In Press in Long, D.L. (Ed.), *APA Handbook of Research Methods in Psychology*

Event-Related Potentials

Steven J. Luck

Center for Mind & Brain and Department of Psychology

University of California, Davis

Event-related potentials (ERPs) are electrical *potentials* generated by the brain that are *related* to specific internal or external *events* (e.g., stimuli, responses, decisions). They can be recorded noninvasively from almost any group of research participants, and they can provide information about a broad range of cognitive and affective processes. Consequently, the ERP technique has become a common tool in almost all areas of psychological research, and students and researchers must be able to understand and evaluate ERP studies in the literature. However, these studies often involve a set of terms and concepts that are unfamiliar to many psychologists; several technical issues must be understood before a student or researcher can read and evaluate ERP studies. The goal of this chapter is to provide students and researchers with this background information so that they can be informed consumers of ERP studies in their area of interest. More detailed works are available for those who would like to learn more or who would like to conduct their own ERP experiments (Handy, 2005; Luck, 2005).

This chapter begins with an example of a particular ERP component—the N170 wave—and describes how it has been used to address issues ranging from perception and attention to development and neurodevelopmental disorders. This will be followed by an overview of the major ERP components, which will provide both a vocabulary and a sense of the topics that are commonly explored with ERPs. The next sections describe how ERPs are generated in the brain and how the neural generator site of a given ERP can be localized. This is followed by a discussion of the basic technical issues involved in recording and analyzing ERPs, using a study of impaired cognition in schizophrenia patients as a concrete example. The chapter ends with a set of concrete questions that should be asked when reading and evaluating an ERP study.

Example 1: The N170 Component and Face Processing

Figure 1 shows the results of an experiment focusing on the *N170* component, a negative-going wave over visual cortex that typically peaks around 170 ms after stimulus onset. In a typical N170 paradigm, photographs of faces and various types of non-face objects are briefly flashed on a computer monitor and the participants passively view the stimuli. In the ERP waveforms shown in Figure 1A, the X axis represents time relative to stimulus onset (measured in milliseconds [ms]) and the Y axis represents the magnitude of the neural response (in microvolts [μV]). In the scalp map shown in Figure 1B, the color indicates the voltage measured at each electrode site during the time period of the N170 (with interpolated values between the individual electrode sites).

In the early days of ERP research, waveforms were plotted with negative upward and positive downward (largely due to historical accident). Many researchers now use the more common Cartesian convention of plotting positive upward, but this is not universal, so it is important to check which convention is used in a given ERP waveform plot. The waveforms in this chapter are all plotted with positive upward.

The N170 component is notable because it is larger when the eliciting stimulus is a face compared to when the stimulus is a non-face object such as an automobile (see review by Rossion & Jacques, in press). The difference between faces and non-face objects begins approximately 150 ms after the onset of the stimulus; this simple fact allows us to conclude that the human brain is able to distinguish between faces and other objects within 150 ms. The scalp distribution helps us to know that this is the same component that is observed in similar studies of the N170, and it suggests that the N170 generator lies in visual cortex (but note that conclusions based on scalp distributions are not usually definitive).

Many researchers have used the N170 to address interesting questions about how faces are processed in the brain. For example, some studies have asked whether face processing is automatic by testing whether the face-elicited N170 is smaller when the faces are ignored. The results of these experiments indicate that face processing is at least partially automatic (Carmel & Bentin, 2002) but can be modulated by attention under some conditions (e.g., when the faces are somewhat difficult to perceive -- Sreenivasan, Goldstein, Lustig, Rivas, & Jha, 2009). Other studies have used the N170 to ask whether faces are processed in a specialized face module, or whether the same neural process is also used when people process other sorts of complex stimuli for which they have extensive expertise. Consistent with a key role for expertise, these studies have shown that bird experts exhibit an enhanced N170 in response to birds, dog experts exhibit an enhanced N170 in response to dogs, and fingerprint experts exhibit an enhanced N170 in response to fingerprints (Busey & Vanderkolk, 2005; Tanaka & Curran, 2001). Developmental studies have used N170 to track the development of face processing, showing that face-specific processing is present early in infancy but becomes faster and more sophisticated over development (Coch & Gullick, in press). Studies of neurodevelopmental disorders have shown that the N170 is abnormal in children with autism spectrum disorder (Dawson, Carver, Meltzoff, Panagiotides, McPartland, & Webb, 2002).

This example makes several important points. First, it shows that ERPs can be used to address important questions across a wide range of basic science and clinical domains. Second, it illustrates the precise temporal resolution of the

technique. ERPs reflect ongoing brain activity with no delay, and an ERP effect observed at 150 ms reflects neural processing that occurred at 150 ms. Consequently, ERPs are especially useful for answering questions about the timing of mental processes. Sometimes this timing information is used explicitly, by asking whether two conditions or groups differ in the timing of a given neural response (just as one might ask whether reaction time differs across conditions or groups). In other cases, the timing information is used to ask whether a given experimental manipulation influences sensory activity that occurs shortly after stimulus onset or higher-level cognitive processes that occur hundreds of milliseconds later. For example, ERPs have been used to ask whether attentional manipulations influence early sensory processes or whether they instead influence postperceptual memory and decision processes (see, e.g., Luck & Hillyard, 2000). More broadly speaking, ERPs are commonly used to determine which specific cognitive process is influenced by a given experimental manipulation. For example, reaction times (RTs) are slowed when people perform two tasks at the same time compared to when they perform a single task, and ERPs have been used to show that this does not reflect a delay in discriminating the identity of the stimuli (Luck, 1998), but instead reflects a slowing in determining which response is appropriate for the stimulus (Osman & Moore, 1993). ERPs can also be used to assess the anticipatory processes that occur prior to a stimulus (Brunia, van Boxtel, & Böcker, in press) and the performance monitoring processes that occur during and after a behavioral response (Gehring, Liu, Orr, & Carp, in press).

A third important point is that the high temporal resolution of the ERP technique is accompanied by relatively low spatial resolution. The topographic map of the N170 component shown in Figure 1B is very coarse compared to the maps of face-related brain activity provided by functional magnetic resonance imaging (fMRI). For reasons that will be detailed later in this chapter, it is difficult to localize ERPs purely on the basis of the observed scalp distribution, and converging evidence (e.g., lesion data) is usually necessary to know with certainty the neuroanatomical origins of a given ERP effect. For example, several converging sources of evidence indicate that the N170 is generated along the ventral surface of the brain near the border between the occipital and temporal lobes, but it is difficult to be certain that this is the source of the effect in most individual N170 experiments. Thus, ERPs are usually most appropriate for answering questions about timing rather than questions about specific brain regions (although there are some clear exceptions to this generalization).

A fourth key attribute of ERPs is that they can be used to "covertly" monitor mental activity in the absence of a behavioral response. For example, the N170 can be used to assess the ability of preverbal infants to discriminate between different types of faces (e.g., male versus female faces). Similarly, dissociations between ERP activity and behavioral responses can sometimes be very informative. For example, ERPs have been used to show that stimuli that cannot be reported (due to inattention or subliminal presentation) have been processed to the point of activating semantic information (Luck, Vogel, & Shapiro, 1996) and premotor response codes (Dehaene, Naccache, Le Clec'H, Koechlin, Mueller, Dehaene-Lambertz, van de Moortele, & Le Bihan, 1998).

In addition to knowing what kinds of issues can be readily explored with ERPs, it is also useful to know what kinds of issues are *not* easily studied with this technique. As will be discussed in detail later, ERPs are extracted from the electroencephalogram (EEG) by averaging together many trials, using a discrete event such as the onset of a stimulus as a time-locking point. ERPs are typically not useful in situations that make it difficult to perform this averaging process. For example, the averaging process cannot be performed if the mental process being studied is not reasonably well time-locked to a discrete, observable event (e.g., spontaneous emotional responses). In addition, tens or hundreds of trials must typically be averaged together for each condition, and some experimental paradigms do not permit this many repetitions of a given condition (e.g., certain paradigms that require deception). ERPs also tend to be most sensitive to processes that unfold over a period of 2 seconds or less, and slower processes are difficult to see in ERPs (e.g., long-term memory consolidation). Finally, as mentioned previously, ERPs are not usually appropriate for answering neuroanatomical questions.

A final implication of the N170 example is that ERP studies usually focus on specific ERP components. To use ERPs, it is important to learn about the major components, because the components are tools that can be used to address many interesting questions. Moreover, a component that reflects one type of process might be very useful for studying other processes. For example, deficits in executive control resulting from aging and from prefrontal lesions have been studied by examining how the impaired control leads to changes in sensory ERP activity (Chao & Knight, 1997). Similarly, language-related ERP components have been used to study how attention influences perception (Luck et al., 1996), and ERP components related to motor preparation have been used to study syntax (van Turennout, Hagoort, & Brown, 1998). Thus, it is important to acquire a basic vocabulary of the major ERP components across domains.

Before we get to these components, however, it is important to ask what is meant by the term *component* in the context of ERPs. An ERP component can be defined, at least approximately, as a voltage deflection that is produced when a specific neural process occurs in a specific brain region. Many components will be elicited by a stimulus in a given task, and the different components sum together to produce the observed ERP waveform. The observed waveform clearly consists of a set of positive and negative peaks that are related to the underlying components, but the relation is imperfect. For example, the voltage recorded at 170 ms does not reflect a single face-selective N170 component, but instead reflects the sum of all of the components that are active at this time. A full discussion of the methods used to isolate ERP components is beyond the scope of this chapter, but this is an important issue in ERP research (for detailed discussions, see Kappenman & Luck, in press-a; Chapter 2 in Luck, 2005).

A Brief Overview of the Major ERP Components

This section covers the major ERP components, providing both a vocabulary for understanding ERP research and an overview of the breadth of research areas in which ERPs have

been used. ERP components can be divided into three main categories: 1) *Exogenous* sensory components that are obligatorily triggered by the presence of a stimulus (but may be modulated to some degree by top-down processes); 2) *Endogenous* components that reflect neural processes that are entirely task-dependent; and 3) *Motor* components that necessarily accompany the preparation and execution of a given motor responses. This section will cover these three classes of components. Given space limitations, the discussion of each component will necessarily be brief, and many minor components will not be discussed at all. For a comprehensive treatment of the broad range of ERP components, see Luck and Kappenman (in press).

Naming Conventions

Before discussing individual components, it is necessary to say a few words about the naming conventions for ERP components. Unfortunately, the naming is often inconsistent and sometimes ill-conceived. The most common convention is to begin with a P or N to indicate that the component is positive-going or negative-going, respectively. This is then followed by a number indicating the peak latency of the waveform (e.g., N400 for a negative component peaking at 400 ms) or the ordinal position of the peak within the waveform (e.g., P2 for the second major positive peak). This seems like a purely descriptive, theory-free approach, but it is not usually used this way. For example, the term *P300* was coined because it was positive and peaked at 300 ms when it was first discovered (Sutton, Braren, Zubin, & John, 1965). In most studies, however, the same functional brain activity typically peaks between 350 and 600 ms, but this component is still often labeled P300. Many investigators therefore prefer to use a number that represents the ordinal position of the component in the waveform (e.g., P3 instead of P300). This can still be confusing. For example, the first major peak for a visual stimulus is the P1 wave, which is observed over posterior electrode sites with a peak latency of approximately 100 ms. This component is not typically visible at anterior scalp sites, where the first major positive peak occurs at approximately 200 ms. This anterior positive peak at 200 ms is typically labeled *P2*, because it is the second major positive peak overall, even though it is the first positive peak in the waveform recorded at the anterior electrode sites.

Using the polarity to label the component is also problematic, because any given component will produce a positive potential on one side of the head and a negative potential on the other side of the head. The polarity will also depend on which electrode serves as the active site and which electrode serves as the reference site (as discussed in more detail later in this chapter). Moreover, some components vary in polarity depending on the experimental conditions (e.g., the *C1* component inverts in polarity for stimuli presented in the upper visual field compared to stimuli presented in the lower visual field).

Another problem is that a given label may refer to a completely different component when different sensory modalities are considered. For example, the auditory P1 wave bears no special relationship to the visual P1 wave. However, later components are largely modality-independent, and the labels for these components refer to the same brain activity whether the stimuli are auditory or visual. For example, N400 refers to the same brain activity whether the eliciting stimulus is auditory or visual.

Although this convention for naming ERP components can be very confusing to novices, experts usually have no trouble understanding exactly what is meant by these names. This is just like the problem of learning words in natural languages: two words that mean different things may sound exactly the same (homophones); two different words may have the same meaning (synonyms); and a given word may be used either literally or metaphorically. This is certainly an impediment to learning both natural languages and ERP terminology, but it is not an insurmountable problem, and in both cases some work is needed to master the vocabulary.

ERP components are sometimes given more functional names, such as the *syntactic positive shift* (which is observed when the participant detects a syntactic error in a sentence) or the *error-related negativity* (which is observed when the participant makes an obviously incorrect behavioral response). These names are often easier to remember, but they can become problematic when subsequent research shows that the same component can be observed under other conditions. For example, some investigators have argued that the error-related negativity is not directly related to the commission of an error and is present (although smaller) even when the correct response is made (Yeung, Cohen, & Botvinick, 2004).

Exogenous Sensory ERP Components

Figure 2 shows the typical ERP components evoked by the presentation of an auditory stimulus (see review by Pratt, in press). If the stimulus has a sudden onset (such as a click), a distinctive set of peaks can be seen over the first 10 ms that reflect the flow of information from the cochlea through the brainstem and into the thalamus. These *auditory brainstem responses* (ABRs) are typically labeled with Roman numerals (waves I – VI). They are highly automatic and can be used to assess the integrity of the auditory pathways. The ABRs are followed by the *midlatency responses* (MLRs) between 10 and 60 ms, which reflect the flow of information through the thalamus and into auditory cortex. The MLRs are influenced both by sensory factors (e.g., age-related hearing decline) and cognitive factors (e.g., attention). The MLRs are followed by the *long-latency responses*, which typically begin with the P50 (P1), N100 (N1), and P160 (P2). The phrase *long-latency response* is a bit confusing, because these are relative short latencies compared to high-level cognitive components, such as P300 and N400. However, the transmission of information along the auditory pathway is very fast, and 100 ms is a relatively late time from the perspective of auditory sensory processing. The long-latency auditory responses can be strongly influenced by high-level factors, such as attention and arousal.

It should be noted that the midlatency and long-latency auditory responses become much smaller when the interval between successive stimuli decreases, with refractory periods that may exceed 1000 ms (this is true for sensory components in other modalities as well). Moreover, the ERP elicited by one stimulus may not be finished before the next stimulus begins when the interval between stimuli is short, which can also confound the results of an experiment. Thus, when

evaluating an ERP study, it is important to assess whether a difference between groups or conditions might be confounded by differences in the interstimulus interval.

When visual stimuli are presented, the initial ERP response does not begin until approximately 50 ms poststimulus. This greater onset latency for visual relative to auditory stimuli is a result of the relatively long period of time required by the retina to accumulate enough photons to produce a reliable response. The typical scalp ERP waveform for a visual stimulus is shown in Figure 3. The waveforms are shown for the most common ERP paradigm, the *oddball paradigm*. In this paradigm (which is similar to the *continuous performance task*), two classes of stimuli are used, a frequently occurring *standard* stimulus and an infrequently occurring *oddball* stimulus. For example, 80% of the stimuli might be the letter X and 20% might be the letter O. Each stimulus is presented briefly (e.g., 100-200 ms), and the interval between successive stimulus onsets is typically 1000-2000 ms. Participants typically count or make a manual response to the oddball stimuli.

The initial sensory response is usually the same for the standards and the oddballs. It begins with the *C1* wave, which is generated in primary visual cortex and is negative for upper-field stimuli and positive for lower-field stimuli (Clark, Fan, & Hillyard, 1994). The C1 wave is strongly influenced by sensory factors but is not usually influenced by the task. The C1 wave is followed by the P1 wave, which is generated in extrastriate areas of visual cortex and is influenced by sensory factors, attention, and arousal (Hillyard, Vogel, & Luck, 1998; Vogel & Luck, 2000). The P1 is followed by the N1 wave, which consists of several distinct subcomponents. That is, several different brain areas produce negative voltages in the same approximate time range, which sum together to produce the overall N1 voltage. The N1 complex includes the N170 component described earlier. It also includes a subcomponent that is present when the participant attempts to discriminate the identity of the stimulus rather than merely detecting the presence of a stimulus (Vogel & Luck, 2000). These N1 subcomponents are also influenced by attention (Hillyard et al., 1998).

Distinct sensory responses are also produced by somatosensory, olfactory, and gustatory stimuli, as reviewed by Pratt (in press).

The P3 family of components

The most common endogenous ERP component is the P3 or P300 wave (see review by Polich, in press). As illustrated in Figure 3, the most distinctive property of the P3 wave is that it is much larger for infrequently occurring stimulus categories than for frequently occurring stimulus categories. It is most often observed in the oddball paradigm, in which the oddball stimuli elicit a larger P3 than the standard stimuli. Two distinctly different P3 components can be observed. The most common is called P3b, and it is sensitive to *task-defined* probability. That is, it is larger for improbable stimuli only when the task requires sorting the stimuli in a way that makes a given stimulus category improbable. Imagine, for example, an experiment in which the stimuli are the digits 0 through 9, with each digit occurring with equal likelihood. If the participant is asked to count occurrences of the letter 4, then the task requires sorting the stimuli into the "4" category and the "non-4" category. The "4" category will have a probability of .1 and the "non-4" category will have a probability of .9, and the P3b component will be much larger for the "4" category than for the "non-4" category (even though the probability of a 4 is equal to the probability of any other individual digit). This dependence on task-defined category means that task-irrelevant stimuli generate very little P3b activity, and probability along task-irrelevant dimensions does not influence P3b amplitude. For example, if 10% of the stimuli are red and 90% are blue, red and blue stimuli will elicit equivalent P3 waves if color is not relevant for the task.

Because P3b amplitude depends on task-defined probability, the difference in amplitude between the oddball and standard stimuli cannot occur until the brain has begun to determine the category of a given stimulus. As a result, factors that influence the time required to perceive and categorize a stimulus strongly influence the onset and peak latency of the P3 wave, and P3 latency is often tightly tied to RT (for an example, see Luck & Hillyard, 1990). However, RT is often influenced by post-categorization factors, such as the complexity of the stimulus-response mapping, and P3 latency sometimes varies independently of RT (Kutas, McCarthy, & Donchin, 1977). Thus, P3 latency can be used to distinguish between pre- and post-categorization processes (but see Verleger, 1997 for a different perspective).

A different P3 subcomponent—called either *P3a* or the *novelty P3*—is elicited by highly distinctive improbable stimuli, even when the task does not require discrimination of these stimuli. For example, if participants are required to count the Xs in a stream of Xs and Os, and photographs of distinctive scenes are occasionally presented, the scenes will elicit a P3a component even if participants are not required to treat them any differently from the frequent O stimuli. The P3a component has a frontal scalp distribution and is reduced in individuals with lesions of prefrontal cortex, whereas the P3b component is largest over central and parietal electrodes and is reduced in individuals with lesions near the temporal-parietal junction.

The N2 family of components

Several anatomically and functionally distinct components contribute to the overall N2 wave (see review by Folstein & Van Petten, 2008). Like the P3b, the N2c subcomponent of the N2 complex is typically larger for infrequent stimulus categories This component appears to reflect the actual process of categorizing the stimulus (whereas the P3b reflects a process that follows stimulus categorization). The N2c is present for both auditory and visual stimuli, but with quite different scalp distributions.

When auditory stimuli are used, the oddballs also elicit a component that was originally called *N2a* but is now called the *mismatch negativity* or MMN (see review by Näätänen & Kreegipuu, in press). Unlike the N2c and P3b components, the MMN is enhanced for rare stimuli even if the stimuli are task-irrelevant. In a typical MMN study, a sequence of low- and high-pitched tones is presented while the participant reads a book. If the pitch difference is discriminable, and one of the two pitches is less probable than the other, then the oddball pitch will elicit an enhanced negative voltage peaking

around 200 ms at anterior electrode sites. This MMN is very useful for determining whether the auditory system can distinguish between different stimulus categories, especially in participants who cannot easily respond behaviorally to indicate the category of a stimulus. For example, the MMN can be used to determine whether infants of a given age can differentiate between two phonemic categories (Coch & Gullick, in press). The MMN is specific for auditory stimuli.

In the visual domain, the *N2pc* component can be used to track the allocation of spatial attention (Luck, in press). The "pc" in N2pc stands for "posterior contralateral," because the N2pc is observed over posterior scalp sites contralateral to the location of an object that is being attended. As participants shift attention from one side of the display to the other, the N2pc shifts from one hemisphere to the other. In addition, the timing of the N2pc can be used to track how long it takes an individual to find a task-relevant object and shift attention to it. When the attended item must be stored in working memory over a delay interval, a sustained voltage is observed over the delay interval (Perez & Vogel, in press). This *contralateral delay activity* is strongly correlated with individual differences in working memory capacity.

There is also an *anterior N2* component that can be observed at frontal and central electrode sites. This component appears to be sensitive to the mismatch between an expectation and a stimulus, and it is often seen when participants are asked to compare sequentially presented stimuli and the two stimuli mismatch. It is also observed on incompatible trials in the Eriksen flankers task and in the Stroop task; in these situations, the mismatch is between two elements of a single stimulus array (see review by Folstein & Van Petten, 2008). Yeung et al (2004) proposed that this component reflects the operation of a conflict detection system and that this same system is also responsible for the error-related negativity (ERN – for a review, see Gehring et al., in press). That is, the ERN occurs when the conflict is so great that an incorrect response occurs. By this account, the anterior N2 and the ERN are actually the same component.

Language-related ERP components

Several ERP components have been discovered that are related to language comprehension (see review by Swaab, Ledoux, Camblin, & Boudewyn, in press). The most widely used language-related component is the N400, which is typically observed for words that are semantically, lexically, or associatively unrelated to preceding words, phrases, or sentences. For example, if sentences are presented one word at a time (in either the visual or auditory modality), the last word of a sentence will elicit a larger N400 if its meaning fits poorly with the sentence than if it fits well (as in "She sat down on the large, fluffy pencil"). The N400 can also be seen with simple word pairs that vary in their degree of relatedness. For example, the ERP elicited by the word "table" will be larger in the pair "bicycle-table" than in the pair "chair-table". Physical and syntactic deviances do not produce a large change in N400 amplitude.

Syntactic anomalies typically produce a P600 component (sometimes called the *syntactic positive shift*). For example, a larger P600 would be elicited by the word "to" in the syntactically incorrect sentence "The broker persuaded to sell the stock" than in the syntactically correct sentence "The broker hoped to sell the stock." Syntactic anomalies may also produce a *left anterior negativity* (LAN) 300–500 ms after the anomalous word. For example, the LAN is observed when the participant is expecting a word in one syntactic category but instead sees or hears a word in a different category (as in the last word of the sentence "He went outside to take a walking"). The LAN is also larger for words that play a primarily syntactic role (e.g., articles, prepositions) than for words that have strong semantic content (e.g., nouns and verbs).

Memory-related ERP components

As with language, several ERP components have been identified that are related to memory. In working memory paradigms, sustained activity can be observed during the retention interval at frontal electrode sites and—when lateralized visual stimuli are used—over the posterior contralateral electrode sites (Perez & Vogel, in press). In long-term memory paradigms, separate ERPs components have been identified that operate during the encoding and retrieval phases of the task. Encoding-related ERPs are often studied by sorting ERPs that were recorded during the encoding phase according to whether or not a given item was later remembered. Any difference in the ERP between stimuli that were later remembered and stimuli that were later forgotten is called a *Dm* effect (difference due to memory) or a *subsequent memory effect*. In most cases, the Dm effect contains a broad positivity from approximately 400–800 ms over centro-parietal electrode sites. However, it may also contain left anterior activity, and the details of the scalp distribution depend on whether the stimuli were words or pictures and on the instructions given to the participants. Thus, Dm is not a single component, but instead reflects many different processes that can influence whether a stimulus is later remembered.

Two main ERP components have been identified that operate at the time of a recognition judgment. These components are typically identified by comparing the waveforms for stimuli that had been presented during encoding (*old* stimuli) and stimuli that had not (*new* stimuli) or by comparing the waveforms elicited by old stimuli that were correctly judged to be old or incorrectly judged to be new. The two main ERP components that have been observed in such experiments correspond closely with two different mechanisms that have been hypothesized to underlie correct recognition performance. First, an item can be recognized as being from the studied set by a *recollection* process that involves a clear memory of the encoding episode, which may include other incidental information about that episode (e.g., the item that immediately preceded the tested item). When participants recognize an item in this manner, the recognized items elicit a positive voltage that is largest over the left parietal lobe from approximately 400-800 ms (called the *left-parietal old-new effect*). It is also possible to correctly report that an item was previously studied because it creates a sense of familiarity, even if the details of the encoding episode cannot be retrieved. When participants recognize an item on the basis of familiarity, the item elicits a somewhat earlier and more anterior positive voltage from approximately 300-500 ms (called the *mid-frontal old-new effect*).

Emotion-related ERP components

ERP studies of emotion have typically used emotion-inducing pictures as stimuli. The emotional content of the stimuli influences many of the components that have already been described. For example, the P1, N1/N170, N2, and P3 components may all be increased for emotion-inducing stimuli relative to neutral stimuli (see review by Hajcak, Weinberg, MacNamara, & Foti, in press). Two emotion-related components have been the focus of most research. First, the *early posterior negativity* is a negative potential over visual cortex in the N2 latency range that is enhanced for emotion-inducing stimuli, particularly those with a positive valence. This component is thought to reflect the recruitment of additional perceptual processing for emotion-inducing stimuli. Second, the *late positive potential* is a positive voltage that typically has the same onset time and scalp distribution as the P3 wave (i.e., onset around 300 ms and parietal maximum). It may extend for many hundreds of milliseconds and may become more centrally distributed over time. The initial portion may actually consist of an enlarged P3 component, reflecting an effect of the intrinsic task relevance of emotion-inducing stimuli. Interestingly, the amplitude of the late positive potential is correlated with subjective arousal ratings for the stimuli, suggesting that it may reflect subjective emotional experience.

Response-related ERP components

If one creates averaged ERP waveforms time-locked to a motor response rather than time-locked to a stimulus, it is possible to see ERP components reflecting the processes that lead up to the response. If a participant is asked to make self-paced responses every few seconds, a large negative voltage is observed over motor cortex that builds up gradually over a period of several hundred milliseconds. This is called the *Bereitschaftspotential* (BP) or *readiness potential* (RP) (see Brunia et al., in press). This component is also present when participants are presented with stimuli and asked to make speeded responses, but the components reflecting stimulus processing become intermixed with the readiness potential in this situation, making it difficult to isolate the response-related brain activity. However, a portion of the readiness potential is larger over the hemisphere contralateral to the response than over the ipsilateral hemisphere, and the difference in voltage between the two hemispheres can be used to isolate the response-specific activity. This difference is called the lateralized readiness potential (LRP), and it has been widely used to study the processes that are involved in selecting an appropriate response following an imperative stimulus (see Smulders & Miller, in press).

ERP components in special populations

The discussion of ERP components up to this point has focused on studies of healthy young adults. However, ERPs have also been widely used to study typical development across infancy and childhood (Coch & Gullick, in press), to study healthy aging and dementia (Friedman, in press), and to study a variety of psychological disorders, including schizophrenia (O'Donnell, Salisbury, Brenner, Niznikiewicz, & Vohs, in press) and affective disorders (Bruder, Kayser, & Tenke, in press). In the context of infants and young children, ERPs are particularly useful because these individuals have relatively poor control over their behavior, and the ERPs can reveal mental processes that are difficult to assess behaviorally. ERPs are relatively well tolerated in infants and young children, for whom fMRI is not a realistic option. In the domain of aging, ERPs are useful for determining whether the overall slowing of responses reflects slowing in specific processes (e.g., perceptual versus motor processes). In the context of mental health disorders, ERPs can be useful in determining exactly which processes are impaired (by determining which components are changed). In addition, ERPs can potentially be used as biomarkers to define specific treatment targets and assess the effectiveness of new treatments (Javitt, Spencer, Thaker, Winterer, & Hajos, 2008; Luck, Mathalon, O'Donnell, Hämäläinen, Spencer, Javitt, & Ulhaaus, submitted). Moreover, many human ERP components have animal homologues, creating opportunities for translating between animal and human research.

Neural Origins of ERPs

In almost all cases, ERPs originate as postsynaptic potentials (PSPs), which occur during neurotransmission when the binding of neurotransmitters to receptors changes the flow of ions across the cell membrane. ERPs are not associated with action potentials except for a few of the very earliest, subcortical sensory responses. When PSPs occur at the same time in large numbers of similarly oriented neurons, they summate and are conducted at nearly the speed of light through the brain, meninges, skull, and scalp. Thus, ERPs provide a direct, instantaneous, millisecond-resolution measure of neurotransmission-mediated neural activity. This contrasts with the blood oxygen level dependent (BOLD) signal in fMRI, which reflects a delayed, secondary consequence of neural activity. Moreover, the close link to neurotransmission make ERPs potentially valuable as biomarkers in studies of pharmacological treatments.

When a PSP occurs within a single neuron, it creates a tiny electrical dipole (an oriented flow of current). Measureable ERPs can be recorded at the scalp only when the dipoles from many thousands of similarly oriented neurons sum together. If the orientations of the neurons in a given region are not similar to each other, the dipoles will cancel out and will be impossible to detect at a distant electrode. The main neurons that have this property are the pyramidal cells of the cerebral cortex, which are the main input-output cells of the cortex. That is, these cells are oriented perpendicular to the cortical surface, and their dipoles therefore add together rather than cancelling out. Consequently, scalp-recorded ERPs almost always reflect neurotransmission that occurs in these cortical pyramidal cells. Nonlaminar structures such as the basal ganglia do not typically generate ERPs that can be recorded from the scalp, nor do interneurons within the cortex. Thus, only a fraction of brain activity leads to detectable ERP activity on the scalp.

ERP components can be either positive or negative at a given electrode site. The polarity depends on a combination of at least four factors: 1) the orientation of the neurons with respect to the recording electrode; 2) the location of the reference electrode; 3) the part of the cell in which the neurotransmission is occurring (the apical dendrites or the basal dendrites); 4) whether the neurotransmission is excitatory or

inhibitory. If three of these factors were known, then the fourth could be inferred from the polarity of the ERP component. However, one almost never knows three of these factors, so it is usually impossible to draw strong conclusions from the polarity of an ERP component.

When the dipoles from many individual neurons sum together, they can be represented quite accurately with a single *equivalent current dipole* that is the vector sum of the individual dipoles. For the rest of this chapter, the term *dipole* will refer to these summed equivalent current dipoles.

The voltage recorded on the surface of the scalp will be positive on one side of the dipole and negative on the other, with a single line of zero voltage separating the positive and negative sides (see Figure 4A). The voltage field spreads out through the conductive medium of the brain, and the high resistance of the skull and the low resistance of the overlying scalp lead to further spatial blurring. Thus, the voltage for a single dipole will be fairly broadly distributed over the surface of the scalp, especially for ERPs that are generated in relatively deep cortical structures, such as the cingulate cortex. This can be seen in the more diffuse voltage distribution for the relatively deep dipole in Figure 4B compared to the relatively superficial dipole in Figure 4A.

Electrical dipoles are always accompanied by magnetic fields, but the skull is transparent to magnetism, leading to less blurring of the magnetic fields. Consequently, it is sometimes advantageous to record the magnetic signal (the magnetoencephalogram or MEG) rather than—or in addition to—the electrical signal (the EEG). However, MEG recordings require extremely expensive equipment and are much less common than EEG recordings.

ERP Localization

When a single dipole is present, one can use the observed scalp distribution to estimate the location and orientation of the dipole with good accuracy unless the dipole is relatively deep in the brain or the data are noisy (for an overview of ERP localization techniques, see Chapter 7 in Luck, 2005). When multiple dipoles are simultaneously active, they simply sum together. That is, the voltage distribution for two dipoles will simply be the sum of the two individual distributions. For example, Figure 4C shows the same dipole as in Figure 4A plus another dipole, and clear voltage foci can be seen over each dipole. Some precision is lost in localizing two dipoles together, but localization can still be reasonably accurate as long as the dipoles are relatively far apart and the noise level is low. However, it can be difficult to separately localize two dipoles that are similar in orientation and fall within several cm of each other. For example, the scalp distribution for the two dipoles in Figure 4D is nearly identical to the distribution of the single dipole in Figure 4A. As more and more simultaneous dipoles are added, it becomes more and more difficult to determine how many dipoles are present and to localize them, especially when the data are noisy. Under these conditions, a set of estimated dipole locations that matches the observed scalp distribution can be quite far from the actual locations.

In many experiments, the number of dipoles could be very large, and localizing ERPs solely on the basis of the observed scalp distribution becomes impossible. Formally speaking, the number of internal generator configurations that could explain an observed voltage distribution is infinite (Helmholtz, 1853). In other words, there is no unique solution to the problem of determining the internal generators solely on the basis of the observed scalp distribution. The only way to localize ERPs in this case is to add external constraints, and this is how existing procedures for localizing ERPs solve the non-uniqueness problem. For example, some common procedures allow the user to simply specify the number of dipoles (Scherg, 1990). Other procedures use structural MRI scans and constrain the dipoles to be in the gray matter. However, this constraint is still not enough to produce a unique solution, so these procedures include additional constraints, such as choosing the solution that is minimizes sudden changes from one patch of cortex to the next (Pascual-Marqui, Esslen, Kochi, & Lehmann, 2002). Although these constraints produce a unique solution, they do not necessarily produce the correct solution.

Generally speaking, the most significant shortcoming of mathematical procedures for localizing ERPs is that they do not typically provide a well-justified margin of error. That is, they do not indicate the probability that the solution is incorrect by more than some number of millimeters. Without a margin of error, it is difficult to judge the credibility of a given localization estimate. In most cases, the strongest claim that can be made is that the observed data are consistent with a given generator location. However, ERP papers often state that the ERPs were "localized" to a certain brain region, as if the localization procedure simply found with certainty the actual location of the generator. Indeed, some papers show the estimated waveforms from specific areas of the brain as if those waveforms were the actual data, without showing the actual observed waveforms from the electrodes. One should be cautious when evaluating studies in which the conclusions rely heavily on these approaches to localizing ERPs, especially when multiple generators are likely to be present.

Although it is usually impossible to definitively localize ERPs solely on the basis of the observed scalp distributions, this does not mean that ERPs can never be localized. Although mathematical localization procedures are usually insufficient, ERPs can be localized using the general hypothesis-testing approach that is used throughout psychology. That is, a hypothesis about the generator location for a given ERP effect leads to a set of predictions, which are then tested by means of experiments. One prediction, of course, is that the observed scalp distribution will be consistent with the hypothesized generator location. However, confirming this prediction is not usually sufficient to have strong confidence that the hypothesis about the generator location is correct. Thus, it is important to test additional predictions. For example, one could test the prediction that damage to the hypothesized generator location eliminates the ERP component. Indeed, researchers initially hypothesized that the P3 component was generated in the hippocampus, and this hypothesis was rejected when experiments demonstrated that the P3 is largely intact in individuals with medial temporal lobe lesions (see review by Polich, in press). Similarly, one could predict that an fMRI experiment should show activation in the hypothesized generator location under the conditions that pro-

duce the ERP component (see, e.g., Hopf, Luck, Boelmans, Schoenfeld, Boehler, Rieger, & Heinze, 2006). It is also possible to record ERPs from the surface of the cortex in neurosurgery patients, and this can been used to test predictions about ERP generators (see, e.g., Allison, McCarthy, Nobre, Puce, & Belger, 1994). This hypothesis-testing approach has been quite successful in localizing ERP components.

Example 2: Impaired Cognition in Schizophrenia

This section will provide a somewhat more detailed discussion of a specific experiment, in which ERPs were used to study impaired cognition in schizophrenia (Luck, Kappenman, Fuller, Robinson, Summerfelt, & Gold, 2009). This will serve both to show how ERPs can be used to isolate specific cognitive processes and to provide a concrete example that will be used in the following sections, which focus on the technical details that one must understand to read and evaluate published ERP studies.

The goal of this example experiment was to ask why RTs are typically slowed in schizophrenia patients when they perform simple sensorimotor tasks. That is, are RTs slowed because of an impairment in perceptual processes, in decision processes, or in response processes? ERPs are ideally suited for answering this question, because they provide a direct means of measuring the timing of the processes that occur between a stimulus and a response. On the basis of prior research, we hypothesized that the slowing of RTs in schizophrenia in simple tasks does not result from slowed perception or decision, but instead results from an impairment in the process of determining which response is appropriate once the stimulus has been perceived and categorized (the *response selection* process).

To test this hypothesis, we recorded ERPs from 20 individuals with schizophrenia and 20 healthy control participants in a modified oddball task (Luck et al., 2009). In each 5-minute block of trials, a sequence of letters and digits was presented at fixation. One stimulus was presented every 1300-1500 ms, and participants made a button-press response for each stimulus, pressing with one hand for letters and with the other hand for digits. One of these two categories was rare (20%) and the other was frequent (80%) in any given trial block. Both the category probabilities and the assignment of hands to categories was counterbalanced across trial blocks.

This design allowed us to isolate specific ERP components by means of *difference waves*, in which the ERP waveform elicited by one trial type is subtracted from the ERP waveform elicited by another trial type (much like difference images in fMRI studies). Difference waves are valuable because they isolate neural processes that are differentially active for two trial types, separating these processes from the many concurrently active brain processes that do not differentiate between these trial types. In the current study, difference waves were used to isolate the P3 wave (subtracting frequent trials from rare trials) and the lateralized readiness potential or LRP (by subtracting ipsilateral electrode sites from contralateral electrode sites, relative to the responding hand). The P3 difference wave reflects the time course of stimulus categorization (e.g., determining whether the current stimulus falls into the rare or frequent category), whereas the LRP difference wave reflects the time course of response selection following stimulus categorization (e.g., determining whether the left button or right button is the appropriate response for the current stimulus). We found that RTs were slowed by approximately 60 ms in patients compared to control participants, and the question was whether this reflects a slowing of perception and categorization (which would be seen in the P3 difference wave) or whether it reflects a slowing of post-categorization response selection processes (which would be seen in the LRP difference wave).

Figure 5 shows the P3 difference waves (rare minus frequent) and the LRP difference waves (contralateral minus ipsilateral for the frequent stimulus category) overlaid for the patients and control participants. These are *grand average* waveforms, meaning that average waveforms were first computed across trials for each participant, and then these waveforms were averaged together for viewing the data. In the grand average waveforms, the P3 wave was virtually indistinguishable for patients versus controls (although the preceding N2 was diminished in the patients). In contrast, the LRP was delayed by 75 ms in onset time and diminished by 50% in amplitude for patients versus controls. Moreover, the degree of amplitude reduction across patients was significantly correlated with the degree of RT slowing. Thus, for a relatively simple perceptual task, the slowed RTs exhibited by the schizophrenia patients appear to result primarily from a slowing of response selection (as evidenced by the later and smaller LRP) rather than a slowing of perception or categorization (as evidenced by no slowing or reduction of the P3).

Recording the Electroencephalogram

We will now turn to the technical details of how ERPs are recorded and analyzed, using the schizophrenia experiment as an example. ERPs are extracted from the EEG, so we will begin by discussing how the EEG is recorded. The EEG is a fluctuating electrical potential (pellets) on the scalp, with a conductive gel or liquid between the electrode and the skin to make a stable electrical connection. The electrical potential (voltage) can then be recorded from each electrode, resulting in a separate waveform from each electrode, with time on the X axis and voltage on the Y axis (see Figure 6B). This waveform will be a mixture of actual brain activity, artifactual electrical potentials produced outside of the brain (by the skin, the eyes, the muscles, etc.) and induced electrical activity from external sources (e.g., video monitors) that are picked up by the head, electrodes, or electrode wires. If precautions are taken to minimize the non-neural potentials, the voltage produced by the brain (the electroencephalogram or EEG) will be relatively large compared to the non-neural potentials.

However, the necessary precautions are not always taken, and studies that fail to control these sources of noise may have poor statistical power. The impact of the noise on the data can be evaluated by examining the prestimulus baseline period in the waveforms. In a well-designed experiment, any differences between conditions prior to stimulus onset must be caused by noise. In Figure 5, for example, the waveforms include a 200-ms prestimulus baseline period, and although the waveforms are not perfectly flat during this prestimulus period, the differences between patients and controls during

this period are much smaller than the P3 and LRP deflections. As a rule of thumb, one should be cautious if a paper reports significant differences between conditions or groups in some poststimulus interval when there are differences of comparable magnitude in the prestimulus interval.

The EEG is quite small (usually under 100 microvolts [μV]), so the signal from each electrode is usually amplified by 1,000-100,000 times (an amplifier *gain* of 5,000 was used in the experiment shown in Figure 5). The continuous voltage signal is then turned into a series of discrete digital values for storage in a computer. In most experiments, the voltage is sampled from each channel at a rate of between 200 and 1000 evenly spaced samples per second (Hz; see Figure 6C). In the experiment shown in Figure 5, the EEG was sampled at 500 Hz (1 sample every 2 ms). In addition, filters are usually used to remove very slow voltage changes (< 0.01–0.1 Hz) and very fast voltage changes (>15–100 Hz), because scalp-recorded voltages in these frequency ranges are likely to be noise from non-neural sources. Frequencies below 0.1 Hz and above 18.5 Hz were filtered from the waveforms shown in Figure 5. Filters can dramatically distort the time course of an ERP waveform and can induce artifactual oscillations when the low cutoff is greater than approximately 0.5 Hz or when the low cutoff is less than approximately 10 Hz, so caution is necessary when extreme filters are used.

The EEG is typically recorded from multiple electrodes distributed across the scalp. The standard nomenclature for electrode sites is shown in Figure 6A. Each electrode name begins with 1-2 letters denoting a general brain region (Fp for frontal pole, F for frontal lobe, C for central sulcus, P for parietal lobe, O for occipital lobe, T for temporal lobe). The letters are followed by a number that reflects the distance from the midline (1 is close the midline; 5 is far from the midline). Odd numbers are used for the left hemisphere and even numbers are used for the right, with 'z' for zero when the electrode is on the midline. Thus, F3 lies over frontal cortex to the left of midline, Fz lines over frontal cortex on the midline, and F4 lies over frontal cortex to the right of midline. Different studies use very different numbers of electrodes. For some studies, almost all of the relevant information can be obtained from 5-6 electrodes; for others, as many as 256 electrodes are needed. Although it might be tempting to assume that more is better, it is actually more difficult to ensure that high quality data are being recorded when the number of electrodes becomes large, and methods for rapidly applying large numbers of electrodes may lead to poorer signal quality and lower statistical power (Kappenman & Luck, in press-b). An intermediate number of electrodes (10-64) is best for most studies. Only 13 scalp sites were used in the study shown in Figure 5. Because ERPs are spatially blurred by the skull, it is very unlikely that an effect will be missed due to insufficient sampling of the scalp unless the number of electrodes is very small.

It is important to note that voltage is the potential for electrical charges to move between two locations, and the EEG is therefore measured as the voltage between two electrodes. One is called the *active* electrode and the other is called the *reference* electrode, and a single reference electrode is typically used for all of the scalp electrodes. The reference is often placed at a location such as the earlobe, the mastoid process (a bony protrusion behind the ear), or the tip of the nose. These sites are sometimes thought to be electrically neutral, with all of the brain activity originating from the active electrode. However, this is a misconception, and there is no electrically neutral location. Thus, it is important to realize that the voltage attributed to a given site is really the potential between two sites, and brain activity at both the active and reference sites contribute to the recorded signal. This problem can be partially solved by using the average across all electrodes as the reference, but the effectiveness of this depends on whether a sufficiently broad range of electrode sites is used (Dien, 1998). Thus, when reading the method section of a published ERP study, it is important to see what reference site was used. The average of the left and right earlobes was used in the study shown in Figure 5.

Artifact Rejection and Correction

There are several common artifacts that are picked up by EEG recordings and require special treatment. The most common of these arise from the eyes. Whenever the eyes blink, a large voltage deflection is observed over the front of the head. This artifact is usually much larger than the ERP signals. Moreover, eyeblinks are sometimes systematically triggered by tasks and may vary across groups or conditions, yielding a systematic distortion of the data. In addition, large potentials are produced by eye movements, and these potentials can confound experiments that use lateralized stimuli or focus on lateralized ERP responses. In most ERP experiments, the participants are instructed to maintain fixation on a central point and minimize eyeblinks. However, most participants cannot avoid blinking entirely, and they may be unable to avoid making eye movements toward lateralized stimuli. Thus, trials containing blinks, eye movements, or other artifacts are typically excluded from the averaged ERP waveforms. In the study shown in Figure 5, for example, three patients and two controls were excluded from the final analysis because more than 50% of trials were rejected (mainly due to blinks). In the remaining participants, an average of 23% of trials was rejected.

This approach has two shortcomings. First, a fairly large number of trials may need to be rejected, thus reducing number of trials remaining in the averaged ERP waveforms. Second, the mental effort involved in suppressing eyeblinks may impair task performance (Ochoa & Polich, 2000). These problems are especially acute in individuals with neurological or psychiatric disorders, who may blink on almost every trial or may perform the task poorly because of the effort devoted to blink suppression. Fortunately, methods have been developed to estimate the artifactual activity and subtract it out, leaving artifact-free EEG data that can be included in the averaged ERP waveforms. Some of these artifact correction techniques are known to make systematic errors in estimating and removing the artifactual activity (see, e.g., Lins, Picton, Berg, & Scherg, 1993), but many of these techniques work reasonably well.

It is important to note, however, that these techniques correct the electrical artifact that is directly produced by an eyeblink or eye movement, but it is impossible to correct for the change in sensory input produced by these events. If an eyeblink causes the eyes to be closed when the stimulus is pre-

sented, then the ERPs will be radically changed by the absence or delay of sensory processing, and removing the electrical potential produced as the eyelid slides over the cornea cannot correct for this change in sensory processing. Thus, the best approach is to reject trials on which blinks or eye movements occurred at a time when they might change the sensory input, but to correct for the artifactual voltage when the timing of the blink or eye movement should not change task performance.

Extracting Averaged ERPs from the EEG

ERPs are typically small in comparison with the rest of the EEG activity, and ERPs are usually isolated from the ongoing EEG by a simple averaging procedure. To make this possible, it is necessary to include *event codes* in the EEG recordings that mark the events that happened at specific times, such as stimulus onsets (Figure 6A). These event codes are then used as a time-locking point to extract segments of the EEG surrounding each event.

To illustrate this, Figure 6 shows the EEG recorded over a 9-s period in an oddball task with infrequent X stimuli (20%) and frequent O stimuli (80%). Each box highlights the 800-ms segment of EEG following one of these stimuli. Figure 6D shows these same segments of EEG, lined up in time. Stimulus onset is time zero. There is quite a bit of variability in the EEG waveforms from trial to trial, and this variability largely reflects the fact that the EEG reflects the sum of many different sources of electrical activity in the brain, many of which are not involved in processing the stimulus. To extract the activity that is related to stimulus processing from the unrelated EEG, the EEG segments following each X are averaged together into one waveform, and the EEG segments following each O are averaged together into a different waveform (Figure 6E). Any brain activity that is not time-locked to the stimulus will be positive at a given latency on some trials and negative at that latency on other trials, and if many trials are averaged together, these voltages will cancel each other out and approach zero. However, any brain activity that is consistently elicited by the stimulus—with approximately the same voltage at a given latency from trial to trial—will remain in the average. Thus, by averaging together many trials of the same type, the brain activity that is consistently time-locked to the stimulus across trials can be extracted from other sources of voltage (including EEG activity that is unrelated to the stimulus and non-neural sources of electrical noise). Other types of events can be used as the time-locking point in the averaging process (e.g., button-press responses, vocalizations, saccadic eye movements, electromyographic activity).

How many trials must be averaged together? That depends on several factors, including the size of the ERP response of interest, the amplitude of the unrelated EEG activity, and the amplitude of non-neural activity. For large components, such as the P3 wave, very clear results can usually be obtained by averaging together 10-30 trials. For smaller components, such as the P1 wave, it is usually necessary to average together 100-500 trials for each trial type to see reliable differences between groups or conditions. Of course, the number of trials that is required to observe a significant difference will also depend on the number of participants and the magnitude of the difference between conditions. Also, as discussed earlier, looking at the prestimulus baseline period in the ERP waveforms can be useful in evaluating whether enough trials were averaged together to minimize noise. In the experiment shown in Figure 5, each participant received 256 oddball stimuli and 1024 standard stimuli. This is more trials than would be typical for a P3 study, but it was appropriate given that we were also looking at the much smaller LRP and that we anticipated rejecting a large percentage of trials due to eyeblinks.

Although the averaging procedure can be extremely useful in extracting consistent brain responses from the EEG, it is based on a key assumption that is not always valid. Specifically, averaging the EEG segments across trials will work well only if the timing of the neural response is the same across trials. Figure 7A shows an example of several single trials in which the latency varies substantially from trial to trial. The average across these trials begins at the onset time of the earliest single trials and ends at the offset time of the latest single trials, and the peak amplitude of the average is much smaller than the peak amplitude of the individual trials. Figure 7B shows an example with less variability in latency, resulting in an average that is less broad and has a greater peak amplitude. Thus, if the averaged ERPs are compared for two conditions in which the single-trial ERPs are of equivalent amplitude, but one condition has greater latency variability, the difference in the peak amplitudes of the averaged waveforms might lead to the incorrect conclusions that these conditions differ in the magnitude of the ERP response when in fact they differ in the timing of the response.

This can be a significant problem in practice, especially when a patient group is compared with a control group, because the patient group might appear to have a smaller amplitude as a result of greater variability in timing. There are several ways to address this problem (see Chapter 4 in Luck, 2005). The simplest is to measure the amplitude of an ERP component as the mean voltage over a broad time range rather than as the peak voltage, because the mean amplitude is not influenced by latency variability (with one exception, described in the next paragraph).

Figure 7C shows a situation that is even more problematic. In this example, each stimulus elicits a sequence of two sinusoidal oscillations. The first oscillation is phase-locked to the stimulus (e.g., the oscillation starts at the same part of the sine wave on each trial), and this oscillation is captured well in the averaged waveform. The second oscillation, however, varies in phase from trial to trial. Consequently, even though the oscillation occurs in the same general time range on each trial, the voltage at a given time point is positive on some trials and negative on other trials, leading to nearly complete cancellation in the averaged waveform. Using mean amplitude to quantify the amplitude of the response in the averaged waveform does not work in this example, because the single-trial waveform has both positive and negative parts; mean amplitude is effective in the face of latency or phase variability only for monophasic ERPs.

There is, however, a solution that works for oscillations such as those shown in Figure 7C. As shown in Figure 7D, it is possible to convert the data into a *time-frequency* representation on each trial, which quantifies the power present in

different frequency bands at each point in time. The power of a frequency band is represented independently of its phase. Consequently, when the time-frequency representations are averaged across trials, the phase variation does not cause cancellation of the power. These time-frequency analyses have become quite popular because they can reveal brain activity that is lost by conventional averaging (for a review, see Bastiaansen, Mazaheri, & Jensen, in press). Although this approach is extremely useful, the results are often over-interpreted. The main problem is that non-oscillating brain activity also produces power in time-frequency analyses, and it can be quite difficult to distinguish between true oscillations and transient, non-oscillating activity. Thus, one should be cautious when a study makes claims about oscillations from time-frequency analyses. That is, these analyses reveal real neural activity that would be obscured by conventional averaging, but they do not usually prove that the neural activity consists of bona fide oscillations.

Quantification of Component Magnitude and Timing

The most common way to quantify the magnitude and timing of a given ERP component is to measure the amplitude and latency of the peak value within some time window. For example, to measure the peak of the P3 wave in the data shown in Figure 5, one would define a measurement window (e.g., 400-700 ms) and find the most positive point in that window. Peak amplitude would be the voltage at this point, and peak latency would be defined as the time of this point (it is also possible to search for negative peaks). This was the simplest approach to measuring ERPs prior to the advent of inexpensive computers, when a ruler was the only available means of quantifying the waveform. This approach is still widely used, but it has several drawbacks. First, there is nothing special about the point where the waveform reaches an extreme value, and the peak does not represent the magnitude or timing of the entire component. Second, because peak measures are based on extremes, they tend to be sensitive to noise. Third, peak measures are not linear, so the peak in an average waveform will not be the same as the average of the peaks from the individual trials. This makes peak amplitude highly sensitive to trial-to-trial latency variability, and it can also result in grand averages that are not representative of the waveforms from the individual participants. Fourth, peak measures can be greatly influenced by overlapping ERP components, making it difficult to know whether a given effect truly reflects the component of interest.

Because of these limitations, other methods for quantifying ERP amplitudes and latencies have been developed. For measuring the magnitude of a component, it is possible to simply measure the mean voltage over a given time window. This captures all or most of a component, not just the most extreme value, and it is less sensitive to noise than peak amplitude. In addition, mean amplitude is a linear measure, so the mean voltage measured from the waveforms on multiple single trials and then averaged together will be equal to the mean voltage measured from the averaged waveform, and trial-to-trial latency variability will have no effect on the measured amplitude (for monophasic components). Thus, mean amplitude is almost always superior to peak amplitude as a measure of the magnitude of a component.

A related measure can be used to quantify component latency. Specifically, it is possible to define the midpoint of a component as the point that divides the region under the waveform into two equal-area subregions. This is called the 50% area latency measure, and it was used to quantify the timing of the P3 wave in the data shown in Figure 5. Measuring the onset latency of a component is more difficult, because the onset is the point at which the signal is infinitesimally greater than the noise. It is possible to use a *25% area latency measure*, which finds the time point that divides the region under the waveform into the first 25% and second 75%. Another approach is to find the peak amplitude and then find the time of the first point that exceeds 50% of that amplitude. This is the approach that was used to measure LRP onset latency in Figure 5. These approaches tend to be both more accurate and more sensitive than other approaches for quantifying component timing (Kiesel, Miller, Jolicoeur, & Brisson, 2008).

Statistical Analysis

In most ERP experiments, an averaged ERP waveform is constructed at each electrode site for each subject in each condition. The amplitude or latency of a component of interest is then measured in each one of these waveforms, and these measured values are then entered into a statistical analysis just like any other variable. Thus, the statistical analysis of ERP data is not usually very different from the analysis of traditional behavioral measures.

One issue, however, is important to consider when reading published ERP studies. Specifically, ERP experiments provide extremely rich data sets, usually consisting of several gigabytes of data. This can lead to both the implicit and explicit use of many statistical comparisons per study, which can dramatically increase the probability of a Type I error (i.e., concluding that a difference is real when it was actually a result of sampling error or measurement error). The explicit use of multiple comparisons arises when, for example, separate statistical analyses are reported for several different components. The implicit use of multiple comparisons occurs when researchers conduct many different analyses and then report only a subset (mainly those that yielded significant results). A related problem occurs when researchers first look at the waveforms and then decide on the time windows to be used for quantifying component amplitudes and latencies. If a time window is chosen because the difference between conditions is greatest in that time window, then this biases the results in favor of statistical significance, even if the difference was caused by noise. An analogous problem arises in studies using large number of electrode sites, when the sites with the largest differences between conditions are chosen for the statistical analyses. With enough electrode sites, it is almost always possible to find a statistically significant difference between two groups or two conditions at a few electrode sites due simply to random noise. Thus, one should be suspicious if unusual, idiosyncratic, and unjustified electrode sites or measurement windows are selected for the statistical analyses.

A second important statistical issue in the analysis of ERP data arises because nearby electrodes are almost always more correlated with each other than distant electrodes. When

electrode site is entered as a within-subjects factor in an analysis of variance (ANOVA), this produces heterogeneity of covariance, which increases the Type I error rate. That is, the actual probability of falsely rejecting the null hypothesis is higher than indicated by the p value. This problem can be addressed in several ways {see chapter 6 in \Luck, 2005 #6038}, but the most common approach is to use the Greenhouse-Geisser epsilon correction, which produces an adjusted p value that more closely reflects the actual probability of a Type I error. Other factors can also produce heterogeneity of covariance, so this adjustment is used even when electrode site is not entered into the ANOVA.

Concluding Comments and a List of Questions to Consider in Evaluating ERP Studies

The ERP technique is extremely valuable for answering questions about the processes that lead up to and follow a behavioral response, providing information that cannot be obtained from any other noninvasive technique. In addition, ERPs are very useful for evaluating cognitive and affective processes in individuals who cannot easily perform complex tasks, and they can be used to reveal processes that are not evident in overt behavior. Moreover, ERPs can be very useful in evaluating pharmacological interventions because they reflect the PSPs generated during neurotransmission. However, many technical factors can prevent a given ERP study from reaching strong conclusions. The following is a summary of questions one should ask about these technical factors when evaluating an ERP study.

1. Are there substantial voltage deflections during the prestimulus baseline period? If so, then the noise level may have been too high or the number of trials averaged together may have been too low, and the reported differences between groups or conditions may be spurious.
2. Could differences in interstimulus interval confound a comparison between conditions or groups, either due to changes in sensory responsiveness or overlapping activity from the previous trial?
3. What reference site was used? It is important to remember that the voltage at a given electrode reflects the potential between that site and the reference electrode
4. What were the filter settings? Extreme filter settings can cause large temporal distortions and artificial oscillations. Be especially cautious if the cutoff for low frequencies is greater than 0.1 Hz.
5. If artifact rejection was used, how many trials were rejected per participant? If artifact correction was used, might blinks or eye movements have changed the sensory input in a manner that confounded the experiment?
6. How many trials were averaged together for each condition? For large components such as P3 and N400, this should typically be 10-50. For small components such as P1 and N1, this should typically be 100-500.
7. Might differences in peak amplitudes in the averaged ERP waveforms be a result of differences in latency variability rather than true differences in the magnitude of the single-trial ERP responses?
8. Does the study imply that the generator source of a given effect is known with certainty? If so, is this well justified?
9. Does the study conclude that oscillations were present in a given frequency band simply because a time-frequency analysis indicated that significant power was present in that frequency band? Even transient, non-oscillating brain responses can produce such effects.
10. Could changes in the ERP waveform that are attributed to changes in a specific ERP component actually be a result of changes in some other component?
11. Were peak measures used to quantify the magnitude and timing of an ERP component? If so, then this may have reduced the accuracy and statistical power of the study.
12. Were unusual, idiosyncratic, and unjustified measurement windows and electrode sites chosen for the statistical analysis? If so, the results may be spurious, and a replication may be necessary for the conclusions to be believable.

.

Author Notes

Preparation of this chapter was supported by grants R01MH076226, R01MH065034, R01MH087450, and R25 MH080794 from the National Institute of Mental Health. Correspondence should be addressed to Steven J. Luck, Center for Mind & Brain, University of California, Davis, CA 95618 (e-mail:sjluck@ucdavis.edu).

References

Allison, T., Goff, W. R., Williamson, P. D., & Van Gilder, J. C. (1980). On the neural origin of early components of the human somatosensory evoked potential. In J. Desmedt (Ed.), *Clinical Uses of Cerebral, Brainstem and Spinal Somatosensory Evoked Potentials* (pp. 51-68): Karger Basel.

Allison, T., McCarthy, G., Nobre, A., Puce, A., & Belger, A. (1994). Human extrastriate visual cortex and the perception of faces, words, numbers, and colors. Cerebral Cortex, 5, 544-554.

Bastiaansen, M., Mazaheri, A., & Jensen, O. (in press). Beyond ERPs: Oscillatory neuronal dynamics. In S. J. Luck & E. S. Kappenman (Eds.), Oxford Handbook of ERP Components. New York: Oxford University Press.

Bruder, G. E., Kayser, J., & Tenke, C. E. (in press). Event-related brain potentials in depression: Clinical, cognitive and neurophysiologic implications. In S. J. Luck & E. S. Kappenman (Eds.), Oxford Handbook of Event-Related Potential Components. New York: Oxford University Press.

Brunia, C. H. M., van Boxtel, G. J. M., & Böcker, K. B. E. (in press). Negative slow waves as indices of anticipation: The bereitschaftspotential, the contingent negative variation, and the stimulus preceding negativity. In S. J. Luck & E. S. Kappenman (Eds.), Oxford Handbook of Event-Related Potential Components. New York: Oxford University Press.

Busey, T. A., & Vanderkolk, J. R. (2005). Behavioral and electrophysiological evidence for configural processing in fingerprint experts. Vision Research, 45, 431-448.

Carmel, D., & Bentin, S. (2002). Domain specificity versus expertise: factors influencing distinct processing of faces. Cognition, 83, 1-29.

Chao, L. L., & Knight, R. T. (1997). Prefrontal deficits in attention and inhibitory control with aging. Cerebral Cortex, 7, 63-69.

Clark, V. P., Fan, S., & Hillyard, S. A. (1994). Identification of early visually evoked potential generators by retinotopic and topographic analyses. Human Brain Mapping, 2, 170-187.

Coch, D., & Gullick, M. M. (in press). Event-related potentials and development. In S. J. Luck & E. S. Kappenman (Eds.), Oxford Handbook of Event-Related Potential Components. New York: Oxford University Press.

Dawson, G., Carver, L., Meltzoff, A. N., Panagiotides, H., McPartland, J., & Webb, S. J. (2002). Neural correlates of face and object recognition in young children with autism spectrum disorder, developmental delay, and typical development. Child Development, 73, 700-717.

Dehaene, S., Naccache, L., Le Clec'H, G., Koechlin, E., Mueller, M., Dehaene-Lambertz, G., et al. (1998). Imaging unconscious semantic priming. Nature, 395, 597-600.

Dien, J. (1998). Issues in the application of the average reference: Review, critiques, and recommendations. Behavior Research Methods, Instruments, & Computers, 30, 34-43.

Folstein, J. R., & Van Petten, C. (2008). Influence of cognitive control and mismatch on the N2 component of the ERP: A review. Psychophysiology, 45, 152-170.

Friedman, D. (in press). The components of aging. In S. J. Luck & E. S. Kappenman (Eds.), Oxford Handbook of Event-Related Potential Components. New York: Oxford University Press.

Gehring, W. J., Liu, Y., Orr, J. M., & Carp, J. (in press). The error-related negativity (ERN/Ne). In S. J. Luck & E. S. Kappenman (Eds.), Oxford Handbook of Event-Related Potential Components. New York: Oxford University Press.

Hajcak, G., Weinberg, A., MacNamara, A., & Foti, D. (in press). ERPs and the study of emotion. In S. J. Luck & E. S. Kappenman (Eds.), Oxford Handbook of Event-Related Potential Components. New York: Oxford University Press.

Handy, T. C. (Ed.). (2005). Event-Related Potentials: A Methods Handbook. Cambridge, MA: MIT Press.

Helmholtz, H. (1853). Ueber einige Gesetze der Vertheilung elektrischer Ströme in körperlichen Leitern mit Anwendung auf die thierisch-elektrischen Versuche. Annalen der Physik und Chemie, 89, 211-233, 354-377.

Hillyard, S. A., Vogel, E. K., & Luck, S. J. (1998). Sensory gain control (amplification) as a mechanism of selective attention: Electrophysiological and neuroimaging evidence. Philosophical Transactions of the Royal Society: Biological Sciences, 353, 1257-1270.

Hopf, J.-M., Luck, S. J., Boelmans, K., Schoenfeld, M. A., Boehler, N., Rieger, J., et al. (2006). The neural site of attention matches the spatial scale of perception. Journal of Neuroscience, 26, 3532-3540.

Javitt, D. C., Spencer, K. M., Thaker, G. K., Winterer, G., & Hajos, M. (2008). Neurophysiological biomarkers for drug development in schizophrenia. Nature Reviews Drug Discovery, 7, 68-83.

Kappenman, E. S., & Luck, S. J. (in press-a). ERP components: The ups and downs of brainwave recordings. In S. J. Luck & E. S. Kappenman (Eds.), Oxford Handbook of ERP Components. New York: Oxford University Press.

Kappenman, E. S., & Luck, S. J. (in press-b). The effects of electrode impedance on data quality and statistical significance in ERP recordings. Psychophysiology.

Kiesel, A., Miller, J., Jolicoeur, P., & Brisson, B. (2008). Measurement of ERP latency differences: A comparison of single-participant and jackknife-based scoring methods. Psychophysiology, 45, 250-274.

Kutas, M., McCarthy, G., & Donchin, E. (1977). Augmenting mental chronometry: The P300 as a measure of stimulus evaluation time. Science, 197, 792-795.

Lins, O. G., Picton, T. W., Berg, P., & Scherg, M. (1993). Ocular artifacts in recording EEGs and event-related potentials. II: Source dipoles and source components. Brain Topography, 6, 65-78.

Luck, S. J. (1998). Sources of dual-task interference: Evidence from human electrophysiology. Psychological Science, 9, 223-227.

Luck, S. J. (2005). An Introduction to the Event-Related Potential Technique. Cambridge, MA: MIT Press.

Luck, S. J. (in press). Electrophysiological correlates of the focusing of attention within complex visual scenes: N2pc and related ERP components. In S. J. Luck & E. S. Kappenman (Eds.), Oxford Handbook of ERP Components. New York: Oxford University Press.

Luck, S. J., & Hillyard, S. A. (1990). Electrophysiological evidence for parallel and serial processing during visual search. Perception & Psychophysics, 48, 603-617.

Luck, S. J., & Hillyard, S. A. (2000). The operation of selective attention at multiple stages of processing: Evidence from human and monkey electrophysiology. In M. S. Gazzaniga (Ed.), The New Cognitive Neurosciences. Cambridge, MA: MIT Press.

Luck, S. J., & Kappenman, E. S. (Eds.). (in press). Oxford Handbook of Event-Related Potential Components. New York: Oxford University Press.

Luck, S. J., Kappenman, E. S., Fuller, R. L., Robinson, B., Summerfelt, A., & Gold, J. M. (2009). Impaired response selection in schizophrenia: Evidence from the P3 wave and the lateralized readiness potential. Psychophysiology, 46, 776-786.

Luck, S. J., Mathalon, D. H., O'Donnell, B. F., Hämäläinen, M. S., Spencer, K. M., Javitt, D. C., et al. (submitted). A roadmap for the development and validation of ERP biomarkers. Manuscript Submitted for Publication.

Luck, S. J., Vogel, E. K., & Shapiro, K. L. (1996). Word meanings can be accessed but not reported during the attentional blink. Nature, 382, 616-618.

Näätänen, R., & Kreegipuu, K. (in press). The mismatch negativity (MMN). In S. J. Luck & E. S. Kappenman (Eds.), Oxford Handbook of Event-Related Potential Components. New York: Oxford University Press.

O'Donnell, B. F., Salisbury, D. F., Brenner, C., Niznikiewicz, M., & Vohs, J. L. (in press). Abnormalities of event-

related potential components in schizophrenia. In S. J. Luck & E. S. Kappenman (Eds.), Oxford Handbook of Event-Related Potential Components. New York: Oxford University Press.

Ochoa, C. J., & Polich, J. (2000). P300 and blink instructions. Clinical Neurophysiology, 111, 93-98.

Osman, A., & Moore, C. M. (1993). The locus of dual-task interference: Psychological refractory effects on movement-related brain potentials. Journal of Experimental Psychology: Human Perception and Performance, 19, 1292-1312.

Pascual-Marqui, R. D., Esslen, M., Kochi, K., & Lehmann, D. (2002). Functional imaging with low-resolution brain electromagnetic tomography (LORETA): a review. Methods & Findings in Experimental & Clinical Pharmacology, 24 Suppl C, 91-95.

Perez, V. B., & Vogel, E. K. (in press). What ERPs can tell us about working memory. In S. J. Luck & E. S. Kappenman (Eds.), Oxford Handbook of Event-Related Potential Components. New York: Oxford University Press.

Polich, J. (in press). Neuropsychology of P300. In S. J. Luck & E. S. Kappenman (Eds.), Oxford Handbook of Event-Related Potential Components. New York: Oxford University Press.

Pratt, H. (in press). Sensory ERP components. In S. J. Luck & E. S. Kappenman (Eds.), Oxford Handbook of Event-Related Potential Components. New York: Oxford University Press.

Rossion, B., & Jacques, C. (in press). The N170: Understanding the time course of face perception in the human brain. In S. J. Luck & E. S. Kappenman (Eds.), Oxford Handbook of Event-Related Potential Components. New York: Oxford University Press.

Scherg, M. (1990). Fundamentals of dipole source potential analysis. In F. Grandori, M. Hoke & G. L. Romani (Eds.), Auditory Evoked Magnetic Fields and Potentials. Advances in Audiology VI (pp. 40-69). Basel: Karger.

Smulders, F. T. Y., & Miller, J. O. (in press). Lateralized Readiness Potential. In S. J. Luck & E. S. Kappenman (Eds.), Oxford Handbook of Event-Related Potential Components. New York: Oxford University Press.

Sreenivasan, K. K., Goldstein, J. M., Lustig, A. G., Rivas, L. R., & Jha, A. P. (2009). Attention to faces modulates early face processing during low but not high face discriminability. Attention, Perception, & Psychophysics, 71, 837-846.

Sutton, S., Braren, M., Zubin, J., & John, E. R. (1965). Evoked potential correlates of stimulus uncertainty. Science, 150, 1187-1188.

Swaab, T. Y., Ledoux, K., Camblin, C. C., & Boudewyn, M. (in press). Language-related ERP components. In S. J. Luck & E. S. Kappenman (Eds.), Oxford Handbook of Event-Related Potential Components. New York: Oxford University Press.

Tanaka, J. W., & Curran, T. (2001). A neural basis for expert object recognition. Psychological Science, 12, 43-47.

van Turennout, M., Hagoort, P., & Brown, C. M. (1998). Brain activity during speaking: From syntax to phonology in 40 milliseconds. Science, 280, 572-574.

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

Vogel, E. K., & Luck, S. J. (2000). The visual N1 component as an index of a discrimination process. Psychophysiology, 37, 190-123.

Yeung, N., Cohen, J. D., & Botvinick, M. M. (2004). The neural basis of error detection: conflict monitoring and the error-related negativity. Psychological Review, 111, 931-959.

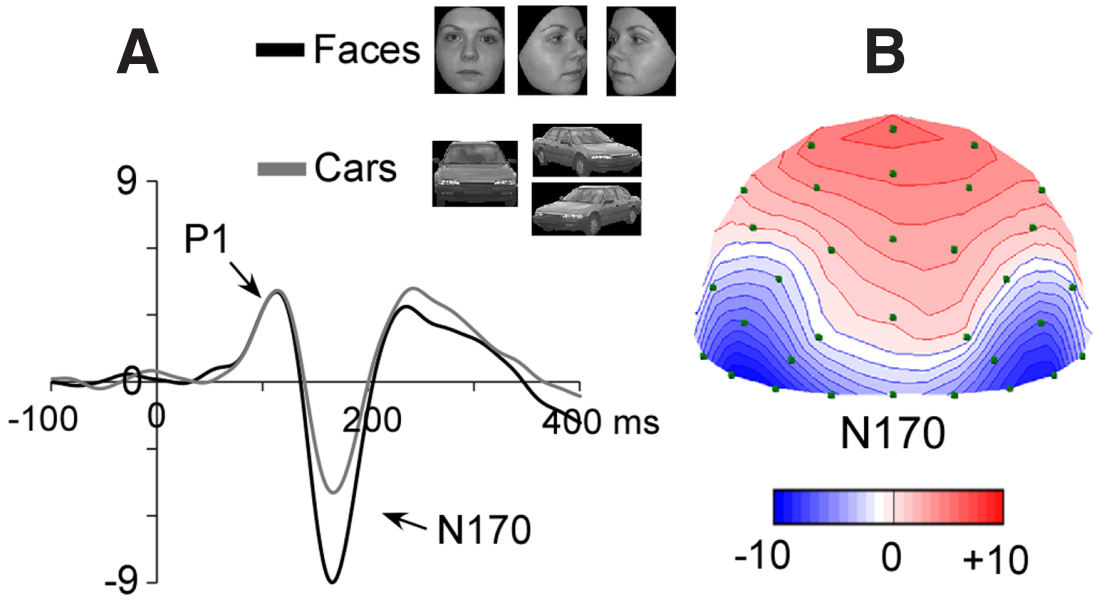

Figure 1. Example N170 experiment, including ERP waveforms from an occipito-temporal electrode site (A) and the scalp distribution of the voltage in the N170 latency range (B). Adapted with permission from Rossion and Jacques (in press). Copyright 2010 by B. Rossion; all rights reserved.

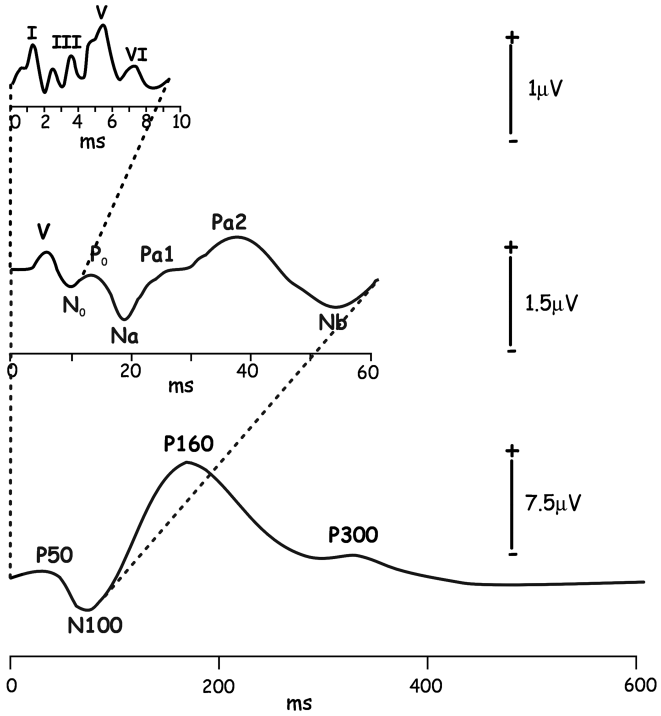

Figure 2. Typical sequence of auditory sensory components. The waveform elicited by a click stimulus is shown over different time ranges with different filter settings to highlight the auditory brainstem responses (top), the midlatency responses (middle), and the long-latency responses (bottom). Adapted with permission from Pratt (in press). Copyright 2010 by H. Pratt; all rights reserved.

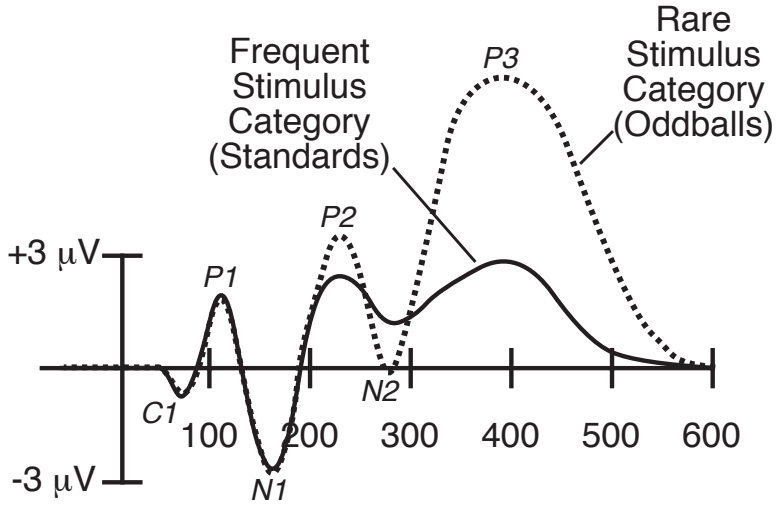

Figure 3. Typical ERP waveforms elicited by standards and oddballs at a posterior electrode site in a visual oddball paradigm.

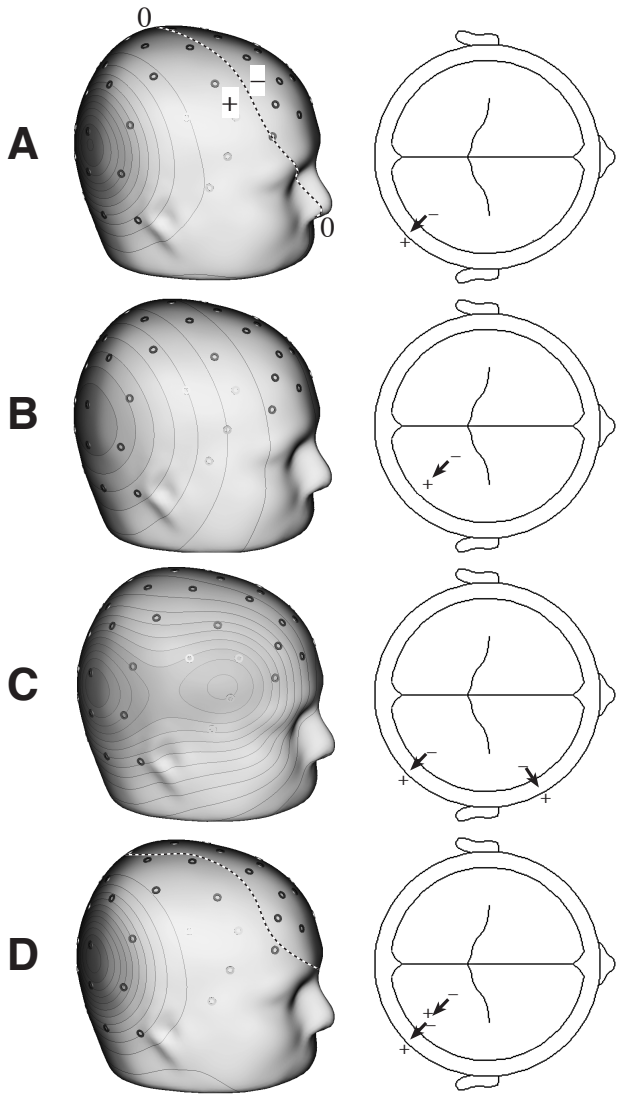

Figure 4. Scalp distributions (left) produced by different dipole configurations (right). Courtesy of Jesse Bengson.

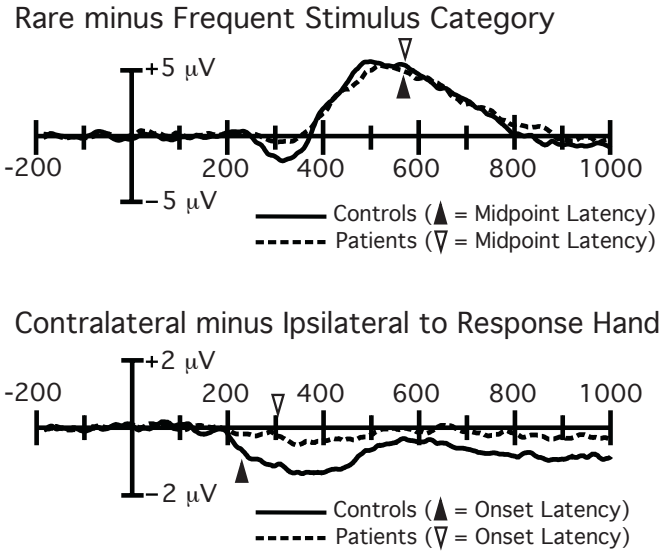

Figure 5. Data from the study of Luck et al. (2009). The P3 was isolated by constructing rare-minus-frequent difference waves at the Pz electrode site (top), and the lateralized readiness potential was isolated by constructing contralateral-minus-ipsilateral difference waves at the C3 and C4 electrode sites (bottom). Triangles show mean latency values.

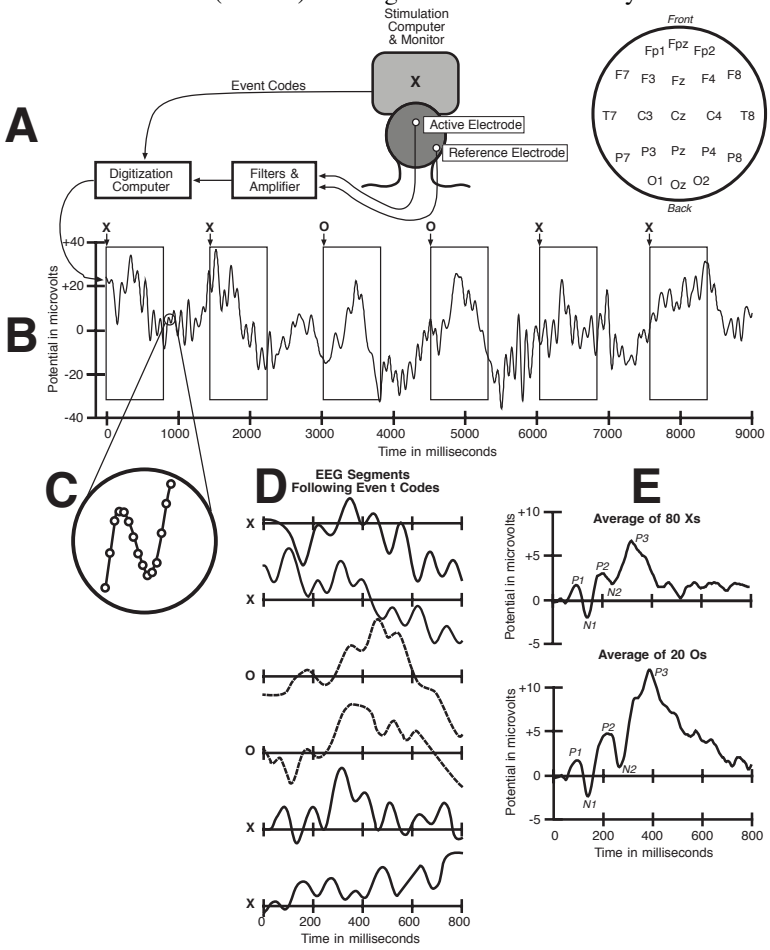

Figure 6. Illustration of the procedures used to measure the EEG and construct averaged ERP waveforms in a typical visual oddball paradigm.

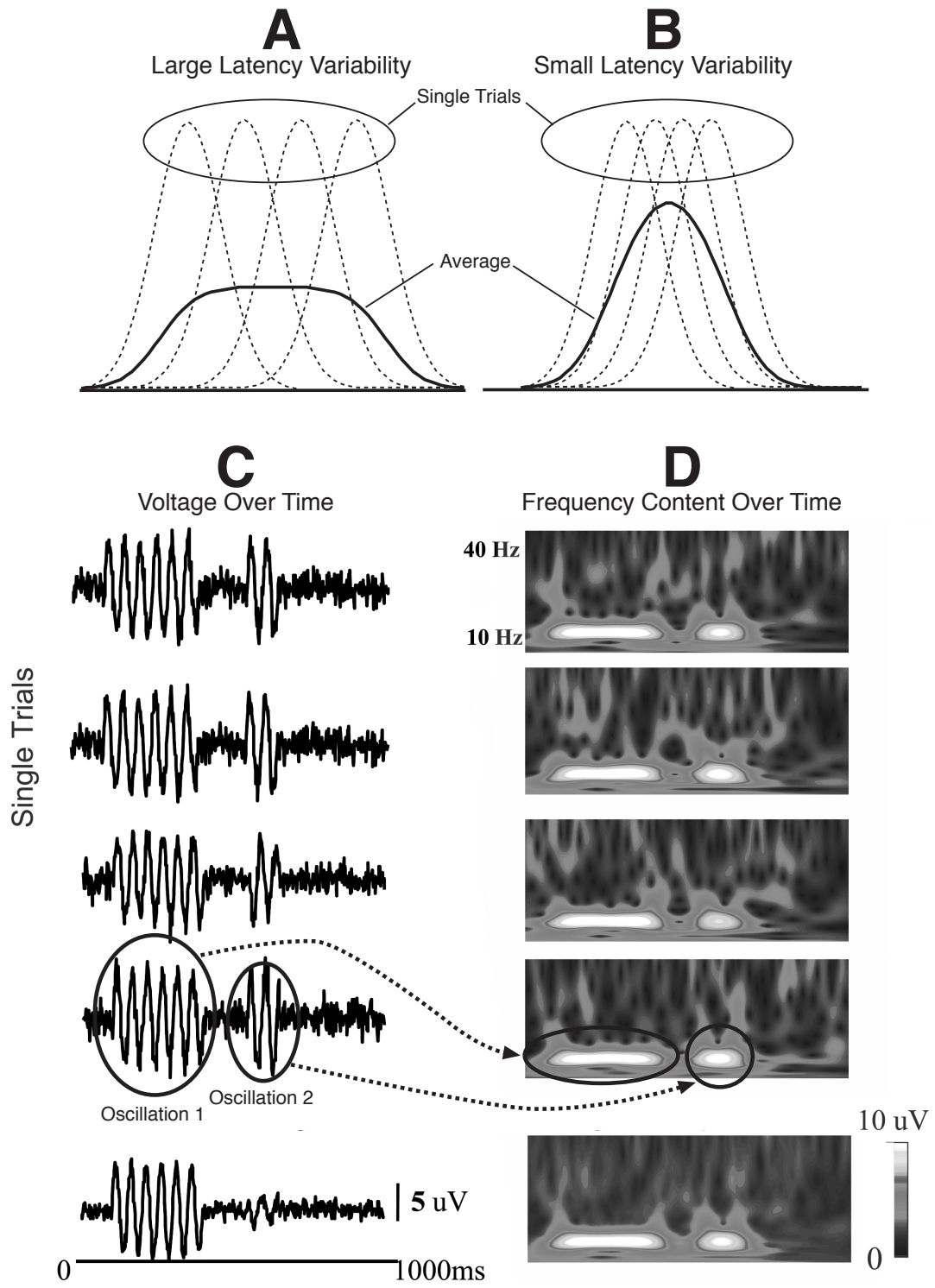

Figure 7. Illustration of the effects of latency and phase variability on averaged ERPs. With a monophasic component, like the P3 wave, a large amount of latency variability in the individual trials leads to a broad averaged waveform with a reduced peak amplitude (A), and reduced latency variability leads to a narrower averaged waveform with a larger peak amplitude (B). When an oscillation is elicited by the stimulus, it will remain in the average if the phase is constant from trial to trial but will virtually disappear from the average if the phase varies randomly (C). The problem of phase variability can be addressed by first converting each single trial into the frequency domain and then averaging across trials (D). The second oscillation remains in this time-frequency average, even though it was largely lost in the conventional average. Panels C and D adapted with permission from Bastiaansen et al. (in press).