THE USE OF BRAIN ELECTROPHYSIOLOGY TECHNIQUES TO STUDY LANGUAGE: A BASIC GUIDE FOR THE BEGINNING CONSUMER OF ELECTROPHYSIOLOGY INFORMATION

Dennis L. Molfese, Victoria J. Molfese, and Spencer Kelly

Abstract. **This article provides a basic background for the professional who is interested in utilizing event-related potential (ERP) approaches to study language processes but has little background in or knowledge about the technique. First, a brief history of the emergence of this technology is presented, followed by definitions, a theoretical overview, and a practical guide to conducting ERP studies. The basis for choice of electrode positions, equipment characteristics (e.g., filter settings), and analyses are also discussed. Finally, examples of language studies that utilize this information in a research study are provided.**

DENNIS L. MOLFESE, Ph.D., is Distinguished University Scholar and Professor and Chair, Department of Psychological and Brain Sciences, University of Louisville.
VICTORIA J. MOLFESE, Ph.D., is Ashland/Nystrand Chair and Director, Center for Research in Early Childhood, University of Louisville.
SPENCER KELLY, Ph.D., is assistant professor, Colgate University.

There has been a long-standing interest by researchers and theorists in brain-behavior relations. This interest has led to the development and adaptation of instruments and methodologies that measure the brain's responses for use in studying developmental issues. Techniques such as electroencephalography (EEG), event-related potentials (ERP), and brainstem-evoked response (BSER) all share a common approach to cortical electrophysiology — scalp electrodes are used to detect electrical activity generated by the brain. These techniques can provide insights into brain-behavior developmental issues that complement and supplement information obtained through more traditional behavioral measures.

This article reviews the history of cortical electrophysiological approaches to investigate brain-behavior relations. An overview of the procedure is provided along with rationales for various components of it. We briefly also comment on how these approaches compare to other techniques, such as Positron Emission Tomography (PET), Magnetic Resonance Imaging (MRI), and functional Magnetic Resonance Imaging (fMRI). A final section will describe the current use of ERP techniques in longitudinal research to investigate the relation between brain functioning in infancy and subsequent cognitive and language development in school-aged children, with a focus on reading and reading disability.

What Are Event-Related Potentials?

In order to understand ERPs (event-related potentials) or Eps (evoked potentials) some general background on the electroencephalography or EEG is useful. In the general EEG (electroencephalogram) technique, electrodes attached to the scalp allow physicians to measure the brain's electrical activity or EEG. The idea behind EEG is straightforward. The human body utilizes electricity in its operation, somewhat like cable cars, computers, and CD players use electricity. One major source and processing station of that electricity is the brain, or more specifically, the neurons that make up much of what we call the brain. Every time a group of neurons perform some function, they generate a small amount of electrical energy, which spreads throughout the brain to other neurons. However, a portion of this electrical signal passes through the brain and travels through the skull where it can be measured on the scalp. By positioning electrodes on the scalp's surface, EEG amplifiers magnify the minute electrical discharges (perhaps 5-10 μV, or millionths of a volt) that occur. Such electrical currents on the scalp thus reflect and indicate the activity level of groups of neurons within the brain. The end result is a readout of the ongoing electrical discharges (in the form of continuous brain waves) that were produced during activity (e.g., sleeping, listening to music, reading) over some period of time.

An EEG field at the scalp can only be recorded if large numbers of neurons are active at the same time, and they must all be aligned in the same orientation. If they are not aligned, that is, they are laminar, then the positive and negative fields of differently aligned neurons tend to cancel each other out. Subcortical brain structures, such as the basal ganglia, may be highly active electrically, but do not appear to contribute to the scalp EEG because they are not aligned with cortical brain structures. Therefore, it is generally assumed that the great majority of the electrical fields contributing to the scalp EEG represent cortical electrical sources. The general belief through two centuries is that measurements of such currents can provide insights not only about the basic neurophysiology of the brain but also behavioral information concerning how and in what manner the brain is involved in sensing, thinking about, and interacting physically and cognitively with the environment throughout development.

One limitation of the EEG is that because it is continuous and ongoing, it is difficult to determine the specific stimuli or events that produce variation or change in the EEG pattern. ERPs, on the other hand, overcome this limitation by simply focusing on a portion (usually about one second) of the ongoing EEG electrical activity that is repeatedly time-locked to the beginning of when a stimulus (e.g., sound, picture) is presented to someone. *Time-locking* refers to the fact that researchers only record the part of an EEG wave that follows the word, sound, or picture stimulus in time. *Repetition* refers to the fact that researchers repeatedly present the same stimulus in order to average out the random and nonstimulus-related background electrical activity that is inherent in the ongoing EEG and does not reflect the brain's response to that stimulus.

Consider the following analogy: Suppose that you were a seismologist who wanted to test an idea about earthquakes by dropping pebbles into a pond. In order to test the impact of a pebble in that pond, you would have to measure the ripples produced *beginning when* the pebble first contacted the water in the pond. You would then continue to record the ripples produced by that impact until they died away. This is the notion of time-locking. However, we know that ponds are already naturally full of ripples from wind and other disturbances (e.g., fish and turtles moving, frogs jumping in, etc.), thus it is difficult to measure the effects of a single pebble drop. In this analogy, the pond's ripples are like ongoing EEG, and the pebbles' ripples are like an ERP. In order to "wash" out (i.e., average out) these background ripples caused by other factors, one would have to sequentially drop many individual pebbles and take the average of each pebble's ripples produced over that time. This kind of repetition leads to "averaging out" information unrelated to the pebble's impact.

This analogy is useful for thinking about how to interpret ERPs. The basic idea behind the ERP methodology is that different stimuli of interest (e.g., words, pictures, thoughts, sounds, etc.) cause different brain waves, just like different-sized and -shaped pebbles cause different waves in a pond. These differences can be used just like any other dependent measure in research on language processing. For example, they can be used like behavioral measures of text comprehension rates, reading time, phonological discrimination, and so on.

Among the many advantages of the technique are its noninvasiveness and the fact that it can be used across the life span with virtually identical procedures. Thus ERP technology has the potential to provide a powerful tool to study changes in brain-behavior relations and functions across the life span.

Research over the past century has demonstrated that the ERP can be used to effectively study both general and specific aspects of the organism's response to eliciting events in the external as well as the internal environment (Molfese, 1978a, 1978b). The ERP can also be used to study an individual's perceptions and decisions during tasks or following a learning situation (Molfese, 1983; Nelson & Salapatek, 1986; Ruchkin, Sutton, Munson, Silver, & Macar, 1981). Given that the EP technique does not require a planned and overt response from which it is recorded, it is particularly well suited for the neuropsychological study of early infant and child language development (Molfese, Freeman, & Palermo, 1975). In addition, as noted above, the ERP can provide very fine temporal information (one ms or less) regarding the brain's response to an eliciting input, such as a speech sound. Finally, the ERP has some gross-level spatial resolution capabilities that permit a basis for speculations concerning the distribution of brain mechanisms that subserve functions such as language.

History

Attempts to record the brain's electrical activity and relate it to behavior date from at least the time of Richard Caton (1875), who recorded evoked potential responses (ERPs) from an electrode placed directly on the surface of a rabbit's exposed brain. Other early brain-recording techniques involved immersion, in which patients were required to place each of their limbs in separate buckets of saline solution. The buckets were used as "electrodes" to detect electrical signals conveyed via cables to an amplifying system that magnified the millionth of a volt signals recorded from the scalp, which could then be penned onto a chart recorder.

By the mid-1920s, plate electrodes were developed that could be applied directly to the skin. Unfortunately, any movement of the plate over the skin, no matter how small, produced large artifacts that distorted or obscured the brain responses. Later, scientists constructed a floater type of electrode that required a conducting paste (electrolyte) to be placed between the skin and the electrode. The electrolyte allowed the small currents on the skin to be more readily transferred to the electrodes and thus recorded. This type of electrode is similar to many of the electrodes in common use today.

Although this latter approach reduced the electrode movement artifacts that often contaminated or obscured the minute evoked potential signal obtained from the contact or plate electrodes, other problems remained that reduced the effectiveness of this technology for studying brain functions. A number of engineering and electrical difficulties continued to limit the potential success of this approach, including problems in finding adequately conductive electrode materials and difficulties in improving the low signal-to-noise ratio (e.g., identifying and isolating the very small brain ERPs from the larger electrical and myographic events generated by other body biosystems — the problem of distinguishing between small pond ripples caused by one pebble from the larger ripples created by wind, rain, and other factors).

In the late 1940s, Dawson (1951) devised a technique to improve the signal-to-noise ratio of the ERP through the use of a capacitance-based computer analogue that summed repetitively elicited event-related potentials. By adding together electrical signals recorded on successive trials, Dawson's device calculated summed event-related potentials that reflected the repetitive information contained in the evoked potential from one time to the next or from trial to trial. These summed or averaged ERPs represented the brain electrical activity common to all the ERPs collected during a recording session. On the other hand, the nonrepetitive signals that reflect random signals or noise failed to contribute systematically to specific portions of the accumulating sum.

The modern EEG and ERP data collection systems offered by many companies today are logical extensions of Dawson's original idea to sum and average event-related potential responses in order to improve the signal-to-noise ratio. Such "low-noise," averaged ERPs provide investigators with the opportunity to see the direct effects of the stimulus and the subject's processing of that stimulus in the brain wave.

An additional issue important to the development and evolution of cortical electrophysiology is the development of analysis techniques useful for evoked potential data. Analysis procedures have developed at an excruciatingly slow pace throughout the past century, as evidenced by the fact that the most widely used methods of data analysis today (e.g., peak amplitude and latency measures) date back to Caton in the late 1800s. However, within the past two decades a number of analyses and ERP technology have emerged, especially since the development and more widespread use of personal computers.

The ERP as an Assessment Tool

The event-related potential (ERP) is a synchronized portion of the ongoing EEG pattern. Basically, evoked potential waveforms are thought to reflect changes in brain activity over time (Rockstroh, Elbert, Birbaumer, & Lutzenberger, 1982). Such differences have traditionally been seen as reflected by changes in the amplitude or height of the wave at different points in its time course or in a change in the latency (time lapsed since stimulus onset) of certain peaks within the ERP. What distinguishes the ERP from the more traditional EEG measure is that the evoked potential is a portion of the ongoing EEG activity of the brain that is time-locked to the onset of a specific event (the stimulus) in the infant's environment. As mentioned, this time-locked feature is at the heart of the strength of the ERP and represents a major advantage over the traditional EEG measure. The ongoing EEG activity reflects a wide range of neural activities related to the myriad neural and body self-regulating systems as well as the various sensory and cognitive functions ongoing in the brain at that time. This intermixing of cognitive, sensory, and other biological signals makes it difficult to separate out one factor from another. On the other hand, because the ERP is time-locked to the onset of an event, researchers are able to evaluate the relationship between the neuroelectrical response and a given event (Callaway, Tueting, & Koslow, 1978; Rockstroh et al., 1982). This relationship can be resolved down to milliseconds or even fractions of a millisecond, if there were physiological processes that operate at this time scale.

ERP waveforms are described visually in a number of ways. One common approach is to identify in some manner the positive and negative peaks (i.e., the point when a portion of the wave reaches its most positive or negative extent or value) that occur in the waveform. This labeling can refer to the sequence in which the peak occurs while at the same time indicating its polarity (i.e., whether it is a "positive" or a "negative" going peak). For example, "N1" would refer to the first negative peak in the waveform while "N2" would refer to the second negative occurring peak. Likewise, "P1" refers to the first positive deflection or peak in the ERP waveform while "P2" refers to the second peak. An alternate and more recent naming scheme for ERP components is to name the positive and negative peaks by their latency (usually defined as the time from stimulus onset). "N100" in this example would refer to the negative peak that occurs 100 ms following stimulus onset. Similarly, "P300" would label the positive peak that occurred 300 ms post-stimulus onset.

Biobehavioral basis of the assessment. The ERP is generally believed to reflect postsynaptic (dendritic) potentials (Allison, Wood, & McCarthy, 1986). Even so, the information recorded at the scalp cannot capture all the generated electrical activity. To reach the scalp, the signals must be produced by fairly extensive sets of activated neurons whose firings must to some extent overlap each other in time. Even so, not all signals reach the scalp for a variety of reasons. It is often difficult to detect a signal because the distance from the cortical regions generating the signal to the scalp is too great relative to the signal's strength. Signals that originate within the brain must travel through a variety of tissues of different densities, conductivity, and composition (e.g., neurons, glial cells, fiber tracts, cerebral spinal fluid, bone, muscle) before they reach the recording electrode placed on the scalp. In addition, the orientation of the cortical columns generating the signal may contribute to whether or not a signal reaches the scalp. If the columns are perpendicular to the scalp, the likelihood of the signal reaching the scalp is good. On the other hand, if the column is parallel to the scalp or at some other angle to it, the signal may not project to the scalp or may project to the scalp some distance away from the electrode that is immediately above it.

The actual ERP signal that is finally detected at the scalp is not an exact and completely stable pattern, reflecting only those discrete neural events directly related to the evoking stimulus, the task, or the subject's state. Clearly, the ERP is only a by-product of the brain's bioelectrical response to such an event, which begins as the stimulus information is transformed by the sensory systems. This signal then progresses through the brainstem into the midbrain, and on upward into the higher centers of the brain. Consequently, the final version of the ERP recorded at the scalp is a composite of a variety of complex factors, only some of which may actually relate directly, or even indirectly, to the variables under manipulation in the experiment.

Choice of electrode placements. The choice of electrode placement on the scalp is an important step in ERP recording. This choice is often driven by hypotheses concerning the relationships between the functioning of different brain regions and the cognitive operations or processes assumed to occur in those areas. Unfortunately, for a variety of reasons, any single scalp electrode does not simply detect information that originates within the brain immediately below it. Instead, each group of neurons creates what is called a "dipole field," which generates positive electricity in one direction (e.g., toward the surface of the cortex) and negative electricity in another. The dipole field can be thought of as a flashlight with two lenses, one pointing up and the other pointing down, each creating a cone of light (actually electricity) that spreads outward until it reaches the surface of the body. Because of this spread of electricity (called "volume conduction"), an electric potential at a given scalp site may not be restricted to nearby brain tissue, but could reflect activity of brain tissue that is far away.

Making matters worse is the fact that the human cortex is highly convoluted (wrinkled). This means that the volume conduction (where the flashlights point) is determined by the wrinkles, which are highly variable from one person to the next. This is why advanced "anatomically constrained" source analysis methods use the subject's actual cortical surface (from the MRI) to estimate how the wrinkled electrical field can be "unfolded" to relate it back to specific regions of cortex.

Another complication is that the volume conduction is changed by the resistivity of brain tissues to passing electrical current. For example, because the skull is thick and hard, it is particularly resistive so that a dipole field "spreads" when it passes through it. Advanced source analysis methods must take this into account, using specific measures or estimates of skull conductivity for each region of the head. This is especially true for infants and young children because their highly variable skulls are incompletely calcified and contain both fontanels and sutures between the skull bones.

What these factors should make clear is that electrodes on the scalp do not necessarily measure electrical activity generated by neurons immediately below them. Rather, scalp electrodes can detect changes in electrical discharges that originate deep within the brain as well as from active areas in other brain regions and the opposite hemisphere.

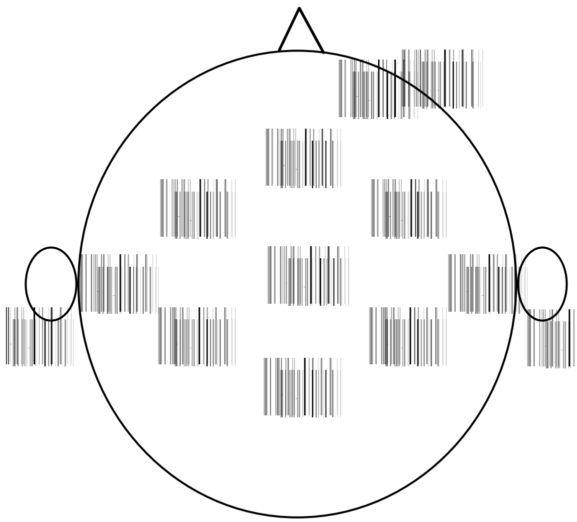
A variety of strategies have been used to select electrode placements sites. Reviews of the ERP and EEG literature suggest that approximately half of the studies use the 10-20 System designed for use with adults and reported by Jasper (1958). This technique relies on proportional measures of the head to determine electrode placements and is useful in attempting to replicate placements done across studies using the same technique. Development of a similar system for infants has been attempted (Blume, Buza, & Okazaki, 1974). However, a number of factors, such as a small sample size and the lack of measurements from both hemispheres from the same infants, limit the usefulness of the Blume et al. approach as a standardized system for electrode placement in infants or children. However, it is clear even from these limited data that the 10-20 System used for electrode placement in adults does not overlie the same cortical regions in young infants. For example, as Blume et al. note, central leads in infants were found to lie over the postcentral gyrus (sensory) whereas such leads were over the precentral gyrus (motor) in adults, two very different functional areas within the brain. In infants the inferior frontal electrode lies inferior to the frontal lobe as opposed to over that area in adults. Additional points of discrepancy were noted between infant and adult placements. These discrepancies further raise issues regarding the legitimacy of comparing recordings between infants, children, and adults.

Placement of electrodes on the individual's head is usually driven by hypotheses concerning the relation between different anatomical brain regions and the cognitive processes assumed to be engaged by the evoking stimulus and the characteristics of the experimental task. For example, the brainstem-evoked response (BSER), which is generally used as a screening technique for sensory processing, is thought to represent brainstem responses that reflect detection of the brief stimuli presented to the participant. This testing procedure involves placing only one active electrode at a central point (C_Z) midway between the left and right sears as well as midway between the upper ridge on the nose (nasion) and the base of the bump at the central back portion of the skull (i.e., the inion).

In research investigating more complex, higher-order processing, such as language processing, electrodes are typically placed over a number of brain regions thought to be actively involved in language perception (i.e., temporal lobes) as well as language integration (e.g., temporal-parietal areas), and language production (i.e., frontal brain regions). An example of these placements is provided in Figure 1, including three central sites: F_Z, (frontal midline), F_L (a left frontal site), F_R (a right frontal site), and six lateral sites over each side of the head: T_3, T_4, C_3, C_4, P_L, and, P_R. Electrode sites with an odd-numbered subscript are positioned over the left side of the head while those with even-numbered subscripts are over the right side. The "L" and "R" subscripts denote non-10-20 electrode sites that the experimenter might utilize to accelerate the placement of electrodes on the scalp or to place electrodes closer to what are believed to be more relevant active processing sites. The former strategy is used by some investigators because some 10-20 electrode locations require several measurements to locate the exact site for placement, a definite disadvantage when testing special populations of children who are not used to remaining motionless for long. With infants or young children, a prolonged electrode application period could ultimately mean that no useful data will be collected. Thus, testing younger participants often demands creativity on the part of the investigator to present ample numbers of stimuli in order to ensure a stable response, reduce the child's anxiety over an unfamiliar situation, minimize fatigue and movements that distort or mask the brain responses, avoid subject fatigue, and keep the child's attention focused on the task over time.

Typical ERP studies place electrodes over bilateral frontal, temporal, central and parietal areas of the brain. This placement is hypothesized to provide information about left- versus right-hemisphere responses to the evoking stimuli and information within each hemisphere concerning functioning of different brain areas. Unfortunately, as noted above, the scalp electrode does not only detect responses from the brain area immediately below the scale location. For example, because of volume conduction, the T_3 electrode site may pick up not only activity produced immediately beneath it in the left-hemisphere temporal region, but also activity produced in other adjacent brain areas. Thus, caution must be used in attributing ERPs recorded at one electrode site to a single area of the brain. Although there are limits to how far one can speculate about the origins of the scalp-recorded ERP signal, a host of computer programs (e.g., BESA), when used in conjunction with theoretical model, offer a means to address questions concerning brain regions responsible for generating such signals.

Figure 1.
Diagram of approximate locations of infant electrode placements for the left side of the head. Corresponding placements were applied to the right side of the head as well.



In infants, with few exception, the number of electrodes used is smaller than in adults for several reasons. Infant tolerance of testing procedures is directly influenced by the time it takes to apply electrodes to their scalps. The more electrodes used, the more time it takes to apply them. For example, because many electrodes are used in the 10-20 System, it requires a long time to set up. However, given the lack of convergence between adult and infant 10-20 Systems noted earlier, investigators have license to modify the system in order to expedite the electrode application process (see Molfese, Morse, & Peters, 1990, for an example).

The use of electrode caps has eliminated some of these time demands. However, traditional caps use blunted needles to abrate the scalp to lower impediances (resistivity) — a practice that causes particular discomfort to infants and children, often resulting in the loss of as much as 50% of the children attempted to be tested (H. Neville, personal communication, January, 2000). Another problem that might preclude the placement of more electrodes on children relates to density. Because infants' heads are smaller, electrodes must be placed closer together compared to adults. If electrodes are placed too close to each other, however, the electrodes themselves might interfere with the recorded signals, thereby creating artifacts. But several technologies offer other solutions. For example, the Geodesic net designed by Dr. Donald Tucker (1991) and his associates, which we use in our own laboratory, offers a means to apply as many as 129 electrodes to the heads of newborn infants in as little as 10 seconds!

The ERP scalp activity recorded at any one electrode is typically referenced to other recording sites. These sites are selected because they are either less electrically active, and consequently of less interest to the investigators (such as the tip of the nose, mastoids, or ear lobes), or are sites on the scalp that may be characterized by comparable but different levels of electrical activity. These latter reference sites are chosen so that the investigator can more directly examine the electrical differences between those recording sites and the other scalp sites. Recent techniques have used a calculated average reference that is based on the average of activity recorded at all electrode sites. A limitation of the average reference method is that it is only accurate when there are a large number of electrodes and they completely cover the head surface, including the face and neck (Junghöfer, Elbert, Tucker, & Braun, 1999).

In addition to scalp and reference electrodes, electrodes are usually placed at supraorbital (i.e., above the middle of one eye over the eye brow) or suborbital (i.e., approximately 2 cm below an eye, on the upper portion of the cheek) and canthal (i.e., to the side of the head that is away from the eye approximately 2-3 cm) positions in relation to one of the participant's eyes to assist in detecting artifacts due to horizontal and vertical eye movements. Such eye movement artifacts or blinks can produce large electrical signals that can distort ERPs recorded at other adjacent electrode sites, even those positioned towards the back of the head (Junghöfer, Elbert, Tucker, & Rockstroh, 2000).

General ERP test procedures. The actual ERP recording procedure involves several steps. First, an individual's head is measured and positions are marked to indicate where electrodes are to be placed. Next, these positions are cleaned with an abrasive such as pumice paste to lower skin impedances, thereby ensuring that the electrodes will be able to conduct a better signal. The abrasive is then removed and a small amount of electrode conducting paste is rubbed onto the scalp. Small disk-shaped electrodes are then filled with the electrode paste and placed on the participant's scalp at these prepared positions. The electrodes are connected via wires to amplifiers that increase the ERP signal by 20,000 to 100,000 times. Given that ERPs are generally very small, on the order of 5 to 10 μV in adults, amplification is needed to provide enough definition of the waveform for further analyses. Amplifiers used in recording systems also contain filters that screen out some of the recording system noise, the ambient electrical noise that surrounds us in our environment (such as the 60 Hz, 120 volt signal that powers our toasters and televisions) as well as the biological background noise that we carry about with us and that the investigator does not want to study. The outputs from these amplifiers are connected to a computer that collects the ERPs from each electrode for each stimulus presented.

Once all the electrodes are in place and connected to the amplifiers and the computer, the stimuli can be presented when the participant is in a reasonably quiet state. Each ERP is made up of a number of time points from stimulus onset until the end of the sampling period. Thus, this time period may range from 0 ms (the point in time when the stimulus begins) until 1,000 or even 2,000 ms after the stimulus onset time. The duration of the ERP is generally up to the investigator but is usually informed by what others have done in similar studies or with comparable subject populations. In addition, investigators can examine the ERP to determine where the variability in the ERP (as evidenced by the standard deviation at each time point) begins to increase and decrease. Usually, there is little variability at the very beginning of the ERP as the stimulus first begins to work its way through the nervous system. There also is a decrease in the size of the standard deviation as the ERP returns to background EEG activity as the time sync between the stimulus and the brain breaks down.

Once the duration of the ERP has been decided, the investigator must decide how frequently to sample the ERP signal across time following the onset of the stimulus. Sampling is necessary because brain waves are analogue (continuous) signals, whereas analyses can be performed only on digital (sampled) signals. As with duration, the decision to select a certain sampling rate can be based on studies investigating similar phenomena or studies of similar populations. Most studies use sampling rates of 4 or 5 ms. If an ERP were sampled at 5 ms intervals, 200 data points would be collected for each ERP during that 1,000 ms period. The 200 points from each of these ERPs from each participant for each electrode can then be submitted to analysis.

Because of the inherent variability in the ERP that results from moment-by-moment changes in an individual's physiology, researchers have a variety of means at their disposal to analyze the collected ERPs. Usually, the ERPs are first recorded to discrete events (e.g., words, pictures, sounds) and then, following artifact rejection, averaged in order to build stable waveforms and improve signal-to-noise ratio. The logic is that the resulting averaged response is more likely to contain the recurring activity that reflects the processing of the stimulus from one time to the next. In contrast, the nonstimulus-related activity that is not time-locked to the onset of the stimulus is expected to average out or be minimized in the averaged waveform of the ERP.

Once averages are obtained, they are subjected to a variety of analysis approaches. Traditionally, the technique of choice has involved amplitude measures taken from various peaks in the waveform. These measures may be made between two adjacent peaks of opposite polarity (e.g., measuring the voltage difference between the most positive peak and immediately preceding or following negative peak), a process referred to as a peak-to-peak measure, or between the average prestimulus baseline signal and a specific positive or negative maximum peak amplitude, a process referred to as a baseline-to-peak measure. Analyses of the ERP are then conducted on the averaged waveforms. These analysis approaches include a range of options such as amplitude and latency (reaction time) measures, area measures, discriminant function procedures, and other multivariate approaches, including principal components analysis, as discussed next.

Multivariate analyses have been found to be useful in a variety of cases for interpretting ERPs. One approach, principal components analysis (PCA), is particularly useful for obtaining information about parts of the brain wave that are similar across subjects. For example, the PCA has produced consistent results in programmatic research across a number of laboratories in studies of cognitive and language issues at different developmental periods (Brown, Marsh, & Smith, 1979; Chapman, McCrary, Bragdon, & Chapman, 1979; Donchin, Teuting, Ritter, Kutas, & Heffley, 1975; Gelfer, 1987; Molfese, 1978a, 1978b; Molfese & Molfese, 1979a, 1979b, 1980, 1985; Ruchkin et al., 1981; Segalowitz & Cohen, 1989). The PCA procedure is similar to a factor analysis except that factors are constructed on the basis of variances rather than correlations (Rockstroh et al., 1982). The PCA procedure is blind to individual experimental conditions and generates the same solution regardless of the order in which the ERPs are entered into it.

The use of PCA to analyze ERPs is fairly straightforward. If a sampling rate of 5 ms is used to collect the ERP over a 1,000 ms period, 200 points will be obtained from each of the ERPs from each participant for each electrode and each stimulus or task condition. The 200 time points serve as the variables in the PCA whereas each ERP itself serves as a case. This collection of ERPs is then transformed into a correlation (or covariance) matrix and the PCA is applied to this matrix. The analysis constructs factors that characterize these areas of variability in the ERPs and account for all or most of the variability in the entire set. The peak for each factor and the area immediately surrounding it in time identifies a temporal region of the ERP where there were changes in amplitude, latency, or slope across some proportion of the ERPs in the data.

A factor score is also generated for each ERP, which indicates how much variability occurred for each ERP submitted to the PCA. These factor scores serve as dependent variables in subsequent analyses that could include discriminant functions (Molfese & Molfese, 1997), regression (Molfese & Molfese, 1986), or analysis of variance (Molfese, Nunez, Seibert, & Ramanaiah, 1976). For example, analyses of variance (ANOVA) have been used to determine if any of the factors (regions of the ERPs) varied systematically as a function of the independent variables in the study (e.g., electrode sites, hemispheres, sex of participant, stimulus conditions). Factor scores can be used in discriminant function analyses to determine whether the scores are useful for discriminating between groups of participants divided according to performance on a specific assessment or evaluation.

Strengths as an assessment tool. ERPs offer a number of strengths as an assessment tool. For example, the procedure can be applied to participants across multiple age groups. Few techniques currently in use can be applied from the newborn through the adulthood period. ERPs are useful for making direct comparisons between infants and adults to address a variety of developmental questions. Although the wave shapes of the ERPs change from infancy to adulthood, one can assess whether brain responses recorded at different ages discriminate reliably between different stimulus, participant, and task conditions obtained concurrently or at different time periods. Moreover, the ERP procedures can be used to obtain response information from participants who have difficulty responding in a normal fashion (as in the case of individuals with brain damage) or who cannot respond because of language or maturity factors (as in the case of young infants and children).

ERPs also are recognized as providing information concerning both between-hemisphere differences and within-hemisphere differences in the brain's electrical activity under specific stimulus conditions. Further, the ERP procedure is useful for providing time-related information. It can indicate the onset of one stimulus relative to another and provide information about the different points in time when such information is detected and processed.

Comparison of ERP strengths and other techniques. Because of their excellent temporal resolution and correlations with specific cognitive/linguistic activities, ERP procedures offer advantages over other brain-imaging procedures such as EEG, brain stem-evoked response (BSER), positron emission tomography (PET), and functional magnetic resonance imaging (fMRI). For example, although the classic EEG measure provides some indication of clinical states such as epileptic seizures, it does not resolve cognitive activities to the level offered by the ERP (Callaway et al., 1978). Thus, while frequency analyses of EEG may indicate attentive or inattentive states (as in the case of alpha activity) or an increase in workload (as in the case of beta activity), it is unable to resolve changes in stimulus parameters, decision making, or short-term memory activity. Likewise, although BSER information can reflect initial sensory detection and brainstem response to very brief evoking stimuli, the temporal duration of the BSER (approximately 10-15 ms) precludes its use for studying longer and later occurring cognitive processes involved in these activities. In addition, although PET and fMRI procedures are good at identifying metabolic brain changes associated with cognitive processes (Shaywitz et al., 1995), they are as yet unable to resolve the temporal order of these processes or the more discrete decisions regarding the processing of stimuli. Moreover, the expense and complexity of setting up an ERP lab is considerably less than the millions of dollars required for the PET, magnetoelectroencephalography (MEG) and fMRI procedures. Finally, the ease of ERP application and testing may be less formidable than other techniques to young participants and their parents.

Constraints on Using the ERP for Assessment

Although ERPs offer many advantages as an assessment technique, there are limits to the interpretation of the results of ERP studies just as there are with other techniques, physiological or behavioral. First, ERP studies share the limitations indigenous to all experimental approaches — one must make a leap from the data obtained in an experiment to interpretation of the data.

Second, although there is something seductive in recording electrical currents thought to originate "directly from the brain," the reality is that the specific origins of these currents and the dynamics that lead to their particular presence at the scalp remain beyond our understanding at this time. The measurement-based placement system reflected by the 10-20 System tried to standardize electrode placement across participants so that placements roughly approximate scalp locations to brain regions. However, for reasons already noted, attributing signals from scalp locations to brain regions is fraught with problems. Further, the scalp electrode does not only detect information that originates immediately below the given electrode position in the brain. Thus, there are limits for how far one can speculate about the brain origins of the scalp-recorded ERP signal. Fortunately, the relationship between ERP signal and cognition and behavior is less tenuous. Indeed, the linkage between the ERPs and specific behaviors is not accidental and can be effectively exploited in carefully designed and executed experimental paradigms as linked to current and later developing behaviors.

A third major limitation concerns the validity of comparing ERP waveform characteristics across different developmental periods. Do similar paradigms used with adult and infant populations tap into the same cognitive or linguistic abilities? The answer is that they most likely do not tap the same abilities, given the large cognitive and linguistic differences that exist between infant and adult populations or even between infants at markedly different times in development. Based on a host of neuroanatomical studies, we know that the brains of infants differ markedly from those of adults in terms of neurogenesis, dendritic development, and myelination, to mention only a few characteristics. Such differences in neural structure state, as well as in the differential development of brain structures across ages, limit our ability to interpret findings from groups that are developmentally disparate. Equally plausible is the alternative explanation that these ERP components tap very different cognitive or physiological mechanisms that generate such similarities. Just as different connectionist models might generate similar behaviors, signals generated within the brain at different stages of development in combination with different cortical densities may produce similar brain responses when recorded at the scalp. Consequently, such similarities (or differences) by themselves do not necessarily mean that the same brain mechanisms are functioning at the same level in theses different populations. To make such a claim, other data must be considered.

Use of ERPs to Predict Language Outcomes

One illustration of the use of ERP techniques can be found in the research of a number of investigators who used this technique to identify later emerging cognitive or language problems. Such research was begun in the late 1960s and early 1970s by Engel and his colleagues (Butler & Engel, 1969; Engel & Fay, 1972; Engel & Henderson, 1972). Butler and Engel (1969) reported the first success in noting correlations between the neonatal visual-evoked potential latencies and later measures related to intelligence. They recorded visual-evoked potentials from 433 newborn infants in response to a series of photic (light) flashes. Although the correlations between Bayley scores at eight months and photic latency were significant, the effects were small and accounted for little of the variance. Jensen and Engel (1971) also reported correlations between neonatal photic latencies and later motor skills (age of walking). When they divided the photic latency response into three regions, they found that infants with the shortest photic latencies were the earliest walkers. Unfortunately, studies attempting to extend the period of predictability into the second year of life and beyond reported limited success (Engel & Fay, 1972; Engel & Henderson, 1973; Henderson & Engel, 1974).

This limited success may have resulted from such studies restricting their analyses to a single early peak or peak latency (usually the N_1 component) in the ERP and using nonlanguage-related stimuli (i.e., a light flash) to elicit the ERPs. Subsequently, researchers speculated that if language-related stimuli were used to elicit the ERPs, and if analyses of the ERP were not confined to a single peak, greater success could be achieved. To this end Molfese and Molfese and their colleagues conducted a number of analyses on different longitudinal studies that measured multiple regions of the ERP collected in response to auditorily presented speech sounds (Molfese, 2000; Molfese & Molfese, 1985, 1986, 1997).

Two examples using ERP data to predict later outcomes will serve as illustrations. One example used the traditional latency measure employed by Engel and his colleagues (Molfese, 2000) while the other used the PCA-ANOVA approach employed by a number of scientists as already described (Molfese, in press).

Example 1. As noted above, Molfese used a traditional latency measure as well as amplitude measures of the newborn infant's brain responses. Moreover, Molfese did not restrict the analyses to a single early ERP peak but used multiple peak latencies in his discriminant function analyses. Further, instead of using a nonlanguage-related stimulus to elicit the brain responses (i.e., a light flash), Molfese used speech syllables to elicit the auditory ERPs. Using linguistic stimuli and analyses based on a more comprehensive sample of brain responses, Molfese found that the neonatally recorded ERPs accurately predicted the reading skills of these children.

Molfese (2000) analyzed ERP data obtained from a total of 48 infants shortly after birth. By eight years of age, this group included 17 dyslexics, 7 poor readers, and 24 controls as described below. Grouping of the children was based on performance on the *Wide Range Achievement Test,* the only scores available for the sample. The dyslexic children at eight years of age had normal full scale IQ (FSIQ) scores (mean FSIQ = 110.0) as measured by the *Wechsler Intelligence Scales for Children-3* (WISC-3; Wechsler, 1991), although their reading scores from the *Wide Range Achievement Test-3* (WRAT; Wilkinson, 1993) were markedly below average (mean = 80.6). The Poor Readers had both low reading scores (mean WRAT = 85.4) and low WISC full scale IQ scores (mean FSIQ = 96.9). Although the FSIQ scores of the Poor Readers and Dyslexics differed from each other, their reading scores did not. The Control children were matched to the full scale IQ scores of the Dyslexic children (mean WISC FSIQ = 111.7), although their WRAT reading scores were higher than those obtained by both the Poor Readers and the Dyslexics (mean WRAT = 103.75).

Auditory event-related potentials (ERPs) were recorded from the left- and right-hemisphere frontal, temporal, and parietal scalp regions (linked ear references) of these 48 infants within 36 hours of birth to a series of two consonant-vowel syllables, /bi/ and /gi/, and nonspeech homologues of these sounds. The children were subsequently tested within two weeks of their eight-year birth date using these ERP procedures as well as a variety of language and cognitive measures, including the reading subtest of the WRAT-3 (mean = 97.66, *s.d.* = 12.6, range = 50-126), which was used to assess general reading performance.

Following artifact rejection for eye and muscle artifacts (rejection levels across infants < 15%), the ERPs were averaged by condition and electrode site. Next, baseline-to-peak amplitude (calculated from the average prestimulus period to a peak within the brain wave) as well as peak latency measures (calculated from stimulus onset to the maximum point of a peak within the brain wave) were calculated for three component peaks for each neonatal ERP. The peak measures served as the dependent measures in a discriminant function analysis to classify children's reading performance at eight years of age. These peaks included (a) the initial negative-positive shift in the ERP in the region from the first large negative peak (N_1, mean peak latency = 174.3 ms, *s.d.* = 31.2, mean baseline-to-peak amplitude = -2.4 μV, *s.d.* = 1.2 μV) to the following positive peak (P_2, mean peak latency = 308.7 ms, *s.d.* = 38.2, mean peak amplitude = 3.3 μV, *s.d.* = 1.2 μV); and (b) a second large negative peak (N_2, mean peak latency = 458.0 ms, *s.d.* = 32.8, mean peak amplitude = -3.5 μV, *s.d.* = 1.2 μV).

Six neonatal ERP responses were used to discriminate between the Control, Dyslexic, and Poor Reader groups at eight years of age. These variables included three amplitude and three latency measures: (a) the second large negative peak amplitude (N_2) recorded at the right-hemisphere frontal electrode site elicited in response to the /gi/ speech syllable; (b) the N_1 amplitude change recorded at the right-temporal hemisphere electrode site elicited in response to the /bi/ nonspeech syllable; (c) the second large positive peak amplitude (P_2) elicited in response to the /bi/ speech syllable; (d) the first large negative peak latency (N_1) to the speech syllable, /gi/, recorded at both the left hemisphere frontal and (e) parietal electrode sites, as well as at (f) the right temporal electrode. The effects generally were characterized by faster latencies for the Control children compared to the Dyslexic and Poor Reading groups as well as larger N_1 amplitudes for the Control infants, whereas the N_2 amplitudes were larger in the Dyslexic and Poor Reading groups. A larger P_2 amplitude, in contrast, characterized the Poor Reading group.

These six measures resulted in the identification of two significant canonical discriminant functions that correctly classified 81.25% of the entire sample (39 of 48 children) at eight years of age. Using the neonatal ERP measures, seven of seven Poor Readers were correctly classified (100%), 13 of 17 Dyslexic children were correctly classified (76.5%), and 19 of 24 of Control children (79.2%). These classifications are approximately two times greater than chance levels. If language interventions designed to prevent reading problems were attempted shortly after birth on the basis of these data, 22 of 24 children in need of intervention at eight years of age could have been targeted to receive intervention beginning at birth while only 5 of 24 children who did not require intervention would have received it. Thus, ERP measures shortly after birth demonstrate high accuracy (identifying nearly 92% of children in need of intervention by eight years) and generate relatively few false positives in predicting reading problems eight years later.

Example 2. In this example Molfese used a different set of consonant-vowel speech stimuli to elicit auditory ERPs — a series of nine consonant-vowel syllables composed of the consonants /b, d, g/ and the vowels /i, a, u/. The analysis approach also differed somewhat from Molfese (2000), in that he used the output of the PCA as the dependent measure in the ANOVA to discriminate between language abilities in these 102 children eight years later. He recorded 5,508 auditory event-related potentials (ERPs) from the left- and right-hemisphere frontal, temporal, and parietal scalp regions (linked ear references). These children were subsequently tested within two weeks of their eighth birth date using these ERP procedures as well as a variety of language and cognitive measures.

Following artifact rejection, the ERPs were averaged by condition and electrode site and submitted to the principal components analysis procedure, which employed a correlation matrix with varimax rotation. Seven factors describing variability in different ERP wave regions characterized 85% of the total variance in the ERP data set. These scores were input to a discriminant function procedure to classify three groups of children defined by their WRAT-3 reading subtest scores at eight years of age: (a) a Low group comprised responses from 11 children who performed one standard deviation below the mean (range = 50-82); (b) an Average group included 75 children who scored within one standard deviation of the population mean (range = 83-113); and (c) a High group included 16 children who scored one standard deviation above the mean (range = 114-126).

Six factor scores representing ERP discriminations of the consonants /b/, /d/, and /g/ at left-hemisphere parietal and temporal electrodes were entered into the discriminant function procedure from two factors, which matched the peak latencies that were previously identified by Molfese and Molfese (1985) in a different longitudinal sample of children. Two discriminant functions classified these children at eight years of age with 66.67% accuracy. Ten of 11 children (90.9%) were correctly classified as members of the Low group, 46 of 75 children (61.3%) were correctly classified as Average, and 12 of 16 children were correctly classified as High performers (75%). *Thus, neonatal brain responses discriminated individual children's level of performance at eight years of age on the WRAT-3 test of reading abilities.* Given that overall WRAT-3 reading scores at eight years were generally within the normal range, these newborn brain response data made relatively fine distinctions in predicting later reading skills. These data extend findings that previously reported strong relationships between neonatal speech discrimination and verbal performance measures at three, five, and eight years and indicate a strong relationship between newborn infants' ability to discriminate speech sounds and their reading level at eight years of age.

Such findings raise the hope that early identification may lead to early successful intervention for reading problems. If infants can be identified shortly after birth as "at risk" for later reading skills, interventions could be started much earlier than is now considered feasible. The additional time to address potential shortfalls in development, when coupled with the early plasticity that cognitive and linguistic systems appear to possess, could potentially lead to elimination of some types of reading disabilities.

In general, studies that restricted their analyses of the evoked potential response to a single early peak in the brain wave or to a specific peak latency (i.e., usually the N_1 component) achieved modest success in short-term prediction, but failed to find a long-term relationship between various measures of intelligence and the ERP waveform. In contrast, the more successful long-term prediction studies analyzed up to four peaks of the ERP waveform and recorded ERPs from scalp locations anterior to the occipital areas.

It is hoped that this review will serve to stimulate more research into critically important, yet largely uncharted area of early language acquisition using measures of the brain involvement in language processing.

REFERENCES

Allison, T., Wood, C. C., & McCarthy, G. M. (1986). The central nervous system. In M. G. H. Coles, E. Donchin, & S. W. Porges (Eds.), *Psychophysiology: Systems, processes, and applications* (pp. 5-25). New York: Guilford.

Blume, W. T., Buza, R. C., & Okazaki, H. (1974). Anatomic correlates of the ten-twenty electrode placement system in infants. *Electroencephalography and Clinical Neurophysiology, 36,* 303-307.

Brown, W. S., Marsh, J. T., & Smith, J. C. (1979). Principal component analysis of ERP differences related to the meaning of an ambiguous word. *Journal of Electroencephalography and Clinical Neurophysiology, 46,* 706-714.

Butler, B., & Engel, R. (1969). Mental and motor scores at 8 months in relation to neonatal photic responses. *Developmental Medicine and Child Neurology, 11,* 77-82.

Cacioppo, J. T., & Tassinary, L. G. (Eds.). (1990). *Principles of psychophysiology.* Cambridge: Cambridge University Press.

Callaway, C., Tueting, P., & Koslow, S. (1978). *Event-related brain potentials and behavior.* New York: Academic Press.

Caton, R. (1875). The electrical currents of the brain. *British Medical Journal, 2,* 278.

Chapman, R. M., McCrary, J. W., Bragdon, H. R., & Chapman, J. A. (1979). Latent components of event-related potentials functionally related to information processing. In J. E. Desmedt (Ed.), *Progress in clinical neuropsychology, Vol. 6: Cognitive components in cerebral event-related potentials and selective attention*. Basel: Karger.

Dawson, G. D. (1951). A summation technique for detecting small signals in a large irregular background. *Journal of Physiology, 119,* 2-3.

Donchin, E., Tueting, P., Ritter, W., Kutas, M., & Heffley, E. (1975). On the independence of the CNV and the P300 components of the human averaged evoked potential. *Journal of Electroencephalography and Clinical Neurophysiology, 38,* 449-461.

Engel, R., & Fay, W. (1972). Visual evoked responses at birth, verbal scores at three years, and IQ at four years. *Developmental Medicine and Child Neurology, 14,* 283-289.

Engel, R., & Henderson, N. (1973). Visual evoked responses and IQ scores at school age. *Developmental Medicine and Child Neurology, 15,* 136-145.

Gelfer, M. (1987). An AER study of stop-consonant discrimination. *Perception & Psychophysics, 42,* 318-327.

Henderson, N., & Engel, R. (1974). Neonatal visual evoked potentials as predictors of psychoeducational testing at age seven. *Developmental Psychology, 10,* 269-276.

Jasper, H. H. (1958). The ten-twenty electrode system of the International Federation of Societies for Electroencephalography: Appendix to report of the committee on methods of clinical examination in electroencephalography. *Journal of Electroencephalography and Clinical Neurophysiology, 10,* 371-375.

Jensen, D. R., & Engel, R. (1971). Statistical procedures for relating dichotomous responses to maturation and EEG measurements. *Electroencephalography and Clinical Neurophysiology, 30,* 437-443.

Junghöfer, M., Elbert, T., Tucker, D. M., & Braun, C. (1999). The polar average reference effect: A bias in estimating the head surface integral in EEG recording. *Clin. Neurophysiol., 110*(6), 1149-1155.

Junghöfer, M., Elbert, T., Tucker, D., & Rockstroh, B. (2000). Statistical control of artifacts in dense array EEG/MEG studies. *Psychophysiology, 37,* 523-532.

Molfese, D. L. (1978a). Left and right hemispheric involvement in speech perception: Electrophysiological correlates. *Perception & Psychophysics, 23,* 237-243.

Molfese, D. L. (1978b). Neuroelectrical correlates of categorical speech perception in adults. *Brain and Language, 5,* 25-35.

Molfese, D. L. (1983). Event related potentials and language processes. In A. W. K. Gaillard & W. Ritter (Eds.), *Tutorials in ERP research: Endogenous components* (pp. 345-368). Amsterdam, the Netherlands: North Holland Publishing Co.

Molfese, D. L. (in press). Newborn brain responses predict language development skills which emerge eight years later. *Brain and Language.*

Molfese, D. L. (2000). Predicting dyslexia at 8 years using newnatal brain responses. *Brain and Language, 72,* 238-245.

Molfese, D. L., Freeman, R., & Palermo, D. (1975). The ontogeny of latralization for speech and nonspeech stimuli. *Brain and Language, 2,* 356-368.

Molfese, D. L., & Molfese, V. J. (1979a). Hemisphere and stimulus differences as reflected in the cortical responses of newborn infants to speech stimuli. *Developmental Psychology, 15,* 505-511.

Molfese, D. L., & Molfese, V. J. (1979b). Infant speech perception: Learned or innate? In H. Whitaker & H. Whitaker (Eds.), *Advances in neurolinguistics: Vol. 4* (pp. 225-240). New York: Academic Press.

Molfese, D. L., & Molfese, V. J. (1980). Cortical responses of preterm infants to phonetic and nonphonetic speech stimuli. *Developmental Psychology, 16,* 574-581.

Molfese, D. L., & Molfese, V. J. (1985). Electrophysiological indices of auditory discrimination in newborn infants: The bases for predicting later language development. *Infant Behavior and Development, 8,* 197-211.

Molfese, D. L., & Molfese, V. J. (1986). Psychophysical indices of early cognitive processes and their relationship to language. In J.E. Obrzut & G.W. Hynd (Eds.), *Child neuropsychology: Vol 1. Theory and research* (pp. 95-115). New York: Academic Press.

Molfese, D. L., & Molfese, V. J. (1997). Discrimination of language skills at five years of age using event related potentials recorded at birth. *Developmental Neuropsychology, 13,* 135-156.

Molfese, D. L., Morse, P. A., & Peters, C. J. (1990). Auditory evoked responses from infants to names for different objects: Cross modal processing as a basis for early language acquisition. *Developmental Psychology, 26,* 780-795.

Molfese, D. L., Nunez, G., Seibert, S., & Ramanaiah, N. (1976). Changes in factors affecting differential hemispheric activity in infants. *Annals of the New York Academy of Sciences, 280,* 821-833.

Nelson, C. A., & Salapatek, P. (1986). Electrophysiological correlates of infant recognition memory. *Child Development, 57,* 1483-1497.

Rockstroh, B., Elbert, T., Birbaumer, N., & Lutzenberger, W. (1982). *Slow brain potentials and behavior.* Baltimore: Urban-Schwarzenberg.

Ruchkin, D., Sutton, S., Munson, R., & Macar, F. (1981). P300 and feedback provided by the absence of the stimuli. *Psychophysiology, 18,* 271-282.

Segalowitz, S., & Cohen, H. (1989). Right hemisphere EEG sensitivity to speech. *Brain and Language, 37,* 220 - 231.

Shaywitz, B. A., Shaywitz, S. E., Pugh, K. R., Constable, R. T., Skudlarski, P., Fulbright, R. K., Bronen, R. A., Fletcher, J. M., Shankweiler, D. P., Katz, L., & Gore, J. C. (1995). Sex differences in the functional organization of the brain for language. *Nature, 373*(6515), 607-609.

Tucker, D. M. (1991). Spatial sampling of head electrical fields: The geodesic sensor net. *Electroencephalography and Clinical Neurophysiology, 79,* 413-419.

Tucker, D. M., Liotti, M., Potts, G. F., Russell, G. S., & Posner, M. I. (1994). Spatiotemporal analysis of brain electrical fields. *Human Brain Mapping, 1,* 134-152.

Wechsler, D. (1991). *Wechsler Intelligence Scales for Children.* New York: The Psychological Corporation.

NOTES

This research was supported by Grant R01 HD17860 from the National Institute of Child Health and Human Development.

Please address correspondence to: Victoria J. Molfese, Center for Research in Early Childhood, School of Education, University of Louisville, Louisville, KY 40292.

Copyright © 2003 EBSCO Publishing