 35 | 85 |
| P007 | 100 | 98 |
| P008 | 99 | 100 |
| P009 | 98 | 100 |
| P010 | 89 | 100 |

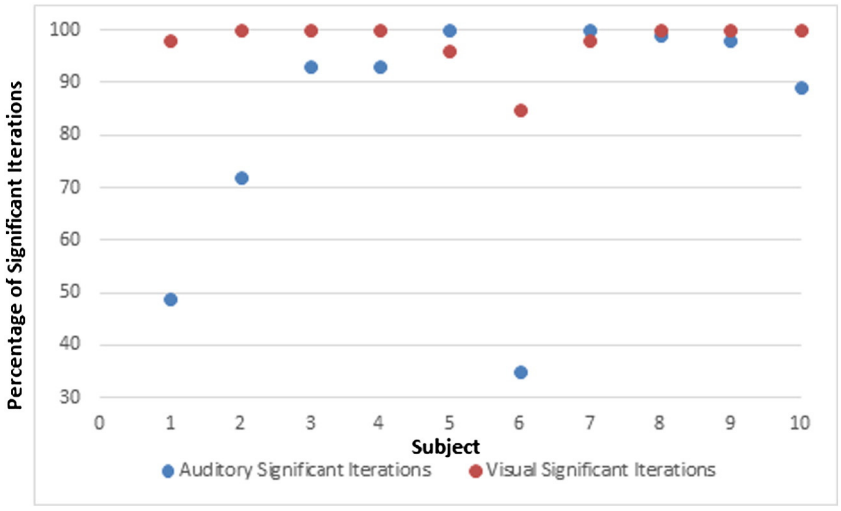


**Fig. 5.** Numbers of iterations in 100 in which the p–p probe P300 was greater than the p–p irrelevant P300, in auditory (blue) and visual (red) modalities. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Apart from the above considerations, there would seem to be little doubt that for CIT protocols which can use a visual presentation mode, such as P300-CITs and most imaging-based CITs, this (visual) mode, with the present protocol, should be preferred, even if benefitting from a modality switch effect as discussed above. (The possible influence of modality switching must be left to a future study.) We are of course aware that in many ANS-based CITs, entire sentences are read aloud to participants; for example "Was the murder weapon a pistol?" We would suggest that even if experimenters or operators believe that it is important to present the entire phrase, the recording epoch begins with the word "pistol," not the beginning of the sentence. If this can be done, there is no reason why a single word stimulus must be presented auditorily, and the present data suggest that all CITs use a visual presentation to achieve the largest responses, whatever they are; ANS or CNS. It is also the case that the use of visual stimuli obviates the need to control for the variability in emotionality and intonation that occurs in a human voice, speaking questions aloud.

The present data are also relevant for the memory literature as our P300 CIT is also a test of recognition of autobiographical (semantic) memory. While our results were mostly consistent with those of Kayser et al. (2003) who made visual versus auditory modality comparisons using an "old–new" recognition protocol, the one difference was our unequivocal finding of greater probe-minus-irrelevant P300 differences in the visual modality, versus their evidence favoring the auditory modality for old versus new P300 differences. Possible reasons for this difference were discussed above, yet it is suggested that future studies look more systematically at the problem. For example, old and new words could be autobiographical and non-autobiographical, as in a new study we are presently planning in which the modalities will be isolated in separate blocks, as in Kayser et al. (2003), versus our present sequence of alternating modalities within one trial sequence/block. (Kayser et al. also used a SOA (2 s), about half the duration of ours, about 4 s).

Reasons why the expected probe amplitude differences obtained were anticipated in the Introduction in terms of the more complete and rapid visual exposure of stimuli, leading to more synchronized arrival of information at the visual cortex, in comparison to auditory processing. The fact that visual stimulation produced both increased auditory P300 latency as well as increased RT suggests that both stimulus evaluation time and response selection (respectively) are faster (Duncan-Johnson, 1981) with visual presentation since the visual stimulus is completely exposed virtually instantaneously, as opposed to the auditory stimuli which require 100–300 ms to fully present.

## References

Duncan-Johnson, C.C., 1981. P300 latency: a new metric of information processing. Psychophysiology 18, 207–215.
Efron, B., 1979. Bootstrap methods: another look at the jackknife. Ann. Stat. 7 (1), 1–26.
Fabiani, M., Gratton, G., Karis, D., Donchin, E., 1987. Definition, identification, and reliability of measurement of the P300 component of the event-related brain potential. Adv. Psychophysiol. 2, 1.
Farwell, L.A., Donchin, E., 1991. The truth will out: interrogative polygraphy ("lie detection") with event-related brain potentials. Psychophysiology 28 (5), 531–547. http://dx.doi.org/10.1111/j.1469-8986.1991.tb01990.x.
Kayser, J., Fong, R., Tenke, C.E., Bruder, G.E., 2003. Event-related brain potentials during auditory and visual word recognition memory tasks. Cogn. Brain Res. 16 (1), 11–25.
Keil, A., Debener, S., Gratton, G., et al., 2014. Committee report: publication guidelines and recommendations for studies using electroencephalography and magnetoencephalography. Psychophysiology 51 (1), 1–21.
Labkovsky, E., Rosenfeld, J.P., 2014. A novel Dual Probe Complex Trial Protocol for detection of concealed information. Psychophysiology http://dx.doi.org/10.1111/psyp.12258 (Article published online: 1 JUL 2014).
Leuthold, H., Sommer, W., 1998. Postperceptual effects and P300 latency. Psychophysiology 35 (01), 34–46.
Meijer, E.H., Smulders, F.T., Johnston, J.E., Merckelbach, H.L., 2007. Combining skin conductance and forced choice in the detection of concealed information. Psychophysiology 44 (5), 814–822.
Richardson, J.T., 2011. Eta squared and partial eta squared as measures of effect size in educational research. Educ. Res. Rev. 6 (2), 135–147.
Rosenfeld, J.P., 2011. P300 in detecting concealed information. In: Verschuere, B., Ben-Shakhar, G., Meijer, E. (Eds.), Memory Detection: Theory and Application of the Concealed Information Test. Cambridge University Press, New York, NY, pp. 63–89.
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 (3), 319–335. http://dx.doi.org/10.1111/j.1469-8986.1991.tb02202.x.
Rosenfeld, J.P., Soskins, M., Bosh, G., Ryan, A., 2004. Simple, effective countermeasures to P300-based tests of detection of concealed information. Psychophysiology 41 (2), 205–219. http://dx.doi.org/10.1111/j.1469-8986.2004.00158.x.
Rosenfeld, J.P., Labkovsky, E., Winograd, M., Lui, M.A., Vandenboom, C., Chedid, E., 2008. The Complex Trial Protocol (CTP): a new, countermeasure-resistant, accurate, P300-based method for detection of concealed information. Psychophysiology 45 (6), 906–919. http://dx.doi.org/10.1111/j.1469-8986.2008.00708.x.
Rosenfeld, J.P., Tang, M., Meixner, J., Winograd, M., Labkovsky, E., 2009. The effects of asymmetric vs. symmetric probability of targets following probe and irrelevant stimuli in the complex trial protocol for detection of concealed information with P300. Physiol. Behav. 98 (1), 10–16.
Rosenfeld, J.P., Ben-Shakhar, G., Ganis, G., 2012a. Detection of concealed stored memories with psychophysiological and neuroimaging methods. In: Nadel, L., Sinnott-Armstrong, W. (Eds.), Memory and Law. Oxford University Press, pp. 263–303.
Rosenfeld, J.P., Hu, X., Pederson, K., 2012b. Deception awareness improves P300-based deception detection in concealed information tests. Int. J. Psychophysiol. 86 (1), 114–121.
Rosenfeld, J.P., Hu, X., Labkovsky, E., Meixner, J., Winograd, M., 2013. Review of recent studies and issues regarding the P300-based, complex trial protocol for detection of concealed information. Int. J. Psychophysiol. 90, 118–134.
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 (6), 695–703.

Please cite this article as: Rosenfeld, J.P., et al., Evidence suggesting superiority of visual (verbal) vs. auditory test presentation modality in the P300-based, Complex Trial Protocol for concealed..., Int. J. Psychophysiol. (2015), http://dx.doi.org/10.1016/j.ijpsycho.2015.02.026

Soskins, M., Rosenfeld, J.P., Niendam, T., 2001. The case for peak-to-peak measurement of P300 recorded at .3 Hz high pass filter settings in detection of deception. Int. J. Psychophysiol. 40, 173–180.

Spence, C., Nicholls, M.E.R., Driver, J., 2000. The cost of expecting events in the wrong sensory modality. Percept. Psychophys. 63, 330–336.

Verleger, R., 1997. On the utility of P3 latency as an index of mental chronometry. Psychophysiology 34 (2), 131–156.

Verschuere, B., De Houwer, J., 2011. Detecting concealed information in less than a second: response latency-based measures. Memory Detection: Theory and Application of the Concealed Information Testpp. 46–62.

Please cite this article as: Rosenfeld, J.P., et al., Evidence suggesting superiority of visual (verbal) vs. auditory test presentation modality in the P300-based, Complex Trial Protocol for concealed..., Int. J. Psychophysiol. (2015), http://dx.doi.org/10.1016/j.ijpsycho.2015.02.026