Chapter 5 Supplement

Electrodes and Skin Potentials

This supplement describes two separate but related issues. First, it describes how more electrodes might actually give you poorer data quality. Second, it describes evidence that recording with high electrode impedances can dramatically reduce your statistical power under some conditions. These are related because high-impedance recordings are particularly valuable as a means of saving time (by avoiding the process of reducing the electrode impedances) when large numbers of electrodes are used.

How Many Electrodes Do You Really Need?

In the early days of ERP recordings, a separate amplifier unit was required for each channel, and the amplifiers and associated circuitry were very expensive. A "multichannel study" in those days meant that three to five active sites were used rather than one or two. Over time, the equipment has become progressively less expensive, and now almost every lab has at least 16 channels, and many labs have 128 or even 256 channels. But how many electrodes do you really need?

In an ideal world, you would want your electrodes to be spaced sufficiently close together that you would have a perfect measurement of the distribution of voltage over the surface of the head. The combination of skull and scalp blurs the voltage, and it is therefore possible to perfectly measure the distribution of voltage with a set of discrete but closely spaced electrodes. If your electrodes are spaced too far apart, you will be missing information. Thus, one would want to record from a large number of closely spaced electrodes in an ideal world.

However, we do not live in an ideal world. In the real world, recording from large numbers of electrodes has some significant costs. The most obvious cost is the price of the electrodes and amplifiers. But this ends up being a relatively small cost in the long run, especially compared with the other costs involved in running a lab (mainly personnel costs). A less obvious cost is the amount of time you will spend copying huge files onto your analysis computer, loading them into your data analysis program, and waiting for the analyses to finish. This may seem trivial, but I have many times observed people deciding not to do a given analysis because of the amount of time required. When I ask one of my students or postdocs to redo all of the analyses in an experiment with a different filter setting, the number of channels plays a major role in whether he or she will be up past midnight.

The least obvious but most important cost is that the quality of your data will almost certainly decline as the number of electrodes increases. To make this cost clear, imagine that you are the quality control inspector in a bicycle factory. If you work for someone like Cervélo, which makes a small number of high-quality bikes each year, you can carefully inspect each bike at each step of the production process, checking the joints, the paint, the cable tensions, and so forth. And if you find a problem, you can halt the production line and make sure that the problem is solved before a large number of substandard bikes have been produced. However, if you work for someone like the Shanghai Phoenix Import & Export Company, which makes approximately 1.5 million bikes each year, you won't be able to give each bike much attention. Consequently, many bikes may leave the factory with minor flaws, and a major problem in the production line may go on for days without being noticed. The company will then need to decide whether to sell the substandard bikes for a deeply discounted price or to throw them all away because they're ruined.

When you are in the laboratory recording the EEG, you are the quality control inspector. You need to monitor the EEG throughout the session (and suppress the temptation to read journal articles or surf the Internet). The more electrodes you record from, the more difficult it is to detect problems and solve them before the data are ruined. If you are recording from just a few channels, you can make sure that you are getting beautiful data. If you are recording from 128 channels, each EEG trace on the monitor will be tiny, and there will be so many traces that it will be difficult for you to notice problems. As a result, your data may be so noisy that some of the key effects do not reach the 0.05 threshold for statistical significance. You may end up having to publish your data in a lower-tier journal or throwing them out completely.

There is another hidden cost associated with recording from large numbers of electrodes. Specifically, the more electrodes you have, the more opportunities you have to obtain spurious results. Imagine, for example, that you are looking for a very small difference in P3 amplitude between two conditions. If you record from 128 electrode sites and conduct an independent *t* test at each electrode site comparing the two conditions, there is a very large probability that you will obtain a significant *p* value at one or more sites even if there is no real effect. You would therefore need to use a correction for multiple comparisons, and an effect would be considered significant at a given site only if the *p* value at that site was less than 0.0004. Consequently, you would very likely miss a small but real effect. Now imagine that you did the same experiment but recorded only from the Pz electrode site, where the P3 component is largest. You would need to do only one *t* test, so you would not need to correct for multiple comparisons, and you would be able to use an uncorrected alpha of 0.05, giving you the power to detect a small effect. This issue is described in detail in chapter 10, which describes the *more-is-less principle*: The more conditions, time points, and electrodes are in your data, the less statistical power you will have.

You might be thinking that recording only from Pz would be problematic if the effect happened to be largest 1.5 cm anterior and 2.2 cm to the left of Pz, because you might miss the effect. Given that ERPs are spatially blurred by the skull and scalp, experimental effects are almost always distributed over several square centimeters of scalp area, so it is unlikely that you will completely miss an effect if you record from an electrode somewhere in the neighborhood of the maximal location of an effect. And if you record from a large number of electrodes to search for this maximal location, the cost of correcting for multiple comparisons (in terms of statistical power) will almost always outweigh the benefits of having an electrode at exactly the right place. I am

not suggesting that you should actually record from only one electrode; my point here is that there are hidden costs associated with recording from large numbers of electrodes, and it is important to choose a number of electrodes that balances the costs with the benefits.

A common approach that is used with large numbers of electrodes is to combine across clusters of electrodes, reducing the number of independent statistical tests that must be conducted. There is nothing wrong with this in principle (assuming that the clusters are not biased by the data; see chapter 10). However, when large numbers of electrodes are used and this kind of clustering is done, it is a little odd to throw away the spatial resolution that was obtained at the cost of reduced data quality.

Before using a larger number of electrodes, you should ask yourself why you need a highly detailed assessment of the scalp distribution. The most obvious answer is that more channels are needed for accurate source localization. However, source localization is difficult on the basis of ERP scalp distributions alone, and accurate localization is virtually impossible with noisy data (see chapter 2 and online chapter 14). As a result, there are very few studies that have had a large impact outside the ERP community that required more than 20 electrodes. In contrast, many high-impact studies of broad interest relied mainly on a small number of electrodes (even if the recordings included many electrodes).

My general advice is to record from between 16 and 32 active electrode sites in most experiments. My lab typically recorded from about 20 channels when we used low-impedance systems, and we now typically record from 32 channels with our high-impedance systems. We occasionally record from 64 channels; this is usually when we've discovered a new effect and want to determine its scalp distribution with a fair amount of precision to guide future research. For example, Marissa Gamble ran an experiment with 32 channels in my lab in which she discovered an auditory analog to the N2pc component, and she then ran a follow-up experiment with 64 channels so that we could precisely establish the scalp distribution of this new effect (Gamble & Luck, 2011). Of course, there are situations in which you might want to record from fewer or more electrodes, but you should always consider both the costs and benefits of doing so.

Effects of Skin Potentials on Statistical Significance

A long time ago, Terry Picton and Steve Hillyard published a study showing that decreasing the impedance of the skin dramatically reduces skin potentials (Picton & Hillyard, 1972), and this was one of many things that I learned about recording clean EEG data when I was in graduate school. Many years later, a friend of mine was considering purchasing an EEG system that was designed to be used with high electrode impedances. I told him that I thought this would lead to problems with skin potentials. He forwarded this concern to the manufacturer, who responded by saying that they did not think this would be a real problem with their modern amplifier designs.

A few years later, I decided to try a high-impedance EEG recording system in my own lab. There are some clear benefits of high-impedance systems (reduced preparation time and reduced likelihood of disease transmission), and I thought it would be worth trying one. I was not convinced that modern amplifier designs would deal with the problem of skin potentials, however, so Emily Kappenman and I decided to run an experiment to test the effects of electrode impedance in our new system (Kappenman & Luck, 2010). We ran an oddball task, and for each subject we

reduced the impedance of the electrodes on one side of the head to 5 kΩ or less (by abrading the skin under each electrode on that side). The electrodes on the other side of the head were not abraded, and the impedance at these electrodes varied naturally between 10 and 190 kΩ. We reduced the impedance for the left hemisphere electrodes in half the subjects and for the right hemisphere electrodes in the other half. Because skin potentials should be more prevalent when the recording environment is warm and humid, each subject was tested in two mini-sessions, once in a cool and dry environment (21.3°C [70.3°F], with 7.67 g/m^3 of humidity), and once in a warm and somewhat humid environment (27.7°C [81.8°F], with 11.5 g/m^3 of humidity). If increased impedance leads to more skin potentials, even in modern EEG amplifiers, then more skin potentials should be observed in the high-impedance hemisphere than in the low-impedance hemisphere, especially under warm and humid conditions. This is exactly what we found.

Figure S5.1A shows an example of 5-min periods of EEG recorded from electrodes with low and high impedances under cool and warm conditions in a single subject. In several of these waveforms, a large but very gradual shift in voltage can be seen over the 5-min period, likely reflecting gradual changes in skin hydration and related factors. Sharp voltage spikes can also be seen—these are eyeblink artifacts (see chapter 6). If you look at the waveform for the high-impedance, warm-temperature recording, you will also see some very large voltage shifts that go up and down over periods of tens of seconds. These are skin potentials. If you look carefully, you can also see smaller skin potentials in the high-impedance, cool-temperature recording.

To aggregate the data across subjects, we computed the frequency content of the EEG and averaged across subjects (see chapter 8 for more information about this type of analysis). Skin potentials are slow shifts, so they would be expected to lead to an increase in low frequencies. As shown in figure S5.1B, increased power was observed in the low frequencies for the high-impedance recordings, and this was especially evident when the recording environment was warm and humid. This is exactly what would be expected if high electrode impedance leads to an increase in skin potentials.

Emily and I wanted to know if this would have a real impact on our experiments. In particular, we were concerned that more low-frequency noise from skin potentials would add variance to our data, decreasing the likelihood that a real result would be statistically significant or increasing the number of trials that we would need to average together for each subject to reach statistical significance. To assess this quantitatively, Emily conducted a set of Monte Carlo analyses, in which she simulated experiments with different numbers of trials by subsampling randomly from the very large number of trials that were actually collected from each subject (for details, see Kappenman & Luck, 2010). In Emily's initial analyses, she found that the likelihood of obtaining a significant result was dramatically reduced for the high-impedance electrodes, especially when the recording environment was warm and humid, but this was partly due to a small number of trials with sudden, very large voltage changes (see figure S5.1A for examples). She therefore performed artifact rejection to exclude these trials from the main analyses.

Some of these analyses were focused on P3 amplitude. When Emily simulated an experiment with a relatively small number of trials, the likelihood of obtaining a significant effect (e.g., a difference in P3 amplitude between rare and frequent trials) was approximately half as great for the high-impedance recordings as for the low-impedance recordings. To reach a given level of statistical power, two to three times as many trials per subject were needed for the high-impedance

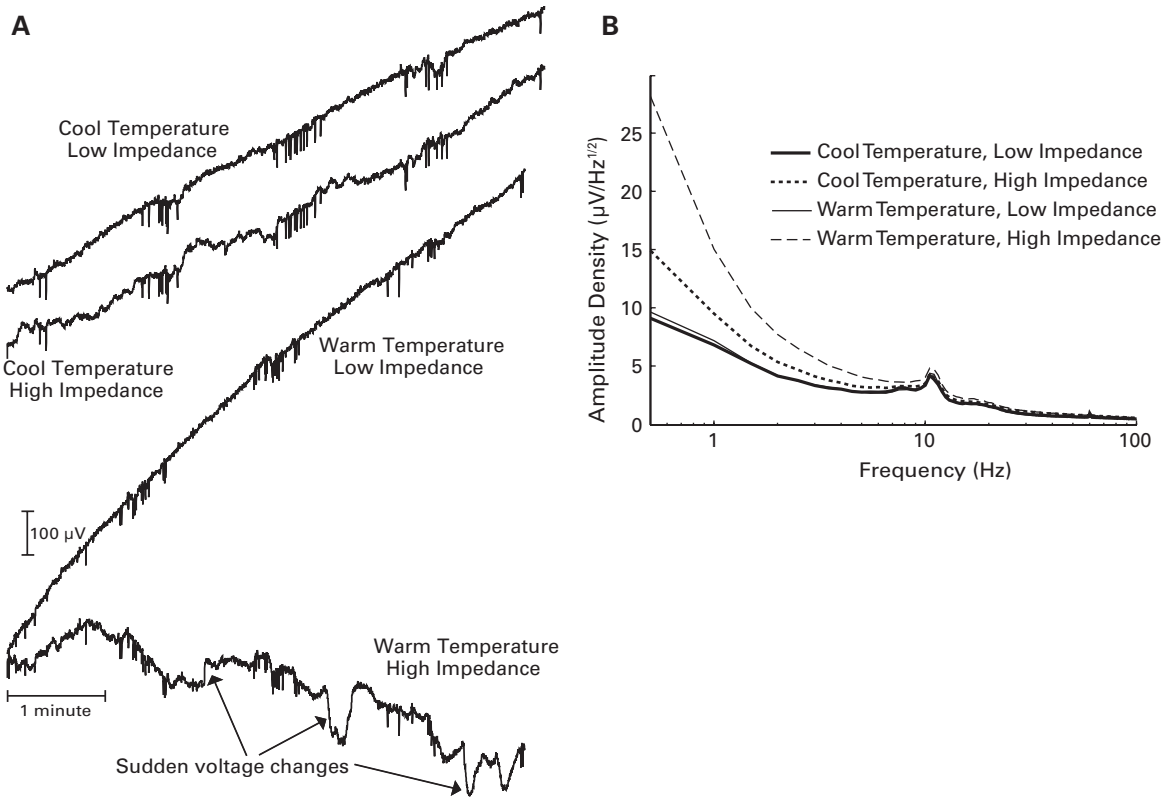

Figure S5.1
(A) Effects of electrode impedance and recording environment on a 5-min segment of EEG. The small voltage spikes are eyeblinks. The very slow changes in voltage reflect changes in the hydration of the skin and similar factors. The voltage variations that occur over periods of tens of seconds are skin potentials. (B) Effects of electrode impedance and recording environment on the power at each frequency, averaged across subjects and time.

recordings compared to the low-impedance recordings. These effects of electrode impedance and recording environment were attenuated somewhat by filtering out the low frequencies in the EEG (and therefore filtering out much of the power in the skin potentials). However, when fairly strong filters were used, visible distortions of the ERP waveforms could be observed (see chapter 7 for a detailed discussion of why filtering can distort ERP waveforms). Moreover, even very strong filters were not enough to completely offset the decline in statistical power observed in the high-impedance recordings.

Emily also conducted some analyses on N1 amplitude. She found much less effect of impedance on statistical power, and statistical power was nearly identical for low- and high-impedance recordings when the recording environment was cool and dry. You are probably wondering why impedance had so little effect on statistical power for N1 amplitude when it had a very large effect

for P3 amplitude. The answer is most likely that the N1 was measured in a period of time that immediately followed the prestimulus baseline period. Specifically, we measured N1 amplitude as the mean voltage 80–100 ms after a 400-ms prestimulus baseline. Any slow changes in voltage arising from skin potentials are unlikely to have much effect over this brief time period. In contrast, P3 amplitude was measured as the mean amplitude from 350 to 650 ms (relative to the prestimulus voltage), providing more time for skin potentials to cause the voltage to drift away from the baseline value (the importance of baselines for amplitude measurement will be discussed in more detail in chapter 9).