

9 Quantifying ERP Amplitudes and Latencies

Overview

This chapter describes methods for quantifying the amplitudes and latencies of ERP components. These are the values that go into your statistical analyses, so it is obviously important that they be both accurate and reliable. There are many very sophisticated approaches that can be used for quantifying ERP components (involving, for example, principal component analysis or dipole source analysis), but here I concentrate on approaches that are relatively simple and robust.

Some of the techniques I describe here are not available in all ERP analysis packages. In fact, this was one of my motivations for putting together my own package, ERPLAB Toolbox (which is freely available at <http://erpinfo.org/erplab>). These techniques are not complicated to implement, so if you don't want to switch to ERPLAB Toolbox, you should pressure the manufacturer of your analysis software to implement them. Although simple, they are often dramatically superior to the more widely available methods, and they are used by many leading laboratories around the world.

Note that all of the approaches described here require defining a *measurement window* during which the amplitude or latency is measured. Selecting an appropriate measurement window is one of the most difficult aspects of amplitude and latency measurement. If you choose the window on the basis of the time period that shows the largest differences between groups or conditions in your data set, you are biasing yourself to find a significant effect even if there is no true difference (because noise may influence the measurement window). Because the method used for choosing the measurement window has implications for statistical significance, I have postponed my discussion of this issue until the next chapter, which focuses on statistical analyses.

Before I get started, however, I want to give you a piece of very important advice: Make sure that you look at the measurements for every ERP waveform that is being measured. That is, don't just run some script that measures the values and then start doing statistical analyses. Compare each measured amplitude or latency value with a plot of the corresponding ERP waveform in every subject. The waveforms often differ markedly among subjects, and a measurement procedure that looks like it will work for the grand averages may fail miserably when applied

to individual subjects. We spent a lot of time developing a visualization tool in ERPLAB Toolbox that allows you to plot each subject's waveforms and see exactly how the amplitude and latency values for those waveforms were computed, and other ERP analysis packages have similar tools. I strongly encourage you to use them!

Basic Measurement Algorithms: Peak Amplitude, Peak Latency, and Mean Amplitude

As illustrated in figure 9.1, the oldest method for measuring ERP amplitudes and latencies is to define a time window and find the maximum point in that time window (either the most positive or most negative point, depending on whether you are searching for a positive peak or a negative peak). The voltage at this point is called the *peak amplitude*, and the time of this point is called the *peak latency*. For example, the P3 wave in figure 9.1A reaches a peak voltage of approximately 17 μ V at 404 ms after stimulus onset. As was described in chapter 8, amplitude is typically measured relative to the average voltage in the prestimulus period. Make sure you don't forget that any noise or systematic distortion in the baseline will have a big impact on your amplitude measurements (for peak amplitude or other measures).

The term *peak* can be ambiguous. For example, to measure the peak of the P2 wave in the data shown in figure 9.1, you might use a time window of 150–300 ms. The maximum voltage in this time window occurs at the edge of the time window (300 ms) due to the onset of the P3

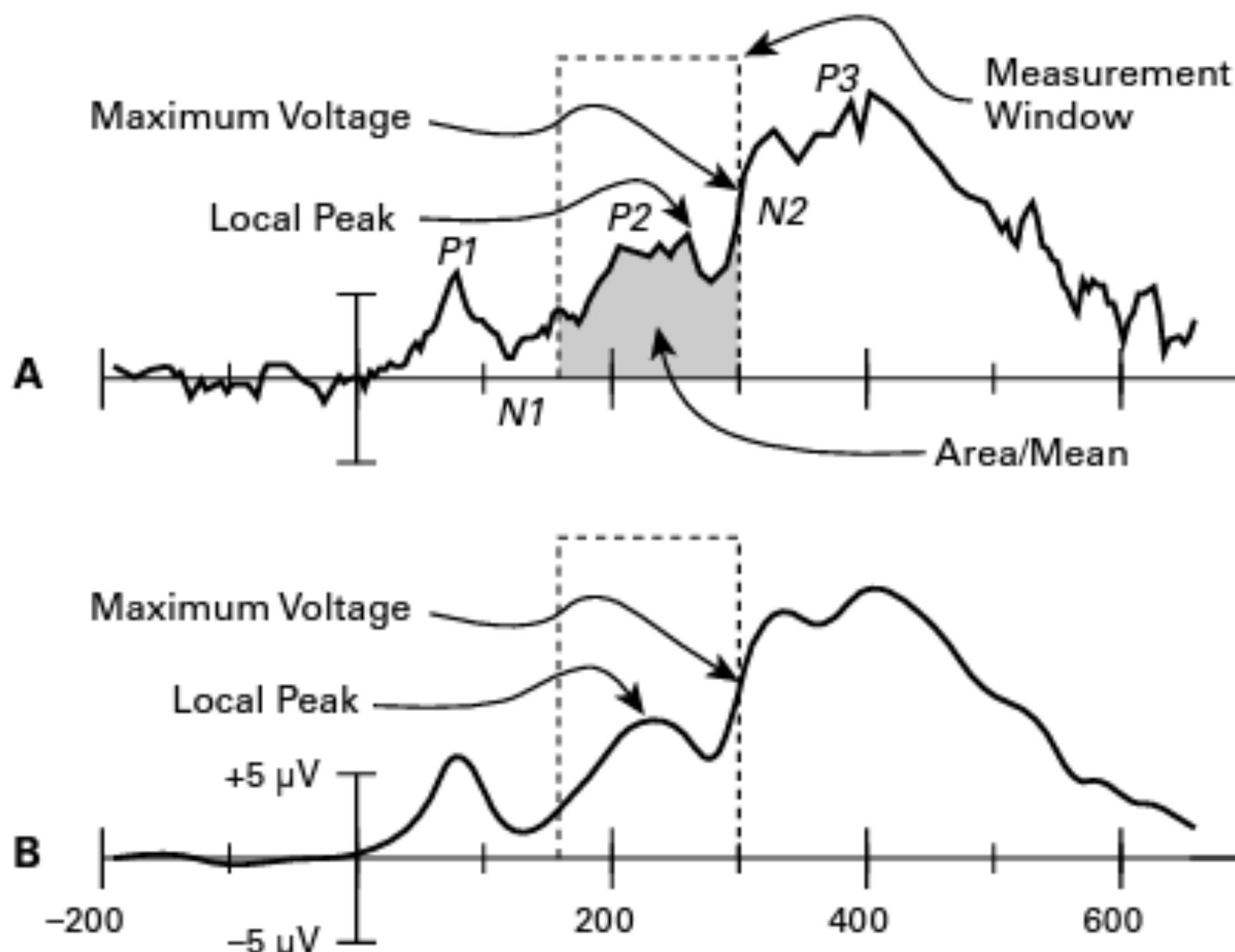


Figure 9.1

(A) Illustration of how the P2 wave can be measured, showing the measurement window (150–300 ms poststimulus), the maximum voltage (i.e., the simple peak), the local peak, and the region used to calculate area or mean amplitude. (B) The same waveform after a low-pass filter was applied to attenuate high-frequency noise.

wave. Consequently, the amplitude of the P2 wave would be measured at 300 ms, which is not very near the actual P2 peak. Clearly, this is not a good way to measure the P2 wave. This problem could be avoided by using a narrower measurement window, but a fairly wide window is usually necessary because of variations in peak latency across electrode sites, experimental conditions, and subjects. When we measure the true maximum value in this manner, this is called a *simple peak amplitude* or *simple peak latency*.

When I encountered this problem in graduate school, I developed a better measure of peak amplitude called *local peak amplitude*, which defines a local peak as the largest point in the measurement window that is surrounded on both sides by lower voltages. If we used this to quantify the P2 in figure 9.1, it would find the amplitude of the point labeled “Local Peak,” which would reflect the P2 much better than the point labeled “Maximum Voltage.” You can also measure *local peak latency*, which is simply the time of the local peak amplitude.¹

Most commercial ERP analysis packages do not implement the local peak algorithm, but it is available in ERPLAB Toolbox and in some commercial packages (e.g., BrainVision Analyzer). When this algorithm is unavailable, people sometimes visually inspect the measurements and manually adjust them according to some informal rule. I recommend against this kind of manual approach because it may lead to bias and it is difficult to describe in a way that can easily be replicated by other researchers. Instead, you should use a package that implements the local peak algorithm or harass the manufacturer of your package so that they implement this simple but useful algorithm. In any case, you should be sure to specify exactly how peaks were defined when you publish your research. For example, you might write, “We measured local peak amplitude (as defined by Luck, 2014) between 150 and 300 ms” (which would make me happy because it will increase my citation count).

Mean amplitude is a common (and usually superior) alternative to peak amplitude. Mean amplitude is computed by simply taking the average voltage over a specified measurement window. That is, you take the voltage at each sample point in the time window and then compute the average of these voltages.

Strengths and Weaknesses of Peak and Mean Measures

In this section, I describe the relative advantages and disadvantages of peak and mean amplitude measures. In the vast majority of cases, mean amplitude is superior to peak amplitude, and it’s important for you to know when and why this is true so that you can use the most appropriate measure in your own research.

Issue 1: Peaks and Components Are Not the Same Thing

The goal of any ERP quantification approach is to provide an accurate measurement of the size and timing of the underlying ERP component with minimal distortion from noise and from other spatially and temporally overlapping components. As discussed in chapter 2, measurements of amplitude and latency from a raw ERP waveform can easily be distorted by overlapping

components. That chapter also described several strategies for avoiding this problem (e.g., using difference waves), but those strategies often rely on the use of an appropriate measurement approach. Chapter 2 also noted that the most important rule for interpreting ERP waveforms is this: *Peaks and components are not the same thing. There is nothing special about the point at which the voltage reaches a local maximum.*

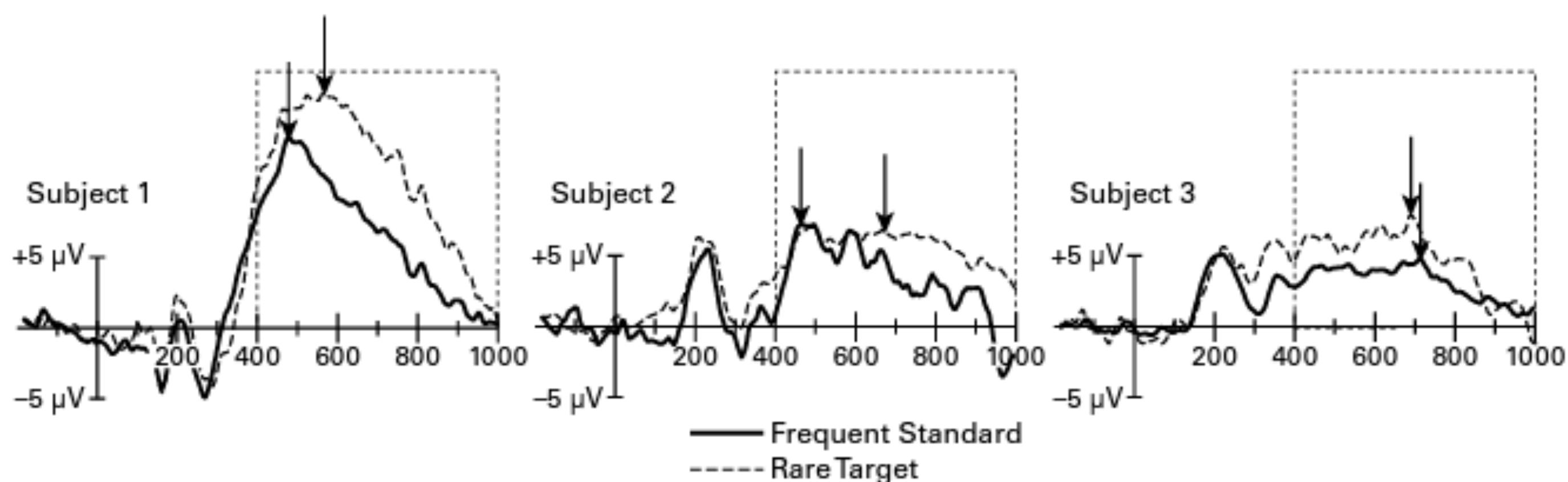
Given that the point at which the voltage reaches a local maximum is not special, why should peak amplitude be used to measure the amplitude or latency of an ERP component? The main reason is largely historical. General-purpose computers were not yet available in the early days of ERP research, and the easiest way to assess amplitudes was to plot the waveforms and use a ruler to measure the peaks (see Donchin & Heffley, 1978). Once peak measurements became standard, they continued to be used even after researchers had access to computers that could perform more sophisticated measurements. Mean amplitude does a better job of treating an ERP component as something that is extended over time, and there has been a gradual trend away from peak amplitude and toward mean amplitude over the past few decades. However, peaks are occasionally useful to measure. The rest of this section will detail the relative advantages and disadvantages of peak versus mean measures.

Issue 2: Sensitivity to High-Frequency Noise

Peak measures are easily distorted by high-frequency noise. In figure 9.1A, for example, the local peak is not in the center of the P2 wave, but is shifted to the right because of a noise “blip.” The effects of high-frequency noise can be mitigated somewhat by filtering out the high frequencies before measuring the peak, as illustrated in figure 9.1B. The local peak in the filtered waveform is closer to the “true” peak. However, this is still the peak in the overall waveform and may not be a particularly good measure of the amplitude of the underlying component.

Another common approach to dealing with high-frequency noise is to find the peak and then measure the average voltage over a window centered on that peak (e.g., a 50-ms window). This does filter out the high-frequency noise after the peak is found, but noise may still lead to the wrong point being chosen as the peak. For example, noise in the waveform causes the local peak for the P2 in the waveform shown in figure 9.1 to be shifted later in time than the true peak; if we found this local peak and then computed the average voltage over a 50-ms period centered on this peak, we would be getting a measure of a 50-ms period that is shifted to the right of the actual peak and is influenced by the falling edge of the N2 wave. I haven’t seen any studies that have directly examined the accuracy of this measure, but I suspect you will get better results by simply applying a low-pass filter first (as in the waveform shown in figure 9.1B) and then measuring the amplitude of this peak.

Mean amplitude measures tend to be less sensitive to high-frequency noise than are peak amplitude measures. This is because all the little “uppies” and “downies” within the measurement window cancel each other out. For example, even though the data in figure 9.2 have been low-pass filtered, there is still quite a bit of noise that can influence the peak (especially in subject 3), but this noise has almost no effect on mean amplitude as long as the measurement

**Figure 9.2**

ERP waveforms elicited by frequent standards and rare targets, recorded from three individual subjects in an oddball paradigm. The dashed region shows a measurement window that could be used for measuring the P3 wave, and the arrows show the peaks in this window.

window is wide enough (e.g., >50 ms). In fact, there is no advantage to applying a low-pass filter prior to measuring mean amplitude, and box 9.1 explains why it is conceptually purer to measure mean amplitude without first applying a low-pass filter.

Issue 3: Measuring the Same Process at Different Times

Even when the high frequencies are filtered out, noise may still have a large impact on peak measures. As an example, consider the three individual subjects shown in figure 9.2, who participated in an oddball paradigm. The waveforms were low-pass filtered with a half-amplitude cutoff of 30 Hz to eliminate high-frequency noise, and then P3 amplitude was measured as the local peak between 400 and 1000 ms. Although subject 1 had a large and sharp P3 deflection, making the peak a reasonable estimate of the amplitude, subject 2 and subject 3 had broader P3 deflections without a very clear peak. Whereas the peak for the standard stimulus was a little before 500 ms in subject 1, the peak for this same stimulus was at approximately 700 ms for subject 3. For a simple oddball task, it is implausible that two neurologically normal people differ by more than 200 ms in the timing of the process that generates the P3 wave. And what would it mean to compare the voltage at 500 ms in one subject with the voltage at 700 ms in another subject? These huge differences in peak latency across subjects make it a little strange to use the P3 peak as a reflection of differences in a single underlying component.

A similar issue arises if we use peak amplitude to compare different conditions. For example, if we compare the amplitude of the P3 peak for the target and standard stimuli in subject 2, we will be comparing a voltage at 475 ms in one condition with a voltage at 675 ms in another condition. It seems unlikely that we are really measuring the same process at these two very different times in this simple oddball task.

Another related issue arises when peak amplitude is used to measure the amplitude of a component at multiple electrode sites. Because voltages propagate instantaneously from the brain to

Box 9.1

Filtering and Mean Amplitude

If you carefully examine the methods sections of papers published by my lab, you will see that we often measure mean amplitudes from data without any low-pass filtering (except for the very mild hardware anti-aliasing filter). However, we low-pass filter the data for figures showing ERP waveforms. This reflects our attempt at being very precise about timing. After all, one of the main virtues of the ERP technique is its high temporal resolution, so it makes sense to be precise about timing.

Why does our approach give us greater precision? As described in chapter 7, low-pass filters cause a temporal spread in the ERP waveforms (see especially figure 7.9B). If our method section said that we measured the mean amplitude between 150 and 250 ms, but we had applied a low-pass filter to the EEG or ERP prior to measurement, our measurements would be influenced by brain activity that happened prior to 150 ms and after 250 ms. Thus, the measurement window described in our methods section would not be a very precise way of describing the timing of our measurements if we filtered prior to measuring the data. Moreover, because high-frequency noise is canceled out in measurements of mean amplitude (assuming the measurement window is wide enough), filtering out the high frequencies prior to measuring mean amplitude doesn't help anything.

In reality, the filters we use are mild enough that we would get almost exactly the same results if we filtered prior to measuring mean amplitude. However, it seems worthwhile to be perfectly clear about the timing of the data that contribute to the analyses, just for the sake of conceptual purity.

However, we do apply a low-pass filter before plotting the data in our figures. This simply makes it easier for the reader to see the differences between the waveforms without being distracted by noise. You will see that we explicitly state that the waveforms have been filtered in our figure captions. Thus, we are being both precise and clear about what we are doing. I would encourage you to take this same approach. If nothing else, it will show the world that you understand what you are doing with your data processing and analysis procedures (although it might also indicate that you are a little bit obsessive-compulsive, just like I am).

all of the electrodes, brain activity from a given underlying component cannot have different latencies at different electrode sites. However, the latency of a peak may vary greatly across electrode sites because of differences in the amplitudes of overlapping components. Thus, if you compare peak amplitudes at different electrode sites, you will typically be measuring the amplitude at different times for the different sites, which makes no sense. One solution to this problem is to find the peak at one electrode site (typically the site where the peak is largest) and then measure the voltage at this time point at the other electrode sites. For example, you might find the P3 peak at the Pz electrode site and then measure the voltage at that time point at all electrode sites. This is a fine solution, as long as you have good reason to believe that the amplitude of the peak is an appropriate measure (which is not typically true).

All of these problems are a result of the fact that peak amplitude does not ordinarily use the same time point to measure a given component across subjects, across conditions, or across electrode sites. In contrast, the mean amplitude in a given time window measures—by definition—the voltages at the same time points in all subjects, in all conditions, and in all channels.

There are occasionally situations in which it would not be appropriate to use the same time window for all subjects or for all conditions. For example, if a component occurs later in one group than in another, it may be problematic to compare mean amplitude over the same time window in both groups. Instead, you may want to find the peak amplitude over a latency window that is broad enough to include all subjects in both groups. However, this sort of situation is usually very tricky even with peak amplitude. For example, if the latencies differ across groups, the latency variability probably also varies, and differences in latency variability invalidate the use of peak amplitude measures (as described earlier in the chapter). In addition, if other components overlap the component you are trying to measure, these components are likely to differ in amplitude between the two measurement windows you are using for the two groups, and this could distort your amplitude measurements. If you are in this situation, you should try to use difference waves to isolate the component of interest (as discussed in chapter 2), and you should consider using the *signed area* measures described later in this chapter (which are insensitive to differences in latency variability).

Issue 4: Biased Versus Unbiased Measures

Another problem with peak amplitude is that it is biased by the noise level (as was mentioned previously in chapter 4). The noisier the data, the larger the maximum amplitude will tend to be. Thus, it is not legitimate to use peak amplitude when comparing waveforms that are based on different numbers of trials or when comparing groups with different noise levels (which is common when comparing patients versus controls or younger versus older subjects).

In contrast, mean amplitude is not biased by the noise level. That is, mean amplitude does not become systematically larger as the number of trials decreases or as some other source of noise increases. Increasing the noise makes mean amplitude measures more variable, decreasing your statistical power, but it does not bias them to become larger. In contrast, peak amplitude becomes both more variable and systematically larger as the noise level increases (all else being equal). Consequently, it is not usually legitimate to compare peak amplitude values across groups or conditions with different noise levels, but it is perfectly fine to compare mean amplitude measurements from waveforms based on different numbers of trials.

Many people have difficulty understanding this distinction between bias and variability, and I see a lot of inappropriate analyses of experiments in which the number of trials differs across groups or conditions. I have therefore provided an online supplement to this chapter that provides a more complete description of what happens to peak and mean amplitude when the number of trials varies across groups or conditions (and what you should do in this situation).

Issue 5: Linear Versus Nonlinear Measures

Some ERP processing operations are linear and others are nonlinear. The measurement of mean amplitude is a linear operation, whereas the measurement of peak amplitude (or peak latency) is nonlinear. You can apply linear processes such as filtering, re-referencing, and averaging in any order, and the result will be the same. In contrast, the order of operations can make a big

difference if one of the processes is nonlinear (see the appendix of this book for further discussion, including a definition of *linear* and *nonlinear* operations).

For example, imagine that you measure the mean amplitude from 150 to 250 ms in the ERP waveforms for 15 subjects, obtaining one value for each subject, and then you compute the average of these 15 values. Now imagine that you measure the mean amplitude from 150 to 250 ms from the grand average of the 15 single-subject ERP waveforms. This measurement from the grand average will be exactly the same as the average of the 15 mean amplitudes that you obtained from the single-subject waveforms. In other words, it doesn't matter whether you measure the mean amplitude from the single subjects and then average them together or whether you average the single-subject waveforms together and then measure the mean amplitude. Because measuring mean amplitude is a linear operation, and making a grand average is a linear operation, the order doesn't matter. The same thing is true when applied to single-trial EEG waveforms and averaged ERP waveforms: Measuring the mean amplitude from 150 to 250 ms on the single-trial EEG waveforms and then averaging these measurements together will give you the same result as averaging the single-trial EEG waveforms into an averaged ERP waveform and then measuring the mean amplitude from 150 to 250 ms. This is a very convenient feature of mean amplitude.

In contrast, peak amplitude is a nonlinear measure, and the order of operations matters a great deal. If you measure the P2 wave by finding the peak amplitude between 150 and 250 ms in each of 15 single-subject ERP waveforms, the average of these peak amplitudes will not be the same as the peak amplitude from a grand average of the 15 single-subject waveforms. This may cause a discrepancy between the grand-average waveforms that you present in your figures and the averaged peak amplitude values that you analyze statistically. Box 9.2 describes how I first noticed this issue as a graduate student. The same thing is true when applied to single-trial EEG waveforms and averaged ERP waveforms: Measuring the peak amplitude from 150 to 250 ms on the single-trial EEG waveforms and then averaging these measurements together will give you a very different result compared to averaging the single-trial EEG waveforms into an averaged ERP waveform and then measuring the peak amplitude from 150 to 250 ms.

Issue 6: Latency Jitter

Because peak amplitude is not a linear measure, peak amplitude in an averaged ERP waveform may radically misrepresent the amplitudes on the single trials. We encountered a special case of this when we discussed latency jitter in chapter 8. As you will recall, when the latency of a component varies from trial to trial in the raw EEG data, the peak amplitude in the averaged ERP waveform will be smaller than the single-trial amplitudes (see figure 8.7 in chapter 8; see also the extended discussion in online chapter 11). If latency variability is greater in one condition than in another, the peak amplitudes will differ between conditions even if there is no difference between conditions in the single-trial peak amplitudes. Thus, peak amplitude is not a valid measure if there are differences in latency variability across groups or conditions.

Box 9.2

Why Don't These Peaks Match?

When you are analyzing data from ERP experiments, you will typically have four, five, or even six factors in your analyses of variance. This is because you will have the same factors as in a behavioral experiment, plus one to two additional counterbalancing factors (if you implement the experimental design suggestions from chapter 4), plus one to two electrode factors (e.g., a left-to-right factor and a front-to-back factor). In most analysis of variance (ANOVA) programs, this requires a data file in which each row is a subject and each column is a single cell of your multifactor design. With four to six factors, and two to six levels of each factor, you will commonly have dozens or even hundreds of columns. To get the ANOVA program to work properly, you need to provide information about the ordering of all these columns. It's really easy to make a mistake, which will lead to completely bogus ANOVA results. Sometimes this is obvious (e.g., when you get an F value of 3432.12 for a counterbalancing factor that shouldn't have much of an effect). But sometimes it's not obvious. When I was in graduate school, I learned that I should always verify the ordering of the factors to avoid errors.

How can you verify this? The simplest way is to look at the table of means provided by the ANOVA program and compare these means with your grand average ERP waveforms. If the means match, then you almost certainly have the factors in the right order.

One day when I was a grad student, I was trying to match up the means from the ANOVA program with the grand averages, and I couldn't get them to match for my peak amplitude measures. However, they matched perfectly for the mean amplitude measures. I double-checked and triple-checked the data files, and everything was in exactly the same order for the peak and mean amplitude measures. Then I started looking at the individual subjects. The peak values for the single subjects in the ANOVA output matched the single-subject averaged ERP waveforms perfectly. Eventually I realized the problem: Measuring the peaks from the individual subjects and then taking the average is not equivalent to averaging the single-subject waveforms together and then measuring the peak from the grand average. In contrast, because mean amplitude and averaging are both linear operations, you get exactly the same result whether you measure the mean amplitudes from the single subjects and then average or measure the mean amplitude from the grand average.

Because mean amplitude is a linear measure, it doesn't matter whether you measure it from the single trials and then average the single-trial values together or whether you measure it from the averaged ERP waveform. Thus, mean amplitude is not influenced by latency jitter as long as your measurement window is sufficiently wide.

Issue 7: Comparing Peaks with Reaction Times

Another problem with peak measures is that peak latencies are difficult to relate to other measures of processing time, such as reaction time (RT). The difficulty arises because the peak of an ERP waveform is analogous to the mode of the RT distribution, not to the mean, and comparing modes and means is a little like comparing apples and oranges. This will be discussed in more detail in the section on "Comparing ERP Latencies with Reaction Times" later in the chapter.

Drawbacks of Mean Amplitude

Although mean amplitude has several advantages over peak amplitude, it is not a panacea. In particular, mean amplitude is still quite sensitive to the problem of overlapping components and can lead to spurious results if the latency of a component varies across conditions. There are some more sophisticated methods for isolating specific components (e.g., dipole source modeling, ICA, PCA, etc.), but these methods are based on a variety of difficult-to-assess assumptions and are beyond the scope of this book. Thus, I would generally recommend using mean amplitude measures in conjunction with the rules and strategies for avoiding the problem of overlapping components that were discussed in chapter 4.

Because the peaks vary in latency across subjects and across electrodes, you might think that measuring the mean amplitude over a fixed window would be problematic. However, these variations in peak latency are often inconsequential for mean amplitude measures. That is, these variations in peak latency are often driven by variations in overlap from other components, not by variations in the timing of the experimental effect. Thus, this is not usually a problem in practice.

Perhaps the most vexing problem that arises in using mean amplitude is defining the measurement window. Consider, for example, the Gedankenexperiment shown in figure 9.3, in which the N2 component is being compared between group A and group B. If we want to measure mean amplitude, what measurement window should we use? If we look at the grand averages, we see that the difference between groups is present from approximately 250 to 400 ms. We

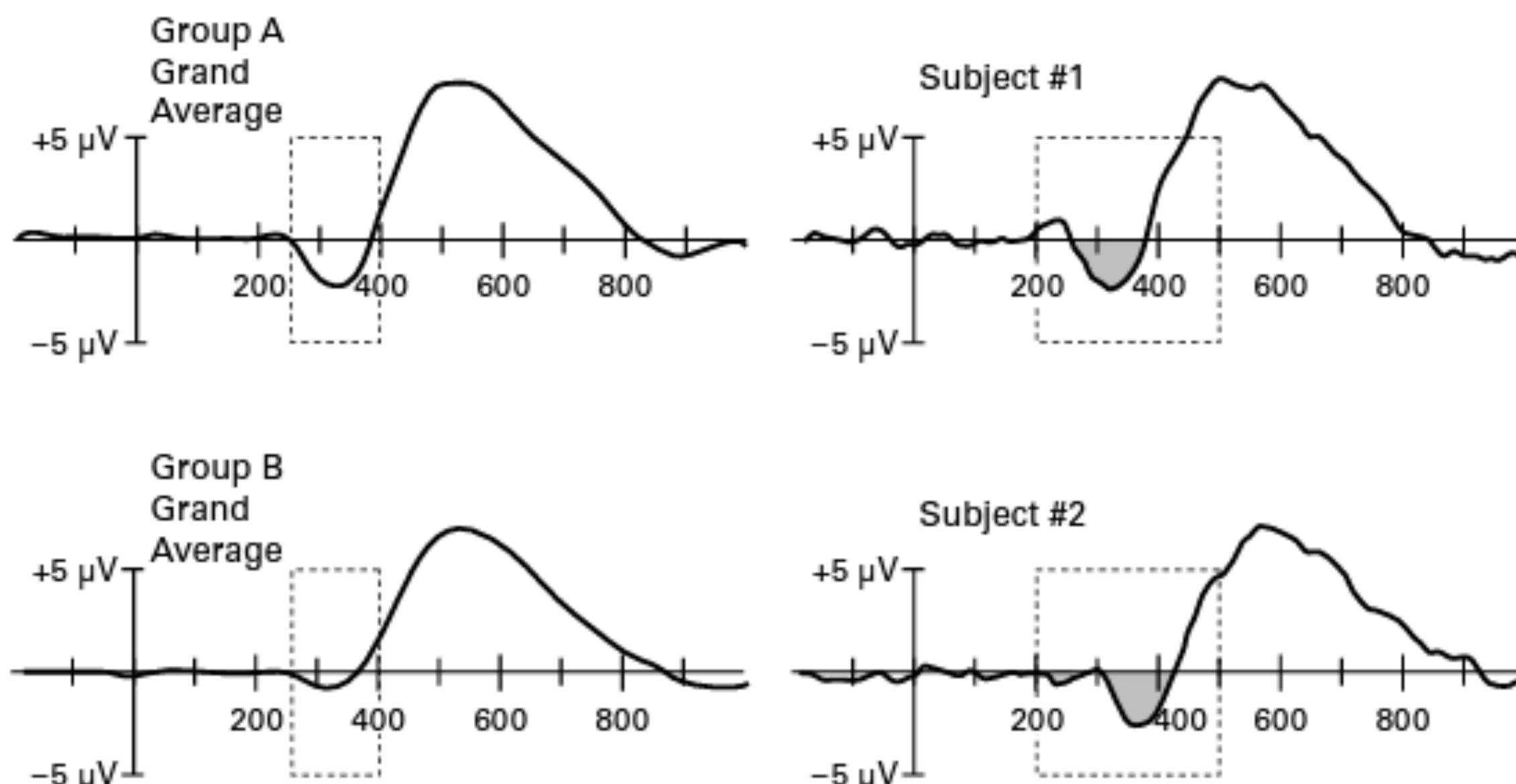


Figure 9.3
Data from a Gedankenexperiment in which the N2 component is compared between two groups of subjects, group A and group B. Grand averages are shown for the two groups, along with single-subject averages of two subjects from group A.

could use this as the measurement window, but because we are choosing the measurement on the basis of the data, we are biased to find a significant difference even if there is no true difference (this issue is described in more detail in chapter 10). A very good alternative is to use previous research to guide the selection of the measurement window. For example, if a previous study comparing these two groups found an N2 difference between 300 and 450 ms, we could use this interval as the measurement window for our new experiment. However, if the stimuli were brighter in the new experiment, the latencies would be a little shorter in the new experiment than in the previous experiment. Indeed, the P3 wave in figure 9.3 starts around 375 ms, and if we measured the mean amplitude from 300 to 450 ms, the P3 during this window would partially cancel the N2, canceling out the difference in N2 amplitude between conditions. We could measure the peak amplitude of the N2 wave, but that would have all the problems described earlier in this section.

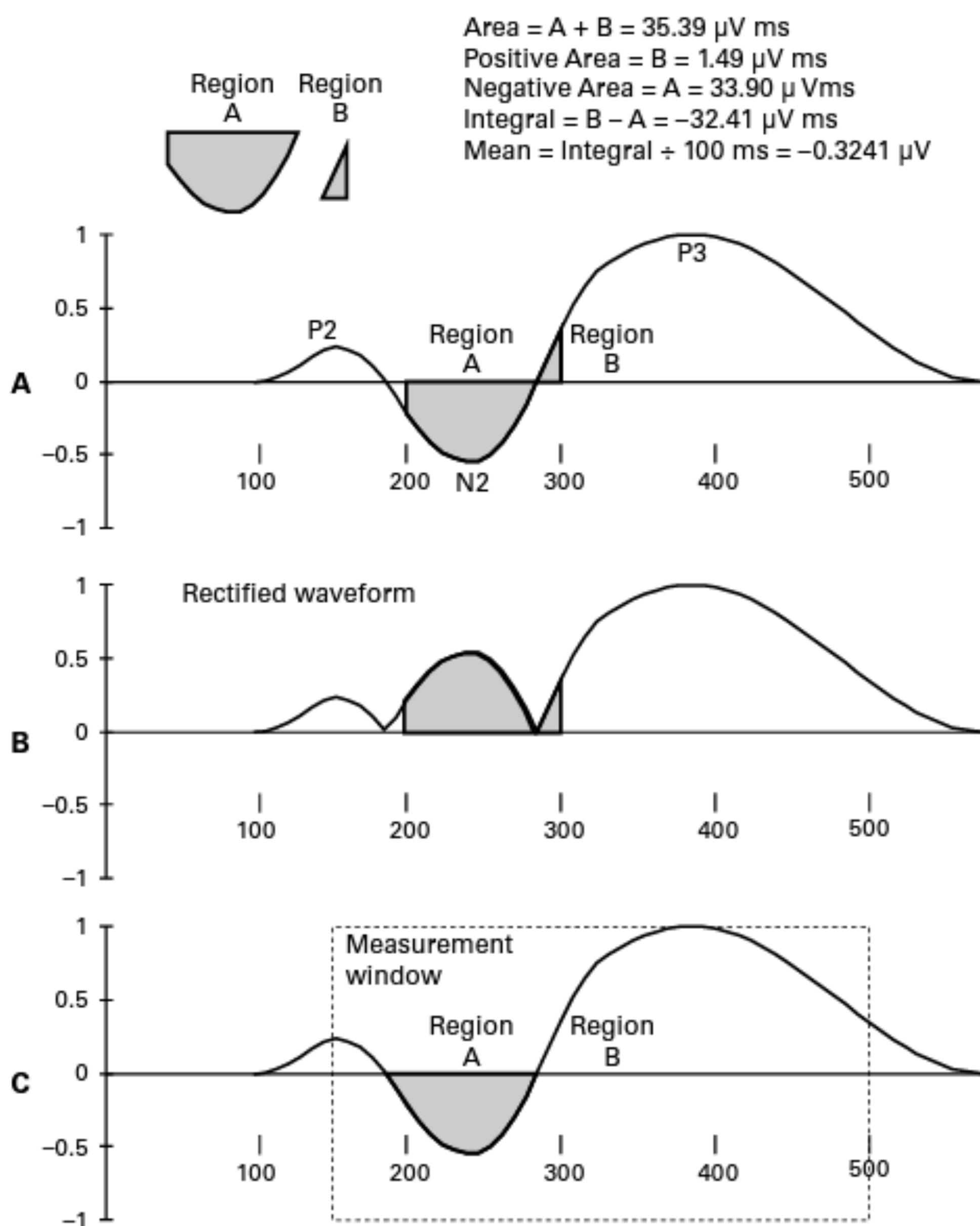
In most cases, using previous experiments to determine the latency window for measuring mean amplitude is a good idea. However, it sometimes fails (as in the hypothetical example shown in figure 9.3). This doesn't mean that mean amplitude is a bad measure or that you shouldn't use previous experiments to select your measurement windows. The point of this example is that mean amplitude is very sensitive to the choice of latency window. This is the main shortcoming of using mean amplitude. Fortunately, area measures can sometimes overcome this limitation, as I will describe in the next section.

Area Amplitude

In the first edition of this book, I wrote that area amplitude is essentially the same as mean amplitude, except that mean involves an additional step of dividing by the duration of the measurement interval. However, when Javier Lopez-Calderon and I were developing the measurement procedures in ERPLAB Toolbox, Javier convinced me that this was not quite correct. Javier then implemented several different types of area measures that turned out to be extremely useful. In this section, I will first define what area amplitude means and how it is different from mean amplitude, and then I will describe how area measures can be advantageous.

Defining Area

Figure 9.4A illustrates the correct way to use the term *area*. To compute areas, geometric regions are defined using the ERP waveform, the baseline, and the edges of the measurement window as the boundaries. In this example, we have defined a measurement window of 200–300 ms to measure the N2 wave. This creates two distinct geometric regions, labeled A and B in the figure. Area is a geometric term, and it is always positive. For example, both region A and region B have a positive area, even though region A is below the baseline and region B is above. The area from 200 to 300 ms in this example is therefore the area of region A plus the area of region B. This is very different from the mean amplitude, in which negatives and positives cancel each other.

**Figure 9.4**

Example of how the term *area* is defined for ERP waveforms. (A) The ERP waveform, baseline, and measurement window (200–300 ms) are used to define two regions, A and B. Each region is a geometric shape and therefore has a positive area, expressed in units of $\mu\text{V}\cdot\text{ms}$. (B) Rectified version of the waveform shown in panel A. Any negative voltages were multiplied by -1 to convert them into positive voltages. (C) Example of the *negative area* in the waveform shown in panel A. A much wider measurement window was possible because the positive regions are excluded when negative area is computed.

The units for area represent the multiplication of the X dimension and the Y dimension. The area of a room, for example, is expressed in square meters (m^2) if the X and Y dimensions are expressed in meters, or in square feet (ft^2) if the dimensions are expressed in feet. Similarly, the area of a region of an ERP waveform is expressed in units that multiply the X dimension (time) and the Y dimension (amplitude). If we measure amplitude in microvolts and time in milliseconds, the area would be given in units of $\mu\text{V}\cdot\text{ms}$. For example, a square region that is 1.5 μV high and 10 ms long would have an area of 15 $\mu\text{V}\cdot\text{ms}$.

When I used the term *area* in the first edition of this book, I was really thinking about the *integral*. The integral between 200 and 300 ms in figure 9.4A would subtract the area below the baseline (region A) from the area above the baseline (region B). The area of region A is 33.90 $\mu\text{V}\cdot\text{ms}$, and the area of region B is 1.49 $\mu\text{V}\cdot\text{ms}$, so the integral would be $B - A = 1.49 - 33.90 = -32.41 \mu\text{V}\cdot\text{ms}$. The mean amplitude is simply the integral divided by the duration of the interval. Thus, an integral of $-32.41 \mu\text{V}\cdot\text{ms}$ divided by a duration of 100 ms is equal to a mean amplitude of $-0.3241 \mu\text{V}$.

To clarify the distinction between the geometric area and the integral, ERPLAB Toolbox uses the term *rectified area* to refer to the geometric area. *Rectification* is just another term for taking the absolute value for each point in the waveform (multiplying each negative value by -1 to make it positive). Figure 9.4B shows what the waveform in figure 9.4A would look like if rectified. Once we rectify the waveform, all the areas are positive, so the integral is now equivalent to the geometric area. Note that rectified area is not usually used to measure ERP components, but rectification is often used in other contexts (e.g., measuring EMG activity), so it is worth knowing this term.

Once you start thinking about area, it becomes obvious that you might want to use only the area of the region below the baseline if you are measuring a negative component or only the area of the region above the baseline if you are measuring a positive component. In figure 9.4A, for example, we might want to quantify N2 amplitude with the area of the negative region (region A). We can therefore define *negative area* as the area of the region (or regions) below the baseline and *positive area* as the area of the region (or regions) above the baseline. More generally, negative area and positive area are cases of *signed area*.

If you use area or integral measures, I encourage you to adopt the specific terminology described in this section, which avoids the ambiguities that otherwise occur with the term *area*.

The Advantage of Signed Area Amplitude

As illustrated in figure 9.4C, an advantage of using negative area to measure the N2 wave is that you can use a very wide latency window without getting any cancellation from the preceding P2 wave or the subsequent P3 wave. In other words, as long as you use a fairly wide measurement window, it doesn't matter very much what measurement window you use.

This can be a huge advantage. Consider, for example, the Gedankenexperiment shown in figure 9.3. If we used the negative area measure, we wouldn't need to worry about the window being so wide that the P3 cancels the N2. We could select a fairly wide window (e.g., 200–500

ms), and it would capture the area of the N2 without any cancellation from the P2 or P3. We would also get approximately the same results with a window of 250–450 ms or 100–600 ms. This eliminates any bias that we might introduce by selecting a narrow window on the basis of the observed time course of the effect. It also gives us the one advantage of peak amplitude, which is the ability to measure something that might occur at different times in different conditions or subjects. But it eliminates several of the disadvantages of peak amplitude. For example, it is relatively insensitive to high-frequency noise. In addition, little noise blips won't cause it to use dramatically different time periods to measure different subjects. Instead, it tends to reflect the real differences across subjects in the timing of components.

The right side of figure 9.3 shows how this works with individual subjects who have somewhat different N2 latencies, using a window of 200–500 ms. Although the N2 is later in subject 2 than in subject 1, the N2 area is fully captured in both subjects. The only downside is that some negative-going noise deflections also contribute to the negative area in subject 2. This is a small price to pay given the advantages of using the negative area.

Risa Sawaki and I used this approach in a study in which a negative-going N2pc component was followed by a positive-going Pd component (Sawaki, Geng, & Luck, 2012). We had no previous experiments that we could use to determine an appropriate measurement window for the Pd component in this experiment. If we chose a narrow window on the basis of the data, this would have biased our results. If we chose a broad window, the N2pc would have partially canceled the Pd. We therefore decided to use the positive area, and it worked great.

The main disadvantage of signed area measures is that they are biased to be larger than the true value. For example, if you measure the positive area, the smallest possible value is zero, so the average across subjects will either be zero or greater than zero. In addition, noisy waveforms will tend to have larger values than clean waveforms (all else being equal). This is the same problem that arises with peak amplitude. Consequently, you cannot ordinarily use positive or negative area to compare groups or conditions with different noise levels. In addition, you can't determine if a component is significantly different from zero by just doing a one-sample t test, as you could with mean amplitude. To deal with this problem, Risa used nonparametric permutation statistics rather than conventional parametric statistics in her study (Sawaki et al., 2012).

Despite this one shortcoming, positive area and negative area have tremendous potential to minimize the problem of selecting the measurement window, and I encourage you to try this approach. But keep in mind that it is a relatively new approach, so you will want to think carefully about it.

Using Fractional Area Latency to Estimate the Midpoint Latency

The vast majority of ERP studies have quantified the timing of a component by measuring the latency of the peak (or local peak) within a given time window. Researchers don't usually provide an explicit justification for using the peak to quantify the latency, and it appears to be used mainly because of tradition (which is a result of the fact that peaks were the only thing that

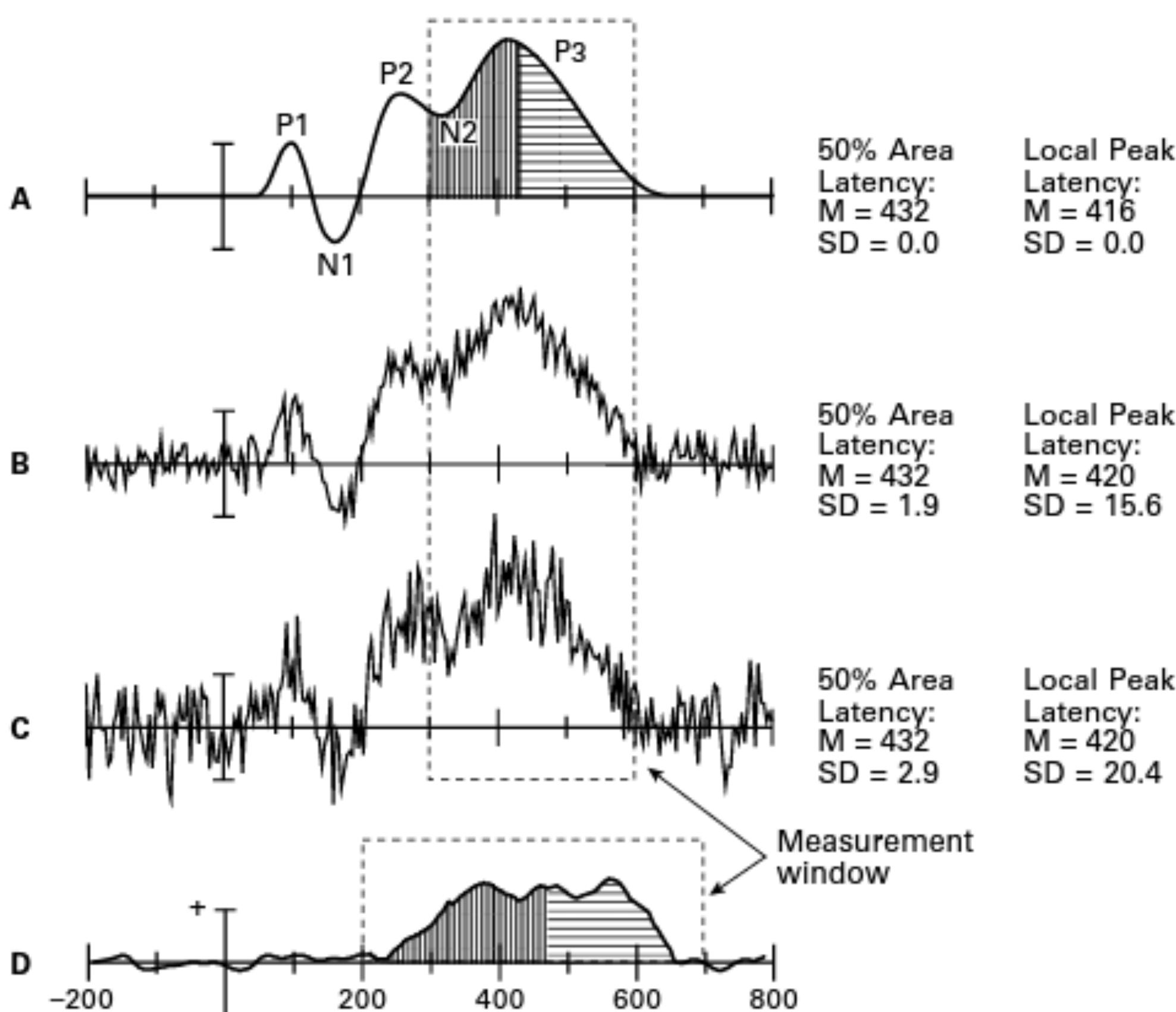
could be easily measured in the early days of ERPs, before powerful computers were available).

As is discussed in detail in online chapter 11, the shape of an averaged ERP waveform reflects the distribution of onset times of the single-trial ERP waveforms. This is why, for example, latency variability influences the shapes of the ERP waveforms. The peak of an ERP waveform is analogous to the mode of the underlying distribution of single-trial waveforms. The mode is not usually used as a measure of central tendency, and this makes it difficult to compare the timing of an ERP peak to other measures of processing time, such as RT. Moreover, the mode is typically less reliable than other measures of central tendency. In this section, I will describe a much better method for assessing the *midpoint* of a component, called *fractional area latency*. This approach involves using area amplitude, and it has many of the same advantages as using area to measure amplitude. In addition, it is much better than peak latency for making comparisons with reaction time.

The fractional area latency measure involves computing the area under the ERP waveform over a given latency range and then finding the time point that divides that area into a prespecified fraction (this approach was apparently first used by Hansen & Hillyard, 1980). Typically, the fraction will be one-half, in which case this would be called a *50% area latency* measure. An example of this is shown in figure 9.5A. The measurement window in this figure is 300–600 ms, and the area under the curve in this time window is divided at 432 ms into two regions of equal area. Thus, the 50% area latency is 432 ms. Just as amplitude can be measured using the geometric (rectified area), the positive area, the negative area, or the integral, you can find the 50% point using any of these definitions of area. Using just the positive area or just the negative area is usually the best approach, because it minimizes contributions from noise or overlapping components that do not have the same polarity as the component of interest.

The latency value that is estimated in this manner will depend quite a bit on the measurement window that is chosen. For example, if the measurement window for the waveform shown in figure 9.5A was shortened to 300–500 ms rather than 300–600 ms, an earlier 50% area latency value would have been computed. Consequently, this measure is appropriate primarily in two (fairly common) situations: (1) when the waveform is dominated by a single large component (e.g., the large P3 wave in the visual search experiment shown in figure 8.8); or (2) when a difference wave has been used to isolate a single component. It may also be useful under other conditions when combined with an automated procedure for determining the measurement window. Figure 9.5D shows an example of measuring the 50% area latency in a rare-minus-frequent difference wave that isolates the P3 wave. A wider measurement window can be used because there are no overlapping components to distort the measurement.

One advantage of the 50% area latency measure is that it is less sensitive to noise than is peak latency. To demonstrate this, I added random (Gaussian) noise to the waveform shown in figure 9.5A and then measured the 50% area latency and the local peak latency of the P3 wave. I did this 100 times for each of two noise levels, making it possible to estimate the variability of the measures. When the noise level was 0.5 μ V (figure 9.5B), the standard deviation of the

**Figure 9.5**

Application of 50% area latency and local peak latency measures to a noise-free ERP waveform (A), an ERP waveform with a moderate amount of noise (B), an ERP waveform with significant noise (C), and a rare-minus-frequent difference wave (D).

peak latency measure over the 100 measurements was 15.6 ms, whereas the standard deviation of the 50% area latency measure was only 1.9 ms. When the noise level was increased to 1.0 μ V (figure 9.5C), the standard deviation of the peak latency measure was 20.4 ms, whereas the standard deviation of the 50% area latency measure was only 2.9 ms. The variability in peak latency measures can be greatly decreased by filtering the data, and a fair test of peak latency should be done with filtered data. When the waveforms were low-pass filtered quite severely with a half-amplitude cutoff at 6 Hz, the standard deviations of the peak latency measures dropped to 3.3 ms and 6.1 ms for noise levels of 0.5 and 1.0 μ V, respectively, but this was still higher than the standard deviations observed with the 50% area latency measure without any filtering. Thus, when severe filtering was performed prior to measuring peak latency (well beyond the range I would ordinarily recommend), 50% area latency was still more reliable than local peak latency.

Kiesel et al. (2008) conducted a set of very rigorous and thoughtful simulations to test several different measures of latency. If your research involves looking for subtle latency differences, I

strongly recommend that you read this paper. Their results were consistent with the simple simulation shown in figure 9.5, showing that fractional area latency tended to be the most reliable way of measuring changes in latency across conditions or groups, leading to the best statistical power (especially when combined with the jackknife approach to statistical analysis, which will be discussed in chapter 10).

Onset Latency

Because peaks were easily measured by early ERP researchers, peak latency became the standard way of measuring timing, and this tradition continues to this day. However, theories of neural, cognitive, and affective processes do not typically focus on the time at which a process reaches its peak. Instead, they typically focus on when a process begins, when it reliably discriminates between alternative input patterns, or its duration. Moreover, as discussed in chapters 2 and 4, the onset time of a difference between two conditions provides an excellent way of assessing the amount of time required for the brain to differentiate between these conditions. Thus, it often makes much more sense to measure the onset or offset of a component (or the onset or offset of a difference wave) rather than the peak. The use of such measures has been increasing in ERP research, which is a good thing.

Challenges in Measuring Onset Latency

Unfortunately, onset latency tends to be more difficult than midpoint latency to measure accurately and reliably. For example, the onset of the P3 wave in figure 9.5A is obscured by the P2 and N2 waves. I can't even make a reasonable guess about the onset time of the P3 wave in this waveform. The only component whose onset is reasonably clear in this waveform is the P1 wave, because it is the very first component. Thus, it is somewhere between difficult and impossible to assess the onset of an ERP component unless (a) it is much larger than the surrounding components or (b) it has been isolated by means of a difference wave.

When a difference wave is used, as in figure 9.5D, the onset is much easier to estimate, at least visually. However, there is still a conceptual problem that must be solved. Specifically, the onset of a difference between conditions is the point at which the difference is infinitesimally greater than zero (or infinitesimally greater than the noise level), which means that the signal-to-noise ratio at this point is essentially zero. People have attempted to address this problem in multiple ways. One intuitively attractive approach is to estimate the slope of the onset period and extrapolate to 0 μ V. I have tried this, and it just doesn't work very well with single-subject ERP waveforms, which are highly variable. In many cases, the rising edge of the waveform is far from linear, making this approach difficult in practice.

Another approach is to find the time at which the waveform's amplitude exceeds the value expected by chance. The variation in prestimulus voltage can be used to assess the amplitude required to exceed chance, and the latency for a given waveform is the time at which this amplitude is first reached (for details, see Osman, Bashore, Coles, Donchin, & Meyer, 1992; Miller,

Patterson, & Ulrich, 1998). Unfortunately, this method is highly dependent on the noise level, which may vary considerably across subjects and conditions.

A related approach is to conduct a *t* test between two conditions or groups and find the time at which two conditions become significantly different from each other. To avoid spurious results arising from the use of a large number of *t* tests, you can find the first time point that meets two criteria: (1) the *p* value is less than 0.05, and (2) the *p* values for the subsequent *N* points are also less than 0.05 (where *N* is usually in the range 3–10). This approach doesn't really find the true onset time of the difference, but rather the point at which the difference is large enough to exceed the statistical threshold. Also, it is somewhat complicated to determine the number of consecutive significant points that should be required to lead to an overall Type I error rate of 5%. Moreover, because significance is tested for individual time points, which tend to be noisy, this approach tends not to have very good statistical power.

Although these techniques for measuring onset latency have significant disadvantages, there are two techniques that work quite well. Before I describe them, however, I would like to stress that all methods for assessing onset latency are easily distorted by high-frequency noise. I recommend using a low-pass filter with a half-amplitude cutoff of approximately 10 Hz and a slope of approximately 24 dB/octave when measuring onset latency. Keep in mind, however, that this can cause a shift in the onset times (see chapter 7). This shift will typically be equivalent across conditions, so it is not usually a problem, but you should be aware of it.

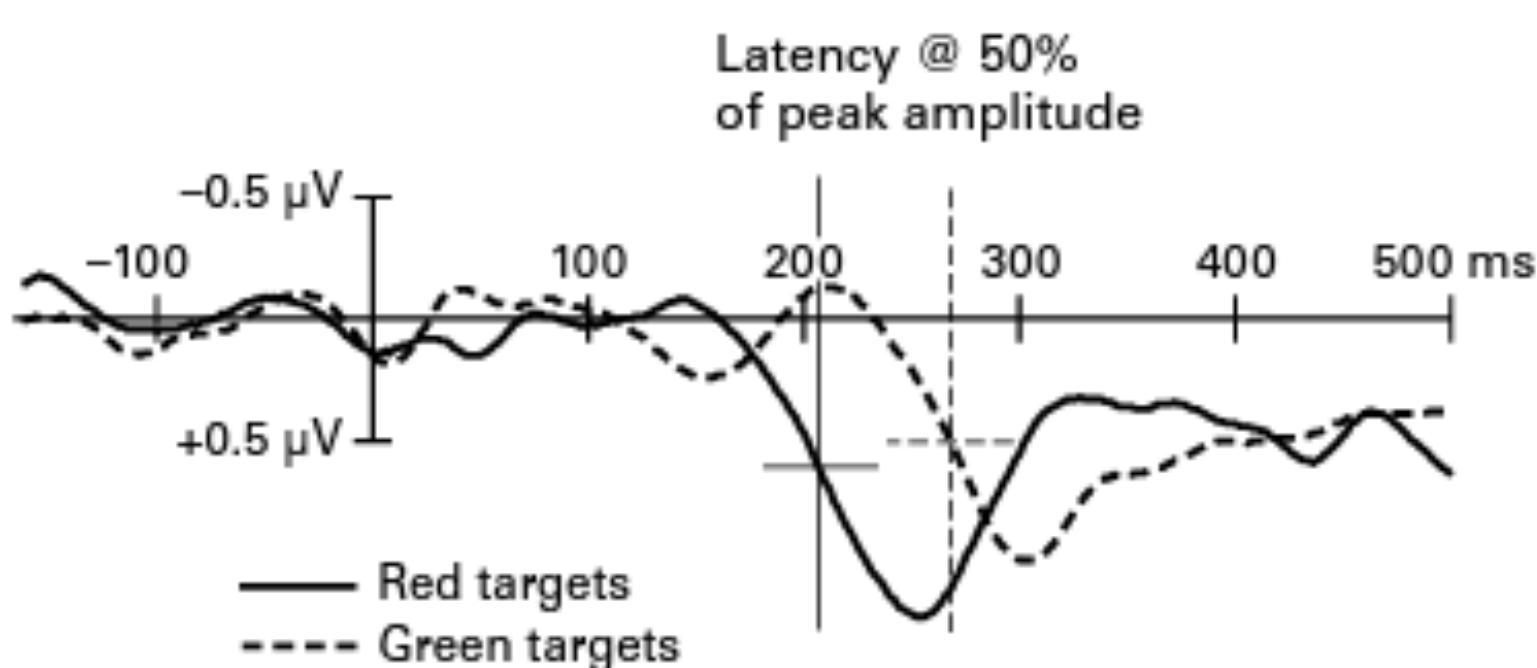
Measuring Onsets with Fractional Area Latency

One powerful approach for measuring onset latency is to use the fractional area latency measure shown in figure 9.5, but using a smaller fraction, such as the point that divides the first 25% of the area from the last 75% (which would be called the 25% area latency). The simulation study of Kiesel et al. found that a 30% area latency measure was highly reliable, although not quite as reliable as the 50% area latency in most cases. So why not just use the 50% area? Many ERP latency effects do not simply consist of a shift of the whole waveform, and the onset may change without much change in the midpoint, so it is sometimes important to measure the onset latency of an effect.

Measuring Onsets with Fractional Peak Latency

My favorite technique for measuring onset latency is called the *fractional peak latency* measure. As illustrated in figure 9.6, you first find the peak amplitude, and then you work backward in time until you find the point in the waveform at which the latency is a certain percentage of the peak value.² This time point is then considered to be the onset time. In most cases, the 50% point yields the highest reliability (Kiesel et al., 2008), in which case this can be called the *50% peak latency* measure.

The waveforms in figure 9.6 are from a visual search experiment in which subjects searched for a red target in some trial blocks and a green target in other trial blocks (Luck et al., 2006). Despite our efforts to equate the salience of the red and green targets, the red targets were quite

**Figure 9.6**

Grand average N2pc difference waves (contralateral minus ipsilateral) for red targets and green targets (from the study of Luck et al., 2006). The 50% peak latency is shown for both waveforms, defined as the latency at which the voltage reaches 50% of the peak (or local peak) voltage.

a bit more salient than the green targets. The figure shows the N2pc activity elicited by the red and green targets, isolated from the rest of the waveform with a contralateral-minus-ipsilateral subtraction (see the section on N2pc in chapter 3). The onset of this difference wave was clearly earlier for the red targets than for the green targets.

This example illustrates two virtues of the fractional peak latency measure. First, N2pc amplitude was a little larger for the red targets than for the green targets, and this is factored out by finding the time at which a percentage of the peak amplitude is reached. Second, the difference wave consisted of a rapid ramp-up from zero to the peak, then a decline, and then a long period of moderate amplitude. If we used fractional area latency instead of fractional peak latency, the long period of moderate amplitude late in the waveform would have influenced our measure of the onset time. In this situation (which is common), fractional peak latency provides a purer measure of onset time.

You might think that the 50% point is too late to provide a good measure of the onset time. However, it's actually a very good measure of the onset time in many cases. Recall from chapter 8 that when the latency of a component varies from trial to trial, the onset time of the average across trials is driven mainly by the trials with the earliest onset times (see figure 8.7). The point at which the average reaches 50% will often be close to the average onset time across trials.

The logic behind this assertion is illustrated in figure 9.7. In this example, I am assuming that the single-trial ERP is a 200-ms square wave (panel A). This is, of course, a little bit unrealistic, but it makes the example simpler. Panel B shows the probability distribution of the single-trial latencies. I am assuming that the most likely onset time is 200 ms, with a gradual fall-off in the probability of earlier and later onset times (with a normal distribution). Given this distribution, the average onset time is also at 200 ms. Panel C shows what the averaged ERP waveform would look like with the single-trial waveform shown in panel A and the distribution of onset times shown in panel B. The point at which the average reaches 50% of the peak amplitude is at 200 ms, which is exactly the average single-trial onset time. Perfect! This does not mean that the

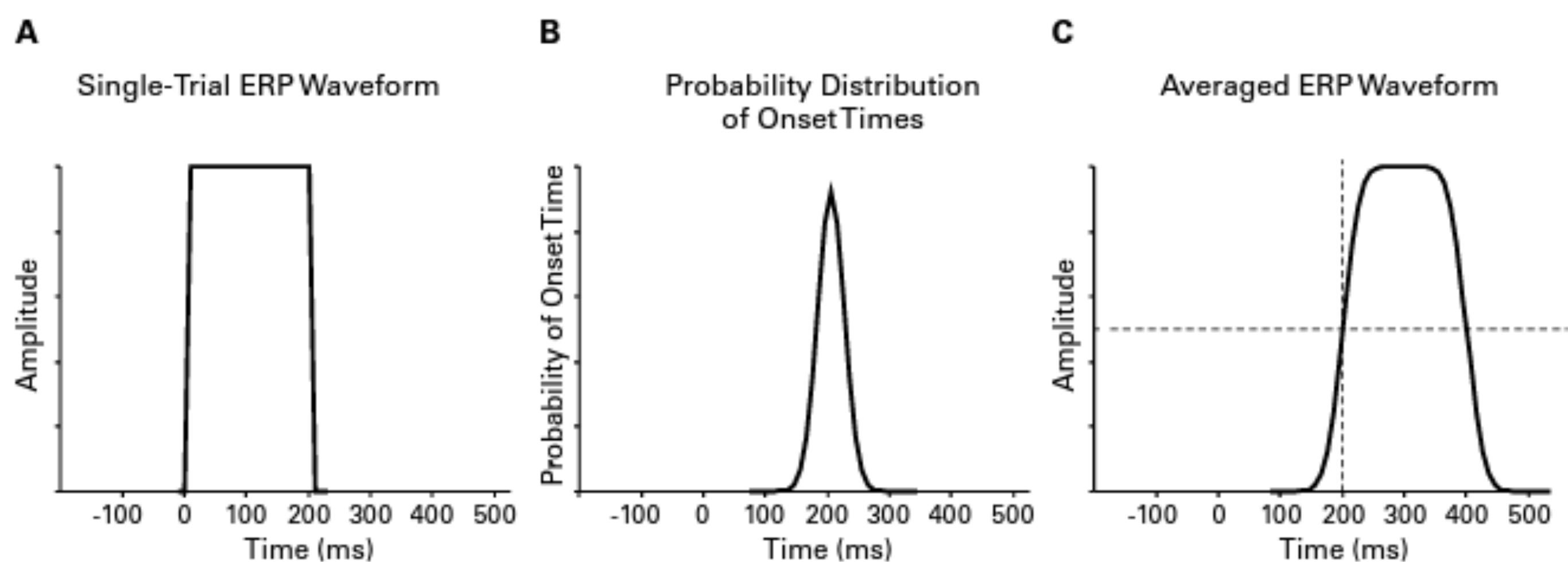


Figure 9.7

Example of how the 50% peak latency point may be a good estimate of the average onset time of a component. (A) Single-trial ERP waveform, which consists of a square wave in this simplified example. (B) Probability distribution of single-trial onset latencies. The average (and most common) onset latency is 200 ms, with a Gaussian distribution around this mean. (C) Averaged ERP waveform that results from combining the single-trial ERP waveform in panel A with the distribution of onset times in panel B.

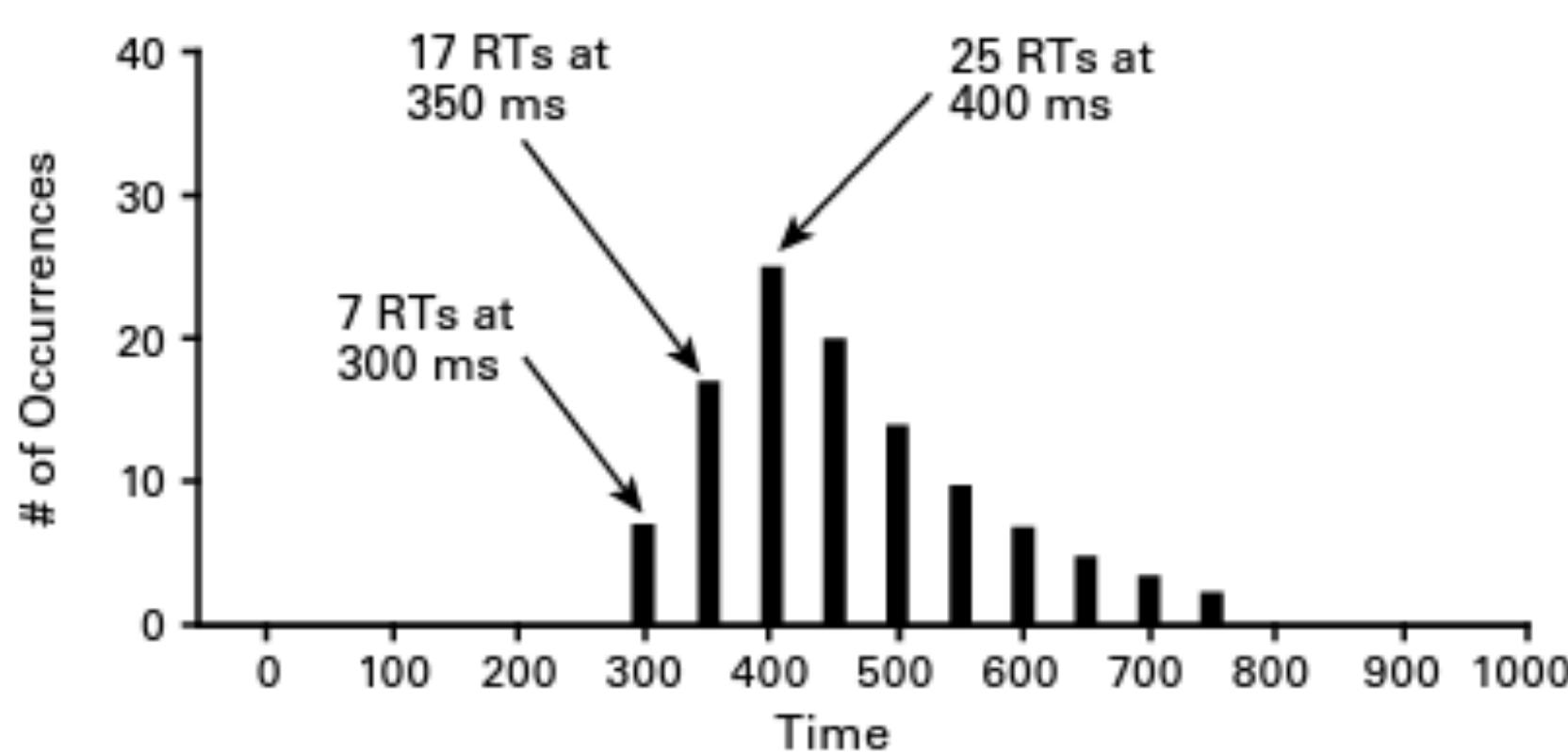
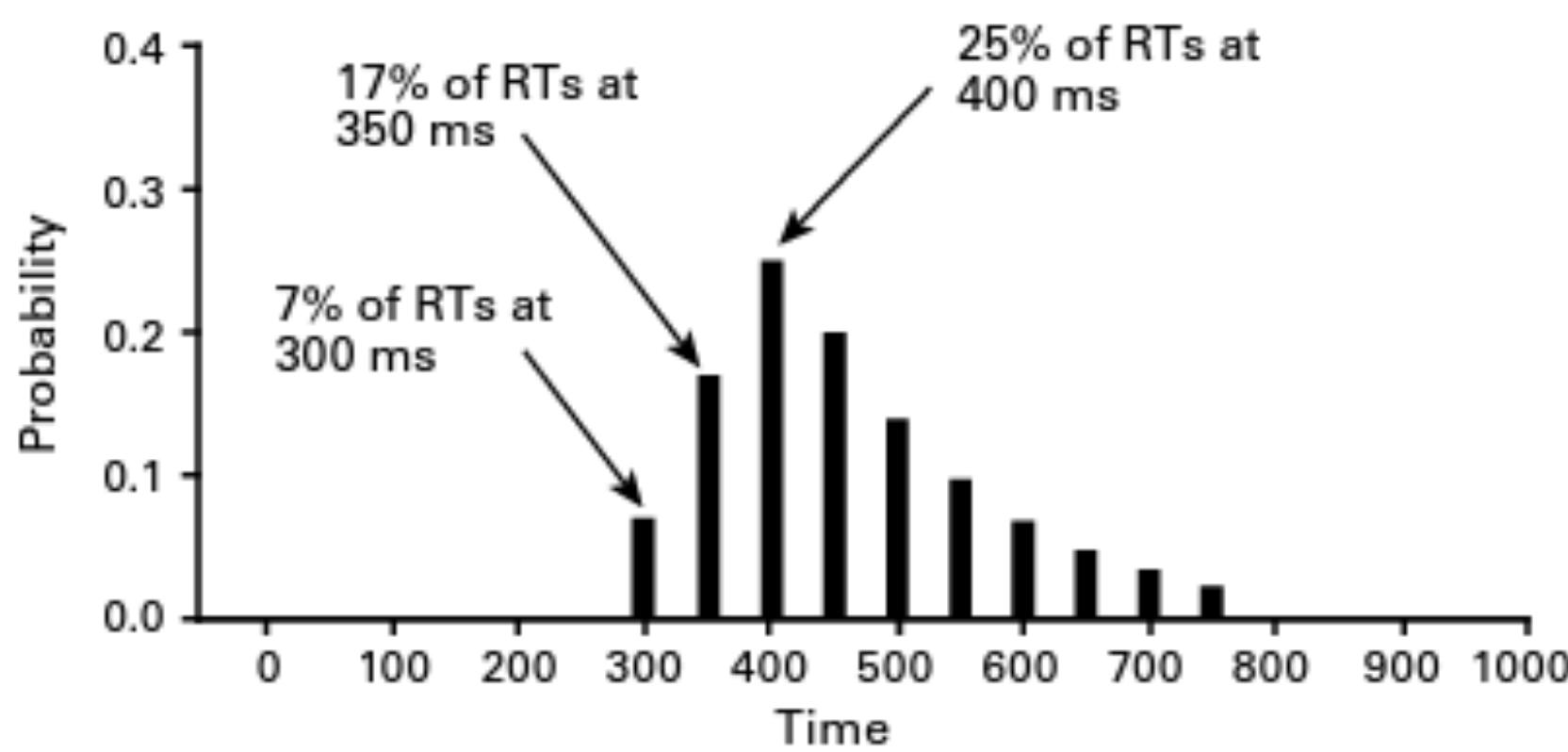
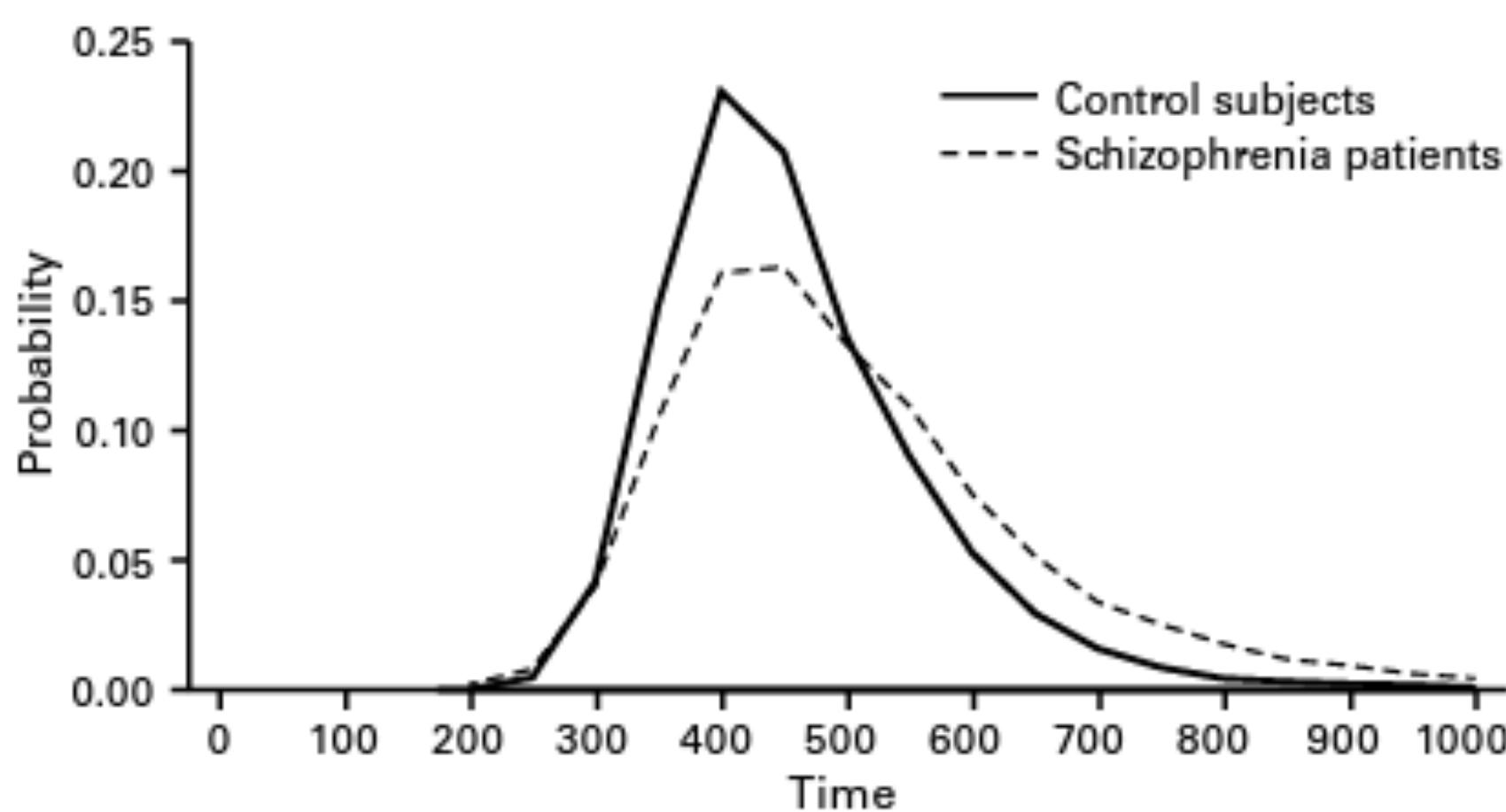
50% peak latency measure will always capture the average single-trial onset time, but it will provide a reasonable approximation under many realistic conditions.³

Kiesel et al. (2008) found that the 50% peak latency measure was not quite as reliable as the 50% area latency measure, but both were quite good. The 50% area latency measure will typically be best when the onset and the midpoint of the component are both shifted across groups or conditions. However, with more complicated patterns such as that shown in figure 9.6, the 50% peak latency measure is likely to be best. In addition, when you want a measure that is likely to be close to the average of the single-trial onset times, the 50% peak latency measure is likely to be best. For examples of how I have used these two measures with the P3 wave, the N2pc component, and the lateralized readiness potential, see Luck and Hillyard (1990), Luck (1998b), Luck et al. (2006), and Luck et al. (2009).

Comparing ERP Latencies with Reaction Times

In many studies, it is useful to compare the size of an ERP latency effect to the size of an RT effect. This seems straightforward, but it is actually quite difficult.

To understand why this is true, you need to understand *frequency distributions* and *probability distributions*. Imagine, for example, that you have recorded RTs from a subject on 100 trials of a given condition. Ordinarily, you would just summarize the RTs with the mean of the 100 trials. However, this throws away a lot of information about the RTs. A frequency distribution provides much more information, showing how often RTs occurred within particular time bins. For example, figure 9.8A uses 50-ms time bins and shows that seven of the 100 RTs occurred in the

A**B****C****Figure 9.8**

(A) Typical frequency distribution of RT. The height of each bar represents the number of responses that occurred within ± 25 ms of the midpoint latency of the bar. (B) Same as panel A, but showing probability rather than frequency (by dividing each bar by the number of trials, which in this example is 100). (C) Actual RT probability distributions from a study of schizophrenia patients and control subjects (Luck et al., 2009). These distributions were aggregated across the subjects in each group, so they reflect both within-subject and across-subject variations in RT.

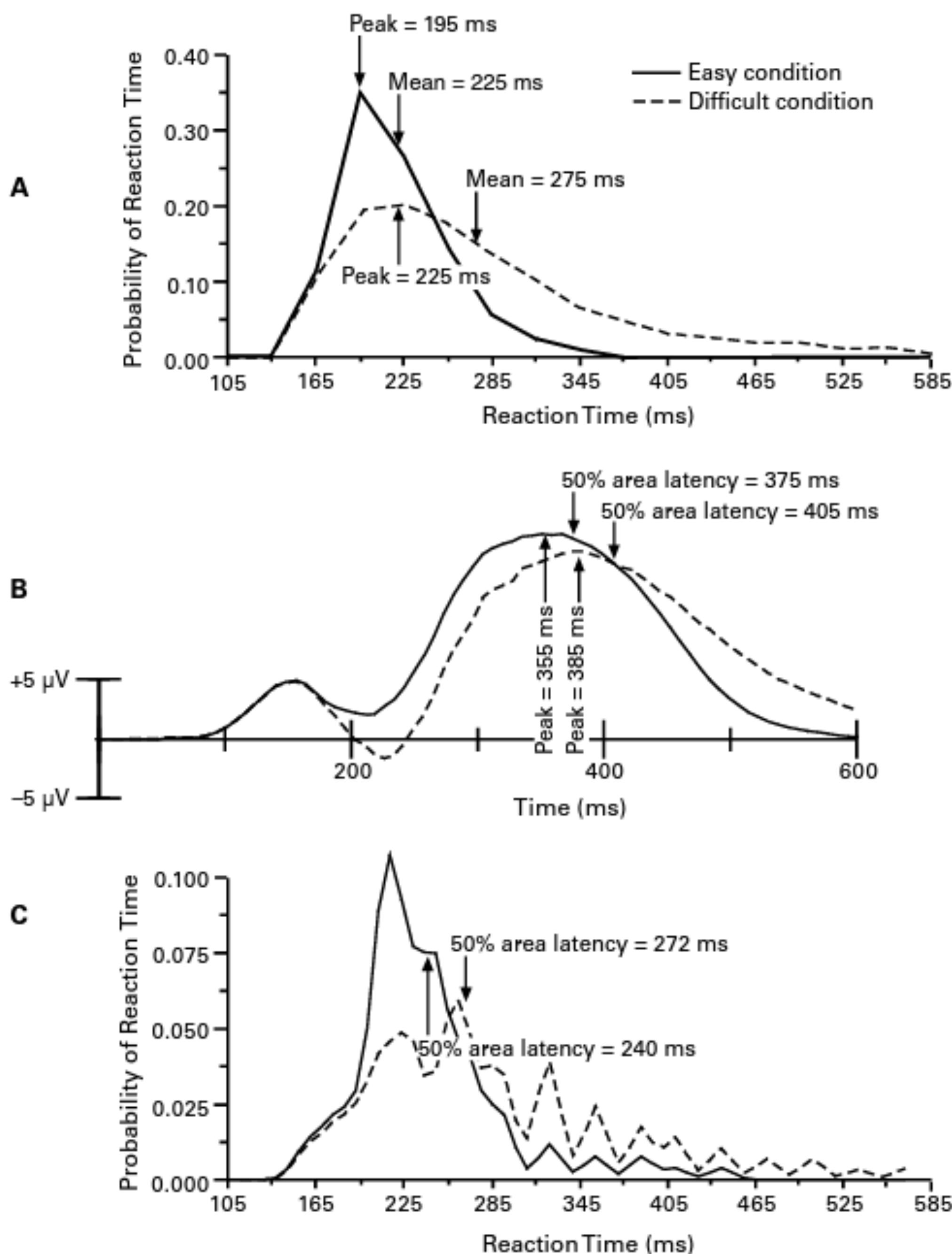
time bin centered at 300 ms (i.e., from 275 to 325 ms), 17 occurred in the time bin centered at 350 ms, 25 occurred in the time bin centered at 400 ms, and so forth. It is often useful to convert these frequencies into probabilities (i.e., the probability that a response occurred within a particular time bin), which you can do by simply dividing the value in each bin by the total number of trials. This gives us a probability distribution, as shown in figure 9.8B.

The probability distribution in figure 9.8B is skewed to the right (i.e., most of the RTs fall to the right of the peak). This is typical for RT, and many differences in RT between groups or conditions consist of an increase in the probability of relatively long RTs rather than a shift in the entire RT distribution. For example, figure 9.8C shows RT probability distributions for a group of schizophrenia patients and a group of control subjects (from the study of Luck et al., 2009). The fastest RTs were similar in the two groups, but patients had more long RTs than controls. The same pattern is often observed for comparisons between different experimental conditions in within-subject experiments.

As chapter 11 will describe in detail, finding the peak of an ERP waveform is like finding the mode (most frequent value) in a probability distribution. If you compare the latency of an ERP peak to a mean RT, this is like comparing a mode with a mean (which is like comparing apples with oranges). In many studies, such as the one shown in figure 9.8C, differences in the mean of the RT distribution will be substantially larger than differences in the mode of the RT distribution. Thus, differences in peak ERP latency will typically be smaller than differences in mean RT, even if the single-trial RT differences and single-trial ERP latency differences are exactly the same.

To make this clearer, figure 9.9A shows the probability distribution of RT in two conditions of a Gedankenexperiment, which we'll call the *easy* and *difficult* conditions. Each point represents the probability of an RT occurring within ± 15 ms of that time. The RT distributions are right-skewed, as usual, and much of the RT difference between the conditions is due to a change in the probability of relatively long RTs rather than a pure shift in the RT distribution. Imagine that the P3 wave in this experiment is precisely time-locked to the response, always peaking 150 ms after the RT. The P3 wave will therefore occur at different times on different trials, with a probability distribution that is shaped just like the RT distribution from the same condition (but shifted rightward by 150 ms). Imagine further that the earlier components are time-locked to the stimulus rather than the response (which will typically be true). The resulting averaged ERP waveforms for these two conditions are shown in figure 9.9B (see chapter 11 for a more extensive discussion of how the shape of the averaged ERP waveform will reflect the probability distribution of single-trial component latencies).

Because most of the RTs occur within a fairly narrow time range in the easy condition, most of the single-trial P3s will also occur within a narrow range, causing the peak of the averaged ERP waveform to occur approximately 150 ms after the peak of the RT distribution (overlap from the other components will influence the precise latency of the peak). Some of the single-trial RTs occur at longer latencies, but they are sufficiently infrequent that they don't have much influence on the peak P3 latency in the averaged waveform.

**Figure 9.9**

Gedankenexperiment comparing an easy condition with a difficult condition. (A) Probability distribution of reaction time, showing the probability of a response occurring in various time bins (bin width = 30 ms) in the easy and difficult conditions. (B) ERP waveforms that would be produced in this experiment if the early components were insensitive to reaction time and the P3 wave was perfectly time-locked to the responses. (C) Probability density waveforms, in which each single-trial reaction time was replaced by a Gaussian waveform (standard deviation = 8 ms), and then all the Gaussian waveforms were averaged together.

The mean RT is 50 ms later in the difficult condition than in the easy condition. However, because much of the RT effect consists of an increase in long RTs, the peak of the RT distribution is only 30 ms later in condition B than in condition A. Because the peak of the P3 wave in the averaged ERP waveform is tied closely to the peak of the RT distribution, P3 peak latency is also 30 ms later in the difficult condition than in the easy condition. Thus, the peak latency of the P3 wave changes in a manner that reflects changes in the peak (mode) of the RT distribution rather than its mean. Consequently, when RT effects consist largely of increases in the tail of the distribution rather than a shift of the whole distribution, changes in peak latency will usually be smaller than changes in mean RT, even if the component and the response are influenced by the experimental manipulation in exactly the same way. Consequently, you shouldn't compare ERP peak latency effects with mean RT effects. Box 9.3 describes my first encounter with this phenomenon as a college student.

How, then, can RT effects be compared to ERP latency effects? The answer is that they must be measured in the same way. One way to achieve this would be to use a peak latency measure for both the ERPs and the RTs (using the probability distribution to find the peak RT). However, the peak of the RT distribution is difficult to estimate reliably, and peak RT effects are likely to be smaller than mean RT effects (just as peak ERP latency effects tend to be smaller than mean RT effects).

An alternative is to use the 50% area latency measure for quantifying the ERP latencies and compare this with median RT. Median RT is the point that separates the fastest half of the RTs from the slowest half, which is almost the same thing as the point that divides the area into two equal halves. I have used this approach in several experiments, and the correspondence between

Box 9.3

Convolution, Peak Latency, and Mean Reaction Time

When I was a student at Reed College, I had a wonderful mentor named Dell Rhodes. Dell was trained as a physiological psychologist, which meant that she started her career in the 1970s studying the brains of rats. As the field of cognitive neuroscience started to emerge, Dell decided to take a new research path, and she spent a sabbatical year learning how to record and analyze ERPs. I started working with her the very next year (in my junior year of college), and I learned a tremendous amount from her as we muddled through the realities of setting up and running an ERP lab.

One day Dell remarked to me that she had been reading a lot of papers that looked at both P3 latency and mean RT, and the mean RT effects (i.e., differences in RT between groups or conditions) were always larger than the P3 latency effects. Both effects were on the same time scale (milliseconds relative to stimulus onset), so it was not obvious why the P3 effects should always be smaller than the RT effects. Dell's comment stuck with me for many years, and I finally figured out the answer to this riddle when I started to think about how peak latency was related to the mode of the RT distribution. The studies Dell had been reading were comparing peak latency with mean RT, and peak latency is analogous to the mode rather than the mean of the RT distribution. This explains why the P3 latency effects were smaller than the RT effects.

P3 latency and RT was excellent (Luck, 1998b; Luck et al., 2009). In most cases, this is the approach I would recommend for comparing ERP latencies with RTs.

You should note, however, that 50% area latency is not perfectly analogous to median RT, because median RT does not take into account the precise values of the RTs above and below the median. For example, a median RT of 300 ms would be obtained for an RT distribution in which half of the values were between 200 and 300 ms and the other half were between 300 and 400 ms, and the same median of 300 ms would be obtained if half of the RTs were between 290 and 300 ms and the other half were between 300 and 5000 ms. However, the correspondence between 50% area latency and median RT is close enough for most purposes, and median RT has the advantage of being a familiar measure.

If you want a measure of RT that is perfectly analogous to the 50% area latency measure, you need to find the point that bisects the *area* of the RT distribution. However, RTs are discrete, instantaneous events with zero area, making it difficult to measure the area of the RT distribution. Figure 9.9C shows how you can convert the data into an ERP-like waveform and then measure the area, using a technique borrowed from single-unit recording studies (see, e.g., Szűcs, 1998). Each individual RT is replaced by a Gaussian function to turn it into a continuous function, and then the average of all the RTs is computed. The result is a waveform that looks a little bit like an ERP waveform, making it possible to calculate the 50% area latency for RT. In the example shown in figure 9.9C, the resulting latencies were 240 ms for the easy condition and 272 ms for the difficult condition, and the 32-ms difference between these latencies was nearly identical to the 30-ms effect that was obtained by measuring the 50% area latency in the ERPs. I don't know of anyone who has tried this approach, but it seems like the optimal way of comparing ERP latencies and RTs.

Suggestions for Further Reading

- Dien, J., Spencer, K. M., & Donchin, E. (2004). Parsing the late positive complex: Mental chronometry and the ERP components that inhabit the neighborhood of the P300. *Psychophysiology*, 41, 665–678.
- Donchin, E., & Heffley, E. F., III. (1978). Multivariate analysis of event-related potential data: A tutorial review. In D. Otto (Ed.), *Multidisciplinary Perspectives in Event-Related Brain Potential Research* (pp. 555–572). Washington, DC: U.S. Government Printing Office.
- Kiesel, A., Miller, J., Jolicœur, P., & Brisson, B. (2008). Measurement of ERP latency differences: A comparison of single-participant and jackknife-based scoring methods. *Psychophysiology*, 45, 250–274.
- Sawaki, R., Geng, J. J., & Luck, S. J. (2012). A common neural mechanism for preventing and terminating attention. *Journal of Neuroscience*, 32, 10725–10736.
- Spencer, K. M., Dien, J., & Donchin, E. (2001). Spatiotemporal analysis of the late ERP responses to deviant stimuli. *Psychophysiology*, 38, 343–358.