Dickson, D. S., & Pelzl, E. (2023). Using time-based electroencephalography to investigate second language. The Routledge Handbook of Second Language Acquisition and Neurolinguistics, 17-30.

Using time-based encephalography to investigate L2

Danielle Dickson (ORCID: 0000-0002-7565-9999), Eric Pelzl (ORCID: 0000-0001-8335-8814)
The Pennsylvania State University
111 Moore Building, University Park, PA 16802

Word Count: 7,299

Abstract

This chapter reviews the event-related potential (ERP) method and its application in second language research. The primary aim is to familiarize readers with the fundamental principles of the technique while highlighting the methodological approaches most frequently used for hypothesis testing in second language research. Practical design considerations, commonly used paradigms, benefits and limitations of ERP measures, and general guidance for interpretation of measures of interest are provided. Special attention will be given to critical measures including the mismatch negativity (MMN), N400, P600, and other waveforms of the ERP. The chapter concludes with a brief discussion of recent innovations and new directions.

Introduction and Critical Definitions: What is EEG?

Electroencephalography (EEG) is the measurement of ongoing electrical activity in the brain (Cohen, 2017, Kirschstein & Köhling, 2009). This signal can be recorded directly in the brain, but in typical laboratory studies involving healthy participants, the EEG signal is measured with sensors placed at the scalp (Luck, 2014). In this chapter, we will be focusing on the time-domain signal of the EEG, meaning the positive and negative-going waveform that unfolds over time (for more on the frequency domain of the same signal, see Mottarella and Prat, this volume).

EEG Data Collection Procedures and Practical Considerations

The procedure for acquiring EEG data entails placing an elastic cap with electrodes embedded in fixed locations on a participant’s head. This approach is considered non-invasive as no chemicals or exterior electromagnetic currents need to be applied. Electrical potentials that are naturally produced by the brain are passively recorded as the person performs tasks and takes in stimuli (Luck & Kappenman, 2012). In order for the signal to be high quality, the impedance between the surface of the scalp and the surface of the electrode sensors must be kept low, typically achieved via application of saline gel-based solutions and gently moving hair out of the way to create a bridge between the scalp and the recording electrodes (Luck, 2014).

Data recorded at the scalp ideally would only include brain activity and no other signal, but that is typically not the case. Instead, external electrical activity from monitors and other electrical equipment can also be unintentionally included in the recorded signal in addition to biological activity unrelated to the neural activity of interest. These irrelevant data signals are considered noise, which researchers mitigate through different strategies (for more on technical recording procedures, see Luck, 2014).

Biological noise includes muscle activity from clenching the jaw, furrowing the brow, or tensing scalp/facial muscles, in addition to eye movements and blinks. For this reason, after the electrode cap is on a participant’s head, but before the experiment begins, investigators usually allow participants to look at their brainwaves, and then demonstrate the effects of muscle activity (e.g., jaw clenching) on the data. This real-time feedback enables participants to understand how to reduce those signals during recording by staying still, fixating centrally during critical recording periods, and blinking according to instructions. EEG experiments often designate periods of time in the experiment when blinks/movement are permitted.

Neurobiological Basis of the EEG Signal

Although it is not necessary to be an expert in neurobiology in order to use EEG methods, a basic understanding of where the EEG signal comes from can be helpful. What the EEG signal captures is a direct reflection of neural activity, typically thought to be dominated not by any individual neuron firing (action potentials), but rather for the most part by the summations of synchronous inhibitory and excitatory activity (postsynaptic potentials) across large populations of neurons close to the scalp (primarily from cortical pyramidal cells) due to their particular arrangement and physical orientation (Kirschstein & Köhling, 2009; Nunez et al., 2016). EEG measures electrical potential, which is a relative comparison of electrically charged activity across locations (e.g., extracellular negativities vs. positivities). Thus, EEG always requires at least two sensors, one of which serves as a comparative reference (and ground for the amplifier) and the other(s) as active/recording electrode(s) whose data is reported (Nunez et al., 2016).

The recorded signal is very small and usually measured in microvolts (µV, one millionth of a volt). For this reason, the signal must be amplified during the recording process, prior to digitization on a computer. The absolute voltage reported at a given sensor at any time point is not typically of interest because it is dependent on what reference electrode is selected. For example, if the reference is right next to the electrode of interest, the voltage measured at that site will likely be small, as the voltage of the reference electrode is essentially subtracted from the voltage of the electrode of interest—but this does not tell us anything about the amount of brain activity there. When interpreting brainwaves, which are dynamic plots of EEG signal across time (see Figure 1, right), the shape, or morphology, of the waves themselves and their relative positioning compared to one another is what matters.

< Figure 1 here>

Event-related Potentials

Even when a participant is sitting still and not doing anything in particular, the continuous EEG signal can carry valuable information (see Mottarella & Prat, this volume). However, many researchers examining language processing are interested in the brain response to their experimental stimuli. These stimuli are the 'events' of the event-related potential (ERP) method. Figure 1 displays how a typical study moves from continuous EEG to analysis of ERPs. Continuous EEG is measured at the scalp across an array of electrodes (left panel). Within this continuous signal, time windows around events of interest are marked (middle panel). The time-locked brain response to these events is then extracted and plotted, averaging across all the instances of particular events of interest. The outcome of this process is the ERP, which is often displayed as a waveform (right panel). With sufficient trials, noise and brain activity unrelated to the experimental event should largely average out, leaving the ERP to reflect the neural processing associated with the experimental stimuli (Luck, 2014).

ERPs provide information about the timing, magnitude, and distribution of neural activity that is tied to specific experimental events. For instance, when a participant sees a word, their neural response to that word will be captured as it unfolds over time, moving in either a positive or negative direction relative to the baseline where it began. ERP waveforms illustrate change in amplitude (measured in microvolts) across time (measured in milliseconds). ERPs also allow us to assess the distribution of that activity across the electrodes positioned on a person's scalp. For instance, we might see that differences in ERP amplitudes at 400 milliseconds are larger at electrodes on the front of the head compared to electrodes at the back of the head. When responses are elicited under similar conditions and share a similar distribution, it might suggest that they share similar neural generators (though localization of those sources is not a strength of EEG, see the section Pros and Cons/Limitations of the ERP Method below). In this way, across locations on the scalp, we can compare when and how ERPs differ among participant groups and/or stimulus conditions. Across experiments, common patterns are recognized and identified as particular ERP components.

ERP Paradigms and Components

In this section, we summarize experimental paradigms and the related ERP components that are most commonly used in second language (L2) research. In some cases, we also identify additional ERP components that have been used infrequently, but hold promise for further investigation in L2 research. For more technical and in-depth discussions of these components, we refer readers to Luck and Kappenman (2012).

Note that components names have come about in a variety of ways, leading to sometimes confusing taxonomies. In general, ERP components are labeled according to their polarity as either N (negative) or P (positive) and then assigned an additional label that indicates either its place in the order of components (e.g., the N1 is the first negative-doing component, the N2 is the second), the timing of its peak (e.g., the N400 peaks at approximately 400ms after stimulus onset), or sometimes a general description (e.g., the mismatch negativity, or the late positive component). Confusion can arise because the same component is sometimes associated with multiple names (e.g., the N1 and N170), or what was originally thought of as a single component may come to be seen as a family of components (e.g., the P3a and P3b). There is no simple solution for understanding naming conventions besides becoming familiar with the relevant literature for each component.

Paradigms and Components for Investigating Linguistic Processes

Investigating L2 Speech Perception with the Mismatch Negativity

For ERP research of L2 speech perception, the mismatch negativity (MMN) has proven to be an extremely useful component. The MMN is a measure of auditory change detection (Näätänen et al., 2019). It occurs as a negative-going deflection approximately 100-200ms after the onset of a detectable difference between auditory stimuli and is typically largest at frontocentral electrodes. As the name implies, the MMN is elicited by a *mismatch*; its measurement entails comparison between responses to frequent *standard* and infrequent *deviant* stimuli. For instance, the MMN might be elicited by comparing responses to frequently occurring high tones and infrequent low tones. The size of the MMN reflects the *perceived* distance between standard and deviant stimuli. Because the MMN reflects perceived differences, it can be used both as a measure of the perception of acoustic differences, and also to target perception of more abstract differences, such as phonetic categories. The targeted differences can be determined entirely by experimental context, or can draw on assumptions about participants' experience outside the lab, for example, as speakers of specific languages. It is this latter possibility that allows for investigation of participants' sensitivity to spoken features of an L2.

The MMN has been used to investigate listeners' sensitivity to a wide array of phonological features, including, but not limited to, voice-onset time (Brandmeyer et al., 2012), vowel quality (Peltola et al. 2003), and tone (Chandrasekaran et al., 2007). The MMN can be elicited in tasks that require an active response from participants, or, in contrast, while participants' attention is focused elsewhere (e.g., while watching a silent movie or reading).

When designing MMN studies, it is important to keep in mind that the onset of the MMN will be determined by the point when the difference between standard and deviant stimuli becomes detectable. For instance, if the contrast is between sounds that occur at the onset of the stimulus (e.g., /ba/ vs. /pa/), the MMN will occur immediately when that difference is detectable. If instead the contrast was between vowel length (e.g., /ba/ vs. /baa/), the MMN would occur at the point where the durational difference becomes apparent.

Investigating Sentence and Word Processing

Language researchers are often interested in how individuals process sentences and words. Figure 1 shows how ERPs time-locked to the onset individual words of interest can be generated. Several experimental paradigms can be used to probe research questions related to the cognitive processing of individual words (e.g., measuring vocabulary or orthographic sensitivity), such as go/no-go, lexical decision, or semantic/visual priming tasks, which are easily adaptable for ERP studies (for example L2 studies, see Pu et al., 2016; Wu & Thierry, 2010).

The typical method for examining how readers process sentences involves presenting a sentence word-by-word, a technique called Rapid Serial Visual Presentation (RSVP, with a typical stimulus onset asynchrony of 500ms). Although most fluent readers can process individual words faster, significantly speeding the rate of presentation can lead to different cognitive outcomes (Wlotko & Federmeier, 2015) as well as slower temporal onsets of critical components (Kutas, 1987). A key feature of this design is that presenting words individually allows researchers to tap into cognitive processing at each word. This enables the examination of stages of processing as they occur, instead of relying on end-stage behavioral measures as a window into prior sentential processing.

In auditory paradigms, speech is typically presented with a more natural flow, rather than word-by-word. This means that ERPs for individual words reflect brain responses that are continuously changing based on the ongoing speech signal, rather than the more discrete responses that result from the time-controlled pacing typical of ERP studies of written language processing. This natural variation in the way spoken words and sentences unfold over time, in addition to the recruitment of auditory rather than visual processing mechanisms, may result in observable differences in the qualities of ERPs to spoken language relative to visual word stimuli (for a comparison across modalities, see Holcomb & Neville, 1990).

In order to more tightly control the temporal alignment of spoken stimuli, efforts have occasionally been made to empirically pre-identify each critical word's isolation point, that is, the precise time within a spoken syllable when the word becomes recognizable (e.g., Van Petten et al., 1999). Hypothetically, each critical word's isolation point would be the most appropriate event for ERP time-locking, but in practice the benefits of time-locking to isolation points has not been found to offset the costs for the amount of work involved. Instead, researchers will usually mark a particular event's onset at the beginning of the critical word or syllable.

Semantic Processing as Measured by the N400

The specific ERP component most famously utilized in sentence processing studies is the N400, due to its power for revealing lexical/semantic processing. It manifests in response to any potentially meaningful stimulus regardless of modality. That is, N400s have been reported in response not just to words, but also to pictures, gestures, environmental smells—any stimuli that might convey meaning (e.g., Federmeier & Kutas, 2001; for review, see Federmeier, 2022; Kutas & Federmeier, 2011). When looking at an ERP waveform, the N400 appears as a negative deflection around 400ms after the stimulus occurred (see Figure 2). By comparing the size of the

N400 in response to different conditions (e.g., a target word after high or low constraint context), we can compute the N400 effect.

< Figure 2 here>

In the context of sentence processing experiments, it is important to recognize that the N400 manifests as part of the standard response to a word. Given no contextual support, the default amplitude of the N400 is large. Critically, the amplitude of the N400 reduces when provided with supportive context (e.g., Kutas & Hillyard, 1980; Szewczyk & Federmeier, 2022). For example, if a reader is provided the sentential context of, *The children went outside to fly a*…, the context builds an expectation for the word *kite,* and the N400 is reduced in amplitude for that contextually congruent word relative to a word that was not supported by the preceding context, like *sock* (see Figure 2). N400 responses are also studied in other types of word processing paradigms (e.g., in priming, with a reduction in amplitude for a semantically primed or repeated word).

In L2 research, the N400 has been used to investigate a wide array of linguistic processes. It has been harnessed to examine questions related to vocabulary acquisition and development (e.g., Pu et al., 2016), influence of proficiency on semantic processing (e.g., Moreno & Kutas, 2005), effects of L2 learning on first language (e.g., Bice & Kroll, 2015), sensitivity to L2 phonology (e.g., Sebastian-Gallés et al., 2006), predictive processing (see Kaan, this volume), and many other topics.

The P600

The N400 and its connection to semantic processing of information is often contrasted with another ERP component identified in language processing studies, the P600. This component is positive-going and is typically reported around 600ms. It is utilized by language researchers due to its sensitivity to syntactic properties of sentence/phrase processing. Namely, when a word is encountered that poses challenges to syntactic processing, a P600 often results (Osterhout & Holcomb, 1992). The challenge might be a syntactic violation (e.g., *The cat was/were purring*), or it might be brought about by other syntactic difficulties, such as a so-called garden path sentence (Osterhout et al., 1994).

The functional significance of the P600 is an active area of research (see Sassenhagen & Bornkessel-Schlesewsky, 2015; Sassenhagen et al., 2014). Although originally identified as an index specifically of syntactic processing, there is substantial evidence that it has a more general functional significance and falls under an umbrella of effects, including the late positive component (LPC) and/or the P3 (see the section The P300 below), which is related to updating of context in memory and is particularly sensitive to probability (for review, see Leckey & Federmeier, 2020).

The P600 has been used frequently in L2 work to test the syntactic knowledge and sensitivity of multilinguals in their languages (for reviews, see Alemán Bañon et al., this volume; Biondo et al., this volume). A core question of interest in L2 research is how learners process syntax in their L2, and the P600 has proven very informative at testing hypotheses in this domain. Some influential work in this domain has used ERPs and the P600 to demonstrate that there are substantial individual differences in L2 grammatical processing across individuals (Tanner et al., 2012; Tanner & Van Hell, 2014). In addition to the P600, a component called the left anterior negativity (LAN), has been more specifically associated with morphosyntactic processing (e.g., verb agreement) and is sometimes also reported (e.g., Dowens et al., 2010) and is not considered further here.

Paradigms and Components for Investigating General Cognitive Processes

In addition to the language-related components reviewed above, there are also ERP components that can be leveraged to explore other aspects of L2 learning and bilingual processing. The list of components that might be mentioned here is long. Thus, we focus on just three components that have been used effectively in L2 research: the N2, the P300, and the error-related negativity (ERN).

The N2

The N2 is so-named because it is the second negative peak in the ERP waveform, occurring after an earlier N1. It is sometimes called the N200,”, as its peak is typically observed anywhere between 200 and 350ms after stimulus onset. The N2 can be categorized into various subtypes according to its distribution and the conditions under which it is elicited (for a thorough review, see Folstein & Van Petten, 2007). In bilingual and L2 research, it is typically the anterior N2 (or N2b) that is of primacy interest. In this case, the N2 effect is a negative-going deflection that is larger when participants withhold a response compared to when they provide a response during go/no-go tasks.

The N2’s utility in go/no-go tasks is a natural fit for many techniques used to investigate executive function and language switching in bilingual research (e.g., oddball, flanker, or Stroop tasks; see Moreno et al., 2014, for an example study). By comparing N2 effects between conditions or groups, researchers can make inferences about the effects of bilingualism on executive function during language selection or other cognitively demanding tasks (for more on cognitive control in L2 neurocognition, see Guo & Ma, this volume).

The P300

The N2 is often followed by a P300 (P3b).[1] The P300 is a positive-going deflection that is largest over parietal electrodes and is elicited during active stimulus evaluation and categorization processes. The nature of a task and its stimuli play a large role in determining the latency of the P300 (see discussion in Luck, 2014, chapter 3). So, despite the P300 label, it is not necessarily the case that its peak will be centered at 300ms. One of its most critical features for experimental design is that the P300 is sensitive to the probability of a *task-relevant* property of a stimulus such that it will be larger for rare (less probable) stimuli than for frequent (more probable) stimuli (Polich, 2007). Thus, stimuli are often carefully controlled to occur with equal frequency rather than risk contamination from P300-related brain responses to differences in probability of occurrence.

Although the P300 is among the most studied of ERP components, its specific functional significance remains a matter of debate (for reviews, see Polich, 2007, 2012). However, the wealth of P300 studies has provided a detailed understanding of the circumstances under which the P300 will be elicited and the manipulations that typically modulate its latency (e.g., slower when categorization is challenging) and amplitude (e.g., larger when evaluation is easier and full attention is captured). Thus, for L2 questions related to categorization or evaluation speed and/or difficulty, the P300 is a strong candidate measure as its sensitivities are well-documented across domains. In L2 research, the P300 has not been used to its full potential in this regard, and its utilization has been mostly limited to the context of go/no-go tasks in conjunction with the N2 (e.g., Moreno et al., 2014).

The Error-Related Negativity (ERN)

The ERN is a negative-going deflection that peaks approximately 50ms after the onset of a response and is larger for incorrect than correct responses (Falkenstein et al., 1991; Gehring et al., 1993). The ERN is maximal at frontocentral electrodes. An important difference between the ERN and other components reviewed so far is that the ERN is tied to the onset of a participant's *response*, rather than to stimulus onset. This introduces some additional considerations during processing and analysis.

While the ERN has not been used often in L2 research, it has potential to shed light on many otherwise difficult to assess processes. For example, Sebastian-Gallés et al. (2006) examined the ERN in native Spanish speaking learners of L2 Catalan who performed with low accuracy on a lexical decision task that required them to detect Catalan-specific vowels in order to correctly reject pseudowords. Examination of the ERN indicated that participants were not just inaccurate on these decisions, but also insensitive to their inaccuracy. That is, there was no ERN despite their errors. This suggested participants typically did not detect the critical Catalan vowel changes at all. Other L2 studies have used the ERN to measure aspects of bilingual cognitive control (Morales et al., 2015), or learners' certainty about grammatical gender in their L2 (Bultena et al., 2020; Bultena, this volume).

Example Studies

Using the MMN to Examine Individual Differences in L2

Diaz et al. (2008) illustrates how the MMN might be utilized in L2 research. The authors asked whether differences in how well people perceived novel L2 sounds was related to either their basic psychoacoustic perceptual abilities, or their speech-specific perceptual abilities. Participants were first language Spanish speakers who had previously been classified as "good" or "poor" L2 perceivers. This designation was based on their performance on a set of three behavioral tasks targeting Catalan vowel contrasts that were difficult for Spanish speakers to perceive. To test this, EEG was recorded from participants in both groups as they completed a passive oddball listening task, and, at the same time, watched a silent movie. In three acoustic blocks, participants heard simple pure tones with deviants that varied in either pitch, duration, or the order of tones in a two tone sequence. In the phonetic blocks, they heard vowels that contrasted with a familiar Spanish vowel (/o/). In the first language block, deviant stimuli were another Spanish vowel (/e/); in the L2 block, deviants were a Finnish vowel (/ö/). Results suggested that for all acoustic contrasts—even the difficult ones—both good and poor perceivers had comparable MMN responses. However, for the phonetic contrasts, good perceivers had stronger MMNs for both Spanish and Finnish contrasts. The authors interpreted these results as evidence that individual differences in phonetic (rather than acoustic) perception abilities were likely responsible for different outcomes when learning difficult L2 contrasts.

Using the N400 to Detect Early stages of Lexical Learning

McLaughlin et al. (2004) illustrates how the N400 can be used to measure developing L2 sensitivity for new words. Students in a beginning L2 French class, and a control group of participants who were not learning French, completed a lexical decision task with priming. Targets were either real words or pronounceable nonwords; for real word targets, primes were either semantically related (chien *dog* – chat *cat*) or unrelated (maison *home* – soif *thirst*). Typically, larger N400 responses would be expected when people read nonwords compared to real words, or when a target is unrelated rather than related to a semantic prime. All participants completed the task at three separate times during the French course, with EEG recorded during each session. Across testing times, control participants' neural and behavioral data never showed any effect of lexicality nor of target relatedness. In contrast, in their initial testing session after only 14 hours of French classes, L2 French learners began to display N400 effects to pseudowords even though their behavioral judgements were essentially at chance. In later testing sessions, French learners' N400 effects for lexicality and target relatedness grew, suggesting a development of semantic sensitivity to the words in French as they learned it. This study demonstrates the potential for ERPs to capture L2 sensitivity that behavioral methods might miss, and also to track lexical development longitudinally.

Pros and Cons/Limitations of the ERP Method

Advantages of the ERP method

While both ERPs and typical reaction time measures can provide millisecond information, ERPs provide several major benefits to the researcher over and above reaction times. Because they pair high temporal resolution with information about changes in amplitude, ERPs allow us to evaluate participant responses at multiple points in time. For instance, changes in amplitude during the first few hundred milliseconds may give us insight into early perceptual processes, while later portions of the response might reflect more controlled decisions. Furthermore, by leveraging accumulated knowledge about the functional significance of specific ERP components, we can make interpretations that link responses to specific cognitive processes. Finally, another major benefit of ERPs is that they can often capture responses that would be impossible to observe behaviorally. Many components (e.g. the MMN, the N400, see above) can be elicited with tasks that require no overt response.

Constraints on the Use the ERP Method

Although ERPs have much to offer L2 researchers, they do come with their own set of constraints and challenges. One practical constraint is that clean measurement of EEG requires participants to remain relatively motionless. The constraints on physical movement naturally make research on language production more difficult, though studies have examined pre- or post-utterance processes (e.g., in a delayed reading aloud paradigm, Fischer-Baum et al., 2014), or even unverbalized responses. Similarly, paradigms that require large amounts of eye-movement (e.g., visual world) will be more difficult to analyze (though some methods are now incorporating simultaneous use of EEG and eye-tracking, which allows for the recovery of the neural signal, e.g., Plöchl et al., 2012).

As alluded to earlier, ERPs are not the typical method of choice for those who want to understand the neural localization of cognitive processes. Because of the interference of physical matter in the brain and skull, pinpointing the origins of electrical activity in the brain is not straightforward (Kirschstein & Köhling, 2009). This is known as the *inverse problem* (Cohen, 2017). For instance, just because electrophysiological activity is detected on the left side of the scalp does not mean that the neural generators of that activity were necessarily on the left side of the brain. With effort and the right equipment (high-density electrode systems), it is possible to improve upon the source localization capabilities of typical EEG systems. In general, researchers who want to identify the sources of brain activity use other methods, such as MEG or MRI (see Kousaie & Klein, this volume, and Rossi et al., this volume), which provide much better spatial precision.

While the rich multi-dimensional data generated by ERP experiments is one of its strengths, it also poses challenges. The various stages of data processing and statistical analysis require many decisions on the part of the researcher (e.g., what events should be used for time-locking? what electrodes should be included/excluded? *and many more*). The choices researchers make are not neutral and can lead to unintentional biases in results (Luck & Gaspelin, 2017). This makes standards for processing and reporting EEG studies particularly important (Keil et al., 2014).

Finally, despite our enthusiasm for ERPs in L2 research, we also want to offer a word of caution. In an ideal world, ERPs would form a direct link between participant responses and the linguistic or cognitive processes that L2 researchers want to measure. It would be great if we could use specific components to measure phonology, semantics, syntax, and pragmatics, or to cleanly differentiate between implicit and explicit knowledge. But although it is attractive to think about components in this way, it is not accurate. For instance, although the N400 has great utility for investigating many linguistic phenomena, it is not a language response *per se* (Kutas & Federmeier, 2011). By constructing experiments with care, we can make compelling inferences about linguistic knowledge based on ERPs, but we cannot (as of yet) directly measure that knowledge. The "ERPology" of determining what components are linked to which cognitive processes is a field of research on its own, and debates about the nature of most components of interest to L2 researchers are far from settled.

Innovations and Future Directions

ERP methods are always evolving, especially as availability of equipment and increases in computing power make the technique more accessible and available. Here we briefly note some of the recent innovations that could be applied in L2 research.

Statistical Advances

As noted earlier, a major challenge in ERP research is controlling all the small decisions researchers can make which ultimately can lead to different outcomes, and increase the likelihood of finding spurious statistically significant effects. For ERP data analysis, mass univariate tests may provide a solution that takes some of the decisions out of researcher hands and also allows for discovery of effects outside pre-defined windows of interest (Fields & Kuperberg, 2020; Groppe et al., 2011a, 2011b; for pitfalls to avoid, see Sassenhagen & Draschkow, 2019). For mixed-effects regression modeling of ERPs, new tools and tutorials are regularly being developed for a variety of platforms (e.g., *R*: Tremblay & Newman, 2015; *MATLAB*: Ehinger & Dimigen, 2019; *Python*: Urbach & Portnoy, 2021).

EEG for naturalistic speech

ERP methods have typically used isolated sentences or words to investigate language processing. The application of modern statistical and computational techniques to continuous EEG (and MEG) signals has opened the possibility of using narrative stories (e.g., audio books) and other forms of naturally occurring speech (Alday, 2019; Hamilton & Huth, 2020). The ability to use more ecologically valid stimuli could open up new avenues for research of L2 speech comprehension.

Visual Half-field Approach to Hemispheric Processing

Although researchers interested in localizing brain function during cognitive activities most frequently use fMRI, there is a technique of potential interest for L2 research with ERPs that does permit for inferences about localized functions to be made. When items are presented to only one visual field, and the eyes are kept centered, then the visual information is first passed to the contralateral hemisphere. The receptive hemisphere is afforded not just earlier access to the information, but also better-quality representation of the stimulus than can be gained through the brain's delayed and incomplete interhemispheric communication mechanisms. This lateralized presentation design enables researchers to examine left-hemisphere and right-hemisphere biased processing of information (e.g., Wlotko & Federmeier, 2013). For example, instead of presenting critical words of interest to the center of the screen when recording EEG/ERPs, researchers can present them a few centimeters to the right or left, and then they can learn if the right hemisphere processes these L2 words differently than the left hemisphere (for a behavioral example, see Cieślicka & Heredia, 2011).

Further Readings

In addition to the relevant chapters in the present volume, this article provides a general review of ERPs in L2 research.

Steinhauer, K. (2014). Event-related potentials (ERPs) in second language research: A brief introduction to the technique, a selected review, and an invitation to reconsider critical periods in L2. *Applied Linguistics*, *35*(4), 393–417. https://doi.org/10.1093/applin/amu028

Steve Luck's book is an authoritative but also very approachable guide to understanding and conducting ERP research.

Luck, S. J. (2014). *An Introduction to the Event-Related Potential Technique* (2nd ed.). MIT Press.

The ERP CORE (Compendium of Open Resources and Experiments) provides free access to data and code for use in training and honing ERP methods.

Kappenman, E. S., Farrens, J. L., Zhang, W., Stewart, A. X., & Luck, S. J. (2021). ERP CORE: An open resource for human event-related potential research. *NeuroImage*, *225*, 117465. https://doi.org/10.1016/j.neuroimage.2020.117465

These authors provide authoritative guidelines for reporting of ERPs in publications.

Keil, A., Debener, S., Gratton, G., Junghöfer, M., Kappenman, E. S., Luck, S. J., Luu, P., Miller, G. A., & Yee, C. M. (2014). Committee report: Publication guidelines and recommendations for studies using electroencephalography and

magnetoencephalography: Guidelines for EEG and MEG. *Psychophysiology*, *51*(1), 1–21. https://doi.org/10.1111/psyp.12147

Acknowledgements

Eric Pelzl's contribution to this chapter was made possible in part by the National Science Foundation under NSF SBE fellowship 2004279. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

References

Alday, P. M. (2019). M/EEG analysis of naturalistic stories: A review from speech to language processing. *Language, Cognition and Neuroscience*, *34*(4), 457–473. https://doi.org/10.1080/23273798.2018.1546882

Alemán Bañon, J., Fiorentino, R., & Gabriele, A. (this volume). The neurolinguistics of the L2 syntactic system. In K. Morgan-Short and J. G. van Hell (Eds.), *The Routledge Handbook of Second Language Acquisition and Neurolinguistics*. Routledge.

Bice, K., & Kroll, J. F. (2015). Native language change during early stages of second language learning. *NeuroReport*, *26*(16), 966. https://doi.org/10.1097/WNR.0000000000000453

Biondo, N., Molinaro, N., & Mancini, S. (this volume). The neurolinguistics of the L2 morphological system: the role of grammar-related and speaker-related factors. In K. Morgan-Short and J. G. van Hell (Eds.), *The Routledge Handbook of Second Language Acquisition and Neurolinguistics*. Routledge.

Brandmeyer, A., Desain, P. W. M., & McQueen, J. M. (2012). Effects of native language on perceptual sensitivity to phonetic cues. *NeuroReport*, *23*(11), 653–657. https://doi.org/10.1097/WNR.0b013e32835542cd

Bultena, S. (this volume). Feedback in L2 neurocognition. In K. Morgan-Short and J. G. van Hell (Eds.), *The Routledge Handbook of Second Language Acquisition and Neurolinguistics*. Routledge.

Bultena, S., Danielmeier, C., Bekkering, H., & Lemhöfer, K. (2020). The Role of conflicting representations and uncertainty in internal error detection during L2 learning. *Language Learning*, *70*(S2), 75–103. https://doi.org/10.1111/lang.12401

Cieślicka, A. B., & Heredia, R. R. (2011). Hemispheric asymmetries in processing L1 and L2 idioms: Effects of salience and context. *Brain and Language*, *116*(3), 136-150. https://doi.org/10.1016/j.bandl.2010.09.007

Chandrasekaran, B., Krishnan, A., & Gandour, J. T. (2007). Mismatch negativity to pitch contours is influence by language experience. *Brain Research*, *1128*, 148–156. https://doi.org/10.1016/j.brainres.2006.10.064

Cohen, M. X. (2017). Where does EEG come from and what does it mean? *Trends in Neurosciences*, *40*(4), 208-218. https://doi.org/10.1016/j.tins.2017.02.004

Diaz, B., Baus, C., Escera, C., Costa, A., & Sebastian-Gallés, N. (2008). Brain potentials to native phoneme discrimination reveal the origin of individual differences in learning the sounds of a second language. *Proceedings of the National Academy of Sciences*, *105*(42), 16083–16088. https://doi.org/10.1073/pnas.0805022105

Dowens, M. G., Vergara, M., Barber, H. A., & Carreiras, M. (2010). Morphosyntactic processing in late second-language learners. *Journal of Cognitive Neuroscience*, *22*(8), 1870-1887. https://doi.org/10.1162/jocn.2009.21304

Ehinger, B. V., & Dimigen, O. (2019). Unfold: An integrated toolbox for overlap correction, non-linear modeling, and regression-based EEG analysis. *PeerJ*, *7*, e7838. https://doi.org/10.7717/peerj.7838

Falkenstein, M., Hohnsbein, J., Hoormann, J., & Blanke, L. (1991). Effects of crossmodal divided attention on late ERP components. II. Error processing in choice reaction tasks. *Electroencephalography and Clinical Neurophysiology*, *78*(6), 447–455. https://doi.org/10.1016/0013-4694(91)90062-9

Federmeier, K. D. (2022). Connecting and considering: Electrophysiology provides insights into comprehension. *Psychophysiology*, *59*(1). https://doi.org/10.1111/psyp.13940

Federmeier, K. D., & Kutas, M. (2001). Meaning and modality: Influences of context, semantic memory organization, and perceptual predictability on picture processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 27*(1), 202–224. https://doi.org/10.1037/0278-7393.27.1.202

Fields, E. C., & Kuperberg, G. R. (2020). Having your cake and eating it too: Flexibility and power with mass univariate statistics for ERP data. *Psychophysiology, 57*(2). https://doi.org/10.1111/psyp.13468

Fischer-Baum, S., Dickson, D. S., & Federmeier, K. D. (2014). Frequency and regularity effects in reading are task dependent: evidence from ERPs. *Language, Cognition and Neuroscience*, 29(10), 1342-1355. https://doi.org/10.1080/23273798.2014.927067

Folstein, J. R., & Van Petten, C. (2007). Influence of cognitive control and mismatch on the N2 component of the ERP: A review. *Psychophysiology*, *45*(1), 152-170. https://doi.org/10.1111/j.1469-8986.2007.00602.x

Gehring, W. J., Goss, B., Coles, M. G. H., Meyer, D. E., & Donchin, E. (1993). A neural system for error detection and compensation. *Psychological Science, 4*(6), 385–390. https://doi.org/10.1111/j.1467-9280.1993.tb00586.x

Groppe, D. M., Urbach, T. P., & Kutas, M. (2011a). Mass univariate analysis of event-related brain potentials/fields I: A critical tutorial review. *Psychophysiology*, *48*(12), 1711–1725. https://doi.org/10.1111/j.1469-8986.2011.01273.x

Groppe, D. M., Urbach, T. P., & Kutas, M. (2011b). Mass univariate analysis of event-related brain potentials/fields II: Simulation studies. *Psychophysiology*, *48*(12), 1726–1737. https://doi.org/10.1111/j.1469-8986.2011.01272.x

Guo, M., & Ma, F. (this volume). Cognitive control in L2 neurocognition. In K. Morgan-Short and J. G. van Hell (Eds.), *The Routledge Handbook of Second Language Acquisition and Neurolinguistics*. Routledge.

Hamilton, L. S., & Huth, A. G. (2020). The revolution will not be controlled: Natural stimuli in speech neuroscience. *Language, Cognition and Neuroscience*, *35*(5), 573–582. https://doi.org/10.1080/23273798.2018.1499946

Holcomb, P. J., & Neville, H. J. (1990). Auditory and visual semantic priming in lexical decision: A comparison using event-related brain potentials. *Language and Cognitive Processes*, *5*(4), 281–312. https://doi.org/10.1080/01690969008407065

Kaan, E. (this volume). The neurocognition of prediction in L2 processing and learning. In K. Morgan-Short and J. G. van Hell (Eds.), *The Routledge Handbook of Second Language Acquisition and Neurolinguistics*. Routledge.

Kappenman, E. S., Farrens, J. L., Zhang, W., Stewart, A. X., & Luck, S. J. (2021). ERP CORE: An open resource for human event-related potential research. *NeuroImage, 225*, 117465. https://doi.org/10.1016/j.neuroimage.2020.117465

Keil, A., Debener, S., Gratton, G., Junghöfer, M., Kappenman, E. S., Luck, S. J., Luu, P., Miller, G. A., & Yee, C. M. (2014). Committee report: Publication guidelines and recommendations for studies using electroencephalography and magnetoencephalography. *Psychophysiology*, *51*(1), 1–21. https://doi.org/10.1111/psyp.12147

Kirschstein, T., & Köhling, R. (2009). What is the source of the EEG? *Clinical EEG and Neuroscience, 40*(3), 146-149. https://doi.org/10.1177/155005940904000305

Kousaie, S., & Klein, D. (this volume). Using functional neuroimaging to investigate L2. In K. Morgan-Short and J. G. van Hell (Eds.), *The Routledge Handbook of Second Language Acquisition and Neurolinguistics*. Routledge.

Kutas, M. (1987). Event-related brain potentials (ERPs) elicited during rapid serial visual presentation of congruous and incongruous sentences. *Electroencephalography and Clinical Neurophysiology*, *40*(Supplement), 406-411.

Kutas, M., & Federmeier, K. D. (2011). Thirty years and counting: Finding meaning in the N400 component of the event-related brain potential (ERP). *Annual Review of Psychology*, *62*, 621–647. https://doi.org/10.1146/annurev.psych.093008.131123

Kutas, M., & Hillyard, S. A. (1980). Reading senseless sentences: Brain potentials reflect semantic incongruity. *Science*, *207*, 203–205. https://doi.org/10.1126/science.7350657

Leckey, M., & Federmeier, K. D. (2020). The P3b and P600(s): Positive contributions to language comprehension. *Psychophysiology, 57*, e13351. https://doi.org/10.1111/psyp.13351

Luck, S. J. (2014). *An Introduction to the Event-Related Potential Technique* (2nd ed.). MIT Press.

Luck, S. J., & Gaspelin, N. (2017). How to get statistically significant effects in any ERP experiment (and why you shouldn't): How to get significant effects. *Psychophysiology*, *54*(1), 146–157. https://doi.org/10.1111/psyp.12639

Luck, S. J., & Kappenman, E. S. (Eds.). (2012). *The Oxford handbook of event-related potential components*. Oxford University Press. https://doi.org/10.1093/oxfordhb/9780195374148.001.000

McLaughlin, J., Osterhout, L., & Kim, A. (2004). Neural correlates of second-language word learning: Minimal instruction produces rapid change. *Nature Neuroscience*, *7*(7), 703-704. https://doi.org/10.1038/nn1264

Morales, J., Yudes, C., Gómez-Ariza, C. J., & Bajo, M. T. (2015). Bilingualism modulates dual mechanisms of cognitive control: Evidence from ERPs. *Neuropsychologia*, *66*, 157–169. https://doi.org/10.1016/j.neuropsychologia.2014.11.014

Moreno, E. M., & Kutas, M. (2005). Processing semantic anomalies in two languages: An electrophysiological exploration in both languages of Spanish–English bilinguals. *Cognitive Brain Research*, *22*(2), 205-220. https://doi.org/10.1016/j.cogbrainres.2004.08.010

Moreno, S., Wodniecka, Z., Tays, W., Alain, C., & Bialystok, E. (2014). Inhibitory control in bilinguals and musicians: Event related potential (ERP) evidence for experience-specific effects. *PloS One*, *9*(4), e94169. https://doi.org/10.1371/journal.pone.0094169

Mottarella, M. & Prat, C. S. (this volume). Using frequency-based encephalography to investigate L2. In K. Morgan-Short and J. G. van Hell (Eds.), *The Routledge Handbook of Second Language Acquisition and Neurolinguistics*. Routledge.

Näätänen, R., Kujala, T., & Light, G. (2019). The mismatch negativity (MMN): An introduction. In R. Näätänen, T. Kujala, & G. Light (Eds.), *The Mismatch Negativity* (pp. 1–40). Oxford University Press. https://doi.org/10.1093/oso/9780198705079.003.0001

Nunez, M. D., Nunez, P. L., & Srinivasan, R. (2016). Electroencephalography (EEG): neurophysics, experimental methods, and signal processing. In Ombao, H., Linquist, M., Thompson, W. & Aston, J. (Eds.), *Handbook of Neuroimaging Data Analysis* (pp. 175-197), Chapman & Hall/CRC. https://doi.org/10.13140/rg.2.2.12706.63687

Osterhout, L., & Holcomb, P. J. (1992). Event-related brain potentials elicited by syntactic anomaly. *Journal of Memory and Language*, *31*, 785–806. https://doi.org/10.1016/0749-596X(92)90039-Z

Osterhout, L., Holcomb, P. J., & Swinney, D. A. (1994). Brain potentials elicited by garden-path sentences: Evidence of the application of verb information during parsing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *20*(4), 786. https://doi.org/10.1037/0278-7393.20.4.786

Peltola, M. (2003). Native and foreign vowel discrimination as indexed by the mismatch negativity (MMN) response. *Neuroscience Letters*, *352*(1), 25-28. https://doi.org/10.1016/S0304-3940(03)00997-2

Plöchl, M., Ossandón, J. P., & König, P. (2012). Combining EEG and eye tracking: Identification, characterization, and correction of eye movement artifacts in electroencephalographic data. *Frontiers in Human Neuroscience*, *6*. https://doi.org/10.3389/fnhum.2012.00278

Polich, J. (2007). Updating P300: An integrative theory of P3a and P3b. *Clinical Neurophysiology*, *118*(10), 2128–2148. https://doi.org/10.1016/j.clinph.2007.04.019

Polich, J. (2012). Neuropsychology of P300. In S. J. Luck & E. S. Kappenman (Eds.), *The Oxford Handbook of Event-related Potential Components* (pp. 159–188). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780195374148.001.000

Pu, H., Holcomb, P. J., & Midgley, K. J. (2016). Neural changes underlying early stages of L2 vocabulary acquisition. *Journal of Neurolinguistics*, *40*, 55-65. https://doi.org/10.1016/j.jneuroling.2016.05.002

Rossi, E., Voits, T., & DeLuca, V. (this volume). Using structural neuroimaging to investigate L2. In K. Morgan-Short and J. G. van Hell (Eds.), *The Routledge Handbook of Second Language Acquisition and Neurolinguistics*. Routledge.

Sassenhagen, J., & Bornkessel-Schlesewsky, I. (2015). The P600 as a correlate of ventral attention network reorientation. *Cortex*, *66*, A3–A20. https://doi.org/10.1016/j.cortex.2014.12.019

Sassenhagen, J., & Draschkow, D. (2019). Cluster-based permutation tests of MEG/EEG data do not establish significance of effect latency or location. *Psychophysiology*, *56*(6), e13335. https://doi.org/10.1111/psyp.13335

Sassenhagen, J., Schlesewsky, M., & Bornkessel-Schlesewsky, I. (2014). The P600-as-P3 hypothesis revisited: Single-trial analyses reveal that the late EEG positivity following linguistically deviant material is reaction time aligned. *Brain and Language*, *137*, 29–39. https://doi.org/10.1016/j.bandl.2014.07.010

Sebastian-Gallés, N., Rodríguez-Fornells, A., de Diego-Balaguer, R., & Díaz, B. (2006). First- and second-language phonological representations in the mental lexicon. *Journal of Cognitive Neuroscience*, *18*(8), 1277–1291. https://doi.org/10.1162/jocn.2006.18.8.1277

Steinhauer, K. (2014). Event-related potentials (ERPs) in second language research: A brief introduction to the technique, a selected review, and an invitation to reconsider critical periods in L2. *Applied Linguistics*, *35*(4), 393–417. https://doi.org/10.1093/applin/amu028

Szewczyk, J. M., & Federmeier, K. D. (2022). Context-based facilitation of semantic access follows both logarithmic and linear functions of stimulus probability. *Journal of Memory and Language*, *123*, 104311. https://doi.org/10.1016/j.jml.2021.104311

Tanner, D., Mclaughlin, J., Herschensohn, J., & Osterhout, L. (2012). Individual differences reveal stages of L2 grammatical acquisition: ERP evidence. *Bilingualism: Language and Cognition*, 16(02), 367–382. https://doi.org/10.1017/S1366728912000302

Tanner, D., & Van Hell, J. G. (2014). ERPs reveal individual differences in morphosyntactic processing. *Neuropsychologia*, 56, 289–301. https://doi.org/10.1016/j.neuropsychologia.2014.02.002

Tremblay, A., & Newman, A. J. (2015). Modeling nonlinear relationships in ERP data using mixed-effects regression with R examples: Modeling using mixed-effects regression. *Psychophysiology*, *52*(1), 124–139. https://doi.org/10.1111/psyp.12299

Urbach, T., & Portnoy, A. (2021). fitgrid: A Python package for multi-channel event-related time series regression modeling. *Journal of Open Source Software*, *6*(63), 3293. https://doi.org/10.21105/joss.03293

Van Petten, C., Coulson, S., Rubin, S., Plante, E., & Parks, M. (1999). Time course of word identification and semantic integration in spoken language. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *25*(2), 394–417. https://doi.org/10.1037/0278-7393.25.2.394

Wlotko, E. W., & Federmeier, K. D. (2013). Two sides of meaning: The scalp-recorded N400 reflects distinct contributions from the cerebral hemispheres. *Frontiers in Psychology*, *4*, 181. https://doi.org/10.3389/fpsyg.2013.00181

Wlotko, E. W., & Federmeier, K. D. (2015). Time for prediction? The effect of presentation rate on predictive sentence comprehension during word-by-word reading. *Cortex*, *68*, 20-32. https://doi.org/10.1016/j.cortex.2015.03.014

Wu, Y. J., & Thierry, G. (2010). Chinese–English bilinguals reading English hear Chinese. *Journal of Neuroscience*, *30*(22), 7646-7651. https://doi.org/10.1523/JNEUROSCI.1602-10.2010

[1] Here, we use P300 to refer to the P3b, not the frontally distributed P3a component that is elicited under different conditions and has its own response properties (Polich, 2007).

Figure 2.1.

Steps in EEG Data Acquisition and Processing

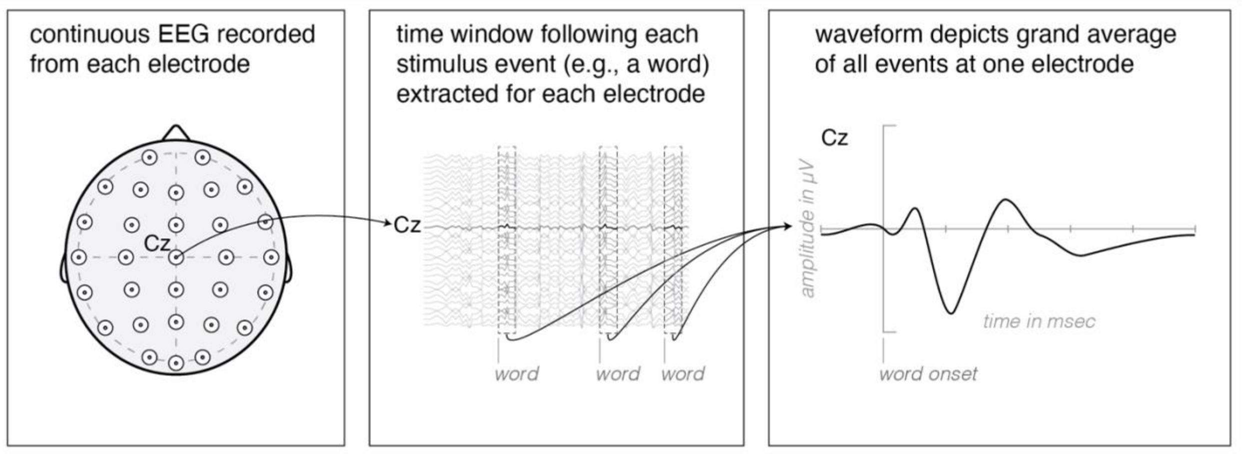

Figure 2.2.

The N400 Effect