International Journal of Language & Communication Disorders

INT J LANG COMMUN DISORD, XXXX 2020,
VOL. 00, NO. 0, 1–13

Review

Understanding event-related potentials (ERPs) in clinical and basic language and communication disorders research: a tutorial

Sean McWeeny† and Elizabeth S. Norton†‡

†Roxelyn and Richard Pepper Department of Communication Sciences and Disorders, Northwestern University, Evanston, IL, USA
‡Institute for Innovations in Developmental Sciences, Northwestern University, Chicago, IL, USA

(*Received January 2020; accepted March 2020*)

Abstract

Background: Event-related potentials (ERPs), which are electrophysiological neural responses time-locked to a stimulus, have become an increasingly common tool in language and communication disorders research. They can provide complementary evidence to behavioural measures as well as unique perspectives on communication disorders. ERPs have the distinct advantage of providing precise information about the timing of neural processes and can be used in cases where it is difficult to obtain responses from participants, such as infants or individuals who are minimally verbal. However, clinicians and clinician–scientists rarely receive training in how to interpret ERP research.
Aims: To provide information that allows readers to better understand, interpret and evaluate research using ERPs. We focus on research related to communication sciences and disorders and the information that is most relevant to interpreting research articles.
Method: We explain what ERPs are and how ERP data are collected, referencing key texts and primary research articles. Potential threats to validity, guidelines for interpreting data, and the pros and cons using of ERPs are discussed. Research in the area of paediatric language disorders is used as a model; common paradigms such as the semantic incongruity N400 and auditory mismatch negativity are used as tangible examples. With this foundation of understanding ERPs, the state of the field in terms of how ERPs are used and the ways they may inform the field are discussed.
Main Contribution: To date, no review has focused on ERPs as they relate to clinical or communication research. The main contribution of this review is that it provides practical information geared toward understanding ERP research.
Conclusions: ERPs offer insights into neural processes supporting communication and can both complement behaviour and provide information that behavioural measures cannot. We encourage readers to evaluate articles using ERPs critically, effectively pushing the field forward through increased understanding and rigor.

Keywords: event-related potential (ERP), neuroscience, mismatch negativity, N400.

What this paper adds

- ERPs have become more prevalent in research relevant to communication sciences and disorders. In order for clinicians to review and evaluate this research, an understanding of ERPs is needed. This review adds to the field by providing an accessible description of what ERPs are, a description of what ERP components are, and the most relevant commonly used components, as well as how ERP data are recorded and processed. With this foundational understanding of how ERPs work, guidelines for the interpretation of ERP data are given. Though few ERP studies currently have direct implications for clinical practice, we discuss several ways through which ERPs can impact clinical practice in future, by providing information that cannot be obtained by behaviour alone about the aetiology of disorders, and as potential biomarkers of disorder or treatment response.

Address correspondence to: Elizabeth S. Norton, Northwestern University, Evanston, IL 60208, USA; e-mail: enorton@northwestern.edu

ISSN 1368-2822 print/ISSN 1460-6984 online
DOI: 10.1111/1460-6984.12535

Introduction

Communication requires a host of complex neural processes, many of which are still not fully understood. When the brain is processing a stimulus, such as a spoken or written word, it produces characteristic electrical signals that can be recorded from sensors on the scalp. These neural signals in response to stimuli, called event-related potentials (ERPs), can index various sensory, cognitive and motor processes such as detecting a stimulus, evaluating or categorizing it, detecting an error or anomaly in the stimulus, and making a motor response such as pressing a button or articulating a word. ERPs are essentially measurements of small electrical voltage changes produced by the brain, with very precise timing. They can provide insights that behavioural measures cannot, such as what happens just milliseconds after a stimulus is presented (before a behavioural response can occur), or processing that happens subconsciously (without explicit effort or attention). They can also provide information into the aetiology of disorders; for example, if a group of patients with a certain disorder look like typical controls in terms of their ERPs that reflect early sensory processing but different in their later appraisal of stimuli, that indicates that the disorder may cause disruption after the phase of early sensory processing. ERPs are also very useful for studying populations who may not be able to make overt responses or follow directions in behavioural tasks, such as infants and young children or patients who have motor, language or cognitive deficits. Assessing the brain can also provide information on the aetiology of communication disorders and may reveal biomarkers of a disorder that behavioural measures alone cannot. For these and other reasons, ERP studies have become increasingly common in speech, language and hearing research.

ERPs can be challenging to understand and interpret without some background in this area, which is not typically part of clinical education and training in communication sciences and disorders. The primary goal of this review is thus to provide speech, language, and hearing clinicians and scientists with a foundation from which they can understand and evaluate research using ERPs. As with any method, ERP research needs to be rigorously assessed not only by the peer-review process but also by the 'consumers' of scientific research. In the same way that a reader needs to understand an intervention approach or statistical procedure to evaluate fully a research paper that uses it, understanding the basics of ERPs can help readers evaluate the conclusions of papers that use them. The goals of this review are to describe: (1) what ERPs are and where they come from; (2) suggestions to evaluate the quality and validity of an ERP publication's methods and interpretation of data; and (3) how ERPs can be effectively used in research with clinical relevance, using work in the area of paediatric language as an example.

Note that several methods similar or related to ERPs are outside the scope of this review. Electroencephalography (EEG) is the continuous signal from which ERPs are derived, and often involves measuring the oscillatory activity of the brain (for a review of EEG in cognitive development, see Bell and Cuevas 2012; for a review of EEG and language, see Weiss and Mueller 2003). Magnetoencephalography (MEG) is a measure of magnetic changes in the brain, similar to EEG (for a review of the MEG/EEG methods, see Puce and Hämäläinen 2017). Electrocorticography (ECoG) records electrical activity directly from the brain's surface (for a review, see He 2015). Also outside the current scope are the auditory brainstem response (ABR) and frequency following response (FFR); although both can be considered ERPs, their interpretation and characteristics differ sufficiently to exclude from this review (for a review, see Skoe and Kraus 2010). Because the ABR's neural origins are known, and it is highly reliable/replicable, it has been extremely clinically useful; building on years of research, newborn hearing screening using ABR is common and required across most of the United States (National Center for Hearing Assessment and Management 2011).

What is an event-related potential (ERP)?

Recall that an ERP is a measurement of the brain's electrical response to a stimulus recorded at the scalp. However, to understand how to collect those measurements, we must first consider how electrical activity is recorded from the living human brain. Electrodes (electrical sensors) are placed on the scalp to record EEG, which is a continuous measure of the electrical activity of the brain, similar to an electrocardiogram (EKG) of the heart. ERPs represent average activity from the continuous EEG signal that is synchronized (i.e., time-locked) to a certain stimulus (e.g., an auditory tone or a printed word).

The electrical activity recorded with EEG comes from neurons in the brain, which function by creating electrical signals. The electrical signals from a neuron are tiny in magnitude, so the signals that can be measured from the scalp reflect thousands of neurons changing their electrical activity in similar patterns. ERPs mostly reflect a certain type of activity from one type of neuron arranged in a certain layout in the cortex (postsynaptic potentials of cortical pyramidal neurons oriented perpendicular to the scalp; for more details on the neural origin of ERPs, see Buzsáki *et al.* 2012 and Luck 2014), but understanding this is not paramount to understanding most applications of ERPs.

The amplitude of electrical activity recorded over time is displayed as brainwaves (typically with a measure

of voltage on the *y*-axis and time on the *x*-axis) and serves as the basis for EEG and ERP signals. To get from continuous EEG to the brain's electrical response to an event (i.e., an ERP), the EEG signal must be time-locked to a stimulus. However, the signal from any one stimulus, or event, captures brain signals that resulted from that event, but also a substantial amount of noise (this includes unrelated neural processes, measurement error and environmental electrical noise). Even in perfect recording conditions with no noise from measurement error or the environment, unrelated brain signals are captured in the recording. Thus, presenting the stimulus many times and averaging the responses together isolates the activity that is related to the event of interest.

ERP components and paradigms

What are ERP components?

Experimental paradigms for ERPs use stimuli that elicit a series of components, which are sensory, perceptual, cognitive or motor processes that appear as a series of positive and negative changes in the waveform (the plotted ERP voltage) over time. Put simply, a specific ERP component is the brain's electrical response to a specific stimulus feature. Most components are named for the direction of their voltage deflection (P for positive, N for negative) and the approximate timing (e.g., N400 component peaks around 400 ms) or order of the peak (N1 for first negative-going peak), though others are abbreviations of their names (such as the mismatch negativity, MMN). Components have an onset, a peak or plateau, and an offset. For example, the onset of a component might be when the first few thousand neurons in primary auditory cortex are systematically responding to a sound, and the offset would be when auditory cortex quiets to baseline levels.

ERP components reflect how processing in the brain unfolds over time, from early sensory processing components precede later cognitive processing components that require attention and evaluation of the stimulus. Early components reflecting sensory and perceptual processing of a stimulus are often called 'pre-attentive', meaning they occur without or before conscious attention from the participant. Later components typically reflect cognitive processing and require conscious attention, though some cognitive ERPs can appear without the subject consciously perceiving a stimulus (Luck *et al.* 1996). Passive or unattended paradigms are well suited to paediatric populations, as they do not require any overt responses or compliance with directions, and they can be collected while the child is attending to something more engaging, such as watching a movie. Nonetheless, well-designed and engaging paradigms that require active participation or attention have been successfully utilized in numerous ERP experiments with children (e.g., Henderson *et al.* 2011).

Examples of ERP paradigms to guide the review

Throughout this review, we will use two classic ERP paradigms in the area of language processing to give concrete examples of concepts. We will first introduce these experimental paradigms and then how the ERP that they elicit is measured.

Semantic incongruity paradigm and N400 component

The N400 semantic incongruity paradigm based on a landmark paper in which Marta Kutas and Steve Hillyard established that ERPs can index detection of a semantic incongruity in language (Kutas and Hillyard 1980). In this now widely used paradigm, a participant reads a sentence that is presented one word at a time on a computer screen; most of the sentences are semantically appropriate or expected, but some have an incongruous or unexpected final word. An example semantically appropriate sentence would be 'It was his first day at work.' In the original experiment, 25% of the sentences had a semantic incongruity, that is, the final word of the sentence is unexpected from the context, such as 'He spread the warm bread with socks.' The brain's response to the final word is very different in this semantically incongruous condition than in the expected condition. This ERP response to semantic incongruity is called the N400 component, as there is a more negative voltage for the unexpected word around 400 ms after the word is presented. The N400 component also occurs when a word is semantically acceptable but not expected. For example, the sentence 'Do not touch the wet dog' would elicit an N400, even though it is not a semantic error; 'Do not touch the wet paint' is the semantically expected ending (Kutas and Hillyard 1984). The N400 occurs regardless of whether the stimuli were presented visually (e.g., Kutas and Hillyard 1980) or auditorily (e.g., McCallum *et al.* 1984) and thus specifically relates to lexical–semantic processing. The N400 may be used to answer questions about the strength of semantic relationships, or whether clinical populations have different patterns of semantic processing. For example, studies have used the N400 component in a modified semantic paradigm to examine whether receptive semantic processing differs in children with autism spectrum disorder who are minimally verbal or non-verbal (e.g., Cantiani *et al.* 2016).

Auditory oddball paradigm and mismatch negativity (MMN) component

The second example is an oddball paradigm in which one stimulus is presented often and one or more other

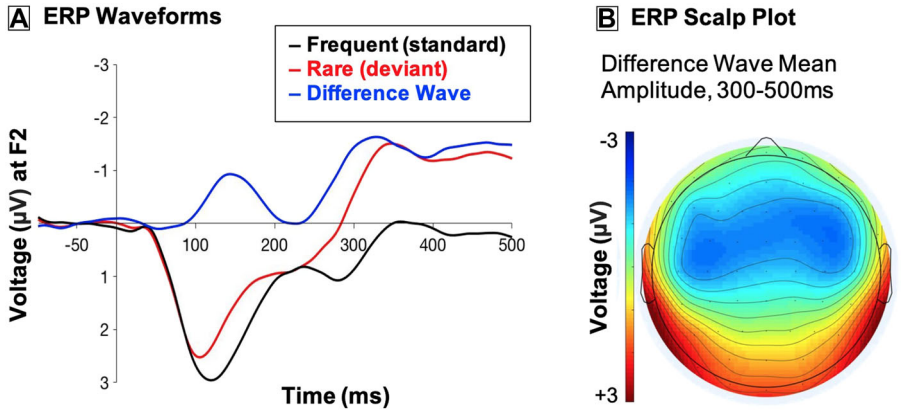

Figure 1. Typical plots present in ERP research: (A) group grand average waveforms (averaged from about 100 child participants) showing two conditions (frequent and rare syllables, in black and red) and the difference or subtraction between them (blue); and (B) group grand average scalp plot showing the spatial distribution of the mean amplitude of the difference wave depicted in (A) within a time window of interest from 300 to 500 ms. The negativity shown in blue extends bilaterally across fronto-central electrode sites. Data are from the authors' laboratory. [Colour figure can be viewed at wileyonlinelibrary.com]

stimuli are presented rarely. For example, the syllable /ba/ is played 90% of the time and the syllable /da/ is played the remaining 10% of the time while the participant watches a silent video. The difference between the frequent stimulus (/ba/) and the oddball or deviant stimulus (e.g., /da/) is a component known as the MMN (Näätänen and Kreegipuu 2011). The rare sound elicits a larger negative voltage response from about 100–250 ms; this difference between the rare and frequent stimuli, shown in figure 1A, is the MMN. If the stimuli probabilities were switched (e.g., /ba/ played 10% and /da/ played 90%), we would see a highly similar pattern of rare versus frequent responses as in the original paradigm, indicating that the differences are not solely driven due to acoustic differences in the two syllables, but depend on more general change detection processes. The MMN can be used to index automatic auditory discrimination of different auditory features (such as tone pitch or duration) and how the brain tracks stimulus probability, which are skills important for language, especially language learning. If the difference between the two sounds is too small for the participant to perceive, the MMN is not observed.

This is only one example of an oddball paradigm, which is a commonly used ERP paradigm that can be modified with multiple deviants, different stimulus probabilities (e.g., 25% deviants), or a task where the participant actively responds to categorize the stimuli. The MMN allows researchers to investigate populations in which auditory discrimination may be impaired. Because it does not require overt attention or responses, it has been widely used in infants and children (Bishop 2007) and shown to differ in groups with or at risk for language and reading disorders (Bishop 2007, Bruder *et al.* 2011). Although it is most commonly used in auditory research, some evidence exists that it exists for visual stimuli as well (Kimura *et al.* 2009).

Other common ERP components

Although it is not necessary to have in-depth knowledge of established components to read ERP research, understanding these components will give a better sense of how they may be applied to research. These descriptions are designed to give a cursory overview of what types of stimuli and paradigms may elicit components related to communication. There are many more components that are not covered here, such as ones related to syntactic incongruity (P600) or error detection (the error-related negativity, ERN). Note that some auditory and visual components share the same name, despite originating from completely different areas of the brain and responding to different stimulus modalities (i.e., hearing versus vision). For further detailed reading, see Luck and Kappenman (2011).

Auditory sensory components

In adults, cortical auditory responses begin with a positive wave around 50 ms called the P1 or P50 component. The P50 is generated in auditory cortex and reflects sensory gating (Pratt 2011), the process through which the brain filters out irrelevant information. P50 is quickly followed by a larger negative wave called the N1, or N100. The N100 reflects a change in auditory stimulation, such as the offset and onset of sounds (Pratt 2011). Some evidence suggests that N1 may be used to index speech segmentation, such as in a statistical learning paradigm (Sanders *et al.* 2002). In young children these components are different, with P1 being larger in children than adults (Čeponiené *et al.* 2002); across

development, they tend to shift in time and change in shape.

Visual sensory components

Although brief flashes of visual stimuli can evoke ERPs as soon as 70 ms after stimulus onset (Allison *et al.* 1977), the main visual ERPs described in research manifest as a series of three peaks. The first of these components, C1, changes polarity based on where in the visual field the stimulus is presented (hence, its name having a C rather than a P or N). It is followed by P1 and N1, which overlap in time (in other words, P1 has not completely offset by the time N1 begins); this a similar pattern to auditory P1 and N1, though from different neural sources. Although C1, N1 and P1 components each peak at different times, they are all generated in visual cortex in short succession (Pratt 2011). All are affected by luminance (i.e., brightness) change, such as a white stimulus appearing on a black background. These three components nicely demonstrate that components can have temporal overlap in their offsets and onsets and thus potentially influence the peak latency of any of the three components (figure 2).

P300/P3/P3a/P3b

Like the MMN, the P3 (also called P300) is typically elicited by an oddball paradigm. However, unlike the MMN, it requires the participant to pay attention to all the stimuli and have an expectation or categorization of the difference between the standard and the rare target stimuli (Polich 2007). It can be elicited using stimuli from visual, auditory and other sensory domains (e.g., tactile; Brouwer and Van Erp 2010, Polich 2007). The P3 wave can be broken up into two unique subcomponents: P3a and P3b. These two subcomponents reflect different aspects of the stimulus; P3b occurs only when the target stimulus is presented (the one the participant is trying to detect), but P3a appears in response to an infrequent stimulus, regardless of whether the participant was looking for it or not (Polich 2011). The exact cognitive processes that P3b represents (e.g., working memory versus a more general process) remains an active area of research and debate (Polich 2011). This illustrates how a single large wave may have multiple generators or underlying subcomponents. The P3 component may be useful in cases in which participants are classifying objects, for example, based on certain object features (Deveney *et al.* 2019).

ERPs from recording to analysis

A familiarity with the basics of the experimental set-up can help contextualize ERP data. Here, we explain how ERP data are recorded and introduce key terms that will show up in most ERP research.

Recording ERPs

To record ERPs, a cap or net holding the electrodes is placed on the scalp. Saline solution or gel is typically applied to conduct the electrical signal from the scalp to the electrodes. This typically takes 5–20 minutes depending on the system and participant. (Newer 'dry' electrode systems that do not require conductive gel or liquid are becoming available, but it is not yet clear whether these systems can provide signal-to-noise ratio that is on par with existing research-quality EEG systems.) Net and cap systems measure the same signals, but have some advantages and trade-offs. Net systems can be quicker to set up as the saline solution can be applied to all electrodes at once and then just placed on the participant's head. Electrodes in a cap system need to be gelled individually, which takes more time, but the data are often cleaner (less noisy), especially with active electrode systems that have signal amplifiers at each site. In the past, the scalp was typically abraded (scraped) to remove a layer of skin to improve signal, but this is not typical with modern systems, which makes EEG recording safer and more comfortable for participants. Additional 'external' electrodes are typically placed near the eyes to detect eye movements and blinks, as the associated muscle movements generate an electrical signal which can contaminate the data of other nearby electrodes. Very young children may not tolerate electrodes on their face, and so experimenters may use electrodes from the forehead to detect blinks. We discuss how to clean blinks (i.e., reducing their effect on the data) in the Preparing Data for Analysis section. External electrodes may also be placed to record a reference or comparison signal, such as on the earlobe or mastoid bones behind the ears. All electrodes are typically plugged in to a small box that amplifies and/or converts the signal to a format that can be recorded by a computer. Unlike magnetic resonance imaging (MRI) or MEG, this equipment is small and can be portable.

Electrodes are typically named and labelled based on their location using what is called the 10–20 system, which provides a standardized map for locating an electrode on the head across people and studies (Homan *et al.* 1987). Electrode locations in the 10–20 system are labelled with a letter indicating where they are on the head (e.g., Fp, frontal pole; F, frontal; FC, frontocentral; C, central; CP, centro-parietal; P, parietal; PO, parieto-occipital; O, occipital; T, temporal; FT, frontotemporal; and TP, temporo-parietal) and number (even numbers on the right, odds on the left, with larger numbers indicating farther distance from the midline, which uses *z* instead of a number). For example, Fz is located

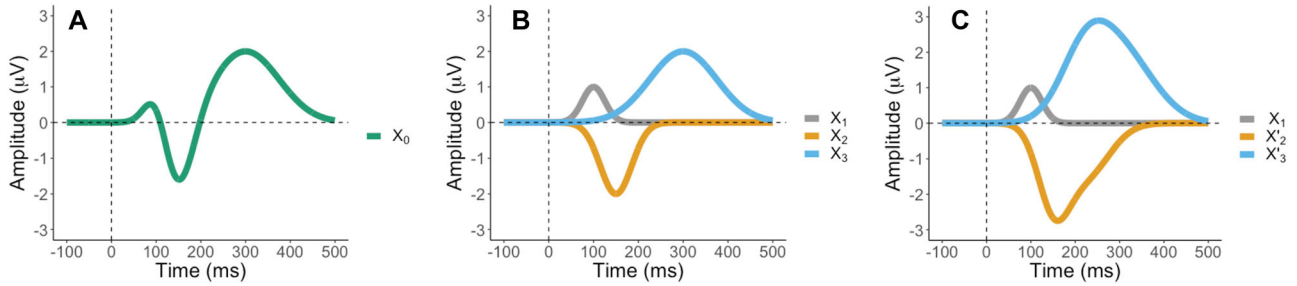

Figure 2. How different underlying component structures can manifest as identical waveforms: (A) a theoretical ERP waveform with three peaks; (B) one possibility for the underlying component structure that when summed together creates the waveform in (A); and (C) another possibility for the component structure of the waveform in (A). Notice the shorter duration of X_2 as compared with X'_2 and the larger amplitude of X_3 as compared with X'_3 in (B). [Colour figure can be viewed at wileyonlinelibrary.com]

over frontal cortex, equidistant from the ears, and F1 is slightly to the side of Fz, over the left hemisphere.

Once the electrodes are placed and connected, the experiment can begin. Researchers typically use a computer program to present stimuli to the participant. Every time a visual or auditory stimulus is presented (though ERPs exist for other sensory modalities, they are rare), a code is sent from the presentation software to the EEG acquisition software that is recording data from the electrodes. This code in the EEG data indicates the exact moment a stimulus was presented, and different codes are typically used to indicate the type of stimulus that was presented. Using the semantic incongruity task example, if we know exactly when each semantically appropriate word and when each semantic incongruity was presented, we can isolate the electrical activity to each type of stimulus (i.e., incongruous or not) by averaging the trials of each type together.

Time and voltage: The units of ERP measurements

ERPs reflect how electrical voltage changes over time, typically over several hundred milliseconds. Voltage and time also characterize the two primary measurements of ERPs: amplitude and latency. Amplitude is how large an ERP voltage is, and latency is when an ERP happens; we discuss how to best measure these below. ERPs measure voltage over time with incredible temporal resolution, often hundreds or even thousands of times per second. This temporal precision is perhaps the greatest strength of ERPs. Because electrical current travels very quickly and the brain likewise processes certain aspects of a stimulus very quickly, the human auditory brainstem can register sound about 1.5 ms after it is played. A cognitive event such as detection of a semantic incongruity occurs much later, 400 ms post-stimulus onset, though that may not seem long!

Most research laboratories use systems that record voltage from many electrodes simultaneously; 32 and 64 electrode caps are common in research, and nets often include 128 or 256 electrodes. (More electrodes are not necessarily always better; Luck, 2014: ch. 5.) Voltage at each electrode is compared with activity at another site such as the nose, earlobe, mastoid processes or an average of all the scalp electrodes. This comparison process is known as referencing, and the site of the reference electrode(s) greatly affects how the data appear. For example, using Cz (the electrode at the very top of the head) as the reference would produce different looking ERPs than the average of all electrodes as a reference, as the reference is subtracted from each electrode. Studies with similar paradigms typically will use the same reference locations as previous work, meaning that one can compare directly across publications; however, when comparing results across publications, it is important to check if the studies use the same reference electrodes.

Preparing data for analysis

ERP analysis involves many steps and many choices, and so understanding the main aspects of analysis is important for interpreting the results from papers. There are established best practices for making these decisions and guidelines for what information about ERP processing that studies need to report (Keil *et al.* 2014). Studies that do not follow these guidelines could report insufficient information for replication, inaccurate or biased data, or inaccurate conclusions about neural differences. Thus, this section gives readers some tools to understand better the steps and decisions that go into making inferences about ERPs and the relative methodological strength of a given publication.

Several steps are needed to go from an EEG recording to a set of ERP waveforms. Here, we describe a typical processing 'pipeline' for generating ERP measurements from EEG data. The first is to filter the EEG data so that only activity that is likely to be relevant to brain processes remains. Much of the activity of neurons is rhythmic, or oscillatory. Thus, a first step is to filter out oscillatory activity that is slower than about 0.01–0.3 Hz (high-pass filter), or faster than about 100 Hz (low-pass filter). High-pass filters at 0.3 Hz and above

can systematically change the shape and timing of ERPs (Tanner *et al.* 2015), so that is something to watch out for. Often, a notch filter, which knocks out a single frequency or small frequency range, is applied to remove electrical noise (e.g., at 50 Hz in Europe and Asia or 60 Hz in North America), as utility electricity is carried at those frequencies and electrodes may pick up that frequency. Electrical noise, also called line noise, is a type of artefact.

An artefact is best described as a non-event-related signal, or noise, that can arise for many reasons. To obtain the cleanest ERPs, artefacts must first be detected, and then either corrected or rejected from the data. Perhaps the most common artefact comes from the participant blinking, which creates a large voltage change due to the muscle activity required to blink. Blinks can be corrected (i.e., their effect subtracted from the data) through a computational process called independent component analysis (ICA). An alternative approach is to just exclude ('reject') moments of data with blinks; however, ICA can increase the number of usable trials, which is typically desirable.

Once ICA is complete, the data are segmented into epochs, which are periods around an event or stimulus onset. Epochs are created based on the codes that are sent when a stimulus is presented, described in the Recording ERPs section. Typically, within an experiment, each epoch is the same length. Epochs typically begin 50–200 ms before the stimulus and last until sometime before the next stimulus is presented. The time before each stimulus is called the pre-stimulus baseline, or just baseline. The average voltage over each trial's baseline is subtracted from the post-stimulus signal in an effort to reduce noise.

Next, any remaining artefacts must be detected in the data and rejected. Artefact rejection and correction is necessary, as large artefacts will distort the final ERP waveforms with large voltage changes due to noise. (The saying 'a few bad apples can spoil the bunch' applies in this case.) Many artefacts in EEG data come from motion by the participant; this is particularly true for children, who tend to move more during recording. Another common type of artefact appears when an electrode's connection to the scalp was poor or lost during the experiment. The information from these electrodes can be deleted and re-created using an average of the electrodes surrounding it in a process known as interpolation. Ideally, this occurs for very few electrodes, as interpolating too many adjacent electrodes will result in an interpolated channel that is likely very different than the actual brain activity that was occurring at that electrode site. After any bad electrodes are interpolated, the data are examined for remaining artefacts. There are lots of ways to do this: one way is to measure how much the measured voltage changes over a 200-ms interval and reject any interval in which data that exceed a given voltage threshold (e.g., 150 μV). This process is then repeated for every 200-ms interval in each epoch. Some researchers choose apply the same threshold for artefact detection to each subject's data, which can speed up this process; others choose to use individualized criteria, as 'clean data' look different in each person, which is an acceptable and often necessary approach for paediatric data (Luck 2014). The data marked as artefacts are then excluded from further analysis. It is common in studies of children for up to 50% of epochs to have artefacts. Thus, it is important to consider how many usable trials are averaged together; too few trials may lead to noisy averages that do not reflect the real neural activity accurately. If different conditions or groups are being compared, it is ideal to make sure that conditions or groups do not have disparate numbers of included trials (see Number of Trials section).

After the data have been cleaned of artefacts, epochs within each condition (e.g., all incongruous trials) are averaged together. As discussed above, averaging together multiple epochs allows one to see the components of interest, because random noise not associated with processing the stimulus will cancel out. Once epochs of a given type are averaged together, the resulting waveform is considered an ERP. Average ERPs of a whole group of participants or comparisons of ERPs between groups are often the main interest in ERP research; thus, publications do not often include the individual participant waveforms calculated in the previous step. Instead, they typically present grand average waveforms which are calculated by averaging the waveforms from every individual in a group. For example, if we were interested in whether early auditory processing differs for toddlers at risk for dyslexia compared with those not at risk, we might compare the grand average for the at-risk group with the grand average for the not-at-risk group (e.g., Plakas *et al.* 2013). These smooth grand average waveforms *are not necessarily what any individual's waveforms look like*, nor what any given trial might look like for any individual. This is because it takes many trials to cancel out unrelated activity (i.e., noise), and many participants sufficiently to cancel out noise and smooth the group's response (i.e., signal). Grand averages can also be compared across different experimental conditions within-subjects, such as semantically expected versus incongruous words, to understand better basic neural processes of semantic comprehension.

Plotting ERPs and interpreting figures

Plotting ERPs allows one to visualize the differences between conditions (e.g., semantic incongruity versus semantically appropriate word) or between groups (e.g., typical development versus autism). Typically, just

before plotting, a 30-Hz low-pass filter is applied to remove high-frequency noise, which makes the waveform appear smoother. Figure 1A shows a typical ERP waveform plot using the auditory oddball paradigm (/ba/ presented 90% of the time and /da/ presented 10% of the time). Time (in ms) is plotted on the *x*-axis, and voltage (in μV) is plotted on the *y*-axis. Here, negative values are plotted up; some figures use this convention, which has been typical in the history of ERP research, although others plot positive up, which is more intuitive for most readers. It is important to pay attention to the axis' labels to confirm which direction is plotted up. The *x*-axis spans from –100 to 500 ms relative to the stimulus onset. As discussed above, the average voltage over the baseline from –100 to 0 ms has been subtracted from the whole waveform, and the baseline section is thus relatively flat. Readers can check for pre-stimulus baselines to be relatively flat compared with the waveforms that occur after stimulus onset; if they are not, this could indicate noisy data or inaccurate processing.

Figure 1A has three ERP waveforms: the waveform to the /ba/ (90%) condition is in black and the ERP waveform to the /da/ (10%) condition is in red. Notice that both the black and red waves have a positive-going wave around 100 ms. Before stimulus onset and early after onset, the two conditions do not differ, which suggests that neural responses are similar between conditions, as we would expect. From 100 to 200 ms and from 250 to 500 ms, there are negative-going peaks in this wave, indicating that one condition had a stronger negative response than the other. The blue waveform is a difference wave; is a subtraction between the two conditions (deviant minus standard waves), which allows readers to visualize more easily the difference between the conditions. The waveforms being compared in a difference wave should have an approximately equivalent number of trials (Luck 2005). From our oddball paradigm example, comparing every frequent sound with every rare (deviant) sound would result in signal that is much noisier for the rare sound, because there are nine times fewer trials of this kind presented. In this case, researchers can use a subset of the frequent sounds (e.g., average together and analyse only some standards).

ERPs can also be plotted using scalp maps, which show how positive or negative the ERP signal is across the scalp at a certain time or time interval. In figure 1B, the mean amplitude (voltage) of the difference wave between 300 and 500 ms is plotted at each electrode and then smoothed in space to create a continuous map of voltage. The colour bar indicates that negative values are plotted in blue; meaning that over the frontal and central part of the scalp, the difference wave is negative. Even though we see an ERP over a certain part of the scalp, this does not necessarily mean that the process happened in that brain area; this will be discussed in depth in the Localization section. Nonetheless, differences in scalp maps between conditions or groups can provide meaningful information.

Quantification or measurement of ERPs

Although looking at the waveforms provides a great deal of information, quantifying or measuring an ERP component is necessary to compare statistically between groups or conditions. Understanding the general strengths and weaknesses of different measurement approaches helps readers to understand better the relative strength of a given paper. As mentioned above, ERPs can be quantified in terms of amplitude and latency; most papers measure one or both of these characteristics of a given component. Though many studies measure features about the peak of a component, there are some drawbacks to doing so. There is nothing inherently meaningful about the time and exact voltage at which a component reaches a maximum voltage (Luck 2014). Furthermore, the exact time and voltage at which peaks occur (i.e., peak latency and peak amplitude) can change based on how much noise is in the EEG signal. This is because noise distorts the waveform, which may create a false peak that does not accurately represent the component. Although peak amplitude and peak latency may provide information about a component, some less-biased alternatives are mean amplitude between two time points. Finding an alternative to peak latency that is robust to noise is more challenging; the approach that is most valid (from options such as fractional area latency, or relative criterion technique) may vary depending on the component of interest (Kiesel *et al.* 2008). Overall, peak measures should be regarded with these considerations in mind.

Even for noise-robust measurements such as mean amplitude, the measurement window and electrode(s) to measure and analyse must be chosen carefully. Ideally, all publications would use measurement windows selected in advance of analysis based on closely related previous literature so as to be unbiased; however, the literature can differ in exactly what time window they quantified their components, or a paradigm or population may have never been studied before. When certain ERP components begin can depend on the duration of a stimulus, the time between stimulus onsets (i.e., stimulus onset asynchrony), as well as the time between stimulus offset and the next stimulus onset (i.e., interstimulus interval). Thus, the time when the component occurs will vary from study to study depending on the stimuli used.

Given the concerns about rigor and reproducibility in neuroscience (e.g., Loken and Gelman 2017), a major concern for all ERP research is the potential

for experimenter bias, either conscious or unconscious. Ideally, analytical decisions such as which electrode site and time window to test, and how to exclude artefacts and outliers, are made in advance of data collection, so that the data cannot influence the results. Practices that should be avoided are 'fishing' for a time window in which there is a difference and then conducting statistical tests only in that window, or testing many time windows and electrodes and not reporting these tests and appropriately controlling for the multiple statistical comparisons. Luck and Gaspelin (2017) illustrate this nicely with an example where statistically significant results were obtained from real ERP data by pure chance by using analytical practices that are common in the field, but which are not rigorous. Their message is clear: in highly multidimensional ERP data, it is extremely easy to find statistically significant results, even though nothing meaningful exists, and peer review may not always catch these instances of false-positive findings. Readers can look for whether authors made analytical choices based on previous literature or a priori hypotheses, rather than seeming to test many possibilities until one reached significance. As more authors pre-register their analytical plans in advance of data collection and journals encourage this rigorous practice, it may become easier to know which findings were truly robust.

Localization

Perhaps the most notable limitation of ERPs is their poor spatial resolution; because of the way electricity moves and spreads through the brain, skull and scalp, voltage at a given electrode does not simply reflect activity that is generated directly below its location. Quantifying the precise anatomical region in the brain that produces an ERP component, also known as localization, is often impossible without making several assumptions about the locations and numbers of generators of that process in the brain that are often unknown. Making claims about a certain ERP's precise location in the brain, or differences by conditions or groups, requires an extremely high burden of proof. Even with analysis technology such as LORETA (Pascual-Marqui *et al.* 2002), the use of high-density electrode arrays, and using each individual's MRI scans to assist in analysis, poor localization remains a fundamental limitation of EEG and ERPs. The sources of ERP signals can be estimated these processes, but we suggest proceeding with caution regarding claims about truly 'localizing' an ERP.

Lateralization

Because localization of ERPs is so limited, researchers interested in spatial questions will sometimes test whether an effect is stronger on one side of the brain than the other, known as lateralization. Lateralization is relevant to questions related to language because language functions are typically left-hemisphere lateralized; thus, one might expect to see larger amplitudes in left versus right hemisphere electrodes. Many studies have reported lateralization being weaker or absent in language-related disorders. Lateralization can be calculated in numerous ways such as using centroids (spatial means) of scalp maps (e.g., Maurer *et al.* 2009) or comparing single or groups of electrodes from left and right hemispheres (e.g., Monjauze *et al.* 2011). Although there is no consensus on the best way to study lateralization, it is nonetheless a valid practice in ERP research.

Considerations for experimental design and interpretation

Number of trials

Researchers face many trade-offs in designing experiments. Ideally, many trials (often, hundreds per condition) are collected so that the cleanest possible average ERP can be achieved. However, this is not always practical for many reasons. The length of an experiment paradigm is an important consideration because children (and even sleep-deprived undergraduates) may have a hard time sitting still and maintaining attention to the experiment over long periods. On the one hand, increasing the number of trials can help increase statistical power for between- and within-group analyses by improving reliability (Boudewyn *et al.* 2018). On the other hand, limiting the length of the paradigm may allow more participants to tolerate the duration of the paradigm. Furthermore, some ERPs may habituate or disappear entirely over time (McGee et al., 2001). There is not a clear rule or guideline of how many trials are 'enough' to include, as this depends on how clean the data are and how large is the component. Efforts should thus be made to determine the reliability of a given ERP, or refer to published research that has determined an appropriate number of trials (e.g., Foti *et al.* 2013). This is challenging, as there seems to be no clear answer for how many trials are needed to measure any given ERP component reliably (e.g., anywhere from 10 to several hundred) across all EEG systems or stimuli parameters, but the number of trials *presented and used/accepted* that are compared per participant group and/or condition should be reported (Boudewyn *et al.* 2018).

Differences across conditions and motor movements

Motor responses are another important consideration in the design of ERP experiments. For example, if the goal of an experiment is to have participants actively categorize a rare stimulus and researchers want to be sure that

this is being done accurately, they may ask participants to press a button to each rare stimulus. Importantly, the ERP response to the rare stimulus will also contain the motor activity associated with the button press. This underscores an important feature of design that readers can again watch for; everything possible should be the same between two types of trials being compared. If trials in one condition have a button press (or other response), the comparison trials should as well, or the source of any differences between trials cannot be determined. Some potential solutions would be to have a button press on every trial (one button for rare and one for frequent, pressed using the same hand or finger), to add a third type of trial that is excluded from analysis in which participants press a button to indicate their attention, or to have no button press at all. Closely matched conditions thus are necessary to isolate the activity of interest with no confounding factors.

Interpreting peaks and components

Even though components follow the general pattern of timing of different processes in the brain, they must be interpreted carefully. A common misconception is that the peak (highest or lowest point between onset and offset) indicates something meaningful about the component or underlying brain processes (Luck, 2014). The same observed waveform (i.e., the plot of voltage change over time) could be made up of multiple sets of components that overlap, so the final waveform peak's amplitude (voltage at the peak) and latency (timing at the peak) may not necessarily be important (figure 2). This is important to keep in mind; for example, it is incorrect to assume that because you see the peak of an ERP in condition A earlier than the peak in condition B, that this is because something about condition A causes the component to happen earlier. There could be other components' onsets or offsets occurring simultaneously that would shift the peaks despite the component of interest remaining stable.

Developmental changes

A particular challenge for paediatric and longitudinal ERP research is that components change over the course of development. A pattern of auditory responses, the P1-N1-P2 complex, is very well established and studied in adults, is not found even by age 9; in children, P1 is the most prominent of these components, whereas N1 is most prominent in adults (Čeponienė *et al.* 2002). Indeed, there is considerable change in the morphology (i.e., shape) and timing of these and other components from infancy to childhood (Čeponienė *et al.* 2002, Wunderlich *et al.* 2006). The MMN can actually be a positive rather than a negative-going wave in infants, and is thus sometimes called the mismatch response (MMR) (van Leeuwen *et al.* 2006). This can make comparisons between different age groups very difficult when discussing each component on its own, especially considering the problem presented in figure 2 of overlapping components that affect each other's shapes. It is also typical for the overall amplitude of ERPs to decrease with age. Readers should consider these factors when studies compare across ages.

Why use ERPs?

Given the complexities and challenges of ERP research, what are the advantages of this technique? Perhaps the clearest reason to use ERPs is that ERPs and behaviour do not have a one-to-one relationship; if they did, there would be no reason to study ERPs when behaviour could be used. For example, the MMN may be a more sensitive indicator of auditory discrimination than behaviour (Stoodley *et al.* 2006). This means that we are measuring something unique with ERPs, that can add information to our understanding of typical and atypical behaviours. Here, we discuss the primary reasons researchers use ERPs.

When behaviour is hard to obtain

One of the best reasons to use ERPs is when behavioural paradigms would be ill-suited to, or difficult to obtain from, the population of interest. A clear example of this is with auditory discrimination among infants. The MMR ERP, an infant corollary of the MMN, is elicited by an oddball paradigm and does not require the attention of the infant (van Leeuwen *et al.* 2006). Like the MMN, the component is believed to represent an objective measure of auditory discrimination, which roughly corresponds to behavioural discrimination thresholds in adults (Näätänen and Kreegipuu 2011). As behavioural tests cannot otherwise assess this skill in very young infants, ERPs offer a window into how this process develops over time. Furthermore, as behavioural auditory discrimination is impaired in some (but not all) individuals with paediatric language delays and disorders (Benasich and Tallal 2002), one might expect this difference to be present in the brain from birth.

As a biomarker when behaviour cannot tell us enough

Another major interest of many researchers is whether ERPs could practically aid in clinical screening, diagnosis or progress monitoring for individuals with a disorder. An ERP used in this way would be considered

a 'biomarker'; a biomarker is a biological substance or process that is objectively quantifiable, reproducible and related to underlying biology implicated in a disorder. Because so many disorders in the field of paediatric communication disorders, such as autism, language disorder and dyslexia cannot be diagnosed until children fail to achieve typical milestones, early and objective neural markers of these disorders have been a focus of much research. Using ERPs in this way has the potential to be an extremely powerful tool, but it requires highly reliable measures. Many studies have found ERP components that relate to specific language impairment/developmental language disorder (e.g., Friedrich *et al.* 2004) or dyslexia (Bruder *et al.* 2011, Halliday *et al.* 2014, Maurer *et al.* 2009). Researchers have also measured the MMR at birth in at-risk individuals and followed them through to school age to see how well ERPs predicted who would develop dyslexia (Jyväskylä Longitudinal Study of Dyslexia; e.g., Guttorm *et al.* 2010). Associations with later language outcomes from these early measures show some promise, but much more research is needed to determine the generalizability of these results (Norton *et al.* 2019).

To date, these biomarker studies have yet to be taken beyond the initial stages of discovery to more clinically meaningful statistical validation and implementation. Once preliminary associations are found, candidate ERP biomarkers must be subjected to cross-validation to test their generalizability and controlled replication studies (for a model of bringing brain measures to clinical utility, see Gabrieli *et al.* 2015). Despite these challenges, ERPs remain an enticing biomarker because they are relatively fast, easy and inexpensive to obtain (compared with, say, a detailed cognitive and language assessment or an MRI), and can be obtained early in life and thus lead to earlier identification and intervention.

Uncovering aetiology of disorders

Most language disorders are neurologically based, but their causes are not fully understood; there is hope that ERPs may be able to help elucidate causes of disorders, rather than just understand their behavioural symptoms. Even if we cannot tell much about an individual owing to low reliability of measures, we may still be able to understand the nature of neural processing in *groups* of people with language differences or disorders with large samples and rigorous practices.

For example, one hypothesis about the cause of dyslexia is that there is a neural impairment in the way that sounds are processed and discriminated. Comparing the MMN ERPs from two groups that are carefully matched (e.g., in age, number of included trials, etc.) can indicate whether this automatic auditory process is altered in language disorders. All studies, regardless of the hypothesis and ERP components used, should discuss how their results tested their hypothesis as explicitly as possible.

Furthering basic understanding of neural processing

ERPs have also been applied at a more basic level to understand how humans process different types of information. The semantic incongruity paradigm has been used extensively to describe how the brain processes linguistic information. Other examples of basic applications of ERPs are how the brain form auditory objects (e.g., Sussman and Steinschneider 2009), how it integrates information from multiple senses (Brett-Green *et al.* 2008) and how it processes emotion in faces (Blau *et al.* 2007). Because these uses are not particularly applicable clinically, we will not discuss them further, but many of the applied uses of ERPs were born out of these basic research applications.

When and why not to use ERPs

Despite the advantages of ERPs in many situations, there are also several situations where it is most appropriate to *not* use this technique. When a behavioural experiment can answer the same question, there is often no reason to 'tack on' ERPs. That is, just because a process can be measured using an ERP does not mean it should; a study should use ERPs when they are the right tool to answer a given question. If the research questions focus on where a process occurs in the brain, ERPs may not be a good technique relative to other methods such as functional magnetic resonance imaging (fMRI) or MEG. ERP studies are less well-suited to certain populations, such as those using cochlear implants, as the electrical activity and placement of the implant receiver may affect the signal.

Clinical implications and conclusions

In our view, research using ERPs is promising, but has not yet made a direct impact on how clinical diagnosis or treatment of speech or language disorders proceeds. Although ERPs have been used in basic research to understand cognitive processes for decades, their applications to clinical questions has been growing as technology becomes more affordable and accessible. Using ERPs to understand clinical disorders is still somewhat in its infancy, and it is important to recognize that ERP research overall has evolved and been refined over decades. It will likely take decades more of rigorous research to bring ERPs to clinical application, as we learn more every year about the most valid and reliable ways to measure ERPs and technology makes them more feasible to collect (including that they are more participant-friendly,

quick to set up and affordable). At the same time, we are also learning more about the heterogeneity of speech and language disorders, and how we can best address this heterogeneity both clinically and scientifically. Nonetheless, we believe it is important for clinicians to interpret ERP results (1) because informed readers will push researchers to improve the state of science, (2) because understanding the shortcomings in existing research can provide a road map of what to look for in the years to come in ERP experiments, and (3) lastly, if ERPs move into clinical practice in future, clinicians can be ready to apply their knowledge of the data to their patients as if it were standardized behavioural assessment data. Although ERPs are not used in the clinic to assess speech and language disorders at this time, we believe that scientific literacy on the topic is key to facilitating research that would help achieve this goal.

Acknowledgments

We acknowledge Steve Luck, whose writing and teaching has guided many of the topics covered in this tutorial. We thank Katherine Wolfert for comments on EEG net setup and Brittany Manning CCC-SLP for discussion on clinical considerations. E.S.N. was supported by the National Institute on Deafness and Other Communication Disorders (NIDCD) under award number R01DC016273. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

References

Allison, T., Matsumiya, Y., Goff, G. D. and Goff, W. R., 1977, The scalp topography of human visual evoked potentials. *Electroencephalography and Clinical Neurophysiology*, **42**, 185–197. https://doi.org/10.1016/0013-4694(77)90025-6.

Bell, M. A. and Cuevas, K., 2012, Using EEG to study cognitive development: Issues and practices. *Journal of Cognition and Development*, **13**, 281–294. https://doi.org/10.1080/15248372.2012.691143.

Benasich, A. A. and Tallal, P., 2002, Infant discrimination of rapid auditory cues predicts later language impairment. *Behavioural Brain Research*, **136**, 31–49. https://doi.org/10.1016/S0166-4328(02)00098-0.

Bishop, D. V. M., 2007, Using mismatch negativity to study central auditory processing in developmental language and literacy impairments: Where are we, and where should we be going? *Psychological Bulletin*, **133**, 651–672. https://doi.org/10.1037/0033-2909.133.4.651.

Blau, V. C., Maurer, U., Tottenham, N. and McCandliss, B. D., 2007, The face-specific N170 component is modulated by emotional facial expression. *Behavioral and Brain Functions*, **3**, 7. https://doi.org/10.1186/1744-9081-3-7.

Boudewyn, M. A., Luck, S. J., Farrens, J. L. and Kappenman, E. S., 2018, How many trials does it take to get a significant ERP effect? It depends. *Psychophysiology*, **55**, e13049. https://doi.org/10.1111/psyp.13049.

Brett-Green, B. A., Miller, L. J., Gavin, W. J. and Davies, P. L., 2008, Multisensory integration in children: A preliminary ERP study. *Brain Research*, **1242**, 283–290. https://doi.org/10.1016/j.brainres.2008.03.090.

Brouwer, A.-M. and Van Erp, J. B. F., 2010, A tactile P300 brain-computer interface. *Frontiers in Neuroscience*, **4**:19. https://doi.org/10.3389/fnins.2010.00019.

Bruder, J., Leppänen, P. H. T., Bartling, J., Csépe, V., Démonet, J.-F. and Schulte-Körne, G., 2011, Children with dyslexia reveal abnormal native language representations: Evidence from a study of mismatch negativity. *Psychophysiology*, **48**, 1107–1118. https://doi.org/10.1111/j.1469-8986.2011.01179.x.

Buzsáki, G., Anastassiou, C. A. and Koch, C., 2012, The origin of extracellular fields and currents — EEG, ECoG, LFP and spikes. *Nature reviews neuroscience*, **13**, 407–420. https://doi.org/10.1038/nrn3241.

Cantiani, C., Choudhury, N. A., Yu, Y. H., Shafer, V. L., Schwartz, R. G. and Benasich, A. A., 2016, From sensory perception to lexical-semantic processing: An ERP study in non-verbal children with autism. *PLOS ONE*, **11**, e0161637. https://doi.org/10.1371/journal.pone.0161637.

Čeponienė, R., Rinne, T. and Näätänen, R., 2002, Maturation of cortical sound processing as indexed by event-related potentials. *Clinical Neurophysiology*, **113**, 870–882. https://doi.org/10.1016/S1388-2457(02)00078-0.

Deveney, C. M., Briggs-Gowan, M. J., Pagliaccio, D., Estabrook, C. R., Zobel, E., Burns, J. L., Norton, E. S., Pine, D. S., Brotman, M. A., Leibenluft, E. and Wakschlag, L. S., 2019, Temporally sensitive neural measures of inhibition in preschool children across a spectrum of irritability. *Developmental Psychobiology*, **61**, 216–227. https://doi.org/10.1002/dev.21792.

Foti, D., Kotov, R. and Hajcak, G., 2013, Psychometric considerations in using error-related brain activity as a biomarker in psychotic disorders. *Journal of Abnormal Psychology*, **122**, 520–531. https://doi.org/10.1037/a0032618.

Friedrich, M., Weber, C. and Friederici, A. D., 2004, Electrophysiological evidence for delayed mismatch response in infants at-risk for specific language impairment. *Psychophysiology*, **41**, 772–782. https://doi.org/10.1111/j.1469-8986.2004.00202.x.

Gabrieli, J. D. E., Ghosh, S. S. and Whitfield-Gabrieli, S., 2015, Prediction as a humanitarian and pragmatic contribution from human cognitive neuroscience. *Neuron*, **85**, 11–26. https://doi.org/10.1016/j.neuron.2014.10.047.

Guttorm, T. K., Leppänen, P. H. T., Hämäläinen, J. A., Eklund, K. M. and Lyytinen, H. J., 2010, Newborn event-related potentials predict poorer pre-reading skills in children at risk for dyslexia. *Journal of Learning Disabilities*, **43**, 391–401. https://doi.org/10.1177/0022219409345005.

Halliday, L. F., Barry, J. G., Hardiman, M. J. and Bishop, D. V., 2014, Late, not early mismatch responses to changes in frequency are reduced or deviant in children with dyslexia: an event-related potential study. *Journal of Neurodevelopmental Disorders*, **6**, 21. https://doi.org/10.1186/1866-1955-6-21.

He, J. B., 2015, Electrocorticogram (ECoG). In D. Jaeger and R. Jung (eds), *Encyclopedia of Computational Neuroscience* (New York, NY: Springer Publishing Company), pp. 1070–1074.

Henderson, L. M., Baseler, H. A., Clarke, P. J., Watson, S. and Snowling, M. J., 2011, The N400 effect in children: Relationships with comprehension, vocabulary and decoding. *Brain and Language*, **117**, 88–99. https://doi.org/10.1016/j.bandl.2010.12.003.

Homan, R. W., Herman, J. and Purdy, P., 1987, Cerebral location of international 10–20 system electrode placement. *Electroencephalography and Clinical Neurophysiology*, **66**, 376–382. https://doi.org/10.1016/0013-4694(87)90206-9.

KEIL, A., DEBENER, S., GRATTON, G., JUNGHÖFER, M., KAPPENMAN, E. S., LUCK, S. J., LUU, P., MILLER, G. A. and YEE, C. M., 2014, Committee report: Publication guidelines and recommendations for studies using electroencephalography and magnetoencephalography. *Psychophysiology*, **51**, 1–21. https://doi.org/10.1111/psyp.12147.

KIESEL, A., MILLER, J., JOLICŒUR, P. and BRISSON, B., 2008, Measurement of ERP latency differences: A comparison of single-participant and jackknife-based scoring methods. *Psychophysiology*, **45**, 250–274. https://doi.org/10.1111/j.1469-8986.2007.00618.x.

KIMURA, M., KATAYAMA, J., OHIRA, H. and SCHRÖGER, E., 2009, Visual mismatch negativity: New evidence from the equiprobable paradigm. *Psychophysiology*, **46**, 402–409. https://doi.org/10.1111/j.1469-8986.2008.00767.x.

KUTAS, M. and HILLYARD, S. A., 1980, Reading senseless sentences: brain potentials reflect semantic incongruity. *Science*, **207**, 203–205. https://doi.org/10.1126/science.7350657.

KUTAS, M. and HILLYARD, S. A., 1984, Brain potentials during reading reflect word expectancy and semantic association. *Nature*, **307**, 161–163. https://doi.org/10.1038/307161a0.

VAN LEEUWEN, T., BEEN, P., KUIJPERS, C., ZWARTS, F., MAASSEN, B. and VAN DER LEIJ, A., 2006, Mismatch response is absent in 2-month-old infants at risk for dyslexia. *NeuroReport*, **17**, 351. https://doi.org/10.1097/01.wnr.0000203624.02082.2d.

LOKEN, E. and GELMAN, A., 2017, Measurement error and the replication crisis. *Science*, **355**, 584–585. https://doi.org/10.1126/science.aal3618.

LUCK, S. J., 2005, Ten simple rules for designing ERP experiments. In T. C. Handy (ed), *Event-related potentials: a methods handbook* (Cambridge, MA: MIT Press), pp. 17–32.

LUCK, S. J., 2014, *An introduction to the event-related potential technique*. 2nd ed. (Cambridge, MA: MIT Press).

LUCK, S. J. and GASPELIN, N., 2017, How to get statistically significant effects in any ERP experiment (and why you shouldn't). *Psychophysiology*, **54**, 146–157. https://doi.org/10.1111/psyp.12639.

LUCK, S. J. and KAPPENMAN, E. S., 2011, *The Oxford handbook of event-related cotential components* (Oxford, England: Oxford University Press).

LUCK, S. J., VOGEL, E. K. and SHAPIRO, K. L., 1996, Word meanings can be accessed but not reported during the attentional blink. *Nature*, **383**, 616–618. https://doi.org/10.1038/383616a0.

MAURER, U., BUCHER, K., BREM, S., BENZ, R., KRANZ, F., SCHULZ, E., VAN DER MARK, S., STEINHAUSEN, H.-C. and BRANDEIS, D., 2009, Neurophysiology in preschool improves behavioral prediction of reading ability throughout primary school. *Biological Psychiatry*, **66**, 341–348. https://doi.org/10.1016/j.biopsych.2009.02.031.

MCCALLUM, W. C., FARMER, S. F. and POCOCK, P. V., 1984, The effects of physical and semantic incongruities on auditory event-related potentials. *Electroencephalography and Clinical Neurophysiology/Evoked Potentials Section*, **59**, 477–488. https://doi.org/10.1016/0168-5597(84)90006-6.

MCGEE, T. J., KING, C., TREMBLAY, K., NICOL, T. G., CUNNINGHAM, J. and KRAUS, N., 2001, Long-term habituation of the speech-elicited mismatch negativity. *Psychophysiology*, **38**, 653–658.

MONJAUZE, C., BROADBENT, H., BOYD, S. G., NEVILLE, B. G. R. and BALDEWEG, T., 2011, Language deficits and altered hemispheric lateralization in young people in remission from BECTS. *Epilepsia*, **52**, e79–e83. https://doi.org/10.1111/j.1528-1167.2011.03105.x.

NÄÄTÄNEN, R. and KREEGIPUU, K., 2011, The Mismatch Negativity (MMN). In *The Oxford handbook of event-related potential components* (Oxford University Press).

NATIONAL CENTER FOR HEARING ASSESSMENT AND MANAGEMENT, 2011, *State early hearing detection and intervention/universal newborn hearing screening mandates* (Utah State University).

NORTON, E. S., GABRIELI, J. D. E. and GAAB, N., 2019, Neural predictors of dyslexia. In L. Verhoeven, C. Perfetti and K. Pugh (eds), *Developmental Dyslexia across Languages and Writing Systems: A Handbook* (Cambridge University Press), pp. 253–276.

PASCUAL-MARQUI, R. D., ESSLEN, M., KOCHI, K. and LEHMANN, D., 2002, Functional imaging with low resolution brain electromagnetic tomography (LORETA): a review. *Clinical Pharmacology*, 12.

PLAKAS, A., VAN ZUIJEN, T., VAN LEEUWEN, T., THOMSON, J. M. and VAN DER LEIJ, A., 2013, Impaired non-speech auditory processing at a pre-reading age is a risk-factor for dyslexia but not a predictor: An ERP study. *Cortex*, **49**, 1034–1045. https://doi.org/10.1016/j.cortex.2012.02.013.

POLICH, J., 2007, Updating P300: An integrative theory of P3a and P3b. *Clinical Neurophysiology*, **118**, 2128–2148. https://doi.org/10.1016/j.clinph.2007.04.019.

POLICH, J., 2011, Neuropsychology of P300. In S. J. Luck and E. S. Kappenman (eds), *The Oxford handbook of event-related potential components* (Oxford, England: Oxford University Press), pp. 159–188.

PRATT, H., 2011, Sensory ERP components. In S. J. Luck and E. S. Kappenman (eds), *The Oxford handbook of event-related potential components* (Oxford University Press), pp. 89–114.

PUCE, A. and HÄMÄLÄINEN, M. S., 2017, A review of issues related to data acquisition and analysis in EEG/MEG studies. *Brain Sciences*, **7**, 58. https://doi.org/10.3390/brainsci7060058.

SANDERS, L. D., NEWPORT, E. L. and NEVILLE, H. J., 2002, Segmenting nonsense: an event-related potential index of perceived onsets in continuous speech. *Nature Neuroscience*, **5**, 700–703. https://doi.org/10.1038/nn873.

SKOE, E. and KRAUS, N., 2010, Auditory brainstem response to complex sounds: a tutorial. *Ear and hearing*, **31**, 302–324. https://doi.org/10.1097/AUD.0b013e3181cdb272.

STOODLEY, C. J., HILL, P. R., STEIN, J. F. and BISHOP, D. V. M., 2006, Auditory event-related potentials differ in dyslexics even when auditory psychophysical performance is normal. *Brain Research*, **1121**, 190–199. https://doi.org/10.1016/j.brainres.2006.08.095.

SUSSMAN, E. and STEINSCHNEIDER, M., 2009, Attention effects on auditory scene analysis in children. *Neuropsychologia*, **47**, 771–785. https://doi.org/10.1016/j.neuropsychologia.2008.12.007.

TANNER, D., MORGAN-SHORT, K. and LUCK, S. J., 2015, How inappropriate high-pass filters can produce artifactual effects and incorrect conclusions in ERP studies of language and cognition. *Psychophysiology*, **52**, 997–1009. https://doi.org/10.1111/psyp.12437.

WEISS, S. and MUELLER, H. M., 2003, The contribution of EEG coherence to the investigation of language. *Brain and Language*, **85**, 325–343. https://doi.org/10.1016/S0093-934X(03)00067-1.

WUNDERLICH, J. L., CONE-WESSON, B. K. and SHEPHERD, R., 2006, Maturation of the cortical auditory evoked potential in infants and young children. *Hearing Research*, **212**, 185–202. https://doi.org/10.1016/j.heares.2005.11.010.