

# **6** Artifact Rejection and Correction

## **Overview**

The signal that you record in an ERP experiment will consist of the EEG plus a variety of non-neural noise sources. These non-neural sources mainly consist of induced electrical signals from the recording environment (e.g., line noise from lights and computers) and biological signals such as blinks, eye movements, muscle activity, and skin potentials. All of these non-EEG signals are considered artifacts. Some of them are small and constant (e.g., line noise and some types of muscle activity), but others are large and transient (e.g., blinks and eye movements). The large and transient artifacts are often eliminated by discarding the trials on which they occur. The smaller and more constant artifacts cannot be eliminated in this way because they are typically present on every trial. However, they can often be attenuated by data processing procedures (e.g., averaging, filtering).

Artifacts can be problematic in three ways. First, they may decrease the signal-to-noise ratio (SNR) of the averaged ERP waveform, decreasing your ability to find significant differences between groups or conditions. Second, some types of artifacts may be systematic rather than random, occurring in some conditions more than others and being at least loosely time-locked to the stimulus so that they are not eliminated by the averaging process. Such artifacts may lead to erroneous conclusions about the effects of an experimental manipulation. For example, some stimuli may be more likely to elicit blinks than others, which could lead to differences in amplitude in the averaged ERP waveforms. Thus, a difference between conditions might be a result of a difference in a motor behavior (blinking) rather than reflecting a difference in brain activity. Third, the most common artifacts in ERP experiments are ocular artifacts related to blinks and eye movements, both of which change the sensory input. If these artifacts differ across conditions, then the sensory input may also differ across conditions, which can be an important confound. This type of confound is often not obvious to the experimenter, and it's on my list of Top Ten Reasons for Rejecting an ERP Manuscript (see online chapter 15).

There are two main classes of techniques for minimizing the effects of artifacts. First, it is possible to detect large artifacts in the EEG and simply exclude contaminated trials from the averaged ERP waveforms (this is called *artifact rejection*). Alternatively, it is sometimes possible

to estimate the influence of the artifacts on the ERPs and use correction procedures to subtract away the estimated contribution of the artifacts (this is called *artifact correction*). In this chapter, I will discuss both approaches, including their advantages and disadvantages. However, I would first like to make a point that should be obvious but is often overlooked. Specifically, it is always better to minimize the occurrence of artifacts rather than to rely heavily on rejection or correction procedures, which always have a cost. This is really just a special case of Hansen's axiom: There is no substitute for good data. In other words, time spent eliminating artifacts at the source will be well rewarded by time saved later. This chapter will therefore also include hints for reducing the occurrence of artifacts.

This chapter provides very detailed and concrete recommendations about artifact rejection. In the first edition of this book, I was frustrated by the fact that most of my recommendations could not be followed in commercial ERP analysis packages, which typically do not provide very sophisticated tools for artifact rejection. This is one of the factors that motivated my lab to create ERPLAB Toolbox, a freely available ERP data analysis package (available at <http://erpinfo.org/erplab>). All of my recommendations about artifact rejection can easily be accomplished with ERPLAB Toolbox.

The description of artifact correction in this chapter is more theoretical. This is partly because there are many different approaches to artifact correction, and it's not practical to discuss all of them. It is also partly because my preferred approach for artifact correction—which uses independent component analysis (ICA)—is complex and rapidly evolving. However, a detailed description of how ICA-based artifact correction works in general, along with an example of how my lab implements it, can be found in the online supplement to chapter 6.

### **The General Artifact Rejection Process**

Artifact rejection has been used since the earliest ERP studies to minimize the effects of artifacts. It has an obvious drawback: If you throw out some proportion of trials, you end up with fewer trials in your averaged ERPs, which can reduce your statistical power. And if the number of rejected trials differs across conditions or groups, this can make the data noisier in some conditions or groups than in others (see the online supplement to chapter 9 for a discussion of the consequences of differences in the number of trials). In many cases, however, it is a simple procedure that works very well, and the benefits greatly outweigh the drawbacks.

Before I get into the details of how to detect specific types of artifacts, I would like to provide a general framework for conceptualizing the artifact rejection process.<sup>1</sup> Detecting artifacts is, in essence, a *signal detection* problem, in which the artifact is treated as the to-be-detected signal. If you don't know about signal detection theory, I would recommend finding a perception textbook and reading about it. It's a fundamental framework for understanding how the mind works, and it is very useful for understanding artifact rejection.

Here I will provide a simple example of signal detection. Imagine that you have lost a valuable ring on a beach, and you have rented a metal detector to help you find it. The metal detector

makes a sound that tells you the extent to which there is evidence of nearby metal, getting louder when there is a lot of mineral content and getting softer when there is not much mineral content. However, this output is quite variable due to random fluctuations in the mineral content of the sand. If you started digging in the sand any time there was a hint of nearby metal, you would make very slow progress because you would start digging every few steps. However, if you only started digging when the metal detector's output was very high, you might miss the ring altogether because it's small and doesn't create a large change in the detector's output. You therefore want to start digging at some intermediate level.

The key aspects of this example are as follows. You are trying to detect something that is either there or not (the ring) based on a noisy, continuously variable signal (the metal detector's output). You select a threshold value, and if the signal exceeds that value, you make a response (digging). In this context, we can define four outcomes for each patch of sand: (1) a *hit* occurs when the sought-after object is present, the signal exceeds the threshold, and you respond (i.e., the ring is under the metal detector, the metal detector's output exceeds a certain value, and you dig); (2) a *miss* occurs when the object is present but the signal fails to exceed the threshold, and you don't respond; (3) a *false alarm* occurs when the object is absent, the signal exceeds the threshold due to random variation, and you respond; (4) a *correct rejection* occurs when the object is absent, the signal doesn't exceed the threshold, and you don't respond. Hits and correct rejections are both correct responses, and misses and false alarms are both errors. Importantly, you can increase the number of hits by choosing a lower threshold (i.e., digging when the metal detector's output is fairly low), but this will also lead to an increase in the number of false alarms. The only way to increase the hit rate without increasing the false alarm rate is to get a better metal detector with an output that better differentiates between the presence or absence of small metal objects and is less influenced by other minerals in the sand.

Now imagine that you are trying to detect blinks in a noisy EEG signal rather than a ring on the beach. When a subject blinks, the movement of the eyelids across the eyeball creates a voltage deflection, and it is possible to assess the presence or absence of a blink by measuring the size of the largest voltage deflection within a given segment of EEG (just like assessing the presence or absence of the ring by examining the output of the metal detector). If the voltage deflection exceeds a certain threshold level, you conclude that the subject blinked and you discard that trial; if the threshold is not exceeded, you conclude that the subject did not blink and you include that trial in the averaged ERP waveform. If you set a low threshold and reject any trials that have even a small voltage deflection, you will eliminate all of the trials with blinks, but you will also discard many blink-free trials, reducing the signal-to-noise ratio of the averaged ERP waveform. If you set a high threshold and reject only trials with very large voltage deflections, you will have more trials in your averages, but some of those trials may contain blinks that failed to exceed your threshold. Thus, simply changing the threshold cannot increase the rejection of true artifacts without also increasing the rejection of artifact-free trials. However, just as you can do a better job of finding a ring in the sand by using a better metal detector, you can do a better job of rejecting artifacts by using a better procedure for measuring artifacts (i.e., a procedure

that does a better job of having a high value when the artifact is present and a low value when it is absent).

When you are doing artifact rejection, it is important to be clear about your goal, which is ultimately to learn the true answer to some scientific question. To get the truth from your data, you need to maximize your statistical power (so that you have the best possible chance of getting a significant  $p$  value if a real effect is present), and you need to avoid confounds (so that a valid conclusion can be drawn from a significant effect). The goal is not to eliminate all artifacts from your data. You should typically eliminate artifacts only when they reduce your statistical power or create confounds. If you discard every trial that contains a hint of an artifact, you will end up throwing out virtually every trial in the experiment (because there is always at least a hint of muscle activity, induced noise, eye movement, etc.). However, if an artifact differs systematically across conditions or groups, it may create a significant but bogus difference between your conditions or groups. Thus, your goal in artifact detection is to discard trials that contain “problematic” artifacts (i.e., artifacts that reduce your statistical power or create confounds) and retain “good” trials (i.e., trials that improve your statistical power without creating confounds). This is not terribly difficult if you spend some time learning about the artifacts that are present in your data, and you think carefully about how the artifacts might influence your data.

In ERPLAB Toolbox, we make a distinction between artifact *detection* and artifact *rejection*. We use the term *detection* to refer to the process of determining that an artifact is present. That is, when the algorithm determines that an EEG epoch exceeds the threshold for rejection, that epoch is given a special mark. At the time of averaging, the marked epochs are excluded from the averages (unless the user specifies otherwise). The artifact-containing epochs are not permanently discarded, and this makes it easy to change the artifact rejection parameters and try again. However, ERPLAB also includes tools for artifact *rejection*, in which periods of data containing artifacts are deleted from the data file. Most people use the term *rejection* to mean the combination of detecting artifacts and excluding them from the average, and this is how I will use this term. Keep in mind, however, that the artifacts are not really being deleted from the EEG file in most cases; they are just being excluded during the ERP averaging process.

In most systems, all channels are discarded for a given trial if an artifact is detected in any channel. This makes sense because an artifact that is present in multiple channels might be harder to detect in some than in others. For example, if an eyeblink artifact is 100  $\mu$ V right above the eyes at the Fp1 and Fp2 electrodes, it would be approximately 36  $\mu$ V at the Fz electrode, approximately 16  $\mu$ V at the Cz electrode, and approximately 10  $\mu$ V at the Pz electrode (Lins, Picton, Berg, & Scherg, 1993a). This artifact would not be big enough to be reliably rejected at the Cz and Pz electrode sites, but it would certainly be big enough to impact a 2- $\mu$ V experimental effect at these sites. Consequently, researchers typically reject all channels for trials with eyeblinks, even though the blinks are detected only at a subset of channels on individual trials. If you excluded trials from the Cz and Pz electrodes only when blinks were detected at those electrodes, you would fail to reject many blinks that are large enough to significantly distort your data. Moreover, if you reject trials only in the electrodes in which the artifact was detected,

**Box 6.1**

## Electrode-Specific Rejection or Interpolation

Some software systems allow you to reject only the channel in which an artifact was detected or to replace the waveform in that channel with interpolated values from the surrounding electrode sites. This approach is particularly popular among people who are recording from very large numbers of electrodes. I have heard people say that they can't throw out all the channels for a given trial when an artifact is detected because they would end up throwing out almost every trial.

In my view, something is deeply wrong with this. If your data quality is so low that at least one channel has an artifact on a high proportion of trials, your data quality is presumably too low to do anything that would require a large number of electrodes (e.g., provide very precise scalp distributions). My recommendation in this situation would be to buy a better EEG recording system or to spend more time making sure that you are getting a good recording from every channel.

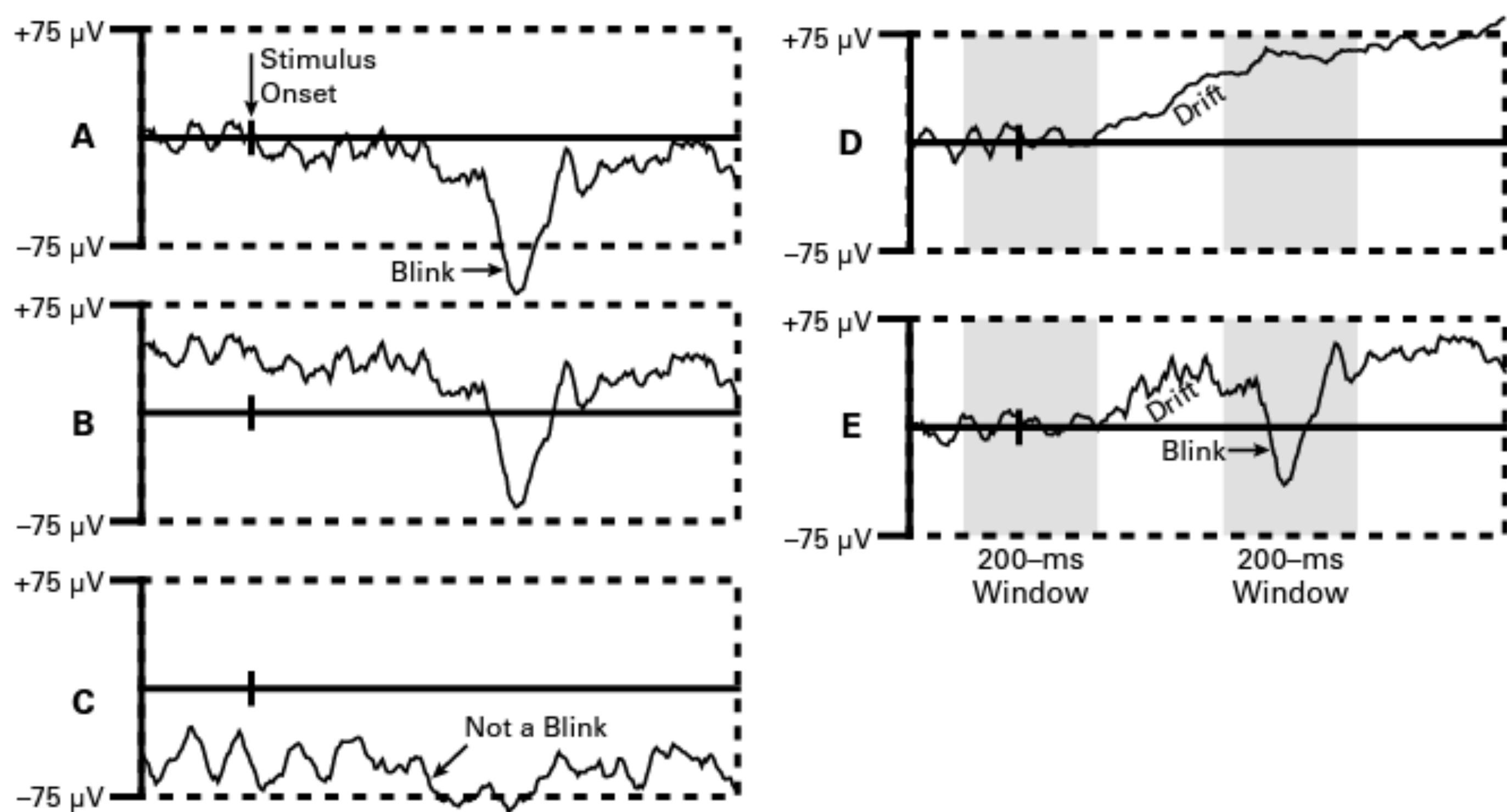
If you have one channel that is bad during the entire recording (e.g., because the electrode is broken), it should be okay to use an interpolated value for that channel. However, this should be a rare event: In most recordings, you should be able to make sure that all channels are delivering high-quality data before you start the task.

your averages will be based on different trials at different electrodes, making it difficult to trust the scalp distributions. Moreover, you would end up with different numbers of trials in your averages at different electrode sites, which might make it problematic to compare amplitudes across electrode sites (see the online supplement to chapter 9). See box 6.1 for some further thoughts.

**Choosing an Artifact Measure**

Many ERP analysis systems provide only a single crude method for rejecting trials with artifacts, in which a trial is rejected if the voltage during the epoch exceeds a user-defined threshold, such as  $\pm 75 \mu\text{V}$ . In essence, this approach uses the absolute value of the maximum voltage in the epoch as the measure of the artifact, rejecting the trial if this voltage exceeds the threshold. I am always shocked that a company will sell an ERP analysis system for thousands of dollars (or thousands of euros, tens of thousands of yuan, hundreds of thousands of yen, etc.) with such primitive artifact detection abilities. This method works okay for blink rejection under some conditions because blinks are very large, but it is totally inadequate for detecting and rejecting more subtle artifacts, such as eye movements. This is one of the reasons that we developed ERPLAB Toolbox, which contains excellent methods for assessing specific types of artifacts. I should note that some commercial packages also have good methods for assessing artifacts, and this should be a factor if you are deciding which package to buy.

If your software forces you to use an absolute voltage rejection approach, it is essential to perform baseline correction prior to artifact rejection. As will be discussed in chapter 8, baseline correction typically involves subtracting the average prestimulus voltage from the entire waveform. Panels A–C of figure 6.1 show why this is necessary if you are rejecting artifacts on the

**Figure 6.1**

(A) Single-trial vertical EOG waveform containing a blink, with baseline correction to eliminate any overall offset in the signal. The dashed lines reflect the voltage threshold for artifact rejection ( $\pm 75 \mu\text{V}$ ). The blink (the large negative-going voltage deflection) brings the signal beyond the threshold, so the trial would be correctly rejected. (B) Same as (A), but without baseline correction. Now the blink does not exceed the threshold for rejection, and the trial would not be rejected (incorrectly). (C) EOG waveform on a trial without a blink, with no baseline correction. The overall offset of the signal causes the trial to be rejected (incorrectly) even though no blink is present. (D) Baseline-corrected EOG waveform on a trial with some voltage drift, which exceeds the threshold for rejection and is therefore rejected (incorrectly). The gray regions indicate two of the many time windows that would be tested with the moving window peak-to-peak amplitude method. The peak-to-peak amplitude is small within each window, so the trial would not be rejected with this method (correctly). (E) Baseline-corrected EOG waveform on a trial with some voltage drift and a blink. The drift partly counteracts the blink, so the trial would not be rejected if a simple voltage criterion were used (incorrectly). The gray regions indicate two of the many time windows that would be tested with the moving window peak-to-peak amplitude method. The peak-to-peak amplitude is large within the window that contains the blink, so the trial could be rejected with this method (correctly).

basis of the absolute voltage. Panel A shows a typical blink that is large enough to exceed the  $75\text{-}\mu\text{V}$  threshold for rejection. Baseline correction was applied to this waveform, so the voltage is centered on  $0\text{ }\mu\text{V}$  prior to the onset of the blink. Panel B shows the same waveform without baseline correction. A voltage offset is present over the entire epoch that, by chance, moves the waveform up enough that this blink no longer exceeds the threshold for rejection. Panel C shows a different epoch of data in which baseline correction was not performed. No blink was present, but a voltage offset brought the voltage close enough to the rejection threshold that ordinary voltage fluctuations caused the waveform to exceed the threshold for rejection. Thus, baseline correction is absolutely essential if you will be using this type of rejection algorithm.

Even if you perform baseline correction, drifts in the data may lead to poor detection of eye-blanks. Figure 6.1D shows a baseline-corrected example in which no blink is present but a drift in the data leads the voltage to exceed the threshold for rejection. Figure 6.1E shows a baseline-corrected example in which a blink is present, but an opposite-polarity drift causes it to stay within the voltage threshold and avoid rejection. The former problem could be addressed by choosing a higher rejection threshold, and the latter problem could be addressed by choosing a lower rejection threshold. Although these solutions would work for these specific trials, increasing the threshold would lead to more misses, and decreasing the threshold would lead to more false alarms. A much better solution would be to use a more sensitive blink detection algorithm.

When Javier Lopez-Calderon was developing the artifact detection tools in ERPLAB Toolbox, he came up with a simple but very effective algorithm for more reliably distinguishing between epochs with blinks and epochs without blinks. It's called the *moving window peak-to-peak amplitude* method.<sup>2</sup> With this method, you define a window width, such as 200 ms (see the gray regions in panels D and E of figure 6.1). The algorithm places this window at the beginning of the epoch and finds the peak-to-peak amplitude within this window (i.e., the difference in amplitude between the most positive and negative points in the window). The window is then shifted rightward by a user-defined amount (e.g., 50 ms), and the peak-to-peak amplitude is determined in this new window. This continues until the whole epoch has been tested (or whatever portion of the epoch the user wants to test). The largest of these peak-to-peak amplitudes is then compared with the threshold for rejection.

Because this method uses the difference between the highest and lowest points in each small window, it is completely insensitive to the overall voltage offset of the epoch. It is also relatively insensitive to slow voltage drifts. Consider, for example, the blink-free data in figure 6.1D. Because there are no blinks, there are no sudden changes in voltage, and the peak-to-peak voltage within any given 200-ms window is small. When a blink is present, however, the peak-to-peak voltage within the 200-ms period containing the blink will be large, even if an opposite-polarity drift is present (as in figure 6.1E). I have found this to be a much more sensitive measure of blinks than the absolute voltage method. With an appropriate threshold, it can detect all true blinks while rarely having false alarms. Another good algorithm will be described later in the chapter.

### Choosing a Rejection Threshold

Once you have chosen an appropriate measure of an artifact, you must choose a threshold that does a good job of balancing misses and false alarms. One option is to pick a threshold on the basis of experience or previous studies and use this value for all subjects. For example, you may decide that all trials with a peak-to-peak EOG amplitude of 50  $\mu$ V or higher will be rejected. However, there is often significant variability across subjects in the size and shape of the voltage deflections produced by a given type of artifact and in the characteristics of the EEG in which these voltage deflections are embedded, and a one-size-fits-all approach is therefore not optimal. Instead, it is usually best to tailor the threshold for each individual subject.

Depending on your data analysis system, you may have control over other parameters as well, such as the time window that is scanned for artifacts. Some systems force you to look for artifacts over the entire epoch, but others let you choose a window within the epoch. You will ordinarily want to reject artifacts at any time in the window, but remember that the goal is not to remove all artifacts but instead to maximize statistical power and avoid confounds. In my lab's N2pc experiments, for example, subjects often make small eye movements in the direction of the target, which leads to a lateralized EOG voltage that can distort the N2pc measurements. These eye movements are fairly small compared to the single-trial noise, but they are still large enough to distort the results in the averaged ERP waveforms. In these cases, we are concerned about eye movements that are consistently directed toward the side of the target stimulus and not with random eye movements. The consistent eye movements never occur prior to 150 ms, and they rarely occur after 500 ms. Thus, it's reasonable to reject small eye movements only between 150 and 500 ms. This allows us to use a low rejection threshold without discarding too many trials. In some cases, we still end up either rejecting too many trials or not catching all of the eye movements, but in those cases we take a close look at the individual subjects and tailor the rejection window for each subject. For example, one subject may make consistent eye movements only between 200 and 300 ms poststimulus and another might make them only between 250 and 400 ms, and we would use different rejection windows for these two subjects.

You may be thinking that this kind of experimenter control over the data processing has the potential to bias the results of an experiment. This is not actually a problem if your experiments exclusively involve within-subject manipulations, because the same parameters will be used in every condition for a given subject.<sup>3</sup> If your experiments involve different groups of subjects, however, the artifact rejection parameters should be set by someone who is blind to the group membership of each subject. This is what we do in our schizophrenia research. It can be inconvenient, because it's difficult to look at the results before the experiment is complete if the person doing the analyses is blind to group membership, but I believe it's still the best approach.

If you set the parameters individually for each subject, the best way to do this is usually by means of visual inspection of the raw EEG. You can do this with the following sequence of steps. First, use your prior experience (or published values) to select an initial threshold. You will then have your software package perform a preliminary artifact detection with these parameters. This allows you to determine visually whether trials with problematic artifacts are not being rejected or if trials without problematic artifacts are being rejected. Of course, this requires that you are able to determine the presence or absence of artifacts by visual inspection. In most cases, this is fairly straightforward, and some hints on how to do this are provided in the next section. After you have tested the effects of this initial threshold, you can adjust the threshold (and/or other parameters) and try the artifact rejection again. You can repeat this until you are able to reject all of the trials that clearly have artifacts without rejecting too many artifact-free trials (as assessed visually). Some types of artifacts also leave a distinctive "signature" in the averaged waveforms (as will be discussed below), so it is also possible to evaluate

whether the threshold adequately rejected trials with artifacts after the data have been averaged.

It can also be useful to ask the subject to make some blinks and eye movements at the beginning of the session so that you can easily see what that subject's artifacts look like. Keep in mind, however, that voluntarily produced blinks are usually larger than spontaneous blinks.

When you are first learning to do artifact rejection, you may spend an hour with the data from each subject, looking through the entire session and trying lots of different parameters. After you think you have the right parameters, you should ask someone with experience to take a look and make sure you've done it right. This will help you to train your visual system to detect artifacts. Once you have more experience, you will typically spend only 5–10 min per subject, looking at a few parts of the session to make sure that the parameters are appropriate for the whole session. Moreover, you will be able to use the same parameters for most subjects in most experiments.

### **Visual Inspection**

Some investigators visually inspect the EEG on each trial to determine which trials contain artifacts, but this process is conceptually identical to the procedure that I just outlined. The only difference is that the experimenter's visual system is used instead of a computer algorithm to determine the extent to which an artifact appears to be present, and a subjective, informal, unspecified, internal threshold is used to determine which trials to reject. The advantage of this approach is that the human visual system can be trained to do an excellent job of differentiating between real artifacts and normal EEG noise. However, a well-designed computer algorithm may be just as sensitive. In addition, computer algorithms have the advantages of being fast, consistent, and bias-free.

In my experience, the use of manual artifact rejection is often a consequence of commercial data analysis packages that have only primitive automated tools for rejection. If I were using one of those packages, I would also end up rejecting artifacts by visual inspection. In almost all cases, I would recommend using a good automated artifact rejection system rather than spending hours trying to identify artifacts by eye. Your visual system will still be used in setting the rejection parameters, but the computer will do most of the tedious work of checking every single trial for artifacts. If you are dealing with difficult-to-obtain data sets with very small numbers of trials (e.g., data from infants or hard-to-find patients who can only tolerate short recording sessions), you may want to do the automated rejection and then visually verify each trial. If you end up using visual inspection, then it is absolutely crucial that the person doing the rejection is blind to groups and conditions.

### **Minimizing and Detecting Specific Types of Artifacts**

In this section, I will discuss several common types of artifacts and provide suggestions for reducing their occurrence and for detecting them when they do occur.

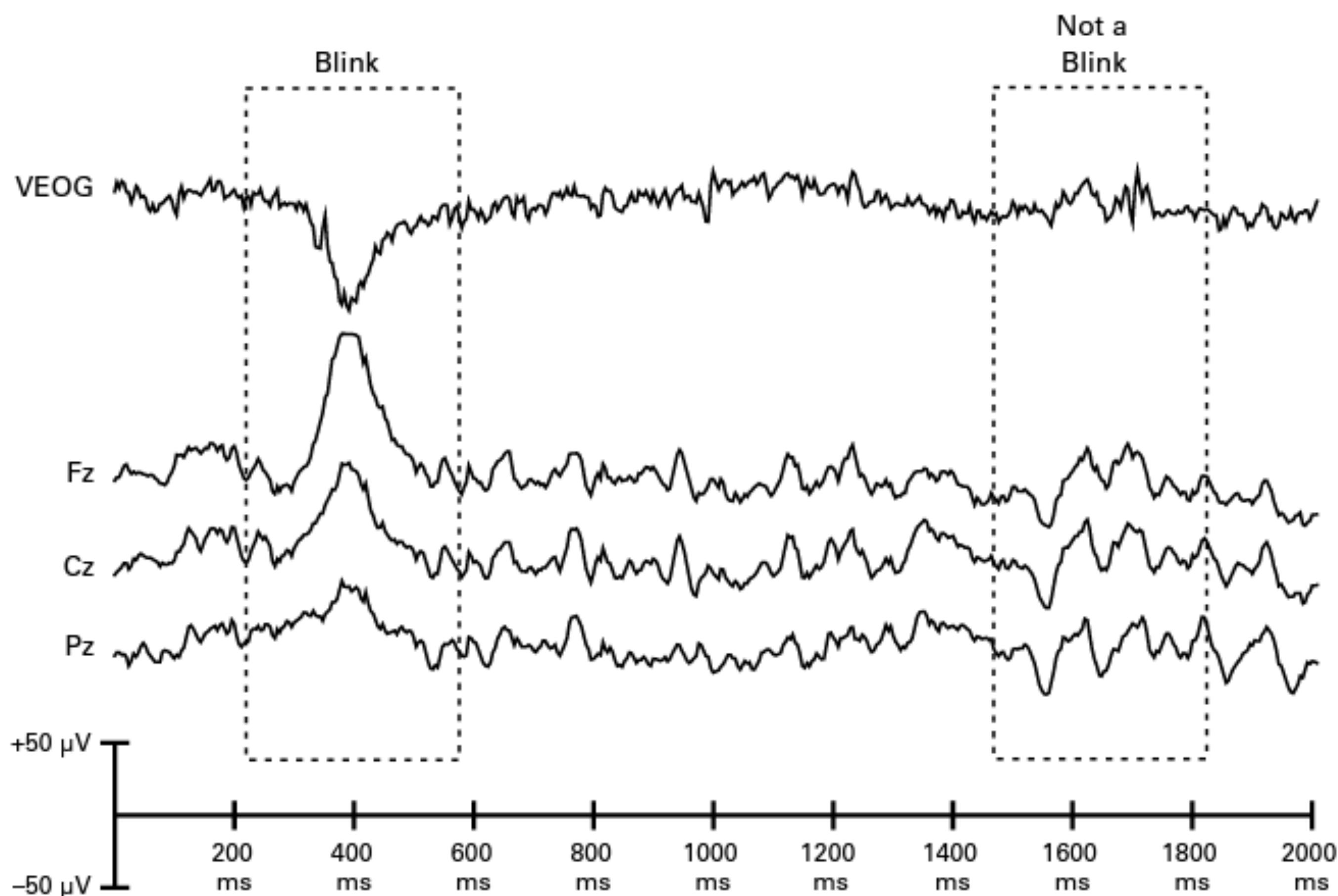
## Blinks

**Understanding Blinks** Within each eye, there is a large, constant electrical potential between the cornea at the front of the eye and the retina at the back of the eye (the *corneal-retinal potential*), which is like a dipole with positive at the front of the eye and negative at the back of the eye. This potential spreads to the surrounding parts of the head, falling off gradually toward the back of the head. This is one source of the voltage offsets that are seen in EEG and EOG recordings (as in figure 6.1B). Ordinarily, this offset is subtracted by the baseline correction procedure, making it effectively invisible. However, anything that changes this potential rapidly will cause a large voltage deflection across much of the scalp. The voltage recorded on the scalp from this dipole is called the *electrooculogram* (EOG). This should not be confused with the *electroretinogram* (ERG), which is the much smaller neural response of the retina to a visual stimulus.

When the eyes blink, the eyelid moves across the eyes, which acts as a variable resistor that changes the EOG voltage recorded from electrodes near the eyes. Other factors (e.g., ocular rotation) also contribute to this voltage, but the movement of the eyelid is the main contributor (for an excellent review, see Plochl, Ossandon, & Konig, 2012). Figure 6.2 shows the typical waveshape of the eyeblink response at a location below the eyes (labeled VEOG) and at several locations on the scalp (all are referenced to a mastoid electrode). The eyeblink response consists primarily of a monophasic deflection of 50–100 µV with a typical duration of 200–400 ms. Perhaps the most important characteristic of the eyeblink response is that it is opposite in polarity for sites above versus below the eye (compare, for example, the VEOG and Fz recordings in figure 6.2). This makes it possible to distinguish between a blink, which would produce opposite-polarity voltage shifts above versus below the eye, and a true EEG deflection, which would typically produce same-polarity voltage shifts above and below the eye. An example of a true EEG deflection is shown on the right side of figure 6.2, where same-polarity deflections can be seen at the VEOG and Fz sites.

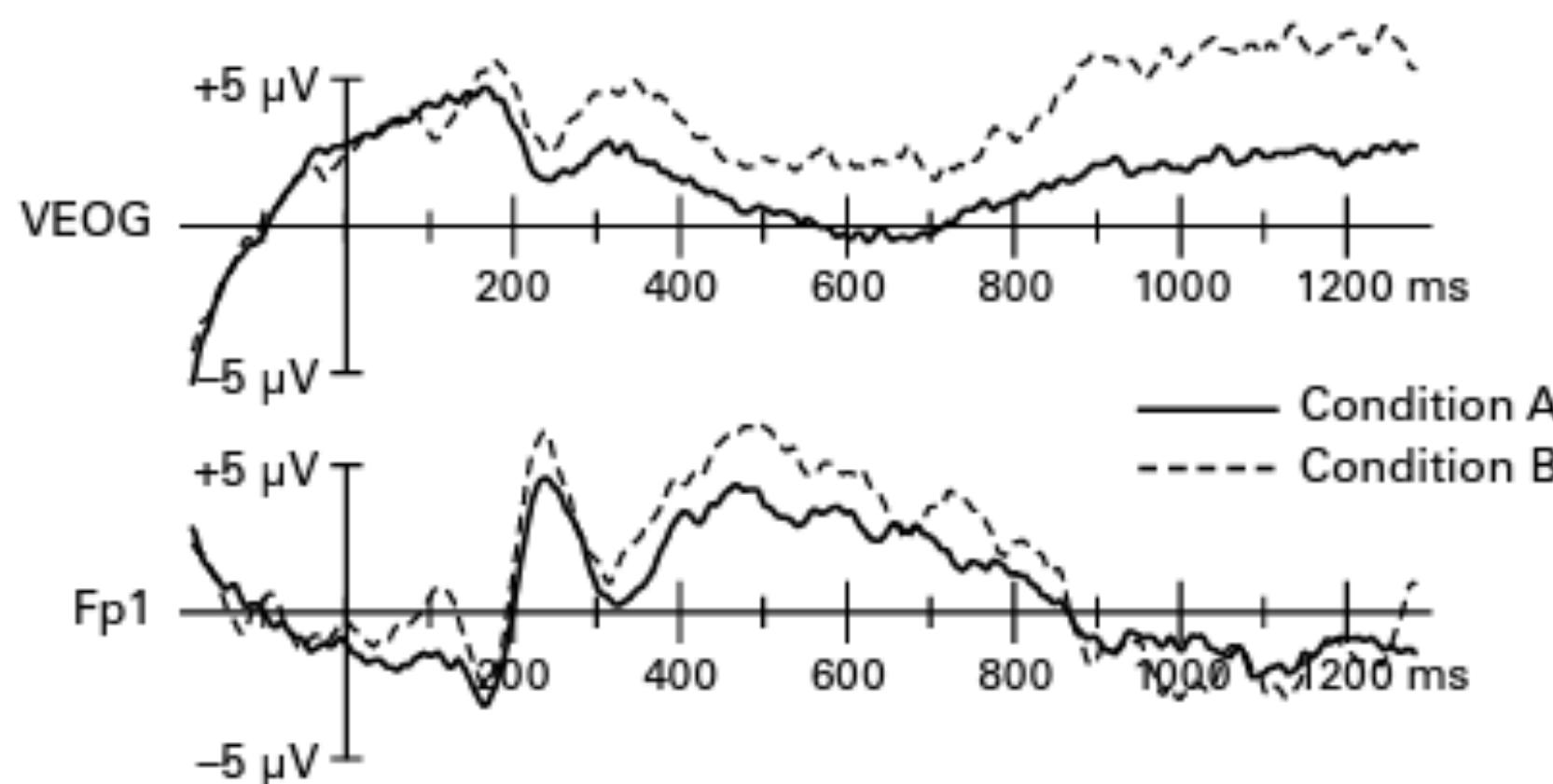
**Reducing the Occurrence of Blinks** Reducing the occurrence of an artifact is always better than rejecting trials with artifacts, and there are several ways in which the number of blinks can be reduced. The first is to ask subjects who normally wear contact lenses—which cause a great deal of blinking—to wear their glasses instead of their contact lenses. These individuals tend to blink more than average even when wearing their glasses, and it is therefore useful to keep a supply of eye drops handy (although these should be used only by individuals who normally use eye drops, and single-use bottles should be used to avoid infection risks). It is also helpful to use short trial blocks of 1–2 min, thus providing the subjects with frequent rest breaks for blinking (this also helps to keep the subjects more alert and focused on the task).

If you see a lot of blinks (or eye movements), it's important to let the subject know. Don't be shy about telling subjects that they need to do a better job of controlling these artifacts. My students tell me that it took them a long time to become comfortable doing this, but you really need to do it, even if it makes you uncomfortable at first.

**Figure 6.2**

Recordings from a vertical EOG (VEOG) electrode located under the left eye and EEG electrodes located at Fz, Cz, and Pz, with a right mastoid reference for all recordings. A blink can be seen at approximately 400 ms, creating a negative deflection at the VEOG electrode and a positive deflection at the scalp electrodes. Note that the deflection is quite large at Fz and then becomes smaller at Cz and even smaller at Pz. The area labeled “Not a Blink” contains moderately large voltage deflections in all of these channels, but these deflections do not reflect a blink because the polarity is not inverted at the under-the-eye VEOG electrode relative to the scalp electrodes.

Some experiments can be designed so that subjects have a well-defined period during the intertrial interval (ITI) in which blinks are allowed. For example, the fixation point might disappear during the ITI, and the subject would be told that blinks are allowed when the fixation point is not present. This can work very well, but you need to be careful about two potential problems. First, if the fixation point appears shortly before the stimulus of interest (e.g., 500 ms before the target), the ERP elicited by the fixation point will overlap the ERP elicited by the stimulus of interest. I have seen many papers in which the prestimulus baseline contained a large and unexpected voltage deflection because of the onset of a fixation point. Second, if a blink occurs immediately before the stimulus on a large proportion of trials, the offset of the blink may be large enough to contaminate the data while being small enough to avoid rejection.



**Figure 6.3**

Recordings from a vertical EOG (VEOG) electrode under the left eye and the Fp1 electrode, which is over the left eye. The offset of a blink is visible in the prestimulus period, appearing as a negative voltage deflection at the VEOG electrode and a positive voltage deflection at the Fp1 electrode. Two different experimental conditions are shown for a single subject (a patient with a neurological condition).

Figure 6.3 shows an example of this from an actual experiment. The data were recorded by a collaborator, Erik St. Louis, from a patient with a neurological disorder. One day, Erik e-mailed me the averaged ERP waveforms shown in this figure and asked what might be causing the obvious distortion in the baseline. It was clear to me that the subject must have frequently blinked just before the onset of the trial, and the tail end of the blink was present during the prestimulus baseline period. Specifically, the offset of the blink was causing a deflection of approximately  $10 \mu\text{V}$  between  $-200$  and  $+200$  ms (relative to stimulus onset). This was far too small to be detected and rejected on the single trials, but it was large enough and consistent enough to cause significant distortion of the ERP waveforms. I will come back to these waveforms in a little bit and tell you what we did to deal with this artifact (the answer may surprise you).

**Detecting Blinks** Blinks are relatively easy to detect on single trials. As I mentioned before, a simple voltage threshold works reasonably well if you have clean data, but it's far from perfect even with clean data and is woefully inadequate in some cases. The moving window peak-to-peak amplitude measure works great, and I highly recommend it (if your software implements it). I will describe another measure—the *step function*—in the section on eye movements, and this function also works nearly perfectly for blink detection. I've never done a formal comparison of these two methods, but I suspect the step function method works a little better than the moving window peak-to-peak amplitude method. However, both work a zillion times better than a simple voltage threshold.

Because the blink potential is negative below the eyes and positive above the eyes, it can be isolated from other EEG activity by recording the difference between an electrode below the eyes and an electrode above the eyes. This also increases the size of the blink response (because a negative value minus a positive value is an even bigger negative value). The best way to do

this is to start with separate recordings from an electrode below the eye and an electrode above the eye, with both electrodes referenced to a common, distant site (e.g., an EOG electrode located below the left eye and the Fp1 site above the left eye, both referenced to a mastoid electrode). You can then create a new channel offline that is the difference between these values (below-the-eye minus above-the-eye). Any brain activity that is in common to these two channels is removed (e.g., most of the EEG), leaving only the blink response and the little bit of EEG that differs between these nearby sites. The blink detection algorithm is then applied to this new channel.

The polarity inversion also makes it possible to determine whether the artifact rejection worked properly. As described earlier, you can look at the averaged ERP waveform at the electrodes under and over the eye (with their original reference) to see if there are any polarity inversions. You might see something like the waveforms in figure 6.3, in which case you may want to take another look at your artifact rejection parameters. You should be especially concerned if an experimental effect has an opposite polarity under versus over the eyes, because this is likely to mean that more blinks survived artifact rejection in one condition than in the other. It's possible for a real effect to show this kind of polarity reversal if the dipole is located at the frontal pole and oriented vertically. However, you should definitely rule out the possibility that it is a blink artifact before accepting it as a real effect.

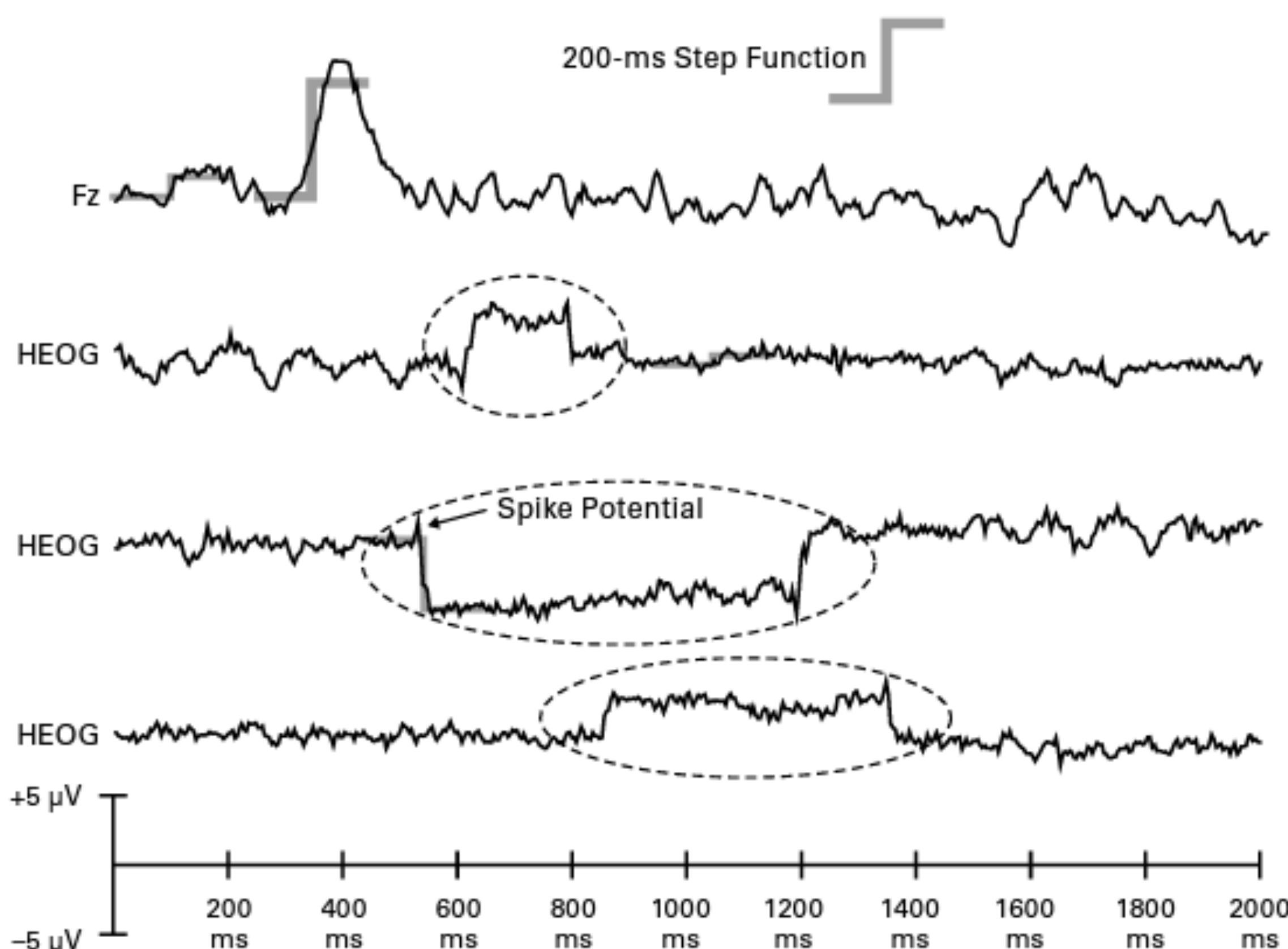
When Erik St. Louis sent me the data shown in figure 6.3, there was no realistic way to adjust the rejection parameters to avoid the blink offset that contaminated the data. The subject blinked prior to the stimulus on almost every trial, so a really low threshold for rejection would have caused almost every trial to be rejected. I told Erik not to worry about it because the experimental effect (the difference between the two conditions) had the same polarity above and below the eyes and could not be a blink artifact. If the data had come from a college student in one of my lab's basic science studies, we probably would have replaced the subject. But the data in figure 6.3 came from a hard-to-find neurological patient, and it was not worth throwing out the data given that the artifact did not impact the experimental effect. The contamination from the blink artifact in this case was much like nondifferential overlap, which can usually be ignored (as discussed in chapter 4 and online chapter 11).

### **Eye Movements**

Like blinks, eye movements are a result of the strong dipole inside the eye. When the eyes are stationary, this dipole creates a static voltage gradient across the scalp, which is removed by baseline correction (and by high-pass filters). When the eyes move, the voltage becomes more positive over the side of the head that the eyes now point toward (for an excellent review, see Plochl et al., 2012). For example, a leftward eye movement causes a positive-going voltage deflection on the left side of the scalp and a negative-going voltage on the right side of the scalp. Horizontal eye movements are most easily observed with bipolar recordings, in which an electrode lateral to one eye is the active site and an electrode lateral to the other eye is the reference site. This is called a *horizontal EOG* (HEOG) recording. Vertical eye movements can be seen

in the same bipolar recordings used to isolate blinks from other EEG activity (i.e., the difference between electrodes below and above the eyes). This is called a *vertical EOG* (VEOG) recording. An eye movement that is not perfectly vertical or perfectly horizontal will show a voltage deflection in both the VEOG and HEOG channels, where the amplitude of the deflection in a given channel reflects the degree to which the eyes moved in the corresponding direction.

Unless the subject is viewing moving objects or making gradual head movements, the vast majority of eye movements will be saccades (sudden shifts in eye position). Examples of saccadic eye movements are shown in the bottom three waveforms of figure 6.4. In the absence of noise, a saccade would consist of a sudden step from one voltage level to another voltage level, where it would remain until the eyes moved again (unless the data have been high-pass filtered, in which case the voltage will fall gradually to zero). In most ERP experiments, subjects make a saccade from the fixation point to some other location and then make another saccade to return



**Figure 6.4**

The step function and its application to an eyeblink recorded from the Fz electrode and to eye movements recorded in a horizontal EOG (HEOG) channel. The Fz recording was referenced to the right mastoid, and the horizontal EOG was recorded with the active electrode adjacent to the right eye and the reference electrode adjacent to the left eye. The areas indicated by the dashed lines are eye movements that move away from the fixation point and then back again, creating a “boxcar” shape. The step function is overlaid with the EOG waveforms in a few places so that you can easily see how the step function matches or mismatches the EOG. The step function would ordinarily be compared with each point in the EOG waveform, not just the few points shown here.

to the fixation point. This leads to a *boxcar*-shaped voltage deflection in the EOG recording (as in the examples in figure 6.4). The length of the boxcar depends on how long the eyes remain at the peripheral location before moving back to the fixation point. This characteristic shape can help you distinguish between real eye movements and noise in the EOG recording.

The average size of the saccade-related deflection has been systematically measured in bipolar recordings by Hillyard and Galambos (1970) and by Lins et al. (1993a), and these studies yielded the following findings: (a) the voltage deflection at a given electrode site is a linear function of the size of the eye movement, at least over a  $\pm 15^\circ$  range of eye movements; (b) an HEOG recording between electrodes at locations immediately adjacent to the two eyes will yield a deflection of approximately 16  $\mu\text{V}$  for each degree of horizontal eye movement; and (c) the voltage falls off in a predictable manner as the distance between the electrode site and the eyes increases (for a list of the propagation factors for a variety of standard electrode sites, see tables V and VI of Lins et al., 1993a).

It should also be noted that eye movements cause the visual input to slip across the retina, which creates a visual ERP response (saccadic suppression mechanisms make us unaware of this motion). This saccade-induced ERP depends on the nature of the stimuli that are visible when the eyes move, just as the ERP elicited by a moving stimulus varies as a function of the nature of the stimulus. Procedures that attempt to correct for the EOG voltages produced by eye movements—which are discussed at the end of this chapter—cannot correct for these saccade-induced ERP responses.

Because of the approximately linear relationship between the size of an eye movement and the magnitude of the corresponding EOG deflection, large eye movements are relatively easy to detect on single trials, but small eye movements are difficult to detect. If a simple voltage threshold is used to detect and reject eye movement artifacts, and a typical threshold of 100  $\mu\text{V}$  is used, eye movements as large as  $10^\circ$  can escape detection (e.g., if the voltage starts at  $-80 \mu\text{V}$ , a  $10^\circ$  eye movement in the appropriate direction will cause a transition to  $+80 \mu\text{V}$ , which would be entirely within the allowable window of  $\pm 100 \mu\text{V}$ ). Of course, a  $10^\circ$  eye movement greatly changes the position of the stimulus on the retina, which can be an important confound, and the resulting voltage deflection is quite large relative to the size of a typical ERP component, even at scalp sites fairly far from the eyes.

The best method for detecting small eye movements that I have seen uses the *step function* shown in figure 6.4 (see box 6.2 for the story of how this method was originally developed). In general, a step function is a flat period of one voltage level followed immediately by another flat period at a lower or higher voltage level. The step function is like the moving window peak-to-peak amplitude algorithm, insofar as a window (e.g., 200 ms) is slid along the data and something is computed within each window. Instead of finding the peak-to-peak amplitude, the step function finds the difference in mean amplitude between the first half of the window and the second half of the window (e.g., between the first 100 ms and the last 100 ms of a 200-ms window). This is an excellent method for finding artifacts that consist of changes in amplitude between adjacent time periods. As you can see from the middle HEOG waveform in figure 6.4,

**Box 6.2**

## The Step Function

When I was in graduate school, I ran lots of N2pc experiments, in which eye movements can be a major headache, and I needed to develop a better way of detecting trials that contained small horizontal eye movements. As figure 6.4 shows, saccadic eye movements have a distinctive shape in HEOG recordings, with a sudden transition from one relatively flat voltage level to a different relatively flat voltage level. That is, they have a step-like shape. I thought I could detect these step-like transitions by looking at the correlation between the HEOG waveform and an actual step function (a 100-ms series of  $-1$  values followed by a 100-ms series of  $+1$  values, as shown in figure 6.4). More precisely, I computed the *cross-correlation function* between the step function and the HEOG waveform. I started by lining up the step function with the initial 200 ms of the HEOG waveform and calculating the correlation (Pearson's  $r$ ) between the step function and this period of the HEOG waveform. I then temporally shifted the step function by one sample period relative to the HEOG waveform and then calculated the correlation again. I did this over and over again, shifting the step function by one sample point each time. This gave me a set of correlations, one for each 200 ms in the HEOG waveform. I then took the highest correlation value and compared that with a threshold to determine whether the trial should be rejected. I reasoned that a high correlation between the step function and any 200-ms section of the HEOG waveform would mean that an eye movement likely occurred in that portion of the HEOG.

When I finally got the program to work, I tested it and found that it worked fairly well but frequently identified sections of HEOG that were fairly flat. When I took a closer look, I saw that these sections did have a step-like shape, but that it was extremely small. In other words, using the correlation worked well for finding the right shape, but it was completely insensitive to the size of the HEOG deflection. I then remembered an important fact that I had read in one of Manny Donchin's papers years before; namely, that you can use covariance rather than correlation when you care about the magnitude of an effect. I changed my program to calculate the covariance between the step function and the HEOG rather than the correlation, and it worked great!

The only problem was that it was very slow. I spent several hours trying to optimize the program, and then I had a realization: The covariance between a step function and some other function is equal to the difference in mean amplitude between the first and second halves of the period of the step function. That is, the covariance for a 200-ms period is simply the difference between the mean amplitude in the first 100 ms of this period and the mean amplitude in the second half of this period. When I rewrote my program to do it this way, I got exactly the same result in 10% of the time.

Now you know how I spent my time in San Diego while everyone else was out surfing.

a step function matches the shape of a saccade-related HEOG deflection very well. It also works well for detecting blinks. This measure is effective for two reasons. First, averaging together the voltage over each 100-ms half of the window filters out any high-frequency noise. Second, computing the difference between successive 100-ms intervals minimizes the effects of any gradual changes in voltage, which corresponds with the fact that a saccadic eye movement produces a sudden voltage change.

The step function can easily detect eye movements of  $2^\circ$  or larger, and it can work reasonably well to detect eye movements as small as  $1^\circ$  under optimal conditions. If even smaller eye movements are a concern in your research (e.g., if you are measuring the N2pc or CDA components), see box 6.3 for additional advice.

**Box 6.3**

## Detecting Small but Consistent Eye Movements

Eye movements are not a big issue for most researchers. For people who study covert shifts of spatial attention, however, eye movements are a major problem. First, when the eyes rotate laterally, the EOG produces a negative voltage over the contralateral hemisphere that contaminates the N2pc and CDA components. Second, if the eyes move prior to the appearance of a lateralized stimulus, this changes the sensory input produced by the stimulus. Even a small eye movement that is consistently directed to a specific location can create a substantial EOG artifact and a meaningful change in stimulus location.

Using a step function, it is often possible to detect eye movements as small as  $1^\circ$  (a 16- $\mu$ V step) on individual trials, but the SNR of the EOG signal is not good enough to detect smaller eye movements without an unacceptably large number of false alarms. However, it is sometimes possible to use averaged EOG waveforms to demonstrate that a given set of ERPs is uncontaminated by very small systematic eye movements. Specifically, if different trial types would be expected to elicit eye movements in different directions, virtually unlimited resolution can be obtained by averaging together multiple trials on which the eye movements would be expected to be similar. For example, if an experiment contains some targets in the left visual field (LVF) and other targets in the right visual field (RVF), one can compute separate averaged EOG waveforms for the LVF and RVF targets and compare these waveforms. The same thing can be done in a cuing paradigm by averaging left-cue trials with right-cue trials (time-locked to the cue). Any consistent differential eye movements will lead to differences in the averaged EOG waveforms, and even very small eye movements can be observed (if they are present on a large proportion of trials) due to the improvement in the SNR produced by the averaging process.

This procedure will not allow individual trials to be rejected, nor will it be useful for detecting eye movements that are infrequent or not consistently related to the side that contains the target. However, it can be useful when combined with the rejection of individual trials with large eye movements in a two-tiered procedure. The first tier consists of the rejection of individual trials with large saccades ( $>1^\circ$ ) by means of the step function. Residual EOG activity can then be examined in the averaged EOG waveforms, and any subjects with differential EOG activity exceeding some criterion (e.g., 1.6  $\mu$ V, corresponding to  $0.1^\circ$ ) can be excluded from the final data set (see, e.g., Woodman & Luck, 2003b).

Eye movements are caused by contraction of the extraocular muscles. This contraction mainly occurs at the onset of the eye movement, and very little muscle activity is needed to maintain the eye position once the saccade has ended. Consequently, a spike of activity from the extraocular muscles (the *spike potential*) is often seen at the onset of each eye movement. You can see this at the beginning and end of each boxcar-shaped deflection in figure 6.4. This artifact is not ordinarily a significant problem, but it can create the appearance of gamma-frequency oscillations in time–frequency analyses (Yuval-Greenberg, Tomer, Keren, Nelken, & Deouell, 2008). This general issue will be discussed more in online chapter 12.

It should be noted that the techniques described above are useful for detecting saccades but are not usually appropriate for detecting slow shifts in eye position or for assessing absolute eye position. To assess these, it is usually necessary to use a video-based eye tracker.

### Skin Potentials and Other Slow Voltage Shifts

As described in chapter 5, skin potentials arise when sweat begins to accumulate in sweat glands, changing the impedance of the skin and therefore causing a change in the standing electrical potential of the skin over a period of many seconds. Some examples can be seen in figure 6.5.

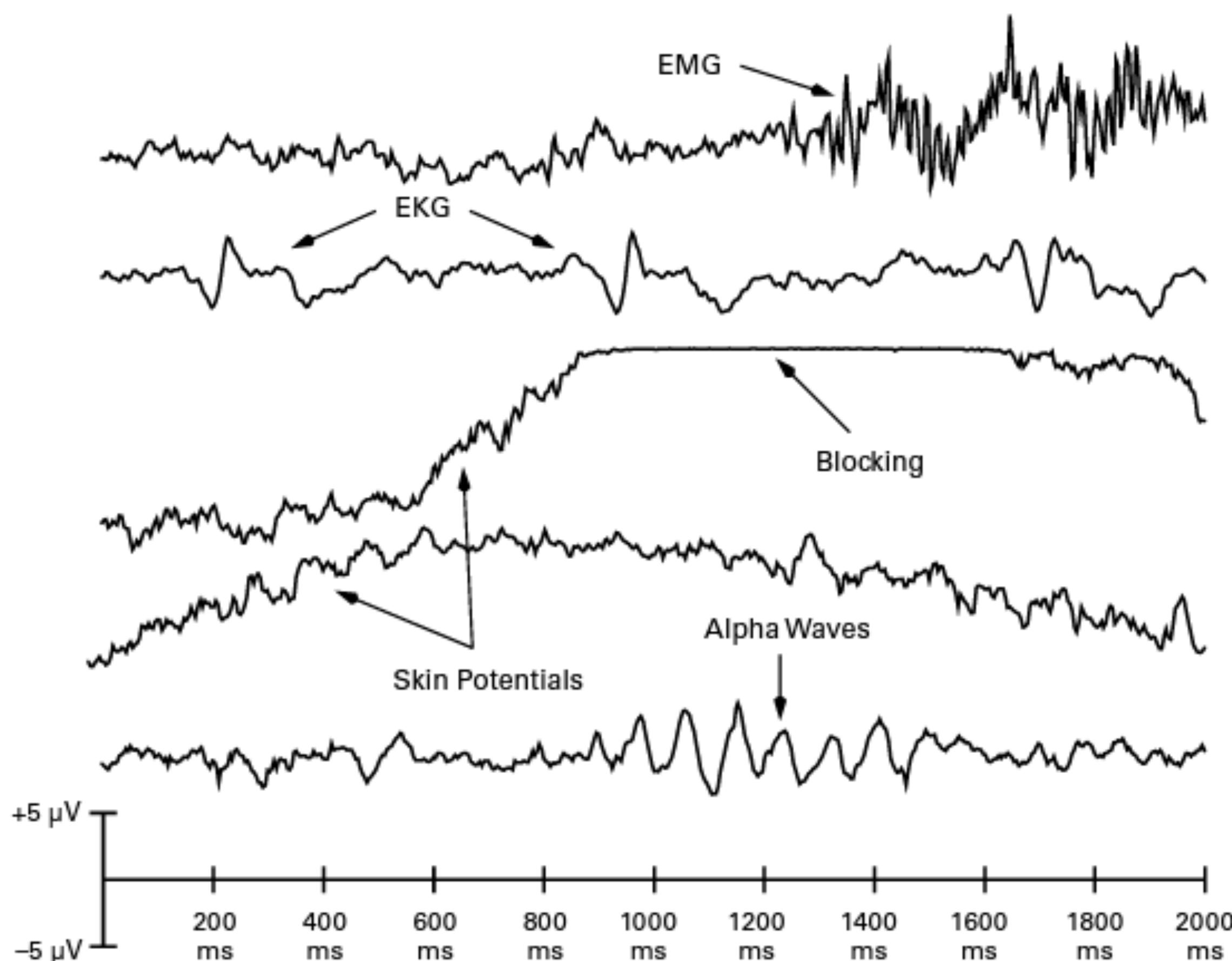
Voltage shifts can also be caused by slight changes in electrode position, which are usually the result of movements by the subject. A change in electrode position will often lead to a change in impedance, thus causing a sustained shift in voltage. This type of artifact can be reduced by making sure that the subjects are comfortable and do not move much. If electrodes are placed at occipital sites, the subjects should not lean the back of the head against the back of the chair.

If the voltage shifts are small, slow, and random, they shouldn't distort the averaged ERPs very much, and there is no need to reject the trials. As discussed in chapter 7, these slow potentials can be minimized by high-pass filters. However, a movement in the electrodes will sometimes cause the voltage to change suddenly to a new level, and this can be detected by means of the moving window peak-to-peak amplitude method or a step function method (you'll want to keep the threshold fairly high to avoid rejecting trials with large ERP deflections, like a large P3 wave).

### Amplifier and ADC Saturation/Blocking

Slow voltage shifts may sometimes cause the amplifier or ADC to saturate, which causes the EEG to be flat for some period of time (this is also called *blocking*). If this happens frequently, you should use a lower gain on the amplifier. This should happen rarely or never in systems with 24 or more bits of resolution (see chapter 5).

As illustrated in figure 6.5, amplifier blocking is relatively easy to spot visually because the EEG literally becomes a flat line. You could reject trials with amplifier saturation by finding trials in which the voltage exceeds some value that is just below the amplifier's saturation point, but in practice this would be difficult because the saturation point may vary from channel to channel and may even vary over time. Another possibility would be to determine if there are a large number of points with identical voltages within each trial, but this isn't quite optimal

**Figure 6.5**

Examples of artifacts on different trials in EEG recordings. The EMG, blocking, and skin potential artifacts were recorded at Cz with a right mastoid reference. The EKG artifacts were recorded at the left mastoid with a right mastoid reference. The alpha waves were recorded at O2 with a right mastoid reference.

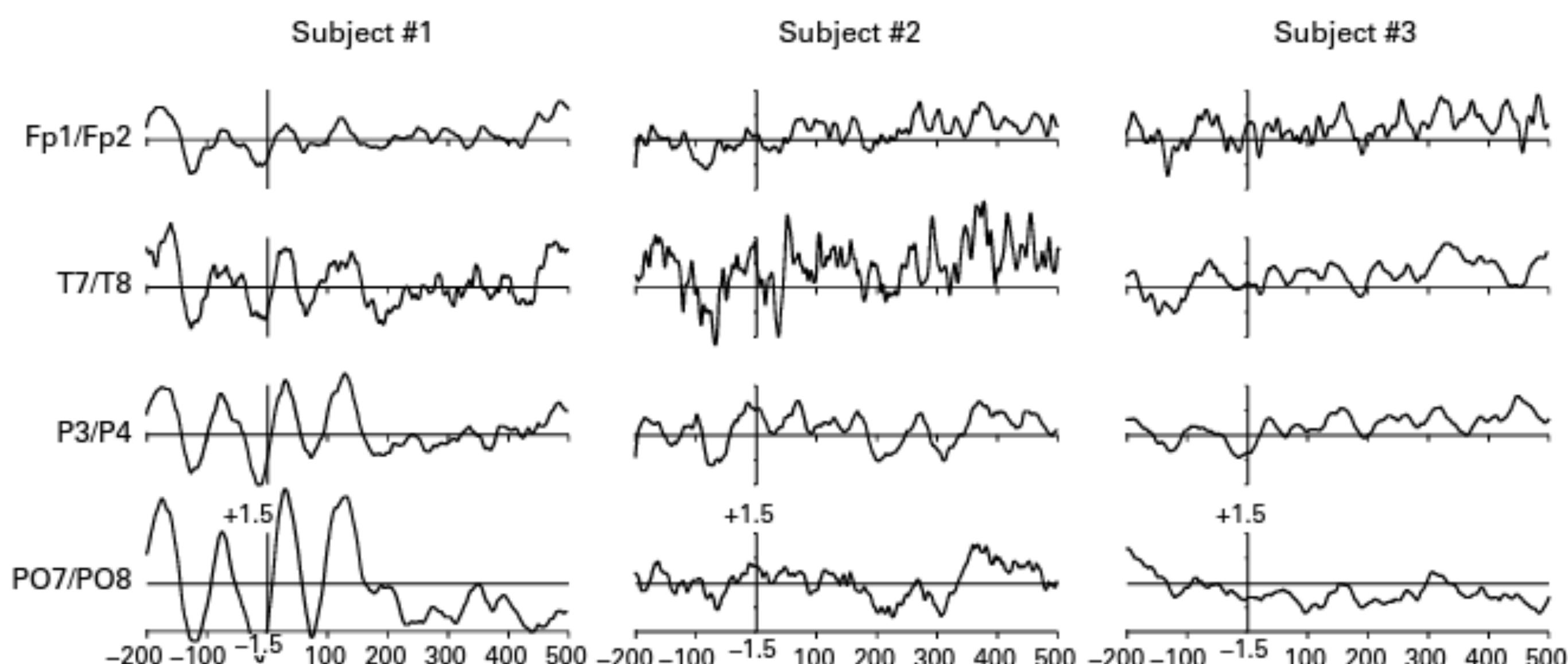
because the voltages might not be exactly the same from moment to moment. A better procedure is to use an algorithm that I call the *X-within-Y-of-peak* method, which was developed by Jon Hansen at UCSD. This method first finds the maximum EEG value within a trial (the peak) and then counts the number of points that are at or near that maximum. X is the number of points, and Y defines how close a value must be to the peak to be counted. For example, you might want to reject any trial in which 30 or more points are within 0.1  $\mu$ V of the peak (i.e., X = 30 and Y = 0.1  $\mu$ V). Of course, the same function must be applied for both the positive peak voltage and the negative peak voltage, and it should be applied to every channel if your system is prone to blocking. The same method can be used to detect other causes of flat-lined data, such as when an electrode becomes intermittently disconnected.

**Alpha Waves** Alpha waves are EEG oscillations at approximately 10 Hz that are typically largest at posterior electrode sites and occur most frequently when subjects are tired or have

their eyes closed (see the bottom waveform in figure 6.5). The best way to reduce alpha waves is to use well-rested subjects and give them interesting tasks to perform, but some individuals have substantial alpha waves even when they are fully alert. Alpha waves can be particularly problematic when a constant stimulus rate is used, because the alpha rhythm can become entrained to the stimulation rate such that the alpha waves are not reduced by the averaging process. Thus, it is useful to include a jitter of at least  $\pm 50$  ms in the intertrial interval (as discussed in chapter 4).

It is not usually worthwhile to reject trials with alpha waves because you will typically end up rejecting almost all trials in some subjects and few or no trials in other subjects. Moreover, alpha oscillations are not always noise, and they may contribute to ERP effects in an important way (see Mazaheri & Jensen, 2008; Bastiaansen, Mazaheri, & Jensen, 2012; van Dijk, van der Werf, Mazaheri, Medendorp, & Jensen, 2010).

Figure 6.6 shows what alpha looked like after averaging across many trials in a subject who had a very large alpha artifact (subject 1 in the figure). As is common, the alpha in this subject was large during the intertrial interval (and therefore during the prestimulus baseline) and became temporarily suppressed within 200 ms of stimulus onset. In addition, the alpha was largest at the back of the head and was quite small at the Fp1 and Fp2 electrode sites. This is a fairly extreme case (the subject with the largest alpha from an experiment with 16 subjects).



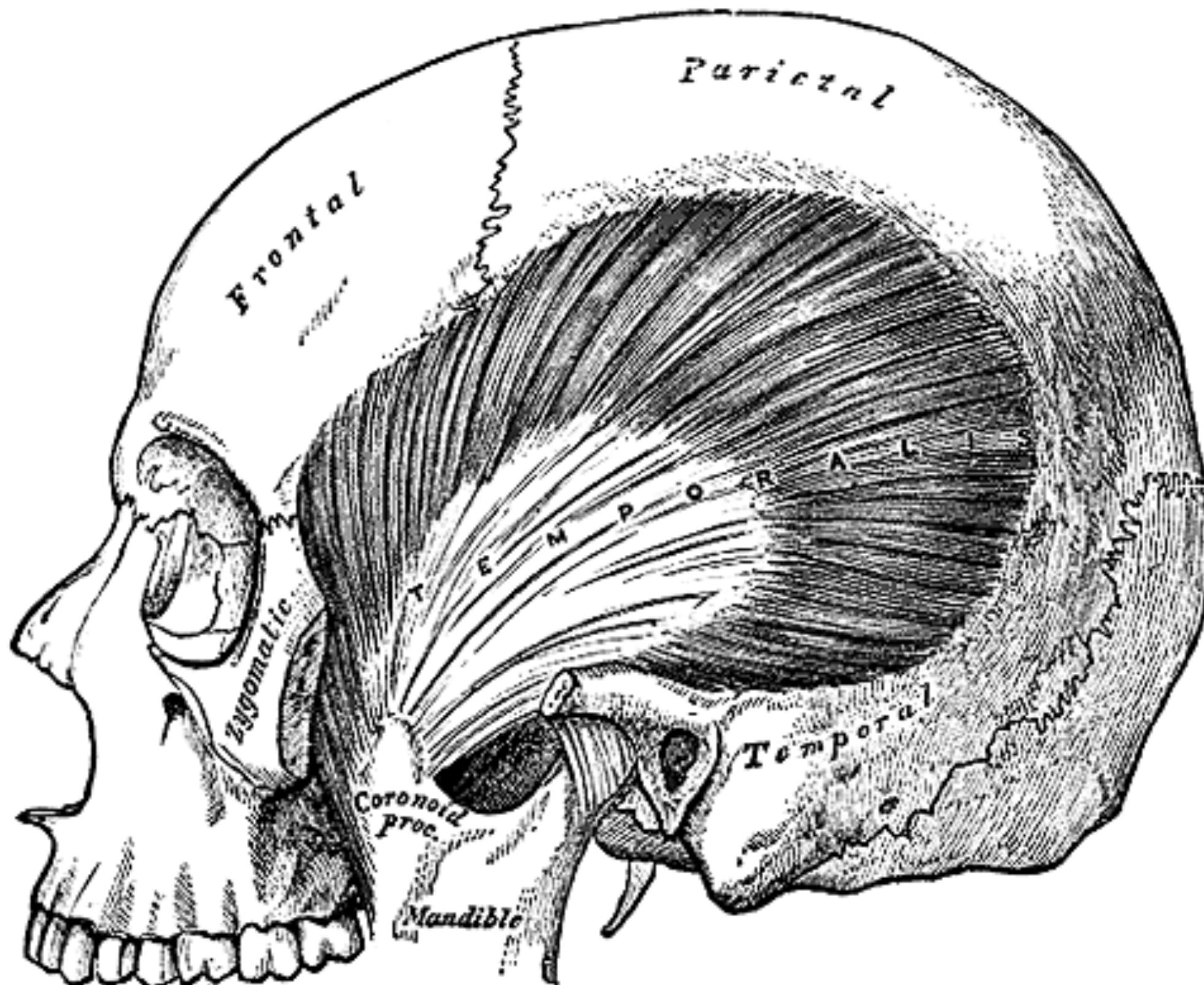
**Figure 6.6**

Examples of artifacts that are still visible after averaging in three electrode sites for three different subjects. These are contralateral-minus-ipsilateral difference waves in an N2pc experiment (data courtesy of Emily Kappenman). You can see alpha activity in the posterior electrodes in subject 1, EMG activity at the T7/T8 electrodes in subject 2, and EMG activity at the Fp1/Fp2 electrodes in subject 3.

### Muscle and Heart Activity

The electrical potential created during the contraction of a muscle is called the electromyogram, or EMG (see top waveform in figure 6.5). The EMG typically consists of rapid voltage fluctuations, and much of the EMG can be eliminated by using a low-pass filter with a half-amplitude cutoff somewhere between 30 and 100 Hz. The best way to reduce EMG is to have subjects relax the relevant muscles. You therefore need to know which muscles cause EMG artifacts at which electrode sites.

The temporalis muscles are powerful muscles that we use to contract our jaws, and they are located right under the T7 and T8 electrodes (see figure 6.7). If you see a lot of EMG in this region of the head, you can ask the subject to relax his or her jaw and avoid teeth-clenching. The temporalis muscles are so large, however, that you will see small but consistent high-frequency EMG artifact at T7 and T8 throughout the session in some subjects, even when they try to relax. Subject 2 in figure 6.6 shows this pattern, with high-frequency noise localized to the T7 and T8 electrodes. And if a subject is chewing on something, you will see enormous bursts of EMG activity at T7, T8, and surrounding electrodes.



**Figure 6.7**

Temporalis muscles. From Gray, H. (1918). *Anatomy of the Human Body*, 20th ed. Philadelphia: Lea & Febiger (now in the public domain).

The muscles of the forehead are also a common source of EMG noise, largest over anterior electrode sites. These muscles can be activated when the electrode cap pulls at the forehead or if the subject is furrowing his or her brow in concentration. Subject 3 in figure 6.6 shows this pattern, with high-frequency noise localized to the Fp1 and Fp2 electrodes. EMG from the forehead muscles can usually be minimized by asking the subject to relax these muscles.

The muscles of the neck are the remaining common source of EMG noise. If a mastoid reference is used, this activity may be picked up by the reference electrode and therefore appears in all channels that use this reference. If a different reference site is used, EMG noise arising from the neck appears at the most inferior occipital and temporal electrode sites. It can usually be minimized by asking the subject to sit straight upright rather than leaning the head forward. Neck EMG can also be minimized by having the subject sit back in a recliner with the head leaning against the recliner, but this can cause artifacts in the occipital electrodes.

It is not usually necessary to reject trials with EMG, assuming that appropriate precautions have been taken to minimize the EMG. However, if it is necessary to reject trials with EMG activity, EMG can be detected by calculating the amount of high-frequency power in the signal (e.g., power above 100 Hz).

It should also be noted that some stimuli will elicit reflexive muscle twitches, and these are particularly problematic because they are time-locked to the stimulus and are therefore not attenuated by the averaging process. These also tend to be sudden, high-frequency voltage changes, but they are usually limited to a very short time period and are therefore difficult to detect by examining the high-frequency power across the entire trial. To reject these artifacts, it is best to look for sudden shifts in voltage during the brief time period during which they are likely to occur (usually within 100 ms of stimulus onset).

The beating of the heart (the electrocardiogram, or EKG) can also be observed in EEG recordings in some subjects, and its distinctive shape is shown in figure 6.5. The EKG propagates to the head through the carotid arteries and is usually picked up by mastoid electrodes. If a mastoid is used as a reference, the EKG is seen in inverted form in all of the electrode sites. The EKG can sometimes be reduced by slightly shifting the position of the mastoid electrode, but usually there is nothing that can be done about it. In addition, this artifact usually occurs approximately once per second during the entire recording session, so rejecting trials with EKG deflections will usually lead to the rejection of an unacceptably large proportion of trials. Fortunately, this artifact is almost never systematic, and it will simply decrease the overall SNR. Because you can't realistically reject trials with EKG artifacts, artifact correction is usually the best approach if the EKG is causing serious problems with your data.

The previous paragraph raises an important point: If you see an artifact or some type of noise equally in all of your EEG channels, it is probably being picked up by the reference electrode. Most artifacts and noise sources will be more prominent at some electrodes than at others, but any signals picked up by the reference electrode will appear in inverted form in all electrodes that use that reference. However, if you are using bipolar recordings for some of your channels (e.g., for EOG recordings), these recordings will not have artifacts or noise arising from the

main reference electrode. This can help you identify and eliminate the sources of noise and artifacts.

### Speech-Related Artifacts

I had always heard that you cannot record clean EEG data while the subject is speaking, but I didn't know why. I thought it must be a result of the muscle activity caused by moving the mouth. Several years ago, Emily Kappenman and I designed an experiment in which we hoped to use a verbal response. To minimize mouth movement, we used /d/ and /t/ as the verbal responses. The EEG was hopelessly contaminated with weird-looking artifacts. When Emily presented the data in a psychology department brown-bag seminar, one of the language-oriented graduate students explained why our data were so noisy. It turns out that there is a strong electrical gradient between the base of the tongue and the tip of the tongue. Consequently, when the tongue moves up and down in the mouth, it creates large voltages that propagate to the surface of the head. These voltages are called *glossokinetic artifacts*. The /d/ and /t/ sounds involve substantial tongue movements, and this is why we were seeing large artifacts.

This places rather unfortunate limitations on language production studies. It is possible to use verbal word-speaking paradigms if the analyses focus on the waveform prior to the spoken output, but it is not usually practical to look at the ERPs after speech onset. It seems plausible that artifact correction could remove this artifact, but I haven't yet seen anyone do it.

### Sporadic Artifacts of Unknown Origin

From time to time, you will see periods of "crazy" voltage fluctuations. This often happens at the end of a break when you tell the subject that you are about to start a trial block. You then see all kinds of large voltage deflections in many channels for a few seconds, and then all the channels return to normal. This also happens occasionally in the middle of a block, usually for just a few seconds. In most cases, you will have no idea why the artifact occurred, and it may be from a combination of things (e.g., a yawn and a stretch followed by a clearing of the throat).

In my lab, we call these *commonly recorded artifactual potentials* (C.R.A.P.; see box 6.4). They are easy to reject with almost any algorithm.

### Some Practical Advice about Artifact Rejection

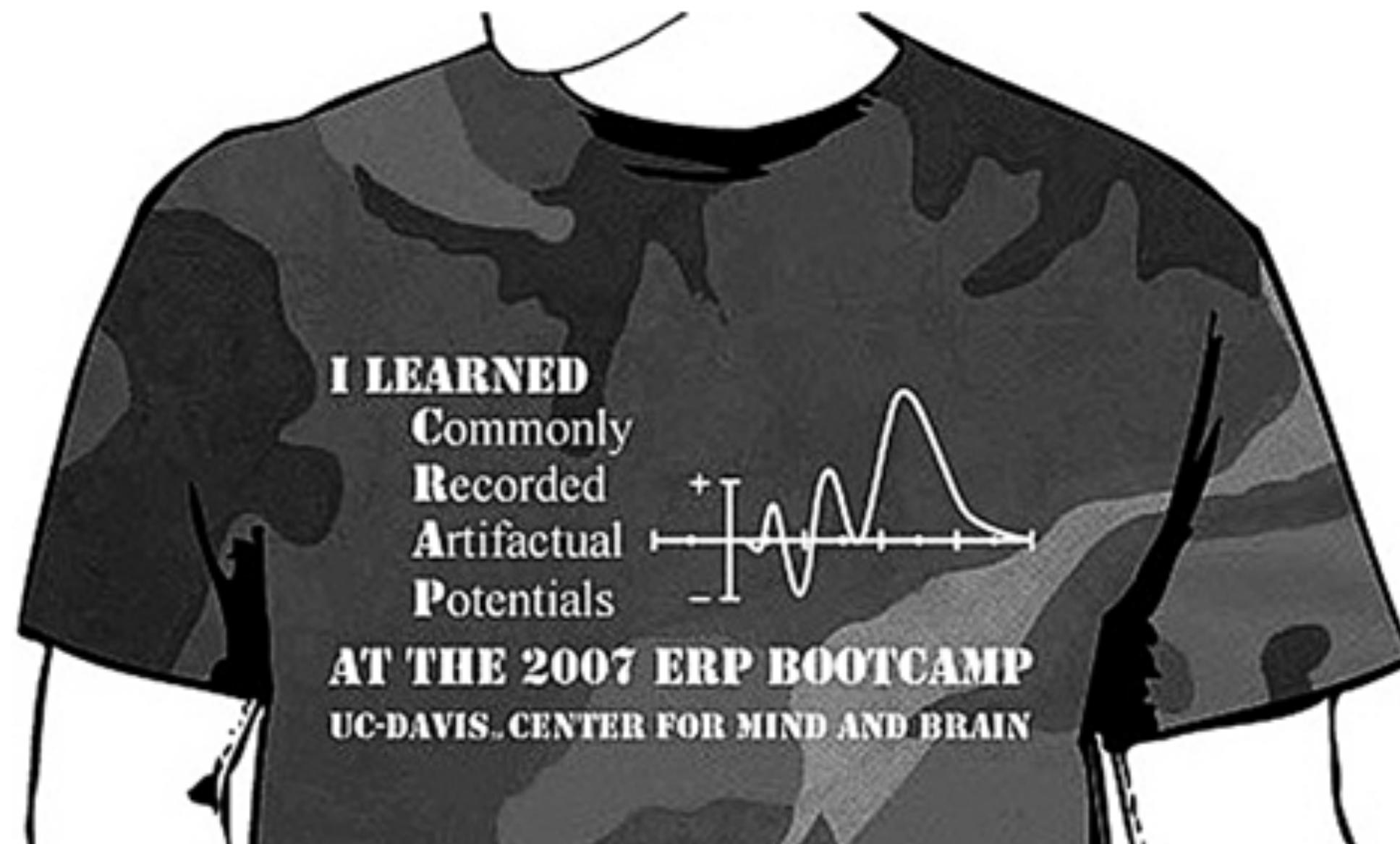
Before transitioning to artifact correction, I'd like to provide some very concrete advice about performing artifact rejection.

First, as described earlier in the chapter, I recommend setting the artifact rejection parameters individually for each subject on the basis of visual inspection of the EEG and adjusting the parameters as necessary to optimize rejection. Once you have become practiced at doing this, it will not take much time, and it will be very worthwhile. If your experiments involve between-group comparisons, however, you need to make sure that the parameters are set by someone who is blind to group membership.

**Box 6.4**

## Commonly Recorded Artifactual Potentials

When describing miscellaneous artifacts to students many years ago, I would refer to them as “crap” in the data. One day, I was giving a guest lecture in a grad course on ERPs, and I said something like “*crap* is a technical term” (just to be funny). I then suggested that someone needed to figure out how C.R.A.P. could be an acronym for some reasonable phrase. Tamara Swaab, who was teaching the class, came up with “commonly recorded artifactual potentials.” I thought this was great, and now I regularly use this acronym. My lab integrated this into the design for the T-shirt we created for the very first UC-Davis ERP Boot Camp.



Second, look at the averaged ERPs to determine whether the rejection is working (especially with blinks and eye movements, as described earlier). It is also useful to create averages that include only the trials with artifacts. This will allow you to see how the artifacts differ across conditions, which may give you hints about how artifacts that escaped detection may be affecting your data.

Third, don't worry that you are reducing your SNR by rejecting trials with artifacts unless you are throwing out more than about 20% of your trials. As was discussed in chapter 5, the square root law means that a 20% reduction in the number of trials leads to only an 11% reduction in the SNR.

Fourth, test for specific artifacts in the channels in which the artifacts are most easily observed (e.g., test for blinks in the VEOG channel and test for eye movements in the HEOG channel). In addition, test for C.R.A.P. in all channels by setting some reasonably high rejection criterion (e.g., a 200- $\mu$ V threshold for the moving window peak-to-peak test).

Fifth, keep track of the percentage of trials that was rejected for each subject, and report this in your journal articles. I like to give the average percentage across subjects and the range (separately for each group, when appropriate). In most cases, you can collapse across conditions when giving these percentages. In experiments where the number of remaining trials may be very small (e.g., studies of infant cognition), you may want to report this information separately for each condition.

Finally, you will probably want to completely exclude subjects for whom a large percentage of trials has been rejected, but you should do this in a way that avoids biasing your results. For example, you will occasionally have a subject who blinks on almost every trial. When this happens in my lab, I tell my students to send the subject home before the experiment is complete, and then the subject is automatically excluded from the final data set. However, if the blinks are happening during the intertrial interval but end up contaminating the baseline, you might still end up rejecting 90% of trials. Obviously you don't want to include a subject like this in your final data analyses. But what if you reject 80% of trials? Or 60%? Or 30%? Or 5%?

Imagine the following scenario. You spend 18 months collecting data for a very important experiment. You do all of the data processing and finally run your statistical analysis. Your conclusions (and your future in academia) hinge on getting a significant three-way interaction. You nervously scan through the output of your statistics program, and your heart sinks when you see that the *p* value for this three-way interaction is the most hated number in all of science, namely 0.06. What do you do? You might take a look at your single-subject ERPs to see if something went wrong with one of the subjects. Imagine that you notice that one subject has noisy and strange-looking waveforms. You then realize that this subject had a very large number of blinks and that 48% of trials were rejected. Obviously, this subject should be excluded from the statistical analyses. You then repeat your statistical analysis, and now the *p* value for your three-way interaction is 0.04. A big smile crosses your face. You write a paper describing your results, noting that you excluded one subject who had artifacts on 48% of trials. Your paper gets published, because reviewers agree that a subject with artifacts on 48% of trials should be excluded, and you get a promotion and a big new grant.

This is a realistic example of how some subjects end up being excluded. It falls within the norms of the field, and it is not a case of ethical misconduct. In some sense, it is obvious that you should exclude a subject with artifacts on 48% of trials. The problem is that the exclusion of the subject was a result of looking at the overall results of the experiment (i.e., it was post hoc). If you exclude subjects only when you find that your effects are not significant, you will produce a systematic bias in your results and end up with more than a 5% false positive rate (i.e., you will reject the null hypothesis when it is true more than 5% of the time). This is obviously a problem, and it does not serve the goal of finding the truth. A *p* value should reflect the probability of a Type I error (a false positive), and anything that makes the *p* values better (smaller) than the actual probability of a Type I error is a misuse of statistics (see box 6.5 for a brief rant about the exclusion of “outlier” subjects).

**Box 6.5**

## Outliers

Many years ago, I reviewed a journal submission in which the authors were looking at the correlation between two variables. I have no memory of what the variables were, so let's just call them X and Y. The correlation between these variables was very important for the conclusions of the paper, but it was not significant unless one "outlier" subject was removed. This outlier subject had X and Y values that were within the typical X range and the typical Y range, but the X,Y combination was quite different from the rest of the subjects. In other words, the subject was an "outlier in multidimensional space." The authors computed some statistic indicating that this X,Y combination was truly beyond the range one would typically expect (e.g., the analog of something like 2 standard deviations away from the mean). Of course, in a truly random sample from an infinite distribution, there is always the chance of obtaining an outlier like this.

The problem with removing an outlier like this is that, in practice, it biases the results and increases the probability of a Type I error (i.e., rejecting the null hypothesis when the null hypothesis is true). I say this with considerable certainty because I have been reading journal articles for about 30 years, and I have never EVER seen a study in which the authors looked at the data and then decided to remove an outlier that would have helped their data if they had kept it. That is, I have never seen a case where the effect was significant (in the predicted direction) when all the data were included, and yet the authors decided to remove an outlier even though the effect was no longer significant with the outlier excluded.

This can and does happen when a criterion for excluding subjects is established prior to the experiment and carried out automatically. However, I have never seen someone do this post hoc. In contrast, I have seen many cases of people excluding outliers that hurt their effects post hoc.

There is growing awareness that practices such as post hoc outlier removal can dramatically increase the rate of false positives in psychology, neuroscience, and related areas (see, e.g., Simmons, Nelson, & Simonsohn, 2011). You can help solve this problem by not letting people get away with this kind of post hoc outlier removal (and, of course, I'm sure you will never do it yourself).

To avoid this problem, I strongly recommend that you adopt a criterion for excluding subjects that you apply to all of your experiments in an *a priori* manner. In my lab's basic science experiments with college students, for example, we automatically exclude any subject for whom more than 25% of trials were rejected, and we automatically include any subject for whom 25% or less of trials were rejected (collapsed across all trial types). We do this irrespective of the way the waveforms look and irrespective of the outcome of the statistical analyses. In our experiments on schizophrenia, we see a lot more artifacts (in both the patients and the control subjects), so we exclude subjects for whom more than 50% of trials were rejected. By adopting a preset criterion, we are just as likely to exclude subjects who "help" or "hurt" our effects, so excluding subjects does not bias us toward obtaining significant results. I encourage you to adopt a standard criterion of this nature and to report it in the methods sections of your papers (e.g., "We always exclude subjects for whom more than 25% of trials were rejected because of artifacts; three subjects were excluded for this reason in the present study").

### Basics of Artifact Correction

In this section, I will describe the reasons why it is sometimes better to subtract away the voltages produced by artifacts (artifact correction) rather than to reject trials with artifacts. I will then describe the rationale behind three major classes of artifact correction approaches. Finally, I will discuss some of the potential pitfalls with artifact correction. An online supplement to this chapter provides a detailed description of one specific method, based on independent component analysis.

#### Why Artifact Correction May Be Useful

Artifact rejection is a relatively crude process because it completely eliminates a subset of trials from the ERP averages. As discussed by Gratton, Coles, and Donchin (1983), there are three potential problems associated with rejecting trials with ocular artifacts. First, in some cases, discarding trials with eye blinks and eye movements might lead to an unrepresentative sample of trials (e.g., you might end up including only trials on which subjects are in a particularly alert state). Second, there are some groups of subjects (e.g., children and psychiatric patients) who cannot easily control their blinks and eye movements, making it difficult to obtain a sufficient number of artifact-free trials. Third, there are some experimental paradigms in which blinks and eye movements are integral to the task, and rejecting trials with these artifacts would be counterproductive. Under these conditions, it would be useful to be able to subtract away the voltages due to eye blinks and eye movements rather than to reject trials with these artifacts.

Ochoa and Polich (2000) identified a related problem: If you instruct subjects not to blink while they are performing a task, you are essentially giving them two tasks to perform at the same time (the task you are explicitly studying and the task of suppressing blinks). If you have ever been a subject in an ERP task in which you were asked to minimize blinking, you know that this takes considerable mental effort. My own experience is that trying not to blink makes me focus attention on the sensations arising from my eyes, which makes me want to blink more rather than less. Ochoa and Polich directly tested the possibility of dual-task interference by giving subjects an oddball task and either telling them to avoid blinking or saying nothing at all about blinking. The P3 had a smaller amplitude and a longer latency when subjects were asked to avoid blinking. The subjects in this experiment were college students, and it seems likely that an even larger effect would be observed in subjects who have lower levels of overall cognitive ability. For example, any differences in P3 between a patient and a control group could reflect the interactive effect of the disorder and the instruction to minimize blinking.

It can be even more difficult for subjects to control their eye movements. Subjects don't have very direct information about where their eyes are pointing and when their eyes move, and this makes it difficult for them to avoid making eye movements. In some cases, it is useful to train subjects to maintain fixation, which can be done with a very simple task that does not involve an eye tracker (Guzman-Martinez, Leung, Franconeri, Grabowecky, & Suzuki, 2009).

### General Approaches to Artifact Correction

To deal with these issues, various procedures have been developed to estimate the artifactual potentials generated by blinks and eye movements and to subtract them from the EEG. Some of these procedures work for other types of artifacts as well. Artifact correction procedures fall into three categories: (1) regression-based procedures, (2) dipole localization procedures, (3) statistical component isolation procedures. Before you read about these procedures, you might want to take a look at box 6.6, which provides some general reasons to be cautious about artifact correction.

Gratton et al. (1983) developed a regression-based procedure to estimate the artifactual potentials generated by blinks and eye movements and to subtract them from the EEG. The basic idea is that an artifact generated in the eyes propagates to each scalp site in a very predictable and

#### Box 6.6

##### Why Artifact Correction Is Scary

I find artifact correction to be a bit scary. Any procedure that subtracts an estimated value from my data makes me nervous. For example, baseline correction and re-referencing are extremely simple procedures that merely subtract a constant value from the entire waveform, and this book contains many, many pages describing the problems that can result from these subtractions (see chapters 5 and 8). Artifact correction procedures are much more complicated than baseline correction and re-referencing procedures, making it more difficult to anticipate exactly how they might distort the ERP waveforms. I am therefore concerned that the cure might be worse than the disease.

Moreover, I have been unimpressed by the evidence that has been provided about the accuracy of artifact correction procedures. Mathematical justifications are unsatisfying because it is hard to know how well the assumptions of the procedure are met by real data. Many attempts at validation simply apply the procedure to existing data and show that the results look sensible. However, these studies do not have a *ground truth* that can be used to assess the accuracy of the correction. In other words, there is no way to know what the true, artifact-free waveforms should look like in these studies, and consequently there is no way to tell how much distortion is produced by the correction. The best studies use real EEG data but add experimenter-created artifact waveforms to these data. This makes it possible to see how well the known artifact is removed (for a really great example of this approach in a different domain, see Kiesel, Miller, Jolicœur, & Brisson, 2008).

I haven't been fully convinced by any of the validation studies of artifact correction that I've read (although it's possible that I missed some). One problem is generality: The fact that a procedure works with one set of subjects, conditions, and artifacts doesn't mean that it will work with other subjects, conditions, and artifacts. In particular, there are some situations that might be particularly difficult to correct, and most studies don't try to "break" the procedure by testing these challenging situations. In many cases, the studies were conducted by the people who developed the procedure, and so it is natural that they did not try to "break" the procedure. However, I am naturally skeptical about such studies, and I much prefer studies that were conducted by independent third parties. The bottom line is this: Take a skeptical perspective when reading studies claiming to provide evidence of the validity of an artifact correction procedure, and don't assume that their results will generalize to your studies unless their subjects, conditions, and artifacts are similar to yours.

quantifiable manner. Consequently, the artifact-related voltage recorded at a given electrode site will be equal to the size of the artifact recorded at the eyes multiplied by a propagation factor. To correct for eye artifacts, therefore, you can simply estimate the propagation factor between the eyes and each of the scalp electrodes and subtract a corresponding proportion of the recorded EOG activity at the eyes from the ERP waveform at each scalp site. For example, Lins et al. (1993a) found that 47% of the voltage present in a bipolar EOG recording propagated to the Fpz electrode, 18% to the Fz electrode, and 8% to the Cz electrode. To subtract away the EOG contribution to the averaged ERP waveforms at these electrode sites, it would be possible to subtract 47% of the EOG waveform from the Fpz electrode, 18% from the Fz electrode, and 8% from the Cz electrode.

Although the development of this procedure was an important step forward, it has a significant shortcoming. Specifically, the EOG recording contains brain activity in addition to true ocular activity, and, as a result, the subtraction procedure ends up subtracting away part of the brain's response as well as the ocular artifacts (see, e.g., Lins, Picton, Berg, & Scherg, 1993b; Plochl et al., 2012). Consequently, I recommend against using the Gratton et al. (1983) method or similar regression-based methods. Thirty years have passed since the introduction of this method, and better techniques are now widely available.

A second approach uses dipole modeling to create a more detailed biophysical model of the artifact and its propagation through the head (Berg & Scherg, 1991a, 1991b). Although I am not usually a big fan of dipole modeling (see online chapter 14), the locations of the ocular dipoles are already known, eliminating much of the uncertainty that is ordinarily involved in dipole modeling. However, the accurate use of this approach may require considerable effort. For example, Lins et al. (1993b) recommended that recordings should be obtained from at least seven electrodes near the eyes, and a set of calibration runs must be conducted for each subject. In addition, this approach typically assumes that a vertical eye movement and the effect of the eyelid passing over the eye have the same scalp distribution, which is not true (see, e.g., Plochl et al., 2012).

A third approach uses the statistical properties of the data to identify a set of components, each of which is characterized by a scalp distribution, and then uses these components to isolate and then subtract the artifact-related voltages. Different statistical methods can be used estimate these components, including principal component analysis (PCA), independent component analysis (ICA), and second-order blind inference (SOBI).

These methods assume that each artifact has a fixed scalp distribution in a given subject. For example, eyelid closure would have one scalp distribution, an eye movement would have another scalp distribution, the EKG would have yet another scalp distribution, and so forth. For many artifacts, this is a valid assumption. The scalp distribution of the EEG at any moment is assumed to consist of the weighted sum of the scalp distribution of the artifact plus the scalp distributions of all the other brain signals. By examining the relationships of the voltages at each electrode site at each time point, these techniques are able to find a small number of scalp distributions that, when summed together, can account for the EEG at each time point. ICA has become the most commonly used of these statistical methods, largely because it is widely available and

relatively easy to use in the free EEGLAB Toolbox package (Delorme & Makeig, 2004). You can find a more detailed description of how ICA-based artifact rejection works, along with an example from a real experiment, in the online supplement to chapter 6.

These methods are not limited to ocular artifacts. They can work with any artifact that has a consistent scalp distribution, including EKG artifacts and induced line-frequency noise from nearby electrical devices. However, they will not work with artifacts that have a variable scalp distribution (as may be the case with skin potentials and movement artifacts).

### Potential Pitfalls and General Advice

No matter what method you use, you should be aware of two potential pitfalls with artifact correction.

First, no artifact correction technique has been demonstrated to work perfectly in all situations. I read several reviews and validation studies to prepare for writing this section, and I found evidence for and against virtually every major artifact correction approach. Different methods appear to work best in different situations. If you decide to use artifact correction in your own research, you should maintain an appropriate level of skepticism about whatever method you use, and you should take steps to assess the validity of the results.

One way to do this is to compare artifact correction with artifact rejection. If your data look similar with both approaches—although noisier with rejection owing to the smaller number of trials—then the correction procedure probably didn't distort your data very much. Carly Leonard, a postdoc in my lab, took this approach the first time we tried using ICA to correct blink and eye movement artifacts. She found comparable results with correction and rejection, although the data were noisier with rejection, and she said so in the methods section of the paper (Leonard et al., 2012). If your results look different between correction and rejection, this may mean that the correction is distorting the data. Alternatively, it could mean that the brain was working somewhat differently on trials with versus without artifacts, and removing trials with artifacts therefore led to an incomplete view of what the brain actually does. If you find yourself in this situation, you will need to think carefully about how to proceed. There is a nice saying in Spanish for difficult situations like this, in which you might need some divine intervention: *Vaya con Dios*.

A second potential pitfall is that artifact correction, even if it works perfectly, cannot account for the changes in sensory input caused by blinks and eye movements. For example, if a subject blinks or makes an eye movement at the time of a visual stimulus, then this stimulus may not be seen properly, and this cannot be accounted for by artifact correction techniques. It should be obvious that closing or moving the eyes will change the processing of a stimulus. An experiment has actually quantified this, showing that reaction times were increased by approximately 200 ms when a blink occurred at the time of stimulus onset (Johns, Crowley, Chapman, Tucker, & Hocking, 2009). Stimulus duration was 400 ms, so the subjects were able to detect the stimuli even when they blinked. If the stimulus duration had been shorter, however, subjects presumably would have completely missed the stimuli.

The obvious solution to this problem is to reject trials with blinks or eye movements that occur within a few hundred milliseconds of stimulus onset prior to performing artifact correction. An alternative solution would be to demonstrate that very few blinks or eye movements actually occurred during this time period. Although these solutions seem obvious, I have seen very few published studies that addressed the problem of blinks changing the sensory input. This may partly reflect the fact that it is not always easy to do both correction and rejection in commercial ERP analysis software. When we were designing ERPLAB Toolbox (which relies on EEGLAB for ICA-based artifact correction), one of our goals was to make this possible.

Should you use artifact correction, or should you just stick with the old-fashioned rejection approach? In the first edition of this book, I said the following: "I would recommend against using artifact correction procedures unless the nature of the experiment or subjects makes artifact rejection impossible." Since that time, my lab has used ICA-based correction in a number of experiments, and we have looked very closely at the results for evidence of problems. As a result, my advice has now changed. I now think it's reasonable to use one of the newer correction techniques to correct for blinks in almost any experiment. Blinks are very large, which tends to make these methods work well, and I have seen no evidence of substantial distortions of the data with ICA-based blink correction.

There are three situations in which I would recommend caution with blink correction. The first is when your experimental effect consists of a subtle effect with a blink-like scalp distribution and a relatively long duration ( $\geq 200$  ms). A subtle distortion from the blink correction algorithm could be problematic and difficult to detect in this situation. The second is when the blinks are highly consistent in their timing (e.g., when the subject blinks at the time of the response or in reaction to an intense stimulus); the statistical methods for finding blink-related activity may lump brain activity with the blinks in this situation. If you are in either of these situations, you might try using correction but testing it very carefully for signs that it is distorting your data (e.g., by comparing rejection and correction).

The third and more common situation in which caution is advised for blink correction is when the blinks differ significantly across groups or conditions, especially during time periods when significant ERP effects are found. How will you know if you are in this situation? One way to determine this is to do the opposite of artifact correction: Instead of removing the component(s) associated with the blink, you can remove all of the other components. If you then average the data, the resulting waveform will show you the artifact-related activity in each group and condition. This seems like something that should be done for every experiment that uses artifact correction. If you find that the artifacts do differ across groups or conditions, you can look at the data in greater detail and determine whether the major effects in your experiments could be explained by inaccuracy in the artifact correction (*Vaya con Dios, amigo!*).

I am less enthusiastic about using correction for other types of artifacts. For example, we have found that ICA works imperfectly for removing eye movement artifacts, as described in the online supplement to this chapter. In Carly Leonard's study, which I mentioned earlier (Leonard et al., 2012), we ended up using ICA to remove eye movement artifacts despite the fact that it

worked imperfectly. It was “good enough,” meaning that several aspects of the results convinced us that the imperfections did not impact our conclusions (e.g., the time course of the eye movements was different from the time course of the ERP effects). But it would not have been good enough if we had been looking for more subtle effects or if we needed precise scalp distribution information. If you use correction for eye movements, make sure that you have a way to determine whether it might be leading you to incorrect conclusions. Plochl et al. (2012) found that ICA worked well for correcting eye movements, but I suspect this reflects the particular nature of their experiments, in which the task involved making large eye movements to a small number of locations. In addition, the nature of the study (which involved real eye movements rather than simulated data) made it difficult to quantify how well the correction worked.

There is controversy about whether these approaches are adequate for removing EMG artifacts (see, e.g., McMenamin, Shackman, Maxwell, Greischar, & Davidson, 2009; McMenamin et al., 2010; Olbrich, Jodicke, Sander, Himmerich, & Hegerl, 2011). The problem may be that the scalp distribution is not sufficiently constant, especially if multiple muscle groups are involved. Similarly, skin potentials may not have a consistent scalp distribution, making them difficult to correct with these methods. I would expect that line-frequency noise and EKG artifacts would be relatively easy to correct, and previous studies have provided some evidence (Jung et al., 2000; Ille, Berg, & Scherg, 2002).

I have sometimes seen people remove dozens of different artifact-related components from the EEG, without a clear understanding of what these components were. “They just looked like artifacts,” the experimenter told me. This seems like a bad idea. As noted in box 6.6, it is a dangerous thing to subtract complex information from your data, and you should do it only when you understand exactly what you’re removing.

In general, I find that remarkably little attention is paid to the details of artifact correction in journal articles. Authors tend to say very little about how they did the correction, and reviewers let them get away with this. Readers of journal articles seem to just accept the assertion that the artifact correction worked. I think we all need to think a little more critically when we read articles that use artifact correction, and editors and reviewers should force authors to give the details of what they did and, in many cases, provide evidence that it did not distort their results.

### Suggestions for Further Reading

- Berg, P., & Scherg, M. (1994). A multiple source approach to the correction of eye artifacts. *Electroencephalography & Clinical Neurophysiology*, 90(3), 229–241.
- Frank, R. M., & Frishkoff, G. A. (2007). Automated protocol for evaluation of electromagnetic component separation (APECs): Application of a framework for evaluating statistical methods of blink extraction from multichannel EEG. *Clinical Neurophysiology*, 118, 80–97.
- Gratton, G., Coles, M. G. H., & Donchin, E. (1983). A new method for off-line removal of ocular artifact. *Electroencephalography and Clinical Neurophysiology*, 55, 468–484.
- Groppe, D. M., Makeig, S., & Kutas, M. (2008). Independent component analysis of event-related potentials. *Cognitive Science Online*, 6.1, 1–44.

- Hillyard, S. A., & Galambos, R. (1970). Eye movement artifact in the CNV. *Electroencephalography and Clinical Neurophysiology*, 28, 173–182.
- Lins, O. G., Picton, T. W., Berg, P., & Scherg, M. (1993). Ocular artifacts in EEG and event-related potentials I: Scalp topography. *Brain Topography*, 6, 51–63.
- 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.
- Jung, T. P., Makeig, S., Humphries, C., Lee, T. W., McKeown, M. J., Iragui, V., et al. (2000). Removing electroencephalographic artifacts by blind source separation. *Psychophysiology*, 37, 163–178.
- Jung, T. P., Makeig, S., Westerfield, M., Townsend, J., Courchesne, E., & Sejnowski, T. J. (2000). Removal of eye activity artifacts from visual event-related potentials in normal and clinical subjects. *Clinical Neurophysiology*, 111, 1745–1758.
- Plochl, M., Ossandon, J. P., & Konig, P. (2012). Combining EEG and eye tracking: identification, characterization, and correction of eye movement artifacts in electroencephalographic data. *Frontiers in Human Neuroscience*, 6, 278.
- Talsma, D. (2008). Auto-adaptive averaging: Detecting artifacts in event-related potential data using a fully automated procedure. *Psychophysiology*, 45, 216–228.
- Verleger, R., Gasser, T., & Moecks, J. (1982). Correction of EOG artifacts in event-related potentials of the EEG: Aspects of reliability and validity. *Psychophysiology*, 19(4), 472–480.