Protocol

Disruption of immediate memory and brain processes: an auditory ERP protocol

Tom Campbell\textsuperscript{a,b,c,*}, István Winkler\textsuperscript{b,d}, Teija Kujala\textsuperscript{a,b,c}

\textsuperscript{a}Helsinki Collegium for Advanced Studies, University of Helsinki, Finland
\textsuperscript{b}Cognitive Brain Research Unit, University of Helsinki, Finland
\textsuperscript{c}Helsinki Brain Research Centre, University of Helsinki, Finland
\textsuperscript{d}Hungarian Academy of Sciences, Budapest, Hungary

Accepted 1 November 2004
Available online 7 January 2005

Abstract

An event-related potential (ERP) protocol is described that can be used to investigate those sound-evoked neural processes that may be implicated in disrupting immediate memory. Conventional electroencephalogram (EEG) is recorded during the performance of a task that involves ignoring irrelevant sounds while trying to hold in memory lists of numbers. Specific bioelectric measures are made to prevent the contamination of recordings by the movements of articulators. An approach is also outlined which controls the timing of ERP components to sounds with different envelopes. Using this approach, it has been shown that the neural processes involved in the elicitation of the auditory N1 ERP response may be involved in the disruption of memory for serial order produced by irrelevant sound.

© 2004 Published by Elsevier B.V.

Theme: Neural basis of behavior
Topic: Cognition

Keywords: Irrelevant sound effect; Auditory event-related potentials; Working memory

1. Type of research

1.1. Investigating ERP and EEG during cognitive (working memory) and auditory perceptual tasks

Working memory is the capacity to maintain and manipulate information temporarily in memory, typically in the service of some particular goal or task, and the influence of these tasks on human brain processes has been investigated using electroencephalogram (EEG) \cite{26} and event-related potentials (ERPs), \cite{41–43}.

1.2. Investigating the neural basis of memory disruption by irrelevant sounds

See Refs. \cite{9,41–43}.

1.3. Auditory event-related potentials to speech sounds

ERPs elicited in the scalp-recorded EEG by auditory stimuli enable noninvasive investigation of the auditory information processing within the human brain to a temporal resolution in the order of milliseconds (for reviews, see Refs. \cite{24,65}). Each response within the auditory ERP is characterized by a particular time course. The genuine onset of speech sounds in notoriously ambiguous and this can obfuscate the subsequent analysis of ERPs. However, it is expected that within the first 10–12 ms after the sound’s onset, responses known as brain stem evoked potentials (BAEPs) occur, followed by middle-
latency responses (MLRs) between 10 and 40 ms, after which emanate the long-latency responses (LLRs).

Responses are typically labeled according to the polarity of the potential difference between the reference and active electrode at the vertex (Cz); P for positivity, N for negativity. First, MLRs run a temporal sequence No, Po, Na, Pa (P30, peaking ca. 30 ms), Nb; followed by LLRs, P1 (peaking ca.50 ms), N1 (peaking ca.100ms), P2 and then N2. Upon repeated presentation of a sound, LLRs are particularly subject to an attenuation which is subject to recovery after a period of silence that is termed refractoriness ([1,2,50]).

2. Time required

Participant preparation and explanation for protocol: 25 min.
Data collection: 11.5 min per experimental condition per participant.
Data analysis routines: 20 min to run per participant on modern hardware.
Total time required: 1 h 40 min per participant, assuming 3 experimental conditions.

3. Materials

3.1. Special equipment

Equipment is required for presentation of auditory and visual stimuli, together with that for the recording and analysis of multi-channel digital EEG from human volunteers. Geodesic electrode arrays with an adequate representation of electrodes on the lower parts of the head are recommended [63] (Lectron Inc., Finland), as is an electrically and acoustically shielded chamber.

3.2. Chemicals and reagents

Reagents for EEG electrode preparation are required.

4. Detailed procedure

4.1. Participants

This protocol should be administered to a large number of participants with intact hearing, alongside normal or corrected-to-normal vision. A record of the handedness, sex, and age of the participants is also advised (in the data reported, 18 right-handed participants; 3M, 18–27 years of age).

4.2. Preparation

Participants are informed about the nature and possible consequences of the experiment (task, EEG recording) after which they give written informed consent in accordance with the Declaration of Helsinki. EEG is set up so as to be recorded with a 30-channel array of electrodes evenly distributed across the scalp, [63], as illustrated in Fig. 1. The reference electrode is attached to the tip of the nose. A bipolar EOG channel–composed of two electrodes attached laterally to the outer canthi of the eyes–is set up to monitor horizontal eye movements. No additional EOG channel is required to measure vertical eye movements, which are monitored using the prefrontal electrodes (Fp1, Fp2, Fpz) on the cap against the common reference. Equipment for recording these bioelectric potentials is configured such that they are amplified within frequency limits (0–30 Hz) and digitized online at digitization rates appropriate to the ERP components of theoretical interest. In order to record mouth and tongue movements, an additional pair of
electrodes is placed on the submandibular surface and the right masseter muscle. The signals picked up by this bipolar montage should be amplified and filtered (0–30 Hz) differently from the EEG and EOG channels.

4.3. Task

The structure of one trial from the task is shown in Fig. 2. Five seconds before the onset of each trial, a fixation cross (+) appears in the center of the computer screen. After 2 s, a 100-ms tone is presented to warn participants that a list of to-be-remembered items is about to be presented. The screen is blank for 200 ms following the offset of the fixation cross and then a list of to-be-remembered material is presented. Such a list, for example, may include the digits 1–9 in a random order. These to-be-remembered items are presented for 800 ms with an inter-item interval of 200 ms when the screen is blank. Participants are required to attend to these digits.

A 10.5-s retention interval follows list presentation, during which the word “WAIT” appears in the center of the screen. The precaution is taken that the timing of onsets and offsets of stimuli are not in audiovisual alignment. During this period, participants are required to silently rehearse the list items and to ignore any sound that they hear.

After the offset of the word “WAIT”, the screen is blank again for 1 s, and then the word “WRITE” appears for 10 s. During this period, participants are required to write down the digits in a strictly left to right fashion, without correction, while attempting to preserve the correct serial order and position of items, leaving a “/” when uncertain about an item. A response grid in which there is a column to write a response for each serial position is recommended to reduce contamination of responses by knock-on errors [23]. The “WRITE” period is followed by the fixation cross that precedes the next trial. Each trial thus lasts 36.5 s in total.

Participants initiate the first trial by saying that they are ready; subsequent lists are then presented at regular intervals. If more than one condition is presented, the order of conditions is randomized, each condition receiving 20 trials.

To-be-remembered items are presented in the center of the computer screen, at a size of 45 × 55 mm in white Helvetica font on a black background. Viewing distance is 1.50 m. The irrelevant stimuli are delivered binaurally via headphones. During the procedure, participants are seated in an acoustically and electrically shielded room.

Behavioral responses are scored with a strict serial position criterion. The irrelevant sound effect is characterized by an increase in the mean probability of an error scored with this criterion as a consequence of the auditory stimulation. The size of the effect depends on the makeup of the sequence of task irrelevant auditory stimuli presented during the retention interval [5,7–10,13,22,27–44,52,53,59–62].

4.4. Auditory stimulus construction

It is recommended that the ignored sequences consist of large numbers of sounds separated by brief intervals of silence and that speech sounds are piloted to ensure that they elicited homogeneous ERP responses (see Section 6.2.1). In the data reported, irrelevant sound sequences consisted of 30 meaningless syllables–either jus[jus], kaes[kaes], tam[tam], nev[nev] or poi[poi]–of 280 ms duration separated by a silent interstimulus interval of 70 ms. Sequences differed in the number of different types of syllable delivered within a sequence ("token set size"). Three different token set sizes were tested: a 1-token (AAAAA...), a 2-token (ABABA...), and a 5-token condition (ABCDEABCDE...).

4.5. ERP analysis

EEG is filtered offline and epochs starting 100 ms before the onset of each sound are taken. Epochs with EEG or EOG exceeding ±50 μV in any channel are omitted from further analyses. ERPs are averaged separately for each experimental condition.

ERPs elicited by the irrelevant items are digitally re-referenced to the average of all scalp electrodes offline [64]. ERP components are measured by integration windows centered on the peak of the corresponding wave in the group averaged responses. Amplitude measurements are referred to the mean voltage during the 100-ms
prestimulus period, though variants of this approach are possible (see Fig. 3). To locate the brain areas involved in generating the ERP components, source modeling (e.g., BESA, etc.) can be used. Other approaches include scalp current density analysis, or alternative methods such as the Hjorth-transform [25], available in Neuroscan 3.2.43, which constitutes an approximation to a discrete Laplacian which allows one to reconstruct radially oriented current sources and sinks.

5. Results

The results showed that increments in set size from 1 to 2 as well as from 2 to 5 resulted in an increase of the N1 amplitude (Fig. 3). Hjorth transforms [25] revealed that token set size related increases in N1 amplitude did not change the cortical generator configuration of N1 response (Fig. 4). Furthermore, increases in set size from 2 to 5, but not from 1 to 2, resulted in memory disruption (Fig. 5).

The interpretation offered is that memory disruption occurs when the N1 signal exceeds a certain threshold. This only occurred in the 5-token condition of the experiment described.

6. Discussion

6.1. Overall assessment of the protocol

This protocol is an ERP approach to determine which components may be related to the disruption of immediate memory produced by irrelevant ignored sound.

---

Fig. 3. Grand-averaged ($N = 18$) ERPs elicited by irrelevant sound items at selected electrodes. Windows around P1, N1, and P2, shown for Fz and Cz, represent the windows of integration within which amplitude measurements were taken. A line is drawn between the averaged peaks for P1 and P2 within these windows. The measure of N1 used is the shaded area between this line and the ERP wave within the window of integration.
The protocol can be implemented within an electrically and acoustically shielded ERP laboratory equipped for auditory and visual stimulation. It is, of course, possible to use different electrode caps or smaller numbers of electrodes. However, the use of small numbers of electrodes precludes the use of Hjorth transforms [25] or other methods of source localization, and an adequate distribution of electrodes beneath the horizontal plane is required for applying the common average reference, which offers both a valid and reliable measure of the auditory N1 [63,64]. For the relative advantages of alternative approaches to the reference problem, see Fisch and Spehlmann [18]. Recourse to reference-independent Hjorth-transform maps (see Fig. 4) may resolve ambiguities of interpretation related to the choice of reference [25]. In addition to the well-known advantages of the ERP method, recording EEG has an advantage in this task over other noninvasive brain research methods: writing and repositioning during the response phase of the task do not produce substantial artefact.

6.2. Alternative/support protocols

6.2.1. Recording, editing, and piloting auditory stimuli

The experimental hypothesis often concerns the influence of the envelope of the irrelevant sounds, which are commonly speech sounds. In our experiment, speech stimuli were high-quality digital recordings (44100 Hz, 16-bit) of the utterances of a native speaker of the participants’ native language. Sounds should be balanced for acoustic factors that are not related to the experimental hypothesis.

6.2.1.1. Recording. High-quality digital mono-recordings (32-bit, 44100 Hz) of utterances may be made via a microphone directly to a computer. Audio software for recording and editing such as Goldwave, http://www.goldwave.com, permits the sounds to be stored in a waveform format, which is a record of the microphone’s gain as a function of time in response to these utterances. This waveform is a series of numbers recorded at regular time intervals known as samples, which represent the deflection of the microphone receiver with respect to a zero point at rest. The waveform may be edited and used for reproduction of the sound through headphones attached to a computer.

Recordings are made of utterances spoken with a monotone by one talker, repeating a cycle of all the different sounds in fixed order. For 2 min, this sequence is repeated at a regular rhythm at which the talker may breathe while...
talking, such that utterances are separated by an intervening period of silence. The purpose of repeating the cycle is to reduce the influence of rises and falls in intonation at the start and end of a sequence of utterances. From the recordings, it is necessary to identify a cycle, somewhere in the middle of the recording, where all the sounds that are spoken have roughly similar objective duration and even subjective pitch. Sounds where the utterance was so loud that the peaks of the waveform are clipped should be avoided. If there is no such cycle, then new recording sessions are required until such a cycle is found. The objective duration may be measured by inspection of the sound waveforms with audio-editing software. Subjective pitch may be gleaned by listening to the sound recordings. At this point, each utterance should be placed into a separate sound file, removing periods of silence.

6.2.1.2. Editing

6.2.1.2.1. Balancing for fundamental frequency. Inspection of sound waveforms should reveal that the vowel portion of the utterances is characterized by a periodic oscillation. This duration reflects the period of oscillation of vibrating vocal folds during phonation. There should be a section of the waveform where the period of oscillation seems to be relatively constant. With the digital-editing software, the period of this oscillation may be measured; the fundamental frequency is the inverse of this period in seconds. To verify that the fundamental frequency rather than that one of its harmonics is measured, the period of this oscillation may be measured; the waveform for the utterance, and then to take the square root of this value to yield a positive $G_{RMS}$ for that utterance as described in Eq. (1).

$$G_{RMS} = 2 \sqrt{\frac{\sum_{x=1}^{n} G(x)^2}{n}}$$

$G_{RMS}$ should be calculated for each utterance and the $G_{RMS}$ of the utterance with the lowest $G_{RMS}$ should be identified as the intended $G_{RMS}$. For each utterance, to reduce the actual $G_{RMS}$ of the utterance to the intended $G_{RMS}$, for each sample in that utterance, the gain of that sample should be multiplied by the intended $G_{RMS}$ of that utterance and then divided by the actual $G_{RMS}$. A subjective check that the loudness of all the utterances is equivalent is recommended at this stage. Looping instances of one utterance played at short constant stimulus onset asynchronies should be played back through the laboratory apparatus, in an acoustically shielded chamber through headphones coupled to a sound level meter, set on a slow response mode, such that the sound level displayed reaches a stable level after a several presentations of the utterance. The waveform of each utterance may be then be empirically adjusted until the utterances are of equivalent intensity when played back in this way (for instance, with the Volume change effect in Goldwave).

Caution should be taken throughout that the utterances still maintain the same syllabic identity throughout, and this should be verified by native listeners.

6.2.1.2.3. Controlling duration. When the sounds are not of the desired duration, it is necessary to select one oscillation in the vowel section of the utterance that has a period which corresponds to the fundamental frequency. At the end of that selection of the waveform, it is necessary to insert looped versions of this oscillation into the recording of the utterance, such that there are no sudden changes in the waveform. The correct number of oscillations should be inserted until the duration of the stimulus exceeds the desired duration slightly. Sounds should be edited such that the onset and the offset of the waveform pass through zero-crossings and is of the desired duration, by removing a section of the waveform and, where necessary, by linear attenuation of the gain of the waveform over a 5 ms window such that the waveform attenuates towards a zero-crossing at onset and offset (e.g., by use of fade-in and fade-out features, respectively, in Goldwave).

6.2.1.2.3. Controlling intensity. It is necessary to equalize the RMS gain $G_{RMS}$ of the waveforms for each utterance, which is related to the overall intensity of the utterance. An expedient approach is, for each sample $s$, to multiply the gain $G(s)$ by itself, then take the mean of this squared gain $G(s)^2$ over each of the $n$ samples which constitute the waveform for the utterance, and then to take the square root of this value to yield a positive $G_{RMS}$ for that utterance as described in Eq. (1).
randomized such that each successive train contains a different sound. In total, 400 stimuli per sound envelope is sufficient to identify the timing of the ERP components for a given sound envelope.

As appropriate, bioelectric potentials are digitized online, EEG is then filtered offline and epochs averaged after artifact rejection (the epoch related to the first sound in each block and epochs with EEG or EOG exceeding ±50 μV in any channel) for each sound envelope.

As illustrated in the upper panel of Fig. 6, the ERPs to the “kaes” sound were delayed with respect to the other stimuli, as reflected in a delayed latency of the peak of the N1 component. This delay was taken to reflect that the stop release burst in “kaes” was not sufficiently intense to influence auditory ERPs and occurred primarily in response to the onset of voicing. This EEG pilot procedure was repeated in an identical fashion with the exception that the relative intensity of the stop release burst in “kaes” was increased and that a different participant was used. As depicted in the lower panel of Fig. 6, the ERP response elicited by “kaes” is now temporally aligned with that of the other syllables. This simple iterative process of digital audio-editing and psychophysiological piloting thus proved an expedient solution to the problem of temporally aligning ERPs to auditory stimuli with different envelopes.

The ERP findings of the pilot procedure should be considered. If there are discrepancies in the latency of components of interest in response to particular sounds, then potential causes of that discrepancy (e.g., envelope) should be identified and altered. It is necessary to rejoin the process of digital audio-editing and psychophysiological piloting.

Fig. 6. Individual ERPs to five syllables. The stop release burst in “kaes” [kæs] was of a greater relative intensity in the lower than in the upper panel. Different participants were used in each measurement.
of editing at a stage such that this alteration of discrepant stimuli (e.g., envelope) is otherwise corrected for other factors (e.g., intensity and duration) that may be influenced by that alteration before re-piloting the stimuli.

A qualitative criterion is adopted for when the auditory stimuli may be considered as ready for use in the irrelevant sound effect procedure. That is, the components of interest should temporally align within the ERP responses to each utterance, as measured from one participant in this pilot procedure (e.g., Fig. 6, lower panel).

6.2.2. Other distraction paradigms

Memory disruption is a form of distraction. For other ERP approaches to auditory distraction, see: Refs. [3,4,14–17,21,51,54–57,66].

6.2.3. Alternative brain research methods

Measurement of evoked magnetic fields can complement ERP findings of the irrelevant sound effect protocol described here [9,61,62]. Magnetoencephalographic recordings offer the temporal resolution of ERP combined with superior possibilities for locating component generators. However, magnetoencephalography (MEG) is less sensitive to the radially oriented component of electromagnetic fields, thus neglecting part of the neural activity. However, because a major part of sound-evoked neural activity (such as the supratemporal N1 response) occurs within the Sylvian fissure, MEG is especially well suited for studies investigating auditory event-related responses (see Ref. [49] for a review).

An alternative protocol derives ERP responses elicited by the to-be-remembered items rather than the irrelevant sounds [41–43]. Studies using this protocol presented the irrelevant sound alongside the to-be-remembered material revealing effects of maximal amplitude between 300 and 400 ms after the onset of the to-be-remembered list items. However, caution is warranted when relating the ERP effects observed in these studies to the disruption produced by irrelevant sounds that are presented a matter of seconds after the list of to-be-remembered items [7–9,41–43]. Indeed, some of the crucial theoretical issues are at stake when considering the relation between the processing of to-be-remembered and irrelevant material within the brain.

A different promising approach that has tackled some of these theoretical issues related to memory disruption is the coherence analysis of EEG data. Coherence is an index of the degree of linear synchronization of activity of different channels measured at different locations on the scalp [48,58]. When two channels show high coherence, this indicates that the activity in those channels is either related to a unitary brain system or multiple “coupled” systems that co-operate at that time. EEG coherence analyses have offered new insights that show the disruption of memory produced by the irrelevant sound effect occur alongside a decrease in synchrony within the gamma-band (35–47 Hz) [35].

Hemodynamic techniques, such as functional Magnetic Resonance Imaging (fMRI) and Positron Emission Tomography (PET), offer excellent spatial resolution at the sacrifice of temporal resolution. While offering slightly more accurate temporal resolution than PET, fMRI, however, suffers an additional disadvantage in this context in that loud scanner noise is typically heard by the participant during fMRI recordings. Scanner noise has been shown to influence the amplitudes and latencies of certain ERP components [46,47] and would be predicted to attenuate the extent of the irrelevant sound effect [11,34]. However, findings from subtractive PET designs have indicated that ignored sound suppresses a network of brain areas [19,20] and when that sound disrupts performance, then the dorsolateral prefrontal cortex is among the activated brain regions [6,21].

6.3. Troubleshooting

The procedure involves EEG recordings of brain activity; therefore, the normal technical issues associated with recording EEG and ERP apply to this protocol. If conventional criteria of artefact rejection and subject exclusion proved insufficient, experienced independent ERP experts, with no knowledge of the experimental hypothesis, should identify which individual ERPs should be excluded from the group analyses due to suspected artefact or technical failure.

The issues that are specific to this protocol are:

1. It should be ensured through an iterative process of piloting and sound editing that these components of interest temporally align for all the sound envelopes used.
2. This procedure involves a cognitively demanding task. It is recommended that the participants read written experimental instructions before the experiment and queries should be fully answered prior to preparation for EEG recordings. It is especially important to encourage the participant to relax in the experimental situation in order to reduce the possibility of artefacts due to muscle tension and teeth-gritting. Relaxation may be promoted by seating the participant in a recliner under controlled conditions of dim artificial lighting during the procedure.
3. Some participants are inclined to articulate the to-be-remembered items during the retention interval. This tendency should be politely discouraged before the onset of the experiment, to preclude the rejection of epochs due to muscle artefacts.
4. Evidence has indicated that there are long-term individual differences in the susceptibility to disruption by irrelevant sound [12]. The use of large samples of participants is thus recommended to bring the irrelevant sound effect to significance. Initial null results with small samples can be misleading, when similar stimuli with a larger sample can produce a significant disruption. Indeed, a focus on those participants that are particularly susceptible to disruption may prove to be a pragmatic approach [35].
7. Essential literature references

Original article: Campbell et al., Cogn. Brain Res. 2003, 18, 39–47 [9].

8. Quick procedure

8.1. Preparation

Participants give their informed consent and are prepared for bioelectric recordings with equipment configured to measure mouth movements, 30-channel EEG, and EOG in the conventional manner and then positioned for these recordings within an electrically and acoustically shielded chamber.

8.2. Task

Participants are instructed to ignore the auditory stimulation presented via headphones, while performing an immediate memory task related to visual material presented to them on a computer screen. Concurrent bioelectric measurements are made. The performance in the active task is measured and compared for the different experimental conditions.

8.3. ERP analysis

Excluding epochs contaminated by artefact, event-related potentials to the irrelevant sounds are computed and compared across the different experimental conditions.

Acknowledgments

This research has been supported by The University of Helsinki, the Hungarian National Research Fund (OTKA T034112), and by a Marie Curie Fellowship of the European Community programme “Improving the Human Research Potential and the Socio-Economic Knowledge Base” under contract number HPMF-CT-2000-00902. Alexander Degerman and Teemu Rinne are acknowledged for taking the MRI that led to Fig. 1. Thanks are due to Markus Klaske and Teemu Peltonen for technical assistance, to Kati Haataja for translation of experimental instructions, to Manuel Martín-Loeches, Patrick May, Risto Näätänen, and Hannu Tiitinen for comments and suggestions.

References


