7 Preprocessing Steps Necessary and Useful for Advanced Data Analysis

7.1 What Is Preprocessing?

Preprocessing refers to any transformations or reorganizations that occur between collecting the data and analyzing the data. Some preprocessing steps merely organize the data to facilitate analyses without changing any of the data (e.g., extracting epochs from continuous data), other preprocessing steps involve removing bad or artifact-ridden data without changing clean data (e.g., removing bad electrodes or rejecting epochs with artifacts), and some preprocessing steps involve modifying otherwise clean data (e.g., applying temporal filters or spatial transformations).

eye blinks, muscle movement etc..

This chapter is not a cookbook for preprocessing data but rather contains discussions of different preprocessing steps that you should consider before you start preprocessing your data. Many preprocessing choices and steps depend on details of the experiment design, the equipment used to collect the data, the analyses you plan on performing, and idiosyncratic protocols and preferences that you or your research group have developed.

It is a good idea to keep track of all the details of preprocessing for each subject, such as which trials were rejected, which electrodes were interpolated, and which independent components were removed from the data. This way, your results can be replicated from the raw data if necessary. Further, because some preprocessing choices may introduce biases to the data, you should use the same preprocessing procedures for all conditions to minimize the possibility that any biases that may have been introduced will spuriously cause condition differences.

7.2 The Balance between Signal and Noise

EEG data contain signal and noise. Appropriate preprocessing will attenuate the noise in the data. Unfortunately, however, signal and noise are often mixed together and may be difficult

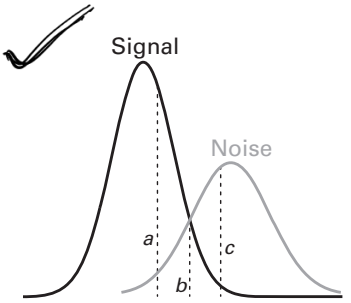

Figure 7.1
Theoretical depiction of signal and noise in EEG data as a signal detection problem. Different preprocessing strategies lead to different balances between the amount of noise versus signal that is retained in data. At one extreme (dashed vertical line a), the data contain nearly no noise, but this comes at the expense of losing signal. At the other extreme (dashed vertical line c), nearly all of the signal in the data is retained, but this comes at the expense of keeping noise in the data as well. There is no optimal position to take in this theoretical function; it depends on the kinds of analyses you are performing, how many trials you have, and how tolerant your analyses are to noise.

to disentangle completely. This leads to a trade-off between signal and noise (figure 7.1): removing a lot of putative noise is also likely to remove some signal, and leaving as much signal in the data as possible will likely mean that noise remains as well. In some cases noise can clearly and unambiguously be dissociated from signal, but in many other cases, the distinction is less clear. Amplifier saturations, for example, produce large spikes that are several orders of magnitude bigger than the neurally generated EEG. There is no ambiguity that this noise spike should be removed from the data.

In other cases, whether a feature of the EEG data is considered noise or signal depends on the researcher and the goal of the study. For example, some researchers apply a low-pass filter of 30 Hz because higher-frequency activity is likely to contain noise and muscle artifacts. Other researchers have built their careers focusing only on activity above 30 Hz. Another example is that many researchers spend their entire careers studying ERPs, whereas other researchers routinely subtract the ERP from the data because it is a potential artifact for some analyses that assume local stationarity. Thus, in some cases, one scientist's noise is another scientist's signal.

Your criterion for balancing signal and noise will influence each step of your preprocessing protocol, ranging from temporal filtering to trial rejection to independent components analysis-based rejection to oculomotor-based trial rejection. Where you set this criterion also depends on how much data you have and on how difficult the data are to acquire. If you have hundreds or thousands of trials, you have the luxury of setting stringent criteria to

remove many trials that may contain noise. On the other hand, if the data were difficult to obtain, for example intracranial recordings in humans, or if the experiment design precludes having many trials in some conditions, you might be willing to tolerate some noise in order to retain as much signal as possible.

Fortunately, time-frequency-based analyses tend to increase the signal-to-noise characteristics of the data, particularly for single-trial analyses and particularly for relatively low frequencies (below around 20 Hz). This is shown in section 18.12.

→ because averaging increases the strength of signal and cancels out noise

↳ why particularly for single trial??

7.3 Creating Epochs

EEG data are recorded continuously and therefore are represented as a two-dimensional (2-D) matrix (time and electrodes). To facilitate investigating task-related changes in the EEG, the continuous data are cut into segments surrounding particular experiment events. Typically, this corresponds to the start of the trial. After epoching, the time-domain data are stored in a 3-D matrix (electrodes, time, and trials). Epoching is not necessary for resting-state datasets, although the continuous data can be segmented into nonoverlapping segments of a few seconds to facilitate analyses.

When epoching, you must decide which event to use for time locking—that is, what to call "time = 0." In some experiment designs this is straightforward: the stimulus onset at each trial is time = 0. Other experiment designs might require some decision making at this step. Experiments in which stimulus-related activity and also response-related activity (which is temporally variable with respect to stimulus onset) are of interest, or experiments in which several stimuli are presented with variable delays, have multiple events that could be used as the time = 0 event. In these cases you could time-lock the data to the earliest event in each trial, which might be convenient for baseline normalization, or you could time-lock the data to the event on which you will focus most of the analyses. The time series data can be temporally shifted during analyses, so the decision of what to use as the time = 0 event will not necessarily limit your analyses. For example, if you time-lock the epochs to stimulus onset, it is possible during the analyses to re-time-lock the data to the response button press without going back to the raw data. If there are events in the trial that have variable timing with respect to the time = 0 event, make sure that all events are in the epoch and that there is sufficient buffer zone (discussed below) for edge artifacts at the end of the epoch.

→ means choosing which event in a trial becomes t=0 (the alignment point) After time locking every trial is lined up so that the chosen event happens at the same time point across all trials

↳ This lets to average/compare trials properly

The other decision you need to make when epoching—which may have significant consequences for the quality of the time-frequency decomposition, particularly for lower frequencies—is how much time to include before and after the time = 0 event. How long the epochs should be depends, of course, on the experiment—the epochs need to be at least as long as

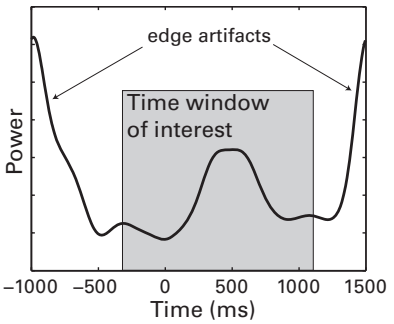

Figure 7.2
Edge artifacts resulting from discontinuous breaks in the time series between trials can contaminate the results if there are insufficient buffer zones to allow those edge artifacts to subside. In this case the edge artifacts are easily identifiable, and it is also clear that those artifacts subside before the time window of interest (gray area). In general, edge artifacts will contaminate up to three cycles of activity, but this could be less or more depending on the magnitude of the edges.

the duration of the trial. But the length of the epochs is also related to the kinds of analyses you want to perform. If you will compute only ERPs, the epochs can be as long as the time period you want to analyze plus a baseline period. For example, –200 ms until 800 ms relative to stimulus onset might suffice.

However, if you plan on performing time-frequency-based analyses, you should create longer epochs. The reason is to avoid contaminating your results with edge artifacts. Edge artifacts result from applying temporal filters to sharp edges such as a step function, and produce a high-amplitude broadband power artifact that can last hundreds of milliseconds. These artifacts will always be present when there are noncontinuous breaks, which happens at the first and last points of the EEG epochs (see figure 7.2 and chapter 14). At the first and last points of each epoch, the time series theoretically goes to zero and continues at zero for infinity. Thus, the transition between the outer "zeros" and the real data are sharp edges.

One way to avoid contaminating your results with edge artifacts is to have long epochs that allow the edge artifacts to subside before and after the experiment events in which you are interested. These extra "buffer zones" at the start and end of the trial will be discarded after the time-frequency decomposition. For example, if you intend to analyze the time period from 0 to 1000 ms, you might take epochs from –1500 to +2500 ms (thus, an extra 1500 ms on either end). How much of a buffer zone you need depends mainly on the frequencies that you intend to extract from the data. Edge artifacts typically last two or three cycles, although this depends on the magnitude of the edge. Thus, the lower the frequency

band that you will extract from the data, the more buffer zone you will need to be confident that the edge artifact has subsided. For example, at 0.5 Hz, the edge artifact can be several seconds long, whereas at 100 Hz, the edge artifact will be only a few tens of ms long. Fortunately, edge artifacts are easy to identify in time-frequency power results. If you are unsure how much of a buffer zone to include, you can analyze one subject and inspect the results carefully to see whether edge artifacts are contaminating the time period of interest. As a general rule of thumb three cycles at the lowest frequency you will analyze should be sufficient for a buffer zone (e.g., 1500 ms for 2-Hz activity).

One caveat to taking large epochs is that, depending on the amount of time between trials, you may have overlapping (and redundant) data in each epoch. This is not problematic for analyses because the overlapping data will be discarded, but it can introduce biases if you perform independent components analysis because the independent components analysis will be run on some time points more often than on other time points. If you have epochs with overlapping data, and if you plan on using independent components analysis to analyze data (that is, analyzing the component time courses instead of the electrode time courses), be sure not to expose the independent components analysis to the same data more than once.

Having sufficient buffer zones is necessary if you will perform time-frequency decomposition via complex Morlet wavelet convolution or the filter-Hilbert method. For time-frequency decomposition via the short-time FFT or multitaper, long buffer zones are not necessary because the time-frequency decomposition occurs on a more temporally local scale (this is explained in chapter 15). Nonetheless, it is still a good idea to take larger epochs than you would for ERP-only analyses.

If you are reanalyzing a dataset that has already been epoched and cannot be reepoched from the continuous data, and if you are concerned that the epochs are too short and the time-frequency results might be contaminated by edge artifacts, you can use a "reflection" approach, whereby the EEG data from each trial and electrode are reversed and put in the beginning and end of the trial. This makes the epoch three times as long, thus creating a buffer zone for edge artifacts. Remember to discard the reflected data after analyses, because the reflected data are backward versions of real data and cannot be interpreted as "pretrial" or "posttrial" activity. See figure 7.3 for an overview of how time series reflection works.

Reflection should be used as a measure of necessity, not as a substitute for including buffer zones in epoching. The reason is that time-frequency analyses involve some temporal smoothing, which you will read more about in chapters 10–14. The temporal smoothing may cause activity to leak out of the reflected data and into the time region of interest. Imagine, for example, that your epochs were cut from –100 to +1000 ms and there is low-frequency activity from around 200–600 ms. When the data are reflected, the low-frequency

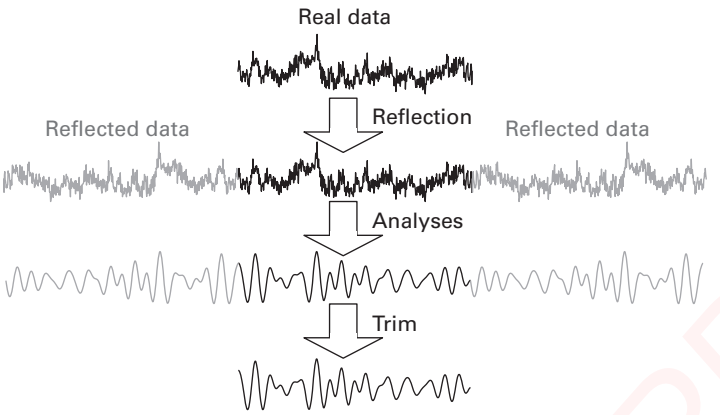

Figure 7.3
Data reflection procedure. Data are reversed in time, concatenated to both ends of the real-data time series, analyses are performed, and then the reflected data are trimmed. This procedure attenuates edge artifacts and can be useful if the epochs are cut too short for planned analyses.

activity toward the beginning of the trial may spuriously enhance the low-frequency activity in the nonreflected data. If this leaks into in the prestimulus period, this could lead to a misinterpretation of the results. Thus, when epochs are created during preprocessing, long epochs with buffer zones are best, but reflecting short epochs is a better strategy than contamination due to edge artifacts.

You might think that instead of taking long epochs with buffer zones or reflecting data you can simply apply a taper to attenuate the data at the beginning and at the end of each trial (similar to the way time segments are tapered for the short-time FFT or multitaper; chapters 15–16). This is generally not a good idea. If you attenuate the data at the start of the epoch, you will decrease the time-frequency power during the baseline period, which will spuriously and inappropriately enhance the trial-related power in the task period. Even if the attenuation occurs before the start of the baseline period, because temporal filtering (done either through wavelet convolution or the filter-Hilbert method) involves setting the activity at each time point to be weighted sum of activity from previous time points, the filtered signal during the baseline period will still be artificially attenuated. Thus, tapering the entire epoch time period is not recommended.

7.4 Matching Trial Count across Conditions

It is ideal for all conditions to have the same number of trials. Depending on the analysis performed, differences in trial count may have little impact or may have a significant impact

on the results. In general, analyses based on phase are more sensitive to trial count than are analyses based on power or on the ERP. For phase-based analyses, a small number of trials will introduce a positive bias in the results, such that conditions with fewer trials are more likely to show larger effects. Power-based analyses may also have some positive bias because raw power values can only be positive, and thus noise is more likely to increase than decrease power. The ERP is not specifically negatively or positively biased by low trial count because the time-domain voltage values can take negative or positive values. However, low trial count decreases signal-to-noise ratio and therefore makes the ERP more sensitive to outliers or nonrepresentative data. For low trial count with ERPs, taking the mean amplitude in a time range is more robust to noise compared to peak times (Clayson, Baldwin, and Larson 2012; Luck 2005).

If there are small differences in trial count across conditions (e.g., 84 trials in condition A and 87 trials in condition B), you do not need to be concerned about potential biases introduced by condition differences in trial count. However, if there are large differences in trial count across conditions, particularly if one condition has fewer than 30 trials (e.g., 84 trials in condition A and 23 trials in condition B), you should be concerned. Whenever possible, try to design the experiment such that there are similar numbers of trials in each condition. However, this is not possible with all experiment designs; if your experiment necessarily entails an unbalanced trial count across trials, you can consider matching trial count across conditions.

Matching trial count generally involves identifying the condition with the fewest trials and selecting trials from other conditions such that all conditions end up with an equal number of trials. You will have to decide whether removing some good data is worth minimizing potentially spurious results due to trial count differences. There are three ways to match trial count across conditions.

One way to match trial count is to select the first *N* trials from each condition, where *N* is the number of trials in the smallest condition. This approach generally should not be taken because it biases some conditions toward having more trials earlier in the experiment when subjects are less tired, more motivated, and less practiced on the task.

A second way to match trial count is to select trials at random. This has the advantage that there is no bias in terms of when the trial occurred in the experiment but has the disadvantage that reanalyzing the same data multiple times may yield slightly different results. Thus, if you randomly select trials, it would be a good idea to store in a variable which trials were selected.

A third way to match trial count is to select trials based on some relevant behavioral or experiment variable such as reaction time. Here the idea is to select a subset of trials from all conditions such that the distributions of reaction times from the retained trials are similar

across conditions. This approach helps to equalize time on task and other general cognitive factors that contribute to reaction time (e.g., attention, engagement, motivation) and may thus help to rule out some alternative explanations of condition differences. The main disadvantage is that if there are reaction time differences between conditions, matching reaction times across conditions may bias trial selection from different regions of the reaction time trial distribution in different conditions. Selective sampling may be based on any other relevant behavioral measure such as saccade speed, pupil response, subjective difficulty rating, or the selective sampling may be based on a stimulus property such as luminance or location of a stimulus.

Trial count matching is not necessary across subjects unless you compare different groups of subjects such as patients versus controls or if you want to correlate the EEG results across subjects with a behavioral variable that might be related to trial count. Chapters 18, 19, and 26 show examples of the effects of trial count on time-frequency results.

If you match trial count across conditions, you should report this in the Methods section of the paper and, if relevant, also report the behavioral results before and after trial selection.

7.5 Filtering

→ removes slow drifts

Filtering data can help remove high-frequency artifacts and low-frequency drifts, and notch filters at 50 Hz or 60 Hz help attenuate electrical line noise. Most time-frequency decomposition methods (wavelet convolution, FFT, filter-Hilbert) involve applying a set of temporal filters to the data. Thus, filtering the data might not be necessary if you will focus on time-frequency dynamics. For example, there is no need to low-pass filter the time-domain data at 40 Hz if you will then perform time-frequency analyses to extract power from 2 to 20 Hz. Applying a high-pass filter at 0.1 or 0.5 Hz to the continuous data is useful and recommended to minimize slow drifts. High-pass filters should be applied only to continuous data and not to epoched data. This is because the edge artifact of a 0.5-Hz filter may last up to 6 s, which is probably longer than your epochs. Filtering ERPs is discussed in chapter 9, and details about how to design and implement bandpass filters in Matlab are presented in chapter 14.

7.6 Trial Rejection

Removing trials that contain artifacts prior to analyses is an important preprocessing step. It is also perhaps the preprocessing step most open to interpretation and idiosyncratic preferences. Some trials should unambiguously be rejected, and other trials will incite disagreement about whether they should be rejected. Some researchers argue for manual trial rejection

based on visual inspection, whereas others argue for automatic trial rejection so that the researcher never needs to look at the data. Automatic rejection procedures are fast and free of user bias, and will result in the same trials being rejected regardless of who handles the data. On the other hand, automatic procedures may use criteria that are appropriate for some subjects but not for others and may produce both Type I and Type II errors (that is, trials you think should be retained are rejected, and trials you think should be rejected are retained). I have tried several automatic algorithmic procedures and found them unsatisfactory and therefore prefer manual trial rejection based on visual inspection. But this is based on my experiences and my preferences for seeing raw data. Your decision of whether to use manual or automatic rejection should be based on your experiences, preferences, and data. Chapter 8 has more discussion about identifying and removing artifacts. If you will perform time-frequency decomposition of the data, you should be aware that sharp edges in the data are more detrimental to time-frequency decomposition than they are to ERPs. Relatively small sharp edges may be undetected by automatic algorithms but can have adverse effects on time-frequency decomposition results.

because wavelets/filters spread out the artifact across time-containing most of the data so even small sharp edges (that an automatic algo might miss) can badly affect time frequency Analysis results

7.7 Spatial Filtering

mathematically combining signals from electrodes to highlight some brain activity & suppress others

There are three main reasons to apply spatial filters to your data. The first is to help localize a result. For example, if you want to confirm that an activity peak corresponds to left motor cortex, you can apply a surface Laplacian or fit a single dipole. → ??

The second reason to apply spatial filters is to isolate a topographical feature of the data by filtering out low-spatial-frequency features. For example, if your task involves visual stimuli that require spatial attention, it may be difficult to separate visual processing in occipital cortex from attention-related processing in parietal cortex. In this case the surface Laplacian or distributed source imaging may help minimize the spatial overlap between occipital and parietal responses, thereby increasing confidence in functional/anatomical distinctions.

The third reason to apply spatial filters is as a preprocessing step for connectivity analyses. Some spatial filters such as the surface Laplacian or distributed source imaging help to minimize volume conduction—a possible artifact that may contaminate some kinds of connectivity analyses. Not all spatial filters (e.g., PCA) will address volume conduction artifacts, and not all connectivity measures require spatial filters.

The stage of preprocessing in which you should apply the spatial filter (e.g., before or after time-frequency decomposition) depends on what kind of filter you will use and what the goal of filtering is. Here are a few examples. If you are performing an ERP study on response preparation and want to see whether an ERP peak is consistent with a source in the motor cortex,

a single dipole can be fit to the grand-averaged ERP. If you are using the surface Laplacian to minimize volume conduction for connectivity analyses, the Laplacian should be applied to the single-trial time-domain data before time-frequency decomposition is performed and connectivity is computed. If you are performing a PCA on time-frequency power, the PCA can be performed either on single trials within a subject or on the trial-averaged power across subjects, depending on whether you are performing a subject-level or group-level PCA.

7.8 Referencing

Referencing is an issue only for EEG; MEG is a reference-free measurement. The issue is that the voltage values recorded from each electrode are relative to a voltage value recorded elsewhere. Where should this "elsewhere" electrode be? In theory, it could be anywhere, such as the subject's toe (or the experimenter's toe) or a wall in the experiment room. These are poor choices because the activity recorded from the reference electrode will be very different from the activity recorded from the electrodes on the scalp. Any activity present in the reference electrode will be reflected as activity in all other electrodes. Thus, choose your reference electrode(s) carefully and make sure the reference electrodes are properly placed and have a good, clean signal during subject preparation. Noise in the reference electrodes will turn into noise in the scalp electrodes. Note that the surface Laplacian is reference independent.

Averaged mastoids (the bone behind the ear) or earlobes are typical reference electrodes. (The sample data online are referenced to linked earlobes.) These are good choices because the reference electrodes are close to the other electrodes but record less brain activity. But mastoids and earlobes are not perfect references because they are close enough to the brain to measure neural activity, for example, from lateral temporal areas. For a large number of electrodes (>100), or if electrodes are placed on the neck and face, an average reference is often recommended. Referencing to one lateralized site is not recommended because this will introduce a lateralization bias in the data. The reference electrode should not be close to an electrode where you expect your main effects. For example, Cz is a poor choice of reference if your analyses involve response errors (which elicit maximal activity around electrode Cz or FCz). An example of the same data with different referencing schemes is shown in figure 7.4.

For intracranial recordings there are several options for a reference electrode, including an electrode in the bone if available, the average of all electrodes in each grid/probe, or another intracranial electrode. Using an external scalp electrode as a reference for intracranial data might be a suboptimal choice because any scalp activity measured by the reference electrode (possibly including muscle activity) will contaminate the intracranial recordings.

Because referencing is a linear transformation of the data, data can be re-referenced offline. Thus, the electrode that serves as the reference during recording is not very important. On

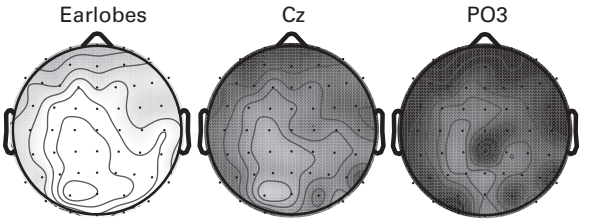

Figure 7.4
The effect of different reference electrodes on the same data. Earlobes refers to the average of electrodes placed on the two earlobes. In many situations, using one of the scalp electrodes as the reference is suboptimal.

the other hand, make sure you know which electrode is used as the online reference; otherwise you might think that you have a bad electrode that records only a flat line.

Some electrodes have a bipolar reference, meaning one electrode is measured relative to another. This is often the case for eye electrodes (one above the eye and one below the eye are referenced to each other, leaving only one signal; and one to the left of the eye and one to the right of the eye, again referenced relative to each other leaving only one signal), EMG, or EKG.

The choice of referencing is a widely discussed topic in EEG research with varied opinions, many of which conclude by noting that no reference is perfect, and the one you should use depends on a variety of factors including how many electrodes you have, where the electrodes are placed, what analyses will be performed, and what kinds of cognitive tasks you use, and thus, from which brain regions you hope to elicit activity (Dien 1998; Junghofer et al. 1999; Schiff 2005).

7.9 Interpolating Bad Electrodes

Interpolation is a process by which data from missing electrodes are estimated based on the activity and locations of other electrodes (see figure 7.5). Most interpolation algorithms use a weighted distance metric such as nearest-neighbor, linear, or spline. The more electrodes you have, the more accurate the estimation will be. Interpolation is used when there is a bad electrode that is either completely flat or that measures noise that is many orders of magnitude larger than real brain signal.

Ideally, you would not need to interpolate because all of your electrodes provide good data. In reality, however, interpolation is often necessary in a minority of cases, particularly with high-density recording systems or with old and overused electrodes. While recording the data during the experiment, it is a good idea to scan the data and look for bad electrodes.

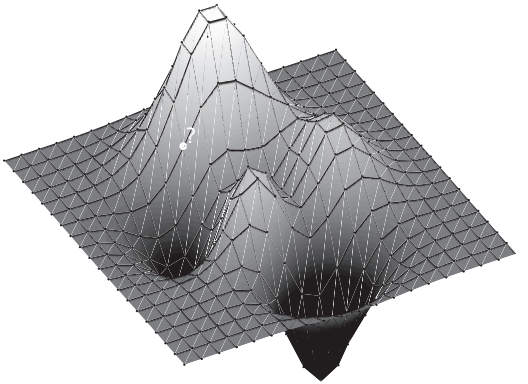

Figure 7.5
Topographical illustration of interpolation. Displayed is a smooth topographical landscape (analogous to a scalp-measured voltage) that is discretely sampled (black dots; analogous to electrodes). Interpolation involves estimating the activity at the white "electrode," given the activities and distances of all other electrodes. This topography was generated with the Matlab `peaks` function.

If you notice a bad electrode during the experiment, pause the experiment if possible and try to fix or replace the bad electrode. It is preferable to reject a few trials but keep all of your electrodes than to interpolate a bad electrode for the duration of the experiment.

The problem with interpolated electrodes is that they do not provide unique data; they are a perfect weighted sum of the activity of other electrodes. This reduces the rank of the data matrix, which may lead to problems in analyses that require the matrix inverse (taking the pseudoinverse is usually an appropriate solution).

The alternative to interpolating bad electrodes is simply to ignore them by removing them from the data or by setting their activities to NaNs and using Matlab's nanmean function during group-level analyses. Removing an electrode entirely from the dataset may cause some confusion when averaging across subjects because, for example, one subject will have 63 electrodes while another subject will have 64 electrodes.

Interpolation can be particularly important for some spatial filters such as the surface Laplacian or source reconstruction or if you will re-reference to the average of all electrodes: the activity of one bad electrode may contaminate the clean signal of other electrodes.

When deciding whether to interpolate an electrode, inspect the data carefully. In some cases there is a true brain signal recorded by the electrode but a lot of noise on top of that signal. If this is the case, try to filter out the noise without interpolating the electrode. One way to determine whether a noisy electrode contains brain signal is to apply a low-pass filter to the data at 30 Hz. If the low-frequency activity from that electrode looks similar to that of

the surrounding electrodes, it is likely that the unique signal from that electrode can be salvaged, and interpolation may not be necessary. If the low-frequency activity from that electrode looks completely different from activity of surrounding electrodes, particularly if the magnitude of the data is much smaller or much larger than that of surrounding electrodes, it is unlikely that any real brain signal was recorded by that electrode, and then it could be interpolated.

Most or all EEG analysis packages have routines for interpolating electrodes. The online Matlab code for this book also provides a spherical spline method, which is based on the same method used for the surface Laplacian, presented in chapter 22.

7.10 Start with Clean Data

There is no data analysis substitute for clean data. The fanciest and most sophisticated data analyses will give you bad results if you start with bad data. On the other hand, if you have very clean and high–signal-to-noise data, even modest analyses can provide compelling and insightful results.

Preprocessing can help turn good data into very good data, but no amount of preprocessing will turn low-quality and noisy data into very good data. Do not rush into a recording if you are unsatisfied with the data quality, and do not be afraid to pause the experiment if the data quality suddenly decreases during a recording and you think you can fix it. Explain to subjects (if they are human and awake) the importance of collecting clean data; they will try to help give you clean data. The next chapter discusses additional strategies for acquiring clean data.

8 EEG Artifacts: Their Detection, Influence, and Removal

If you ask EEG researchers to list the types of artifacts that contaminate EEG data, most will identify blinks, muscle movements, brief amplifier saturations, and line noise as the main artifacts that plague EEG data. These are true sources of artifacts, and they should be removed from the data before analyses are performed, but they are not the only artifacts that contaminate EEG data. There are also cognitive artifacts that should be identified and removed from your data. These artifacts might not have any noticeable correlate in the EEG, but they may contaminate your results nonetheless. This chapter provides an overview and discussion of the different types of artifacts in your data, how to identify them, and some ways to deal with them. It is not necessarily the case that you need to reject all the types of trials mentioned in this chapter. Some of these suggestions may be appropriate for your data, and others may be of no concern for your data. However, you should consider these points and think carefully about the procedures you use for identifying and removing artifacts in the data.

One important concept that is worth stressing here—and which newcomers to EEG often struggle with—is that EEG is *not* a noise-free measurement. There is no way to remove all of the noise from EEG, in part because it is unknown what all of the sources of noise are. Noise can also be difficult to identify. For example, high-frequency activity may look like noise at the single-trial level, and non-phase-locked activity may look like noise when one is computing an ERP. Certainly, you should try to minimize noise, but you should also realize and accept that your data will never be completely noise free. Most EEG data analyses are robust to the levels of noise that are typically observed in EEG data (after preprocessing), particularly with dozens or hundreds of trials per condition.

8.1 Removing Data Based on Independent Components Analysis

Independent components analysis is a source-separation technique that decomposes the EEG time series data into a set of components that attempt to identify independent sources of

variance in the data. It is often described in terms of audio recordings at a cocktail party: imagine there are many voices talking and many microphones placed around the room. By considering weighted combinations of the microphones' recordings, you could isolate the sound coming from individual voices. In this case the result of an independent components analysis is a set of weights for each microphone such that the weighted sum of all microphones best isolates the voice from one person. In the context of EEG the independent components analysis provides a set of weights for all electrodes such that each component is a weighted sum of activity at all electrodes, and the weights are designed to isolate sources of brain electrical signals (Jung et al. 2000; Makeig et al. 2004). Some aspects of independent components analysis are similar to principal components analysis, which you will learn more about in chapter 23.

(1) Independent components analysis can be used either to clean EEG data by identifying components that isolate artifacts and then subtracting those components from the data, or (2) as a data reduction technique by analyzing component time series instead of electrode time series. Nearly all of the time series analysis methods presented in this book can be applied to independent components in addition to electrode time series.

Instead of analyzing all electrodes, we can analyze fewer components often representing meaningful brain sources more clearly

When independent components analysis is used as a preprocessing tool, components can be judged as containing artifacts based on their topographies, time courses, and frequency spectra. Components containing blink artifacts are probably the easiest to identify. They have an anterior distribution, and their time courses are largely flat with occasional very high-amplitude spikes, corresponding to the artifacts of the eye muscles as they close and open (figure 8.1). Other components might identify EMG or line noise.

Keep in mind that the decomposition of the data is based purely on statistical properties; an independent components analysis cannot determine which properties of the data to consider "signal" and which properties to consider "noise." In practice, components are likely to contain both signal and noise. This can make it difficult to know which components to remove from the data. In general, if the component time course shows a task-related ERP-looking deflection, it may contain signal. However, any non-phase-locked signal would also not be apparent in the ERP, so the absence of an ERP is not proof that the component contains no signal. You should be cautious about removing components that seem to contain signal. In general it is good to take a conservative approach and remove components from the data only if you are convinced that those components contain artifacts or noise and no or very little signal. In the best-case scenario you would remove only one component corresponding to blink artifacts.

The maximum number of components that can be isolated in the EEG data is the number of electrodes you have. If you have many electrodes (more than 100), it might be useful to

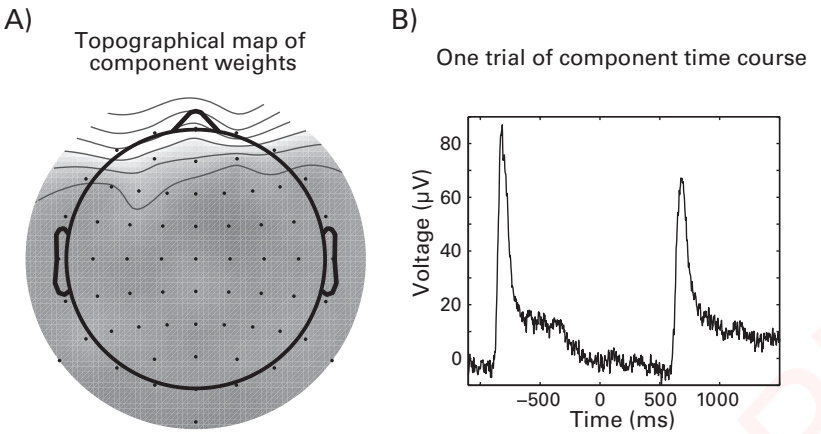

Figure 8.1
Example topographical map and example single-trial time course of an independent component that isolated a blink artifact. Panel A shows that the weights from this component are maximal at anterior electrodes, and panel B shows the time course of this component from one trial. You can see that on this trial, the subject blinked before and after the trial. The time course is a weighted sum of the activity of all electrodes, and the weights are defined by the results of the independent components analysis.

extract a smaller number of components than there are electrodes, in part because it will speed the analysis and in part because there are unlikely to be over 100 independent sources in the brain that are active and that can be statistically isolated at the level of scalp EEG.

Many EEG analysis programs offer utilities for performing independent components analysis. The Matlab toolbox eeglab is the toolbox that provides the most active development of independent components analysis.

8.2 Removing Trials because of Blinks

Should you reject trials that contain blinks? Blinks clearly introduce artifacts in the EEG data, but there are arguments for and against removing these trials.

Blink artifacts do not destroy the brain-generated EEG signal but, rather, linearly sum on top of the brain-generated EEG. There are several methods that successfully attenuate the oculomotor artifacts while sparing brain activity. The two most commonly used methods are independent components analysis (Jung et al. 2000) and regression-based techniques (Gratton, Coles, and Donchin 1983). Independent components analysis seems to work better for removing blinks and other oculomotor artifacts (Hoffmann and Falkenstein 2008; Plochl,

Ossandon, and Konig 2012). The success of these algorithms at removing oculomotor artifacts suggests that trials containing blinks should not be rejected.

However, there might be situations in which blink artifacts can be successfully removed but the trial should be rejected anyway. For example, relatively long blinks can be removed by independent components analysis, but if the subject closed his or her eyes for several hundred milliseconds, she or he may have been too tired on that trial to be focused on the task. Another example is tasks that involve briefly presented visual stimuli. A poorly timed blink might mean that the subject did not see the stimulus. Thus, depending on your experiment, trials with blinks could be rejected based on specific criteria such as the duration of the blink or the proximity of the blink to the visual stimulus presentation.

Some researchers specifically instruct subjects to inhibit blinking during the trial and to blink only during specified times or during the intertrial interval. This approach eliminates the need for correcting the data or rejecting trials, but it may introduce other problems. First, active suppression of blinking is a demanding task that relies on cortical oculomotor networks (Berman et al. 2012). Inhibiting blinks may thus introduce task-unrelated but stimulus-locked activity in frontoparietal oculomotor circuits. Second, inhibiting blinks can be cognitively demanding and distracting, particularly with long trials or a long experiment. Subjects may become so preoccupied with inhibiting their blinks that they sacrifice attention to the task. Third, if subjects are permitted to blink during the intertrial interval, then this time period might not be suitable as a baseline for normalization of time-frequency dynamics.

If you allow subjects to blink whenever they need to, they should avoid time-locking their blinks to experiment events such as button presses or stimulus offsets to facilitate a clean statistical isolation of blink artifacts from the experiment events to which those blinks are time-locked. For example, if subjects always blink whenever they press a button, it will be difficult for an independent components analysis to isolate the blink activity from the response-related activity (because these will not be independent components).

8.3 Removing Trials because of Oculomotor Activity

Blinks are not the only source of artifacts introduced by the eyes. There are also saccades and microsaccades that can contaminate EEG data, particularly at frontal and lateral frontal electrodes (or at posterior electrodes if the reference electrode is on the face). These artifacts can be minimized through experiment design by having visual stimuli at a central location on the experiment monitor, thus minimizing the need for subjects to look around. Having an easy-to-see fixation spot on the monitor at all times will also help subjects prevent eye

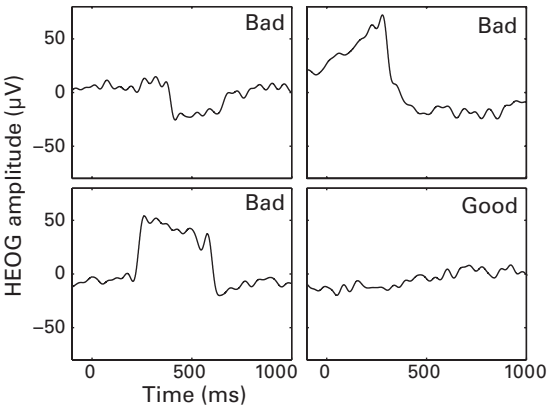

Figure 8.2
Horizontal EOG activity indicates eye movements after stimulus onset (time = 0). The first three panels are taken from trials that were removed prior to analyses.

movements. Finally, instructing subjects that small eye movements cause artifacts in the data will help minimize the frequency of saccades.

Having an eye tracker is ideal for trial rejection based on eye movements. If you do not have an eye tracker, you can use horizontal and vertical electrooculogram (EOG) electrodes. These electrodes are good at detecting relatively large eye movements, although they may lack the sensitivity of eye trackers to detect small eye movements or microsaccades. Figure 8.2 shows examples of horizontal EOG electrode activity. The panels marked "Bad" show trials that were removed prior to analyses.

Microsaccades can be more difficult to detect without an eye tracker because the small eye movements may be indistinguishable from noise in all but the cleanest EOG data. The potential impact of microsaccades on high-frequency EEG activity is a recent topic of discussion (Fries, Scheeringa, and Oostenveld 2008; Hassler, Barreto, and Gruber 2011; Yuval-Greenberg and Deouell 2009; Yuval-Greenberg et al. 2008). Microsaccades can be minimized by having small stimuli such that subjects do not need to saccade to see the entire stimulus, and by presenting the stimulus on the screen for only a short period of time (this may not be feasible in all experiments). There are algorithmic approaches for detecting and removing microsaccade artifacts from EEG data (Hassler, Barreto, and Gruber 2011; Keren, Yuval-Greenberg, and Deouell 2010; Nottage 2010).

How detrimental eye movements are to the data also depends on the referencing scheme and on your planned analyses. If you use a nose reference, eye movements might influence

the data more than if you use an earlobe reference. If your main hypotheses concern anterior frontal or lateral frontal regions, EOG artifacts are a serious concern; if your main hypotheses concern midcentral electrodes, EOG artifacts are less of a concern because they are less likely to be measured at these electrodes.

Some spatial filtering techniques can help isolate potential EOG artifacts. The surface Laplacian, for example, will help prevent the spread of oculomotor artifacts to activity at other electrodes. This is discussed further in chapter 22.

You might argue that artifacts from horizontal and vertical eye movements can, like blinks, be identified and removed with independent components analysis. Although this is likely to be the case if subjects produce enough eye movements, there is a larger concern than the oculomotor artifact itself: if the subjects are supposed to be fixating throughout the experiment, trials in which subjects broke fixation indicate that the subjects were not fully engaged in the task on that trial.

8.4 Removing Trials Based on EMG in EEG Channels

Trials with excessive EMG activity in the EEG channels should be removed. EMG is noticeable as bursts of 20- to 40-Hz activity, often has relatively large amplitude, and is typically maximal in electrodes around the face, neck, and ears. EMG bursts are deleterious for EEG data if you plan on analyzing activity above 15 Hz. EMG bursts also indicate that the subject moved, sneezed, coughed, or giggled during that trial. Thus, even if you will not examine frequencies above around 15 Hz, you might still want to remove trials with large EMG bursts because the subject may have been engaged in activities other than the task during that trial. A burst of EMG activity during a trial is shown in figure 8.3 (this trial was removed prior to analyses).

Some subjects show low-amplitude EMG activity continuously throughout the experiment, even after you instruct them to relax their face, neck, and shoulders. What to do with these datasets depends in part on your experiment. If you have specific hypotheses about beta-band activity, you might be unwilling to tolerate a dataset with continuous EMG and therefore will exclude this dataset from group analyses. Sometimes the EMG activity is well localized to lateral temporal or anterior electrodes; if you plan on testing beta-band activity over motor electrodes that show little EMG artifact, you may be willing to keep that dataset in the group analyses.

As you will learn in chapter 18, baseline normalization of time-frequency power is based on relative changes in power before versus after trial onset. This means that if the EMG activity has constant amplitude before and after trial onset, that EMG activity will be removed

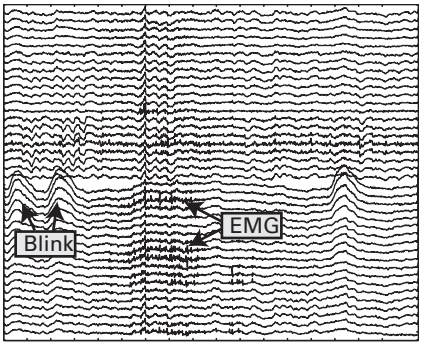

Figure 8.3
One trial of EEG data showing EMG activity on several electrodes. EMG activity is recognizable as brief bursts of high-frequency activity. This snippet of EEG data also shows blink artifacts.

during baseline normalization. Furthermore, if the continuous EMG activity is present in all conditions, it will subtract out during condition comparisons.

Sometimes continuous EMG activity can be well isolated with independent components analysis. If you plan on subtracting an EMG-containing component from your data, make sure the component does not also contain a brain-generated signal.

Note that if you create long epochs to have a buffer zone for edge artifacts (as discussed in section 7.3), it is not necessary to reject trials based on artifacts present in the buffer zone because these data will be discarded after time-frequency decomposition.

8.5 Removing Trials Based on Task Performance

So far, this chapter has discussed artifacts that contaminate the EEG signal or that are visually recognizable in the EEG data. There are other artifacts that might have little or no visually recognizable correlate in the EEG data but that may nonetheless have negative consequences for the results. These artifacts can be called *cognitive noise* or *cognitive artifacts*, and you should consider removing these trials from your analyses.

If your task involves responses that can be accurate or not, you will likely want to remove or at least separate error trials. You might also want to remove posterror trials if you are concerned about posterror behavior changes that might influence task performance and associated brain activity. You should also consider removing trials in which subjects do not make a response if they were instructed to do so, trials with more responses than were required, trials with very fast reaction times (for a finger button press, less than 200 ms), or trials with very

slow reaction times (for example, reactions times that are slower than three standard deviations from each subject's median reaction time). In these trials it is likely that the subject was not fully engaged in the task.

Depending on your task there might be additional justifications for removing trials that are suspected to contain cognitive noise. For example, if you have rest breaks that last several tens of seconds, subjects might not be fully reengaged in the task on the first trial after the break. Another example is if subjects perform a few tens of trials with one set of instructions and then switch to another set of instructions; the first trial after each switch will involve a cognitive set shift and a switch cost. These trials generally have longer reaction times and lower accuracy rates. If you are not studying these switch costs, you might want to exclude switch trials from the data.

8.6 Removing Trials Based on Response Hand EMG

If your experiment and lab setup allow, consider recording EMG from the muscles subjects use to indicate the response (e.g., from the fingers used to press buttons). This will allow you to identify partial error trials in your data. Partial errors occur when the subject twitches the muscle of the incorrect response, although he or she pressed the correct button. Correct trials that contain partial errors elicit patterns of brain activity that look more like errors than they do like correct responses (Cohen and van Gaal 2012; Coles et al. 1985; Coles, Scheffers, and Fournier 1995). In other words, correct trials containing partial errors contaminate "pure" correct trials. Partial errors can also be identified using force grip response devices. Figure 8.4 shows an example of a partial-error EMG response.

Here is the method we use to identify partial errors (Cohen and van Gaal 2012); this algorithm may need adjustment depending on the type and quality of your data. First, the *Z*-transform of the derivative of the EMG signal from each hand is taken and then rectified (that is, taking the absolute value); this eliminates hand- and subject-specific differences in impedance and signal amplitude. A partial error is identified when this *Z*-derivative signal of the hand not used to make the response exceeds two standard deviations (that is, a *Z*-score of 2) in the time between stimulus onset and the actual button press. The magnitude of this EMG peak must be more than two times larger than the largest EMG peak from –300 ms to stimulus onset (this eliminates trials in which noisy EMG produces apparent partial errors). Regardless of how you identify partial errors, you should visually inspect trials that are and are not identified as partial errors to confirm that the algorithm works well for your data.

Recording EMG from the thumb muscle tends to be easier than EMG from the other fingers. This is partly because the thumb muscles are bigger and therefore easier to find. You can

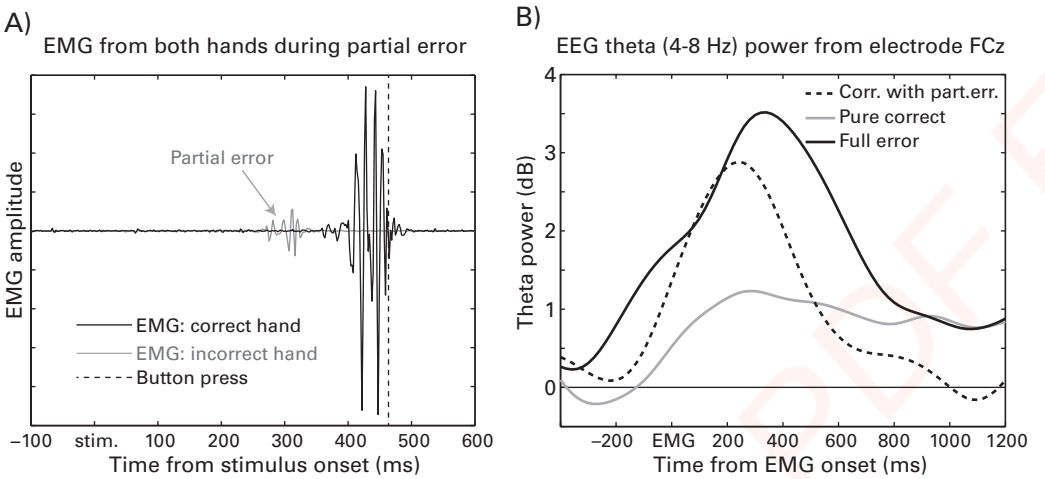

Figure 8.4
Partial errors can be detected with EMG recordings and are useful to identify correct trials with error-like brain responses. Panel A shows example EMGs from right and left thumbs showing a partial error—a muscle twitch of the incorrect hand although only the correct button was pressed in this trial. Correct trials that contain partial errors elicit error-like brain processes, as can be seen in panel B. Theta-band power from electrode FCz is time-locked to EMG onset; each line is the average of all trials from one subject. Partial errors can be identified and removed from the dataset or separately analyzed.

also improve the quality of the EMG data by having subjects use response buttons that take some effort to press. This will require more muscle engagement, which will thus produce bigger and cleaner EMG responses. In contrast, a mouse button requires little physical exertion to press and will produce a small EMG activation.

8.7 Train Subjects to Minimize Artifacts

It is safe to assume that your human volunteer subjects want to provide you clean, high-quality data, particularly if you stress the importance of their data to the overall research. If the subjects are causing artifacts in the EEG data, most likely they are not aware that they are causing artifacts. Many EEG artifacts can therefore be minimized with proper training. After setting up the EEG cap, show the subject her EEG data in real time on a computer monitor that she can see. Explain that EEG data contain both brain activity and noise from muscles. You can have her blink, clench her jaw, tense her neck/shoulder muscles, talk, smile, wiggle her ears, and so on. When subjects know what kinds of behaviors produce EEG artifacts, they can minimize those behaviors during the task.

8.8 Minimize Artifacts during Data Collection

As mentioned in the previous chapter, there is no preprocessing or analysis trick that will turn low-quality, noisy data into beautiful results. The best way to get your preprocessing protocol to turn good data into very good data is to start with good data. Training subjects before the task is part of ensuring that you will have good data.

While running the experiment, keep an eye on the real-time EEG data as they are being acquired. At least once every 30 s you should quickly glance at the EEG data and check that they look OK. If you notice artifacts in the data, first check whether there is a rest period (large artifacts during rest breaks are usually OK because these periods are typically not analyzed). If the artifacts are present during the task, consider pausing the experiment at the next possible opportunity to determine what is causing the artifacts. If the artifacts are coming from the subject being tired or uncomfortable, turning the lights on and talking to the subject for a minute before resuming the task might help to reengage the subject's attention and reduce the source of the artifacts.